

# On the Reconstruction of Concentration Distributions from Comparison of Deterministic Predictions to Observational Data

Igor C. Furtado<sup>\*</sup>, Bardo E. J. Bodmann, Marco T. Vilhena

Pos-Graduate Program in Mechanical Engineering, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

---

**Abstract** The dispersion of pollutants in the atmospheric boundary layer is a stochastic process but many approaches make use of deterministic models, such as the advection-diffusion equation, that determines average values. Comparison between observation and model prediction show a significant spread of values due to the stochastic character of the pollution dispersion phenomenon. Measured data though represent only one sample of an unknown distribution. Thus, the present article is a first attempt to reconstruct at least some of the pollutant concentration distribution properties from the comparison of deterministic predictions to observed concentrations under specific micro-meteorological conditions. The experimental data are the findings of the Copenhagen campaign. We show the scatter plot of observed versus predicted ground level concentrations from which distributional properties are extracted by determining the distance of each plot point from the bisector, proposing a parametrization for the probability function and fit the discrete set of data points. The probability density function obtained from the probability distribution shows a narrow peak centered at zero besides a second smaller but displaced contribution. A reconstructed distribution symmetrically around zero signifies, that the model describes the average values of the distribution with fairly good fidelity and the width could be used as an approximation for the second statistical moment. The distribution which is not centered at the origin indicates either missing physics in the model, or failures in the measuring procedure. The reconstructed distribution with correlation less than one shows the aforementioned stochastic character of the phenomenon. Although applied to a specific experiment and using one deterministic model the reconstruction method is general and can be applied to other scenarios in an analogue fashion.

**Keywords** Dispersion of pollutants, Stochastic phenomenon, Deterministic predictions, Probability density function

---

## 1. Introduction

The dispersion of pollutants in the atmospheric boundary layer is a stochastic process which varies unpredictably over a time interval. Although it is a random phenomenon, in many approaches is modeled by deterministic models, in our case given by the advection-diffusion equation, that determine average values.

A variety of studies have shown that in almost all cases a significant spread of values in the observed-predicted concentration plot occurs which on the one side may have influences by model uncertainties but on the other hand stems from the stochastic character of the dispersion phenomenon.

Independent of the model nature and its predictability, it is noteworthy, that measured data in general may not be repeated with equal micro-meteorological conditions, so that

often a measured concentration is only one sample of an unknown distribution, with its related statistical moments.

The model choice in favor of the advection-diffusion equation is based on formal progress in solving this equation in an (semi-)analytical fashion and although predicting only average values shows fairly good agreement with measurements. The short coming of lacking information of higher statistical moments in the model is to its simplified closure by K-theory. This theory was extended by parameterizations that besides the vertical eddy diffusion profile also takes into account the changes in the diffusion parametric functions with distance from the pollution source.

In this line the present article is a first attempt to reconstruct at least some of the pollutant concentration distribution properties from the comparison of deterministic predictions to observed concentrations for observations with determined micro-meteorological conditions. The same conditions are used for the deterministic dispersion simulations. In principle, the presented procedure may be applied to any other dispersion model, so that in the present discussion, focus is put on the reconstruction procedure and less on the pertinent question as to what is the most

---

\* Corresponding author:

igorjara@gmail.com (Igor C. Furtado)

Published online at <http://journal.sapub.org/ajee>

Copyright © 2016 Scientific & Academic Publishing. All Rights Reserved

appropriate model for describing the phenomenon in consideration.

## 2. The Model

The model used in this approach is the advection-diffusion equation obtained by combining the continuity equation with Fick's Law:

$$\frac{\partial \bar{c}}{\partial t} + \bar{u} \frac{\partial \bar{c}}{\partial x} = \frac{\partial}{\partial y} \left( K_y \frac{\partial \bar{c}}{\partial y} \right) + \frac{\partial}{\partial z} \left( K_z \frac{\partial \bar{c}}{\partial z} \right) \quad (1)$$

where  $x$  is the direction of the mean wind,  $\bar{c}$  is the mean concentration,  $\bar{u}$  is the mean wind,  $K_y$  and  $K_z$  are the turbulent diffusion coefficients and in this case  $K_z$  dependent on the height  $z$  and time  $t$ . Equation (1) is subject to the initial condition of zero concentration and zero flow in the contours of the area of interest. The source is  $\bar{u}\bar{c}(0, y, z, t) = Q\delta(y - y_0)\delta(z - H_s)$  in  $x = 0$ , where  $Q$  is the rate of emission of the pollutant,  $H_s$  is the height of the source,  $\delta$  is the Dirac delta function,  $y_0$  is characterized by the crosswind coordinate of the source and  $h$  is the height of the boundary layer.

