Effect of Sampling Duration on the Estimate of Pollutant Concentration Behind a Heavy-Duty Vehicle: a Large-Eddy Simulation

Jingwei Xie[1], Chun-Ho Liu[1], Yuhan Huang[2,3], Wai-Chuen Mok[1,2,3]

[1]Department of Mechanical Engineering, The University of Hong Kong, Hong Kong

[2]School of Civil and Environmental Engineering, University of Technology Sydney, Australia

[3]Jockey Club Heavy Vehicle Emission Testing & Research Centre, Vocational Training Council, Hong Kong

Highlights

- Large-eddy simulation is conducted for the tailpipe dispersion of an on-road truck
- Fluctuating concentrations are tightly related to the turbulence in the near wake
- Sampling accuracy is affected by sampling duration and fluctuating concentrations
- Concentration reading is more accurate if sampling points are closer to the tailpipe
- Sampling at dominant frequencies is necessary to reduce sampling uncertainty

Effect of Sampling Duration on the Estimate of Pollutant Concentration Behind a Heavy-Duty Vehicle: a Large-Eddy Simulation

Jingwei Xie[1], Chun-Ho Liu[1], Yuhan Huang[2,3], Wai-Chuen Mok[1,2,3]

[1]Department of Mechanical Engineering, The University of Hong Kong, Hong Kong

[2]School of Civil and Environmental Engineering, University of Technology Sydney, Australia

[3]Jockey Club Heavy Vehicle Emission Testing & Research Centre, Vocational Training Council, Hong Kong

Abstract

Plume chasing is cost-effective, measuring individual, on-road vehicular emissions. Whereas, wake-flow-generated turbulence results in intermittent, rapid pollutant dilution and substantial fluctuating concentrations right behind the vehicle being chased. The sampling duration is therefore one of the important factors for acquiring representative (average) concentrations, which, however, has been seldom addressed. This paper, which is based on the detailed spatio-temporal dispersion data after a heavy-duty truck calculated by large-eddy simulation (LES), examines how sampling duration affects the uncertainty of the measured concentrations in plume chasing. The tailpipe dispersion is largely driven by the jet-like flows through the vehicle underbody with approximate Gaussian concentration distribution for $x \leq 0.6h$, where x is the distance after the vehicle and h the characteristic vehicle size. Thereafter for $x \geq 0.6h$, the major recirculation plays an important role in near-wake pollutant transport whose concentrations are highly fluctuating and positively shewed. Plume chasing for a longer sampling duration is more favourable but is logistically impractical in busy traffic. Sampling duration, also known as averaging time in the statistical analysis, thus has a crucial role in sampling accuracy. With a longer sampling (averaging) duration, the sample mean concentration converges to the population mean, improving the sample reliability. However, this effect is less pronounced in long sampling duration. The sampling accuracy is also influenced by the locations of sampling points. For the region $x > 0.6h$, the sampling accuracy is degraded to a large extent. As a result, acceptable sample mean is hardly achievable. Finally, frequency analysis unveils the mechanism leading to the variance in concentration measurements which is attributed to sampling duration. Those data with frequency higher than the sampling frequency are filtered out by moving average in the statistical analyses.

(283 words)

Keywords: Measurement uncertainty; plume chasing; plume meandering; sampling reliability; tailpipe dispersion; turbulent wake

1. Introduction

Vehicular exhaust consists of greenhouse gases and toxic pollutants that would lead to various detrimental health concerns, including respiratory symptoms, disease and cancer (Smit et al., 2019; Tayarani & Rowangould, 2020). In Hong Kong, road transport contributed 20%, 53%, 10% and 13% to the annual emissions of nitrogen oxides (NO_X), carbon monoxide (CO), respirable suspended particulate (PM_{10}) and fine suspended particulate ($PM_{2.5}$), respectively (HKEPD, 2019). Vehicular pollutant is crucial to pedestrian-level air quality because of its close proximity to stakeholders (Smit et al., 2019). Hence, roadside pollutant concentrations are usually much higher compared with ambient ones. The impact is more severe in cities due to huge population. Most air-pollution-related premature deaths are pertinent to vehicular exhaust (Caiazzo et al., 2013). It was estimated that vehicular exhaust resulted in 385,000 premature deaths and around US$ 1 trillion in health damages worldwide in 2015 (Anenberg et al. 2019). In recent years, the increasing traffic volumes and high-rise, dense buildings further worsen the roadside air pollution problem (Huang et al., 2021). Thus, proper control of vehicular exhaust, in particular the reliable identification of heavy on-road emitters, should be enacted.

In view of depreciation, inappropriate maintenance, tampering or breakdown of engine components, in-use vehicles often violate emission regulations (Huang et al., 2020a). Road conditions, such as slope and traffic congestion, influence the emission directly (Davison et al., 2020; Smit & Kingston, 2019), which, however, are hardly modelled in laboratories. In this connection, on-road sampling techniques, including portable emission measurement system (PEMS), on-road remote sensing, exhaust plume chasing, together with tunnel and roadside ambient measurements, have been developed (Huang et al. 2018). Tunnel and roadside ambient measurements are designed for group sampling but not individuals (Smit et al., 2010). Remote sensing is a non-intrusive way to identify heavy on-road emitters. However, its functionality is degraded by the short sampling episode (less than the turbulence timescale, seconds, in vehicular wakes) and the constraints of sampling locations (Wu et al. 2017). Among others, PEMS and plume chasing enable long-term (minutes) emission-data collection for a specific vehicle. The applicability of PEMS for fleetwide measurements is limited by its long turnover time (Franco et al., 2013). Practically, plume chasing realizes the on-road measurements of individual vehicles (Ježek et al., 2015). Another vehicle, which is equipped with rapid-response pollutant analysers, follows the target vehicle for (continuous) data collection during real-world driving conditions. Plume chasing is high throughput (compared with PEMS), facilitating massive on-road data collection for vehicle-fleet exhaust and emission technology (Wang et al. 2020). In view of road safety, a minimum separation is required between the two vehicles (roughly 10 m). Nonetheless, this shortcoming can be overcome by towing a mobile laboratory after the targeted vehicle (Morawska et al., 2007).

Implementation of plume chasing, on the other hand, is complicated by the turbulent wake behind the target vehicle (Yang et al. 2018). After tailpipe exhaust, the plume undergoes dilution in two regimes (Morawska et al. 2007). Within the near-wake regime, the tailpipe discharge momentum and vehicle-induced turbulence dominate the initial, rapid plume dispersion. Afterward, in the far-field regime, the plume dispersion is driven by the prevailing wind (Chan et al., 2001). In view of intermittency, the sampling duration $\Delta\tau$ should be long enough to capture representative statistical properties. The current recommended sampling duration for plume chasing is at least 2 minutes ($\Delta\tau \geq 352h/U_\infty$ where h and U_∞ are the characteristic size and speed of the vehicle, respectively; Wang et al., 2020), which, however, is hardly realizable. Some of the emission parameters, such as engine power and vehicle speeds, are seldom constant. Moreover, the vehicle pair must travel a long distance together for one single test that arouses logistic concern. Apparently, a shorter sampling duration for plume chasing (but reliable readings) would be beneficial. Whereas, there is no study available for the drawback especially the sampling inaccuracy. The uncertainty of plume chasing in response to shortening sampling duration is analysed in this paper to bridge the knowledge gap.

The data sensitivity to sampling duration in long-term, ambient air pollutant measurements has been studied for years. In annual averaging, it was mainly caused by synoptic scales or seasonal factors but not intermittency nor turbulent wakes (Brown and Woods 2014). On top of equipment precision, the measurement accuracy depends on sampling duration $\Delta\tau$ and (unsteady) concentrations ϕ (Ballesta 2005, Brown et al. 2008). Venkatram (2002) examined the effect of sampling duration based on a binomial model of pollutant concentrations. However, the setting was oversimplified that barely represented the real-world situation. The power law

$$\frac{\bar{\phi}_{max}}{\bar{\Phi}} = \left(\frac{\Delta\tau}{\Delta T}\right)^{-p} \tag{1}$$

was suggested to describe the dependence of maximum mean concentration $\bar{\phi}_{max}$ on (a shorter) sampling duration $\Delta\tau$ (peak-to-mean ratio; Santos 2019). Here, $\bar{\Phi}$ is the mean concentration over a longer sampling duration ΔT and p (a real number between 0 and 1) is the exponent (Singer 1963). Theoretically, ΔT is long enough for asymptotically converged $\bar{\Phi}$ though it is hardly defined in non-stationary turbulence (Santos et al. 2009). Nonetheless, it is practically employed as the reference to estimate the maximum mean concentration $\bar{\phi}_{max}$ based on a shorter sampling duration $\Delta\tau$ (Wilson 2010). However, the aforementioned studies have focused on (short-term) maximum but not the uncertainty induced by finite sampling duration. Apart from gaseous pollutants, similar findings have been arrived based on the transport of aeolian sediment (Ellis et al., 2012; Webb et al., 2019). Previous studies related to vehicular exhaust, by and large, have focused on the uncertainty of emission factor (EF; pollutant-to-carbon-dioxide concentration ratio) rather than mean concentration (Tong et al., 2022; Wang et al., 2020; Zheng et al., 2016). EF could avoid temporal variance and turbulence interference, however, its validity is based on stoichiometric combustion. While it is hard to maintain on-road complete combustion at all times, instrumentation issues, such as slow response time and/or short sampling duration, occur very often (Park et al. 2011), deviating from the above assumption. Besides, the determination of EF for individual pollutants from tailpipe emission depends on the reliable pollutant measurements in both plume and ambient (Wen et al., 2019). These uncertainties would degrade the quality of remote sensing. Under this circumstance, statistically robust measures of mean pollutant concentrations in plume chasing are alternative solutions.

Given a time trace of length *T*, the ideal sampling duration $\Delta\tau$ should be long enough such that the mean concentration $\overline{\phi}$ is independent (quasi-steady state). Under homogeneous and stationery turbulence, this independence is achievable provided that the sampling duration is longer than the time scale of turbulence eddies (Santos et al., 2009). In the light of the intermittent vehicular wake in plume chasing, a proper sampling duration is hardly defined. Therefore, it is necessary to quantify the uncertainty of mean-concentration measurements induced by the sampling duration in plume chasing. Large-eddy simulation (LES) is adopted in this paper so the influence other than vehicular wake is excluded.

