

# A statistical model for the prediction of wind-speed probabilities in the atmospheric surface layer

G. C. Efthimiou<sup>1a</sup>, D. Hertwig<sup>b</sup>, S. Andronopoulos<sup>a</sup>, J.G. Bartzis<sup>c</sup>, O. Coceal<sup>d</sup>

<sup>a</sup>Environmental Research Laboratory, INRASTES, NCSR Demokritos, Patriarchou Grigoriou & Neapoleos Str., 15310, Aghia Paraskevi, Greece.

<sup>b</sup>Department of Meteorology, University of Reading, Reading, UK.

<sup>c</sup>University of Western Macedonia, Dept. of Mechanical Engineering, Sialvera & Bakola Str., 50100, Kozani, Greece.

<sup>d</sup>National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK.

## Abstract

Wind fields in the atmospheric surface layer (ASL) are highly three-dimensional and characterized by strong spatial and temporal variability. For various applications such as wind comfort assessments and structural design, an understanding of potentially hazardous wind extremes is important. Statistical models are designed to facilitate conclusions about the occurrence probability of wind speeds based on the knowledge of low-order flow statistics. Being particularly interested in the upper tail regions we show that the statistical behavior of near-surface wind speeds is adequately represented by the Beta distribution. By using the properties of the Beta probability density function in combination with a model for estimating extreme values based on readily available turbulence statistics, it is demonstrated that this novel modelling approach reliably predicts the upper margins of encountered wind speeds. The model's basic parameter is derived from three substantially different calibrating datasets of flow in the ASL originating from boundary-layer wind-tunnel measurements and direct numerical simulation. Evaluating the model based on independent field observations of near-surface wind speeds showed a high level of agreement between the statistically modelled horizontal wind speeds and measurements. The results show that, based on the knowledge of only a few simple flow statistics (mean wind speed, wind speed fluctuations and integral time scales), the occurrence probability of velocity magnitudes at arbitrary flow locations in the ASL can be estimated with a high degree of confidence.

## Keywords

Atmospheric surface layer; Beta distribution; Direct numerical simulation; Extreme wind speeds; Field experiment; Probability density function; Weibull distribution; Wind tunnel

## 1. Introduction

---

<sup>1</sup> Corresponding author. Tel: +30 2106503405. E-mail address: [gefthimiou@ipta.demokritos.gr](mailto:gefthimiou@ipta.demokritos.gr) (G. C. Efthimiou)

34 Knowledge of the probability density function (p.d.f.) of wind speed in the atmospheric  
 35 boundary layer (ABL) is necessary for many applications such as the estimation of the wind  
 36 energy potential at a particular location (Sarkar et al., 2011) or wind comfort and safety  
 37 studies (e.g. Janssen et al., 2014). Furthermore, safety considerations in the design of  
 38 structures at exposed locations (e.g. bridges, radio masts or wind turbines) require a reliable  
 39 assessment of the maximum expected wind speeds (Brabson and Palutikof 2000).

40 For the prediction of wind gusts, i.e. sudden, brief increases of local wind speeds, various  
 41 methodologies have been proposed in the past. Zhang et al. (2013), for example, performed  
 42 an analysis of the characteristics of 1-Hz wind-speed data continuously sampled in the air  
 43 layer below 2 m. They proposed an empirical model to predict the fluctuating wind gusts of  
 44 the streamwise velocity based on friction velocity, mean wind speed and standard deviation at  
 45 2 m. Sallis et al. (2011) used a machine-learning approach to determine meaningful and  
 46 robust results of wind gusts and proposed an algorithm for application to real-time climate  
 47 data. Brasseur (2001) proposed a new wind gust estimate method where the determination of  
 48 gusts is fully based on physical considerations. The proposed approach assumes that surface  
 49 gusts result from the deflection of air parcels in the upper levels of the boundary layer, which  
 50 are brought down by turbulent eddies. The method takes into account the mean wind and the  
 51 turbulent structure of the atmosphere.

52 Over the past years, research activity in the area of wind-speed distribution modelling has  
 53 increased considerably. For the prediction of wind-speed distributions statistical models that  
 54 provide information about the local occurrence probabilities at a certain site are preferably  
 55 employed. For this task, the choice of a suitable p.d.f. is crucial. A number of previous  
 56 studies compared statistical distributions with measurements in order to examine how well  
 57 the p.d.f.s describe the statistical properties of the measured wind speed. An overview of  
 58 recent studies is presented in Table 1.

59 **Table 1.** Overview of recent studies (in chronological order) that use different statistical distributions to assess  
 60 wind speed occurrence probabilities.

Publication	Distributions	Averaging time interval of data analyzed
Donk et al. (2005)	Weibull	1 h
Carta et al. (2009)	Generalized Gamma, Gamma, Weibull, singly truncated from below normal, two components mixture Weibull, Rayleigh, beta, inverse Gaussian, lognormal	1 h
He et al. (2010)	Weibull	1 h
Morgan et al. (2011)	Gamma, lognormal, Rayleigh, log Pearson type III, Generalized Rayleigh, Generalized Gamma, Pearson type III, Weibull, Generalized normal, Wakeby,	10 min

	Kappa, bimodal Weibull mixture	
Waewsak et al. (2011)	Weibull	10 min
Aidan (2011)	Normal, Gamma, Weibull, Rayleigh	1 month
Sarkar et al. (2011)	Weibull, extreme value distribution of type I (Gumbel)	1 h
Odo et al. (2012)	Weibull	1 d
Kollu et al. (2012)	Weibull-extreme value distribution (GEV), Weibull-lognormal, GEV-lognormal	10 min
Masseran et al. (2013)	Lognormal, Weibull, Rayleigh, exponential, Gamma, inverse Gaussian, Burr, inverse Gamma	1 h
Datta and Datta (2013)	Weibull, exponentiated Weibull	1 a
Nemeş (2013)	Weibull	1 h
Indhumathy et al. (2014)	Weibull	1 h
Kidmo et al. (2015)	Weibull	1 h
Petković (2015)	Weibull	-
Men et al. (2016)	Gauss	30 min
Karthikeya et al. (2016)	Weibull	10 min
Carneiro et al. (2016)	Weibull	10 min

61

62 Table 1 shows that a large number of different p.d.f.s were previously compared with wind  
63 speed data, with the Weibull distribution overall being the most popular choice.

64 In order to assess extreme wind speeds, extreme value theory can be used (e.g. Palutikof et al.  
65 1999; Holmes and Moriarty 1999; Simiu et al. 2001). However, in this case, the successful  
66 modelling of the upper tail can often lead to an inadequate representation of the main part of  
67 the wind-speed distribution. Most of the studies listed in Table 1 used 1-h or 10-min  
68 averages. Steinkohl et al. (2010) analyzed the wind-speed measurements on a finer time scale  
69 in the so-called micrometeorological range. Their dataset consisted of observations measured  
70 in the ABL on two different days with a sampling frequency of 1 Hz. They focused on the  
71 modelling of the tail of the wind-speed distribution by using the ‘peaks-over threshold’  
72 approach of extreme value theory.

73 Other studies use nonparametric estimators of wind speed. Rozas-Larraondo et al. (2014), for  
74 example, studied a new method based on nonparametric multivariate locally weighted

75 regression for improving wind speeds forecast by a numerical weather prediction model.  
 76 Wind direction data were used to build different regression models, as a way of accounting  
 77 for the effect of surrounding topography. Recently, D'Amico et al. (2014) presented a new  
 78 nonparametric model to predict wind speeds based on semi-Markov chains. They found the  
 79 model to be able to reproduce the statistical behavior of wind speeds accurately for different  
 80 time scales when used as a forecast tool. Francisco-Fernández and Quintela-del-Río (2013)  
 81 applied nonparametric curve estimation methods to analyze time series of wind speeds,  
 82 focusing on extreme events exceeding a chosen threshold. Nonparametric methods to directly  
 83 estimate quantities such as the probability of exceedance, the quantiles or return levels or the  
 84 return periods were proposed. Moreover, bootstrap techniques were used to develop  
 85 pointwise and simultaneous confidence intervals for these functions.

### 86 **1.1. Aim of this study**

87 In this study we aim at demonstrating that occurrence probabilities of wind speeds in the  
 88 atmospheric surface layer (ASL) can be estimated based on the knowledge of low-order flow  
 89 statistics readily available from field measurements and the choice of a p.d.f. that has a finite  
 90 range. In quantitative terms we aim to predict the horizontal wind speed  $V$  within the time  
 91 interval  $\Delta\tau$  that is encountered at an arbitrary location,

$$92 \quad V(\Delta\tau) = \frac{1}{\Delta\tau} \int_{\Delta\tau} v(t) dt \quad (1)$$

93 where  $v(t)$  is the instantaneous wind speed.  $V(\Delta\tau)$  is in meteorological terminology the  $\Delta\tau$   
 94 gust.

95 The time interval  $\Delta\tau$  can signify a measurement time interval or an averaging period, as is the  
 96 case in this study. However,  $\Delta\tau$  can be assigned further significance when associating it with  
 97 typical exposure times to certain wind speeds, e.g. based on the time an individual is expected  
 98 to stay at a particular location. While mean velocities in urban areas are typically low,  
 99 turbulence levels can be significant and wind gusts can by far exceed the time-averaged wind  
 100 speeds encountered at street level, which can cause discomfort for pedestrians or result in  
 101 structural damage to buildings. The time scales associated with such gust episodes are as low  
 102 as times associated with typical pedestrian walking speeds (i.e. a few s).

103 The derivation of a new p.d.f. for  $V(\Delta\tau)$  and the associated maximum expected wind speed  
 104  $V_{max}(\Delta\tau)$  will be the main effort in the present investigation. A suitable p.d.f. should satisfy  
 105 the following criteria:

- 106 1. It should describe all parts of the wind-speed distribution accurately.
- 107 2. It should have a finite upper extreme.
- 108 3. It should be applicable to a wide range of turbulent flows in the ASL.

109 One probability distribution that fulfils these criteria is the Beta distribution.

110 The methodology employed in this study is described in detail in the next section. The  
 111 experimental and numerical wind speed databases are introduced in Sect. 3. Based on this  
 112 calibrating data, in Sect. 4 the basic parameter of the proposed model is estimated. Finally, in  
 113 Sect. 5 the methodology is validated based on independent field data.

## 114 **2. Methodology**

115 The present methodology is based on the studies by Bartzis et al. (2008) and Bartzis et al.  
 116 (2015), in which the cumulative distribution function (c.d.f.) of scalar concentrations from  
 117 point source releases of airborne materials was modelled. In the present study we apply the  
 118 same methodology to the prediction of wind speed p.d.f.s and c.d.f.s.

119 For this, we make the hypothesis that the maximum wind speed that can be encountered in  
 120 the ASL takes a finite value and that this maximum wind speed  $V_{max}(\Delta\tau)$  can be predicted by  
 121 the Bartzis et al. (2008) model. We present a detailed justification for the applicability of the  
 122 model in the following.

### 123 **2.1 Probability density function selection and parameterization**

124 It is assumed that the p.d.f. for the time-averaged horizontal wind speed  $V(\Delta\tau)$  at a certain  
 125 location is given by the Beta distribution (e.g. Gupta and Nadarajah, 2004)

$$126 \quad pdf(x) \propto x^{\alpha-1}(1-x)^{\xi-1}; \quad 0 \leq x \leq 1. \quad (2a)$$

127 In the case of our study  $x$  is given by a normalized form

$$128 \quad x = \frac{V(\Delta\tau)}{V_{max}(\Delta\tau)}, \quad (2b)$$

129 and hence  $x$  ranges from 0 to 1.

130 The exponents  $\alpha$  and  $\xi$  are estimated from the wind speed mean, variance and the maximum  
 131 value based on the general relationships for Beta distributions,

$$132 \quad \alpha = \frac{1}{1+\eta} \left( \frac{\eta}{I} - 1 \right), \quad (3a)$$

$$133 \quad \xi = \eta\alpha, \quad (3b)$$

$$134 \quad \eta = \frac{V_{max}(\Delta\tau) - \bar{V}}{\bar{V}}, \quad (3c)$$

135 where  $V_{max}(\Delta\tau)$  is the maximum time-averaged wind speed in the interval  $\Delta\tau$  or in  
 136 meteorological terminology the maximum  $\Delta\tau$  gust in a large ensemble or a long time series.

