Impact of the numerical domain on turbulent flow statistics: scalings and considerations for canopy flows

Atharva S. Sathe[1], Marco G. Giometto[1] †

[1]Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY 10027

(Received xx; revised xx; accepted xx)

Large Eddy Simulations (LES) are widely used to study the effects of surface morphology on turbulence statistics, exchange processes, and turbulence topology in urban canopies. However, as LES are only approximations of reality, special attention is needed for the computational model setup to ensure an accurate representation of the physical processes of interest. This paper shows that the choice of the numerical domain can significantly impact the accuracy of turbulent flow statistics, potentially causing a mismatch between numerical studies and experimental data. The study examines the influence of cross-stream aspect ratio (YAR), streamwise aspect ratio (XAR), and scale separation (SS) on first and second-order flow statistics and turbulence topology. It is found that domains with a low YAR underestimate the velocity variance, while those with a low XAR overestimate the variance value. The study proposes a new approach based on the Buckingham Pi theorem to evaluate the effect of SS, as the existing method has major limitations for canopy flows. The results suggest that domains with small SS underpredict the variance value. To minimize the artificial impact of the numerical domain on turbulent flow statistics, the study recommends guidelines for future research, including a YAR of 3 or more, an XAR of 6 or more, and an SS of 12 or more. Error tables are presented to allow researchers to select smaller domains than recommended, depending on their research interests in specific parts of the flow.

Key words: Authors should not enter keywords on the manuscript, as these must be chosen by the author during the online submission process and will then be added during the typesetting process (see Keyword PDF for the full list). Other classifications will be added at the same time.

1. Introduction

The urbanization process profoundly impacts the urban boundary layer (UBL) due to impervious man-made structures that alter the aerodynamic and hydrothermal properties of the land surface. These changes affect mass, energy, and momentum transfer with the

† Email address for correspondence: mg3929@columbia.edu

overlying atmosphere, which are the main drivers of urban weather and climate variability. These exchange processes play a crucial role in applications related to urban climate (Oke *et al.* 2017; Oke 1982), urban ecohydrology (Meili *et al.* 2020), air quality (Fernando *et al.* 2001), urban resilience (Gorlé *et al.* 2015), and public health (Lowe & Forsberg 2011), to name a few. The interaction between the urban environment and atmospheric turbulence regulates these exchanges over a broad continuum of scales, ranging from tens of meters over the roof of a building to the kilometer scale over an urban neighborhood (Rotach 1993, 1999). Motivated by the need to address open challenges in these fields and improve our interaction with the environment, the past decades have seen significant efforts to advance our understanding and ability to model turbulent transport in urban settings.

Scientific discovery in the field of microscale meteorology has historically relied on three pillars: field observations (Rotach *et al.* 2005), wind-tunnel experiments (Barlow *et al.* 2004), and numerical simulations (Coceal *et al.* 2006). This paradigm has provided useful insight into how urban morphology affects flow statistics in the UBL, but the alignment between findings from these three fields is not always optimal. An instance of this is where a range of values for the von Kármán constant κ have been proposed by different field measurements and laboratory studies, with values varying from 0.33 to 0.43. This is comprehensively documented by Andreas *et al.* (2006). In addition, Philips *et al.* (2013) have pointed out several challenges in matching parameters of the underlying system, which hinder the accurate alignment of experimental data with numerical simulations. One such obstacle is the use of different methods to compute the repeating parameters, such as friction velocity, which cannot be uniformly applied across different fields. They also demonstrate that the vertical profile of the experimental data can often be accurately matched up to a certain height above the ground, beyond which significant deviations occur. This partial matching approach has also been utilized in other research studies, (see, e.g., Coceal *et al.* 2007; Xie *et al.* 2008), which serves to delimit the region of interest. Another factor contributing to the discrepancy between profiles is the sensitivity of flow statistics to changes in initial and boundary conditions and input parameters. This phenomenon often makes it challenging to establish connections between research findings within the same field (see, e.g., Wang *et al.* 2011).

In the context of numerical simulations, direct numerical simulations (DNS) and large eddy simulations (LES) of open channel flow over surface-mounted cuboids have been the workhorse for studying turbulent transport in the UBL (Coceal *et al.* 2006; Xie & Castro 2006; Leonardi & Castro 2010; Claus *et al.* 2012; Yang & Anderson 2017; Schmid *et al.* 2019; Stroh *et al.* 2020). In these simulations, in addition to the aforementioned sources of discrepancies, one crucial factor affecting the accuracy and reliability of model results is the selection of the numerical domain size (Moin & Kim 1982; Lozano-Durán & Jiménez 2014). Wall-bounded turbulence is characterized by coherent structures with a high correlation in the streamwise direction and a lower but still non-negligible correlation in the cross-stream direction. Thus, excessive periodization in the horizontal directions can compromise the accuracy with which these structures are captured (Moin & Kim 1982). Furthermore, in real-world environments, the scale separation between the inversion layer and the height of the canopies is often significant, and the presence of a free-slip top boundary condition too close to the surface may result in spurious effects encompassing the entire UBL. Hence, it is crucial to exercise caution during the simulation design stage to ensure the precise capturing of statistics in the region of interest.

Past DNSs and LESs have been conducted using a range of computational domains, whose size is typically dictated by the available computational resources (Coceal *et al.* 2006; Xie & Castro 2006; Stroh *et al.* 2020). To facilitate the comparison of the various domain sizes used, the concept of aspect ratio and scale separation is employed in this study. The naming

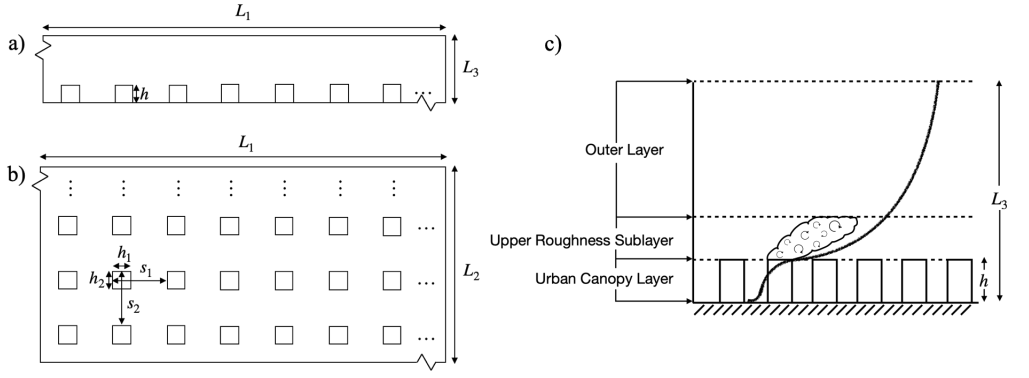

Figure 1: (*a*) Side view and (*b*) top view of the computational domain. (*c*) Marks the regions defined as Urban Canopy Layer (UCL), Upper Roughness Sublayer (URSL) and the Outer Layer (OL).

convention used to describe the dimensions of the computational domain is graphically illustrated in figure 1(a, b), with the subscripts 1, 2, and 3 referring to the streamwise, cross-stream, and vertical directions, respectively. The aspect ratio of a 3D computational domain is defined as L_1/L_3:L_2/L_3:1, where L_1/L_3 defines the streamwise aspect ratio (XAR) and L_2/L_3 defines the cross-stream aspect ratio (YAR). Additionally, the height of the domain is described in terms of the scale separation (SS), defined as L_3/h, where h is the mean height of the underlying surface topography.

One of the early DNS studies of flow over cuboids was performed by Coceal *et al.* (2006) to analyze turbulent flow statistics and unsteady effects in the roughness sublayer (RSL). This study represents a pivotal contribution to the understanding of canopy flow dynamics, achieved through the use of high-resolution DNS. However, as is common in such studies, the need for high resolution necessitated the selection of a smaller domain to ensure computational feasibility. For their open channel flow setup, they used a numerical domain with an aspect ratio of 1:1:1 with an SS of 4. To showcase domain size independence, they compared selected statistics with a domain of aspect ratio 2:2:1 and found the first-order statistics as well as second-order Reynolds stress $\overline{u_1' u_3'}$ to match well. However, it is well known that the profile of $\overline{u_1' u_3'}$ in the bulk of the flow is primarily determined by the imposed pressure gradient and has to vary linearly, as seen from the Navier-Stokes streamwise momentum balance equation; hence the accurate collapse of $\overline{u_1' u_3'}$ for domains with the same boundary layer height does not necessarily indicate the accurate capturing of other second-order moments. Also, as the focus of this study was on the canopy configurations with high packing density, the domain used cannot be deemed as sufficient for the shown statistics to study RSL dynamics in general, as the extent of the RSL, as well as the turbulence characteristics of the RSL depend on the underlying surface configuration (Chung *et al.* 2021). Xie & Castro (2006) performed LES simulations with domain 1:1:1 and SS of 4 and found that their simulations were underpredicting the streamwise RMS velocity (u_{rms}) when compared to corresponding DNS as well as experimental results. Later in this study (§3.1), it will be shown that this underprediction is due to a direct consequence of limiting YAR of the domain and not due to differences between LES and DNS algorithms. Leonardi & Castro (2010) used various domain sizes with SS of 8 and aspect ratios ranging from 1:0.75:1 to 1.25:1.25:1 using DNS. The choice of XARs and YARs was purely driven by the need to accommodate a sufficient number of repeating patterns for different configurations. Schmid *et al.* (2019) used a domain with SS of 4 and aspect ratio 1.5:1.5:1 to study the impact of solid

volume fraction on turbulent flow statistics using LES. Yang & Anderson (2017) used LES to analyze the physics of roughness-induced secondary flows by using domains with SS of 15 and 20 while keeping the aspect ratio of the domain as π:π:1. They showcased that domain with aspect ratio 2π:2π:1 produces similar results. However, this choice of high SS and high aspect ratio to reduce the artificial impacts of the numerical domain resulted in fewer nodes being used to resolve the cubes, which introduces an additional source of error. Stroh *et al.* (2020) used DNS to study the polarity of secondary flows by using a domain with an SS of 23.25 and an aspect ratio of 8:4:1. These studies demonstrate an apparent disparity in the employed domain sizes. From these observations, we infer the presence of a general trend towards maintaining a similar extent of the domain in both the streamwise and cross-stream directions. However, due to the asymmetrical nature of the turbulent flow structures and their extended presence in the streamwise direction compared to the cross-stream direction, it remains uncertain whether these domains will have an artificial impact on the flow statistics.

The presence of roughness-induced secondary flows, a topic which has received increased attention over the past decade (Willingham *et al.* 2014; Anderson *et al.* 2015; Vanderwel & Ganapathisubramani 2015; Yang & Anderson 2017; Chung *et al.* 2018; Wangsawijaya *et al.* 2020; Stroh *et al.* 2020; Salesky *et al.* 2022), also calls for special attention when designing the domain size. When the cross-stream spacing between the roughness elements is sufficiently large, it results in streamwise-aligned time-invariant counter-rotating vortices predominantly occupying the RSL. The size of these vortices is influenced by both the spacing of roughness elements in the cross-stream direction and the height of the domain. As demonstrated (see §3.3), these circulations significantly affect the flow dynamics and necessitate a specialized approach to evaluate the effect of SS, as the height of the domain plays a critical role in governing these flows.

In the context of channel flow over aerodynamically smooth surfaces, analysis done by Comte-Bellot (1963) and Schumann (1973) guided early numerical studies to determine the optimal domain size to reduce the artificial impact of periodic boundary condition in the horizontal directions (Moin & Kim 1982). Comte-Bellot (1963) conducted two-point correlation measurements of velocity fluctuations and found that the correlation became negligible at a separation of 3.2δ in the streamwise direction and 1.6δ in the cross-stream direction, where δ is the height of the half channel. Schumann (1973) and Moin & Kim (1982) later suggested that to reduce the artificial impact of periodic boundary conditions, the size of the simulation domain should be approximately twice as large as these dimensions. Lozano-Durán & Jiménez (2014) conducted an extensive domain size analysis for plane channel flow using DNS at $Re_\tau = 4200$. They showed that the computational box with aspect ratio 2π:π:1 was able to capture the one-point statistics with satisfactory accuracy. This aspect ratio of the domain aligns with the arguments provided by Schumann (1973) and Moin & Kim (1982). Zheng *et al.* (2021) conducted a series of LES to examine the impact of domain size on pollutant dispersion in street canyons with periodic boundary conditions applied only in the cross-stream direction. The study recommends a SS of 7.5 with a width of at least $0.33L_3$, an upstream domain length of $0.67L_3$, and a downstream domain length of $1.33L_3$. These guidelines, however, are based on the 2.5D geometry of cross-stream-aligned bars and cannot be generalized to LES simulations of open channel flow over cuboids or more general surface morphologies. As a result, there are currently no comprehensive guidelines for determining the appropriate size of the numerical domain for studying the UBL using an open channel flow setup with LES.

The appropriateness of the domain size also depends on the specific region of interest under investigation. In the existing literature, it is commonly observed that researchers prefer smaller domain sizes when focusing on regions close to the surface, as capturing accurate statistics for the entire domain is not always necessary (Anderson 2016; Zhang *et al.* 2022). In

this study, we introduce the Urban Canopy Layer (UCL), Upper Roughness Sublayer (URSL), and Outer Layer (OL) as illustrated in Figure 1(c) to facilitate the examination of flow statistics on a per-layer basis. Here, URSL is defined as a distinct component of the RSL, separate from the UCL, to avoid overlap when comparing flow statistics. Notably, we intentionally omit the inertial sublayer in our error analysis, as the study examines diverse packing densities and scale separations, where the presence of an inertial sublayer is not always guaranteed. We discuss this aspect in § 3.4. Hence, we incorporate the inertial sublayer, whenever present, in the OL for the purpose of our investigation.

This study investigates the impact of numerical domain size in these three distinct layers and addresses the aforementioned knowledge gap by providing extensive guidelines for researchers based on the packing density of the underlying configuration and the region of interest in a given study. The aim is to equip researchers with the essential data necessary for determining the optimal size of their numerical domain in LES simulations of UBL flows, thereby allowing them to predict any changes to their statistical profiles that may occur due to limitations in domain size.

The structure of paper is organized as follows: Section 2 outlines the methodology employed in this study, which includes the details of the simulation algorithm (§2.1) and the dimensional analysis and simulation setup (§2.2). The findings and observations from the simulations are presented in Section 3. Finally, Section 4 provides the conclusions drawn from the study.