This paper is organized as follows. Section 1 (this section) is the introduction. Next, the LES setup and the statistical methods for sampling uncertainty are described in Section 2. In Section 3, the fluctuating concentrations behind the tailpipe within the near wake are analysed. Their power spectra are then employed to elucidate the relationship between uncertainty and sampling duration. Finally, the findings and the conclusions are summarized in Section 4.

2. Methodology

2.1 Mathematical Model

LES is an appealing tool investigating the spatio-temporal dynamics of flows and pollutant dispersion (eddy-resolving). It explicitly calculates much of the conservation of momentum and mass while models small portion of subgrid-scale (SGS) fluxes at a reasonable computational load (Chan et al. 2008). LES is advantageous in terms of calculating the unsteadiness and intermittency of flows and tailpipe dispersion (Li et al. 2007) so is adopted.

The LES used in this paper is the open-source computational fluid dynamics (CFD) code OpenFOAM 6 (Weller et al. 1998). It was validated in our previous study (Xie et al. 2020). The SGS motions are modelled by the Smagorinsky model (Smagorinsky 1963). The computational domain sizes 31.8*h* (streamwise) × 3.9*h* (spanwise) × 10.3*h* (height) while the model of the heavy-duty vehicle sizes 3.86*h* (length) × 0.89*h* (width) × 1.09*h* (height). The logarithmic law-of-the-wall (log-law) is used to model the flows near all the solid boundaries (the ground and the truck body). At the domain top and the spanwise extent, Neumann boundary conditions (BCs; $\partial \overrightarrow{\psi} / \partial \vec{n} = 0$ where $\vec{n}$ is the normal to the boundary) for both flows and pollutant transport are applied. Dirichlet BCs of constant wind speed U_∞ and zero pollutant $\phi = 0$ are prescribed at the inflow. Turbulence is not prescribed at the inflow but is only induced by the flows around the vehicle. This configuration helps focus on tailpipe dispersion driven by wake-induced turbulence. An open BC is adopted at the outflow so all the pollutants are removed from the computational domain without any rebound. A point source of pollutant with a constant emission rate $\dot{Q}$ is placed at the tailpipe exit ($x = y = z = 0$) to simulate vehicular exhaust. Here, x, y and z are the streamwise, spanwise and vertical coordinates, respectively. The spatial domain is discretized into 3.38 million unstructured hexahedra using the mesh generation utility *snappyHexMesh* (OpenFOAM 2018). Its mesh is refined toward the vehicle surfaces and the ground. The minimum and maximum cell volume is about $10^{-7}h^3$ and $10^{-2}h^3$, respectively. The second-order-accurate finite volume method (FVM) is used to discretize the gradient, divergence and Laplacian terms. The time increment is $\Delta t = 0.15h/U_\infty$ and the LES is integrated for $T = 510h/U_\infty$ in the time domain using the implicit, second-order-accurate backward differencing.

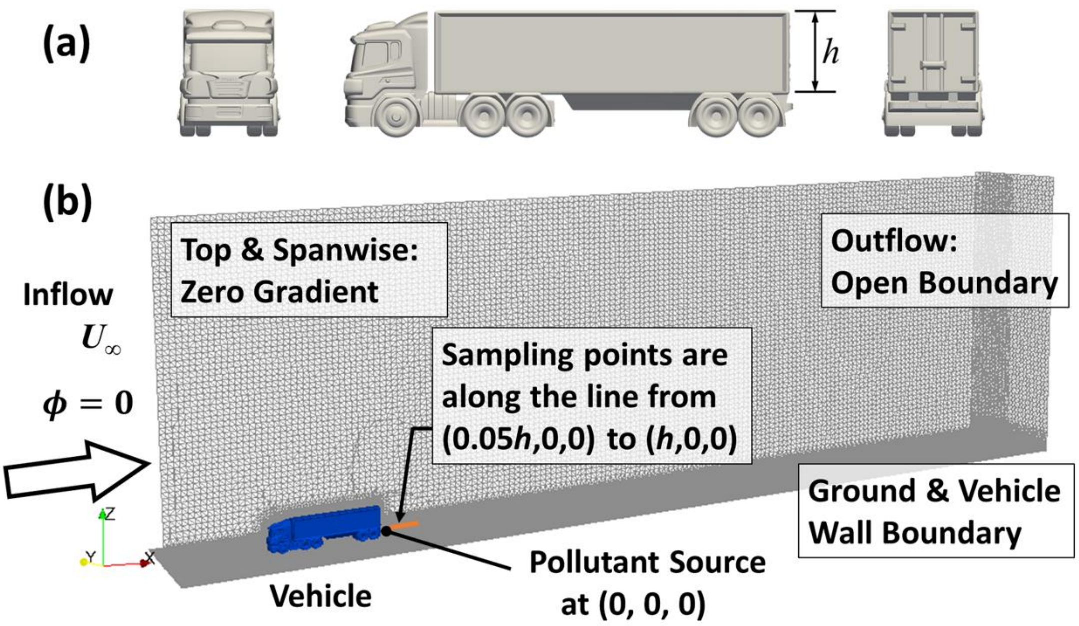

Fig. 1 (a) Digital model of the heavy-duty vehicle together with (b) computational domain and boundary conditions.

2.2 Statistical Method

The gaseous pollutant considered in this paper is passive and chemically inert that could be taken as carbon dioxide CO_2. It is diluted by vehicle-induced turbulence in the wake with characteristic scales of length h and velocity U_∞. In favour of detection sensitivity, sampling within the near wake is suggested where the concentrations are almost ten times larger than those in the far field (Xie et al. 2020). The sampling locations are aligned along the sampling line in the streamwise direction at the tailpipe level from (0.05h, 0, 0) to (h, 0, 0) to mimic plume chasing (Huang et al., 2020b). The definition of variables used in this paper is summarized in Table. S1.

2.2.1 Moving Average

The concentrations ϕ are normalized by the characteristic pollutant concentration Φ_0 (= $\dot{Q}/U_\infty h^2$). Time traces of dimensionless pollutant concentration C_i (= ϕ_i/Φ_0) are probed from the LES dataset where the subscript i is the index of the data sample. Their moving average

$$\overline{C}(\Delta\tau)_i = \frac{1}{n}\sum_{j=0}^{n-1} C_{i+j} \tag{2}$$

represents the sample mean over the sampling duration $\Delta\tau$ (= $(n\text{-}1)\times\Delta t$, where n (> 1) is the number of data points within $\Delta\tau$). Here, the overbar $\overline{\varphi}$ denotes time average. The sampling duration considered is in the range of $10\Delta t \leq \Delta\tau \leq 320\Delta t$. It is noteworthy that $\overline{C}_{i=0}\left(\Delta\tau = T = 510\,h/U_\infty\right)$ is the population mean $\overline{C}$.

2.2.2 Uncertainty Analysis

The relative deviation between the mean of each data subset (moving average) with sampling duration $\Delta\tau$ and the population is

$$\delta(\Delta\tau)_i = \frac{\overline{C}(\Delta\tau)_i - \overline{C}}{\overline{C}}. \tag{3}$$

To consolidate the uncertainty induced by the data subsets with sampling duration $\Delta\tau$ the coefficient of variance

$$CV(\Delta\tau) = \left\{\frac{1}{N}\sum_{i=0}^{N-1}\left[\delta(\Delta\tau)_i\right]^2\right\}^{1/2} = \left\{\frac{1}{N}\sum_{i=0}^{N-1}\left[\frac{\overline{C}(\Delta\tau)_i - \overline{C}}{\overline{C}}\right]^2\right\}^{1/2} \tag{4}$$

is adopted where N is the number of data subsets with sample mean $\overline{C}(\Delta\tau)_i$. It is noteworthy that $CV(\Delta\tau = \Delta t)$ is equal to the fluctuating concentration intensity I.

For demonstration purposes, the sample mean $\overline{C}(\Delta\tau)_i$ is defined acceptable in this paper provided that its tolerance is within ±15% compared with the population mean $\overline{C}$, i.e. $|\delta(\Delta\tau)_i| \leq 15\%$. To obtain a sample mean with specified confidence, the sampling duration $\Delta\tau$ should be long enough so that more than 90% of the data in a new dataset $\overline{C}(\Delta\tau)_i$ are within the acceptable deviation. Therefore, the fraction of acceptable sample mean

$$k(\Delta\tau) = \frac{\text{No. of data subsets in which } \left|\delta(\Delta\tau)_i\right| \leq 15\%}{N} \tag{5}$$

is studied. The specific criterion aforementioned was adopted elsewhere (Li et al. 2017).

3. Results and Discussion

3.1 Fluctuating Concentration

Fig. 2a shows the population mean $\overline{C}$, maximum C_{max} and minimum C_{min} of dimensionless pollutant concentration based on the entire LES dataset of sampling duration T. Liu et al. (2011) found that a dimensionless averaging time T^* (= $T \times U_{ref}/L_{ref}$ where U_{ref} and L_{ref} are the reference scales of wind speed and length, respectively) in the range of 200 to 400 is sufficient for reliable population mean $\overline{C}$ around a high-rise building. The current dimensionless averaging time ($T^* = T \times U_\infty/h = 510$) well exceeds the requirement.

The initial dilution directly behind the tailpipe is dominated by the jet-like flows from the truck underbody. Simultaneously, the spanwise and vertical dispersion is attributed to the turbulence and instability generated by the shear within the jet-like flows (Xie et al. 2020). In this connection, the population mean concentration $\overline{C}$ decreases monotonically for $x \leq 0.6h$ then keeps at a low level thereafter (Fig. 2a). For example, the population mean concentration

$\overline{C}$ at x = 0.4h is almost 14 times larger than that at x = 0.8h. The region close to x = 0.8h is characterized by the strong entrainment and the upward flows within the major recirculation, driving the jet-like flows back to the vehicle body (Fig. S1a and S1b). A similar sharp drop in concentration (an order of magnitude) after a vehicle was also reported by Chang (2009).

