

Predicting Concentration Fluctuations of Locally Emitted Air Pollutants in Urban-like Geometry Using Deep Learning

Bálint Papp^{1*}, Gergely Kristóf¹

¹ Department of Fluid Mechanics, Faculty of Mechanical Engineering, Budapest University of Technology and Economics, H-1521 Budapest, P. O. B. 91, Hungary

* Corresponding author, e-mail: papp.balint@gpk.bme.hu

Received: 11 September 2023, Accepted: 04 December 2023, Published online: 09 January 2024

Abstract

The accurate quantification of concentration fluctuations is crucial when evaluating the exposure to toxic, infectious, reactive, flammable, or explosive substances, as well as for the estimation of odor nuisance. However, in the field of Computational Fluid Dynamics (CFD), the industry currently relies predominantly on steady-state RANS turbulence models for simulating near-field pollutant dispersion, which are only capable of producing the time-averaged concentration field. This paper presents a regression relationship for calculating the standard deviation of the local concentration based on the mean concentration and the downstream distance from a point source, over a city-like surface, in the case of the wind direction perpendicular to the streets. The desired peak values and other statistical characteristics can be predicted by assuming a gamma distribution which is fitted based on the average and standard deviation. To obtain the regression function, a deep neural network model was used. The model was trained using time-resolved concentration data obtained from wind tunnel experiments. The validation results show that the concentration fluctuations predicted by the DNN-based model are in satisfactory agreement with the measurement data in terms of the skewness, the kurtosis, the median, and the peak concentrations. Furthermore, the present paper suggests a workflow for estimating the concentration fluctuations based on RANS CFD results, as well as recommendations for generating further training data for specific applications.

Keywords

fluid dynamics, urban air quality, wind tunnel measurement, deep neural network (DNN), artificial intelligence (AI)

1 Introduction

Outdoor air pollution poses a major health risk, and it is responsible for one in every nine deaths worldwide [1]. To assess the impact of air pollutants on the human body, so-called toxicity models were developed [2]. The most straightforward models assume a linear relationship between the received dose (i.e., the time integral of the concentration) and its effects, but more elaborate dose-toxicity models require the peak concentration values, which can be 10–1000 times higher than the mean [3], as well as time statistics, such as the probability of the concentration exceeding a specific threshold and the expected mean time above this threshold; therefore, knowing the time-average of the concentration is not enough. Moreover, it was reported by [4] that for the assessment of the exposure to infectious, reactive, flammable, or explosive substances, as well as for the estimation of odor nuisance, the magnitude of the concentration fluctuations is vital input data.

Computational urban dispersion studies, reviewed by [5–7], often utilize the Reynolds Averaged Navier-Stokes (RANS) approach for turbulence modeling. These simulations can only provide the time-averaged wind velocity and pollutant concentration field – but in the matter of days or even hours. However, RANS simulations are unable to capture the key features of urban flows and near-field dispersion, e.g., the unsteadiness of large-scale flow structures, and the anisotropy of turbulent scalar fluxes. On the other hand, scale-resolving turbulence models, such as Large Eddy Simulation (LES), can provide significantly more accurate results compared to RANS regarding the mean concentration distribution [7, 8], with the additional benefit of the capability of predicting concentration fluctuations. The computational demand of scale resolving CFD simulations, however, can be higher than that of RANS calculations by two orders of magnitude.

In spite of the rapid increase in computational resources and claims made in the past decades that LES would eventually render RANS obsolete, the RANS approach still remains widely used for simulating near-field pollutant dispersion in research and in the engineering practice as well [9], due to its robustness, its speed, and the availability of Best Practice Guidelines, such as [10–13].

In recent years, artificial neural networks (ANNs) have been successfully implemented in many air pollution forecasting applications. The corresponding literature was reviewed by [14, 15], listing applications using different ANN types to predict either the mean value or a time series of the concentration (of CO_x , NO_x , SO_2 , O_3 , as well as particulate matter) based on various meteorological parameters, including wind speed and direction, ambient pressure and temperature, relative humidity, solar radiation, rainfall, day of the week/month/year, source intensity and concentration, as well as geometrical coordinates. Numerous studies have shown that ANN models can be effectively used to define functional relationships between dependent and independent variables. A variant of ANNs, deep neural networks (DNNs), can be trained to accurately predict the complex relationship of numerous input parameters and the pollutant concentration via multiple hidden layers and can be applied to similar tasks to the ones listed above, as reported by [16, 17].

The present paper proposes a DNN-based regression function for predicting concentration fluctuations of locally emitted air pollutants. The streamwise distance from the source and the local mean concentration is the input part, and the measured concentration's standard deviation is the output part of the training dataset, which covers three urban-like surface structures with different variability of building height and building alignment. The wind direction is perpendicular to the streets in all cases. The model's accuracy undergoes a meticulous quantitative analysis, with the derived regression function illustrated through graphical representations, facilitating a deeper understanding and application of the findings.

2 Methods

2.1 Building configurations

In the present paper, three periodically repeated building arrangements of equal total volume are employed to model the urban environment:

1. *Uniform street canyons (UC)* of $H/W = 1$ height-to-width aspect ratio, forming parallel streets perpendicular to the wind direction.

