22nd International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
10-14 June 2024, Pärnu, Estonia

CONCENTRATION FLUCTUATIONS OF A PASSIVE SCALAR IN A REGULAR STREET NETWORK

Karine Klippel[1], Elisa Valentim Goulart[1], Torsten Auerswald[2], Neyval Costa Reis Junior[1], Omduth Coceal[2]

[1]Federal University of Espirito Santo, Brazil
[2]University of Reading, UK

Abstract: We analyze data generated from a direct numerical simulation (DNS) of passive scalar dispersion in a regular street network comprised of rectangular buildings to characterize concentration fluctuations and their dependence on spatial location. The scalar was released from a point source near the ground in an intersection and the forcing wind direction was 45 degrees to the street direction. The moments and probability density functions (PDFs) of concentration were computed from the DNS data and the empirically derived PDFs were fitted to four commonly used probability distributions in urban dispersion applications: gamma, beta, lognormal, and Weibull. The visual fit to the histograms was satisfactory for all distributions at most sampled locations. Performance metrics such as FAC2, FB, NMSE, RE, and R for percentiles ranging from 75 to 99 (which represent the upper tail of the data distribution), gave results consistently within the best-expected range for all distributions, in agreement with prior findings in the literature. No single distribution exhibited significantly superior performance. The mean and variance of the distributions computed from the data compared well, in general, with those obtained from the theoretical distributions. However, the skewness and kurtosis showed large discrepancies, especially for the lognormal distribution. All the distributions modelled the empirical PDFs within the plume center better compared to the plume edges.

Key words: *Direct numerical simulations (DNS), street network, stochastic concentration*

INTRODUCTION

Given the risks associated with pollutant exposure to human health and the environment in urban areas, it is essential not only to document and model average concentrations but also to analyze their fluctuations to assess, for example, the impact of extreme concentration exposure. This analysis requires knowledge of concentration statistics such as variance and higher order statistical moments, and ideally the full probability density function (PDF) of concentration. Many previous studies have explored this problem. For example, Cassiani et al. (2020) noted that most PDF models proposed for point sources in the literature are two-parameter distributions, such as the clipped-normal, gamma, Weibull, lognormal, and beta distributions. Recent field and laboratory experiments have leaned towards the gamma distribution. For instance, Efthmiou et al. (2016) tested gamma and lognormal distributions for modeling the upper tail of concentration distributions from experimental datasets, with the gamma distribution showing superior performance and leading to the development of a RANS-gamma model. Efthmiou et al. (2017) used gamma and beta distributions to assess the impact of stability conditions and source heights on statistical models derived from field experiments. The models captured the effect of stability conditions on the 99th percentile, although neither distribution precisely estimated skewness and kurtosis. Overall, beta distributions performed better across all conditions. Expanding on previous studies, this study aims to characterize concentration fluctuations in an urban area by analyzing moments and probability density within an urban street network, using data from direct numerical simulations (DNS). We investigate whether position within the plume and the type of street where the measurement point is located significantly impact the results.

METHODOLOGY

Direct numerical simulations (DNS) were performed by the University of Reading within the DIPLOS (Dispersion of Localized Releases in a Street Network) project (Auerswald et al., 2024). The domain, shown in Figure 1a, consists of a regular array of rectangular buildings elongated in y-direction and with an aspect ratio of $h/L = 0.5$ and $h/w = 1$, where h is the building height, $L = 2h$ is the length and $W = 1h$ is the width. The horizontal extent of the domain is $24h \times 24h$ and the domain height is $8h$. The simulation was run in parallel using computational block units of size $1h \times 1h \times 1h$ each resolved by $32 \times 32 \times 32$ grid points, with periodic horizontal boundaries. A sponge layer (shown in Figure 1a) was included to reduce the scalar concentration to zero in a certain distance around the source and prevent the scalar field from re-entering through the periodic boundaries. The simulations were run with a time step of $\Delta t = 0.00025\,T$ and data were collected every 20 time steps ($0.005\,T$), with $T = 1.23$ the non-dimensional time scale. The wind direction was 45 degrees to the street directions. Passive scalars were released from a ground source at an intersection, identified in Figure 1a. The releases were continuous, i.e. the release time was much larger than the travel time from the source to the receptor, and the sources were located at 4 different locations as indicated in Figure 1a. For each release an independent scalar field was simulated and its evolution in the time-varying DNS wind field integrated forward in time. The sources were implemented in the approximate shape of a circle consisting of 52 grid cells. A constant release rate per volume of $q_S = 1000$ was prescribed which results in a source rate of $Q = q_S \cdot V \cdot N = 1.59$.