2. Methodology

2.1. *Simulation algorithm*

A large suite of LES of flow over cuboid arrays is performed in this study using an in-house code (Albertson & Parlange 1999*a*,*b*; Bou-Zeid *et al.* 2005; Chamecki *et al.* 2009; Anderson *et al.* 2015; Fang & Porté-Agel 2015; Li *et al.* 2016; Giometto *et al.* 2016). The filtered Navier-Stokes equations are solved in their rotational form (Orszag & Pao 1975) to ensure the conservation of mass and kinetic energy in the inviscid limit, i.e.,

$$\frac{\partial \tilde{u}_i}{\partial x_i} = 0 \,, \tag{2.1}$$

$$\frac{\partial \tilde{u}_i}{\partial t} + \tilde{u}_j\left(\frac{\partial \tilde{u}_i}{\partial x_j} - \frac{\partial \tilde{u}_j}{\partial x_i}\right) = -\frac{1}{\rho}\frac{\partial \tilde{p}^*}{\partial x_i} - \frac{\partial \tau_{ij}^{SGS}}{\partial x_j} - \frac{1}{\rho}\frac{\partial \tilde{p}_\infty}{\partial x_1}\delta_{i1} + \tilde{F}_i \,, \tag{2.2}$$

where $\tilde{u}_1$, $\tilde{u}_2$, and $\tilde{u}_3$ are the filtered velocities along the streamwise x_1, cross-stream x_2, and wall-normal x_3 directions, respectively and ρ is the reference density. The deviatoric component of the subgrid-scale (SGS) stress tensor (τ_{ij}^{SGS}) is evaluated via the Lagrangian scale-dependent dynamic (LASD) Smagorinsky model (Bou-Zeid *et al.* 2005). Extensive validation of the LASD model has been carried out in both wall-modeled simulations of unsteady atmospheric boundary layer flow (Momen & Bou-Zeid 2017; Salesky *et al.* 2017) and in simulations of flow over surface-resolved urban-like canopies (Anderson *et al.* 2015; Li *et al.* 2016; Giometto *et al.* 2016; Yang 2016). Validation for the setup used in this study is also shown in appendix A. Viscous stresses are neglected in the current study, and the skin friction is evaluated via an inviscid equilibrium logarithmic law of the wall for flow over aerodynamically rough surfaces (Giometto *et al.* 2016). Neglecting viscous stresses is valid under the assumption that SGS stress contributions are predominantly from the pressure field. Here, $\tilde{p}^* = \tilde{p} + \frac{1}{3}\rho\tau_{ii}^{SGS} + \frac{1}{2}\rho\tilde{u}_i\tilde{u}_i$ is the modified pressure, which accounts for the trace of SGS stress and resolved turbulent kinetic energy. The flow is driven by

a spatially uniform pressure gradient. The magnitude of friction velocity u_τ is calculated based on imposed pressure gradient such that $(\nabla p/\rho)V_f = u_\tau^2 A_s$, where V_f is the volume of the fluid in the open channel and A_s is the surface area. This allows the friction velocity to be an input parameter for this study. While different definitions of friction velocities are employed in the literature (Tian *et al.* 2023), we use the pressure gradient based definition of friction velocity in this study given its widespread usage in the open channel flow literature (Bou-Zeid *et al.* 2004; Philips *et al.* 2013; Fang & Porté-Agel 2015; Yang & Anderson 2017; Stroh *et al.* 2020). The wall-parallel directions have periodic boundary conditions, while the upper boundary has free-slip boundary conditions, which can be expressed as $u_3 = 0, \partial u_1/\partial x_3 = 0$ and $\partial u_2/\partial x_3 = 0$. The lower surface represents an urban landscape with uniformly distributed cuboids. To resolve roughness elements, a discrete forcing immersed boundary method (IBM) is used (Mittal & Iaccarino 2005; Chester *et al.* 2007; Giometto *et al.* 2016), where an artificial force F_i is employed to bring the velocity to zero within the cuboids. An algebraic equilibrium wall-layer model, based on the law of the wall, is applied over a narrow band at the fluid-solid interface, i.e. on the surfaces of the cuboids, as well as on the solid base wall.

The spatial derivatives in the wall-parallel directions are computed by utilizing a pseudo-spectral collocation method that relies on truncated Fourier expansions (Orszag 1970). Conversely, in the wall-normal direction, a second-order staggered finite difference scheme is implemented. The time integration process involves the adoption of a second-order Adams-Bashforth scheme. To deal with non-linear advection terms, the 3/2 rule is utilized for de-aliasing (Canuto *et al.* 2007; Margairaz *et al.* 2018). Additionally, to ensure the enforcement of the incompressibility condition (2.1), a fraction step method (Kim & Moin 1985) is employed. The simulations are run for $200T$, where T is the large eddy turnover time defined as $T = L_3/u_\tau$ to ensure temporal convergence of first and second-order statistics. The time step employed in these simulations is selected to maintain a Courant-Friedrichs-Lewy (CFL) number below 0.1, ensuring numerical stability.

A large number of domain sizes are considered to study the impact of YAR, XAR, and SS. The size of the computational domain is $[0, L_1] \times [0, L_2] \times [0, L_3]$, with L_3/h taking values $\{4, 8, 12, 16\}$. Here h is the height of cuboids, kept constant and equal to 1 across all simulations. L_2/L_3 takes values $\{1.5, 3.0, 4.5, 6.0\}$ while L_1/L_3 takes values $\{3.0, 6.0, 9.0, 18.0, 27.0\}$. An aerodynamic roughness length of $z_0 = 10^{-6}h$ is prescribed at the cube surfaces and the lower surface via the wall-layer model. With the chosen value of z_0, the SGS pressure drag is a negligible contributor to the overall momentum balance (Yang & Meneveau 2016). The flow is in fully rough aerodynamic regime with a roughness Reynolds number $Re_\tau \equiv u_\tau h/\nu = 10^5$. The domain is discretized using a uniform Cartesian grid where each cube is resolved using $n_1 \times n_2 \times n_3 = 4 \times 4 \times 8$ for cases shown in table 2 and 3, and $n_1 \times n_2 \times n_3 = 6 \times 6 \times 12$ for cases shown in table 4 and 5, where n_i denotes the number of collocation nodes per cube edge. In the case of lowest packing density accompanied by pronounced secondary flows, we have noted that the numerical instability of the compressed grid, necessary for boundary layer height-based scaling, has led to the generation of unrealistic flow patterns. Consequently, in our simulations involving packing density of 0.007 and SS of 4 and 8, we have utilized a grid resolution of $n_1 \times n_2 \times n_3 = 4 \times 4 \times 8$, as detailed in table 5. The error attributed to grid compression was notably lower compared to the errors observed in the statistics, thereby validating the appropriateness of employing grid compression for the analysis of cases scaled with boundary layer height-based scaling. The chosen grid resolution ensures that the study is computationally feasible while providing adequate resolution to capture the flow dynamics with large domains. The analysis presented in appendix A and B shows that the chosen grid resolution yields flow statistics that are accurate up to second-order moments, based on the scope of this study.

Variable	Name	Variable	Name
U	mean streamwise velocity	h_2	cross-stream width of cuboids
u_τ	friction velocity	h_1	streamwise width of cuboids
L_2	cross-stream extent of the domain	s_2	cross-stream length of repeating unit
L_1	streamwise extent of the domain	s_1	streamwise length of repeating unit
L_3	boundary layer height	x_3	wall-normal distance
h	vertical height of cuboids		

Table 1: Variables determining flow characteristics for open channel flow simulations of flow over cuboids.

2.2. *Dimensional Analysis and Setup of Simulations*

This section discusses the setup of simulations and scaling arguments for flow statistics based on a Buckingham Pi theorem rationale. As mentioned in the Introduction, the study aims to analyze the impact of domain geometry on flow statistics, with a lens on the YAR (L_2/L_3), XAR (L_1/L_3), and SS (L_3/h) parameters. To achieve this objective, a suite of LES of flow over cuboid arrays is conducted, programmatically varying input parameters for the problem. Table 1 shows the quantities governing flow statistics; these quantities encompass two fundamental dimensions, length L and time T, so the considered flow system can be completely characterized by a total of $11 - 2 = 9$ Pi groups (Buckingham 1914).

Based on the choice of repeating parameters, two different scaling relations can be obtained for the flow statistics. The merits and limitations of each are discussed in the next sections.

2.2.1. *Canopy length-based scaling*

In the canopy length-based scaling, the vertical height of cuboids (h) and friction velocity (u_τ) are chosen as repeating parameters. While all length scales are normalized by h, special considerations are needed for L_1 and L_2 as the flow structures in the OL scale with the boundary layer height. By combining Pi groups, L_1 and L_2 can be scaled appropriately with L_3. Therefore, for example, the normalized mean streamwise velocity can be written in terms of non-dimensional groups as:

$$U/u_\tau = f\left(\frac{L_3}{h}, \frac{L_2}{L_3}, \frac{L_1}{L_3}, \frac{h_2}{h}, \frac{h_1}{h}, \frac{s_2}{h}, \frac{s_1}{h}, \frac{x_3}{h}\right) . \tag{2.3}$$

In order to study the impact of YAR (L_2/L_3) on the non-dimensional mean streamwise velocity, the set of simulations in table 2 are chosen where for a particular packing density, only the non-dimensional group L_2/L_3 is varied across cases. This variation is achieved by varying the cross-stream length of the domain L_2 while keeping the boundary layer height L_3 constant. In order to minimize the impact of SS, the largest available value of L_3 is chosen. All the simulations have $h_2/h = h_1/h = 1$.

A similar analysis is carried out to study the impact of XAR using the set of simulations in table 3. The variation in L_1/L_3 is achieved by varying L_1 while keeping L_3 constant. Again, the largest value of SS (L_3/h) is chosen to minimize the impact of the blockage effect. While the largest L_2/L_3 among the available values is chosen for domains with $L_1/L_3 = 6$ and 9, $L_2/L_3 = 3.0$ is chosen for cases with $L_1/L_3 = 3$, since 3:3:1 is a very common aspect ratio of the domain found in canopy flow literature.

To study the impact of SS on flow statistics, set of simulations in table 4 are chosen where

	Cross-stream aspect ratio cases					Resolution
λ_p	L_3/h	L_2/L_3	L_1/L_3	s_2/h	s_1/h	$N_1 \times N_2 \times N_3$
0.25	16	1.5	6	2	2	$384 \times 96 \times 128$
0.25	16	3.0	6	2	2	$384 \times 192 \times 128$
0.25	16	4.5	6	2	2	$384 \times 288 \times 128$
0.062	16	1.5	6	4	4	$384 \times 96 \times 128$
0.062	16	3.0	6	4	4	$384 \times 192 \times 128$
0.062	16	4.5	6	4	4	$384 \times 288 \times 128$
0.062	16	6.0	6	4	4	$384 \times 384 \times 128$
0.028	16	1.5	6	6	6	$384 \times 96 \times 128$
0.028	16	3.0	6	6	6	$384 \times 192 \times 128$
0.028	16	4.5	6	6	6	$384 \times 288 \times 128$
0.028	16	6.0	6	6	6	$384 \times 384 \times 128$
0.007	16	1.5	6	12	12	$384 \times 96 \times 128$
0.007	16	3.0	6	12	12	$384 \times 192 \times 128$
0.007	16	4.5	6	12	12	$384 \times 288 \times 128$

Table 2: Set of simulations to study the impact of cross-stream aspect ratio of the numerical domain on flow statistics. The Pi groups are mentioned in the table based on (2.3). For all the simulations, $h_2/h = h_1/h = 1$.

	Streamwise aspect ratio cases					Resolution
λ_p	L_3/h	L_2/L_3	L_1/L_3	s_2/h	s_1/h	$N_1 \times N_2 \times N_3$
0.25	16	3.0	3	2	2	$192 \times 192 \times 128$
0.25	16	4.5	6	2	2	$384 \times 288 \times 128$
0.25	16	4.5	9	2	2	$576 \times 288 \times 128$
0.25	16	4.5	18	2	2	$1152 \times 288 \times 128$
0.25	16	4.5	27	2	2	$1728 \times 288 \times 128$
0.062	16	3.0	3	4	4	$192 \times 192 \times 128$
0.062	16	4.5	6	4	4	$384 \times 288 \times 128$
0.062	16	4.5	9	4	4	$576 \times 288 \times 128$
0.062	16	4.5	18	4	4	$1152 \times 288 \times 128$
0.062	16	4.5	27	4	4	$1728 \times 288 \times 128$
0.028	16	3.0	3	6	6	$192 \times 192 \times 128$
0.028	16	4.5	6	6	6	$384 \times 288 \times 128$
0.028	16	4.5	9	6	6	$576 \times 288 \times 128$
0.028	16	4.5	18	6	6	$1152 \times 288 \times 128$
0.028	16	4.5	27	6	6	$1728 \times 288 \times 128$

Table 3: Set of simulations to study the impact of streamwise aspect ratio of the numerical domain on flow statistics. The Pi groups are mentioned in the table based on (2.3). For all the simulations, $h_2/h = h_1/h = 1$.

	Dimensionless groups for h scaling					Resolution
λ_p	L_3/h	L_2/L_3	L_1/L_3	s_1/h	s_2/h	$N_1 \times N_2 \times N_3$
0.25	4	3.0	6	2	2	$144 \times 72 \times 48$
0.25	8	3.0	6	2	2	$288 \times 144 \times 96$
0.25	12	3.0	6	2	2	$432 \times 216 \times 144$
0.25	16	3.0	6	2	2	$576 \times 288 \times 192$
0.062	4	3.0	6	4	4	$144 \times 72 \times 48$
0.062	8	3.0	6	4	4	$288 \times 144 \times 96$
0.062	12	3.0	6	4	4	$432 \times 216 \times 144$
0.062	16	3.0	6	4	4	$576 \times 288 \times 192$
0.062	24	3.0	6	4	4	$864 \times 432 \times 288$
0.028	4	3.0	6	6	6	$144 \times 72 \times 48$
0.028	8	3.0	6	6	6	$288 \times 144 \times 96$
0.028	12	3.0	6	6	6	$432 \times 216 \times 144$
0.028	16	3.0	6	6	6	$576 \times 288 \times 192$
0.028	24	3.0	6	6	6	$864 \times 432 \times 288$
0.007	4	3.0	6	12	12	$144 \times 72 \times 48$
0.007	8	3.0	6	12	12	$288 \times 144 \times 96$
0.007	12	3.0	6	12	12	$432 \times 216 \times 144$
0.007	16	3.0	6	12	12	$576 \times 288 \times 192$

Table 4: Set of simulations to study the impact of scale separation L_3/h of the numerical domain on flow statistics using canopy length-based scaling. The Pi groups are mentioned in the table based on (2.3). For all the simulations, $h_2/h = h_1/h = 1$.

for a particular packing density, only L_3/h is varied across cases. This variation in L_3/h was achieved by varying the boundary layer height L_3 while keeping the canopy height h constant. It is later shown that $L_2/L_3 = 3$ and $L_1/L_3 = 6$ are large enough such that they do not artificially alter the flow statistics. Hence, these values are chosen while varying the SS.

2.2.2. *Boundary layer height-based scaling*

In boundary layer height-based scaling, the boundary layer height (L_3) and friction velocity (u_τ) are chosen as repeating parameters. While all length scales are normalized by L_3, special considerations are needed for h_1 and s_1. As the displacement distance is determined by the extent to which flow can penetrate the canopy layer, the parameter is significantly influenced by the height of the roughness element (h), gaps between two elements in the streamwise direction (s_1) and the portion of the gap occupied by the roughness element (h_1). Thus, to preserve the displacement distance, it's more appropriate to scale s_1 and h_1 with canopy height h, which can be achieved from a combination of the new set of Pi groups. Also, the normalized parameter h/L_3 can be inverted to have a consistent SS definition throughout the paper.