2. *Aligned towers (AT)*: $1.5H$ and $0.5H$ tall building segments, forming street canyons of variable building height, with a full overlap of the tall building segments in streamwise direction.
3. *Staggered towers (ST)* of $1.5H$ and $0.5H$ tall building segments, with no streamwise overlap of the tall buildings in consecutive rows.

The building configurations are depicted in Fig. 1. The reference building height is $H = 100$ [mm], the street width and the building breadth (in streamwise direction)

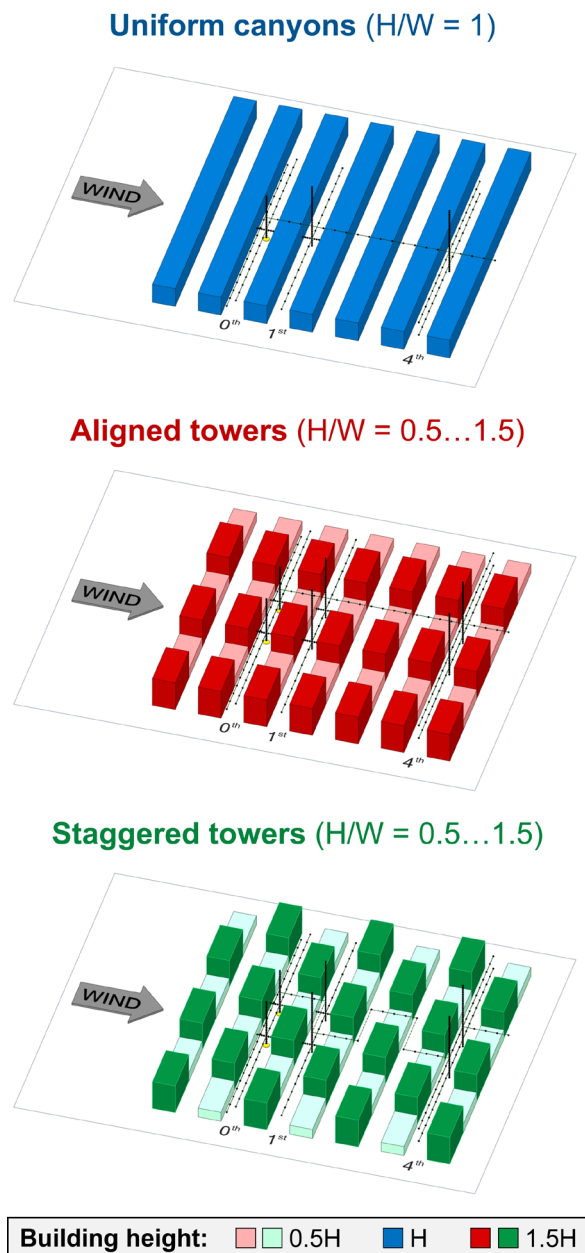


Fig. 1 Periodic building arrangements used in the experiments. (The positions of the sampling points are denoted by black dots. The location and size of the sources are shown in yellow color in the 0th canyon.)

is $W = B = 100$ [mm], and the lateral size of the tower segments is $T = 208.33$ [mm]. The building models were constructed considering a 1:200 model scale. As the streets were constructed perpendicular to the wind direction, the investigated scenario corresponds to the critical wind direction regarding ventilation.

2.2 Wind tunnel experiments

The dispersion of air pollutants was investigated via wind tunnel experiments. The experiments were carried out in the closed-circuit horizontal wind tunnel of the Theodore von Kármán Wind Tunnel Laboratory at the Department of Fluid Mechanics of the Budapest University of Technology and Economics (BME). The wind tunnel has a circular cross-section of 2.6 [m] diameter at the open test section of 3.8 [m] length, and it is equipped with a 2.5 [m] wide horizontal table (Fig. 2).

The incoming flow was homogeneous with a low turbulence intensity (<1%) thus the boundary layer could adapt to the building arrangements. The flow was neutrally stratified, and it was found that the boundary layer over the buildings can be considered fully developed after 11 street canyons; thus, after that, it can be considered periodic. This canyon is labeled as the 0^{th} (*source*) *canyon* from now on. Two identical tracer gas sources were flush mounted on the base plate of the model in the middle of the 0^{th} canyon with a lateral offset of half a tower width, i.e., one of the sources was located downstream of a short ($0.5H$) and another downstream of a tall ($1.5H$) tower. The applied tracer gas representing traffic-induced air pollutants, or an accidental gas release was pure methane (CH_4), emitted continuously from one of the two point-like sources at a time. The flow rate was controlled by pre-calibrated smart mass flow rate meters.

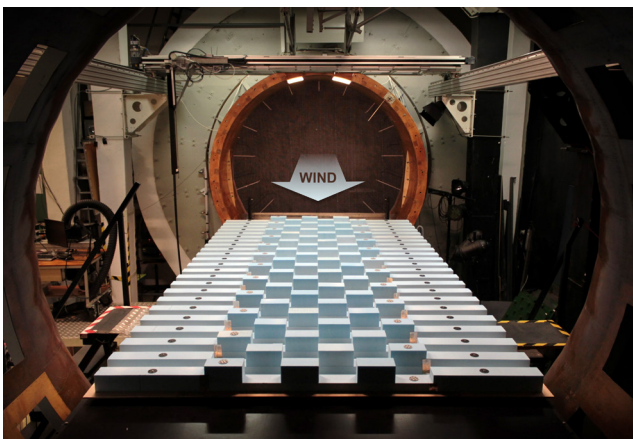


Fig. 2 The staggered towers (ST) configuration in the wind tunnel

The concentration field was sampled in a total of 1138 gauging points for the three building arrangements in total, illustrated in Fig. 1, using a Cambustion HFR400 FFID (Fast Flame Ionization Detection) analyzer, for 60–150 [s] in each measurement point, with a sampling frequency of 1000 [Hz] and a response time of 2.05 [ms], having a $\Delta_{rel} = 12\%$ relative and $\Delta_{abs} = 0.28 c^*$ absolute measurement uncertainty. The complete list of the measurement instrumentation can be found in [18].