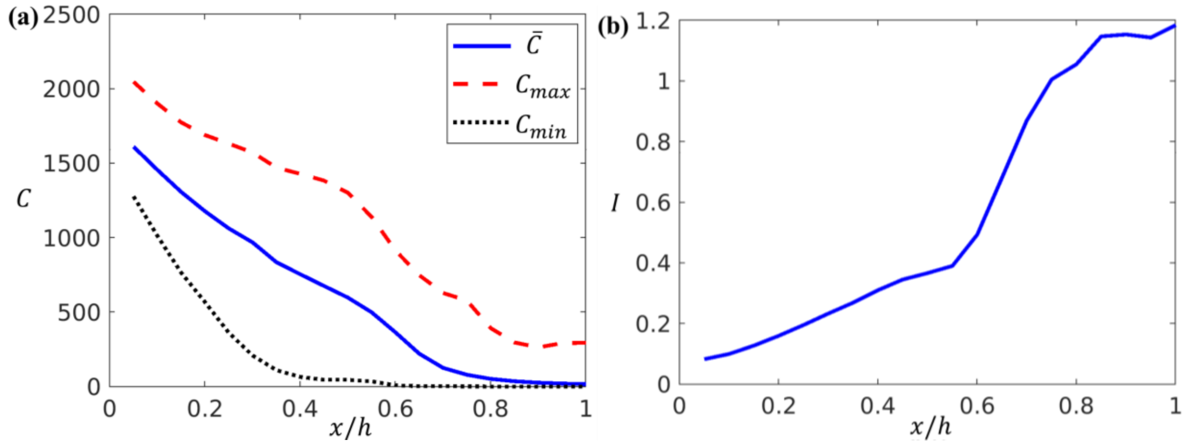

Fig. 2 Longitudinal profiles of (a) population mean $\overline{C}$, maximum C_{max} and minimum C_{min} of dimensionless pollutant concentration together with (b) fluctuating concentration intensity I along the sampling line.

In addition to the upward flows within the major recirculation, the fresh-air entrainment from the sides of the vehicle quickly dilutes the pollutant (Fig. S1a). It is concurred by the close-to-zero minimum concentration C_{min} for $0.6h \leq x \leq h$ (Fig. 2a). Likewise, the maximum concentration C_{max} elevates for $x \geq 0.6h$ though the population mean concentration $\overline{C}$ keeps decreasing behind the tailpipe (Fig. 2a). For example, the maximum concentration C_{max} is almost 6 times larger than the population mean $\overline{C}$ at x = 0.8h. It is in turn suggested that the pollutant dispersion behind x = 0.6h is intermittent in response to the strong shear-generated turbulence within the major recirculation.

The fluctuating concentration intensity increases gradually in the streamwise direction for $x \leq 0.6h$ then soars thereafter, resulting in an elevated level ($I \geq 1$) towards the end of the near-wake region (Fig. 2b). The peaked I is over unity so the fluctuations are comparable to the mean $\overline{C}$. The initial increase in I is attributed to the recirculating flows. Moreover, the larger eddies augment turbulent mixing (widening plume coverage). A similar plume development with increasing I in the near-source region over open terrain was reported elsewhere (Yee and Biltoft, 2004). In the region $0.8h \leq x \leq h$, the flow entrainment and the shear in-between the near and far wakes contribute much to the elevated fluctuating concentrations.

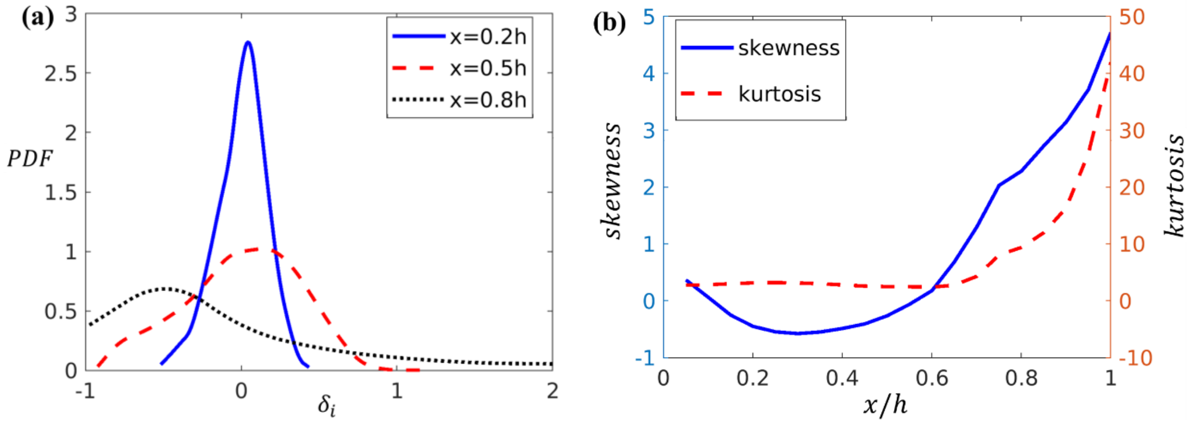

Fig. 3 (a) Probability density function (PDF) of the relative deviations δ_i of instantaneous pollutant concentrations at the sampling points $x = 0.2h$, $0.5h$ and $0.8h$ directly behind the tailpipe. (b) Skewness and kurtosis of δ_i along the sampling line.

The probability density functions (PDFs) of relative deviation $\delta(\Delta\tau)_i$ at $x = 0.2h$, $0.5h$ and $0.8h$ are depicted in Fig. 3a. In view of the gradually augmented fluctuating concentration, the range of the PDF spreads with increasing distance behind the truck. Besides, the PDF of

$\delta(\Delta\tau)_i$ at $x = 0.2h$ and $0.5h$ is close to Gaussian distribution but at $x = 0.8h$ is positively skewed. As such, most of the measured instantaneous concentrations at $x = 0.8h$ are lower than the population mean $\overline{C}$. The asymmetric PDFs are concurred by the skewness and kurtosis of $\delta(\Delta\tau)_i$ which are close to zero and 3, respectively, for $x \leq 0.6h$ (Fig. 3b). Thereafter, the increasing skewness and kurtosis indicate the positively skewed and leptokurtic PDFs. The sharp change in maximum relative deviation δ_{max} from $x = 0.5h$ to $x = 0.8h$ is also notable (Fig. 3a). It is attributed to the turbulence generated by the major recirculation and the entrainment. Likewise, the minor increase in maximum δ_{max} (the upper range of δ_i shown in Fig. 3a) for $0.2h \leq x \leq 0.5h$ is attributed to the plume development driven by the recirculation.

3.2 Uncertainty Analysis

The mean concentrations at different sampling duration $\overline{C}(\Delta\tau)_i$ are calculated by moving average based on the instantaneous (dimensionless) concentration C_i. The averaging time $\Delta\tau$, which is also the sampling duration, cannot be too long, owing to the finite total sampling period T. Otherwise, the data are insufficient for the analysis of sample mean, leading to substantial inaccuracy. The sampling duration is limited to $\Delta\tau \leq 320\Delta t$ that is roughly 9% of the total sampling period T. It also satisfies the requirement of averaging time $\Delta\tau$ that is less than one-third of the entire time trace T (Janik et al, 2012). Fig. 4 compares the time traces of the relative deviations $\delta(\Delta\tau)_i$ for different sampling duration at $x = 0.8h$. It is found that the instantaneous relative deviations δ_i are highly fluctuating whose maximum is up to $\delta_{max} = 650\%$. It is in turn implied that the maximum concentration C_{max} is up to 6.5 times larger than the population mean $\overline{C}$. The variation of the relative deviations for sample mean $\delta(\Delta\tau > \Delta t)_i$ is less than that of the instantaneous value $\delta(\Delta\tau = \Delta t)_i$. Increasing the sampling duration $\Delta\tau$ reduces the uncertainty of the sample mean concentrations. For example, the maximum relative deviation $\delta(\Delta\tau)_{max}$ for $\Delta\tau = 160\Delta t$ is only 60%. The improved accuracy is attributed to the low pass of sample mean $\overline{C}(\Delta\tau)_i$ with period shorter than the averaging time (sampling duration) $\Delta\tau$ by applying moving average. The average of sample mean concentrations $\overline{C}(\Delta\tau)_i$ obtained by moving average is very close to the population mean $\overline{C}$. However, parts of the fluctuating signal, especially those short-term extremities, are filtered out (Fig. 4).