Therefore, for example, the normalized streamwise velocity can be written in terms of non-dimensional groups as:

$$U/u_\tau = f\left(\frac{L_3}{h}, \frac{L_2}{L_3}, \frac{L_1}{L_3}, \frac{h_2}{L_3}, \frac{h_1}{h}, \frac{s_2}{L_3}, \frac{s_1}{h}, \frac{x_3}{L_3}\right). \tag{2.4}$$

One may also choose to normalize h_1 with s_1 and h_2 with s_2 to preserve the extent of

	Dimensionless groups for L_3 scaling							Resolution
λ_p	L_3/h	L_2/L_3	L_1/L_3	h_1/h	h_2/L_3	s_1/h	s_2/L_3	$N_1 \times N_2 \times N_3$
0.25	4	3.0	6	1	0.0625	2	0.125	$144 \times 288 \times 48$
0.25	8	3.0	6	1	0.0625	2	0.125	$288 \times 288 \times 96$
0.25	12	3.0	6	1	0.0625	2	0.125	$432 \times 288 \times 144$
0.25	16	3.0	6	1	0.0625	2	0.125	$576 \times 288 \times 192$
0.062	4	3.0	6	1	0.0625	4	0.25	$144 \times 288 \times 48$
0.062	8	3.0	6	1	0.0625	4	0.25	$288 \times 288 \times 96$
0.062	12	3.0	6	1	0.0625	4	0.25	$432 \times 288 \times 144$
0.062	16	3.0	6	1	0.0625	4	0.25	$576 \times 288 \times 192$
0.028	4	3.0	6	1	0.0625	6	0.375	$144 \times 288 \times 48$
0.028	8	3.0	6	1	0.0625	6	0.375	$288 \times 288 \times 96$
0.028	12	3.0	6	1	0.0625	6	0.375	$432 \times 288 \times 144$
0.028	16	3.0	6	1	0.0625	6	0.375	$576 \times 288 \times 192$
0.007	4	3.0	6	1	0.0625	12	0.75	$96 \times 192 \times 32$
0.007	8	3.0	6	1	0.0625	12	0.75	$192 \times 192 \times 64$
0.007	12	3.0	6	1	0.0625	12	0.75	$432 \times 288 \times 144$
0.007	16	3.0	6	1	0.0625	12	0.75	$576 \times 288 \times 192$

Table 5: Set of simulations to study the impact of scale separation L_3/h of the numerical domain on flow statistics using boundary layer height-based scaling. The Pi groups are mentioned in the table based on (2.4).

roughness element in the repeating unit. The Pi groups presented in (2.4) ensure that the pairs (h_1, s_1) and (h_2, s_2) are normalized by the same length scale, h and L_3 respectively. This automatically preserves h_1/s_1 and h_2/s_2 across cases, eliminating the need to modify these Pi groups further.

To study the impact of SS on flow statistics, a new set of simulations is proposed in table 5 based on boundary layer height-based scaling. Variation in L_3/h is achieved similarly by varying the boundary layer height L_3 while keeping the canopy height h constant. For the cases with $L_3/h = 16$, surface geometry contains regularly arranged cubes. However, in order to preserve h_2/L_3 across different SSs, the cross-stream extent of the cuboids h_2 must be adjusted, which results in distortion of the cube geometry. Therefore, as we decrease the domain height, the cuboids become slender in the cross-stream direction, while the streamwise extent of the cuboid remains the same, as it scales with the canopy height h. The motivation for implementing this scaling technique arises from the inadequacies of traditional canopy length-based scaling for canopy flows, which fails to isolate the effects of SS accurately. This alternative approach provides more precise isolation of SS impacts across all packing densities, as explained in § 3.3 and shown in figure 11.

3. Results and observations

This section examines the impact YAR, XAR, and SS on selected turbulent flow statistics. Statistics are discussed on a per-layer basis for the three layers depicted in figure 1(c). To estimate the height of the RSL (x_{3r}), we utilize a formula proposed by Chung *et al.* (2021),

i.e.,

$$x_{3r} = \frac{s_2}{2} + d\,, \tag{3.1}$$

where d is the aerodynamic displacement height of the given surface. Values for d are chosen from the values reported for square configurations in Kanda *et al.* (2004). This estimate is useful in predicting the extent of the RSL *a priori*; however, it tends to overestimate the height of RSL for densely packed configurations. For the purpose of our study, such shifts in the prediction of the extent of RSL have no significant impact on the error magnitudes, thus justifying the use of (3.1). Additionally, an analysis of the existence of an inertial sublayer is also presented in this section for cases with varying scale separations and packing densities.

In this study, the operation of time-averaging is denoted by $\overline{(\cdot)}$, while the process of spatial averaging in the horizontal directions is denoted by $\langle\cdot\rangle$. The averaging operation in the UCL is defined as a superficial average, where the flow statistics are normalized by the total volume, which includes the solid canopy elements (Schmid *et al.* 2019). A fluctuation from space and time-averaged quantity is denoted by the symbol $(\cdot)'$. It is important to note that all the second-order statistics under discussion are computed using the resolved portion of the flow field. The present study does not include a detailed examination of SGS stresses. This decision is based on their limited contribution, comprising less than 2% of the total Reynolds stress (Tian *et al.* 2023). Moreover, given that SGS stresses arise from small-scale motions, it is anticipated that the influence of domain boundary conditions on these stresses will be of negligible significance.

3.1. *Impact of cross-stream aspect ratio*

This section discusses the impact of YAR on first and second-order flow statistics as well as on the structure of turbulence through two-point correlation maps. To investigate the influence of YAR, simulations were conducted using three YAR values: 1.5, 3.0, and 4.5. These simulations were performed at four different packing densities, as outlined in table 2. Additionally, for packing densities of 0.062 and 0.028, simulations were carried out with a YAR value of 6.0. The discrepancy between the profiles obtained with YAR 4.5 and YAR 6.0, concerning the first and second-order statistics considered in this subsection, does not exceed 1% across all layers. This satisfactory agreement between the results obtained using YAR 4.5 and YAR 6.0 suggests that the data derived from YAR 4.5 can confidently serve as the ground truth for the subsequent analysis presented in this subsection. Consequently, the following analysis exclusively focuses on the cases corresponding to YAR values of 1.5, 3.0, and 4.5.

Figure 2 shows profiles of mean streamwise velocity for different YAR values and packing densities. Differences in the profiles can be solely attributed to the artificial effects of the cross-stream width of the domain. Table 6 shows the error norms in different parts of the boundary layer, which are also visualized in figure 4. The results indicate that the velocity profile of the narrow domain (i.e. YAR 1.5) can estimate this quantity within 2% error when compared to the velocity profile of the largest domain across all the layers and all the packing densities. Marginal improvements are seen in the error magnitudes when YAR is increased to 3.0.

Figure 3 shows profiles of resolved mean streamwise variance for the same cases considered in figure 2 and errors in the different parts of the boundary layer are shown in table 6. It is observed that in UCL and URSL, the narrow domain is capable of predicting the resolved variance within 10% of the largest domain, except for the case with packing density of 0.028, where the narrow domain results in a noticeable deviation in URSL, leading to an error of 17%. In the OL, the error in this quantity exceeds 14% for all cases except for the

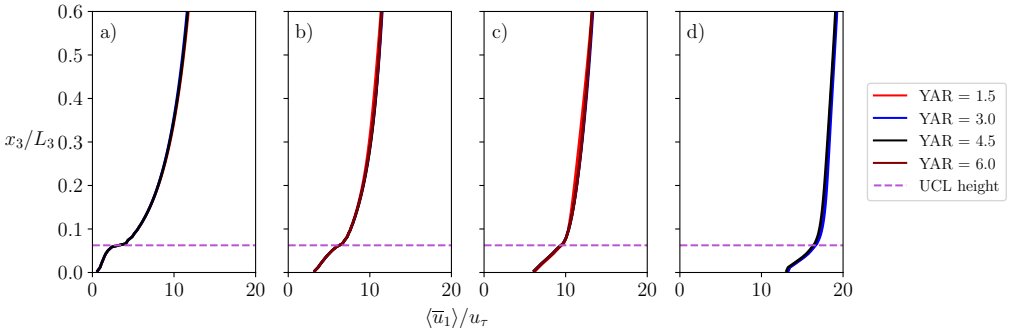

Figure 2: Mean streamwise velocity profiles for different packing densities, (a) 0.25, (b) 0.062, (c) 0.028, (d) 0.007. The vertical profiles for each packing density correspond to different cross-stream aspect ratio cases mentioned in table 2.

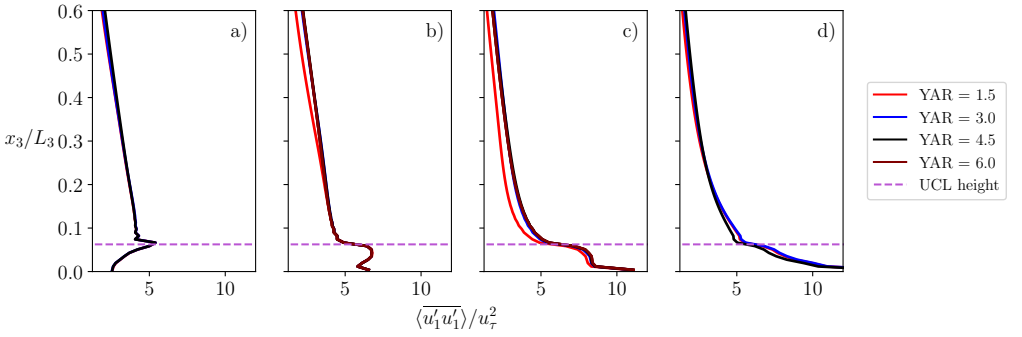

Figure 3: Resolved mean streamwise variance profiles for different packing densities, (a) 0.25, (b) 0.062, (c) 0.028, (d) 0.007. The vertical profiles for each packing density correspond to different cross-stream aspect ratio cases mentioned in table 2.

		$\langle \overline{u}_1 \rangle / u_\tau$			$\langle \overline{u_1' u_1'} \rangle / u_\tau^2$			L_{22}		
λ_p	YAR	UCL	URSL	OL	UCL	URSL	OL	UCL	URSL	OL
0.25	1.5	0.0036	0.0032	0.0116	0.0060	0.0043	0.0543	0.0261	0.0528	0.2048
0.25	3.0	0.0012	0.0004	0.0050	0.0008	0.0031	0.0388	0.0065	0.0163	0.0605
0.062	1.5	0.0014	0.0017	0.0159	0.0036	0.0066	0.1433	0.0414	0.0549	0.3490
0.062	3.0	0.0008	0.0010	0.0094	0.0013	0.0013	0.0203	0.0169	0.0063	0.0787
0.028	1.5	0.0192	0.0165	0.0165	0.0417	0.1722	0.2069	0.0428	0.0325	0.2574
0.028	3.0	0.0020	0.0011	0.0061	0.0088	0.0287	0.0263	0.0067	0.0088	0.0514
0.007	1.5	0.0136	0.0094	0.0105	0.0597	0.0563	0.1382	0.0514	0.0647	0.2187
0.007	3.0	0.0157	0.0159	0.0104	0.0721	0.0612	0.0522	0.0622	0.0243	0.0336

Table 6: Relative error (l_2 norm) of mean streamwise velocity, resolved mean streamwise variance and resolved transverse integral length scale in urban canopy layer (UCL), upper roughness sublayer (URSL) and outer layer (OL) for simulations with different cross-stream aspect ratios (YAR). Results from the largest domain (YAR 4.5) are considered as ground truths.

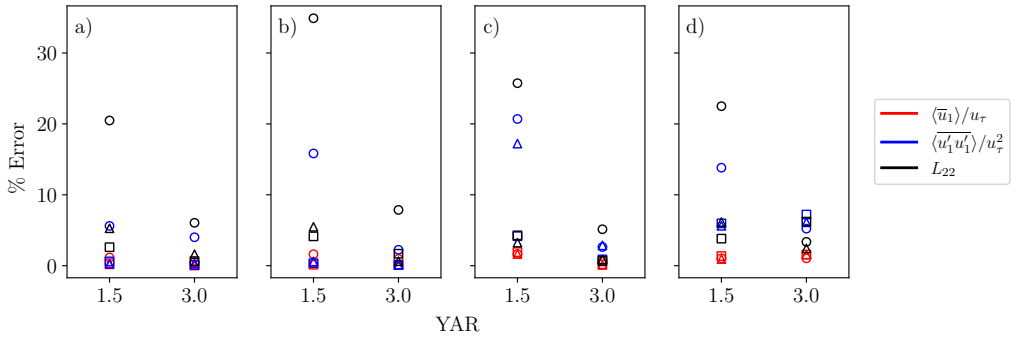

Figure 4: Visualization of error values in table 6 for different packing densities, (a) 0.25, (b) 0.062, (c) 0.028, (d) 0.007. Error values for different layers are represented by distinct symbols: Square - Urban canopy layer (UCL), triangle - Upper roughness sublayer (URSL), circle - Outer layer (OL).

densely packed case, for which the error remains within 6%. This observed error can be attributed to the tendency of the narrow domain to underestimate the value of variance. In contrast, the domain with YAR 3.0 can predict this quantity with an error magnitude that is approximately 7% or lower when compared to the profiles of the largest domain across all the layers and all the packing densities, indicating a reduced influence of artificial periodization in the cross-stream direction. This also indicates that the periodic boundary condition in the cross-stream direction has a lesser impact on the first-order statistics compared to the second-order statistics. In order to investigate the underlying cause of the observed statistical shifts in the narrow domain, we now use two-point correlation to assess the impact of restricting cross-stream width of the domain on the topology of turbulence.

Figure 5 shows two-point correlation (R_{11}) contours and instantaneous flow field fluctuations for different YAR at $x_3/L_3 = 0.6$. For brevity, only the cases with packing density of 0.028 are shown here. This packing density is chosen to qualitatively assess the reason behind the narrow domain noticeably underpredicting the resolved mean streamwise variance, as seen in figure 3(c). The color bar is not shown here as the values are not used for inference; however, it is kept constant for all the flow field visualizations to get an appropriate sense of fast (red) and slow (blue) turbulent streaks. The two-point correlation between any two quantities is defined as

$$R_{\alpha\beta}(\Delta x_1, \Delta x_2, x_3) = \frac{\overline{u'_\alpha(x_1, x_2, x_3) u'_\beta(x_1 + \Delta x_1, x_2 + \Delta x_2, x_3)}}{\sigma_{u_\alpha} \sigma_{u_\beta}} \tag{3.2}$$

where σ_{u_α} is the standard deviation of the resolved fluctuating field u'_α. It is important to note that the presence of repeated indices in this context does not denote summation. From figure 5(a, c, e), we see that the streamwise extent of correlation for the narrow domain is much smaller compared to cases with YAR 3.0 and 4.5. This observation is strongly supported by the resolved streamwise instantaneous flow field fluctuations shown in figure 5(b, d, f). For the cases with YAR 3.0 and 4.5, we observe long streamwise turbulent structures of the order of the corresponding domain extent, justifying a more significant streamwise correlation. However, as shown in figure 5(b), no such structures are observed for the case with YAR 1.5. This shows that the narrow cross-stream width of the domain can significantly alter the growth of turbulent flow structures in the streamwise direction.