137 We obtain  $V_{max}(\Delta\tau)$  from the model proposed by Bartzis et al. (2008) (see Sect. 2.2),  $\bar{V}$  is the  
 138 mean wind speed and  $I$  is the wind speed fluctuation intensity given by

139 
$$I = \frac{\sigma_V^2}{\bar{V}^2} \quad (4a)$$

140 with

141 
$$\sigma_V^2 = \overline{V'^2} \quad , \quad (4b)$$

142 where  $\sigma_V^2$  is the variance and  $V'$  is the fluctuation, which are quantities that are routinely  
 143 available from different types of atmospheric flow models by solving the relevant equations  
 144 (e.g. Hertwig et al., 2012; Koutsourakis et al., 2012) or from experimental measurements.  
 145 Hence, in order to use the model the extreme value  $V_{max}(\Delta\tau)$  needs to be estimated. This is  
 146 described below.

## 147 **2.2 Extreme value analysis**

148 When adopting any finite range p.d.f. for the wind speed  $V(\Delta\tau)$ , the ability to estimate the  
 149 extreme value  $V_{max}(\Delta\tau)$  forms a prerequisite. It should be noted that the experimental  
 150 maximum  $V_{max}^{meas}(\Delta\tau)$  cannot be a priori expected to be the “true” extreme value as a direct  
 151 consequence of statistical uncertainties associated with limited measuring (or simulation)  
 152 times. For this reason we use the theoretical approach proposed by Bartzis et al. (2008) in  
 153 order to approximate the expected (“true”)  $V_{max}(\Delta\tau)$ , which is modelled by

154 
$$V_{max}(\Delta\tau) = \bar{V} \left[ 1 + b \left( \frac{\Delta\tau}{T_V} \right)^{-\nu} I \right] \quad (5)$$

155 where  $T_V$  is the wind speed integral time scale derived from the wind speed autocorrelation  
 156 function  $R_V(\tau)$  via

157 
$$T_V = \int_0^{\infty} R_V(\tau) d\tau \quad (6a)$$

158 and  $R_V(\tau)$  is defined as

159 
$$R_V(\tau) = \frac{\overline{V'(t)V'(t+\tau)}}{\overline{V'^2}} \quad . \quad (6b)$$

160 Eq. 5 was developed initially for the estimation of maximum concentrations of airborne  
 161 pollutants released from point sources. The theoretical background for the development of  
 162 Eq. 5 is based on the application of the following equation for a stationary time series of  
 163 infinite length:

164 
$$\frac{V_{max}(\Delta\tau)}{V_{max}(\Delta T)} = \left( \frac{\Delta\tau}{\Delta T} \right)^{-n} \quad (7)$$

165 Eq. 7 is used for the calculation of the maximum expected wind speed in a time interval  $\Delta\tau$   
166 when the maximum wind speed in a time interval  $\Delta T$  is known. It follows a similar functional  
167 form as proposed by Bartzis et al. (2008) for concentrations. Following the same argument  
168 presented in Bartzis et al. (2008), here referring to wind speed instead of concentration, Eq. 7  
169 is applied in deriving Eq. 5 based on the following assumptions:

- 170 (1) The maximum wind speed  $V_{max}(\Delta T)$  tends to the mean wind speed  $\bar{V}$  as the time  
171 interval  $\Delta T$  increases.
- 172 (2) The time interval by which the wind speed  $V_{max}(\Delta T)$  approximates  $\bar{V}$  is analogous to  
173 the integral time scale of the wind speed  $T_V$  (Eq. 6a).
- 174 (3) When the time interval  $\Delta\tau$  increases, the wind speed  $V_{max}(\Delta\tau)$  approximates zero  
175 according to Eq. 7, while it should tend to  $\bar{V}$ . This accounts for the additional term of  
176 unity in Eq. 5.
- 177 (4) The ratio  $V_{max}(\Delta T)/\bar{V}$  depends on the fluctuation intensity  $I$  (Eq. 4a).

178 Under these considerations the parameters that determine the extreme wind speed at a certain  
179 location are the fluctuation intensity  $I$  and the integral time scale  $T_V$ , together with the  
180 constants  $b$  and  $\nu$ .

181 The relationship presented in Eq. 5 was previously used successfully to predict maximum  
182 time-averaged pollutant concentrations from near-ground emission sources based on  
183 numerical results from simulations performed with computational fluid dynamics (CFD)  
184 models (Efthimiou and Bartzis 2011, 2014; Efthimiou et al. 2011a, 2011b, 2015). The  
185 parameters  $b$  and  $\nu$  in Eq. 5 can be derived empirically and typically exhibit a wide range of  
186 values as demonstrated in previous studies. This is a result of the combination of limitations  
187 of the model, experimental errors, insufficient stationarity of the time series and the finite  
188 duration of the analyzed signal used to derive these values. Previous studies on the dispersion  
189 of airborne material in atmospheric flows suggested indicative values of  $b = 1.5$  and  $\nu = 0.3$ .

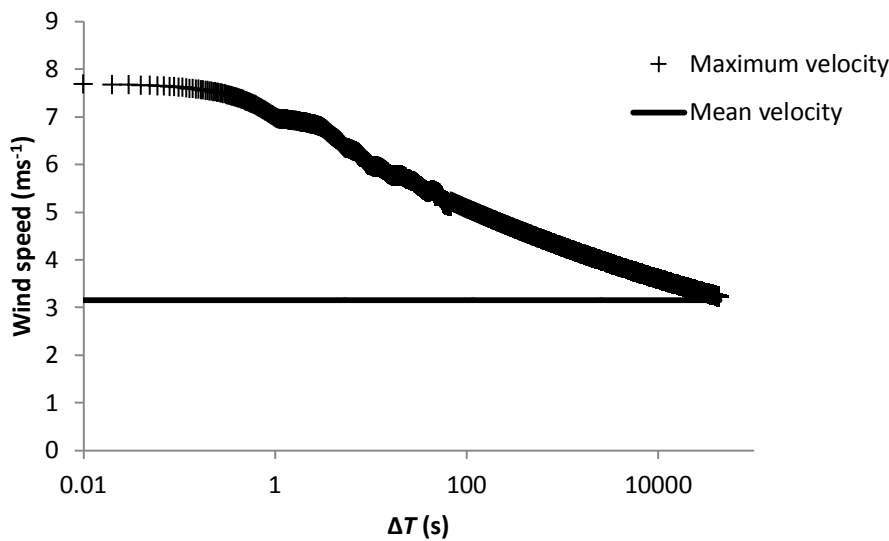
190 Several previous studies of wind gusts in the ASL have demonstrated that local gusts scale  
191 with the standard deviation of wind speed observed at the site (e.g. Beljaars 1987; Kristensen  
192 1991), a parameter which indirectly includes information of surface roughness characteristics  
193 and effects of atmospheric stratification. This dependence is also included in the model  
194 shown in Eq. 5, via the fluctuation intensity  $I$ . Further information about the temporal scale of  
195 the phenomena is added by including a direct link to the local auto-correlation time scale  $T_V$ .  
196 The rationale for adopting the Bartzis et al. (2008) concentration model for the prediction of  
197 wind speed extremes (Eq. 5) is further based on the following:

- 198 1. The wind speed and the concentration are scalars, real numbers and take positive  
199 values.
- 200 2. There is relation between the wind and the concentration which is expressed through  
201 the Schmidt number.
- 202 3. It is expected that both variables have finite extreme values in the ABL.
- 203 4. The assumptions that were mentioned before for the construction of Bartzis et al.  
204 (2008) model are also considered to be valid for the wind speed.

205 The last point is examined by analyzing sample data. For this purpose a random wind speed  
 206 time series was selected from one of the experimental test cases used in this study (BL3-0  
 207 case described in Sect. 3.1.1.), which represents measurements in a rough-wall boundary  
 208 layer modelled in the wind tunnel.

209 The first assumption is: “The maximum wind speed  $V_{max}(\Delta T)$  tends to be equal to  $\bar{V}$  as the  
 210 time interval  $\Delta T$  increases.” The maximum time-averaged wind speed  $V_{max}(\Delta T)$  of the  
 211 random wind speed time series is plotted against  $\Delta T$  in Fig. 1. Also the mean velocity is  
 212 presented. The horizontal axis is plotted in logarithmic scale. It is clear that  $V_{max}(\Delta T)$   
 213 approximates  $\bar{V}$  with the increase of  $\Delta T$ .

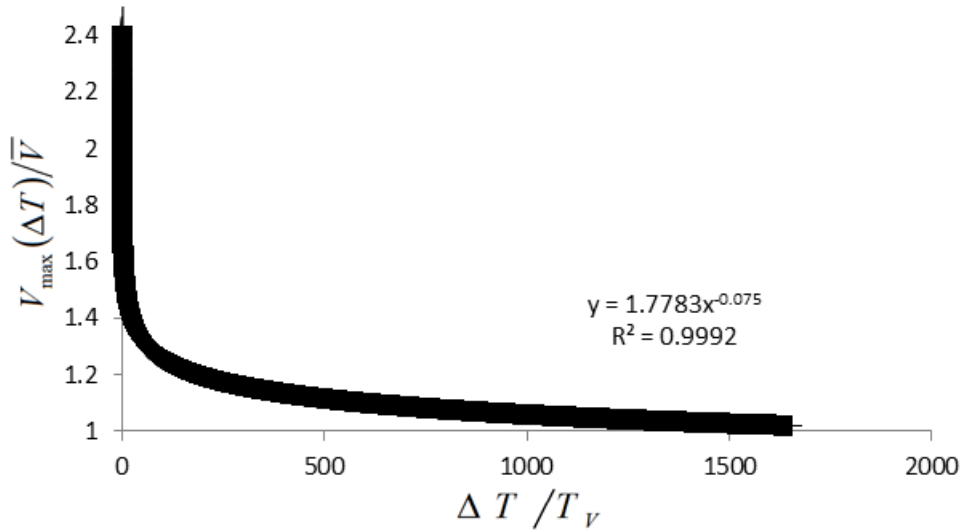
214



215 **Fig 1** The maximum time-averaged wind speed and the mean wind speed are plotted versus the time interval  
 216  $\Delta T$ . The results correspond to a random wind speed time series of the BL3-0 case.  
 217

218 The second assumption is: “The time interval by which the wind speed  $V_{max}(\Delta T)$   
 219 approximates  $\bar{V}$  is analogous to the integral time scale of the wind speed  $T_V$  (Eq. 6a).” The  
 220 ratio  $V_{max}(\Delta T)/\bar{V}$  of the random wind speed time series is plotted against  $\Delta T/T_V$  in Fig. 2.  
 221 There is a clear correlation ( $R^2 = 0.9992$ ) and a power-law function fits the data very well.  
 222

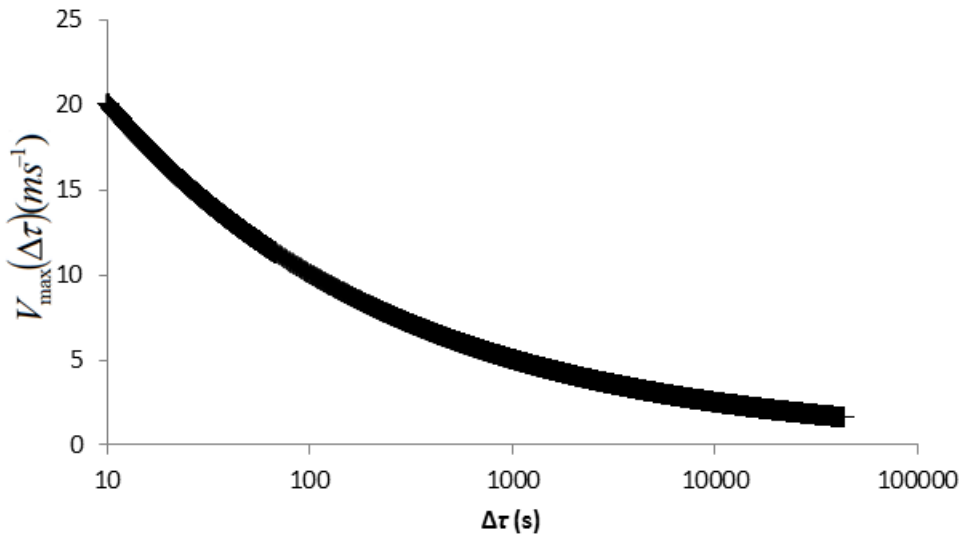




223

224 **Fig 2.** The ratio  $V_{\max}(\Delta T)/\bar{V}$  is plotted versus the ratio  $\Delta T/T_v$  for the same dataset as in Fig. 1.