The success of a deterministic model is strongly influenced by the way the turbulent parameters are related to the understanding of the phenomenon. The diffusion coefficient, with temporal dependence, derived by Degrazia et al. [1] is based on diffusion theory, statistical Taylor approach and a spectral model of kinetic energy. This methodology derived from convective and moderately unstable conditions, provide diffusion coefficients characteristic described in terms of speed and duration of energy containing eddies and scales can be expressed by algebraic formulation as:

$$\frac{K_\alpha}{w_* h} = \frac{0.58c_i\psi^{2/3}(z/h)^{4/3}X \left[ 0.55(z/h)^{2/3} + 1.03c_i^{1/2}\psi^{1/3}(f_m^*)_i^{2/3}X \right]}{\left[ 0.55(z/h)^{2/3}(f_m^*)_i^{1/3} + 2.06c_i^{1/2}\psi^{1/3}(f_m^*)_i X \right]^2} \quad (2)$$

Here  $\alpha = x, y, z, i = u, v, w$ , and

$$c_i = \alpha_i(0.5 \pm 0.05)(2\pi\kappa)^{-2/3}$$

$\alpha_i = 1, \frac{4}{3}, \frac{4}{3}$  respectively for  $u, v$  and  $w$  [2],  $\kappa = 0.4$  is the von Karman constant,  $(f_m^*)_i$  is the normalized frequency of the spectral peak,  $h$  is the height of the atmospheric boundary layer,  $w_*$  is the convective velocity scale,  $\psi$  is the non-dimensional molecular dissipation rate function and  $X = \frac{xw_*}{\bar{u}h}$  is the non-dimensional travel time, where  $\bar{u}$  is the horizontal mean wind speed.

For horizontal homogeneity the convective boundary layer evolution is driven mainly by the vertical transport of heat. As a consequence of this buoyancy driven ABL, the vertical dispersion process of scalars is dominant when compared to the horizontal. Therefore, the present analysis will focus on the travel time dependent vertical eddy diffusivity. This vertical eddy diffusivity can be derived from Eq. (2) by assuming that  $c_w = 0.36$  and

$$(f_m^*)_w = 0.55 \left( \frac{z}{h} \right) \left[ 1 - \exp\left( \frac{-4z}{h} \right) - 0.0003 \exp\left( \frac{8z}{h} \right) \right]^{-1}$$

for the vertical component [3].

Furthermore, considering the horizontal plane, we can idealize the turbulent structure as a homogeneous one with the length scale of energy-containing eddies being proportional to the convective boundary layer height  $h$ . Thus, for the lateral eddy diffusivity we used the asymptotic behaviour of Eq. (2) for large diffusion travel times with  $c_v = 0.36$  and  $(f_m)_v = 0.66 \frac{v}{h}$  [4].

Finally, the dissipation function  $\psi$  according to [5, 6] has the form

$$\psi^{(1/3)} = \left[ \left( 1 - \frac{z}{h} \right)^2 - \left( \frac{z}{L} \right)^{-2/3} + 0.75 \right]^{1/2}$$

