

# 1 AUGMENTING SIMULATIONS OF AIRFLOW AROUND BUILDINGS 2 USING FIELD MEASUREMENTS

3 Didier G. Vernay <sup>a,c,\*</sup>, Benny Raphael <sup>b</sup> and Ian F.C. Smith <sup>a,c</sup>

4 a) Future Cities Laboratory, ETH Zurich, Zurich, Switzerland.

5 b) Civil Engineering Department, Indian Institute of Technology, Madras, India.

6 c) Applied Computing and Mechanics Laboratory, School of Architecture, Civil and Environmental Engineering (ENAC),  
7 EPFL, Lausanne, Switzerland.

8 \* Corresponding author: Didier G. Vernay, Future Cities Laboratory, Singapore-ETH Centre 1 CREATE Way #06-01  
9 CREATE Tower Singapore 138602. Tel.: +65 9723 1127, E-mail: didier.vernay@epfl.ch

10 **Abstract** – Computational fluid-dynamics (CFD) simulations have become an important tool for the  
11 assessment of airflow in urban areas. However, large discrepancies may appear when simulated  
12 predictions are compared with field measurements because of the complexity of airflow behaviour  
13 around buildings and difficulties in defining correct sets of parameter values, including those for inlet  
14 conditions. Inlet conditions of the CFD model are difficult to estimate and often the values employed  
15 do not represent real conditions. In this paper, a model-based data-interpretation framework is proposed  
16 in order to integrate knowledge obtained through CFD simulations with those obtained from field  
17 measurements carried out in the urban canopy layer (UCL). In this framework, probability-based inlet  
18 conditions of the CFD simulation are identified with measurements taken in the UCL. The framework  
19 is built on the error-domain model falsification approach that has been developed for the identification  
20 of other complex systems. System identification of physics-based models is a challenging task because  
21 of the presence of errors in models as well as measurements. This paper presents a methodology to  
22 estimate modelling errors. Furthermore, error-domain model falsification has been adapted for the  
23 application of airflow modelling around buildings in order to accommodate the time variability of  
24 atmospheric conditions. As a case study, the framework is tested and validated for the predictions of  
25 airflow around an experimental facility of the Future Cities Laboratory, called “BubbleZERO”. Results  
26 show that the framework is capable of narrowing down parameter-value sets from over five hundred to  
27 a few having possible inlet conditions for the selected case-study. Thus the case-study illustrates an  
28 approach to identifying time-varying inlet conditions and predicting wind characteristics at locations  
29 where there are no sensors.

30 *Keywords:* Computational Fluid Dynamics (CFD); airflow modelling; field measurements; system  
31 identification; multi-model reasoning

## 32 1) INTRODUCTION

33 Urban populations are growing and therefore, understanding urban climate behaviour has become an  
34 increasingly important research field. Urban climate has an impact on the comfort and health of  
35 residents. The energy consumption of buildings is influenced by the convective heat flux at the building  
36 façade and therefore by the wind pattern around buildings (Defraeye et al., 2011). Furthermore, energy  
37 demand of buildings can also be reduced by harnessing airflow for natural ventilation (Ghiaus and  
38 Allard, 2005). Therefore, cities and buildings should be planned and designed according to  
39 characteristics of their climatic environments.

40 Computational fluid-dynamics (CFD) simulations have increasingly been used to simulate the airflow  
41 in urban areas (Al-Sallal and Al-Rais, 2011; Van Hooff and Blocken, 2010). CFD simulations  
42 numerically solve the fluid-flow equations of motion. In general, the equations are time dependent; this

43 means that flow variables at each point have to be computed at several points in time. Directly solving  
44 the equations using numerical methods take much computer memory and time and therefore,  
45 simplifications are made to mathematical models. In steady Reynolds-averaged Navier-Stokes (RANS)-  
46 based models, the fluid-flow equations of motion are averaged over time. This results in steady-state  
47 equations that are easier to solve. Standard k-epsilon and revised k-epsilon models are examples of  
48 RANS-based models. In these models, two additional transport equations are employed to model  
49 turbulent properties of the flow. For example, in the standard k-epsilon model, the new transported  
50 variables are the turbulent kinetic energy ( $k$ ) and turbulent dissipation rate ( $\epsilon$ ). Another popular  
51 approach for modelling turbulence is Large Eddy Simulation (LES) in which time-dependent variations  
52 of flow quantities are computed. LES solves large eddies of flow and model the small eddies with a  
53 subgrid-scale model. While, this is computationally more efficient than direct numerical solution, it  
54 takes significantly greater computation time than RANS-based models.

55 CFD simulations provide high-spatial resolution of data and allow efficient parametric studies for  
56 evaluating design configurations (Van Hooff and Blocken, 2010). Due to the increased use of CFD  
57 simulations for urban airflow modelling, several sets of best practice guidelines (BPG) have been  
58 established (Franke et al., 2011; Tominaga et al., 2008b). However, the accuracy of CFD simulations  
59 remains a major concern due to uncertainties associated with i) modelling complex phenomena present  
60 in urban environments (Allegrini et al., 2013; Mochida and Lun, 2008; Murakami, 2006), ii) the  
61 representation of complex geometrical structures of urban sites and iii) numerical challenges at wall  
62 boundaries (Blocken et al., 2007). Even a sophisticated model may not be accurate because of  
63 uncertainties in input parameter values such as inlet conditions. For a steady-state model, other  
64 significant sources of uncertainties are associated with the time variability of atmospheric conditions  
65 (Schatzmann and Leitl, 2011). CFD simulations are thus often validated only with wind-tunnel  
66 experiments under controlled experimental conditions (e.g. Ramponi and Blocken, 2012). However,  
67 time variability of atmospheric conditions cannot be captured in a conventional wind tunnel in which  
68 fixed inlet conditions are employed. In urban environments, many combinations of values of variables  
69 such as wind speed, wind direction and turbulent kinetic energy may occur upstream of the area of  
70 interest over a short period. The time variability of atmospheric conditions may be estimated using  
71 Large Eddy Simulation (LES) with time-dependent inlet conditions. Jiang and Chen found better results  
72 using LES with varying wind directions at the inlet rather than using LES with fixed wind directions  
73 (Jiang and Chen, 2002). However, LES was only performed during a real-time period of 10-20 min  
74 with specific inlet conditions.

75 Challenges appear if field measurements are used to validate steady-state models (Schatzmann and  
76 Leitl, 2011). The averaging period of measurement data should be short enough in order to capture the  
77 variations at the inlet of the computational domain. However, if a short averaging period is employed,  
78 errors between predicted and measured values are expected because of the stochastic time variations of  
79 flow characteristics due to low frequency turbulence (Schatzmann and Leitl, 2011).

80 Alternatively, field measurements are employed in order to quantify the airflow in urban areas.  
81 Measurement values in the Urban Canopy Layer (UCL) depend on the location of the sensor because  
82 of high spatial variability of airflow. Significant variations might be observed if measurements are  
83 carried out even a few meters apart from each other (Schatzmann and Leitl, 2011). Therefore, field  
84 measurements cannot provide the whole image of the airflow pattern as CFD simulations can do.