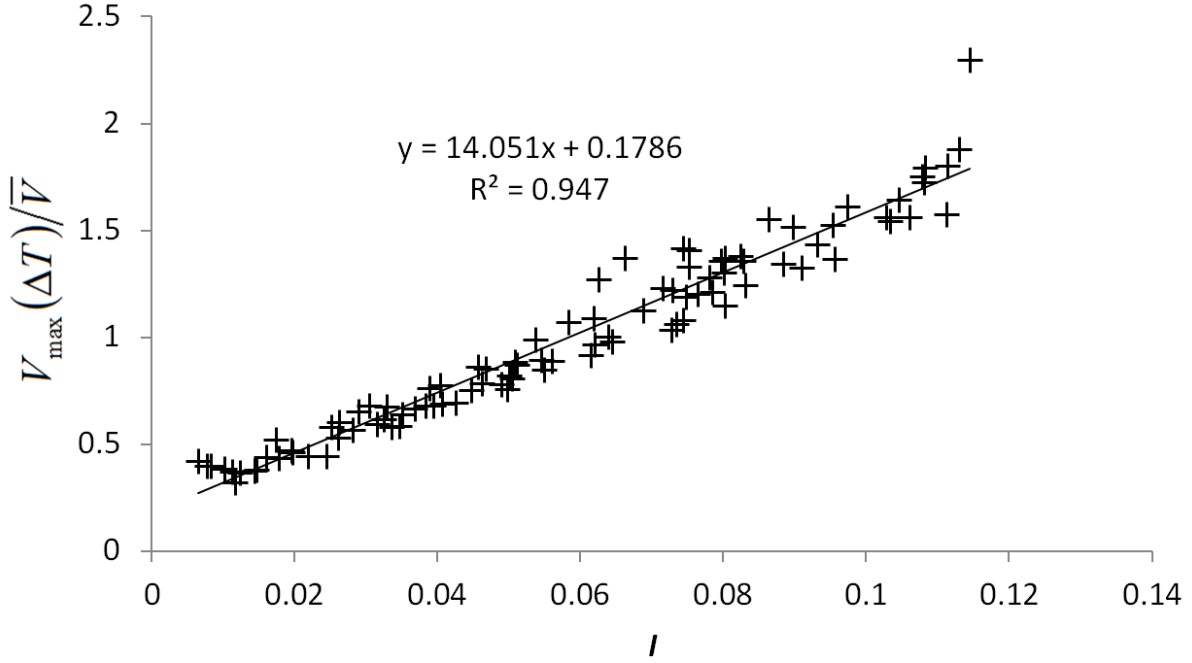
225 The third assumption is: “When the time interval  $\Delta\tau$  increases, the wind speed  $V_{\max}(\Delta\tau)$   
 226 approximates zero, while it should tend to  $\bar{V}$ .” This is illustrated in Fig. 3.



227

228 **Fig 3** The maximum time-averaged wind speed  $V_{\max}(\Delta\tau)$  is plotted versus the time interval  $\Delta\tau$  for the same  
 229 dataset as in Fig. 1.

230 The fourth assumption is: “The ratio  $V_{\max}(\Delta T)/\bar{V}$  depends on the fluctuation intensity  $I$  (Eq.  
 231 4a).” In order to substantiate this assumption based on the data used in this study, the entire  
 232 set of experimental wind speed time series of the wind-tunnel boundary-layer flow case BL3-  
 233 0 were analyzed. The ratio  $V_{\max}(\Delta T)/\bar{V}$  is plotted against the fluctuation intensity  $I$  in Fig. 4,  
 234 clearly illustrating a strong linear relationship ( $R^2 = 0.947$ ).



235

236 **Fig 4** The ratio  $V_{\max}(\Delta T)/\bar{V}$  is plotted versus the fluctuation intensity  $I$  for all horizontal wind speed time  
 237 series available from the rough-wall boundary-layer flow experiment in the wind tunnel.

238 Finally, a further test is performed for the applicability of the method.  $V_{\max}(\Delta\tau)$ ,  $\bar{V}$ ,  $I$  and  $T_V$   
 239 used in Eq. 5 can be calculated from the experimental wind speed time series. In this case,  $\Delta\tau$   
 240 denotes the measurement time interval. The indicative value of the parameter  $\nu$  in Eq. 5 is 0.3  
 241 if the equation is used to estimate peak concentrations. If we assume the same value to be  
 242 valid for wind speed data, Eq. 5 can be rewritten as follows:

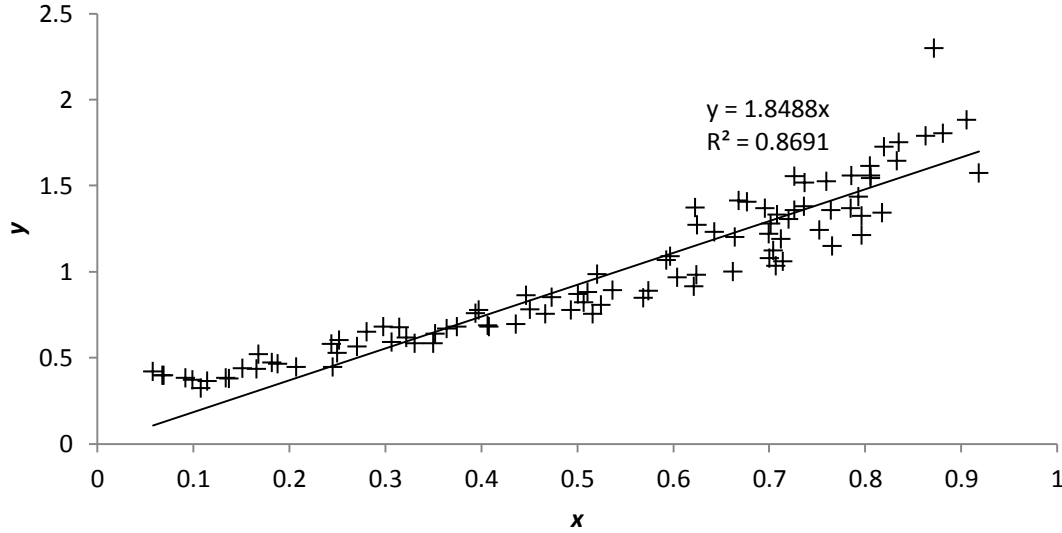
$$243 \quad \frac{V_{\max}(\Delta\tau)}{\bar{V}} - 1 = bI \left( \frac{\Delta\tau}{T_V} \right)^{-0.3} \quad (8)$$

244 The form of Eq. 8 is equivalent to the linear equation  $y = b x$  where:

$$245 \quad y = \frac{V_{\max}(\Delta\tau)}{\bar{V}} - 1 \quad (9a)$$

$$246 \quad x = I \left( \frac{\Delta\tau}{T_V} \right)^{-0.3} \quad (9b)$$

247 If Eq. 5 is valid for wind speed then  $y$  should clearly correlate with  $x$ .  $y$  and  $x$  are plotted in  
 248 Fig. 5. As for Fig. 4, the data points in Fig. 5 are from all the horizontal wind speed time  
 249 series of the laboratory experiment (BL3-0 case). The correlation coefficient  $R^2$  is 0.87,  
 250 which indicates that there is a significant linear relation. Furthermore it is noteworthy that the  
 251 parameter  $b$  takes a value of 1.85 which is close to the indicative value of 1.5 that had earlier  
 252 been determined for concentration data.

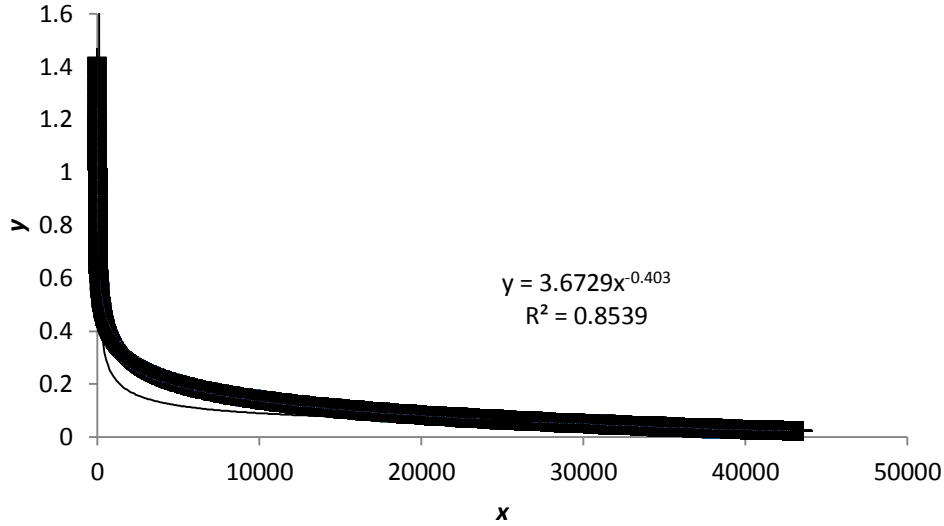


253  
 254 **Fig 5** Examination of the validity of Eq. 5 for wind speed data based on all horizontal wind speed time series  
 255 available from the rough-wall boundary-layer flow experiment in the wind tunnel.

256 However, since we are interested in the p.d.f.'s upper bound we go beyond the indicative  
 257 maximum values measured in the experiments and instead focus on the extreme value  
 258  $V_{max}(\Delta\tau)$  reached within a time interval of infinite duration. It is noted that naturally such an  
 259 extreme value cannot be verified experimentally. When comparing the measured peak value  
 260  $V_{max}^{meas}(\Delta\tau)$  with the expected extreme value  $V_{max}(\Delta\tau)$  for a specific sensor location, in theory  
 261 the relation  $V_{max}^{meas}(\Delta\tau) \leq V_{max}(\Delta\tau)$  will always hold true due to the ultimately finite length of  
 262 the measured signal.

263 It is proposed that the extreme value  $V_{max}(\Delta\tau)$  can be approximated based on Eq. (5) given the  
 264 values of the parameters  $b$  and  $\nu$ . The present strategy is to fix the value of  $\nu = 0.3$  (the  
 265 indicative value from Bartzis et al., 2008) and allow the  $b$  parameter to be estimated from  
 266 suitable calibrating data. This reduces the problem of estimating  $V_{max}(\Delta\tau)$  to estimating a  
 267 single parameter,  $b$ .

268 Bartzis et al. (2008) have presented a method to estimate the parameter  $\nu$ , where  $(V_{max}(\Delta\tau)/\bar{V}) - 1$   
 269 is plotted versus  $\Delta\tau$ .  $V_{max}(\Delta\tau)$  is the experimental maximum time-averaged wind speed  
 270 which changes at every  $\Delta\tau$ . In other words at every  $\Delta\tau$  a new wind speed time series is  
 271 constructed and the maximum wind speed is calculated. For the following test  $\Delta\tau$  ranges from  
 272 0.01 to 45000 s.  $V_{mean}$  is the experimental mean wind speed for  $\Delta\tau$  equal to the measured time  
 273 interval i.e. the original time series that is measured by the instrument. For the following test  
 274  $\Delta\tau = 0.01$  s. For this test we selected a random time series from the BL3-0 wind-tunnel  
 275 experiment. Results are presented in Fig. 6. The equation  $y = \alpha*(x^{-\beta})$  was fitted to the data,  
 276 where  $y$  is  $(V_{max}(\Delta\tau)/\bar{V}) - 1$  and  $x$  is  $\Delta\tau$ . In our case, the parameter  $\beta$  is the desired  $\nu$   
 277 parameter. For this dataset  $\beta$  was found to be equal to 0.403, which is very close to the  
 278 previously determined value of 0.3. Uncertainties arising after fixing the parameter  $\nu$  are  
 279 reflected in the value of  $b$  for each particular location.