As these coherent structures scale with the separation distance from the wall and as figure 5 only illustrates the case where $x_3/L_3 = 0.6$, a more detailed analysis is needed to comment

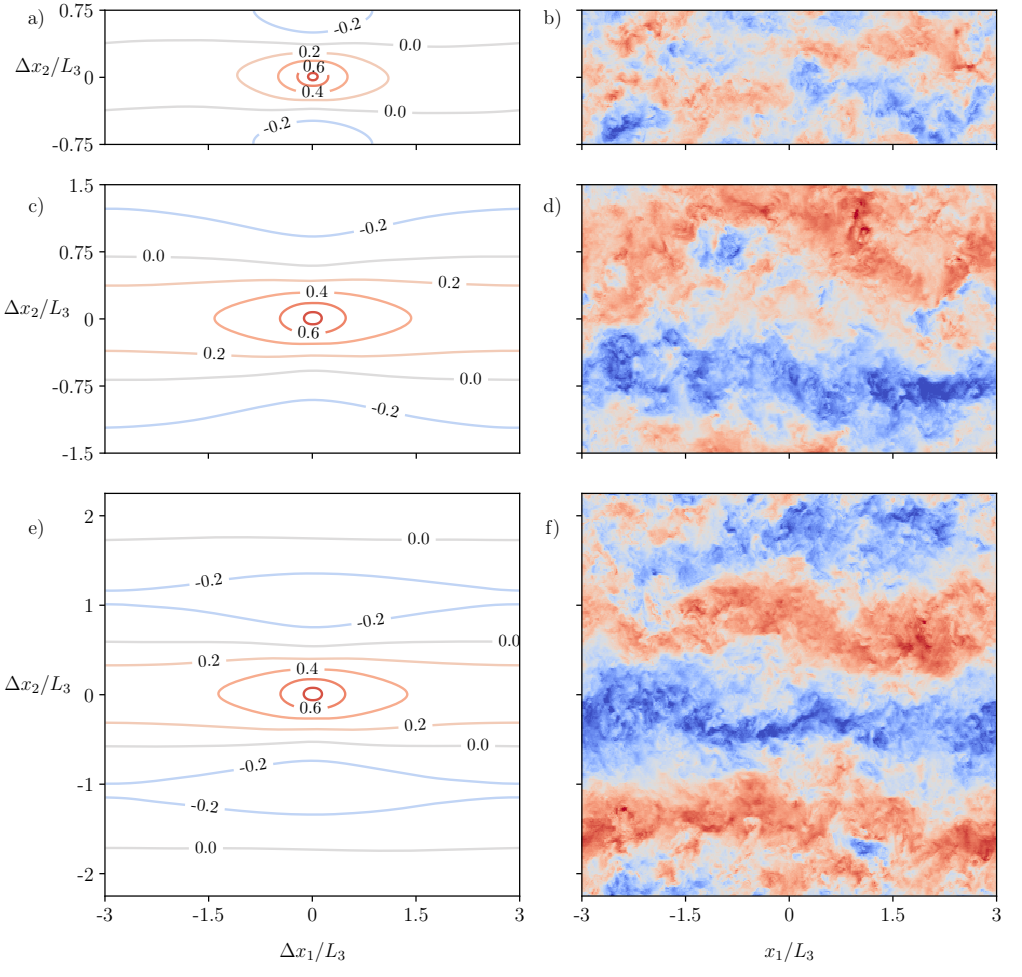

Figure 5: Two point correlation R_{11} contours (a, c, e) and streamwise resolved instantaneous flow field fluctuations (b, d, f) for cases with packing density 0.028. The cross-stream aspect ratio is varied as: (a, b) 1.5, (c, d) 3.0, (e, f) 4.5. The wall-parallel slice shown in all the figures is taken at $x_3/L_3 = 0.6$.

on the suitability of the domain with YAR 1.5 to accommodate a pair of these structures at different vertical positions and across all packing densities (Tomkins & Adrian 2003; Ganapathisubramani *et al.* 2005; Coceal *et al.* 2007). To address this matter, we analyze the typical width of such structures and investigate the ability of the domain with YAR 1.5 to accommodate fast and slow turbulent streaks at different vertical locations.

Figure 6 shows the total width of a fast and slow streak pair, which were observed in figure 5(d, f), as a function of height for cases with YAR 3.0 and 4.5. The width of a structure is computed as twice the cross-stream width over which R_{11} drops from 1 to 0. This width is then doubled to get the total width of the fast and slow streak pair. Figure 6 shows that as the size of streamwise coherent structure increases with height, the domain with YAR 1.5 is not sufficient to accommodate a pair of fast and slow streaks at $x_3/L_3 = 0.6$. This explains why no streamwise coherence was observed in figure 5(b). We also see that till $x_3/L_3 \approx 0.8$, the domain with YAR 3.0 is sufficient to accommodate a fast and slow streak pair even as the cross-stream extent of the domain is increased to YAR 4.5. A rapid increase in the structure

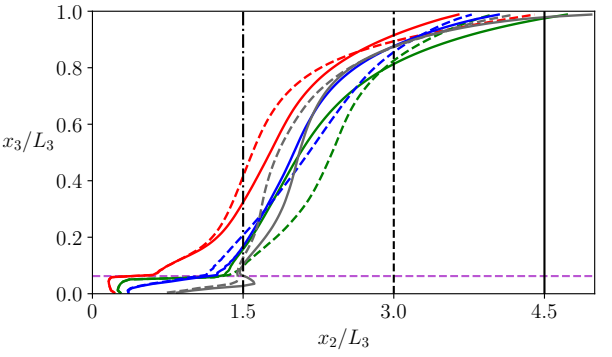

Figure 6: Width of turbulent streamwise coherent structures consisting of fast and slow streak pair. Cases with cross-stream aspect ratio (YAR) of 4.5 are shown in solid lines and YAR of 3.0 in dashed lines. Colors correspond to different packing densities: 0.25 - red, 0.062 - green, 0.028 - blue, 0.007 - grey. Black vertical lines indicate the width of the domain: dash-dot - YAR 1.5, dashed - YAR 3.0, solid - YAR 4.5. Purple horizontal line (dashed) indicates height of the urban canopy layer.

size is observed beyond $x_3/L_3 \approx 0.8$ due to the free-slip boundary condition applied at the top of the computational domain, as it inhibits the inclined growth of the structures, conforming them to a planar configuration (Ganapathisubramani *et al.* 2005). Since canopy flow studies in the open-channel flow setup do not typically focus on this region of the boundary layer, YAR 3.0 can be considered good enough to capture these coherent structures in the region below $x_3/L_3 \approx 0.8$. A noticeable deviation can be seen in the width of streamwise coherent structures between cases with YAR 3.0 and 4.5. However, as seen from figure 2, figure 3 and table 6, the impact of this deviation does not significantly alter the first and second-order statistics. From figure 6 we also see that the vertical locations at which the width of the fast and slow streak pair exceeds the width of the domain with YAR 1.5 is different for different packing densities. For the case with highest packing density (i.e. 0.25), the crossing point lies at $x_3/L_3 \approx 0.37$. For packing densities 0.062 and 0.028, the crossing point lies at $x_3/L_3 \approx 0.15$, whereas this value is $x_3/L_3 \approx 0.07$ for packing density 0.007. Although these structures are seen to be increasing at a similar rate across all packing densities, the different vertical locations of these crossing points are a result of differences in the width of these structures near the top of the canopy layer. As observed by Coceal *et al.* (2007), the size of these structures near the canopy top is influenced by the geometry of obstacles, and their potential for growth depends on the configuration of said obstacles. This explains why different error magnitudes were observed in figure 3 across different packing densities for YAR 1.5, as the same domain width may or may not be able to accommodate these structures at a particular height based on the underlying surface configuration.

So far, the analysis has shown that insufficient cross-stream width of a numerical domain can inhibit the growth of streamwise coherent structures. To analyze the impact of YAR on the cross-stream coherent structures, resolved transverse integral length scale L_{22} is shown in figure 7 as a function of height. Errors in the profiles in different parts of the boundary layer are shown in table 6. The integral length scale in this study is defined as

$$L_{\alpha\alpha}(x_3) = \int_0^\infty R_{\alpha\alpha}(\Delta x_1 \delta_{\alpha 1}, \Delta x_2 \delta_{\alpha 2}, x_3) d\Delta x_\alpha \tag{3.3}$$

Thus, L_{22} characterizes the length of instantaneous flow structures in the cross-stream direction. Note that the presence of repeated indices in this context does not imply summation.

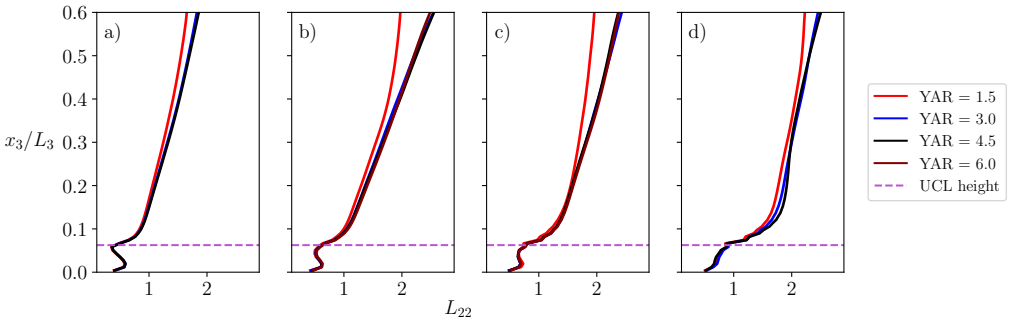

Figure 7: Resolved transverse integral length scale for different packing densities, (a) 0.25, (b) 0.062, (c) 0.028, (d) 0.007. The vertical profiles for each packing density correspond to different cross-stream aspect ratio cases mentioned in table 2.

To discard the noise present around the correlation value 0, a cutoff value of 0.2 is used to compute the resolved transverse integral length scale (Ganapathisubramani *et al.* 2005). The profile of the narrow domain in the OL exhibits significant deviation across all packing densities, as shown in figure 7, with errors exceeding 20% in all cases. Whereas in UCL and URSL, a maximum of 7% error is observed for the narrow domain. In contrast, the domain with YAR 3.0 is able to predict the length scale within 8% of the values of the profiles with the largest domain across all the layers and packing densities, indicating a reduced influence of cross-stream periodization on the spanwise growth of coherent structures. It is crucial to acknowledge that the two-point correlation function in the RSL cannot be considered independent of the position vector due to the heterogeneity of the flow field. In the RSL, strong signatures from the mean flow patterns affect the values of the integral length scale. Nevertheless, accepting this limitation permits the assessment of domain size impact in these layers based on the observed deviations since the mean flow patterns should have the same effect under identical surface configurations and flow conditions.

The extent of R_{22} is often used to see how far the flow field is correlated in the cross-stream direction. For the turbulent channel flow simulation, Moin & Kim (1982) showed that the transverse correlation becomes zero around $1.6L_3$ for a large domain. Based on this, they estimated that a cross-stream domain length of $3.2L_3$ is sufficient to accommodate coherent structures, which is in agreement with the presented results. However, the extent of transverse correlation does not always provide a complete picture. As shown in figure 5, the destruction of coherent structures for the narrow domain will also result in a de-correlated flow field, wrongly indicating the domain to be sufficient for de-correlation to occur.

3.2. *Impact of streamwise aspect ratio*

This section discusses the impact of XAR of the numerical domain on first and second-order flow statistics as well as on the structure of turbulence through two-point correlation maps.

Long structures seen in figure 5 are also a consequence of periodic boundary condition in the streamwise direction. In order to assess the impact of the interactions of these infinite structures, configurations mentioned in table 3 are simulated where the streamwise extent of the domain is varied systematically.

Figure 8 shows the resolved R_{11} correlation contours, mean streamwise velocity, resolved mean streamwise variance and resolved longitudinal integral length scale for cases with different XARs. For brevity, only the cases with packing density of 0.028 are shown here. From the figure, we see that as the domain is restricted in the streamwise direction, the correlation that infinite structures can sustain increases due to periodic boundary condition.

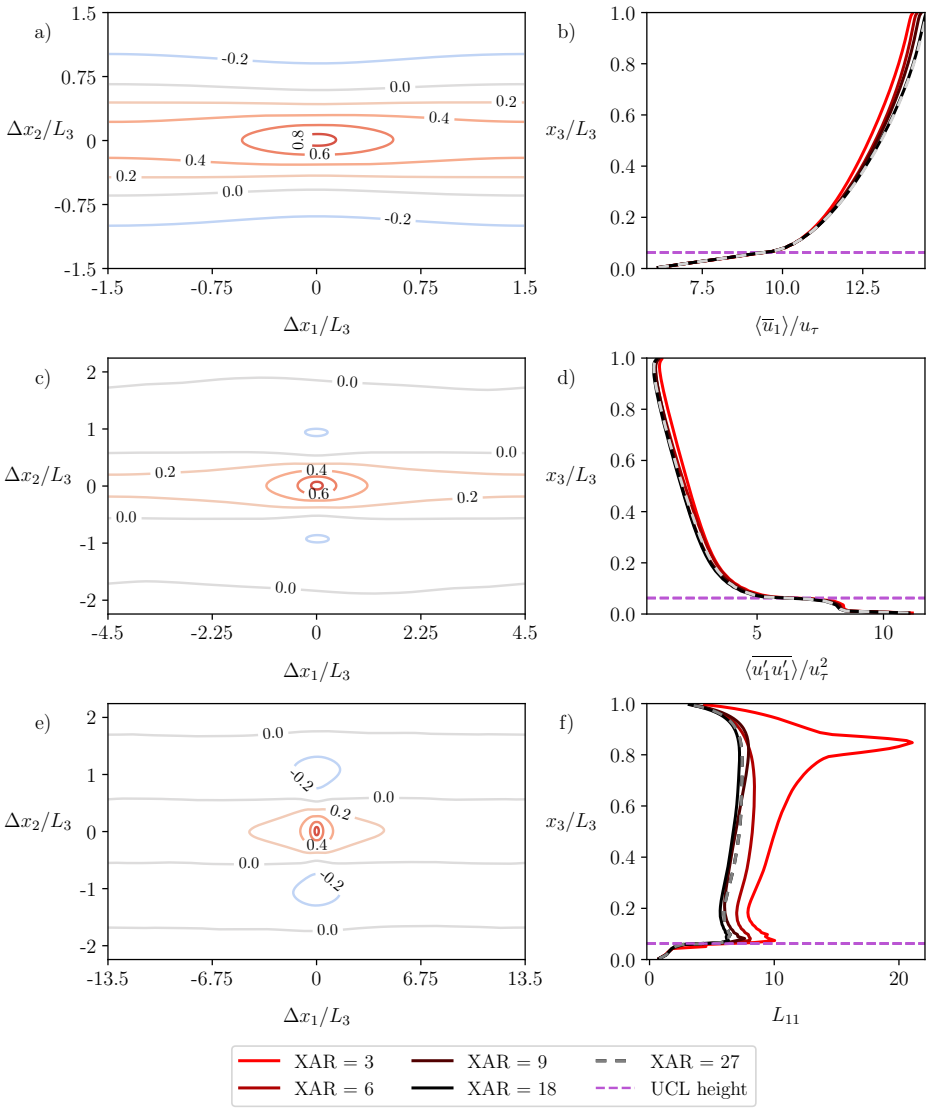

Figure 8: Two point correlation R_{11} contours (a, c, e), mean streamwise velocity (b), resolved mean streamwise variance (d), resolved longitudinal integral length scale (f) for cases with packing density 0.028. The streamwise aspect ratio (XAR) is varied as: (a) 3.0, (c) 9.0, (e) 27.0. The wall-parallel slice shown in (a, c, e) is taken at $x_3/L_3 = 0.6$. Domain configurations for cases with different XAR are shown in table 3.