85 Model-based data-interpretation strategies have the potential to improve the accuracy of CFD  
86 simulations using knowledge obtained from field measurements. In this paper, use of these approaches  
87 involves generating sets of CFD simulations through assigning values of parameters that are not known  
88 precisely to a model class (template model).

89 Many approaches exist for the identification of physics-based models using measurements.  
90 Deterministic approaches, in which an optimum model is found by minimizing the differences between  
91 model predictions and measurements, may not be appropriate to identify parameter values within model  
92 classes because many combinations of parameter values may be consistent with the measurement data.  
93 Such ambiguities are amplified by uncertainties in models and measurements, particularly when there  
94 are several sources of systematic bias. Modelling assumptions often introduce bias.

95 Moreover, deterministic approaches do not usually provide information on the uncertainties of  
96 predictions. Probabilistic approaches, in which many plausible models are found, are more appropriate  
97 for identification of environmental models. Their performance depends on the knowledge of the error  
98 structure (Beven, 2008). The error structure refers to the probability density function of both  
99 measurement and modelling error as well as their spatial correlations. Modelling error originates from  
100 uncertainty associated with the model class.

101 Model-falsification approaches such as Generalized Likelihood Uncertainty Estimation (GLUE)  
102 (Beven, 2008; Beven, 2006) and the error-domain model falsification approach (Goulet et al., 2013;  
103 Goulet et al., 2012) may be used to identify parameter values of CFD models where knowledge of  
104 modelling error is not known precisely. Error-domain model falsification involves falsification of model  
105 instances for which the difference between measured and predicted values is, for any measurement  
106 location, larger than an estimate of maximal values of error defined by combining measurement and  
107 modelling uncertainties at that location. The term model instance refers to an instantiation of a model  
108 class with specific values of parameters. Error-domain model falsification has not yet been applied to  
109 time varying contexts such as airflow around buildings.

110 The first objective of this paper is to present a model-based data interpretation framework that is  
111 appropriate for CFD simulations in urban areas. More specifically, a systematic approach for  
112 determining ranges of possible inlet conditions is described along with a methodology for estimating  
113 modelling errors. In this framework, error-domain model falsification is employed and adapted in order  
114 to deal with the time variability of airflow. Modelling errors are estimated by comparing responses of  
115 simulations using RANS-based model with those of LES. Responses of LES are not perfectly correct.  
116 Therefore, only lower bound estimates of modelling errors are determined in this paper. The second  
117 objective is to illustrate this framework using a case study of the “BubbleZERO” facility.  
118 “BubbleZERO” is an experimental facility of the Future Cities Laboratory, Singapore-ETH Centre for  
119 Global Environmental Sustainability located at the National University of Singapore.

120 The structure of the paper is as follows. In the next section, an overview of the accuracy of CFD  
121 simulation for the predictions of wind characteristics in urban areas is provided. A model-based data  
122 interpretation framework is proposed in Section 3. Section 4 introduces the case study as well as the set  
123 of numerical simulations used in the model-based data interpretation framework. In Section 5, the  
124 sensor setup for the case study is described. Section 6 presents a methodology to evaluate modelling  
125 errors. Section 7 completes the case study by applying the model-based data interpretation framework  
126 to measurement and simulation datasets. The paper ends with a discussion of results, limitations and  
127 plans for future work.

## 128 **2) SOURCES OF UNCERTAINTIES IN NUMERICAL SIMULATIONS**

129 Several studies have evaluated the performance of RANS-based models for prediction of mean flow  
130 quantities around buildings. These studies usually compare responses of numerical simulations with  
131 wind tunnel experiments. The advantage of wind tunnel experiments for validation of CFD simulations  
132 is that values of simulation parameters and boundary conditions are well known. Hence, if numerical  
133 errors are negligible, conclusions related to the performance of the turbulence model can be made. An

134 extensive validation study was carried out by Yoshie et al. for the development of the Architectural  
135 Institute of Japan (AIJ) guidelines (Yoshie et al., 2007). Comparison between simulations using RANS-  
136 based models and wind tunnel experiments have been performed for four building geometries, 1)  
137 Airflow around a single idealized building with ratio length:width:height=1:1:2, 2) Airflow around an  
138 idealized high-rise building with ratio length:width:height=1:1:4 surrounded by low-rise buildings, 3)  
139 Airflow in the urban area of Niigata in which CAD data were used, 4) Airflow in the urban area of  
140 Shinjuku in which CAD data were used.

141 In the study of the airflow around a single idealized building with ratio length:width:height=1:1:2,  
142 responses of CFD simulations using the standard k-epsilon model or revised k-epsilon models have  
143 been compared with wind tunnel measurements carried out by Meng and Hibi (1998). It was found that  
144 predicted values of amplification factor of wind speeds,  $U/U_0$ , are in good agreement ( $\pm 10\%$ ) with  
145 measured values in regions of high wind speeds ( $U/U_0 > 1$ ). The amplification factor of wind speeds is  
146 defined as the ratio between the local wind speed,  $U$ , to the wind speed,  $U_0$ , that would occur without  
147 the presence of buildings. However, in low wind speeds regions ( $U/U_0 < 1$ ), CFD simulations  
148 underestimate the amplification factors of wind speeds by a factor 5 or more (Blocken et al., 2011).

149 Revised k-epsilon models provide slightly more accurate responses in regions of high wind speeds but  
150 less accurate responses in regions of low wind speeds. The reverse flow on the roof is not reproduced  
151 with the standard k-epsilon model while the reverse flow is slightly overestimated with revised k-  
152 epsilon models (Yoshie et al., 2007). Overestimation of the region of reverse flow in the wake of the  
153 building is observed for all RANS-based models.

154 The same measurement data were used by Tominaga et al. (2008a). In this study, measurements were  
155 compared with responses of simulations using RANS-based models (standard k-epsilon model and  
156 revised k-epsilon models) as well as responses of LES with and without inflow turbulence. The  
157 estimation of the size of the reverse-flow region in the wake of the building is improved if LES is  
158 employed (Tominaga et al., 2008a). Responses of LES with inflow turbulence were found to be in good  
159 agreement with wind tunnel experiments when predicting mean velocity and turbulent kinetic energy  
160 in the wake of the building.

161 Yoshie et al. had similar conclusions for the three other building geometries (Yoshie et al., 2007): in  
162 regions of high wind speeds, predicted amplification factors of wind speeds are fairly accurate ( $\pm 10$ -  
163 20%). However, large errors in the predictions of amplification factors are observed in regions of low  
164 wind speeds (factor 4-5 or more).

165 Blocken and Carmeliet compared responses of simulations using the Realizable k-epsilon model (Shih  
166 et al., 1995) (RANS-based model) with sand-erosion wind-tunnel experiments carried out by Beranek  
167 (1979) for three configurations of parallel buildings (Blocken and Carmeliet, 2008). A grid-sensitivity  
168 analysis has been performed and high-order discretisation schemes were employed in order to reduce  
169 numerical errors. The results confirmed that performance of the RANS-based model in the predictions  
170 of amplification factors is higher in regions of high wind speeds ( $\pm 10\%$  accuracy). Simulations using  
171 the RANS-based model significantly underestimate the amplification factor in regions of low wind  
172 speeds (factor 4 or more).