We consider four probability distributions that are most commonly applied in the dispersion literature for point sources (Cassiani et al., 2020): 2-parameter gamma, lognormal and Weibull, and 4-parameter beta. Originally, the beta distribution is defined within the interval [0, 1], however, since some data may not fall within this interval, we include two additional parameters, 'loc' and 'scale', to shift and scale the distribution to fit within this interval (SciPy, 2024). The parameters were estimated from the data in each measurement point using the maximum likelihood estimation method (MLE), which involves finding the best-fitting parameters by minimizing the negative log-likelihood function.

We use the metrics defined by Chang and Hanna (2004) for evaluating the performance of air quality models. These metrics include factor of two observations (FAC2), normalized mean squared error (NMSE), fractional bias (FB), relative error (RE) and correlation coefficient (R). Efthimiou et al. (2016) and Efthimiou et al. (2017) used these measures to evaluate the performance of the statistical models based on the gamma, lognormal and beta distributions. Using the same methodology as Efthimiou et al. (2016), we calculate the inverses of the cumulative density function (ICDF) from 75th to 99th percentiles of concentration, which correspond to the upper tail of the data. We test which model distribution agrees better with the empirical distribution from the numerical simulations.

RESULTS

Figure 1b shows the time series of normalized concentration and the associated histogram within the stationary part of the time series for three sampling locations. The locations were chosen to display the three different shapes of histogram we identified within the domain. The first histogram, shape is approximately exponential, which is in general located near the source and in between the edge of the plume and outside of it. As expected, both skewness and kurtosis are very large and positive. The second one is asymmetrical, and is located in the edge of the plume. The third histogram resembles a bell curve, generally located in the center of the plume. In this case, skewness is near 0 and kurtosis is near 3, which are the expected values for the normal distribution. This behavior is similar for all height levels analyzed. At the point nearest to the source (not showed here), (x, y, z) = (3.5h, 5.5h, 0.125h), the distribution assumes an asymmetric shape, resembling a transition between exponential and bell-shaped distributions.

In the areas with an exponential-like histogram shape, we observe a time series with a wide range of concentrations, with a higher incidence of zero concentrations and a high incidence of peaks. As we move towards the center of the plume and the shape becomes more bell-shaped, the range of concentrations decreases, with fewer extreme concentration values and no instances of zero concentration. For the points

in Figure 1b, the normalized mean concentrations are 8.1e-02, 4.2e-03, and 1.0e-2, respectively. The respective variances are 1.6e-02, 4.6e-06, and 3.7e-06, highlighting that in the first case, the mean and variance are of the same order of magnitude, indicating large concentration fluctuations.