The voltage time signal produced by the FFID analyzer was first calibrated to obtain the tracer gas concentration (in part per million, ppm) using two calibration mixtures of known concentrations (100 ppm and 4950 ppm). The background concentration of the wind tunnel was also measured frequently along with the calibration of the device, with a maximum time interval of 30 minutes. The normalized concentration time series were obtained as

$$c^*(t) = \frac{c(t) - c_{BG}}{10^6} \cdot \frac{u_{ref,UC}}{Q/A}, \quad (1)$$

in which $c^* [-]$ is the normalized concentration, t [s] is time, c [ppm] is the measured concentration, c_{BG} [ppm] is the mean wind tunnel background concentration corresponding to the current measurement point, $u_{ref,UC} = 4.74$ [m/s] is the reference mean velocity (taken at $z/H = 2$ in the uniform canyons configuration), $Q = 6.2 \cdot 10^{-6} \dots 6.9 \cdot 10^{-5}$ [m³/s] is the volume flow rate of the tracer gas, and $A = 2 \cdot S \cdot T$ [m²] is the reference area (constant for all three geometries), in which $S = 2H$ is the distance of two consecutive streets.

As the approach flow velocity was identically 9 [m/s] for the three different geometries (UC, AT, ST), using the same velocity and area for normalizing the results for all of them allows a fair comparison of the resulting concentration fields, as this approach also includes the effect of surface drag corresponding to the installation length used in the model: in the case of rougher surfaces – i.e. when roof height heterogeneity is present – a better mass transfer is achieved at the cost of a greater deceleration of the wind speed. Note that the normalized concentration results are also compensated for the minor temporal changes in the bulk velocity.

2.3 Deep neural network (DNN)

A deep neural network (DNN) model was employed to investigate the correlation between the standard deviation of the normalized tracer gas concentration (c_{std}^*) and two input parameters: the normalized streamwise distance from the point source (x/H) and the mean normalized concentration (c_{mean}^*). The density of the measurement grid

was insufficient for the proper resolution of the concentration distribution close to the source; thus the measurement points closer than $0.7H$ to the active sources were removed from the training dataset, resulting in a remainder of 1015 sampling points with a maximum mean concentration of $c^* = 180$.

The DNN, developed using the Keras Sequential framework, was structured with three internal layers, each housing 20 neurons and utilizing the LeakyReLU activation function (with $\alpha = 0.1$ slope coefficient). A final output layer with linear activation was implemented to predict the target variable (c_{std}^*). The network comprised 925 trainable parameters and was optimized using the Adam optimizer, minimizing the mean squared error (MSE) as the loss function. Using a learning rate of 0.001, the training process converged after 200 epochs.

2.4 Analytical estimation of the pointwise concentration probability distribution

The probability distribution of the concentration was modeled with the gamma distribution, which is a widely used model for describing concentration fluctuations in near-field pollutant dispersion studies; see [4].

In the present study, the parameters of the gamma distribution were determined based on the mean and standard deviation of the concentration. The mean value corresponded to the measurement in all cases, and the standard deviation was either determined by the DNN model or – and as a reference – taken directly from the locally measured concentration time series.

The probability density function (PDF) of the gamma distribution is given by

$$f(c^* | a, b) = \frac{1}{b^a \cdot \Gamma(a)} \cdot (c^*)^{a-1} \cdot \exp\left(-\frac{c^*}{b}\right), \quad (2)$$

in which a [-] is the shape parameter, b [-] is the scale parameter, and Γ denotes the Gamma function. The shape and scale parameters are fitted based on the mean (c_{mean}^*) and standard deviation (c_{std}^*) of the normalized concentration using Eq. (3):

$$a = \left(\frac{c_{mean}^*}{c_{std}^*}\right)^2, \text{ and } b = \frac{(c_{std}^*)^2}{c_{mean}^*}. \quad (3)$$

3 Results and discussion

3.1 Measured concentration statistics

In this section, the characteristic vertical concentration profiles are presented (out of the 1138 measurement points) and

both the near-field and the far-field results are discussed. The statistics of the pointwise concentration probability distribution are plotted in Fig. 3, from the source canyon (0th) and from the subsequent downstream (1st) street canyon.

For the uniform canyons (UC) case, the vertical mean concentration distribution is rather homogeneous at the middle of the canyon, above the source ($x/H = 0$) with a maximum at roof height, as the canyon vortex, the axis of which is horizontal, sweeps the pollutants towards the upstream building at street level, and then transports them vertically upwards near the wall. The location of the maximum standard deviations both in the source and the first downstream street canyon are closely linked to the location of the highest turbulence, i.e., the shear layer at roof height.