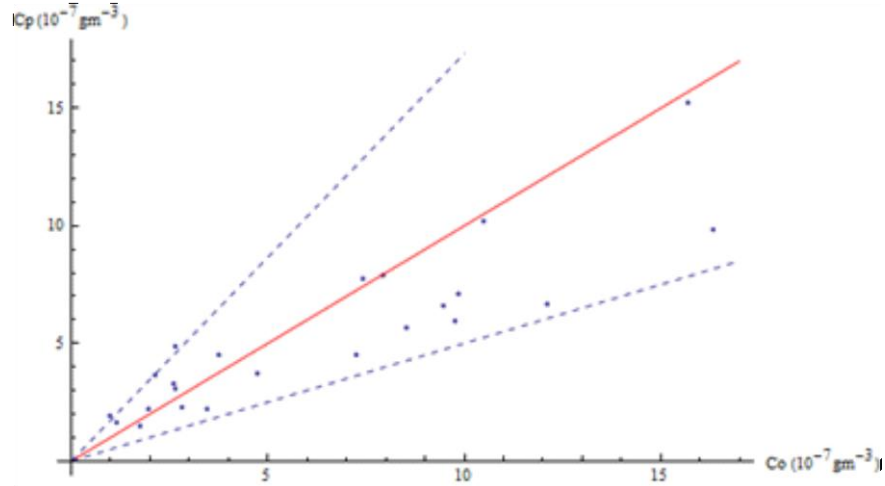
where  $L$  is the Monin-Obukhov length in the surface layer.

### 2.1. Results 3D-GILTT

Equation (1) was solved analytically in [7] by combining the idea of the decomposition method and GILTT. Upon applying the idea of the decomposition reduces the advection-diffusion equation with time dependence of the diffusion coefficient into a set of recursive equations depending only on the diffusion coefficient in the spatial variable  $z$  (besidest), which is then directly solved by the aforementioned method.

**Table 1.** Pollutant concentrations for nine runs at various positions of the Copenhagen experiment and model prediction by the 3D-GILTT approach with time dependence of eddy diffusivity. Concentration is divided by the emission rate

Run	Distance (m)	Observed ( $10^{-7} s.m^{-3}$ )	Predictions ( $10^{-7} s.m^{-3}$ )
1	1900	10.5	10.21
1	3700	2.14	3.65
2	2100	9.85	7.09
2	4200	2.83	2.28
3	1900	16.33	9.80
3	3700	7.95	7.88
3	5400	3.76	4.50
4	4000	15.71	15.24
5	2100	12.11	6.68
5	4200	7.24	4.48
5	6100	4.75	3.73
6	2000	7.44	7.78
6	4200	3.47	2.20
6	5900	1.74	1.46
7	2000	9.48	6.60
7	4100	2.62	3.32
7	5300	1.15	1.65
8	1900	9.76	5.97
8	3600	2.64	4.87
8	5300	0.98	1.91
9	2100	8.52	5.64
9	4200	2.66	3.04
9	6000	1.98	2.19



**Figure 1.** Observed ( $C_o$ ) and predicted ( $C_p$ ) scatter plot of centerline concentration using the Copenhagen dataset. Data between dotted lines correspond to ratio  $C_o/C_p \in [0.5, 2]$

Evaluation of the performance of the 3D-GILTT and simulation of the dispersion of contaminants in the atmospheric boundary layer is compared to the performance of the solutions against experimental concentrations of the Copenhagen campaign [8].

The vertical wind speed profile used in the simulations is described by the power law that follows

$$\frac{\bar{u}}{\bar{u}_1} = \left(\frac{z}{z_1}\right)^n \quad (3)$$

where  $\bar{u}$  and  $\bar{u}_1$  are the average speeds of the horizontal wind at heights  $z$  and  $z_1$ , respectively and  $n$  is an exponent which is related to the intensity of turbulence [9], where according to the author  $n = 0.1$  is valid for a power wind profile in unstable conditions. Table 1 shows the comparison of 3D-GILTT simulated results against experimental data.

Figure (1) shows the scatter plot of ground level concentrations observed versus model simulations by 3D-GILTT and normalized by the emission rate. For statistical comparisons of performance between the Copenhagen experimental data and the results from the GILTT the following statistical indices are used for evaluation.

$$\text{NMSE}(\text{normalized mean square error}) = \overline{(C_o - C_p)^2} / \overline{C_o C_p},$$

$$\text{COR}(\text{correlation coefficient}) = 0.5(C_p/C_o),$$

$$\text{FS}(\text{fractional standard deviations}) = (\sigma_o \sigma_p) / 0.5(\sigma_o + \sigma_p)$$

where the subscripts  $o$  and  $p$  refer to observed and predicted quantities, respectively, and the overbar indicates an averaged value. An acceptable result should reflect small values for the indices NMSE and FS and above 0.8 for the index COR. Table 2 displays the model performance evaluated by the statistical indices.