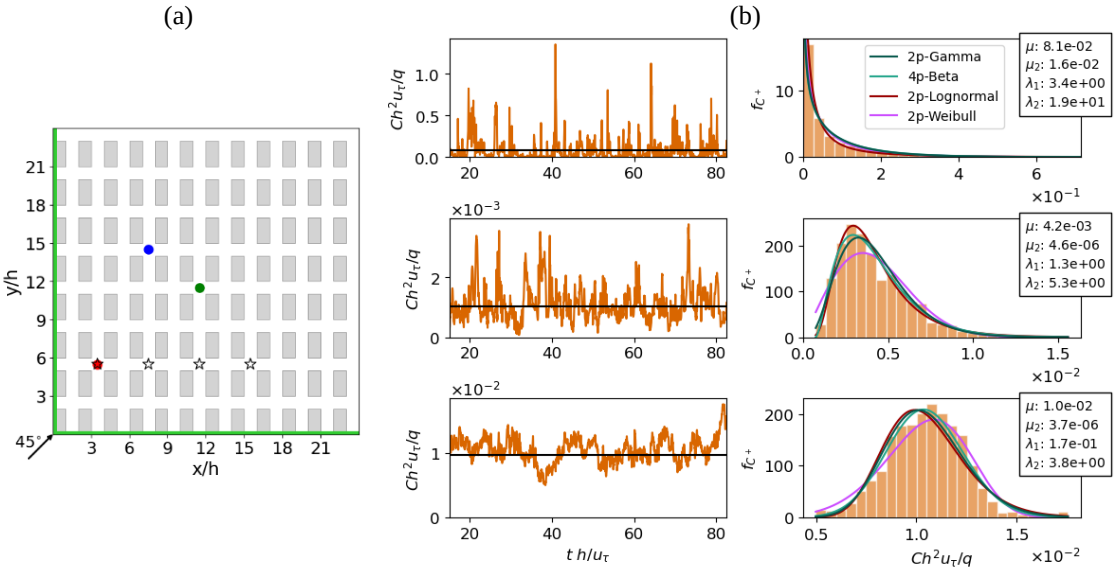

Figure 1. a) Schematic of the computational domain. The stars represent the source locations of the ensemble members. The circles indicate the location of the measurement points of **b**. The green lines represent the sponge layer for the source at $x/h = 3.5$ and $y/h = 5.5$. **b)** Stationary time series of concentration and the corresponding histograms and PDFs for 2p-gamma, 4p-beta, 2p-lognormal and 2p-Weibull distributions at z = 0.5h and: first row) x = 3.5h, y = 5.5h (indicated by a red circle in **a**); second row) x = 11.5h, y = 11.5h (indicated by a green circle in **a**); third row) x = 7.5h, y = 15.5h (indicated by a blue circle in **a**). μ, μ_2, λ_1 and λ_2 represent the four first moments mean, variance, skewness and kurtosis, respectively, for each data set.

Overall, regardless of the location within the domain, all four distributions visually demonstrated a good fit to the data. However, for the points presented in Figure 1b, the Weibull distribution performed less well at the asymmetric point, while the gamma, beta, and lognormal distributions exhibited peaks displaced more to the left than the histogram in the bell-shaped point. Using the FAC2, FB, NMSE, RE, and R metrics to evaluate the performance of the models for the 75th to 99th percentiles of the inverse cumulative density function (ICDF), corresponding to the upper tail or the highest concentration values, as done by Efthimiou et al. (2016), we find that on average, all metrics are very close to their optimal values, as shown in Table 1. For FAC2 and R, all distributions resulted in values larger than 0.99. For FB, gamma and Weibull were the best, whereas beta showed a positive value, indicating a slight overestimation on average, and lognormal was the opposite, showing a slight underestimation. NMSE and RE were near zero for all distributions.

Table 1. General performance of 2p-Gamma, 4p-Beta, 2p-Lognormal and 2p-Weibull distributions for 75th – 99th percentiles of the inverse cumulative density function (ICDF) of concentrations. The results represent the mean values considering all points within the street network.

PD	FAC2	FB	NMSE	RE	R
2p-Gamma	0.9952	0.006959	0.03563	0.04537	0.9925
4p-Beta	0.9956	0.01255	0.05017	0.005957	0.9931
2p-Lognormal	0.9987	-0.01804	0.04457	0.01443	0.9916
2p-Weibull	0.9999	0.004904	0.01513	0.00848	0.9920

However, at specific points (see QQ-Plot in Figure 2) often the models deviate from the DNS for extreme concentration values. We observe this especially for the lognormal distribution near the source (x = 3.5h, y = 5.5h, z = 05.h), which reaches much higher concentration values. In this case, the lognormal distribution consistently overestimates the higher concentrations, while the beta and Weibull distributions

underestimate them. NMSE for this measurement point is 0.043, 0.035, 0.27, and 0.019 for the gamma, beta, lognormal, and Weibull distributions, respectively. This confirms that the lognormal performed poorer than the other distributions for this case. The kurtosis at this point is very high, of order 10^4, indicating that the lognormal distribution has a tail which is not representative of the data; this could be a limitation when using this distribution. The gamma and beta in general give results closest to the DNS values across the entire concentration range.