173 Incorrect definition of parameter values/inlet conditions may also lead to uncertain predictions of wind  
174 characteristics. Inlet conditions of CFD simulations are difficult to estimate. Even if a measurement  
175 station is employed to measure the inlet boundary conditions, the measurements may not be  
176 representative of the overall conditions at the inlet because of the high spatial variability of wind  
177 characteristics in the UCL (Schatzmann and Leitl, 2011). Even above the buildings, spatial variability

178 is observed because the airflow is constantly adapting to the change of surface characteristics  
179 (Schatzmann and Leitl, 2011).

180 Buildings are modelled with a certain degree of geometrical simplification in CFD simulations. An  
181 equivalent roughness is imposed at building surfaces. The equivalent surface roughness of the  
182 surrounding buildings is difficult to estimate and may have a strong influence on predictions of wind  
183 speeds (Blocken and Persoon, 2009). In a previous study, field measurements have been used in order  
184 to estimate the value of the roughness of the surrounding buildings that leads to the best fit (Blocken  
185 and Persoon, 2009). However, as mentioned previously, while calibration approaches force matches of  
186 model predictions with measurement data at locations of sensors, they may not be accurate at other  
187 locations.

### 188 3) METHODOLOGY

#### 189 3.1) Description of the system variables

190 This Section presents the model-based data-interpretation framework used to identify plausible  
191 parameter values of CFD simulations through measurements carried out in the urban canopy layer. The  
192 framework is an adaptation of the error-domain model falsification approach developed by Goulet  
193 (Goulet et al., 2013; Goulet et al., 2012).

194 The first step is to define the parameters that need to be identified  $\theta = [\theta_1, \dots, \theta_{max}]$  with measurements  
195  $y$  and their prior ranges of values. In this study, boundary conditions such as the inlet wind speed and  
196 direction are treated as parameters of the model. Then, a population of model instances  $M(\theta_j)$  is  
197 generated through assigning a set of parameter values  $\theta_j$  to a model class  $M$  with  $j \in \{1, \dots, n_m\}$ , where  
198  $n_m$  corresponds to the number of model instances generated.

199 When the right set of parameter values  $\theta^* = [\theta_1^*, \dots, \theta_{max}^*]$  is assigned to the model class  $M$ , the  
200 predictions of the model instance  $M(\theta^*)$  correspond to the real quantity  $Q$  plus a modelling error  $\epsilon_{model}$ .  
201 Modelling errors are errors associated with the model class  $M$  and they cannot be accounted for through  
202 varying values of parameters  $\theta$ . Furthermore, measurement  $y$  is equal to the real quantity  $Q$  plus a  
203 measurement error  $\epsilon_{measure}$ . This relation is expressed in Eq. (1).

$$204 \quad M(\theta^*) + \epsilon_{model} = Q = y + \epsilon_{measure} \quad (1)$$

205 Therefore, by rearranging the terms of Eq. (1), the difference between a model prediction and  
206 measurement is equal to the difference between measurement and modelling errors. This relation is  
207 expressed in Eq. (2).

$$208 \quad M(\theta^*) - y = \epsilon_{measure} - \epsilon_{model} \quad (2)$$

209 Modelling error and measurement error are seldom known. Only bounds of plausible values of errors  
210 (imprecise probability) can be estimated.

211 Model instances are considered to be candidate models if the residual of the difference between  
212 measured and predicted values fall inside the intervals  $[T_{min}, T_{max}]$ . Bounds of these intervals are  
213 obtained by computing the difference between measurement and modelling errors using elementary  
214 rules of interval arithmetic (Moore, 1966). This is expressed in Eq. (3).

$$215 \quad \epsilon_{measure} - \epsilon_{model} = [\epsilon_{measure,min}, \epsilon_{measure,max}] - [\epsilon_{model,min}, \epsilon_{model,max}] \\ 216 \quad = [\epsilon_{measure,min} - \epsilon_{model,max}, \epsilon_{measure,max} - \epsilon_{model,min}]$$

$$217 \quad = [T_{min}, T_{max}] \quad (3)$$

### 218 3.2) Identification of parameter values of CFD simulations

219 In this paper, the model parameters  $\theta$  that need to be identified with field measurements are the wind  
 220 speed and the wind direction at the inlet of the CFD simulation. The quantities compared are measured  
 221 and predicted horizontal wind speeds  $v_h$  as well as wind directions  $\vartheta$  at sensor locations. The range of  
 222 modelling error is estimated at each sensor location  $i \in \{1, \dots, n_s\}$ , for each model instance  $j \in$   
 223  $\{1, \dots, n_m\}$  and for each compared quantity  $k \in \{v_h, \vartheta\}$ , where  $n_m$  corresponds to the number of model  
 224 instances and  $n_s$  corresponds to the number of sensors. At each sensor location, measured values of  
 225 horizontal wind speeds  $v_h$  and wind directions  $\vartheta$  are compared with model predictions. Model instances  
 226  $M(\theta_j)$  are candidates if, for each and every sensor location  $i \in \{1, \dots, n_s\}$  and for each and every  
 227 compared quantities  $k \in \{v_h, \vartheta\}$ , the difference between measurement and model predictions falls  
 228 inside the intervals  $[T_{min,i,j,k}, T_{max,i,j,k}]$ . This corresponds to the situation where Eq. (4) is satisfied.

$$229 \quad \forall i \in \{1, \dots, n_s\} \text{ and } \forall k \in \{v_h, \vartheta\} : T_{min,i,j,k} \leq M(\theta_j)_{i,k} - y_{i,k} \leq T_{max,i,j,k} \quad (4)$$

230 Where  $M(\theta_j)_{i,k}$  is the predicted value of quantity  $k \in \{v_h, \vartheta\}$  at sensor location  $i \in \{1, \dots, n_s\}$  by the  
 231 model instance  $j \in \{1, \dots, n_m\}$  and  $y_{i,k}$  is the measured value of quantity  $k$  at sensor location  $i$ .