Figure 8(a) shows that the infinite structure can sustain a positive correlation of 0.4 throughout the domain for the case with XAR 3.0. This value drops to 0.2 as the streamwise extent of the domain is increased, as shown in figure 8(c) for XAR 9.0. As the XAR is increased further to 27.0, the domain can no longer sustain a positive correlation of 0.2 at $x_3/L_3 = 0.6$ as shown in figure 8(e). The same is observed with negative correlation contours where the infinite structures can sustain a -0.2 correlation throughout the domain for cases with XAR 3.0 and 6.0, which is not observed for the case with XAR 9.0 and beyond. Figure 8(b, d, f) shows mean streamwise velocity and variance, as well as the integral length scale L_{11},

		$\langle\overline{u}_1\rangle/u_\tau$			$\langle\overline{u_1'u_1'}\rangle/u_\tau^2$		
λ_p	XAR	UCL	URSL	OL	UCL	URSL	OL
0.25	3.0	0.0059	0.0078	0.0351	0.0075	0.0262	0.1328
0.25	6.0	0.0015	0.0028	0.0073	0.0011	0.0017	0.0305
0.25	9.0	0.0017	0.0004	0.0022	0.0046	0.0053	0.0226
0.25	18.0	0.0004	0.0000	0.0062	0.0006	0.0014	0.0303
0.062	3.0	0.0040	0.0029	0.0413	0.0193	0.0567	0.1988
0.062	6.0	0.0023	0.0024	0.0111	0.0087	0.0155	0.0141
0.062	9.0	0.0025	0.0016	0.0160	0.0031	0.0051	0.0045
0.062	18.0	0.0007	0.0006	0.0025	0.0043	0.009	0.0058
0.028	3.0	0.0047	0.0055	0.0298	0.0298	0.0542	0.1198
0.028	6.0	0.0006	0.0016	0.0171	0.0148	0.038	0.0456
0.028	9.0	0.0011	0.0005	0.0109	0.0104	0.0135	0.0265
0.028	18.0	0.0008	0.0002	0.0014	0.0031	0.0185	0.0307

Table 7: Relative error (l_2 norm) of mean streamwise velocity and resolved mean streamwise variance in urban canopy layer (UCL), upper roughness sublayer (URSL), and outer layer (OL) for simulations with different streamwise aspect ratios (XAR). Results from the largest domain (XAR 27.0) are considered as ground truths.

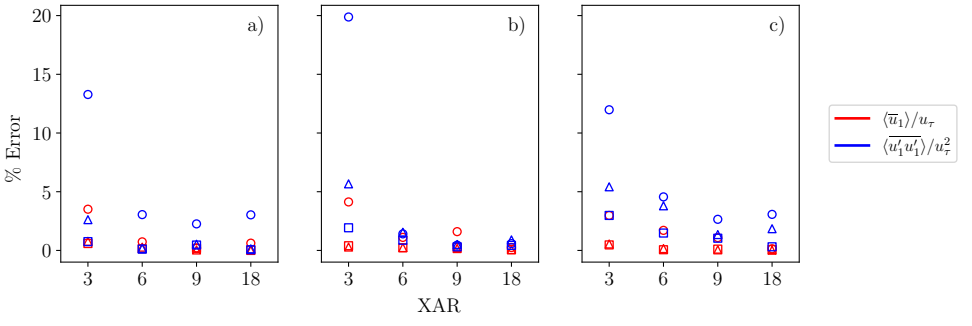

Figure 9: Visualization of error values in table 7 for different packing densities, (a) 0.25, (b) 0.062, (c) 0.028. Error values for different layers are represented by distinct symbols: Square - Urban canopy layer (UCL), triangle - Upper roughness sublayer (URSL), circle - Outer layer (OL).

which characterizes the length of instantaneous flow structures in the streamwise direction and is computed in accordance with (3.3), using a cutoff value of 0.5 (Ganapathisubramani *et al.* 2005). The increased cutoff value, compared to the 0.2 used for L_{22}, ensures that all analyzed cases, spanning various domains and packing densities, demonstrate a correlation value below the chosen contour threshold. From these statistics, we see that the strength of correlation resulting from periodization influences the first and second-order statistics. The cases with smaller streamwise extent tend to increase the correlation of the infinite structures throughout the domain, which coincides with increased resolved variance and slower mean streamwise velocity. The decrease in mean streamwise velocity is likely the result of increased turbulent mixing.

The case with the shortest domain (i.e. XAR 3.0) was found to produce a mean streamwise

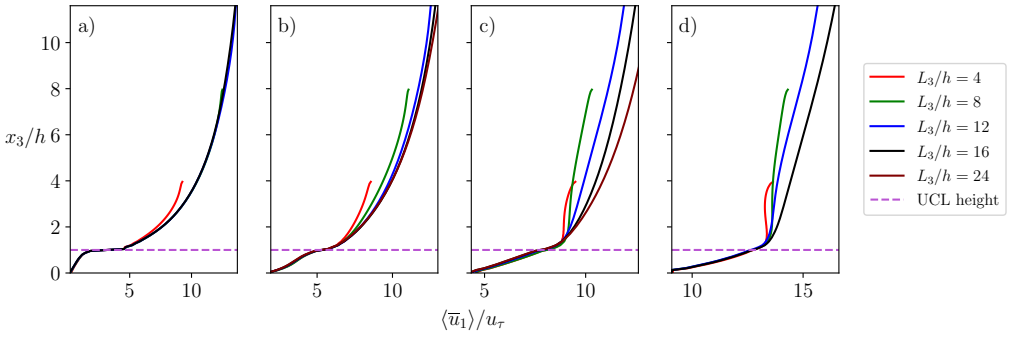

Figure 10: Mean streamwise velocity profiles for different packing densities, (a) 0.25, (b) 0.062, (c) 0.028, (d) 0.007. The vertical profiles for each packing density correspond to different scale separation cases mentioned in table 4.

velocity prediction that is within 4% of the values obtained from the largest domain (i.e. XAR 27.0) across all layers and packing densities, as shown in table 7. Figure 9 provides a visual representation of these error values.. The maximum error observed in the resolved mean streamwise variance for the UCL and URSL remains limited to 6% for all cases with the same domain. In contrast, the resolved mean streamwise variance error in the OL can increase up to 20% for the shortest domain. On the other hand, the case with XAR 6.0 is able to predict both the statistics within 5% of the values obtained from the largest domain across all the layers and packing densities, indicating a reduced influence of artificial periodization on first and second-order statistics. This also indicates that the periodic boundary condition in the streamwise direction has a lesser impact on the first-order statistics compared to the second-order statistics.

It is interesting to note that the impact of a restricted streamwise and cross-stream domain extent on flow statistics is entirely the opposite. When the cross-stream width of the domain is restricted, it inhibits the growth of coherent structures, which can lead to lower variance and higher mean streamwise velocity. Conversely, when the streamwise length of the domain is restricted, it enhances the strength of coherent structures due to artificial periodization, resulting in higher variance and lower mean streamwise velocity.

3.3. *Impact of scale separation*

This section discusses the impact of SS (L_3/h) of the numerical domain on first and second-order flow statistics. Here, two different scalings mentioned in §2.2.1 and §2.2.2 are discussed in order to isolate the impact of SS appropriately.

Initially, simulation configurations are selected based on canopy length-based scaling discussed in §2.2.1 to achieve different SSs. The configurations are mentioned in table 4. This is the conventional way to test the impact of SS, where the domain height is varied systematically without changing the surface.

Figure 10 shows profiles of mean streamwise velocity for different SSs and packing densities. For the case with the highest packing density shown in figure 10(a), all the velocity profiles from SS 8 to 16 collapse quite well. However, this trend is not observed when the packing density of the canopy surface is systematically decreased. Figure 10(b, c, d) show significant deviation in the mean velocity profile when the SS varies from 8 to 16. This significant difference in the velocity profiles is observed for the sparsely packed cases because varying L_3/h while keeping s_2/h constant changes a key parameter s_2/L_3, which controls the size and strength of secondary flows in sparse, regularly aligned canopies (Willingham

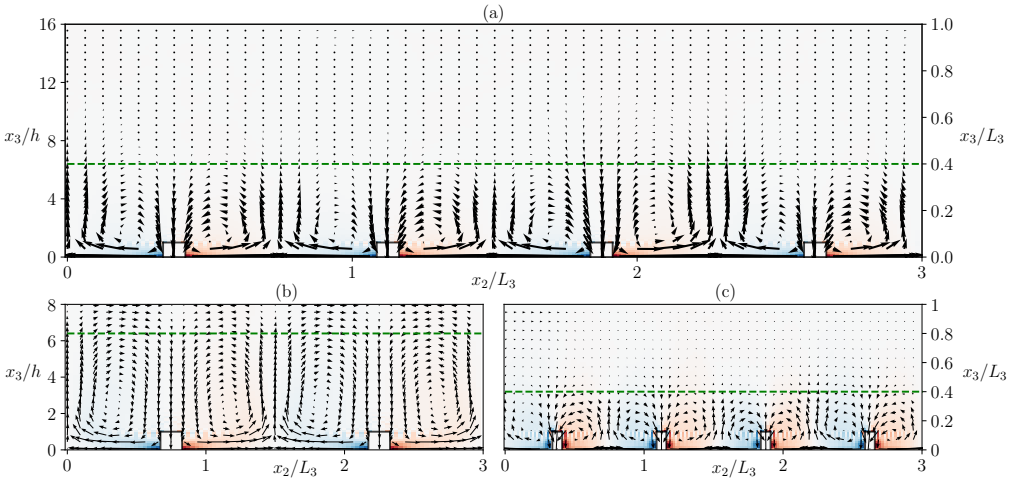

Figure 11: Flow configuration for sparsely arranged canopies based on two different sets of Pi groups. The packing density of the surface is 0.007 in all cases. The scale separation is varied as (a) 16, (b, c) 8. Configuration in (b) is scaled down from (a) based on canopy length-based scaling, whereas configuration in (c) is scaled down from (a) based on boundary layer height-based scaling. The green reference line matches x_3/h for (a) and (b) and x_3/L_3 for (a) and (c).

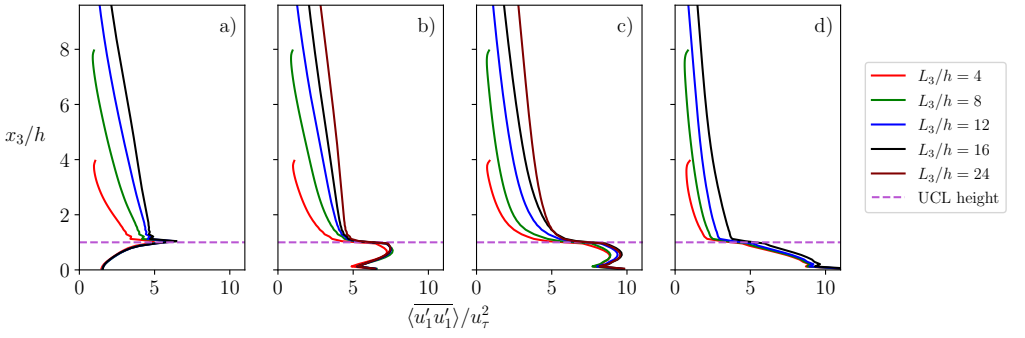

Figure 12: Resolved mean streamwise variance profiles for different packing densities, (a) 0.25, (b) 0.062, (c) 0.028, (d) 0.007. The vertical profiles for each packing density correspond to different scale separation cases mentioned in table 4.

et al. 2014; Yang & Anderson 2017; Vanderwel & Ganapathisubramani 2015). For example, when the domain height is decreased from $16h$ to $8h$ for the sparse configuration with packing density 0.007, for which s_2/h is equal to 12, s_2/L_3 changes from 0.75 to 1.5. When s_2/L_3 is 0.75, it results in the generation of moderately strong secondary flows, whereas when the parameter is increased to 1.5, it results in strong secondary flows, which occupy the entire half-channel height. Figure 11 displays this effect. When the base configuration shown in figure 11(a) is scaled down using canopy length-based scaling, the resulting flow configuration shown in figure 11(b) is quite different. A dashed green line is drawn for reference at $x_3/h = 6.4$. We can clearly see that at this height, the flow configuration is entirely different, and the magnitude of this difference is directly related to the size and strength of secondary flows in the base configuration. Hence, the deviation observed in figure 10(b, c, d) cannot be solely attributed to the effect of SS. These results highlight that,

		$\langle\overline{u}_1\rangle/u_\tau$		$\langle\overline{u_1'u_1'}\rangle/u_\tau^2$	
λ_p	SS	UCL	URSL	UCL	URSL
0.25	4	0.0684	0.0280	0.0663	0.3004
0.25	8	0.0431	0.0066	0.0184	0.1273
0.25	12	0.0149	0.0026	0.0073	0.0395
0.062	8	0.0303	0.0272	0.0208	0.1584
0.062	12	0.0080	0.0040	0.0051	0.0528

Table 8: Relative error (l_2 norm) of mean streamwise velocity and resolved mean streamwise variance in urban canopy layer (UCL) and upper roughness sublayer (URSL) for simulations with different scale separations (SS) mentioned in table 4. Results from the domain with SS 16 are considered as ground truths.

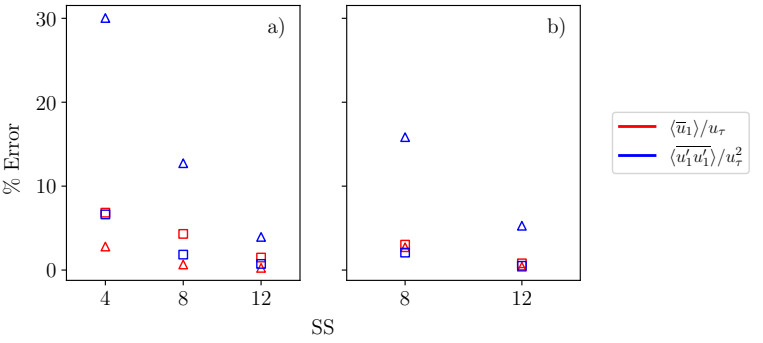

Figure 13: Visualization of error values in table 8 for different packing densities, (a) 0.25, (b) 0.062. Error values for different layers are represented by distinct symbols: Square - Urban canopy layer (UCL), triangle - Upper roughness sublayer (URSL).

for sparse configurations which induce secondary flows, the set of Pi groups stated in (2.3) cannot be used to isolate the impact of SS. For the densely packed case, the strength of the secondary flows is very weak due to the limiting cross-stream gap, and the surface essentially behaves as a conventional rough surface (Yang & Anderson 2017). This is why decreasing the parameter s_2/L_3 with increasing domain height does not have any impact on already weak secondary flows, which justifies the good collapse of streamwise velocity profiles observed in figure 10(a) across a large range of SS values.