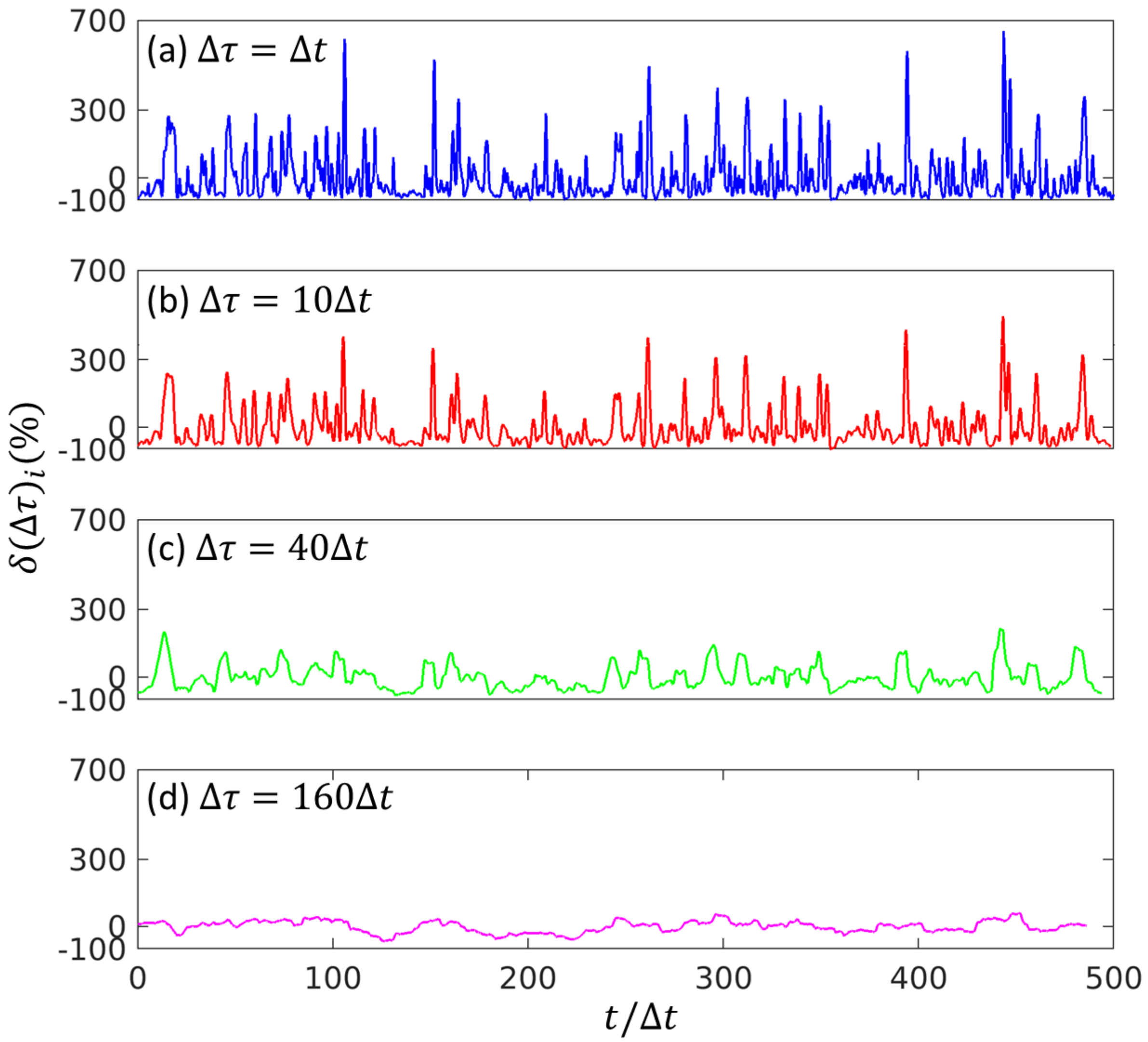

Fig. 4 Time traces of the relative deviation $\delta(\Delta\tau)_i$ of pollutant concentrations. Sampling duration $\Delta\tau$ = (a) Δt (instantaneous values); (b) $10\Delta t$; (c) $40\Delta t$; and (d) $160\Delta t$ by moving average behind the tailpipe at $(x, y, z) = (0.8h, 0, 0)$.

Cumulative density functions (CDFs) measure the coefficient of variation $CV(\Delta\tau)$ by the slope of the curve core (Fig. 5). Steeper gradient suggests a smaller coefficient of variation, and vice versa. Moreover, it depicts the maximum relative deviation $\delta(\Delta\tau)_{max}$ (along the x-axis of Fig. 5) at which the CDF reaches unity (Santos et al., 2005). Like Fig. 3, the CDF (Fig. 5) shows that the relative deviation for $\Delta\tau = \Delta t$ at $x = 0.8h$ is positively skewed, signifying frequent low-concentration and occasional high-concentration events. The sampling duration $\Delta\tau$ has a strong influence on the maximum relative deviation $\delta(\Delta\tau)_{max}$ which decreases by almost ten times when the sampling duration is increased from $\Delta\tau = \Delta t$ to $\Delta\tau = 160\Delta t$. As shown in Fig. 5, the curve core steepens with extending sampling duration $\Delta\tau$, indicating a smaller coefficient of variation $CV(\Delta\tau)$ as well as more accurate sample mean concentrations $\overline{C}(\Delta\tau)_i$. It is noticed that, with increasing sampling duration $\Delta\tau$, the CDF gradually converges close to normal distribution that is in line with that reported elsewhere (Venkatram 2002).

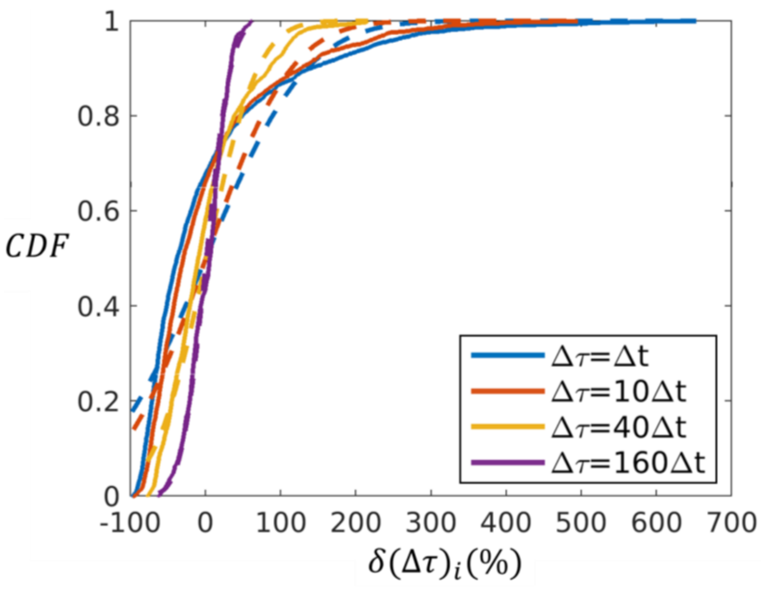

Fig. 5 Cumulative density functions (CDFs) of the relative deviation $\delta(\Delta\tau)_i$ of instantaneous concentrations ($\Delta\tau = \Delta t$) and the sample mean concentrations averaged over $\Delta\tau = 10\Delta t$, $40\Delta t$ and $160\Delta t$ at $(x, y, z) = (0.8h, 0, 0)$. Dashed lines denote the CDFs of corresponding normal distribution with the same mean and standard deviation.

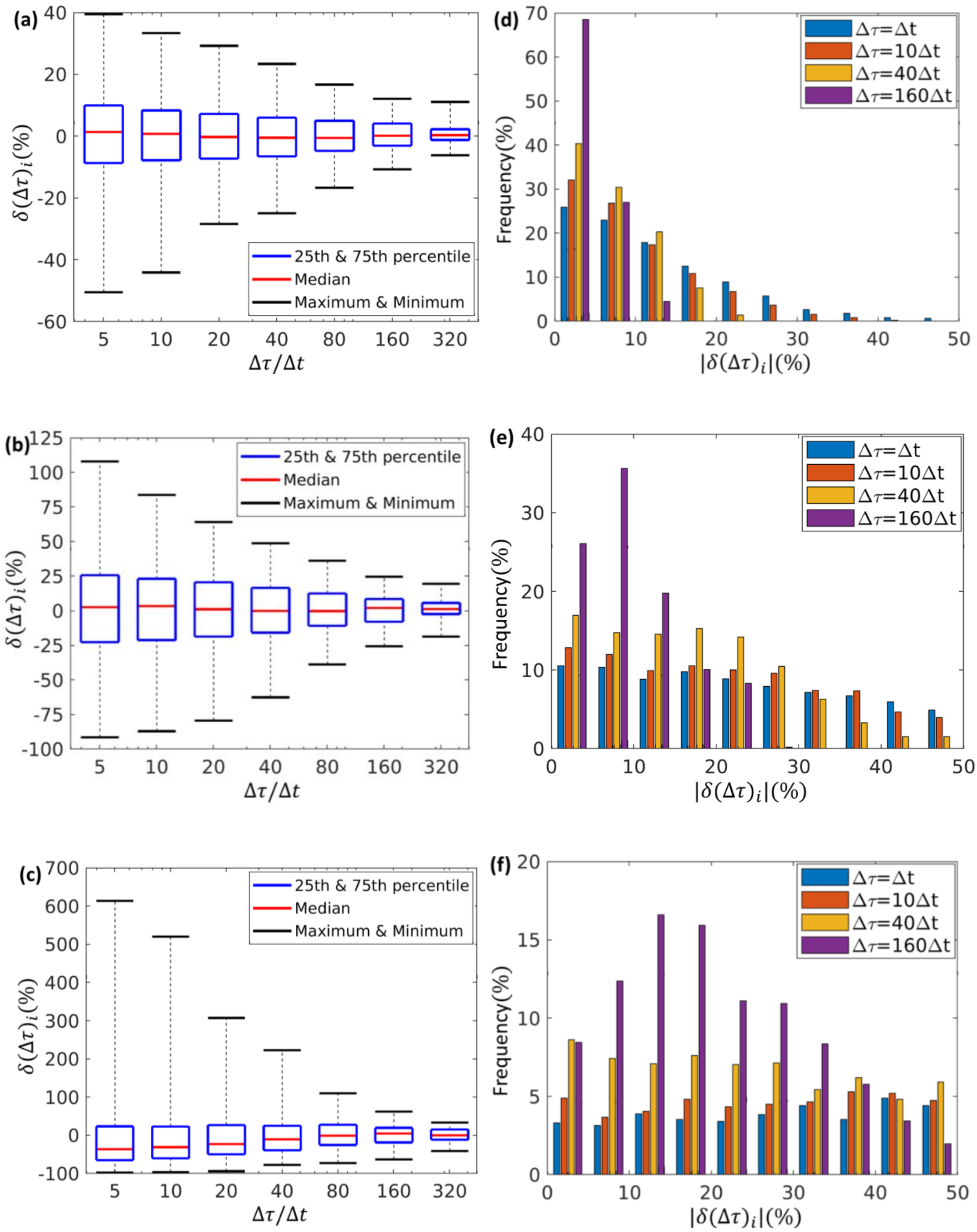

Fig. 6 Relative deviations $\delta(\Delta\tau)_i$ of sample mean concentrations obtained by moving average over different sampling duration $\Delta\tau$ at x = (a) 0.2h; (b) 0.5h; and (c) 0.8h. Also shown are the frequency distribution of the absolute relative deviation $|\delta(\Delta\tau)_i|$ at x = (d) 0.2h; (e) 0.5h; and (f) 0.8h with $\Delta\tau = \Delta t$, 10Δt, 40Δt, and 160Δt.