232 Model instances that do not satisfy Eq. (4) are falsified.

### 233 3.3) Predictions of flow quantities where there are no sensors

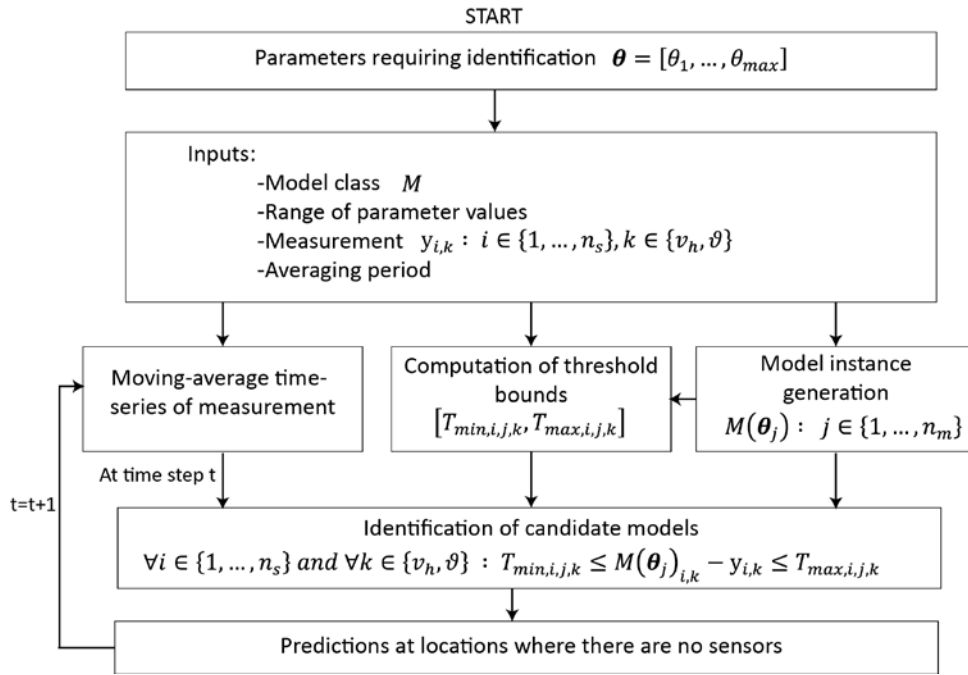
234 The final candidate model set is employed to make predictions where there are no sensors. Predictions  
 235  $P_{k,l}$  of flow quantities  $k \in \{v_h, \vartheta, \dots\}$  at locations where there are no sensors  $l$  are calculated using Eq.  
 236 (5).

$$237 \quad P_{k,l} = M(\theta_j)_{l,k} + \epsilon_{model,j,l,k} \quad (5)$$

238 Where  $M(\theta_j)_{l,k}$  is the predicted value of quantity  $k \in \{v_h, \vartheta, \dots\}$  at a location where there are no  
 239 sensors,  $l$ , by the candidate model  $j \in \{1, \dots, n_{cand}\}$ . The term,  $n_{cand}$  is the candidate-model number.  
 240  $\epsilon_{model,j,l,k}$  is the modelling error of the model instance  $j \in \{1, \dots, n_{cand}\}$  in the prediction of flow  
 241 quantity  $k \in \{v_h, \vartheta, \dots\}$  at locations where there are no sensors  $l$ .

242 Distributions of predictions of flow quantities  $k \in \{v_h, \vartheta, \dots\}$  at locations where there are no sensors  
 243 are obtained after falsification. Modelling error is then added to these distributions. Monte Carlo  
 244 techniques are employed to draw samples from modelling-error distributions. Bounds of modelling  
 245 error are the same as those considered for the identification of parameter values. A uniform distribution  
 246 is assumed between bounds. It is common to use a uniform distribution in absence of more precise  
 247 information (Goulet and Smith, 2011). For each candidate model (each plausible set of parameter  
 248 values), 1000 samples are generated by adding a random value of modelling error  $\epsilon_{model,j,l,k}$  to the  
 249 prediction of the candidate model  $M(\theta_j)_{l,k}$ . This results in distributions of values of flow quantity  $k \in$   
 250  $\{v_h, \vartheta, \dots\}$  at locations where there are no sensors. Finally, ranges of predictions of flow quantity  $k \in$   
 251  $\{v_h, \vartheta, \dots\}$  at locations where there are no sensors are calculated using a confidence level of 95%. Figure  
 252 1 summarizes the steps leading to the identification of inlet conditions given observations from  
 253 measurements and the subsequent predictions of flow quantities where there are no sensors. A moving-  
 254 average time series of horizontal wind speeds  $v_h$  and wind directions  $\vartheta$  is computed at each sensor  
 255 location using the measurement data and the averaging window. The methodology is repeated for all

256 time steps of the time series as shown in Figure 1. Distinct sets of model instances representing sets of  
 257 inlet conditions are identified at each time step.



258

259 Figure 1: Framework for the identification of plausible parameter values of CFD simulations given  
 260 observations from measurements as well as the subsequent predictions of flow quantities

#### 261 4) NUMERICAL SIMULATIONS

262 Computational fluid-dynamics (CFD) simulations have been carried out in order to simulate the airflow  
 263 pattern around the “BubbleZERO”. The “BubbleZERO” is an experimental facility of the Future Cities  
 264 Laboratory (FCL) Singapore-ETH Centre for Global Environmental Sustainability located at the  
 265 National University of Singapore. The commercial code FLUENT 14.5 is used to solve the fluid  
 266 equations of motion using the control volume method.

267 Only the near surroundings, including buildings and vegetation, of the “BubbleZERO” are modelled.  
 268 Other buildings located within the computational domain are modelled implicitly using an increased  
 269 equivalent roughness length  $y_0$ . Best practice guidelines were used to determine the size of the  
 270 computational domain (Franke et al., 2011), creating a computational domain with dimensions  
 271 length×width×height = 220m×140m×40m. CutCell meshing was employed in order to divide the  
 272 computational domain into small sub-regions resulting in a grid with approximately  $3.9 \times 10^5$   
 273 hexahedral cells. CutCell meshing is a method developed for FLUENT that generates a high percentage  
 274 of hexahedral cells with minimal user setup. Hexahedral cells are preferable to tetrahedral cells because  
 275 they introduce smaller truncation errors and provide a better iterative convergence (Franke, 2006). The  
 276 expansion ratio between two adjacent cells is set to 1.1. A grid-sensitivity analysis has been performed.  
 277 Two new grids consisting of  $2.5 \times 10^5$  cells and  $6.9 \times 10^5$  cells have been generated by varying the  
 278 grid settings. The grid-sensitivity analysis has been performed only for one model instance; we assume  
 279 that the grid settings defined after the analysis give good results for the other model instances.  
 280 Horizontal wind-speed predictions at sensor locations have been compared for the three grids. On  
 281 average, over all sensor locations, the difference between the prediction of the coarse grid and the  
 282 prediction of the middle grid is 4.2% of the prediction of the middle grid. Furthermore, the difference  
 283 between the prediction of the middle grid and the prediction of the fine grid is 2.7% of the prediction  
 284 of the fine grid. Predictions of the middle grid do not differ significantly from those of the fine grid.

285 Larger differences are observed between predictions of the coarse grid and predictions of the middle  
 286 grid. Therefore, the middle grid has been selected.

287 The SIMPLE algorithm is employed to deal with the pressure-velocity coupling (Patankar and Spalding,  
 288 1972). The second-order discretization scheme is used to interpolate pressure from the values obtained  
 289 at cell centres. The convergence criteria are based on the scaled residuals for all variables, which are  
 290 set to  $10^{-4}$ . Complexity of trees is reduced by using a tree-canopy model. Trees are modelled using a  
 291 porous zone (Guo and Maghirang, 2012) defined by a level of porosity of 1 and an isotropic inertial  
 292 resistance of  $0.1\text{m}^{-1}$ . Isothermal simulations were performed. This can be justified due to the cloudy  
 293 conditions following rain that occurred during the measurement period. The 3D steady RANS equations  
 294 in combination with the Realizable k-epsilon model have been used to model the turbulent airflow.