For the aligned and staggered tower arrangements (AT, ST), the maximum values of the mean concentrations exceed those of the UC case. In the source canyon, especially high mean concentrations can be observed near the sources, along with similarly high standard deviations. In the first downstream canyon, subjected only indirectly to the emission, the concentration distribution is more homogeneous compared to the UC case, indicating the presence of intense turbulence within the buildings (below $1.5H$ for the towers).

The skewness shows the asymmetry of the distribution, and the kurtosis is a good indicator of the "tailedness" of the distribution, i.e. how often outliers occur. Note that the symmetric Gaussian distribution has a skewness of 0 and a kurtosis 3, and the exponential distribution has a skewness of 2 and a kurtosis of 9. It can be observed in Fig. 3 that the highest values of these third and fourth central moments can be found in two locations:

1. near the sources, where the plume meandering is dominant, and
2. near the edges of the concentration plume.

Both of these locations can be characterized by highly intermittent concentration-time signals, composed of frequent near-zero elements and a few peaks that can be multiple orders of magnitude higher than the mean value. The lower (absolute) value of the skewness and the kurtosis suggest more symmetrical, normal-like distribution, characterizing the inside of the plume, where the pollutants are more evenly mixed due to diffusion.

3.2 Model validation

The shape and scale parameters of the gamma distribution were calculated according to Eq. (3) based on two approaches:

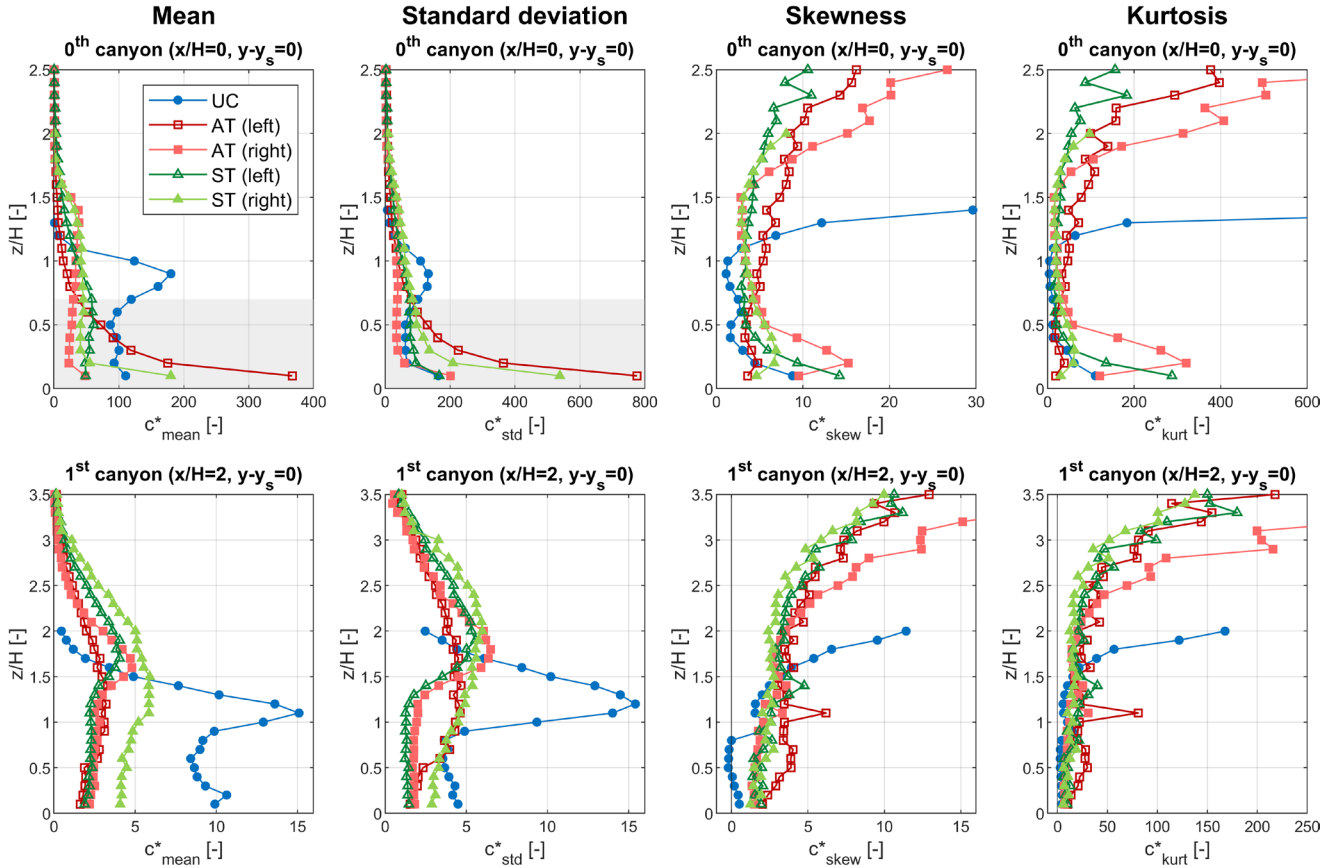


Fig. 3 Vertical profiles of the mean, the standard deviation, the skewness and the kurtosis of the pointwise concentration distributions, measured in the source (0th) and the subsequent downstream canyon (1st) for different geometries. (The total number of measurement points, including the ones in the present profiles, was 1138. Points in the close vicinity of the sources, here denoted by a grey area, were not used for the training of the DNN model.)