**Table 2.** Statistical comparison between the 3D-GILTT model and the Copenhagen dataset

Model	NMSE	COR	FS
3D-GILTT	0.17	0.89	0.32

A critical comment is in order here. From the data fitting point of view the ideal values of vanishing NMSE and FS are desired, whereas a correlation shall be close to unity. Recalling, that for the considered type of problem the measured phenomenon has turbulence characteristics where statistical moments beyond the first one are significant for an adequate distributional description but the underlying theoretical model considers only average values, consequently the best one may expect is a correlation from 0.8 on and above depending on the natural width and kurtosis of the measured distribution, which is not present in the deterministic model.

### 3. Distribution Reconstruction

As a first approach towards a procedure to recover some distributional properties from the comparison with the experimental data set, one determines the distances of each point  $(C_o, C_p)$  to the ideal line, i.e. the bisector. To this end we introduce a simple change of the canonical basis  $\alpha = [(1,0), (0,1)]$  to the orthogonal base with  $\beta = \left[\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{-1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)\right]$ . Thus, the distance from each respective point to the bisector is expressed in the new basis, by

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = [I]_{\alpha}^{\beta} \begin{bmatrix} C_o \\ C_p \end{bmatrix} \quad (4)$$

where  $\Delta_y$  is the distance of  $(C_o, C_p)$  to the bisector line and  $[I]_{\alpha}^{\beta}$  is the matrix representing the change of basis. Table (3) shows the distance of each point resulting from the above procedure.

Once  $\Delta$  is known, one can construct a probability distribution function

$$F_x(X) = P(\omega X(\omega) \subseteq X) \quad (5)$$

where  $F_x(X)$  shall satisfy

- i)  $F_x(-\infty) = 0$ ,
- ii)  $F_x(\infty) = 1$ ,

iii)  $F_X$  is monotonically increasing.

**Table 3.** Distances  $\Delta$  of each point  $(C_o, C_p)$  to line of  $45^\circ$

$(C_o, C_p)$	$\Delta$
(10.5, 10.21)	0.145
(2.14, 3.65)	-0.755
(9.85, 7.09)	1.38
(2.83, 2.28)	0.275
(16.33, 9.80)	3.265
(7.95, 7.88)	0.035
(3.76, 4.50)	-0.370
(15.71, 15.24)	0.235
(12.11, 6.68)	2.715
(7.24, 4.48)	1.380
(4.75, 3.73)	0.510
(7.44, 7.78)	-0.169
(3.47, 2.20)	0.635
(1.74, 1.46)	0.140
(9.48, 6.60)	1.440
(2.62, 3.32)	-0.350
(1.15, 1.65)	-0.250
(9.76, 5.97)	1.895
(2.64, 4.87)	-1.115
(0.98, 1.191)	-0.465
(8.52, 5.64)	1.440
(2.66, 3.04)	-0.189
(1.98, 2.19)	-0.105

Figure (2) shows the probability distribution function. To determine the probability density function, it is necessary to transform our discrete variable  $X$  in a continuous variable, which may be attained introducing a parametrization compatible with the restrictions for  $F_X(-\infty)$ . The

distribution profile presented in figure 2 may be parametrized by an approximation using a superposition of hyperbolic tangent functions. In the present case the implementation needed only the sum of two hyperbolic tangent functions.

$$F_X(X) = \frac{1}{2} (a_1 (\tanh(a_2 X) + 1) + (1 - a_1) (\tanh(a_3 - a_4 X) + 1)) \tag{6}$$

That is,  $F_X$  is defined such that  $a_1$  refers to an amplitude,  $a_2$  and  $a_4$  are factors related to the growth rate of  $X$  and  $a_3$  determines translation in  $X$  for the second function.