In Fig. 6a (x = 0.2h) and 6b (x = 0.5h), the relative deviation $\delta(\Delta\tau)_i$ is almost normally distributed considering its symmetry about zero (Fig. 3a). As shown in Fig. 6c, the positive tail diminishes with increasing sampling duration $\Delta\tau$ so the relative deviation $\delta(\Delta\tau)_i$ at x = 0.8h tends to be normally distributed. It is also found that increasing the sampling duration $\Delta\tau$ narrows the range of the sample mean concentrations $\overline{C}(\Delta\tau)_i$. On the contrary, the extreme sample mean concentration $C(\Delta\tau)_{max}$ rises sharply with shortening sampling duration.

Figs. 6d, 6e and 6f depict the frequency distribution of the absolute relative deviation $|\delta(\Delta\tau)_i|$ at the three sampling locations. As shown in Fig. 2a, the fluctuating concentration intensity I at x = 0.2h and 0.5h is lower than that at x = 0.8h. It is thus implied that at x = 0.2h and 0.5h (Fig. 6d, 6e), the fraction of the data close to the population mean $\overline{C}$, such as $|\delta(\Delta\tau)_i| \leq 10\%$, is much higher than that at x = 0.8h (Fig. 6f) for the same sampling duration $\Delta\tau$. Taking the data subset with sampling duration $\Delta\tau = 160\Delta t$ as an example, almost 90% and 60% of the absolute relative deviation $|\delta(\Delta\tau)_i|$ are less than 10% at x = 0.2h and 0.5h, respectively. On the contrary, only 20% of $|\delta(\Delta\tau)_i| \leq 10\%$ at x = 0.8h where the fluctuating concentration intensity I is much larger. However, the fraction of instantaneous dimensionless concentrations $\overline{C}(\Delta\tau = \Delta t)$ within the same range is only 50% at x = 0.2h, 20% at x = 0.5h and 10% at x = 0.8h. It is hence suggested that for the region close to the tailpipe with low fluctuating concentration intensity I, more accurate sampling is achievable using a shorter sampling duration $\Delta\tau$. The difference in the sampling accuracy $|\delta(\Delta\tau)_i|$ between x = 0.2h and x = 0.8h reduces after applying a longer sampling duration $\Delta\tau$. It is because more short-term extremities are filtered out.

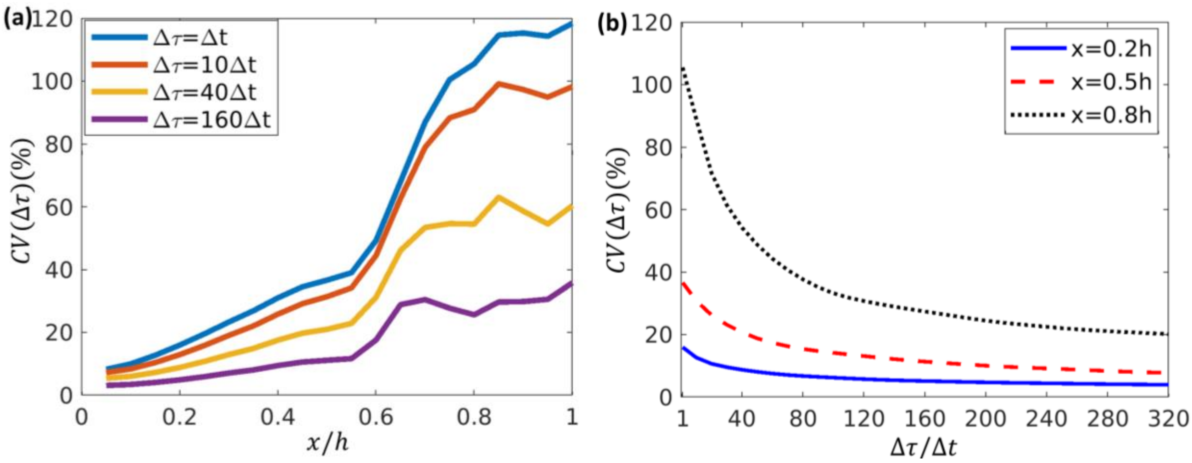

Fig. 7 Coefficient of variance $CV(\Delta\tau)$ for instantaneous concentration ($\Delta\tau = \Delta t$) and sample mean with (a) sampling duration $\Delta\tau = 10\Delta t$; $40\Delta t$; and $160\Delta t$ along the sampling line together with (b) different sampling duration $\Delta\tau$ at the selected sampling locations.

It is observed that increasing the sampling duration $\Delta\tau$ reduces the uncertainty (Fig. 7a) in the entire major recirculation. The improvement is more obvious in the range of $0.6h \leq x \leq h$ where the concentrations are highly fluctuating in response to the underbody flows and sideward entrainment (Fig. S1). Although the sampling duration is extended to $\Delta\tau = 160\Delta t$, the uncertainty in sample mean concentration for $x \geq 0.6h$ soars. Such a phenomenon is attributed to the instantaneous fluctuating dimensionless concentration C_i - $\overline{C}$ in $0.6h \leq x \leq h$ that is tightly driven by the eddies in the major recirculation. Their effect is not negligible unless the sampling duration $\Delta\tau$ is longer than the turbulence time scales. The coefficient of variance $CV(\Delta\tau)$ decreases with increasing sampling duration $\Delta\tau$ (Fig. 7b). Its diminishing gradient indicates the importance of sampling duration $\Delta\tau$ to accuracy. Given a sufficiently long $\Delta\tau \geq 160\Delta t$, the curves flatten so the sampling uncertainty is negligible. Further increasing the sampling duration $\Delta\tau$, however, leads to costly measurement but limited accuracy improvement.

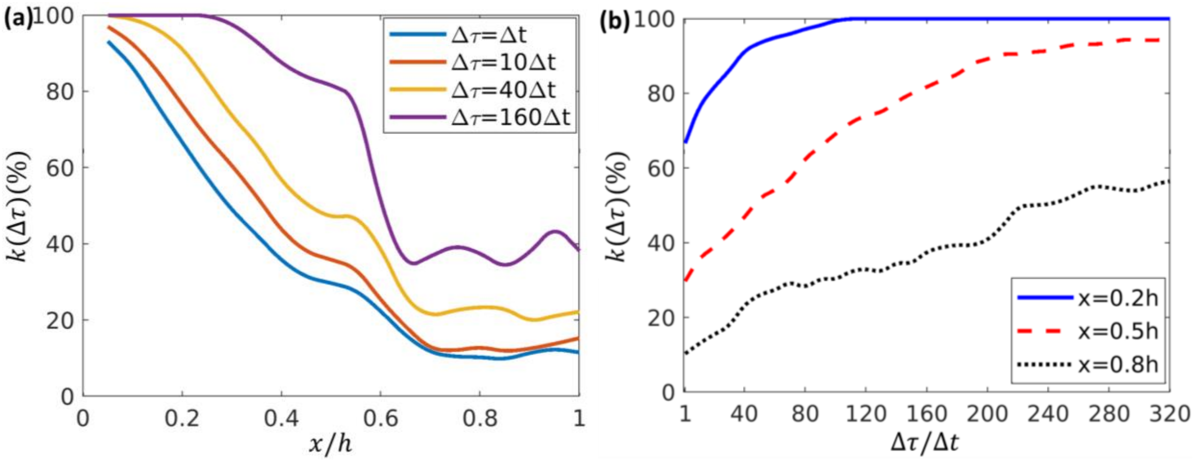

Fig. 8 Fraction $k(\Delta\tau)$ of instantaneous concentrations and sample mean concentrations whose relative deviations are within ±15% ($|\delta(\Delta\tau)_i| \leq 15\%$). (a) Along the sampling line behind the tailpipe and (b) with different sampling duration $\Delta\tau$ at the sampling locations.

A key question is how long the sampling duration $\Delta\tau$ is sufficient for reliable plume chasing. For demonstration purposes, the range of reliable sample mean concentration $\overline{C}(\Delta\tau)_i$ is herewith defined as ±15% of the population mean $\overline{C}$, i.e. $|\delta(\Delta\tau)_i| \leq 15\%$. The fraction $k(\Delta\tau)$ of data sample within this range is compared in Fig. 8. The minimum sampling duration enabling a reliable sample mean is defined as the shortest averaging time under which $k(\Delta\tau) \geq 90\%$ (Li et al. 2017). The fraction $k(\Delta\tau)$ for different sampling duration $\Delta\tau$ decreases in the streamwise direction until $x = 0.6h$ that remains at a low level thereafter (Fig. 8a). When adopting the (longer) sampling duration of $\Delta\tau = 160\Delta t$, the data subset for sampling locations $x \leq 0.4h$ fulfils the 90% criteria but not for $x \geq 0.6h$ where $k(\Delta\tau)$ is only about 40%. Therefore, plume chasing targeting within $x \leq 0.6h$ enables more accurate measurements for the same sampling duration $\Delta\tau$. Indeed, the uncertainty could be further reduced if the sampling points are closer to the tailpipe. The acceptable sampling duration $\Delta\tau$ is shortened to $40\Delta t$ at $x = 0.2h$ that is reduced by 5 times compared with $220\Delta t$ at $x = 0.5h$ (Fig. 8b). Whereas, the 90%

criterion is not achievable at $x = 0.8h$ for the range of sampling duration $\Delta\tau$ tested. Unlike the other two sampling points, the dispersion at $x = 0.8h$ is more affected by energetic eddies whose influence is hardly eliminated by averaging over a finite sampling duration. It is noteworthy that the sampling duration in the current sensitivity test is limited to $\Delta\tau \leq 320\Delta t$ to ensure validity (Fig. 8b). Otherwise, the number of sample mean would be insufficient, degrading the subsequent error analysis of sample mean. Similar concern was reported elsewhere (Janik et al. 2012).