1. the measured concentration statistics ($c_{\text{mean,WT}}^*$, $c_{\text{std,WT}}^*$) as well as
2. the combined output of the wind tunnel experiment and the DNN model ($c_{\text{mean,WT}}^*$, $c_{\text{std,DNN}}^*$).

The resultant distributions were compared to the empirical statistical parameters in terms of the central moments, the median (50th percentile), and the representative minimum and maximum values (1st and 99th percentiles; the rest were considered outliers). The performance metrics proposed by [19] were applied to all above quantities, which range over multiple orders of magnitude, thus, only the correlation coefficient (R , Eq. (4)), the factor of two of the observations ($FAC2$, Eq. (5)), the geometric mean bias (MG , Eq. (6)), and the geometric variance (VG , Eq. (7)) are representative for them among the metrics proposed in the original paper.

$$R = \frac{(\overline{O} - \overline{O})(\overline{P} - \overline{P})}{\sigma_O \sigma_P} \quad (4)$$

$$FAC2 = \frac{1}{n} \sum_{i=1}^N n_i, \quad \text{with}$$

$$n_i = \begin{cases} 1 & \text{if } 0.5 \leq \frac{P_i}{O_i} \leq 2 \\ 1 & \text{if } |O_i| \leq \Delta_{\text{abs}} \text{ and } |P_i| \leq \Delta_{\text{abs}} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$$MG = \exp(\overline{\ln \tilde{O}} - \overline{\ln \tilde{P}}), \quad \text{with } \tilde{O} = \max(O, \Delta_{\text{abs}}) \quad (6)$$

and $\tilde{P} = \max(P, \Delta_{\text{abs}})$

$$VG = \exp(\overline{(\ln \tilde{O} - \ln \tilde{P})^2}), \quad \text{with } \tilde{O} = \max(O, \Delta_{\text{abs}}) \quad (7)$$

and $\tilde{P} = \max(P, \Delta_{\text{abs}})$

In the above equations, P denotes the model predictions, and O denotes the measurement results. Moreover, \tilde{O} and \tilde{P} stand for the averages, while σ_O and σ_P represent the standard deviations, and n denotes the number of elements of the O and P data sets. Note that data points where the measured mean concentration was smaller than

the absolute measurement uncertainty ($\Delta_{abs} = 0.28c^*$) were removed prior to the performance analysis. The formulas for MG and VG accept positive values only, thus 20 further data points, which yielded negative skewness values had to be also removed before the calculation of the MG and VG metrics. (This way, 884 points were used to calculate the performance metrics, except for the skewness, from which an additional 20 points were removed.) The resultant validation metrics are shown in Table 1.

Regarding the *analytical gamma distribution estimation*, the perfect fit of the standard deviation is an inherent feature of the fitting method. The correlation coefficient of the skewness and the kurtosis is 0.637 and 0.340, respectively; and it can be concluded that 86.1% and 70.8% of the estimated skewness and kurtosis values fall into the range designated by the $FAC2$ metric. Both the third and fourth moments are underestimated ($MG = 1.136$ and 1.310 , see Fig. 4 for illustration), and one can observe a minor random error for the skewness and the kurtosis ($VG = 1.364$ and 1.768). It must be noted, however, that although, the fit of the estimated gamma distributions are not perfect, all validation metrics show values sufficiently close to the targets for the 1st, 50th and 99th percentiles ($R = 0.789 \dots 0.998$, $FAC2 = 0.919 \dots 0.993$, $MG = 0.960 \dots 1.132$, $VG = 1.041 \dots 1.150$), suggesting that the concentration minima, the medians, and the maxima are all predicted with good accuracy.

Table 1 Performance metrics

Metric	R	$FAC2$	MG	VG
Target	1	1	1	1
Standard deviation:				
From $c_{std,WT}^*$	1	1	1	1
From $c_{std,DNN}^*$	0.945	0.760	0.831	1.362
Skewness:				
From $c_{std,WT}^*$	0.637	0.861	1.136	1.364
From $c_{std,DNN}^*$	0.540	0.781	0.949	1.599
Kurtosis:				
From $c_{std,WT}^*$	0.340	0.708	1.310	1.768
From $c_{std,DNN}^*$	0.452	0.624	1.057	2.175
Representative minimum (1 st percentile):				
From $c_{std,WT}^*$	0.789	0.934	0.960	1.089
From $c_{std,DNN}^*$	0.140	0.821	0.980	1.138
Median (50 th percentile):				
From $c_{std,WT}^*$	0.990	0.919	1.132	1.150
From $c_{std,DNN}^*$	0.958	0.800	1.136	1.323
Representative maximum (99 th percentile):				
From $c_{std,WT}^*$	0.998	0.993	1.040	1.041
From $c_{std,DNN}^*$	0.939	0.958	0.985	1.250

About the *DNN model*, it must be noted firstly, that the standard deviations are predicted accurately ($R = 0.945$, $FAC2 = 0.760$), although, they are slightly overestimated ($MG = 0.831$), with a minor random error ($VG = 1.362$). To assess the propagation of the error of the fitting method using the mean and the standard deviation, the prediction accuracy for the skewness, the kurtosis as well as the percentiles must be assessed.