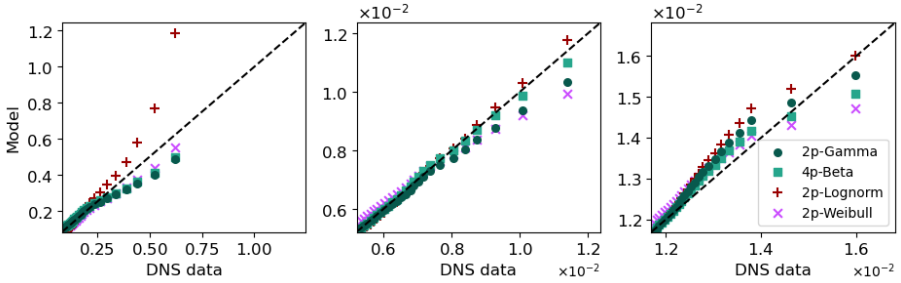

Figure 2. QQ-Plot comparing the 75th to 99th percentiles of the ICDF from DNS data and the four probability distributions for the same points of **Error! Reference source not found.** at z = 0.5h and: first column) x = 3.5h, y = 5.5h; second column) x = 11.5h, y = 11.5h; third column) x = 7.5h, y = 15.5h.

When comparing only the 99th percentile of the ICDF (Table 2), the average performance of the metrics deteriorates, particularly for NMSE and the lognormal distribution. This aligns with the observations from the QQ-Plot in Figure 2. The significant overestimation by the lognormal distribution at certain measurement points resulted in very high NMSE values. While gamma also exhibited large NMSE, FAC2, FB, RE, and R performed very well. Overall, the lognormal distribution yielded the poorest average performance in modeling the 99th percentile concentrations.

Table 2. General performance of 2p-Gamma, 4p-Beta, 2p-Lognormal and 2p-Weibull distributions for 99th percentiles of the ICDF of concentrations. The results represent the mean values considering all points within the plume.

PD	FAC2	FB	NMSE	RE	r
2p-Gamma	0.9917	0.001576	1.5623	-0.03746	0.9971
4p-Beta	0.9868	0.04154	0.3037	-0.05305	0.9986
2p-Lognormal	0.9868	-0.4682	49.8414	0.1069	0.9914
2p-Weibull	0.9967	0.04084	0.3220	-0.08285	0.9988

Evaluating the models' performance in estimating the first four moments (Figure 3), we observe that the mean is estimated very accurately, with only a slight discrepancy observed at certain points for the beta and lognormal distributions. However, the discrepancies are larger for the higher moments. For the variance, the agreement between DNS and models is good at most sampling points, except in some cases where the lognormal distribution significantly overestimates. For skewness, the beta distribution exhibited the best performance overall, although it showed a large underestimation for high skewness values. Gamma and lognormal distributions tended to overestimate smaller skewness values, with lognormal showing particularly significant overestimation for high DNS skewness values. Weibull underestimated skewness in many locations. For kurtosis, all models exhibited significant scatter around the ideal agreement line. For small kurtosis values (approximately less than 4), the agreement is satisfactory. Gamma, beta and Weibull underestimated the highest DNS kurtosis, while lognormal showed the worst performance, with significant overestimation of high kurtosis values.

CONCLUSIONS

In this work we have explored the spatial behavior of the concentration probability density functions (PDFs) in an urban street network using DNS data and compared with common PDF models. The results will aid the development of stochastic street network dispersion models. Overall, the performance of all models for the 75th-99th inverse cumulative distribution function (ICDF) of concentrations was satisfactory. However,

when considering only the 99th percentile, which represents extreme concentrations, the models' performance decreased but remained acceptable, except for the lognormal distribution, which exhibited significant discrepancies between the model and DNS data. For the first four moments, gamma and beta distributions generally demonstrated the best performance. The Weibull distribution also performed well for mean and variance. However, for the highest moments, skewness and kurtosis, the models diverged considerably from the DNS data for many locations, with the lognormal distribution in particular showing very large discrepancies. In summary, it can be concluded that gamma and beta distributions exhibited the best overall performance across the entire domain, while the lognormal distribution performed poorly, particularly in representing the upper tail, i.e., extreme values.