3.3 Fast Fourier Transform

Fast Fourier Transform (FFT) is adopted in this study to investigate the frequency characteristics of the tailpipe dispersion within the near wake. It transforms the data from time domain to frequency domain, providing the power associated with different frequencies (Richards 2003). Fig. 9 shows the power spectra of relative deviations of instantaneous concentrations $\delta(\Delta\tau)_i$ at the three sampling locations. The frequency is normalized in the form of Strouhal number $St = fd/U_\infty$ where f is the frequency and d the trunk width (McArthur et al. 2016). The power generally increases with increasing frequency in the low-frequency regime, reaches its maximum in $0.03 \leq St \leq 0.1$ and decreases thereafter. It finally keeps at a low level ($\leq 10^{-3}$) for $St \geq 1$. The unsteadiness in concentration is directly affected by the flow intermittency. Therefore, the spectra obtained from the concentration data help identify the dominant scales in the near-wake region.

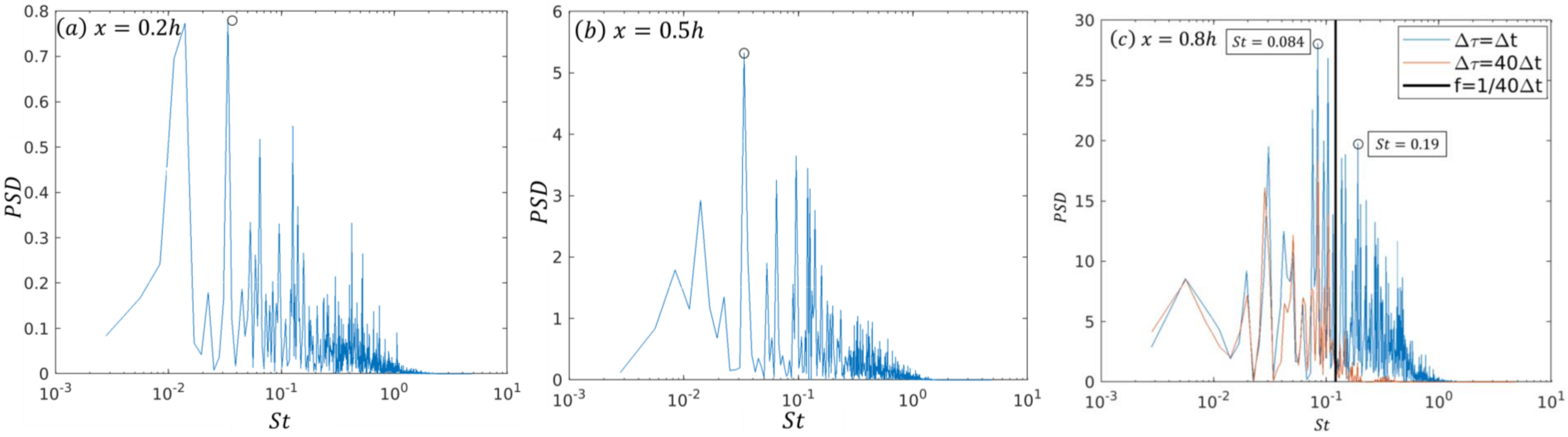

Fig. 9 Power spectra of relative deviation $\delta(\Delta\tau = \Delta t)_i$ for instantaneous concentrations at the selected sampling locations at x = (a) 0.2h; (b) 0.5h; and (c) 0.8h. Also shown in (c) are the power spectra of relative deviations $\delta(\Delta\tau = 40\Delta t)_i$ for sample mean concentrations over $\Delta\tau = 40\Delta t$ obtained by moving average and the corresponding sample frequency $f = 1/\Delta\tau = 1/40\Delta t$ (dark solid line). The primary and secondary peaks for $\Delta\tau = \Delta t$ are highlighted (circles in (c)).

The pollutant transport for $x \leq 0.6h$ is mainly driven by the jet-like flows from the vehicle underbody (Fig. S1a). Therefore, the low-frequency motions, which is peaked at St = 0.033, are more energetic (Fig. 9a and 9b). At $x = 0.8h$, a primary peak and a secondary peak are shown at St = 0.084 and St = 0.19, respectively (Fig. 9c). Alike Volpe et al. (2015), the two peaks are attributed to the wake pumping of major recirculation (St = 0.08) and the vortex shedding initiated at the two vertical edges of the vehicle (St = 0.19). Wake pumping is the key component in the unsteady recirculation. It is induced by the wake lengthening and shortening in response to the increasing entrainment into the near wake together with the vortex shedding induced by the major recirculation (Richards 2003, Rao et al. 2019).

The power spectra of relative deviation of the instantaneous concentrations δ_i and the sample mean concentrations $\delta(\Delta\tau = 40\Delta t)_i$ are compared to examine the effect of moving average in the frequency domain (Fig. 9c). In fact, moving average applies a low-pass filter on the data in the time domain. It damps out the signal with frequency higher than $1/\Delta\tau$ (cut-off frequency). The power of low-frequency signal is less affected. The power spectral density (PSD) of sample mean obtained by moving average decreases by a factor of $\sin^2(\pi f\Delta\tau)/(\pi f\Delta\tau)$ compared with the original ones. The difference diminishes when the product of frequency f and duration $\Delta\tau$ is larger than unity (Arya, S. P., 1999). As shown in Fig. 9c, the sampling duration, which can capture the dominant frequency at $x = 0.8h$, should be higher than $70\Delta t$. Such sampling duration is roughly the transition point for $CV(\Delta\tau)$ to stabilize with increasing $\Delta\tau$ (Fig. 7). More high-frequency signals could be averaged out with longer $\Delta\tau$. If $\Delta\tau$ is long enough to capture the dominant frequency (signals with most energy), the improvement in measurement accuracy would be slowed down. The results suggest that a short sampling duration $\Delta\tau$ only captures the high-frequency signal. Sampling signal of dominant frequency is crucial to the measurement accuracy of mean concentrations.

It is known that the spectra of fluctuating concentration depend on the distance from the point source (Mylne and Mason 1991). In the vicinity of a tailpipe, the turbulence characteristic length scale (wake-induced) is usually longer than the plume width (Xie et al. 2007), resulting in plume meandering. Eddies dominate the transport as long as the plume coverage is comparable to or larger than the turbulence characteristic length scale. Trunk dimension is the characteristic length scale in the near-wake region after a heavy-duty vehicle. The plume transport is mainly driven by the vehicular wake, especially the wake pumping and the vortex shedding from the longer trunk edges.

3.4 Implication to plume chasing

The results reported above collectively show that, in plume chasing deployment, increasing the sampling duration helps filter out parts of fluctuating signal as well as reduce sampling uncertainty. The sample mean often fluctuates substantially for a short sampling duration. As such, it would possibly deviate much from the population mean. The sample mean varies less with extending sampling duration. Therefore, a longer sampling duration is more favourable for a reliable sample mean as well as the tailpipe emission. As shown in the frequency analysis, the sampling accuracy could be affected by the dominant frequency. If the sampling duration is long enough to capture the signal at the dominant frequency (inverse of sampling duration smaller than the dominant frequency), the sampling accuracy would be improved substantially. An even longer sampling duration, which is longer than the inverse of the dominant frequency, however, would slow down the improvement in sampling accuracy.

The threshold sampling duration is defined as the shortest time period over which the fraction of data sample satisfying $|\delta(\Delta\tau)_i| \leq 15\%$ reaches 90%. The threshold sampling durations are $40\Delta t$ and $220\Delta t$, respectively, at $x = 0.2h$ and $x = 0.5h$ after the truck. However, it is much longer at $x = 0.8h$ that is beyond the range of sampling duration being investigated in this paper. For a 4-m-high truck driving at a speed of 10 m sec^{-1}, the sampling duration should be at least 2.4 sec at $x = 0.8$ m or 13.2 sec at $x = 2$ m. A shorter sampling duration is required to obtain a reliable sample mean in the region close to the tailpipe. In view of the elevated fluctuating concentration intensity in the region after $x = 0.6h$, a longer sampling duration is needed. Therefore, in the plume chasing after a heavy-duty vehicle, it is suggested to sample within the region $x \leq 0.6h$. Moreover, the measurements would be more reliable if the sampling point is closer to the tailpipe.

4. Conclusions

In order to investigate the effect of sampling duration $\Delta\tau$ on vehicular pollutant measurement, LES is carried out for a heavy-duty vehicle to collect the spatio-temporal behaviours of pollutant concentrations at the tailpipe level within the near-wake region. The sampling uncertainty is then examined by statistical analysis. Based on the results reported above, the conclusions could be drawn as follows.

- Within the near-wake region, the fluctuating concentration intensity I increases slowly for $x \leq 0.6h$. It experiences a sharp increase thereafter due to the augmentation of fluctuating concentration by the major recirculation. Afterward, a positively skewed distribution after

$x = 0.6h$ is developed, indicating that the instantaneous concentrations could have a notable deviation from the population mean $\overline{C}$.

- A longer sampling duration would result in the loss of the high-frequency fluctuating components, leaving the low-frequency signal. Thus, the coefficient of concentration variance $CV(\Delta\tau)$ would decrease with increasing sampling duration (i.e. more accurate sample mean). However, the improvement in sampling accuracy gradually diminishes if the sampling duration is longer than $160\Delta t$. It is noteworthy that, even a long sampling duration is adopted, the sampling accuracy degrades for $x \geq 0.6h$ because of the elevated fluctuating concentration intensity I.
- Sampling duration also affects the distribution of sample mean concentration $\overline{C}(\Delta\tau)_i$. For a longer sampling duration, the maximum sample mean $\overline{C}(\Delta\tau)_{max}$ and the minimum $\overline{C}(\Delta\tau)_{min}$ approach the population mean $\overline{C}$, improving the sampling accuracy. However, the improvement lessens for prolonging sampling duration $\Delta\tau$. Increasing the sampling duration helps the distribution of sample mean $\overline{C}(\Delta\tau)_i$ that is alike the normal distribution with increasing sampling duration.
- The minimum sampling durations are $40\Delta t$ at $x = 0.2h$ and $220\Delta t$ at $x = 0.5h$. However, at $x = 0.8h$, the minimum sampling duration is even longer than the entire time period T being collected in this study. A shorter sampling duration is needed to acquire a reliable sample mean concentration in the region close to the tailpipe.
- From the FFT analysis, it is found that the variance of sample mean is attributed to the signal with frequency lower than the sampling frequency (= $1/\Delta\tau$). This indicates that sampling at the dominant frequencies could reduce the sampling uncertainty to a large extent.