The correlation coefficients for the skewness and the kurtosis are $R = 0.540$ and 0.452 , respectively, which are slightly smaller and larger than those of the analytical model. The skewness is predicted with reasonable accuracy, but it is overestimated by the DNN model, rather than underestimated, as shown in Fig. 4 ($FAC2 = 0.781$, $MG = 0.949$, $VG = 1.599$). Moreover, it can be seen that the prediction capability of the DNN model regarding the kurtosis is on par with the analytical estimation. It can also be concluded that the percentiles can be modeled with acceptable accuracy, as their metrics – apart from the weak correlation of the very small values of the 1st percentiles – show almost as good agreement with the experimental data as those of the analytical estimates ($R = 0.939 \dots 0.958$, $FAC2 = 0.800 \dots 0.958$, $MG = 0.980 \dots 1.136$, $VG = 1.138 \dots 1.323$).

3.3 DNN-based concentration statistics

The DNN model can be utilized to calculate the standard deviation for any given streamwise coordinate and mean concentration data pair as long as the inputs are in the range of the corresponding training data. Fig. 5 shows the c_{std}^* distribution for the valid input range of $z/H = -0.5 \dots 12$ and $c_{mean}^* = 0 \dots 180$.

As shown in the left-hand-side plot of Fig. 5, in the near-field, i.e., the 0th (source) canyon, the concentration fluctuations are almost linearly proportional to the mean concentration.

Since the mean concentrations generally decrease with increasing downstream distance (see the valid range in the middle plot of Fig. 5), the standard deviation in the far-field is more conveniently plotted as the function of $(x/H) \cdot c_{mean}^*$. The standard deviation displays monotonous increase for each curve; with decreasing maximum values of c_{std}^* with increasing distance from the source.

4 Conclusions and outlook

In this paper, a deep neural network-based method for predicting urban concentration fluctuations was demonstrated. The DNN model takes the local mean concentration (c_{mean}^*) and the streamwise position (x/H) as inputs