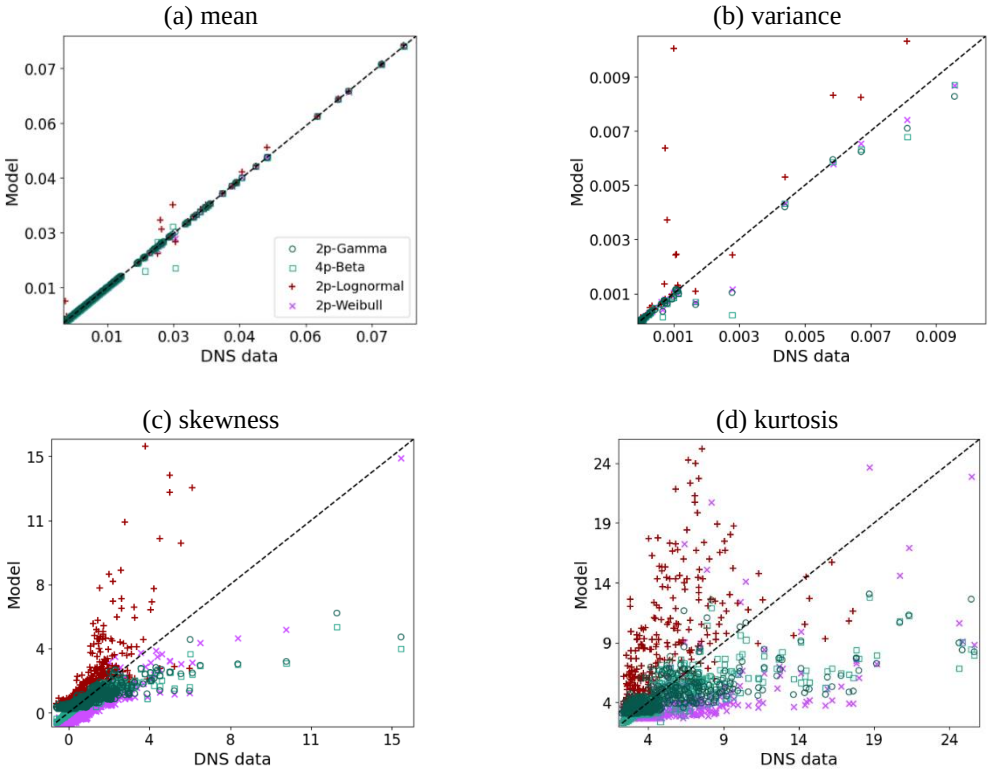

Figure 3. Comparison of the first four moments calculated from the DNS data with those estimated from theoretical distributions: a) mean, b) variance, c) skewness and d) kurtosis.

REFERENCES

Auerswald, T., Klippel, K., Thomas, T. G., Goulart, E. V., Carpentieri, M., Hayden, P. and Coceal, O., 2024: Effect of Flow Variability on Dispersion of Continuous and Puff Releases in a Regular Street Network. Boundary-Layer Meteorology, 190(4), 20.

Cassiani M., Bertagni M.B., Marro M. and Salizzoni P., 2020: Concentration fluctuations from localized atmospheric releases. Boundary-Layer Meteorology, 177(2):461–510.

Chang, J. C., and Hanna, S. R., 2004: Air quality model performance evaluation. Meteorology and Atmospheric Physics, 87(1), 167-196.

Efthimiou G., Andronopoulos S., Tolias I. and Venetsanos A., 2016: Prediction of the upper tail of concentration distributions of a continuous point source release in urban environments. Environmental Fluid Mechanics, 16:899–921.

Efthimiou G., Andronopoulos S., Bartzis J., 2017: Evaluation of probability distributions for concentration fluctuations in a building array. Physica A: Statistical Mechanics and its Applications, 484:104–116.

SciPy, 2024: scipy.stats.beta—SciPy v1.13.0. Accessed: 2024-05-03, from https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.beta.html#scipy.stats.beta.