Considering the outcomes illustrated in figure 10(a), it becomes apparent that the influence of domain height on the mean velocity profile throughout the domain, via the generation of secondary flows, becomes negligible when the ratio $s_2/L_3 \lesssim 0.25$. Thus, achieving converged mean velocity profiles appears feasible by adhering to the following SS criterion:

$$\frac{L_3}{h} \gtrsim 4\frac{s_2}{h}\,. \tag{3.4}$$

Consequently, (3.4) postulates that in scenarios with a packing density of 0.062, where $s_2/h = 4$, the mean velocity profiles will converge satisfactorily beyond an SS value of 16. In contrast, for packing density 0.028, where $s_2/h = 6$, deviations in the mean velocity profiles will persist even at an SS of 16. To test the validity of (3.4), two additional simulations are conducted with SS 24 for packing densities 0.062 and 0.028. The inference drawn from

the equation aligns with the observations in figure 10(b, c), where the mean streamwise velocity profile with SS 16 is within 1% of the profile with SS 24 for a packing density of 0.062. Meanwhile, for a packing density of 0.028, discrepancies in the velocity profile remain noticeable at both SS values of 16 and 24. Hence, (3.4) provides us with the means to make informed decisions regarding the appropriate SS level required to completely mitigate the impact of secondary flows on the mean velocity profiles.

Figure 12 depicts profiles of resolved mean streamwise variance for various SSs and packing densities. The figure demonstrates that the rate of variance decay is significantly affected by the top boundary condition. Furthermore, the change in the parameter s_2/L_3 also affects the variance values in the RSL. Yang & Anderson (2017) showed that surfaces with s_2/L_3 considerably below 1 behave as conventional rough surfaces and exhibit weaker secondary circulations. Since the secondary flows are weak in such cases, the RSL statistics are predominantly affected by the wake flow from the canopies. Turbulence scales in the wake flow are primarily influenced by the dimensions of the canopy, which are preserved in canopy length-based scaling (Raupach *et al.* 1991). As a result, the turbulence features of the RSL remain similar for these cases, enabling comparisons across different SSs. Hence, this scaling can still be used to evaluate the impact of SS on RSL statistics for cases with packing densities of 0.25 and 0.062 and for scale separations where s_2/L_3 is less than or equal to 0.5. It is worth noting, however, that instances with $s_2/L_3 \approx 0.5$ also display minor secondary scale circulations, indicating that the RSL traits may not be identical, but these will not contribute significantly to the flow statistics. The error values of these cases in the UCL and URSL are presented in table 8 and visualized in figure 13. The maximum error observed for the mean streamwise velocity is less than 7% across all scale separations. As for the resolved variance, the maximum error observed in the UCL is also less than 7% across all scale separations. In the URSL, the cases with a scale separation of 12 capture this statistic with a maximum error of approximately 5%, while the error can reach up to 15% for the cases with a scale separation of 8 and 30% for the case with a scale separation of 4. The discussion in the previous two paragraphs demonstrates that the canopy length-based scaling is unable to accurately isolate the impact of SS on turbulent flow statistics in the sparse cases which generate secondary flows as well as in the outer layer. Although scaling x_3 with L_3 instead of h enables comparison of statistics in the outer layer for selected cases, it still does not facilitate comparison of the statistics for sparse cases or packing density 0.062 with scale separation 4. Thus, to overcome the limitations of canopy length-based scaling and to study the impact of SS across all the packing densities and the outer layer, a change in the repeating parameter determining the length scale is required.

A different scaling was proposed in §2.2.2, where the domain height was chosen as a repeating parameter. This results in a different set of Pi groups presented in (2.4). The effect of this change in the repeating parameter can be appreciated in figure 11. When the base configuration shown in figure 11(a) is scaled down using boundary layer height-based scaling shown in figure 11(c), a similar flow configuration is achieved. A green dashed line is drawn for reference at $x_3/L_3 = 0.4$, which accurately compares the extent of secondary circulation despite having different SS. Figure 11(c) also demonstrates that the underlying surface configuration changes and the cuboids become slender when the SS is decreased as per boundary layer height-based scaling. However, it should be noted that this change in configuration preserves the frontal area fraction of the surface, which results in a similar z_0 value throughout the cases for a fair comparison.

As the length scales of eddies in the RSL are predominantly associated with the canopy lengths, a direct comparison of statistics is not possible in this region, as the boundary layer height-based scaling distorts the surface. However, we can still compare statistics in the OL as it has turbulent eddies independent of canopy scales. Also, the OL turbulence is most

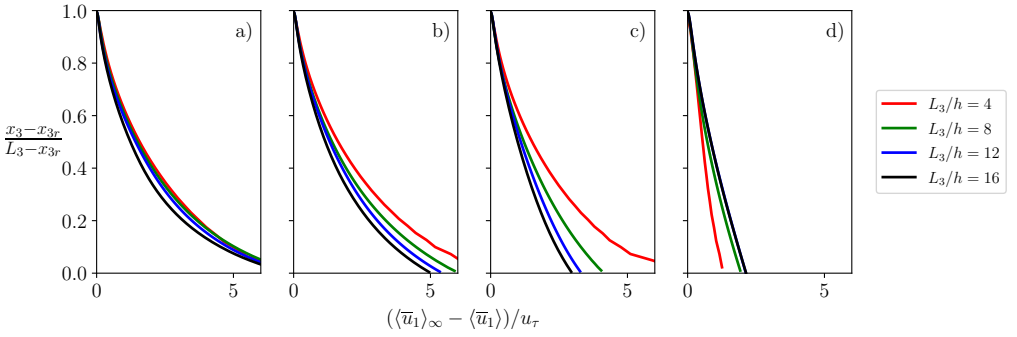

Figure 14: Mean streamwise velocity defect profiles for different packing densities, (a) 0.25, (b) 0.062, (c) 0.028, (d) 0.007. The vertical profiles for each packing density correspond to different scale separation cases mentioned in table 5.

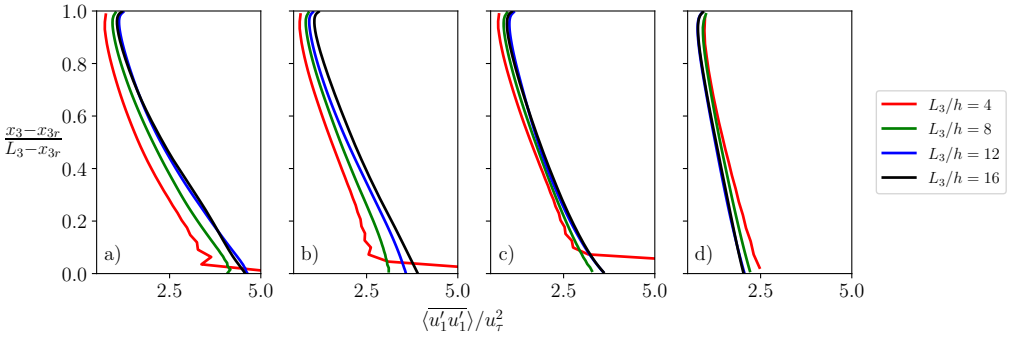

Figure 15: Resolved mean streamwise variance profiles for different packing densities, (a) 0.25, (b) 0.062, (c) 0.028, (d) 0.007. The vertical profiles for each packing density correspond to different scale separation cases mentioned in table 5.

	SS - $\langle\overline{u}_1\rangle/u_\tau$			SS - $\langle\overline{u_1'u_1'}\rangle/u_\tau^2$		
λ_p	4	8	12	4	8	12
0.250	0.1413	0.1269	0.0793	0.2366	0.1047	0.0223
0.062	0.3939	0.2026	0.1004	0.2865	0.1751	0.0775
0.028	0.8532	0.3279	0.1192	0.2773	0.0720	0.0219
0.007	0.4091	0.1268	0.0029	0.1836	0.0950	0.0143

Table 9: Relative error (l_2 norm) of mean streamwise velocity and resolved mean streamwise variance for simulations with different scale separation and packing densities mentioned in table 5. Results from the case with the largest scale separation ($L_3/h = 16$) are considered as ground truths. The statistics are compared in the outer layer (OL).

affected by the SS due to its close proximity to the no-stress boundary condition. Thus, minimizing the impact of SS in the OL ensures that the impact of SS is minimal in UCL and URSL.

In order to accurately match equivalent points in the outer layer across cases, a new scaling is introduced, which maps the extent of the outer layer from 0 to 1. A non-dimensional

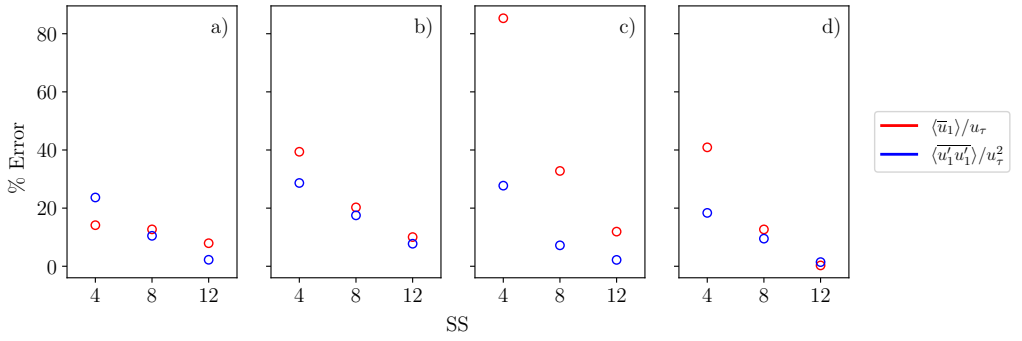

Figure 16: Visualization of error values in table 9 for different packing densities, (a) 0.25, (b) 0.062, (c) 0.028, (d) 0.007. Circle represents error values in the outer layer (OL).

function $f(x_3, x_{3r}, L_3)$ is defined as:

$$f(x_3, x_{3r}, L_3) = \frac{x_3 - x_{3r}}{L_3 - x_{3r}}, \tag{3.5}$$

where x_{3r} is the height of the RSL, which is calculated from (3.1).

Figure 14 shows the mean streamwise velocity defect for different SSs and packing densities. Simulation setups for the shown cases can be found in table 5. A converging trend is observed across different packing densities, which was absent in figure 10. Figure 15 shows the resolved mean streamwise variance for the same cases. We observe that the errors between the streamwise velocity and resolved streamwise variance profiles shown in table 9 are relatively smaller for the densely packed case as well as the most sparse configuration considered in this study. Figure 16 provides a visual representation of these error values. A physical explanation for this behavior can be provided by examining the characteristics of the RSL in the distorted surfaces. In canopy flows that do not generate secondary flows, the RSL is dynamically influenced by length scales associated with roughness elements (Raupach *et al.* 1991). Therefore, the change in the dimensions of cuboids required to preserve the s_2/L_3 Pi group changes the RSL characteristics of the surface. When the SS is low, the OL is not truly independent of influence from the roughness elements, and this change in the RSL turbulence also affects statistics in the OL. For the case with high packing density, the RSL does not extend significantly beyond the UCL, as can be inferred from the magnitude of dispersive fluxes (not shown), and OL independence is quickly achieved. In sparse canopies that generate secondary flows, the RSL is predominantly occupied by the counter-rotating vortices, which are preserved when the surface is scaled as per the boundary layer height-based scaling. Hence, the boundary layer height-based scaling tends to preserve the RSL characteristics for the sparse cases generating secondary flows. This explains the observed lower shift in streamwise velocity and resolved streamwise variance profiles for highly dense and highly sparse cases. From the table, we see that the domains with SS 12 predict both the quantities with less than 12% error in the OL, and this error magnitude is likely to be lesser in UCL and URSL given their relatively more significant separation from the top boundary. In contrast the error for cases with SS 4 and 8 can be substantial and we refer the reader to table 9 for specific values. Also, a monotonic increase is observed in the value of resolved variance as the SS is increased, except for the case with strong secondary flows. This shows that domains with smaller SSs tend to inhibit the growth of turbulent structures, which contribute to the variance magnitude. This behavior is clearly linked to the free-slip upper boundary condition, which is known to dampen velocity fluctuations, especially those

in the vertical direction. In cases where strong secondary flows are present, the reversal in the trend indicates that the patterns of strong mean flow in the RSL affect the turbulence in the OL, leading to a higher variance value. However, when the SS is increased beyond 12, the statistics show excellent collapse, suggesting the recovery of OL independence.

3.4. *Existence of inertial sublayer in canopy flows*

In this section, the existence of an inertial sublayer is examined for cases with different SS available from the suite of simulations presented in table 4. In urban boundary layer flow, the inertial sublayer exists between the RSL and OL and is the region where most of the turbulent kinetic energy is generated (Jiménez 2004). A logarithmic rise in the velocity within the region characterizes this layer. In the flow over roughness elements, the logarithmic profile can be described in terms of aerodynamic roughness length z_0, which quantifies the ability of the surface to absorb momentum, as

$$\frac{\langle \overline{u}_1 \rangle}{u_\tau} = \frac{1}{\kappa} \ln\left(\frac{x_3 - d}{z_0}\right), \tag{3.6}$$

where d is the aerodynamic displacement height of the given surface. As the existence of an inertial sublayer is not always guaranteed, we use a modified form of (3.6) in this study, as

$$\frac{\langle \overline{u}_1 \rangle}{u_\tau} = \frac{1}{\kappa} \ln\left(\frac{x_3 - d}{L_3}\right) - \beta, \tag{3.7}$$

where β is a dimensionless constant defined as

$$\beta = \frac{1}{\kappa} \ln\left(\frac{z_0}{L_3}\right). \tag{3.8}$$

Figure 17 depicts the mean streamwise velocity across multiple cases with varying packing densities and SSs in accordance with (3.7). In this study, the value of κ is chosen as 0.384 as recommended by Monkewitz *et al.* (2008). The solid black line in the figure serves as a reference to highlight the logarithmic profile of rough wall flow with a reference roughness length $z_{0_{\text{ref}}} = 0.25$ and $d = 0$. The matching of slope of the profiles with the reference line indicates the existence of an inertial sublayer. The extent of deviation from the reference line depends on the magnitude of z_0. For the existence of logarithmic profiles, a layer sufficiently distant from the surface is necessary, such that the canopy scales do not impact the flow, and from the boundary layer height, such that L_3 is not a dominant length scale. Hence, SS becomes an essential parameter to determine whether the characteristics of the true inertial sublayer can be retrieved. In this study, an upper limit of $0.15L_3$ is considered for the inertial sublayer as beyond this height, the boundary layer height L_3 becomes a dominant scale (Jiménez 2004; Marusic & Monty 2019). However, some researchers have recommended a larger value of $0.3L_3$ (Pope 2000). The dashed lines in Figure 17 indicate the upper limit of $0.15L_3$ for cases with $\lambda_p = 0.25$ and 0.062, and the colors correspond to the velocity profiles. The solid horizontal lines represent the height of the canopy in these cases.