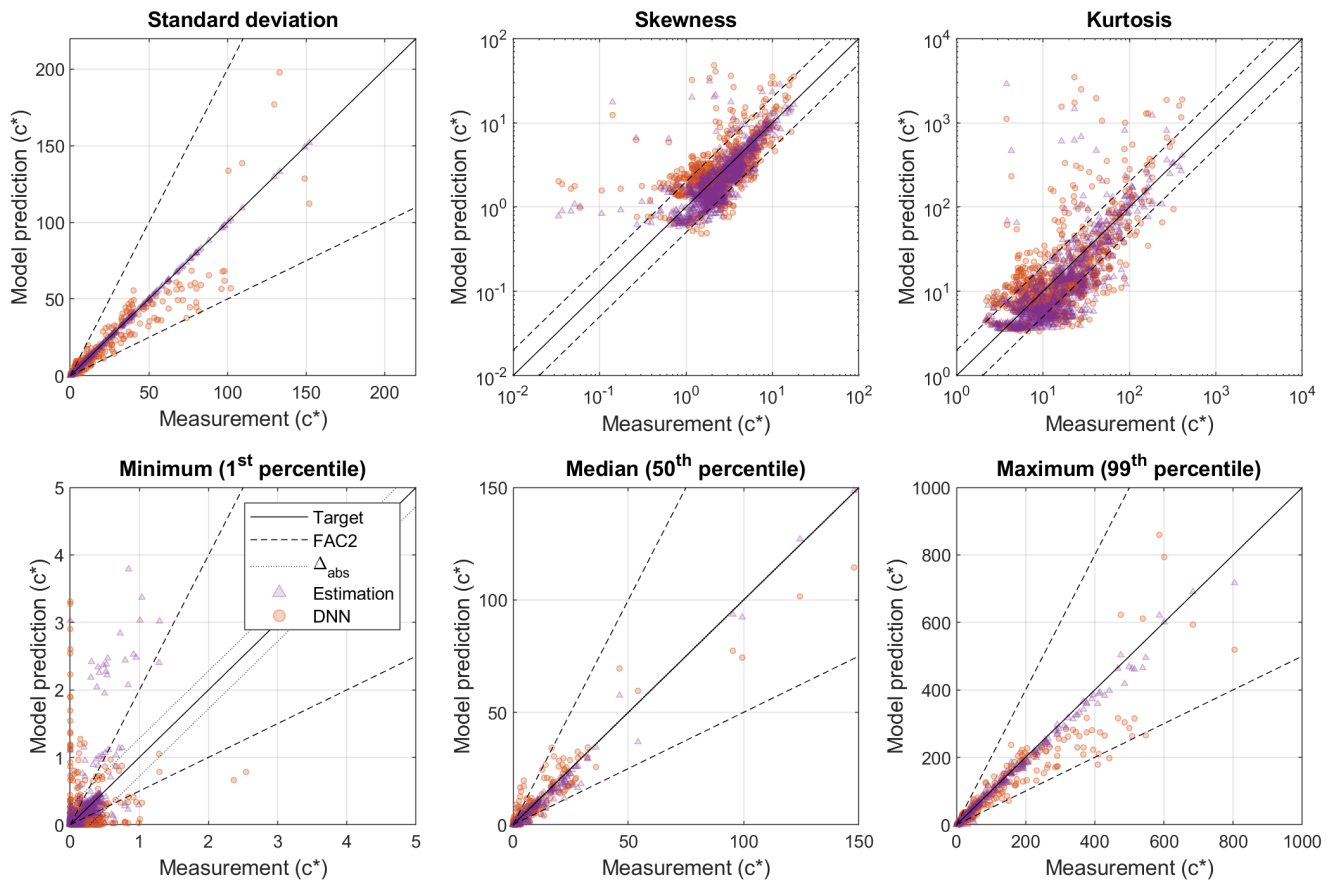


Fig. 4 Scatter plots illustrating the correlation between the concentrations predicted by the analytical estimation and the wind tunnel results (purple triangles), as well as between the prediction of the deep neural network (DNN) and the wind tunnel results (orange circles).

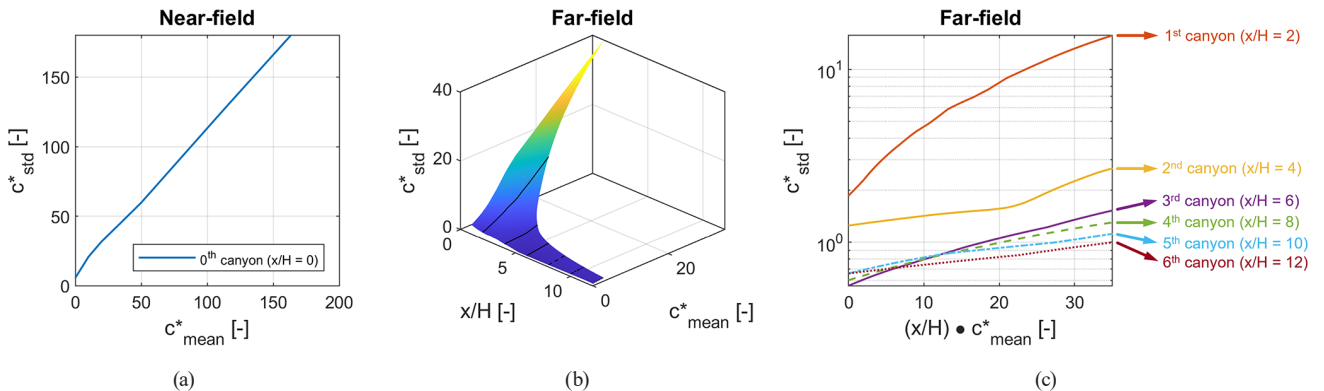


Fig. 5 Concentration fluctuation results predicted by the DNN model. (a) Standard deviation as the function of the mean concentration in the middle of the source canyon. (b) Standard deviation as the function of the streamwise coordinate and the mean concentration in the far-field. (c) Standard deviation as the function of the mean concentration and the streamwise coordinate, at the middle of the downstream street canyons. (The same curves are plotted in the middle diagram in black.)

and gives the local concentration standard deviation (c_{std}^*) as output. It was shown that based on the mean and the DNN-predicted standard deviation, the pointwise concentration probability distribution can be modeled with good accuracy using the gamma distribution (see Section 2.3 for the method and Table 1 for the performance metrics).

The presented method can be an effective addition to steady-state CFD models, such as the ones applying the industry standard RANS turbulence models, for forecasting concentration fluctuations in a short time and with limited computational demands. To successfully predict the concentration peaks, the following workflow is proposed:

1. Train a DNN model in advance, using concentration statistics of various building configurations, which are together representative of the investigated area.
2. Apply a steady-state CFD model to the specific geometry to obtain the mean normalized concentration distribution of the urban area.
3. The standard deviation of the concentration at a given location can be estimated based on its streamwise position and the local mean concentration.
4. The concentration fluctuations (skewness, kurtosis, median, peak) can be estimated by fitting a gamma distribution based on the CFD-generated mean value and the DNN-predicted standard deviation.

In order to train a DNN model capable of producing a representative function for predicting the concentration fluctuations, the desired training dataset must comply with the following recommendations:

- A sufficiently fine, evenly placed sampling grid must be applied, covering both the near-field and the far-field.
- A sufficient number of periodic geometries (in terms of type and parameter set) must be investigated to provide a sufficient estimation of the urban area. (See [20] for such examples.)
- Multiple wind directions representative of the meteorological conditions of the urban area must be investigated.