280

281 **Fig 6**  $(V_{max}(\Delta\tau)/\bar{V}) - 1$  is plotted versus the time interval  $\Delta\tau$ . The results correspond to a random wind speed  
 282 time series of the BL3-0 case.

283 It is noted that for the parameters  $b$  and  $b^{meas}$  obtained from Eq. (5) using  $V_{max}(\Delta\tau)$  and  
 284  $V_{max}^{meas}(\Delta\tau)$  respectively, in theory  $b^{meas} \leq b$  always holds true since  $V_{max}^{meas}(\Delta\tau) \leq V_{max}(\Delta\tau)$ .

285 With this approach we implicitly hypothesize the existence of a single value of  $b$  that can  
 286 qualify as the upper bound of all  $b^{meas}$  values obtained at any location in the ASL flow at  
 287 which the corresponding value of  $V_{max}^{meas}(\Delta\tau)$  is detected.

288 To close the model, it remains to estimate the value of the parameter  $b$  in Eq. 5.

289 In summary there are two groups of equations.

290 (1) The equations for the construction of the Beta distribution (Eqs. 2a, 2b, 3a, 3b and  
 291 3c).

292 (2) The equation for the estimation of the maximum time-averaged wind speed (Eq. 5).

293 The equations of the first group use the equation of the second group through the parameter  $\eta$   
 294 (Eq. 3c).

### 295 **2.3 Application of the method**

296 The methodology used in this study includes the following steps:

297 1. Estimation of the parameter  $b$  from an analysis of experimental and numerical datasets. In  
 298 our study we use near-surface turbulent flow signals that are available from a dense sensor  
 299 network, offering sufficient coverage of a diverse set of ASL flow scenarios, also including  
 300 data measured within and above urban environments. The wind speed time series should be  
 301 of high time resolution, statistically stationary and have a sufficiently long duration to ensure  
 302 that relevant statistics are derived with high levels of statistical confidence.

303 2. Calculation of the mean wind speed, variance and time scale parameter  $T_V$  from each  
304 measured wind speed time series from the calibrating data.

305 3. Identification of the peak value  $V_{\max}^{meas}(\Delta\tau)$  and estimation of the corresponding local  
306 parameter  $b^{meas}$  from Eq. (5) for each wind speed time series. In this study, the selected time  
307 interval  $\Delta\tau$  is equal to the time resolution of the experimental/simulation data.

308 4. Estimation of the parameters for the wind speed Beta p.d.f. as described by Eqs. (2a, b) and  
309 (3a, b, c) for each location. The values of the required input variables  $\bar{V}$ ,  $\sigma_V^2$  and  $T_V$  are  
310 derived from the reference datasets (see point 2. above). The single missing parameter is the  
311 coefficient  $b$ .

312 5. Derivation of a suitable value for  $b$ .

313 6. Evaluation of the accuracy and robustness of the single value for  $b$  derived from the  
314 calibrating data by testing the model based on independent wind speed measurements from  
315 field experiments.

316 It is emphasized that we look for a p.d.f. that produces a relatively simple but adequate  
317 approximation. Thus, the present methodology is based on the assumption that if a theoretical  
318 c.d.f. can reproduce the real threshold wind speeds for various probabilities then the p.d.f.  
319 used is considered a good approximation.

### 320 **3. Data and Test Cases**

321 For the estimation of the parameter  $b$  and the construction of the statistical model we use data  
322 from wind-tunnel experiments and numerical simulations. The performance of the statistical  
323 model is evaluated in Sect. 5 based on hourly wind speed measurements taken at various field  
324 sites.

#### 325 **3.1. Wind-tunnel experiments**

326 The first two flow datasets analyzed in this study to derive the model coefficient  $b$  stem from  
327 boundary-layer wind-tunnel measurements conducted at the Environmental Wind Tunnel  
328 Laboratory (EWTL) of the University of Hamburg. The datasets are part of the CEDVAL-  
329 LES reference database that offers time-averaged statistics as well as time-resolved data for  
330 different types of boundary-layer flow and dispersion scenarios under neutral stability  
331 conditions. The validation datasets are freely available and described in detail in Fischer et al.  
332 (2010).

333 CEDVAL-LES contains data for various levels of geometric complexity and surface  
334 roughness characteristics. In the database, the term “complexity” refers to the configuration  
335 of the flow scenario covered in the experiment, ranging from complexity 0 denoting simple  
336 rough-wall boundary-layer flows, over flows around isolated obstacles or within obstacle  
337 arrays (complexities 1 or 2, respectively) to flows in semi-idealized urban environments  
338 (complexity 3) or realistic city layouts (complexity 4). Two of the available cases were  
339 selected for this study: **(1)** boundary-layer flow over a very rough surface (CEDVAL-LES

340 complexity 0; case reference: BL3-0) and (2) urban flow within and above a semi-idealized  
341 city geometry (complexity 3; “Michel-Stadt” case reference: BL3-3). In both cases the flow  
342 was physically modelled under a scale of 1:225 and point-wise velocity measurements were  
343 conducted by means of 2D laser Doppler anemometry (LDA). The setup of the very rough  
344 boundary-layer flow (BL3-0) was used as the inflow boundary layer for the semi-idealized  
345 city case (BL3-3).

### 346 **3.1.1 Boundary layer over very rough surface**

347 With a power-law profile exponent of  $\alpha = 0.27$  and a roughness length of  $z_0 = 1.53$  m, the  
348 boundary-layer flow (complexity 0 case) shows roughness characteristics of flow above an  
349 urban environment. In this case, the buildings are not directly represented, but their  
350 aerodynamic effect on the approach flow boundary layer is physically modelled by means of  
351 floor-roughness elements.

352 At sufficient distance from the tunnel inlet and from the floor-roughness elements, the  
353 boundary layer was verified to be horizontally homogeneous. Measurements are only taken  
354 above the blending height, where only the integrated effect of the surface roughness is  
355 represented in the flow characteristics. The 2D-LDA was operated consecutively in two  
356 measuring modes to acquire two components of the velocity vector at a time: the streamwise  
357 ( $U$ ) and spanwise ( $V$ ) velocities ( $UV$ -mode) and the streamwise and vertical ( $W$ ) velocities  
358 ( $UW$ -mode). Data are available in terms of vertical profiles and horizontal transects  
359 perpendicular to the mean inflow direction. All velocity data were scaled to a full-scale  
360 reference height of  $z_{ref} = 100$  m (444.44 mm model scale) with reference wind speeds  $U_{ref}$   
361 ranging between 4.75 and 6 m s<sup>-1</sup>. With a measurement duration of 3 min (corresponding to  
362 11 h full scale) per locations, the derived velocity statistics offered a high level of statistical  
363 representativeness.

364 For the present study, wind speed time series with a full-scale resolution of  $\Delta\tau = 0.01$  s were  
365 analyzed at 96 points for  $UV$  measurements. It should be noted that all time-series were  
366 resampled to the same time resolution.

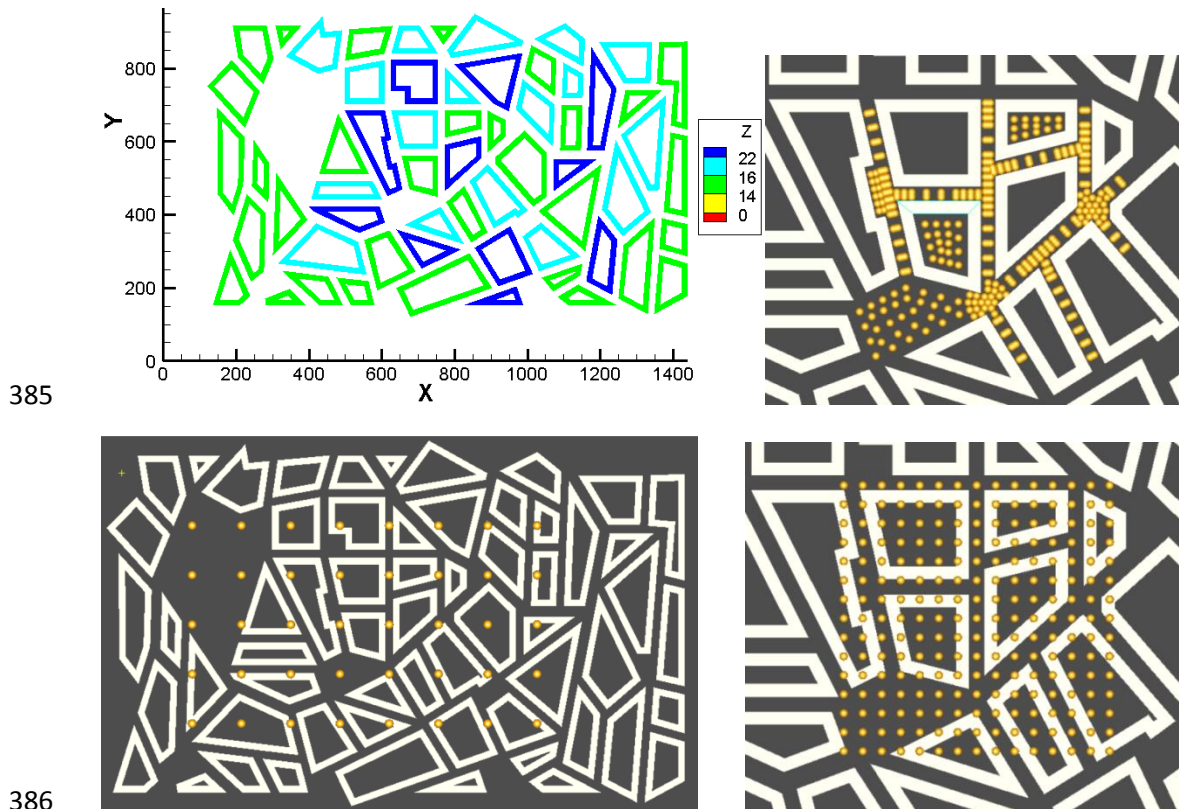
### 367 **3.1.2 Flow in a semi-idealized city**

368 The semi-idealized urban geometry includes typical features of Northern and Central  
369 European cities like courtyards, oblique road arrangements, squares and complex  
370 intersections. Three building heights are included in the model: 15, 18 and 24 m full-scale  
371 (see. Fig. 7 top left). All buildings had flat roofs.

372 Velocity measurements were conducted with the LDA in  $UV$ -mode, providing information  
373 about the horizontal winds within and above the city. The streamwise reference velocity at a  
374 height of 100 m was kept at 6 m/s and was monitored during each measurement run. The  
375 reference velocity was verified to be sufficiently high to guarantee Reynolds number  
376 independence of derived flow statistics within and above the urban canopy. Three groups of  
377 wind speed measurements were used in this study: (1) 40 vertical profiles distributed at  
378 various points throughout the city (Fig. 7, bottom left), (2) detailed measurements on dense

379 horizontal grids at heights of 2, 9 and 18 m full-scale within the main city area (340 m x 340  
380 m), containing 383 measurement points per level (Fig. 7, top right); (3) measurements on a  
381 coarse horizontal grid above the city centre at full-scale heights of 27.5 and 30.2 m,  
382 containing 252 data points at each height (Fig. 7, bottom right).

383 The total number of signals analyzed is 2,158. Again the time series were resampled to a  
384 resolution of 0.01 s.



387 **Fig 7** Top left: the idealized city domain (dimensions given in full scale, dimensions  $X$ ,  $Y$  and  $Z$  are in meters);  
388 top right: densely-spaced measurement locations within the urban canopy layer at elevations of 2 m, 9 m and 18  
389 m; bottom left: vertical profile locations; bottom right: densely-spaced measurement locations above roof top at  
390 heights of 27 m and 30 m.