Figure 17(a, b) shows that a logarithmic rise in the velocity is noticeable for cases with packing densities of 0.25 and 0.062 and SSs of 12 and 16. However, as the SS decreases to 8, the extent of the logarithmic layer is substantially reduced compared to the SSs of 12 and 16. For this SS, the logarithmic rise is only observed for the case with a packing density of 0.25 around the $0.15L_3$ mark. This occurs because the height extent of the RSL for a densely packed configuration (e.g., $\lambda_p = 0.25$) is smaller than that for configurations with relatively sparse arrangements (e.g., $\lambda_p = 0.062$), which can be observed from the extent to which the dispersive fluxes are significant (not shown). Hence, for the case with $\lambda_p = 0.25$, the extent

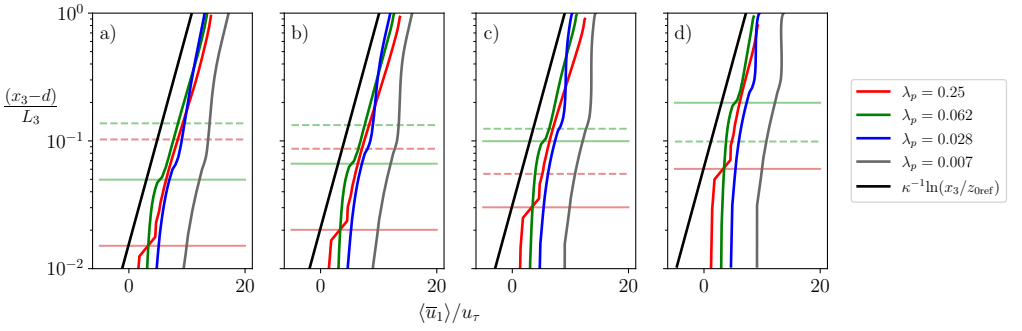

Figure 17: Mean streamwise velocity profiles for different scale separations, (a) 16, (b) 12, (c) 8, (d) 4. The vertical profiles for each scale separation correspond to different packing density cases mentioned in table 4. The horizontal lines correspond to the region between canopy height ($x_3/h = 1$, solid) and the theoretical limit of the extent of inertial sublayer ($x_3/L_3 = 0.15$, dashed) for packing densities 0.25 and 0.062.

of the RSL does not entirely occupy the significant portion of the region where the inertial sublayer can exist. However, the same is not true for the case with $\lambda_p = 0.062$. For cases with a SS of 4, the height of the canopy exceeds the upper limit of the extent of the inertial sublayer, and thus, the inertial sublayer is not observed for any case. For packing density 0.25 and 0.062 and SS 16, the corresponding z_0 values are measured at 0.068 and 0.086, respectively. Across different SSs where the logarithmic profile is observed, the z_0 values remain within 0.5% of those observed for SS 16.

Cases with packing densities of 0.028 and 0.007 have been excluded from the above discussion because these configurations generate secondary flows (see § 3.3). The size and strength of the secondary flows are significantly influenced by the height of the boundary layer, as s_2/L_3 is one of the crucial parameters governing secondary flows. Thus, for the cases with secondary flows, the height of the boundary layer L_3 directly impacts the flow velocity in the RSL, which are occupied by the counter-rotating vortices. Consequently, if L_3 affects the velocity at the wall as well as near the top boundary, there is no layer in-between where the impact of L_3 on the velocity can be neglected. As a result, the basic requirement of independence from L_3 required for the existence of an inertial sublayer does not hold, and thus we do not observe an inertial sublayer for these cases in figure 17. This behavior is consistent with findings from secondary flow research by Willingham *et al.* (2014).

These findings highlight a crucial aspect of canopy flows, where the existence of an inertial sublayer is not solely determined by the SS as in the case of smooth wall flows but also depends on the packing density of the underlying surface. Figure 17 illustrates that, for a given SS, the inertial sublayer may or may not exist depending on the underlying canopy configuration. Specifically, for a densely packed configuration, the flow may exhibit an inertial sublayer, while a sparsely arranged canopy may not exhibit such a layer for a particular SS.

4. Conclusion

In this study, we investigate the impact of numerical domain size on turbulent flow statistics for canopy flows spanning a wide range of packing densities. Specifically, we consider the impact of three relevant length scales: YAR (L_2/L_3), XAR (L_1/L_3), and SS (L_3/h). Furthermore, we explore the question of the existence of an inertial sublayer for a wide range of cases with different packing densities and scale separations. Our findings reveal

that poorly designed domains can have a significant impact on turbulent flow statistics and turbulent coherent structures. We outline the main findings of this study as follows:

(i) Impact of cross-stream aspect ratio (L_2/L_3): Narrower domains, characterized by YAR considerably below 3.0, can be inadequate to accommodate a pair of fast and slow turbulent streaks, thereby artificially destroying the growth of turbulent structures in the streamwise direction. Additionally, a decrease in the growth of cross-stream structures is observed by analyzing the resolved integral length scale L_{22} in narrower domains. Moreover, the statistics indicate that narrower domains tend to underpredict the value of resolved streamwise variance across a wide range of packing densities. Overall, it is concluded that domains with YAR 3.0 or more are sufficient to reduce the artificial impact of cross-stream periodization and to accurately capture the first and second-order statistics. Detailed information about the specific errors in first and second-order statistics in UCL, URSL, and OL can be found in table 6.

(ii) Impact of streamwise aspect ratio (L_1/L_3): Shorter domains, characterized by XAR considerably below 6.0, experience excessive periodization, resulting in an artificial strengthening of the turbulent coherent structures in the streamwise direction. As a result, the coherent structures may exhibit longer correlation values throughout the domain. Additionally, the statistics reveal that the shorter domains tend to overpredict the value of resolved streamwise variance across a wide range of packing densities. Overall, it is determined that domains with XAR 6.0 or more are sufficient to reduce the artificial impact of streamwise periodization and to accurately capture the first and second-order statistics. Detailed information about the specific errors in first and second-order statistics in UCL, URSL, and OL can be found in table 7.

(iii) Impact of scale separation (L_3/h): This study demonstrates that the conventional method to test the impact of scale separation has major limitations for canopy flows, especially for configurations where the parameter s_2/L_3 exceeds 0.5. To overcome the limitations of the existing method, a new set of Pi groups is proposed that can relatively accurately isolate the effects of scale separation. Using the novel L_3 scaling approach, we observe that domains with limited scale separation tend to underestimate the resolved variance values in the outer layer. In addition, our findings reveal that a scale separation of 12 and above is adequate to reduce the artificial impact of the top boundary condition on flow statistics in the UCL, URSL, and till, up to at least $0.6L_3$ in the OL. Detailed information about the specific errors in first and second-order statistics can be found in table 8 and 9.

(iv) Existence of inertial sublayer: Conventionally, scale separation is considered the sole parameter to determine the presence of an inertial sublayer in a flow field. However, our study shows that for canopy flows, the existence of an inertial sublayer depends not only on scale separation but also on the arrangement of the underlying surface. This is because the extent of the RSL depends on the underlying surface configuration and also because certain arrangements generate secondary flows which occupy the entire RSL. We found that for moderately dense ($\lambda_p = 0.062$) and dense ($\lambda_p = 0.25$) cases, a logarithmic rise in the streamwise velocity profile could be recovered for scale separations of 12 and beyond. However, for a scale separation of 8, only the densely packed case ($\lambda_p = 0.25$) exhibited the characteristic logarithmic rise. For sparse configurations which generate secondary flows, it is observed that the inertial sublayer does not exist for any scale separation. Scaling justification is provided in order to support the observed results for secondary flow cases.

Overall, our results indicate that a domain with an SS of 12 or larger, YAR of 3.0 or larger, and XAR of 6.0 or larger is suitable for minimizing the artificial impacts of the numerical domain. However, researchers can use the error values reported in tables 6, 7, 8 and 9 to choose smaller domain than recommended based on their region of interest and research purpose. It is important to note that our study only considers the aligned configuration of

canopy elements, but we expect our recommendations to be valid for staggered as well as other configurations based on the physical justifications provided in each section. We recommend that researchers match their configurations with an aligned configuration that has a similar extent of RSL.

Acknowledgements. The authors would like to thank Dr. Weiyi Li for his comments and suggestions on the manuscript. This work used the Anvil supercomputer at Purdue University through allocation ATM180022 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296. The authors also acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high performance computing resources that have contributed to the research results reported within this paper.

Funding. This material is based upon work supported by, or in part by, the Army Research Laboratory and the Army Research Office under grant number W911NF-22-1-0178.

Declaration of interests. The authors report no conflict of interest.

Appendix A.

In this section, we compare our LES results with independent DNS, LES and wind tunnel data to showcase validity of the results presented in this study. For this purpose, a staggered configuration with packing density 0.25 is simulated. The DNS results of Coceal *et al.* (2006) with 32 nodes per cube (DNS_C32) and 64 nodes per cube (DNS_C64) are chosen, while the LES result of Anderson *et al.* (2015) (LES_A) is chosen to perform the validation. Wind tunnel data from Cheng & Castro (2002) (WT_CC) is chosen for the same configuration. The domain size chosen for this validation consists of SS 4, YAR 3 and XAR 6. The SS of 4 is chosen such that statistics in UCL,URSL and OL can be compared with DNS_C32, DNS_C64 and LES_A results using a single scaling. The choice of YAR and XAR is consistent with the recommendation of this study. We maintained a resolution of $n_1 \times n_2 \times n_3 = 6 \times 6 \times 12$ per cube, mirroring the resolution employed in the study.

Figure 18 shows first and second-order statistics for the chosen configuration. Here, friction velocity (u_τ) is based on the imposed pressure gradient in the LES simulation, while the shear velocity (u_*) for the LES results is obtained as

$$u_* = \sqrt{-\langle \overline{u_1' u_3'} \rangle}_{\max} . \tag{A 1}$$

Here, the symbol $(\cdot)'$ is used to denote temporal fluctuations. Note that the contribution from the subgrid stresses has been added in the second-order statistics and intrinsic averaging scheme as been used in UCL to match the statistics with Coceal *et al.* (2006). This specific definition of shear velocity was adopted due to its consistency with both wind tunnel experiments and DNS data, a finding similarly reported by Tian *et al.* (2023). This relation gives $u_* = 0.958 u_\tau$ for the considered configuration.

In conclusion, figure 18 demonstrates our LES algorithm's proficiency in capturing both first and second-order statistics.

Appendix B.

This section presents an analysis of the influence of grid resolution on turbulent flow statistics, specifically on streamwise velocity, resolved streamwise variance, and resolved Reynolds shear stress. To conduct this study, a computational domain with SS of 4, YAR of 4.5, and XAR of 6.0 is selected. The domain is discretized with different resolutions such that $n_1 \times n_2 \times n_3 = 4 \times 4 \times 8, 6 \times 6 \times 12, 8 \times 8 \times 16$, where n_i represents the number of collocation

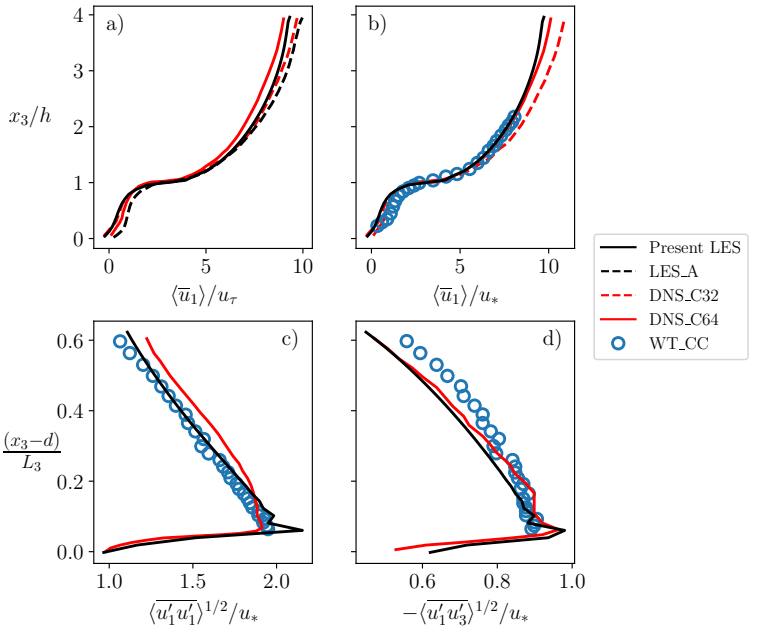

Figure 18: (a) Mean streamwise velocity normalized by friction velocity (u_τ), (b) mean streamwise velocity normalized by shear velocity (u_*), (c) r.m.s. value of streamwise Reynolds stress normalized by shear velocity (u_*) and (d) r.m.s. value of Reynolds shear stress profile normalized by shear velocity (u_*) for staggered arrangement with packing density 0.25. LES_A: LES result from Anderson *et al.* (2015), DNS_C32/64: DNS result from Coceal *et al.* (2006) using 32/64 nodes per cube, WT_CC: Wind tunnel data from Cheng & Castro (2002).

Resolution	$\langle\overline{u}_1\rangle/u_\tau$	$\langle\overline{u_1'u_1'}\rangle/u_\tau^2$	$\langle\overline{u_1'u_3'}\rangle/u_\tau^2$
$4\times4\times8$	0.0292	0.0919	0.0538
$6\times6\times12$	0.0183	0.0687	0.0269

Table 10: Relative error (l_2 norm) of mean streamwise velocity, resolved mean streamwise variance and Reynolds shear stress for different resolutions. Results from the case with the highest resolution ($8\times8\times16$) are considered to be ground truths.

nodes per cube edge. Although this domain is not sufficient to accurately capture the turbulent flow statistics, the aim of this section is to demonstrate that the flow field is not significantly impacted by the choice of grid resolution, indicating that the chosen domain is appropriate for this purpose. The results presented in figure 19 reveal that the resolutions of $4\times4\times8$ and $6\times6\times12$ can predict the trends in the profiles with satisfactory accuracy based on the scope of this work. The errors associated with these profiles are summarized in table 10. Error values are modest compared to corresponding variations in flow statistics resulting from XAR, YAR, and SS. Since the existence of the inertial sublayer necessitates the accurate capture of flow statistics, a higher resolution of $6\times6\times12$ is selected for the analysis of the impact of scale separation in §3.3. For §3.1 and 3.2, a lower resolution of $4\times4\times8$ is chosen to ensure the computational feasibility of this study.