The aforementioned findings collectively enrich our understanding of how the sampling uncertainty of plume chasing varies with the sampling duration $\Delta\tau$ behind the tailpipe within the near-wake region. In this study, the variation of uncertainty is mainly attributed to the pollutant source and the turbulence of vehicular wake. In practice, vehicular emission in the wake region could also be affected by some other factors, such as engines, acceleration, and brakes. This paper only focuses on the turbulence effect induced by the vehicle body. Further studies could combine with the sampling uncertainty in field measurements to advance the contribution from different factors. Although it is advised to sample pollutant concentrations within the near-wake region, practically the safe distance apart should be at least 10 m ($\geq 2h$ in this paper). In this connection, it is worthy to look into the sampling accuracy beyond the near wake.

Acknowledgment

This research is conducted in part using the research computing facilities and/or advisory services offered by Information Technology Services (ITS), The University of Hong Kong (HKU). Technical support from Ms. Lilian Y.L. Chan, Mr. W.K. Kwan and Mr. Bill H.T. Yau is appreciated. This work was supported by the General Research Fund (GRF) of the Hong Kong Research Grants Council (RGC) RGC HKU 17210115 as well as the Environment and Conservation Fund (ECF) Project 26/2018 of the Hong Kong (HK) SAR.

References

Anenberg, S., Miller, J., Henze, D., & Minjares, R. (2019). *A global snapshot of the air pollution-related health impacts of transportation sector emissions in 2010 and 2015*. International Council on Clean Transportation: Washington, DC, USA.

Arya, S. P. (1999). *Air Pollution Meteorology and Dispersion*. New York: Oxford University Press.

Ballesta, P. P. (2005). The uncertainty of averaging a time series of measurements and its use in environmental legislation. *Atmospheric Environment*, **39**(11), 2003-2009.

Brown, R. J., Hood, D., & Brown, A. S. (2008). On the optimum sampling time for the measurement of pollutants in ambient air. *Journal of Automated Methods and Management in Chemistry*, **2008**(4), 814715.

Brown, R. J., & Woods, P. T. (2014). Proposals for new data quality objectives to underpin ambient air quality monitoring networks. *Accreditation and Quality Assurance*, **19**(6), 465-471.

Caiazzo, F., Ashok, A., Waitz, I. A., Yim, S. H., & Barrett, S. R. (2013). Air pollution and early deaths in the United States. Part I: Quantifying the impact of major sectors in 2005. *Atmospheric Environment*, **79**, 198-208.

Chan, T., Luo, D., Cheung, C., & Chan, C. (2008). Large eddy simulation of flow structures and pollutant dispersion in the near-wake region of the studied ground vehicle for different driving conditions. *Atmospheric Environment*, **42**(21), 5317-5339.

Chan, T. L., Dong, G., Cheung, C. S., Leung, C. W., Wong, C., & Hung, W. (2001). Monte Carlo simulation of nitrogen oxides dispersion from a vehicular exhaust plume and its sensitivity studies. *Atmospheric Environment*, **35**(35), 6117-6127.

Chang, V. W.-C., Hildemann, L. M., & Chang, C.-h. (2009). Dilution rates for tailpipe emissions: effects of vehicle shape, tailpipe position, and exhaust velocity. *Journal of the Air & Waste Management Association*, **59**(6), 715-724.

Davison, J., Bernard, Y., Borken-Kleefeld, J., Farren, N. J., Hausberger, S., Sjödin, Å., Tate, J. E., Vaughan, A. R., & Carslaw, D. C. (2020). Distance-based emission factors from vehicle emission remote sensing measurements. *Science of the Total Environment*, **739**, 139688.

Ellis, J. T., Sherman, D. J., Farrell, E. J., & Li, B. (2012). Temporal and spatial variability of aeolian sand transport: Implications for field measurements. *Aeolian Research*, **3**(4), 379–387.

Franco, V., Kousoulidou, M., Muntean, M., Ntziachristos, L., Hausberger, S., & Dilara, P. (2013). Road vehicle emission factors development: A review. *Atmospheric Environment*, **70**, 84-97.

Huang, Y., Organ, B., Zhou, J. L., Surawski, N. C., Hong, G., Chan, E. F., & Yam, Y. S. (2018). Remote sensing of on-road vehicle emissions: Mechanism, applications and a case study from Hong Kong. *Atmospheric Environment*, **182**, 58-74.

Huang, Y., Yu, Y., Yam, Y.-s., Zhou, J. L., Lei, C., Organ, B., Zhuang, Y., Mok, W.-c., & Chan, E. F. C. (2020a). Statistical evaluation of on-road vehicle emissions measurement using a dual remote sensing technique. *Environmental Pollution*, **267**, 115456.

Huang, Y., Ng, E. C. Y., Surawski, N. C., Yam, Y.-S., Mok, W.-C., Liu, C.-H., Zhou, J. L., Organ, B., & Chan, E. F. C. (2020b). Large eddy simulation of vehicle emissions dispersion: Implications for on-road remote sensing measurements. *Environmental Pollution*, **259**, 113974.

Huang, Y., Lei, C., Liu, C. H., Perez, P., Forehead, H., Kong, S., & Zhou, J. L. (2021). A review of strategies for mitigating roadside air pollution in urban street canyons. *Environmental Pollution*, **280**, 116971.

HKEPD (2019). *2017 Hong Kong Emission Inventory Report*, Environmental Protection Department, The Government of the Hong Kong Special Administrative Region. Retrieved from https://www.epd.gov.hk/epd/sites/default/files/epd/data/2017_Emission_Inventory_Report_Eng.pdf

Janik, M., Łoskiewicz, J., Tokonami, S., Kozak, K., Mazur, J., & Ishikawa, T. (2012). Determination of the minimum measurement time for estimating long-term mean radon concentration. *Radiation Protection Dosimetry*, **152**(1-3), 168-173.

Ježek, I., Drinovec, L., Ferrero, L., Carriero, M., & Močnik, G. (2015). Determination of car on-road black carbon and particle number emission factors and comparison between mobile and stationary measurements. *Atmospheric Measurement Techniques*, **8**(1), 43-55.

Li, X. X., Liu, C. H., & Leung, D. (2007). Large-eddy simulation of flow field and pollutant transport insider urban street canyons with high aspect ratios, *7th Symposium on Urban Environment*; 10 to 13 September 2007; San Diego, USA.

Li, Z., Che, W., Frey, H. C., Lau, A. K., & Lin, C. (2017). Characterization of PM2. 5 exposure concentration in transport microenvironments using portable monitors. *Environmental Pollution*, **228**, 433-442.

Liu, X., Niu, J., & Kwok, K. C. (2011). Analysis of concentration fluctuations in gas dispersion around high-rise building for different incident wind directions. *Journal of Hazardous Materials*, **192**(3), 1623-1632.

McArthur, D., Burton, D., Thompson, M., & Sheridan, J. (2016). On the near wake of a simplified heavy vehicle. *Journal of Fluids and Structures*, **66**, 293-314.

Morawska, L., Ristovski, Z., Johnson, G., Jayaratne, E., & Mengersen, K. (2007). Novel method for on-road emission factor measurements using a plume capture trailer. *Environmental Science & Technology*, **41**(2), 574-579.

Mylne, K. R. & Mason, P. (1991). Concentration fluctuation measurements in a dispersing plume at a range of up to 1000 m. *Quarterly Journal of the Royal Meteorological Society*, **117**(497), 177-206.

OpenFOAM. (2018). OpenFOAM 6. Retrieved from https://openfoam.org/

Park, S. S., Kozawa, K., Fruin, S., Mara, S., Hsu, Y. K., Jakober, C., Winer, A., & Herner, J. (2011). Emission factors for high-emitting vehicles based on on-road measurements of individual vehicle exhaust with a mobile measurement platform. *Journal of the Air & Waste Management Association*, **61**(10), 1046-1056.

Rao, A. N., Zhang, J., Minelli, G., Basara, B., & Krajnović, S. (2019). An LES investigation of the near-wake flow topology of a simplified heavy vehicle. *Flow, Turbulence and Combustion*, **102**(2), 389-415.

Richards, K. (2003). *Computational Modelling of Pollution Dispersion in the Near Wake of a Vehicle*. PhD thesis, University of Nottingham.

Santos, J., Griffiths, R., Reis Jr, N., & Mavroidis, I. (2009). Experimental investigation of averaging time effects on building influenced atmospheric dispersion under different meteorological stability conditions. *Building and Environment*, **44**(6), 1295-1305.

Santos, J., Griffiths, R., Roberts, I., & Reis Jr, N. (2005). A field experiment on turbulent concentration fluctuations of an atmospheric tracer gas in the vicinity of a complex-shaped building. *Atmospheric Environment*, **39**(28), 4999-5012.

Santos, J. M., Reis, N., Castro, I., Goulart, E. V., & Xie, Z.-T. (2019). Using large-eddy simulation and wind-tunnel data to investigate peak-to-mean concentration ratios in an urban environment. *Boundary-Layer Meteorology*, **172**(3), 333-350.

Singer, I. A., Imai, K., & Campo, R. G. D. (1963). Peak to mean pollutant concentration ratios for various terrain and vegetation cover. *Journal of the Air Pollution Control Association*, **13**(1), 40-42.

Smagorinsky, J. (1963). General circulation experiments with the primitive equations: I. The basic experiment. *Monthly Weather Review*, **91**(3), 99-164.

Smit, R. & Kingston, P. (2019). Measuring on-road vehicle emissions with multiple instruments including remote sensing. *Atmosphere*, **10**(9), 516.

Smit, R., Kingston, P., Neale, D., Brown, M., Verran, B., & Nolan, T. (2019). Monitoring on-road air quality and measuring vehicle emissions with remote sensing in an urban area. *Atmospheric Environment*, **218**, 116978.

Smit, R., Ntziachristos, L., & Boulter, P. (2010). Validation of road vehicle and traffic emission models–A review and meta-analysis. *Atmospheric Environment*, **44**(25), 2943-2953.

Tayarani, M. & Rowangould, G. (2020). Estimating exposure to fine particulate matter emissions from vehicle traffic: Exposure misclassification and daily activity patterns in a large, sprawling region. *Environmental Research*, **182**, 108999.