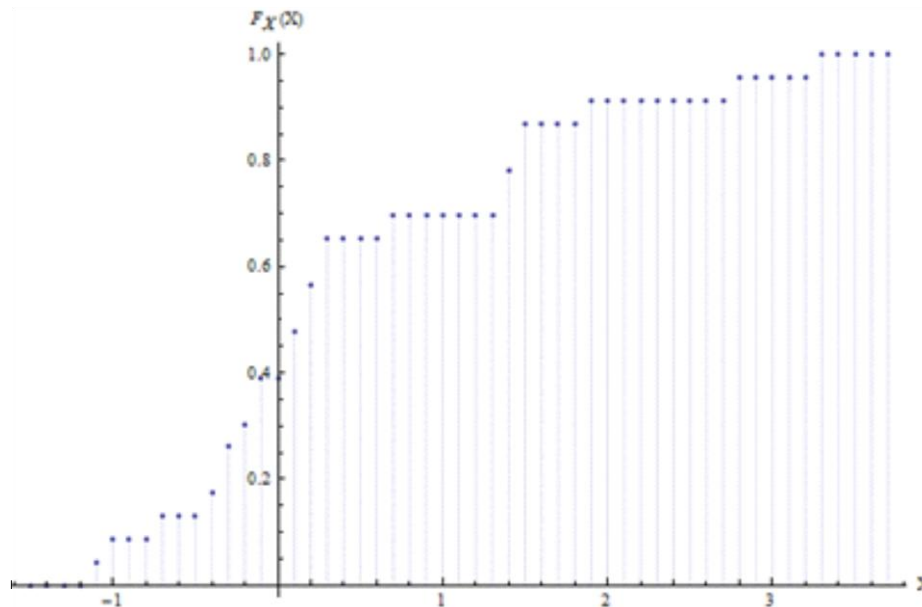
A best fit for the discrete set of data points may be obtained upon minimizing the squares of differences between the data and the parametrized probability distribution function. A variety of minimizing procedure are at hand, for simplicity we use as an optimization the method of least squares. The least squares problem is to find the parameter set  $a$  that minimizes the expression

$$G(a) = \frac{1}{2} \sum_{k=1}^m (r_k(a))^2 = \frac{1}{2} |r(a)|^2 \tag{7}$$

where  $a = [a_1, a_2, a_3, a_4]^T$  and  $r(x)$  represents the difference between the data set and the proposed function, assuming  $m$  points to be adjusted to the curve.

The solution of the least squares problem was implemented using Newton's method although it is known to work with local convergence only. Due to that fact a preprocessing was applied for the iterative process such that the initial values for the parameter set of the Newton iteration was within the radius of convergence.

The approximation to Newton's iterative process was established graphically and the resulting interpolation parameter set is given by  $a_1 = 0.787834$ ,  $a_2 = 1.33617$ ,  $a_3 = 2.31281$  and  $a_4 = 1.09579$ . Figure (3) shows the agreement of the approximate function with the set of predicted data.



**Figure 2.** Probability distribution function

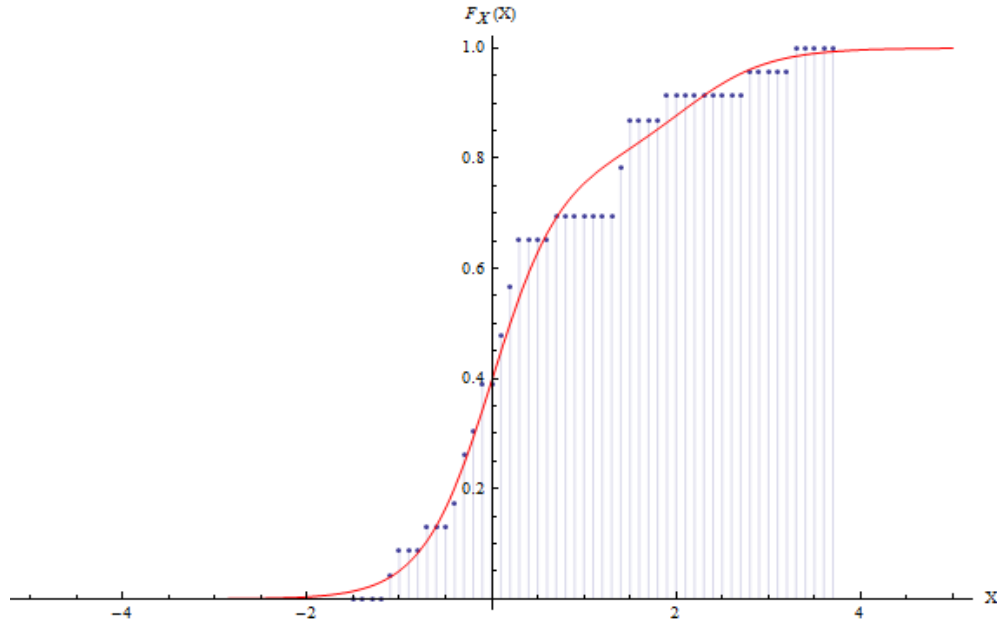


Figure 3. Comparison of the distribution function proposed and the data points from comparison