References

- [1] World Health Organization "Ambient air pollution: A global assessment of exposure and burden of disease", World Health Organization, 2016. ISBN 9789241511353 [online] Available at: <https://iris.who.int/handle/10665/250141> [Accessed: 01 December 2023]
- [2] Gunatilaka, A., Skvortsov, A., Gailis, R. "A review of toxicity models for realistic atmospheric applications", *Atmospheric Environment*, 84, pp. 230–243, 2014. <https://doi.org/10.1016/j.atmosenv.2013.11.051>
- [3] Gurka, R., Liberzon, A., Sarathi, P., Sullivan, P. J. "Diffusion of scalar concentration from localized sources in turbulent flows", *Reviews in Chemical Engineering*, 26(1–2), pp. 13–27, 2010. <https://doi.org/10.1515/REVCE.2010.004>
- [4] Cassiani, M., Bertagni, M. B., Marro, M., Salizzoni, P. "Concentration fluctuations from localized atmospheric releases", *Boundary-Layer Meteorology*, 177(2), pp. 461–510, 2020. <https://doi.org/10.1007/s10546-020-00547-4>
- [5] Lateb, M., Meroney, R. N., Yataghene, M., Fellouah, H., Saleh, F., Boufadel, M. C. "On the use of numerical modelling for near-field pollutant dispersion in urban environments – A review", *Environmental Pollution*, 208, pp. 271–283, 2016. <https://doi.org/10.1016/j.envpol.2015.07.039>
- [6] Tominaga, Y., Stathopoulos, T. "Ten questions concerning modeling of near-field pollutant dispersion in the built environment", *Building and Environment*, 105, pp. 390–402, 2016. <https://doi.org/10.1016/j.buildenv.2016.06.027>
- [7] Toparlak, Y., Blocken, B., Maiheu, B., van Heijst, G. J. F. "A review on the CFD analysis of urban microclimate", *Renewable and Sustainable Energy Reviews*, 80, pp. 1613–1640, 2017. <https://doi.org/10.1016/j.rser.2017.05.248>
- [8] Tominaga, Y., Stathopoulos, T. "CFD simulation of near-field pollutant dispersion in the urban environment: A review of current modeling techniques", *Atmospheric Environment*, 79, pp. 716–730, 2013. <https://doi.org/10.1016/j.atmosenv.2013.07.028>
- [9] Blocken, B. "LES over RANS in building simulation for outdoor and indoor applications: A foregone conclusion?", *Building Simulation*, 11(5), pp. 821–870, 2018. <https://doi.org/10.1007/s12273-018-0459-3>
- [10] Franke, J., Hellsten, A., Schlünzen, H., Carissimo, B. "Best practice guideline for the CFD simulation of flows in the urban environment", COST European Cooperation in Science and Technology, hal-04181390, 2007. [online] Available at: <https://hal.science/hal-04181390/document> [Accessed: 01 December 2023]

- [11] Tominaga, Y., Mochida, A., Yoshie, R., Kataoka, H., Nozu, T., Yoshikawa, M., Shirasawa, T. "AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings", *Journal of Wind Engineering and Industrial Aerodynamics*, 96(10–11), pp. 1749–1761, 2008.
<https://doi.org/10.1016/j.jweia.2008.02.058>
- [12] Blocken, B., Gualtieri, C. "Ten iterative steps for model development and evaluation applied to Computational Fluid Dynamics for Environmental Fluid Mechanics", *Environmental Modelling & Software*, 33, pp. 1–22, 2012.
<https://doi.org/10.1016/j.envsoft.2012.02.001>
- [13] Blocken, B. "Computational Fluid Dynamics for urban physics: Importance, scales, possibilities, limitations and ten tips and tricks towards accurate and reliable simulations", *Building and Environment*, 91, pp. 219–245, 2015.
<https://doi.org/10.1016/j.buildenv.2015.02.015>
- [14] Cabaneros, S. M., Calautit, J. K., Hughes, B. R. "A review of artificial neural network models for ambient air pollution prediction", *Environmental Modelling & Software*, 119, pp. 285–304, 2019.
<https://doi.org/10.1016/j.envsoft.2019.06.014>
- [15] Masood, A., Ahmad, K. "A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance", *Journal of Cleaner Production*, 322, 129072, 2021.
<https://doi.org/10.1016/j.jclepro.2021.129072>
- [16] Zhang, B., Rong, Y., Yong, R., Qin, D., Li, M., Zou, G., Pan, J. "Deep learning for air pollutant concentration prediction: A review", *Atmospheric Environment*, 119347, 2022.
<https://doi.org/10.1016/j.atmosenv.2022.119347>
- [17] Zaini, N., Ean, L. W., Ahmed, A. N., Malek, M. A. "A systematic literature review of deep learning neural network for time series air quality forecasting", *Environmental Science and Pollution Research*, 29(4), pp. 4958–990, 2022.
<https://doi.org/10.1007/s11356-021-17442-1>
- [18] Papp, B., Kristóf, G., Istók, B., Koren, M., Balczó, M., Balogh, M. "Measurement-driven Large Eddy Simulation of dispersion in street canyons of variable building height", *Journal of Wind Engineering and Industrial Aerodynamics*, 211, 104495, 2021.
<https://doi.org/10.1016/j.jweia.2020.104495>
- [19] Chang, J. C., Hanna, S. R. "Air quality model performance evaluation", *Meteorology and Atmospheric Physics*, 87(1), pp. 167–196, 2004.
<https://doi.org/10.1007/s00703-003-0070-7>
- [20] Papp, B., Kristóf, G. "Building Patterns Favorable for Air Quality: A Paramater Study Using LES", [pdf] In: *Proceedings of the Conference on Modelling Fluid Flow CMFF'22*, Budapest, Hungary, 2022, pp. 443–456. ISBN 978-963-421-881-4 Available at: https://www.cmff.hu/papers/CMFF22_Final_Paper_PDF_84.pdf [Accessed: 01 December 2023]
- [21] Kristóf, G., Papp, B. "Application of GPU-based large eddy simulation in urban dispersion studies", *Atmosphere*, 9(11), 442, 2018.
<https://doi.org/10.3390/atmos9110442>