295 Vertical profiles of mean wind speed  $U$ , turbulent kinetic energy  $k$  and turbulence dissipation rate  $\varepsilon$  are  
 296 imposed at the inlet of the computational domain using a user-defined function (UDF) in FLUENT. The  
 297 flow conditions at the inlet of the computational domain are described in Eq. (6), Eq. (7) and Eq. (8)  
 298 which have been derived from neutral atmospheric conditions.

$$299 \quad U(y) = \frac{u_{ABL}^*}{\kappa} \ln\left(\frac{y + y_0}{y_0}\right) \quad (6)$$

$$300 \quad k(y) = \frac{u_{ABL}^{*2}}{\sqrt{C_\mu}} \quad (7)$$

$$301 \quad \varepsilon(y) = \frac{u_{ABL}^{*3}}{\kappa(y + y_0)} \quad (8)$$

302 where  $y$  is the height coordinate,  $\kappa$  is the von Karman constant,  $u_{ABL}^*$  is the atmospheric boundary layer  
 303 (ABL) friction velocity and  $C_\mu$  is a model constant.

304 The standard-wall function is used to treat the near-wall behaviour of airflow (Launder and Spalding,  
 305 1974). A non-dimensional wall distance  $y^+ > 30$  is achieved at all computational nodes adjacent to  
 306 wall surfaces as recommended by Franke et al. (2011). In FLUENT, the sand-grain roughness  $k_s$  is  
 307 employed to describe the surface roughness. For the standard wall function implemented in FLUENT,  
 308 a relationship between  $k_s$  and the roughness length  $y_0$  (commonly used in wind-engineering problems)  
 309 has been established by Blocken et al. (2007). This relation is expressed in Eq. (9).

$$310 \quad k_s = \frac{9.793y_0}{C_s} \quad (9)$$

311 where  $C_s$  is the roughness constant. In FLUENT, the value of  $k_s$  cannot be larger than  $y_p$ , which is the  
 312 distance between the centroid of the wall-adjacent cell and the wall.

313 The roughness length  $y_0$  of the atmospheric boundary layer with implicitly modelled buildings is set to  
 314  $y_0 = 0.45\text{m}$ , which represent dense, low buildings (Wieringa, 1992). Therefore, values of sand-grain  
 315 roughness and roughness constant have been set to  $k_s = 0.73\text{m}$  and  $C_s = 6$  in FLUENT. The roughness  
 316 length imposed at the ABL surface is the same as the roughness length used to compute the inlet wind  
 317 profiles in order to avoid stream gradient due to roughness modification in the upstream part of the  
 318 computational domain (Blocken et al., 2007). The building surfaces are defined with a zero sand-grain  
 319 roughness ( $k_s = 0\text{m}$ ). The sides and the top of the computational domain are modelled using symmetry  
 320 boundary conditions. Hence, zero normal velocity is imposed at the sides and top of the computational



321 domain. The outlet of the computational domain was modelled assuming zero pressure boundary  
 322 conditions.

323 Airflow responses depend upon the inlet conditions of the CFD simulation. In urban areas,  
 324 representative values of inlet conditions are difficult to estimate or measure (Schatzmann and Leitl,  
 325 2011). In the model-based data interpretation framework, representative values of inlet conditions are  
 326 identified from measurements taken in the UCL. The parameters requiring identification  $\theta$  are the wind  
 327 direction at the inlet  $\vartheta_{inlet}$  as well as the reference wind speed at the inlet at 16m height  $U_{ref}$ . A  
 328 population of model instances  $M(\theta_j)$  are generated through assigning sets of parameter values  $\theta_j =$   
 329  $[\vartheta_{inlet,j}, U_{ref,j}]$  to a model class  $M$ .

330 A simple-grid sampling is employed in order to generate model instances uniformly within the  
 331 parameter space. Table 1 presents the ranges of values of parameter requiring identification as well as  
 332 their discretization intervals. The simulations were executed using 12 processes in parallel on a  
 333 Windows Server 2012 containing four Hexa-Core Intel Xeon E54607 2.20GHz processors and 64 GB  
 334 memory. A total of 504 simulations have been executed in batch mode which requires 48 hours of  
 335 simulations using FLUENT. The outside box of the computational domain changes its orientation  
 336 between runs in order to simulate different inlet wind directions  $\vartheta_{inlet}$ . This leads to the generation of  
 337 a new grid for each inlet wind direction. For each model instances, horizontal wind speeds  $v_h$  and wind  
 338 directions  $\vartheta$  are predicted at sensor locations. Figure 2 presents horizontal wind speeds at 1.5m height  
 339 predicted with two model instances. Those two model instances are characterized by different wind  
 340 directions at the inlet  $\vartheta_{inlet}$ .

341 *Table 1: Minimal and maximal bounds of model parameters requiring identification and size of intervals used in the simple-*  
 342 *grid sampling*

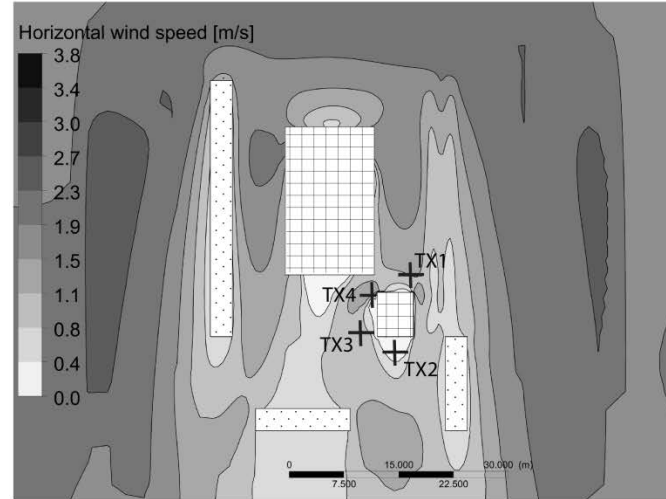
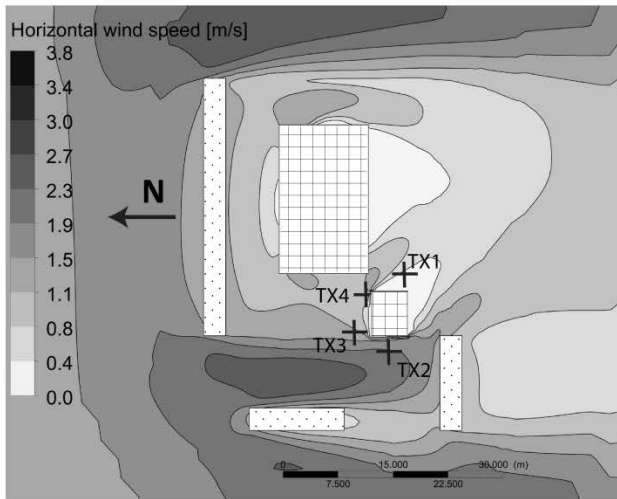
<b>Parameter requiring identification</b>	<b>Minimal Value</b>	<b>Maximal Value</b>	<b>Discretization intervals</b>
Wind direction at the inlet $\vartheta_{inlet}$ [°]	0	360	10
Reference wind speed at the inlet $U_{ref}$ (at 16m height) [m/s]	0.5	7	0.5