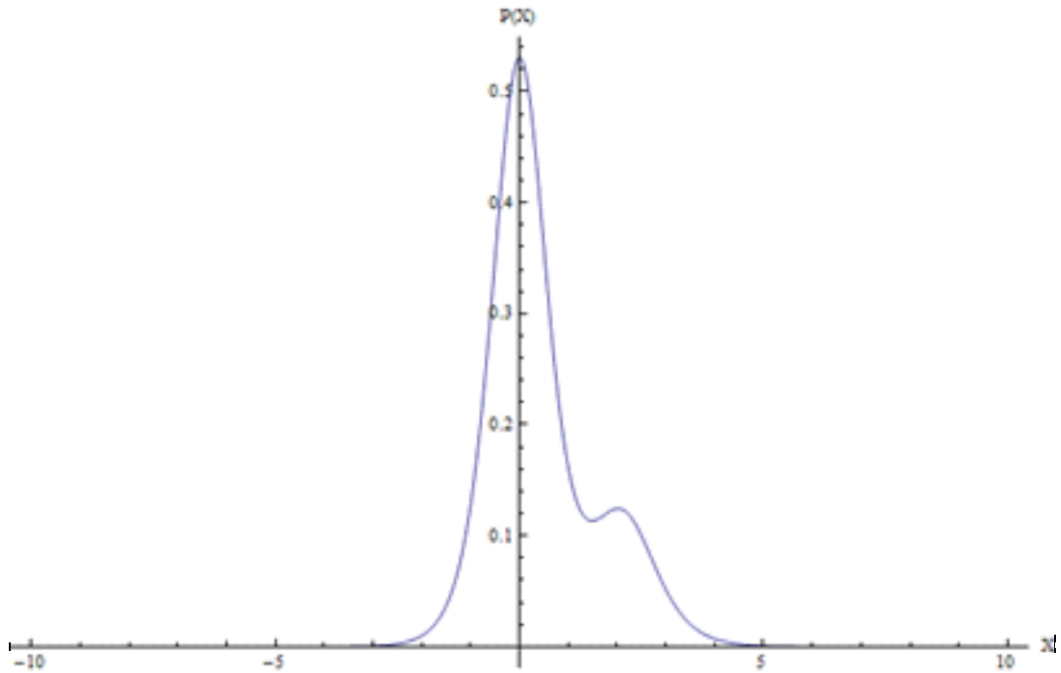


Figure 4. Probability density function P(X)

From these findings one can calculate the probability density function by the well known relation between the probability distribution and the probability density function,

$$P(x) = \frac{dF_X(X)}{dX} \tag{8}$$

where  $P(\pm\infty) = 0$  holds and further  $P(X)$  is normalized.  $\int_{-\infty}^{\infty} P(x)dx = 1$ .

The probability density function is shown in Fig. (4). By inspection one observes a narrow peak centered at zero. If the reconstructed distribution was symmetrically around zero, one could conclude that the model describes the

average values of the distribution with fairly good fidelity and the width could be used as an approximation for the second statistical moment. However, figure 2 shows also a second contribution with a maximum approximately at  $X=2$ . This distribution which is not centered at  $X=0$  indicates that there is either some physics missing in the model, or that there occurred errors in the measuring procedure. Without further information on data acquisition or turbulence parameterization no conclusion may be drawn so far. In any case the reconstructed distribution clearly shows the aforementioned stochastic character of the phenomenon and explains why one should expect values for  $COR < 1$ .