### 391 3.2 Direct numerical simulations

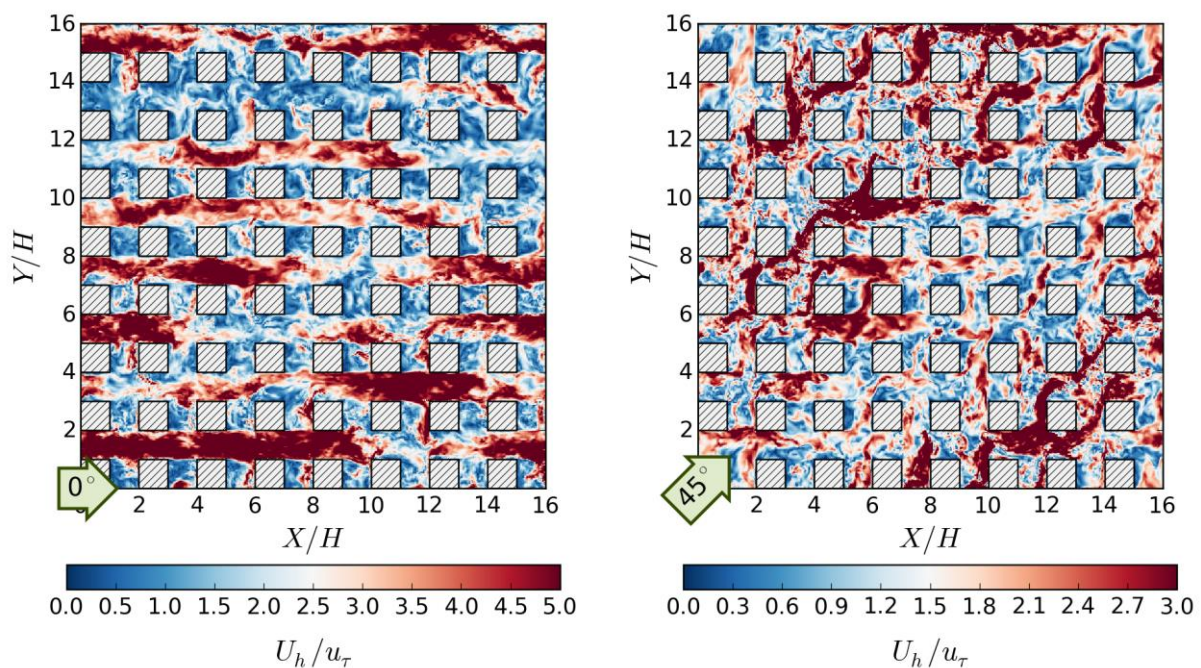
392 The second type of data analyzed in this study stems from simulations of urban flow fields in  
393 an idealized urban roughness generated by direct numerical simulation (DNS). With the DNS  
394 approach, turbulence is directly resolved down to the small dissipative eddy scales. In order  
395 to facilitate this computationally, the flow is simulated at lower Reynolds numbers compared  
396 to the ones typically encountered in the atmospheric boundary layer. The Reynolds numbers  
397 that can be realized, however, are comparable to those typically achieved in boundary-layer  
398 wind-tunnel experiments as those described in the preceding sections. Compared to other  
399 turbulence-resolving CFD approaches like for example large-eddy simulation (LES), the  
400 accuracy of flow simulations with DNS is not affected by errors resulting from turbulence  
401 modelling. Hence, DNS data, after appropriate accuracy checks, can be used as reference data  
402 similar to experimental measurements. The significant computational requirements involved



403 in performing DNS, however, currently restrict the applicability of the technique to simple  
 404 flow scenarios at low Reynolds numbers.

405 The DNS code that had been used to generate the data analyzed in this study is the research  
 406 code CgLES developed at the University of Southampton specifically for performing  
 407 massively-parallel DNS and LES computations. The code is parallelized with the Message  
 408 Passing Interface (MPI) and a flexible multi-block mapping strategy is used to deal with  
 409 complex geometries. The Navier-Stokes equations are discretized using second-order central  
 410 finite differences in space and a second-order Adams-Bashforth scheme in time based on the  
 411 pressure correction method. The Poisson equation for pressure is solved using a multigrid  
 412 method. A detailed description of the numerical techniques involved in the DNS as well as  
 413 examples of previous studies with the code can be found, e.g., in Yao et al. (2001), Branford  
 414 et al. (2011) or Coceal et al. (2006, 2007, 2014).

415 Flow simulations were conducted in a geometry comprised of 64 cubical obstacles of height  
 416  $H$  that were set up in a regular array consisting of 8 rows of 8 obstacles. The computational  
 417 domain was of size  $16H \times 16H \times 8H$ . Periodic boundary conditions were prescribed in  
 418 horizontal directions, free-slip conditions at the upper domain boundary and no-slip and  
 419 impermeability conditions at the bottom of the domain and at solid surfaces. The flow was  
 420 driven by a constant body force that resulted in a roughness Reynolds number of  $Re_\tau = u_\tau H/\nu$   
 421  $= 500$ , where  $u_\tau$  is the total wall friction velocity and  $\nu$  is the kinematic viscosity. Sensitivity  
 422 studies for this setup (Coceal et al. 2006, 2007) demonstrated that the selected resolution of  
 423 the uniform grid of  $H/32$  is sufficient to adequately resolve the flow down to the dissipative  
 424 scales. Data is available for two wind directions ( $0^\circ$  and  $45^\circ$ ) under neutral stratification. Fig.  
 425 8 shows snapshots of the instantaneous horizontal wind speed magnitudes at a height of  $Z =$   
 426  $0.5H$  for both wind directions, illustrating the complexity of the canopy layer flow field.



427



428 **Fig 8** Plan views of the computational domain of the DNS including contours of instantaneous snapshots of the  
 429 non-dimensional horizontal velocity magnitude,  $U_H/u_\tau$ , at half the building height ( $Z = 0.5H$ ). Left:  $0^\circ$  forcing  
 430 direction; right:  $45^\circ$  forcing direction.

431 Wind speed time series at 896 locations overall (for both wind directions) were analyzed in  
 432 this study, each of them having a length of  $140T$ , where  $T$  is the eddy turnover time defined  
 433 as  $T = H/u_\tau$ . As done with the wind-tunnel data, the DNS time series were resampled to a  
 434 non-dimensional time resolution of 0.01. The data locations are distributed in two horizontal  
 435 planes covering the entire simulation domain at two different heights:  $Z = 0.5H$  and  $Z = 1.5H$ .  
 436 Time series were extracted in the centre of each street and intersection. The DNS  
 437 computation were performed on a supercomputer and required a total spin-up time of 12 days  
 438 on 124 nodes for both the  $0^\circ$  and  $45^\circ$  runs, after which the simulations ran for 10 days ( $0^\circ$   
 439 case) and 13 days ( $45^\circ$  case) on 248 nodes to collect flow time series and statistics.

440

#### 441 4. Estimation of the $b$ parameter

##### 442 4.1 Wind-tunnel flow fields

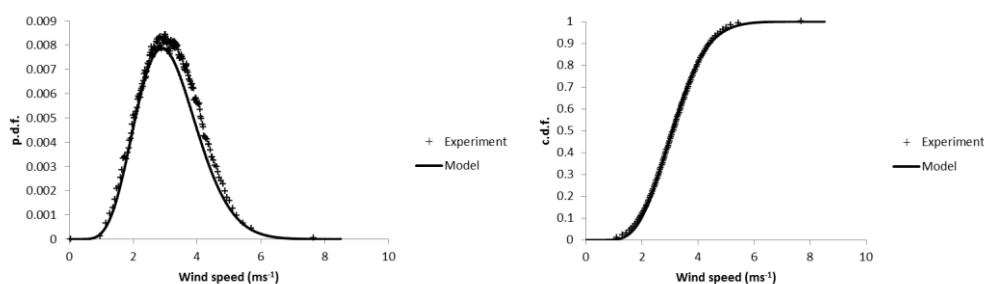
##### 443 4.1.1 Boundary-layer flow

444 The autocorrelation time  $T_V$  is calculated from the autocorrelation function  $R_V(\tau)$  on the  
 445 interval from 1.0 to 0. The wind speed time series are considered to be characterized by a  
 446 sufficiently high temporal resolution.

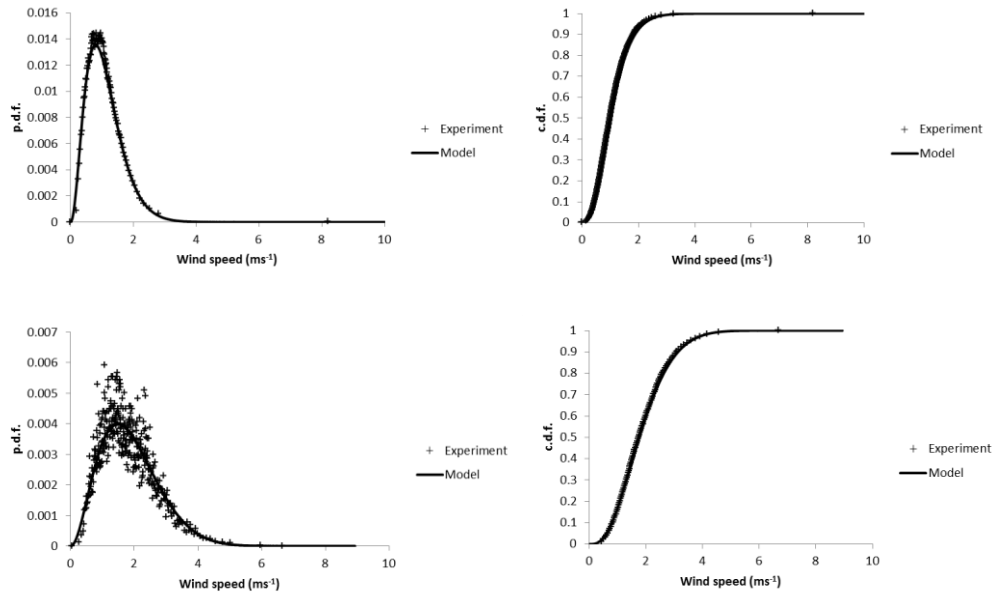
447 At each sensor location, an experimental peak  $V_{\max}^{meas}(\Delta\tau)$  is identified and a  $b^{meas}$  value is  
 448 estimated from Eq. 5 using  $V_{\max}^{meas}(\Delta\tau)$ . The values of  $b^{meas}$  range from 1.5 to 7.2. The value of  
 449 5, however, is only exceeded three times out of 96 (i.e. 3.1%) and seems to point more to  
 450 ‘outliers’ behavior.

451 It is evident that if the proposed model is valid and a single value for  $b$  exists this value has to  
 452 be greater than 5. This dataset and the two other datasets (discussed in detail in the following  
 453 Sects. 4.1.2. and 4.2) clearly indicate that a value of  $b = 6$  is appropriate for these flow  
 454 scenarios.

455 For a randomly selected sensor of the BL3-0 case, the Beta and the experimental p.d.f.s and  
 456 c.d.f.s are plotted in Fig. 9 (first row). The agreement between the model curve and the  
 457 experimental data points is very good.



458



459

460

461 **Fig 9** Experimental and theoretical p.d.f.s and c.d.f.s using the Beta distribution for a randomly selected sensor  
 462 of the BL3-0 case (first row); for a randomly selected sensor of the BL3-3 case (second row) and for a randomly  
 463 selected sensor of the DNS experiment (third row). The parameter  $b$  is set to 6.

#### 464 4.1.2 Semi-idealized city

465 As a next step, the robustness of using a value for  $b$  equal to 6 was examined by analyzing  
 466 flow data acquired in the semi-idealized city wind-tunnel model.

467 As for the boundary-layer flow data, the  $b^{meas}$  is obtained based on the analysis of all 2,158  
 468 signals. The  $b^{meas}$  ranges from 0.63 to 7.1. Again, only at three of the available measurement  
 469 positions a  $b^{meas}$  value larger than 5 was obtained. The proposed value of  $b$  (equal to 6) is also  
 470 relatively close to the maximum value of 7.1.