Tong, Z., Li, Y., Lin, Q., Wang, H., Zhang, S., Wu, Y., & Zhang, K. M. (2022). Uncertainty investigation of plume-chasing method for measuring on-road NOx emission factors of heavy-duty diesel vehicles. *Journal of Hazardous Materials*, **424**, 127372.

Venkatram, A. (2002). Accounting for averaging time in air pollution modeling. *Atmospheric Environment*, **36**(13), 2165-2170.

Volpe, R., Devinant, P., & Kourta, A. (2015). Experimental characterization of the unsteady natural wake of the full-scale square back Ahmed body: flow bi-stability and spectral analysis. *Experiments in Fluids*, **56**(5), 1-22.

Wang, H., Wu, Y., Zhang, K. M., Zhang, S., Baldauf, R. W., Snow, R., Deshmukh, P., Zheng, X., He, L., & Hao, J. (2020). Evaluating mobile monitoring of on-road emission factors by comparing concurrent PEMS measurements. *Science of the Total Environment*, **736**, 139507.

Webb, N. P., Chappell, A., Edwards, B. L., McCord, S. E., Van Zee, J. W., Cooper, B. F., Courtright, E. M., Duniway, M. C., Sharratt, B., Tedela, N., & Toledo, D. (2019). Reducing Sampling Uncertainty in Aeolian Research to Improve Change Detection. *Journal of Geophysical Research: Earth Surface*, **124**(6), 1366–1377.

Weller, H. G., Tabor, G., Jasak, H., & Fureby, C. (1998). A tensorial approach to computational continuum mechanics using object-oriented techniques. *Computers in Physics*, **12**(6), 620-631.

Wen, Y., Wang, H., Larson, T., Kelp, M., Zhang, S., Wu, Y., & Marshall, J. D. (2019). On-highway vehicle emission factors, and spatial patterns, based on mobile monitoring and absolute principal component score. *Science of the Total Environment*, **676**, 242-251.

Wilson, D. J. (2010). *Concentration Fluctuations and Averaging Time in Vapor Clouds*. John Wiley & Sons.

Wu, Y., Zhang, S., Hao, J., Liu, H., Wu, X., Hu, J., Walsh, M. P., Wallington, T. J., Zhang, K. M., & Stevanovic, S. (2017). On-road vehicle emissions and their control in China: A review and outlook. *Science of the Total Environment*, **574**, 332-349.

Xie, J., Liu, C.-H., Mo, Z., Huang, Y., & Mok, W.-C. (2020). Near-field dynamics and plume dispersion after an on-road truck: Implication to remote sensing. *Science of the Total Environment*, **748**, 141211.

Xie, Z.-T., Hayden, P., Robins, A. G., & Voke, P. R. (2007). Modelling extreme concentrations from a source in a turbulent flow over a rough wall. *Atmospheric Environment*, **41**(16), 3395-3406.

Yang, B., Zhang, K. M., Xu, W. D., Zhang, S., Batterman, S., Baldauf, R. W., Deshmukh, P., Snow, R., Wu, Y., Zhang, Q., Li, Z., & Wu, X. (2018). On-road chemical transformation as an important mechanism of NO_2 formation. *Environmental Science & Technology*, **52**(8), 4574-4582.

Yee, E. & Biltoft, C. A. (2004). Concentration fluctuation measurements in a plume dispersing through a regular array of obstacles. *Boundary-Layer Meteorology*, **111**(3), 363-415.

Zheng, X., Wu, Y., Zhang, S., Baldauf, R. W., Zhang, K. M., Hu, J., Li, Z., Fu, L., & Hao, J. (2016). Joint measurements of black carbon and particle mass for heavy-duty diesel vehicles using a portable emission measurement system. *Atmospheric Environment*, **141**, 435–442.

Supplementary Materials

Table. S1 Nomenclature.

Symbol	Definition
h, *d*	Height and width of the trunk
U_∞	Vehicle speed (inflow velocity in LES)
$\Delta t = 0.15h/U_\infty$	Time increment in LES
$\Delta\tau = (n - 1)\Delta t$	Sampling duration (averaging time in statistical analysis)
n (≥ 1)	Number of data samples in the sampling duration $\Delta\tau$
ΔT	Reference time interval (a longer time period)
p	Exponent of power law in Eq. (1)
$T = 510h/U_\infty$	Total sampling period (length of time-series data)
ϕ_i	Pollutant concentration from the LES results (*i* is the index of data signal)
$\bar{\phi}$	Mean concentration over the sampling duration $\Delta\tau$
$\dot{Q}$	Pollutant emission rate (constant) from the tailpipe
$\Phi_0 = \dot{Q}/U_\infty h^2$	Characteristic pollutant concentration
$C_i = \phi_i/\Phi_0$	Dimensionless pollutant concentration
$\overline{C}(\Delta\tau)_i = \sum_{j=0}^{n-1} C_{i+j}/n$	Sample mean concentration obtained by moving average over $\Delta\tau$ whose first data point starts at point *i*
$\overline{C} = \overline{C}(\Delta\tau = T)_{i=0}$	Population mean concentration over *T*
$\delta(\Delta\tau)_i = \left[\overline{C}(\Delta\tau)_i - \overline{C}\right]/\overline{C}$	Relative deviation between sample mean $\overline{C}(\Delta\tau)_i$ and population mean $\overline{C}$ concentrations
$\delta_i = \delta(\Delta\tau = \Delta t)_i = \left(C_i - \overline{C}\right)/\overline{C}$	Relative deviation of instantaneous concentration C_i
$CV(\Delta\tau) = \left\{\sum_{i=0}^{N-1}\left[\delta(\Delta\tau)_i\right]^2/N\right\}^{1/2}$	Coefficient of variance in sample mean
N	Number of data samples in the subsets of sample mean
$k(\Delta\tau)$	Fraction of acceptable sample mean satisfying $\lvert\delta(\Delta\tau)_i\rvert \le 15\%$
$I = CV(\Delta t)$	Fluctuating concentration intensity
f	Frequency of fluctuating concentration
$St = fd/U_\infty$	Strouhal number

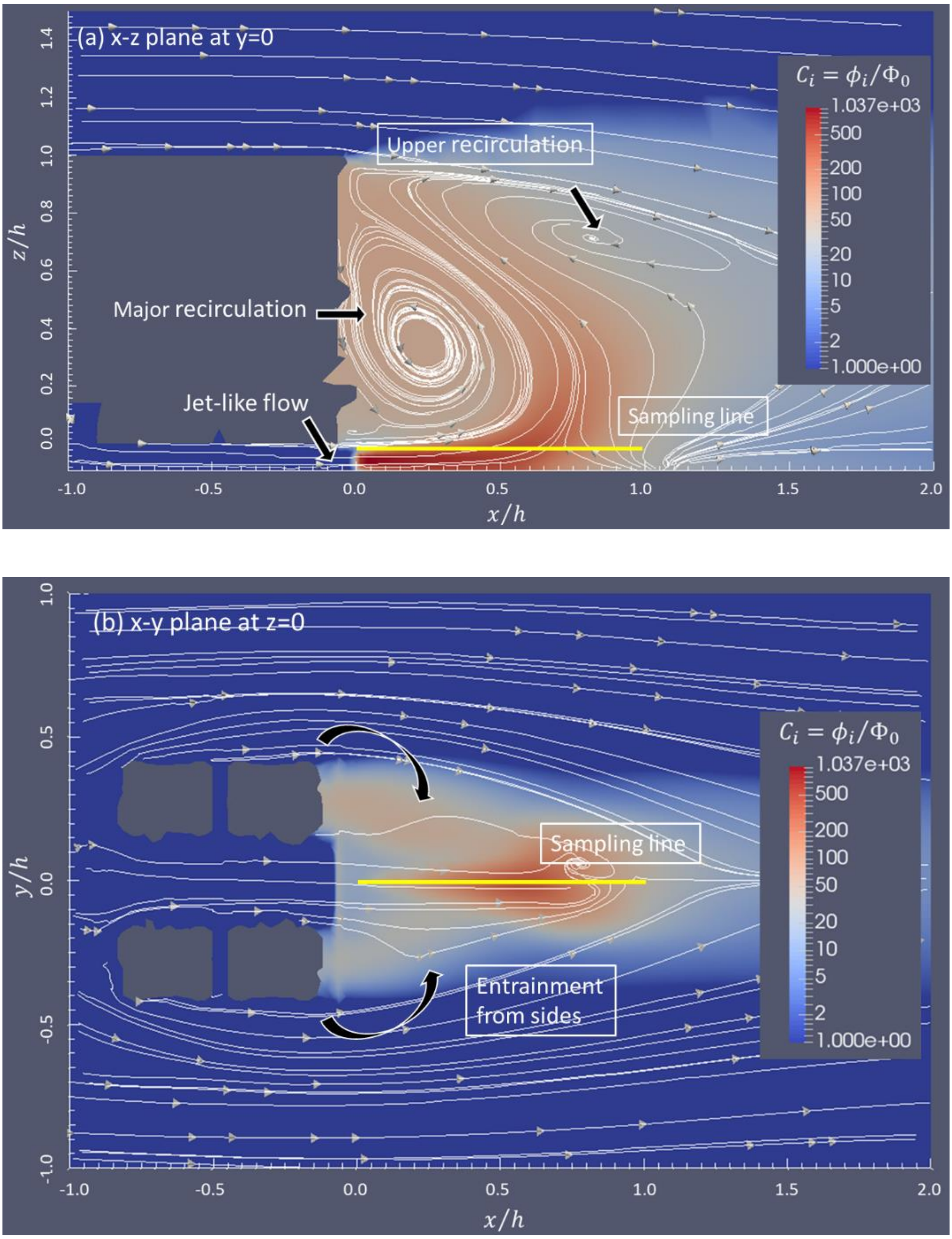

Fig. S1 Shaded contours of dimensionless pollutant concentration C_i (= ϕ_i/Φ_0) overlaid with streamlines on (a) *x*-*z* plane at the centreline (*y* = 0) and (b) *x*-*y* plane at the tailpipe level (*z* = 0).