343

344

a

b



 Buildings
  Trees
  Sensors

345

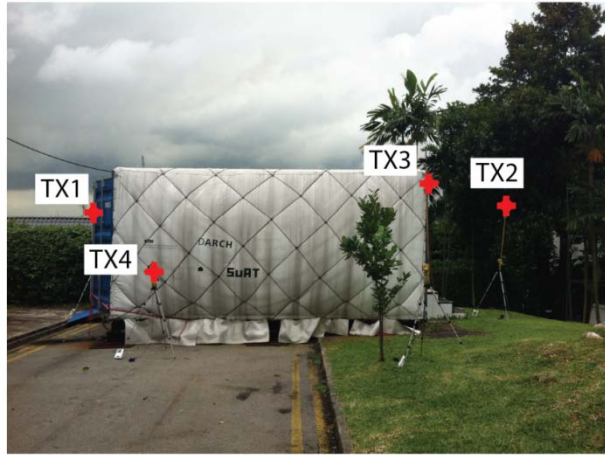
346 Figure 2: Horizontal wind speeds predicted by two model instances at 1.5m height. (a)  $U_{ref} =$   
 347  $3\text{m/s}$  (at 16m height),  $\vartheta_{inlet} = 0^\circ$ . (b)  $U_{ref} = 3\text{m/s}$  (at 16m height),  $\vartheta_{inlet} = 90^\circ$ , Wind from  
 348 the North is at  $0^\circ$ .

349

## 350 5) EXPERIMENTAL SETUP AND FIELD MEASUREMENTS

351 A measurement campaign was carried out around “BubbleZERO” in order to improve our  
 352 understanding of the system. The locations of the sensors are shown in Figure 3. The sensors at locations  
 353 TX1, TX2 and TX4 were used to falsify model instances that are not defined with plausible sets of  
 354 parameter values (inlet conditions). The sensor at location TX3 was used to validate the methodology,  
 355 by treating its measurements as test data and not as “training” data used for falsifying model instances.

356 Measurements were made with four wind-cup anemometer positioned on tripods at heights of 1.5m or  
 357 2.5m above ground level. Sensors measure horizontal wind speeds  $v_h$  as well as wind directions  $\vartheta$ . A  
 358 wireless device was used to record data measured from all the wind cup anemometers simultaneously.  
 359 The precisions of the sensors were 0.1m/s for horizontal wind speed measurements and  $22.5^\circ$  for wind  
 360 direction measurements. The resolutions of the sensors were 0.4m/s. Horizontal wind speeds were  
 361 sampled every 2.25 seconds while wind directions were sampled every second. Measurement data were  
 362 recorded every 10 seconds. The recorded horizontal wind speed corresponded to the average value  
 363 during the recording interval while the recorded wind direction corresponded to the dominant wind  
 364 direction during the recording interval.

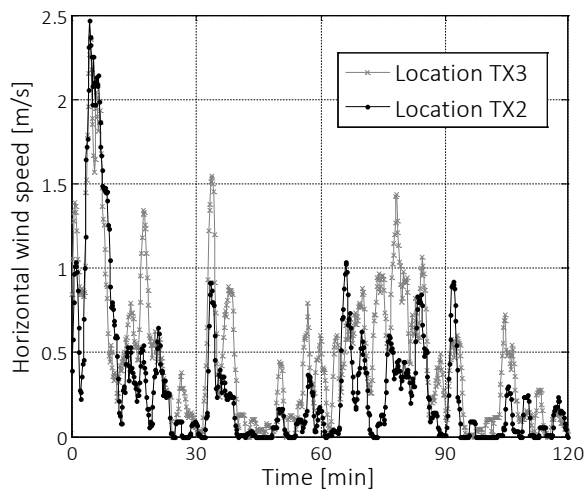


365

366 Figure 3: Picture of the “BubbleZERO” and position of wind-cup anemometers, TX.

367 Figure 4 presents the moving-average time series of measurements at location TX3 and TX2 using an  
 368 averaging window of 90 seconds. The moving-average time series are computed by replacing the  
 369 recorded value (every 10 seconds) with the 90s-averaged value of its neighbour steps. An averaging  
 370 window of 90 seconds has been employed in this paper. The time variability of horizontal wind speeds  
 371 is clearly observed in Figure 4. If a longer averaging window is used, the time variability of atmospheric  
 372 conditions may not have been captured with the proposed model-based data interpretation framework.  
 373 Measurements were carried out on 18<sup>th</sup> December 2012 from 1pm to 3pm following rain. Surface  
 374 heating was negligible due to the cloudy conditions occurring during the measurement campaign.

375 Even though measurements were carried out at locations that are close together, variations of wind  
 376 quantities are observed between measurement stations. Larger spatial variability of wind quantities  
 377 might be observed if the measurement stations were located further away from each other.



378

379 Figure 4: Moving-average time series of horizontal wind speed at locations TX3 and TX2 (averaging  
 380 window of 90 seconds)

## 381 6) MODELLING ERROR

382 An important step that determines the reliability of the identification of parameter values  $\theta =$   
383  $[\vartheta_{inlet}, U_{ref}]$  is the correct estimation of the threshold bounds  $[T_{min,i,j,k}, T_{max,i,j,k}]$ , defined by  
384 combining measurement and modelling error as expressed in Eq. (3). An estimated range of modelling  
385 error in the predictions of horizontal wind speed  $v_h$  and wind direction  $\vartheta$  is required for each model  
386 instance  $j \in \{1, \dots, n_m\}$  and at each sensor location  $i \in \{1, \dots, n_s\}$  in order to define those bounds. The  
387 goal of this Section is to define a relationship between errors associated with a RANS-based model and  
388 input/output variables of this model (e.g. errors related to the amplification factor of wind speeds as  
389 mentioned in Section 2).

390 Ranges of modelling error have been estimated by comparing responses of CFD simulations obtained  
391 with two turbulence models around a single bluff body having dimensions length=6.06m, width=4.88m  
392 and height=2.9m which corresponds to the building of interest in the case study presented in Section 4.  
393 Dimensions of the computational domain are length×width×height = 55.5m×31.06m×14.5m. The first  
394 turbulence model is the same as the one used for the generation of model instances  $M(\theta_j)$ . This  
395 turbulence model is built on the steady RANS equations in combination with the Realizable k-epsilon  
396 model. The same numerical methods as those described in Section 4 are employed. The second  
397 turbulence model is LES using the dynamic Smagorinsky model to model the small eddies of the flow  
398 (Germano et al., 1991). Unlike RANS-based model, LES provides time-dependent predictions of flow  
399 quantities. LES has been found to be in good agreement with wind tunnel experiments (Cheng et al.,  
400 2003; Tominaga et al., 2008a) while RANS-based models have difficulties to reproduce flow separation  
401 and recirculation because of the unsteadiness of the flow field in those regions. However, since it is  
402 computationally prohibitive to perform large numbers of simulations using LES, simulations using a  
403 RANS-based model have been employed in the model-based data interpretation framework. Estimates  
404 ranges of modelling errors are determined from the differences between responses of LES and responses  
405 of the RANS-based model. A grid composed of  $7.1 \times 10^4$  cells is used for the RANS-based simulation  
406 and the LES. A grid-sensitivity analysis has been performed using the RANS-based model in order to  
407 define the grid settings, similarly to the analysis performed in Section 4. Scaled residuals of  $10^{-4}$  for all  
408 variables are used as convergence criteria for the RANS-based simulation and the LES.