471 As for the boundary layer flow case analyzed above, the agreement between the Beta and the  
 472 experimental p.d.f.s and c.d.f.s for an example time series from a randomly selected sensor  
 473 location is very high (Fig. 9, second row).

#### 474 4.2 Direct numerical simulations

475 The analysis of the wind-tunnel data clearly supports the proposed methodological approach.  
 476 In a next step, the DNS data of flow in an idealized urban roughness is analyzed in a similar  
 477 manner.

478 As with the experimental data, the DNS dataset is first analyzed with regard to determining  
 479 an appropriate value of the  $b$  parameter. For each sensor, a peak  $V_{\max}^{meas}(\Delta\tau)$  is identified from  
 480 the simulation data and the corresponding  $b^{meas}$  value is estimated from Eq. (5). The  $b^{meas}$   
 481 range is from 1.1 to 5.56. Only 10 values are larger than 5 and the proposed  $b$  value (equal to  
 482 6) is relatively close to the maximum value 5.56.

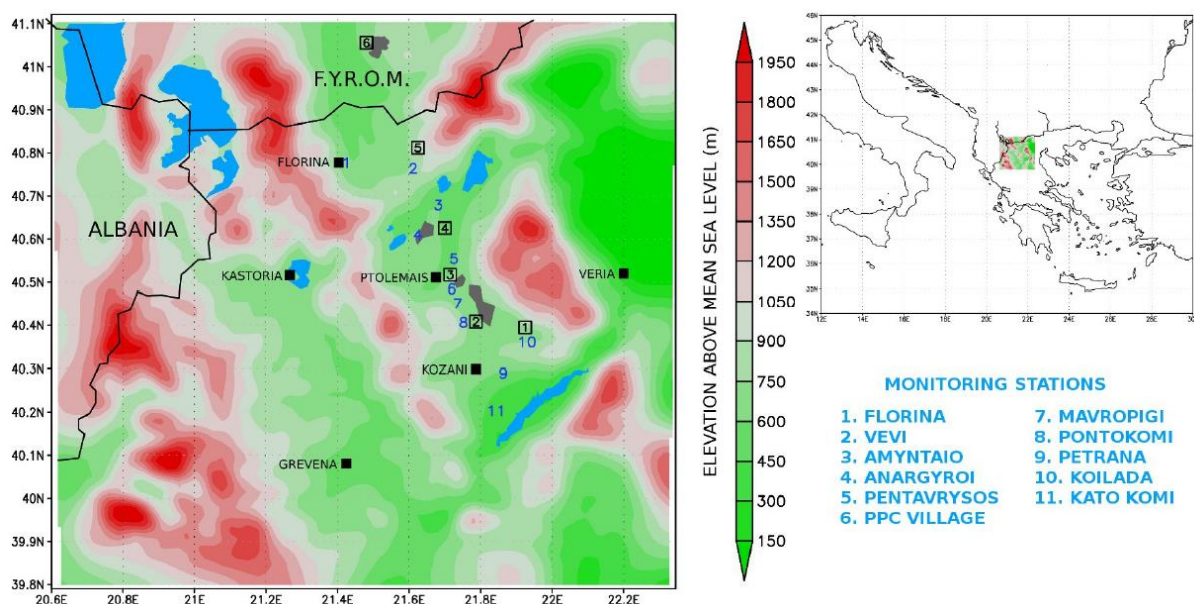
483 The Beta and the experimental p.d.f.s and c.d.f.s from time series at a randomly selected  
 484 point of the DNS flow simulation is presented in Fig. 9 (third row) and again reveal a high  
 485 level of agreements between the statistical model and the reference data.

486 It is worth highlighting that it is rather impressive that two completely different kinds of data  
 487 sources, the wind-tunnel measurements and the DNS, corresponding to very different near-  
 488 surface flow scenarios show such a similar range of values determined for  $b^{meas}$ . This  
 489 provides strong support for the hypothesis that a single representative maximum value of the  
 490  $b$  parameter can be derived, which could then be applied to various types of ASL flows.

### 491 5. Model evaluation based on field data

492 Based on the analysis of the three different calibrating datasets presented above, an upper  
 493 value of the parameter  $b$  equal to 6 was derived. However, the flow scenarios analyzed to  
 494 derive this model parameter represent quite idealized cases of ABL flow (e.g. with respect to  
 495 the isothermal conditions/neutral stratification and stationarity of the flow). In order to  
 496 demonstrate the applicability of the model to real-world ABL flow scenarios, in a next step  
 497 the model is evaluated based on independent field measurements. These consist of hourly-  
 498 averaged in-situ wind-speed measurements from multiple ground-based sensors available  
 499 over the course of several months, which reflect the true variability of the natural atmospheric  
 500 boundary layer in terms of wind-speed trends (e.g. through the propagation of meso-scale  
 501 systems) and stratification effects.

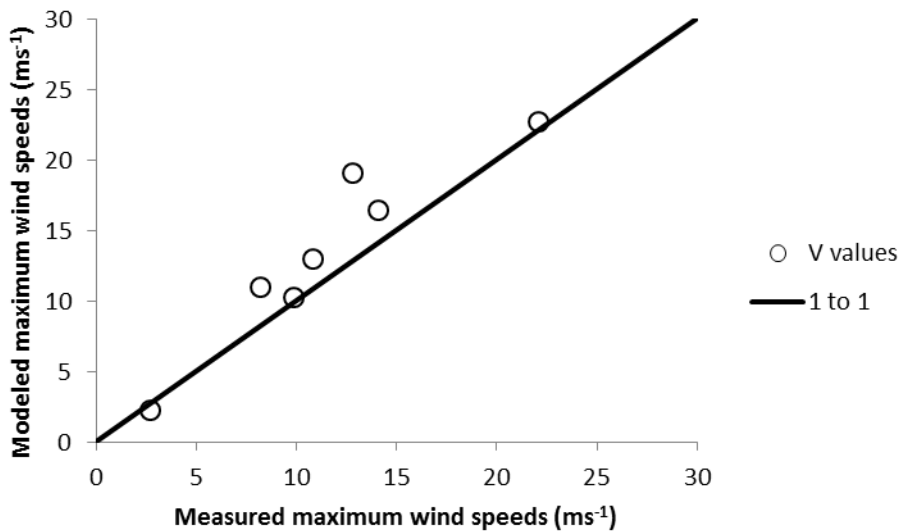
502 For this study, the Greek Public Power Corporation (<https://www.dei.gr/en>) provided hourly  
 503 wind-speed data for the period from 1 January 2012 to 31 August 2012 from 7  
 504 meteorological stations (Vevi, Florina, Koilada, PPC village, Pentabrysos, Petrana and  
 505 Pontokomi) located in the western part of Greece (see Fig. 10). All velocity sensors are  
 506 located in urban areas, the measurement height is 10 m above ground and the averaging  
 507 period of the signals is 1 h.



508

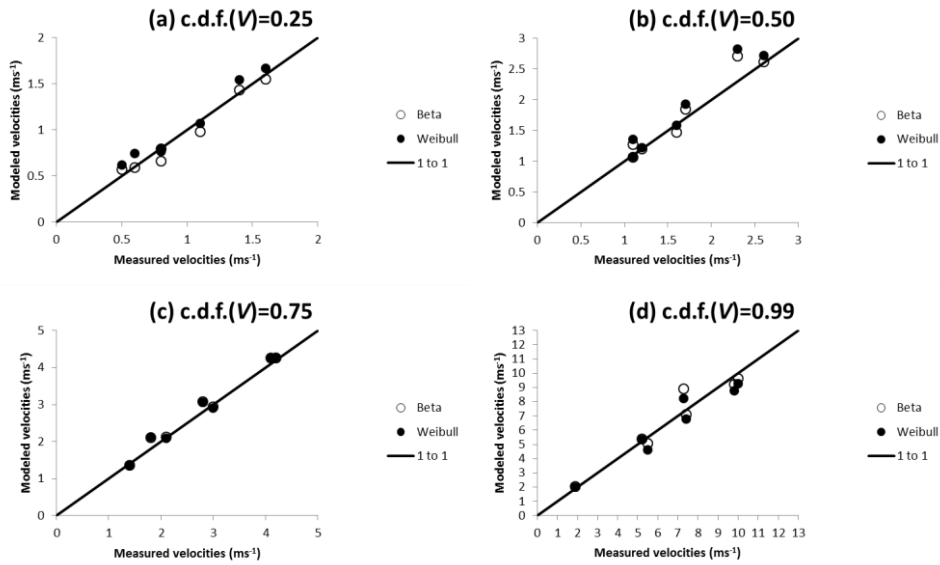
509 **Fig 10** Topography of Western Macedonia, Greece, where the meteorological stations are located. The black  
 510 numbered boxes indicate power plants, gray areas mines. The blue numbers indicate the location of  
 511 meteorological measurement stations, with names listed on the right. Data available for this study are from the  
 512 stations in Vevi, Florina, Koilada, PPC village, Pentabrysos, Petrana and Pontokomi.

513 Initially a comparison is performed in terms of the theoretical value of  $V_{max}(\Delta\tau)$  as derived  
 514 from the Bartzis et al. (2008) model and the measured  $V_{max}^{meas}(\Delta\tau)$  at all stations (see Fig. 11).  
 515 The model provides a success rate of 85.7% (only one value is below the 1:1 line), which  
 516 supports the hypothesis that the proposed theoretical  $V_{max}(\Delta\tau)$  can serve as an upper bound of  
 517 the corresponding measured  $V_{max}^{meas}(\Delta\tau)$ .



518  
 519 **Fig 11** Modelled versus measured peak wind speeds at the 7 field measurement stations. The straight line  
 520 indicated the 1:1 relationship.

521 In the following analysis the performance of the Weibull distribution is also tested. This  
 522 distribution was fitted to the data and its parameters were calculated with the maximum  
 523 likelihood estimation. The results of the 99% threshold (c.d.f.( $V$ )=0.99) are analyzed in order  
 524 to test the performance of the statistical model at the upper tail of the distribution. Based on  
 525 the analysis of the calibrating data discussed in the preceding section, the Beta distribution  
 526 was configured with a  $b$  value of 6. In Fig. 12 scatter plots are presented comparing the  
 527 calculated wind speed  $V(\Delta\tau)$  from the Beta and Weibull distributions and the ones derived  
 528 from the field data corresponding to cumulative probabilities of c.d.f.( $V$ ) = 0.25, 0.50, 0.75  
 529 and 0.99. The Beta distribution performs slightly better than the Weibull distribution in the  
 530 higher wind-speed range for probabilities of 0.25, 0.50 and 0.99.