From the obtained probability density function one may obtain confidence levels for probabilities of sample data. This may be used as a further plausibility test for the reconstructed distribution. Such a procedure is necessary due to the fact that the observed data are only one sample of an unknown distribution, where the reconstructed distribution is a first estimate. The width of the distribution indicates how significant higher stochastic moments are in order to describe the turbulent phenomenon. For instance a narrow width indicates that a deterministic model is sufficient whereas a broad one indicates that a model extension is of need, i.e. additional equations for the higher statistical moments shall be proposed. The confidence level is determined by

$$\int_{\bar{x}-\bar{\Delta}}^{\bar{x}+\bar{\Delta}} P(x)dx = \nu \quad (9)$$

where  $\bar{\Delta}$  is the expectation value, i.e. the value an exact deterministic model would predict. Table 4 also shows that because of the non-centered distribution the confidence interval is not symmetric, due to the already mentioned model or measuring errors. It is noteworthy, that the distribution width and associated confidence level is not a measure for the reliability of the model but a manifestation of the uncertainties of the turbulence parametrization or experimental data quality.

**Table 4.** Confidence levels and corresponding confidence interval

95%	[-1.59809, 2.79609]
80%	[-0.787786, 2.79609]
70%	[-0.37872, 2.27413]
50%	[-0.162961, 1.43876]
30%	[0.0281579, 0.777125]
10%	[0.222279, 0.44912]

## 4. Conclusions

The present work focussed on the pertinent question of describing the stochastic phenomenon of pollution dispersion in the planetary boundary layer by a deterministic model. Although being one of the oldest enigmas of fundamental and applied science, up to now no model exists that is able to reliably describe effects of turbulent flows. Moreover, in the planetary boundary layer thermal effects are absorbed in the eddy diffusivities by their stability regime dependent parametrizations, which have their origin in the closure problem a crucial step along the derivation of the model. As a first attempt to recover the lost statistical moments by the modelling details a comparison of experimental data to the predicted values by the deterministic model was performed. The procedure used a projection of the dispersed data points, i.e. observed versus predicted concentrations, which resulted in a discrete sample that were used to parametrize a probability distribution function. From the analysis of the obtained probability

density function one could identify at least two distributions, the first one symmetric around the bisector, that corresponded to data compatible with the model and the second one that indicated model or measurement errors. Evidently the present discussion leaves open a variety of questions, such as how to discriminate model and measurement failures, though we believe to have done a first step into a new direction.

## ACKNOWLEDGEMENTS

The authors acknowledge the financial support provided by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). Some of the authors thank LGSA, Tevisa and ANEEL P&D.

## REFERENCES

- [1] G.A. Degrazia, D.M. Moreira, C.R.J. Campos, J.C. Carvalho, M.T. Vilhena, Comparison between an integral and algebraic formulation for the eddy diffusivity using the Copenhagen experimental dataset, *Il Nuovo Cimento 25C* (2002) 207–218.
- [2] F.H. Champagne, C.A. Friche, J.C. Larve, J.C. Wyngaard, Flux measurements flux estimation techniques, and fine scale turbulence measurements in the unstable surface layer over land, *J. Atmospheric Sci.* 34 (1977) 515–520.
- [3] G.A. Degrazia, D. Anfossi, J.C. Carvalho, C. Mangia, T. Tirabassi, Turbulence parameterization for PBL dispersion models in all stability conditions, *Atmos. Environ.* 34, 3575–3583, 2000.
- [4] G.A. Degrazia, H.F. Campos Velho, J.C. Carvalho, Nonlocal exchange coefficients for the convective boundary layer derived from spectral properties, *Contrib. Atmos. Phys.* 70 (1997) 57–64.
- [5] J. Højstrup, Velocity spectra in the unstable boundary layer, *J. Atmospheric Sci.* 39 (1982) 2239–2248.
- [6] G.A. Degrazia, C. Mangia, U. Rizza, A comparison between different methods to estimate the lateral dispersion parameter under convective conditions, *J. Appl. Meteorol.* 37 (1998) 227–231.
- [7] Vilhena, M. T., Buske, D., Degrazia, G. A., Quadros, R. S., 2011. An analytical model with temporal variable eddy diffusivity applied to contaminant dispersion in the atmospheric boundary layer. *Physica A* 391 (2012) 2576–2584.
- [8] S.E. Gryning, E. Lyck, Atmospheric dispersion from elevated source in an urban area: comparison between tracer experiments and model calculations, *J. Appl. Meteorol.* 23 (1984) 651–654.
- [9] J.S. Irwin, A theoretical variation of the wind profile power-law exponent as a function of surface roughness and stability, *Atmos. Environ.* 13 (1979) 191–194.