409 Other sources of error, such as those associated with geometry simplification, have not been considered.  
410 Furthermore, LES results are not perfectly correct. Thus, the threshold bounds employed in this paper  
411 are lower bound estimates.

412 Profiles of wind speed  $U$ , turbulent kinetic energy  $k$  and turbulent dissipation energy  $\varepsilon$  imposed at the  
413 inlet of the RANS-based model are defined using Eq. (6), Eq. (7) and Eq. (8). The ground surface and  
414 building surfaces are treated with standard wall functions (Launder and Spalding, 1974) using a zero  
415 roughness length ( $y_0=0$ m). The inlet wind speed at 1.5m height is used as reference for the definition  
416 of the wind speed profile and has been set to  $U_{1.5m} = 2$ m/s. The turbulence characteristics defined at  
417 the inlet of the RANS-based model are reproduced in LES by imposing a time-dependent inlet velocity  
418 profile generated using the vortex method in FLUENT 14.5. Responses of the RANS-based model are  
419 used as initial conditions for the LES simulation. A time step size of 0.1 second is used for the LES  
420 simulation. After 1 hour of real-time, representative mean values of flow quantities are reached.

421 In Section 6.1, responses of the RANS-based model are compared with mean responses of LES in order  
422 to estimate error of RANS-based models in the predictions of mean flow quantities.

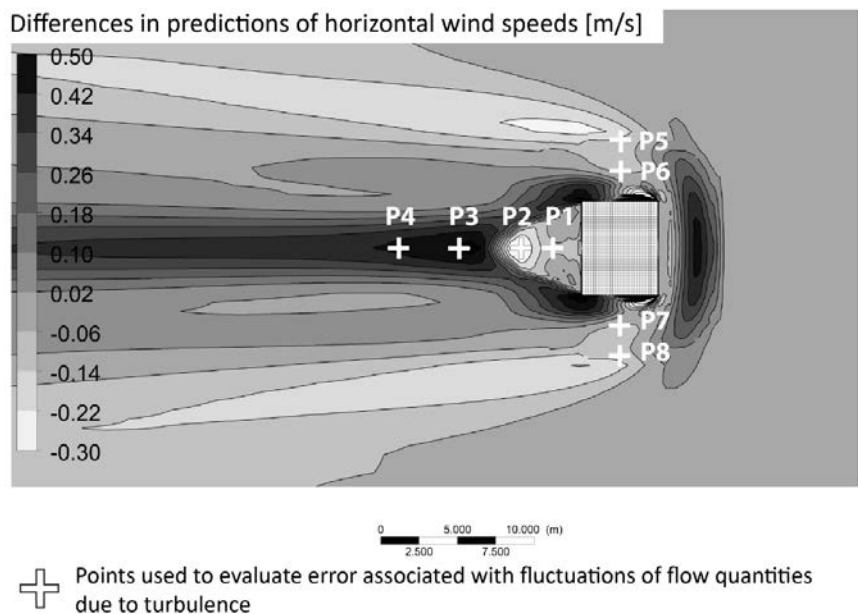
423 At each time step of the time series of measurement, measured values at sensor locations are compared  
424 with model predictions. Measurements need to be averaged over a short period of time in order to

425 capture variation of inlet conditions with the proposed model-based data interpretation framework. In  
 426 Section 6.2, LES is employed to have an estimate of error associated with fluctuations of flow quantities  
 427 in which amplitude depends upon the averaging period chosen. In Section 6.3, ranges of errors are  
 428 estimated for different inlet wind speeds.

429 *6.1) Estimated ranges of error in the prediction of mean flow quantities*

430 The differences between the mean responses of LES and the responses of the RANS-based model are  
 431 employed to estimate error of RANS-based models in the predictions of mean flow quantities.

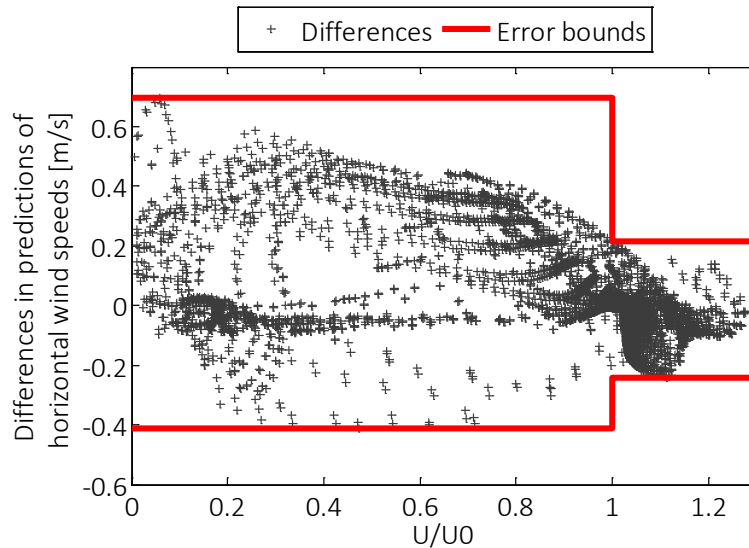
432 Figure 5 presents the spatial distribution of these differences for predictions of mean horizontal wind  
 433 speeds at 1.5m height. In this study, responses of the RANS-based model have been compared to  
 434 responses of LES at 1.5m height. This corresponds approximately to half of the height of the building.



435  
 436 Figure 5: Differences between mean horizontal wind speeds predicted with LES and horizontal wind  
 437 speeds predicted with the RANS-based model at 1.5m height (wind coming from the right).

438 Figure 6 shows values of differences with respect to the amplification factor of wind speeds  $U/U_0$   
 439 predicted with the RANS-based model at nodes located at 1.5m height. The amplification factor of wind  
 440 speeds is defined as the ratio between the local horizontal wind speed,  $U$ , to the horizontal wind speed,  
 441  $U_0$ , that would occur without buildings. In this study,  $U_0$  corresponds to the inlet wind speed at the same  
 442 height as the local horizontal wind speed  $U$ . This assumption is justified because the wind speed profile  
 443 imposed at the inlet is compatible with the roughness imposed at ground level. Therefore, no stream  
 444 gradient due to roughness modification in the upstream part of the computational domain is expected.  
 445 It becomes clear that the range of differences is larger in the region where the amplification factor of  
 446 wind speeds is less than unity ( $U/U_0 < 1$ ).