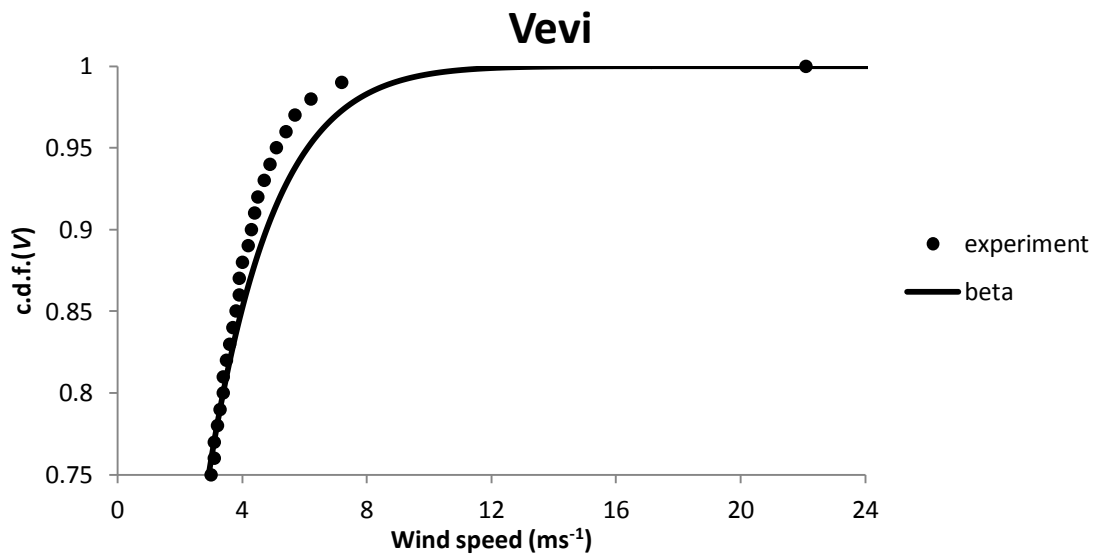


531

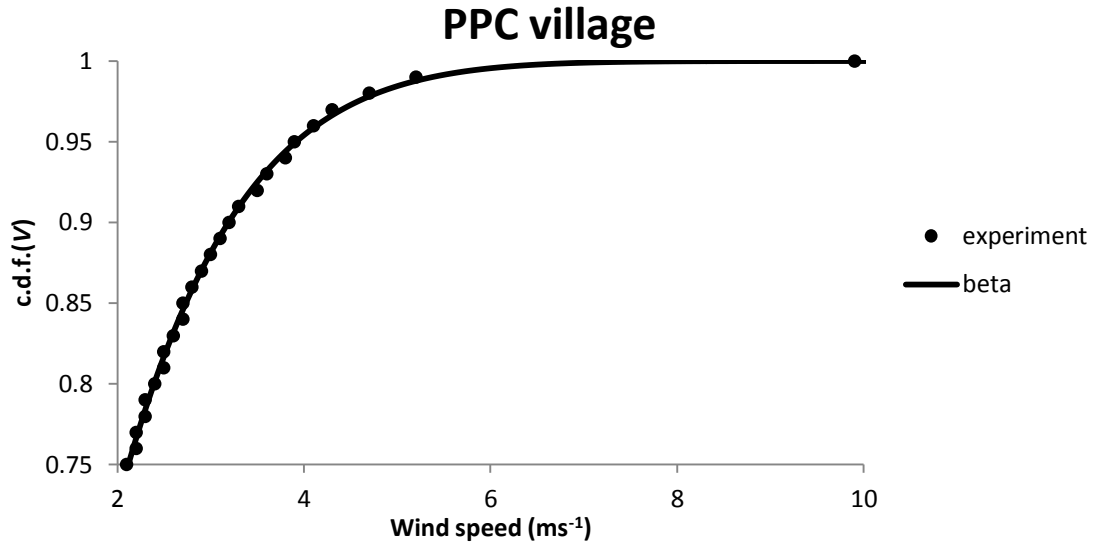
532

533 **Fig 12** Modelled (Beta and Weibull) versus measured wind speeds from the field experiment corresponding to  
 534 (a) c.d.f.( $V$ ) = 0.25, (b) c.d.f.( $V$ ) = 0.50, (c) c.d.f.( $V$ ) = 0.75 and (d) c.d.f.( $V$ ) = 0.99.

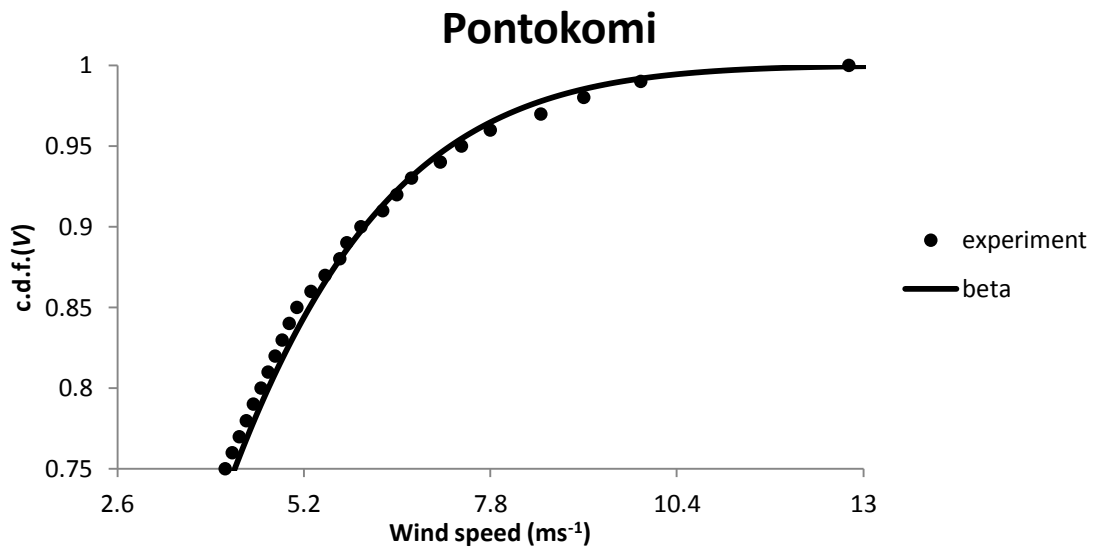
535 The remaining critical question is how individual c.d.f.s derived from the Beta distribution  
 536 compare with the counterparts from the field measurements at each location. In Fig. 13 the  
 537 measured and modelled c.d.f.s for percentiles between 75<sup>th</sup> and 100<sup>th</sup> are shown for three  
 538 stations (Vevi, PPC village and Pontokomi). At these and the other locations not shown here,  
 539 the model presents good agreement with the field data.



540



541



542

543 **Fig 13** Modelled versus measured c.d.f. results for percentiles between the 75<sup>th</sup> and 100<sup>th</sup> for three  
 544 meteorological stations of the Western Macedonia region of Greece.

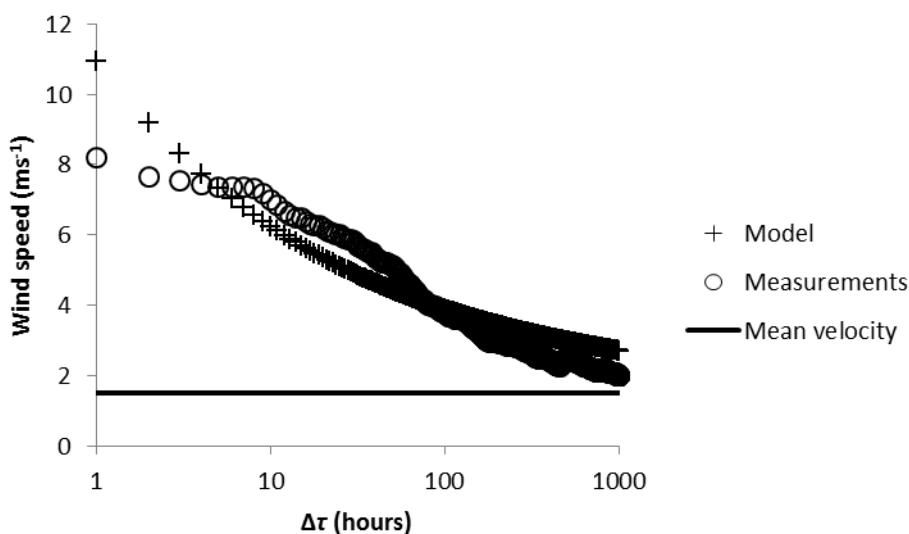
545 Concerning the atmospheric stability it should be noted that the wind tunnel and the DNS  
 546 experiments modelled neutral conditions. However the present field experiment covers all the  
 547 stability conditions and the Beta distribution performs very well under all conditions. Using a  
 548 value of  $b = 6$  when applying the model to independent datasets from field measurements  
 549 shows that the value seems to be a good choice when dealing with ASL wind speeds.

550 **5.1 Effect of the time interval  $\Delta\tau$  on the performance of the model**

551 A further testing of the universal nature of the proposed Beta model is conducted by  
 552 repeating the above analysis for different  $\Delta\tau$ .

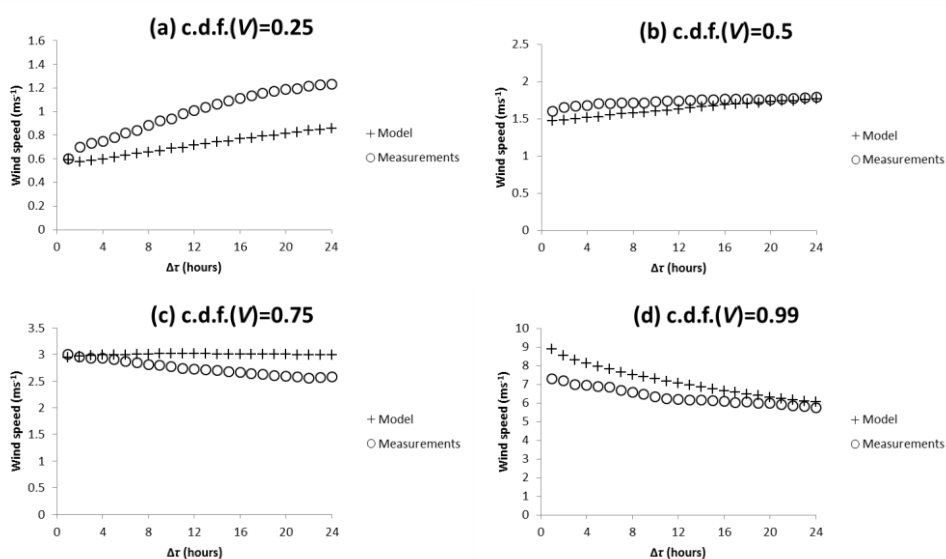
553 In Fig. 14 the peak time-averaged wind speeds based on Eq. 5 and the field experiment are  
 554 plotted for different  $\Delta\tau$  for an example station. The horizontal axis is presented using a  
 555 logarithmic scale. For  $\Delta\tau = 1$  h the model wind speed is higher than the experimental peak

556 wind speed, as expected. For  $\Delta\tau$  between 10 h and 100 h the model slightly underestimates  
 557 the experimental values, while overestimating again after 100 h. For 1000 h the  $V_{max}(\Delta\tau)$  of  
 558 both model and experiment approximate the mean wind speed.



559  
 560 **Fig. 14** Examination of Eq. 5 against field measurements for various  $\Delta\tau$  for data from an example station.

561 In Fig. 15  $V(\Delta\tau)$  obtained from the Beta distribution model and the values derived from one  
 562 of the field stations are presented.  $\Delta\tau$  ranges from 1 h to 24 h. The results correspond to  
 563 cumulative probabilities of c.d.f.( $V$ ) = 0.25, 0.50, 0.75 and 0.99. For c.d.f.( $V$ ) = 0.25 and 0.5  
 564 the wind speed increases with  $\Delta\tau$  while for 0.75 and 0.99 it decreases with  $\Delta\tau$ . This indicates  
 565 that with the increase of  $\Delta\tau$  the distribution becomes more leptokurtic. The model shows the  
 566 same tendency as the experiment. For the specific station the largest deviation of the model  
 567 from the experiment is observed for c.d.f.( $V$ ) = 0.25 and for  $\Delta\tau = 24$  h. On the other hand the  
 568 best performance of the model is observed for c.d.f.( $V$ ) = 0.5 over the entire range of  $\Delta\tau$ .



569  
 570  
 571 **Fig 15** Beta distribution versus measured wind speeds for an example station corresponding to (a) c.d.f.( $V$ ) =  
 572 0.25, (b) c.d.f.( $V$ ) = 0.50, (c) c.d.f.( $V$ ) = 0.75 and (d) c.d.f.( $V$ ) = 0.99.

## 573 **5. Conclusions and Outlook**

574 By using the properties of the Beta p.d.f. in combination with a model for estimating extreme  
575 values based on readily available turbulence statistics (Bartzis et al., 2008), this study  
576 demonstrated that this novel modelling approach can reliably predict the upper margins of  
577 wind speeds encountered in the ASL.

578 The problem itself is quite complex and adequate validation studies require extensive  
579 experimental datasets. Such comprehensive validation efforts exceed the scope of a single  
580 publication, but the work presented here represents a significant first step towards a thorough  
581 testing of the proposed methodology.