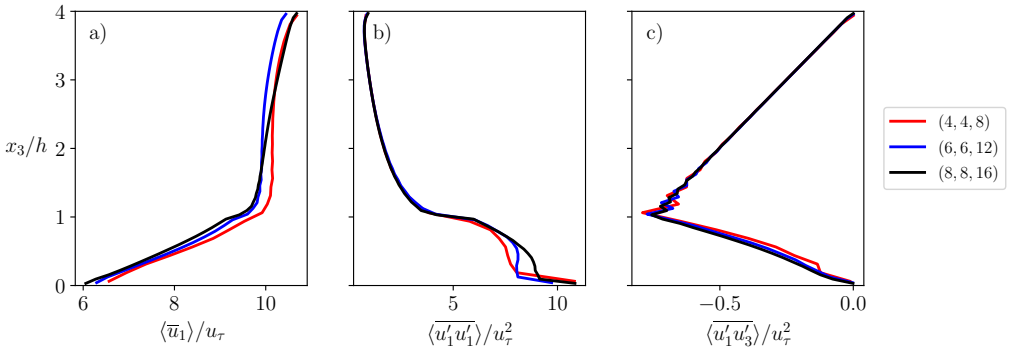

Figure 19: (a) Streamwise velocity, (b) resolved streamwise variance and (c) Reynolds shear stress profiles for a case with a packing density of 0.028, scale separation of 4, and an aspect ratio of 1 : 4.5 : 6, at different resolutions. The legend denotes the resolution as (n_1, n_2, n_3), where n_i represents the number of collocation nodes per cube edge.

REFERENCES

Albertson, J. D. & Parlange, M. B. 1999*a* Natural integration of scalar fluxes from complex terrain. *Adv. Water Resour.* **23**, 239–252.

Albertson, J. D. & Parlange, M. B. 1999*b* Surface length scales and shear stress: Implications for land-atmosphere interaction over complex terrain. *Water Resour. Res.* **35**, 2121–2132.

Anderson, William 2016 Amplitude modulation of streamwise velocity fluctuations in the roughness sublayer: evidence from large-eddy simulations. *Journal of Fluid Mechanics* **789**, 567–588.

Anderson, W., Li, Q. & Bou-Zeid, E. 2015 Numerical simulation of flow over urban-like topographies and evaluation of turbulence temporal attributes. *J. Turbul.* **16**, 809–831.

Andreas, E. L., Claffy, K. J., Jordan, R. E., Fairall, C. W., Guest, P. S., Persson, P. O. G. & Grachev, A. A. 2006 Evaluations of the von kármán constant in the atmospheric surface layer. *J. Fluid Mech.* **559**, 117–149.

Barlow, J. F., Harman, I. N. & Belcher, S. E. 2004 Scalar fluxes from urban street canyons. part i: Laboratory simulation. *Boundary-Layer Meteorol.* **113**, 369–385.

Bou-Zeid, E., Meneveau, C. & Parlange, M. B. 2004 Large-eddy simulation of neutral atmospheric boundary layer flow over heterogeneous surfaces: Blending height and effective surface roughness. *Water Resour. Res.* **40**, W02505.

Bou-Zeid, E., Meneveau, C. & Parlange, M. B. 2005 A scale-dependent lagrangian dynamic model for large eddy simulation of complex turbulent flows. *Phys. Fluids* **17**, 025105.

Buckingham, E. 1914 On physically similar systems; illustrations of the use of dimensional equations. *Phys. Rev.* **4**, 345–376.

Canuto, C., Hussaini, M. Y., Quarteroni, A. & Zang, T. A. 2007 *Spectral methods: evolution to complex geometries and applications to fluid dynamics*. Springer Science & Business Media.

Chamecki, M., Meneveau, C. & Parlange, M. B. 2009 Large eddy simulation of pollen transport in the atmospheric boundary layer. *J. Aerosol Sci.* **40**, 241–255.

Cheng, H. & Castro, I. P. 2002 Near wall flow over urban-like roughness. *Boundary-Layer Meteorol.* **104**, 229–259.

Chester, S., Meneveau, C. & Parlange, M. B. 2007 Modeling turbulent flow over fractal trees with renormalized numerical simulation. *J. Comput. Phys.* **225**, 427–448.

Chung, D., Hutchins, N., Schultz, M. P. & Flack, K. A. 2021 Predicting the Drag of Rough Surfaces. *Annu. Rev. Fluid Mech.* **53**, 439–471.

Chung, D., Monty, J. P. & Hutchins, N. 2018 Similarity and structure of wall turbulence with lateral wall shear stress variations. *J. Fluid Mech.* **847**, 591–613.

Claus, J., Coceal, O., Thomas, T. Glyn, Branford, S., Belcher, S. E. & Castro, I. P. 2012 Wind-Direction Effects on Urban-Type Flows. *Boundary-Layer Meteorol.* **142**, 265–287.

Coceal, O., Dobre, A., Thomas, T. G. & Belcher, S. E. 2007 Structure of turbulent flow over regular arrays of cubical roughness. *J. Fluid Mech.* **589**, 375–409.

Coceal, O., Thomas, T. G., Castro, I. P. & Belcher, S. E. 2006 Mean flow and turbulence statistics over groups of urban-like cubical obstacles. *Boundary-Layer Meteorol.* **121**, 491–519.

Comte-Bellot, G. 1963 *Contribution à l'étude de la turbulence de conduite*. impr. Guirimand.

Fang, J. & Porté-Agel, F. 2015 Large-eddy simulation of very-large-scale motions in the neutrally stratified atmospheric boundary layer. *Boundary-Layer Meteorol.* **155**, 397–416.

Fang, J. & Porté-Agel, F. 2015 Large-eddy simulation of very-large-scale motions in the neutrally stratified atmospheric boundary layer. *Boundary-Layer Meteorology* **155**, 397–416.

Fernando, H., Lee, S.M., Anderson, J., Princevac, Marko, Pardyjak, Eric & Grossman-Clarke, Susanne 2001 Urban fluid mechanics: Air circulation and contaminant dispersion in cities. *Environ. Fluid Mech.* **1**, 107–164.

Ganapathisubramani, B., Hutchins, N., Hambleton, W. T., Longmire, E. K. & Marusic, I. 2005 Investigation of large-scale coherence in a turbulent boundary layer using two-point correlations. *J. Fluid Mech.* **524**, 57–80.

Giometto, M. G., Christen, A., Meneveau, C., Fang, J., Krafczyk, M. & Parlange, M. B. 2016 Spatial characteristics of roughness sublayer mean flow and turbulence over a realistic urban surface. *Boundary-Layer Meteorol.* **160**, 425–452.

Gorlé, C., Garcia-Sanchez, C. & Iaccarino, G. 2015 Quantifying inflow and rans turbulence model form uncertainties for wind engineering flows. *J. Wind. Eng. Ind.* **144**, 202–212.

Jiménez, J. 2004 Turbulent flows over rough walls. *Annu. Rev. Fluid Mech.* **36**, 173–196.

Kanda, M., Moriwaki, R. & Kasamatsu, F. 2004 Large-eddy simulation of turbulent organized structures within and above explicity resolved cube arrays. *Boundary-Layer Meteorol.* **112**, 343–368.

Kim, J. & Moin, P. 1985 Application of a fractional-step method to incompressible navier-stokes equations. *J. Comput. Phys.* **59**, 308–323.

Leonardi, S. & Castro, I. P. 2010 Channel flow over large cube roughness: a direct numerical simulation study. *J. Fluid Mech.* **651**, 519–539.

Li, Q., Bou-Zeid, E., Anderson, W., Grimmond, S. & Hultmark, M. 2016 Quality and reliability of les of convective scalar transfer at high reynolds numbers. *Int. J. Heat Mass Transf.* **102**, 959–970.

Lowe, D., Ebi K. L. & Forsberg, B. 2011 Heatwave early warning systems and adaptation advice to reduce human health consequences of heatwaves. *Int. J. Environ. Res. Public Health* **8**, 4623–4648.

Lozano-Durán, A. & Jiménez, J. 2014 Effect of the computational domain on direct simulations of turbulent channels up to $Re\tau = 4200$. *Phys. Fluids* **26**.

Margairaz, F., Giometto, M. G., Parlange, M. B. & Calaf, M. 2018 Comparison of dealiasing schemes in large-eddy simulation of neutrally stratified atmospheric flows. *Geosci. Model Dev.* **11**, 4069–4084.

Marusic, I. & Monty, J. P. 2019 Attached eddy model of wall turbulence. *Annu. Rev. Fluid Mech.* **51**, 49–74.

Meili, N., Manoli, G., Burlando, P., Bou-Zeid, E., Chow, W. T. L., Coutts, A. M., Daly, E., Nice, K. A., Roth, M., Tapper, N. J., Velasco, E., Vivoni, E. R. & Fatichi, S. 2020 An urban ecohydrological model to quantify the effect of vegetation on urban climate and hydrology (ut&c v1.0). *Geosci. Model Dev.* **13**, 335–362.

Mittal, R. & Iaccarino, G. 2005 Immersed boundary methods. *Annu. Rev. Fluid Mech.* **37**, 239–261.

Moin, P. & Kim, J. 1982 Numerical investigation of turbulent channel flow. *J. Fluid Mech.* **118**, 341–377.

Momen, M. & Bou-Zeid, E. 2017 Mean and turbulence dynamics in unsteady ekman boundary layers. *J. Fluid Mech.* **816**, 209–242.

Monkewitz, P. A., Chauhan, K. A. & Nagib, H. M. 2008 Comparison of mean flow similarity laws in zero pressure gradient turbulent boundary layers. *Phys. Fluids* **20**, 105102.

Oke, T. R. 1982 The energetic basis of the urban heat island. *Q. J. R. Meteorol. Soc.* **108**, 1–24.

Oke, T. R., Mills, G., Christen, A. & Voogt, J. A. 2017 *Urban Climates*. Cambridge University Press.

Orszag, S. A. 1970 Analytical theories of turbulence. *J. Fluid Mech.* **41**, 363–386.

Orszag, S. A. & Pao, Y. 1975 Numerical computation of turbulent shear flows. In *Adv. Geophys.*, , vol. 18, pp. 225–236. Elsevier.

Philips, D. A., Rossi, R. & Iaccarino, G. 2013 Large-eddy simulation of passive scalar dispersion in an urban-like canopy. *J. Fluid Mech.* **723**, 404–428.

Pope, S. B. 2000 *Turbulent flows*. Cambridge University Press.

Raupach, M. R., Antonia, R. A. & Rajagopalan, S. 1991 Rough-wall turbulent boundary layers. *Appl. Mech. Rev.* **44**, 1–25.

925 Rotach, M. W. 1993 Turbulence close to a rough urban surface part I: Reynolds stress. *Boundary-Layer*
926 *Meteorol.* **65**, 1–28.
927 Rotach, M. W. 1999 On the influence of the urban roughness sublayer on turbulence and dispersion.
928 *Atmospheric Environment* **33**, 4001–4008.
929 Rotach, M. W., Vogt, R., Bernhofer, C., Batchvarova, E., Christen, A., Clappier, A., Feddersen,
930 B., Gryning, S., Martucci, G., Mayer, H. & others 2005 Bubble–an urban boundary layer
931 meteorology project. *Theor. Appl. Climatol.* **81**, 231–261.
932 Salesky, S. T., Calaf, M. & Anderson, W. 2022 Unstable turbulent channel flow response to spanwise-
933 heterogeneous heat fluxes: Prandtl's secondary flow of the third kind. *J. Fluid Mech.* **934**, A46.
934 Salesky, S. T., Chamecki, M. & Bou-Zeid, E. 2017 On the nature of the transition between roll and cellular
935 organization in the convective boundary layer. *Boundary-Layer Meteorol.* **163**, 41–68.
936 Schmid, M. F., Lawrence, G. A., Parlange, M. B. & Giometto, M. G. 2019 Volume Averaging for Urban
937 Canopies. *Boundary-Layer Meteorol.* **173**, 349–372.
938 Schumann, U. 1973 *Ein Verfahren zur direkten numerischen Simulation turbulenter Strömungen in Platten-*
939 *und Ringspaltkanälen und über seine Anwendung zur Untersuchung von Turbulenzmodellen*. Ges. f.
940 Kernforschung mbH.
941 Stroh, A., Schäfer, K., Frohnapfel, B. & Forooghi, P. 2020 Rearrangement of secondary flow over
942 spanwise heterogeneous roughness. *J. Fluid Mech.* **885**, 1–12.
943 Tian, Geng, Wan, Minping & Chen, Shiyi 2023 A note on friction velocity and viscous effect for idealized
944 urban canopy flows. *Boundary-Layer Meteorology* **187**, 819–829.
945 Tomkins, C. D. & Adrian, R. J. 2003 Spanwise structure and scale growth in turbulent boundary layers. *J.*
946 *Fluid Mech.* **490**, 37–74.
947 Vanderwel, C. & Ganapathisubramani, B. 2015 Effects of spanwise spacing on large-scale secondary
948 flows in rough-wall turbulent boundary layers. *J. Fluid Mech.* **774**.
949 Wang, Z., Bou-Zeid, E., Au, S. K. & Smith, J. A. 2011 Analyzing the sensitivity of wrf's single-layer urban
950 canopy model to parameter uncertainty using advanced monte carlo simulation. *J. Appl. Meteorol.*
951 *Climatol.* **50**, 1795 – 1814.
952 Wangsawijaya, D. D., Baidya, R., Chung, D., Marusic, I. & Hutchins, N. 2020 The effect of spanwise
953 wavelength of surface heterogeneity on turbulent secondary flows. *J. Fluid Mech.* .
954 Willingham, D., Anderson, W., Christensen, K. T. & Barros, J. M. 2014 Turbulent boundary layer flow
955 over transverse aerodynamic roughness transitions: Induced mixing and flow characterization. *Phys.*
956 *Fluids* **26**, 025111.
957 Xie, Z. & Castro, I. P. 2006 LES and RANS for turbulent flow over arrays of wall-mounted obstacles. *Flow*
958 *Turbul. Combust.* **76**, 291–312.
959 Xie, Z., Coceal, O. & Castro, I. P. 2008 Large-eddy simulation of flows over random urban-like obstacles.
960 *Boundary-Layer Meteorol.* **129**, 1–23.
961 Yang, J. & Anderson, W. 2017 Numerical Study of Turbulent Channel Flow over Surfaces with Variable
962 Spanwise Heterogeneities: Topographically-driven Secondary Flows Affect Outer-layer Similarity
963 of Turbulent Length Scales. *Flow Turbul. Combust.* **100**, 1–17.
964 Yang, X. I. A. 2016 On the mean flow behaviour in the presence of regional-scale surface roughness
965 heterogeneity. *Boundary-Layer Meteorol.* **161**, 127–143.
966 Yang, X. I. A. & Meneveau, C. 2016 Recycling inflow method for simulations of spatially evolving turbulent
967 boundary layers over rough surfaces. *J. Turbul.* **17**, 75–93.
968 Zhang, Wen, Zhu, Xiaowei, Yang, Xiang I A & Wan, Minping 2022 Evidence for raupach et al.'s
969 mixing-layer analogy in deep homogeneous urban-canopy flows. *Journal of Fluid Mechanics* **944**,
970 A46.
971 Zheng, X., Montazeri, H. & Blocken, B. 2021 Large-eddy simulation of pollutant dispersion in generic
972 urban street canyons: Guidelines for domain size. *J. Wind. Eng. Ind.* **211**, 104527.