582 The selected calibrating data for constructing the proposed model are representative of  
583 different scenarios of turbulent wind flow in the ASL: a rough boundary layer without  
584 buildings, a typical European urban micro-environment and an urban-like arrangement of  
585 cubical buildings. The sensor locations cover the boundary layer and the urban canopy-layer  
586 characteristics. The temporal resolution of the wind-speed signals covers a wide range of  
587 fluctuation intensities.

588 The performance of the model was successfully evaluated based on long-term independent  
589 field measurements (hourly averages), which cover the true variability of ABL flows in terms  
590 of wind-speed trends through the propagation of meso-scale systems and stratification effects.  
591 Concerning the atmospheric stability it should be noticed that the wind tunnel and the DNS  
592 experiments were conducted under neutral conditions. The field experiment, however,  
593 covered a wide range of stability conditions that can be encountered in natural ASLs and the  
594 Beta distribution performed very well under all conditions.

595 From the results obtained the following main conclusions are drawn:

596 1. The approximation of the statistical behavior of the abovementioned wind speed variability  
597 with a Beta distribution p.d.f. was shown to be satisfactory.

598 2. The important issue of the extreme value in the Beta distribution is properly addressed by  
599 an adaptation of the Bartzis et al. (2008) model.

600 3. The present work proposes  $b = 6$  and  $\nu = 0.3$  in Eq. 5.

601 4. With an increase of the averaging time interval the wind-speed distributions of the model  
602 and experiment become more leptokurtic.

603 The new model can broaden the capability of ensemble-averaged computational models such  
604 as Reynolds Averaged Navier Stokes–CFD models to estimate the wind-speed p.d.f. provided  
605 that reliable predictions of mean wind speeds, wind-speed fluctuations and integral time  
606 scales are available from these computations.

607

## 608 **Acknowledgements**



609 The authors would like to thank Bernd Leitl of the Environmental Wind Tunnel Laboratory at  
610 the University of Hamburg for providing access to the reference database (CEDVAL-LES;  
611 <https://www.mi.zmaw.de/index.php?id=6339>). The authors would like to thank the Greek  
612 Public Power Corporation for providing access to the field measurements.

## 613 **References**

- 614 Aidan J (2011) Wind Speed distribution and Performance of some selected wind turbines in Jos, Nigeria. *Latin*  
615 *Am J Phys Educat* 5:457-460
- 616 Bartzis JG, Efthimiou GC (2010) Maximum individual exposure estimation using CFD RANS modelling. 13th  
617 conference on harmonisation within atmospheric dispersion modelling for regulatory purposes, 1-4 June,  
618 2010, Paris, France.
- 619 Bartzis JG, Efthimiou GC, Andronopoulos S (2015) Modelling short term individual exposure from airborne  
620 hazardous releases in urban environments. *J Hazard Mater* 300:182–188
- 621 Bartzis JG, Sfetsos A, Andronopoulos S (2008) On the individual exposure from airborne hazardous releases:  
622 the effect of atmospheric turbulence. *J Hazard Mater* 150:76–82
- 623 Beljaars ACM (1987) The influence of sampling and filtering on measured wind gusts. *J Atmos Ocean Technol*  
624 4:613-626.
- 625 Brabson BB, Palutikof JP (2000) Tests of the Generalized Pareto Distribution for predicting extreme wind  
626 speeds. *J Appl Meteorol* 39:1627-1640
- 627 Branford S, Coceal O, Thomas TG, Belcher SE (2011) Dispersion of a point-source release of a passive scalar  
628 through an urban-like array for different wind directions. *Boundary-Layer Meteorol* 139:367-394
- 629 Brasseur O (2001) Development and application of a physical approach to estimating wind gusts. *Monthly*  
630 *Weather Review* 129:5-25
- 631 Carneiro TC, Melo SP, Carvalho PCM, Braga AP de S (2016) Particle Swarm Optimization method for  
632 estimation of Weibull parameters: A case study for the Brazilian northeast region. *Renew Energy* 86:751-  
633 759
- 634 Carta JA, Ramirez P, Velazquez S (2009) A review of wind speed probability distributions used in wind energy  
635 analysis Case studies in the Canary Islands. *Renew Sust Energy Rev* 13:933–955
- 636 Coceal O, Thomas TG, Castro IP, Belcher SE (2006) Mean flow and turbulence statistics over groups of urban-  
637 like cubical obstacles. *Boundary-Layer Meteorol* 121:491-519
- 638 Coceal O, Dobre A, Thomas TG, Belcher SE (2007) Structure of turbulent flow over regular arrays of cubical  
639 roughness. *J Fluid Mech* 589: 375-409
- 640 Coceal O, Goulart EV, Branford S, Thomas TG, Belcher SE (2014) Flow structure and near-field dispersion in  
641 arrays of building-like obstacles. *J Wind Eng Ind Aerodyn* 125:52–68
- 642 D’Amico G, Petroni F, Pratico F (2014) Wind speed and energy forecasting at different time scales: A  
643 nonparametric approach. *Physica A* 406: 59–66
- 644 Datta D, Datta D (2013) Comparison of Weibull distribution and Exponentiated Weibull distribution based  
645 estimation of mean and variance of wind data. *Int J Energy Informat Communicat* 4(4):1-11
- 646 Efthimiou GC, Bartzis JG (2011) Atmospheric dispersion and individual exposure of hazardous materials. *J*  
647 *Hazard Mater* 188:375–383

- 648 Efthimiou GC, Bartzis JG (2014) Atmospheric dispersion and individual exposure of hazardous materials.  
649 Validation and intercomparison studies. *Int J Environ Pollut* 55:76–85
- 650 Efthimiou G, Bartzis JG, Andronopoulos S, Sfetsos A (2011a) Air dispersion modelling for individual exposure  
651 studies. *Int J Environ Pollut* 302–316.
- 652 Efthimiou GC, Bartzis JG, Koutsourakis N (2011b) Modelling concentration fluctuations and individual  
653 exposure in complex urban environments *J Wind Eng Ind Aerodyn* 99:349–356
- 654 Efthimiou GC, Berbekar E, Harms F, Bartzis JG, Leitl B (2015) Prediction of high concentrations and  
655 concentration distribution of a continuous point source release in a semi-idealized urban canopy using CFD–  
656 RANS modeling. *Atmos Environ* 100:48–56
- 657 Fischer R, Bastigkeit I, Leitl B, Schatzmann M (2010) Generation of spatio-temporally high resolved datasets  
658 for the validation of LES-models simulating flow and dispersion phenomena within the lower atmospheric  
659 boundary layer. *CWE2010*, 23-27 May 2010, Chapel-Hill, NC, USA
- 660 Francisco-Fernández M, Quintela-del-Río A (2013) Nonparametric analysis of high wind speed data. *Clim Dyn*  
661 40:429. doi:10.1007/s00382-011-1263-2
- 662 Gupta AK, Nadarajah S (2004) *Handbook of beta distribution and its applications*, Marcel Dekker, New York.
- 663 He Y, Monahan AH, Jones CG, Dai A, Biner S, Caya D, Winger K (2010) Probability distributions of land  
664 surface wind speeds over North America. *J Geophys Res* 115
- 665 Hertwig D, Efthimiou GC, Bartzis JG, Leitl B (2012) CFD-RANS model validation of turbulent flow in a semi  
666 idealized urban canopy. *J Wind Eng Indust Aerodyn* 111:61–72
- 667 Holmes JD, Moriarty WW (1999) Application of the generalized Pareto distribution to extreme value analysis in  
668 wind engineering. *J Wind Eng Indust Aerodyn* 83:1-10
- 669 Indhumathy D, Seshaiyah CV, Sukkiramathi K (2014) Estimation of Weibull parameters for wind speed  
670 calculation at Kanyakumari in India. *Int J Innovat Res Sci Eng Technol* 3:8340-8345
- 671 Janssen WD, Blocken B, van Hooff T (2014) Computational evaluation of pedestrian wind comfort and wind  
672 safety around a high-rise building in an urban area. *7th Intl. Congress on Env. Modelling and Software*, 15-  
673 19 June 2014, San Diego, CA, USA.
- 674 Karthikeya BR, Negi PS, Srikanth N (2016) Wind resource assessment for urban renewable energy application  
675 in Singapore. *Renew Energy* 87:403-414
- 676 Kidmo DK, Danwe R, Doka SY, Djongyang N (2015) Statistical analysis of wind speed distribution based on  
677 six Weibull methods for wind power evaluation in Garoua, Cameroon. *Rev Energ Renouv* 18:105-125
- 678 Kollu R, Rayapudi SR, Narasimham SVL, Pakkurthi KM (2012) Mixture probability distribution functions to  
679 model wind speed distributions. *Int J Energy Environ Eng* 3:27
- 680 Koutsourakis N, Hertwig D, Efthimiou GC, Venetsanos AG, Bartzis JG, Leitl B (2012) Evaluation of the  
681 ADREA-HF LES code for urban air quality assessment, using the CEDVAL LES wind tunnel database. *8th*  
682 *International Conference on Air Quality – Science and Application*, 19-23 March 2012, Athens, Greece.
- 683 Kristensen L, Casanova M, Courtney MS, Troen I (1991) In search of a gust definition. *Boundary Layer*  
684 *Meteorology* 55:91-107
- 685 Masseran N, Razali AM, Ibrahim K (2013). The probability distribution model of wind speed over east  
686 Malaysia. *Res J Appl Sci Eng Technol* 6:1774-1779

- 687 Men Z, Yee E, Lien FS, Chen DWY (2016) Short-term wind speed and power forecasting using an ensemble of  
688 mixture density neural networks. *Renew Energy* 87:203-211
- 689 Morgan EC, Lackner M, Vogel RM, Baise LG (2011) Probability distributions for offshore wind speeds. *Energy*  
690 *Convers Manag* 52:15–26
- 691 Nemeş CM (2013) Statistical analysis of wind speed profile: a case study from Iasi region, Romania. *Int J*  
692 *Energy Eng* 3:261-268
- 693 Odo FC, Offiah SU, Ugwuoke PE (2012) Weibull distribution-based model for prediction of wind potential in  
694 Enugu, Nigeria. *Adv Appl Sci Res* 3:1202-1208
- 695 Palutikof JP, Brabson BB, Lister DH, Adcock ST (1999) A review of methods to calculate extreme wind speeds.  
696 *Meteorol Appl* 6:119–132
- 697 Petkovic´ D (2015) Adaptive neuro-fuzzy approach for estimation of wind speed distribution. *Elect Power*  
698 *Energy Syst* 73:389–392
- 699 Rozas-Larraondo P, Inza I, Lozano JA (2014) A method for wind speed forecasting in airports based on  
700 nonparametric regression. *Weather Forecast* 29:6:1332-1342
- 701 Sallis PJ, Claster W, Hernandez S (2011) A machine-learning algorithm for wind gust prediction. *Comput*  
702 *Geosci* 37:1337–1344
- 703 Sarkar A, Singh S, Mitra D (2011) Wind climate modeling using Weibull and extreme value distribution. *Int J*  
704 *Eng Sci Technol* 3:100-106
- 705 Simiu E, Heckert NA, Filliben JJ, Johnson SK (2001) Extreme wind load estimates based on the Gumbel  
706 distribution of dynamic pressures: an assessment. *Struct Safety* 23:221-229
- 707 Steinkohl C, Davis RA, Kluppelberg C (2010) Extreme value analysis of multivariate high frequency wind  
708 speed data. Working Paper. University of Munich and Columbia University.
- 709 van Donk SJ, Wagner LE, Skidmore EL, Tatarko J (2005) Comparison of the Weibull model with measured  
710 wind speed distributions for stochastic wind generation. *Trans. ASAE* 48:503–510
- 711 Waewsak J, Chancham C, Landry M, Gagnon Y (2011) An analysis of wind speed distribution at Thasala,  
712 Nakhon Si Thammarat, Thailand. *J Sust Energy Environ* 2:51-55
- 713 Yao YF, Thomas TG, Sandham ND, Williams JJR (2001) Direct numerical simulation of turbulent flow over a  
714 rectangular trailing edge. *Theor Comput Fluid Dyn* 14:337-358
- 715 Zhang J, Wang P, Zheng X (2013) A prediction model for simulating near-surface wind gusts. *Eur Phys J E*  
716 36:51