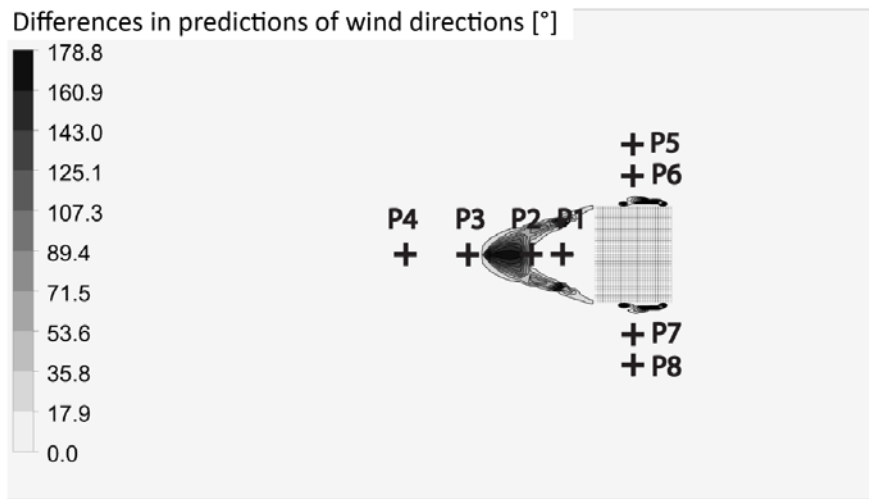
447 Therefore, distinction is made between sensors located in the region of high wind speeds ( $U/U_0 > 1$ ) and  
 448 those located in the region of low wind speeds ( $U/U_0 < 1$ ) in the model-based data interpretation  
 449 framework. For each region, estimated bounds of error in the predictions of mean horizontal wind  
 450 speeds are defined as the maximal values of differences observed in the region. Locations of these  
 451 regions are different for each model instance. Ranges of error at sensor locations, and therefore  
 452 threshold bounds, are adjusted depending on which region the sensor is located.



453

454 Figure 6: Differences between mean horizontal wind speeds predicted with LES and horizontal wind  
 455 speeds predicted with the RANS-based model at 1.5m;  $U_0=2\text{m/s}$

456 The same methodology is repeated in order to estimate errors of RANS-based models in the predictions  
 457 of wind directions. Figure 7 presents differences between mean wind directions predicted with LES and  
 458 wind directions predicted with the RANS-based model at 1.5m height. In some locations, large  
 459 differences are observed. This originates from the over-estimation of the region of reverse flow in the  
 460 wake of the building using RANS-based models. Therefore, RANS-based models may provide a reverse  
 461 flow at some locations in the wake of the building where LES does not.



+

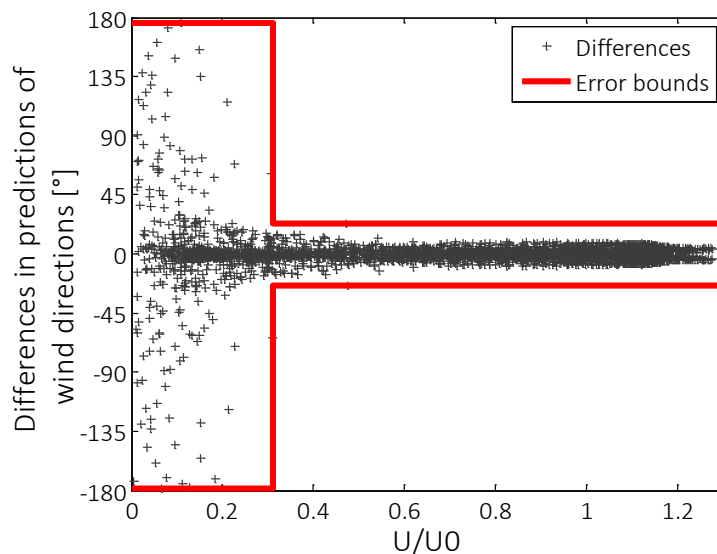
Points used to evaluate error associated with fluctuations of flow quantities due to turbulence

462

463 Figure 7: Differences between mean wind directions predicted with LES and wind directions predicted  
 464 with the RANS-based model at 1.5m height in absolute values

465 Figure 8 describes values of the differences with respect to the amplification factor of wind speeds  
 466  $U/U_0$ . Large range of differences is observed in the region where the amplification factor of wind speeds  
 467 is low  $U/U_0 < 0.33$ . In this region, the range of difference can be up to  $[-180^\circ, 180^\circ]$ . Therefore, sensors  
 468 located in this region will not add to knowledge of airflow behaviour. In the model-based data-

469 interpretation framework, wind directions measured by sensors located in this region are not employed  
470 to falsify model instances. Locations of such region are different for each model instance.



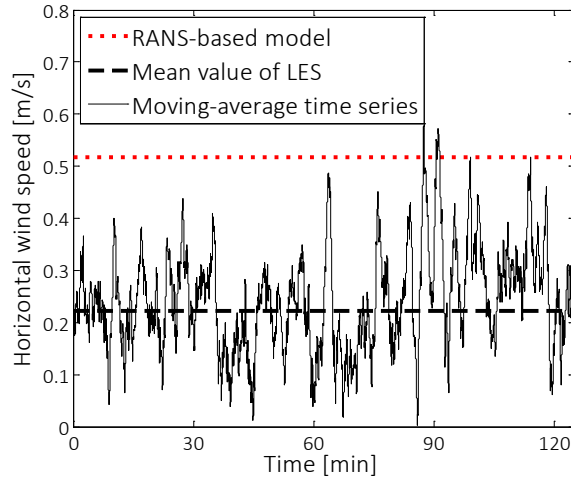
471  
472 Figure 8: Differences between mean wind directions predicted with LES and wind directions predicted  
473 with the RANS-based model at 1.5m height;  $U_0=2\text{m/s}$

474 *6.2) Estimated ranges of error associated with fluctuations of flow quantities due to turbulence*

475 Discrepancies between measured and predicted values appear if the period used to average  
476 measurements is not sufficiently long because of the stochastic fluctuations of flow quantities due to  
477 turbulence.

478 Moving-average time series of horizontal wind speed are predicted at comparison points P1-P8 (Figure  
479 5) using LES in order to estimate a range of error associated with fluctuations of horizontal wind speed  
480 due to turbulence. The comparison point that provides the largest fluctuations is used to estimate bounds  
481 of error. Bounds are defined as the difference between the extreme values of the moving-average time  
482 series predicted with LES and the mean value of LES. No distinction is made between the region of  
483 high wind speeds ( $U/U_0 > 1$ ) and the region of low wind speeds ( $U/U_0 < 1$ ).

484 Figure 9 presents the horizontal wind speed predicted with the RANS-based model as well as the  
485 moving-average time series predicted with LES at comparison point P1. Point P1 is the point that  
486 provides the largest fluctuations of horizontal wind speed. The red-dotted line corresponds to the value  
487 of horizontal wind speed predicted with the RANS-based model. The black-dashed line is the mean  
488 value of horizontal wind speed predicted with LES. The black line corresponds to the moving-average  
489 time series of horizontal wind speeds predicted with LES using an averaging window of 90 seconds.



490

491 Figure 9: Horizontal wind speed predicted with the RANS-based model at point P1 (red-dotted line),  
 492 mean horizontal wind speed predicted with LES (black-dashed line) and the moving-average time series  
 493 of horizontal wind speed predicted with LES (black line); averaging window=90s;  $U_0=2\text{m/s}$

494 The same methodology is employed to evaluate bounds of error associated with fluctuations of wind  
 495 directions due to turbulence. Only comparison points characterized by an amplification factor of wind  
 496 speeds  $U/U_0 > 0.33$  (P3-P8) are considered. In order to deal with the discontinuity of wind direction  
 497 between  $0^\circ$  and  $360^\circ$ , the average value of wind direction  $\bar{\vartheta}$  has been determined using Eq. (10).

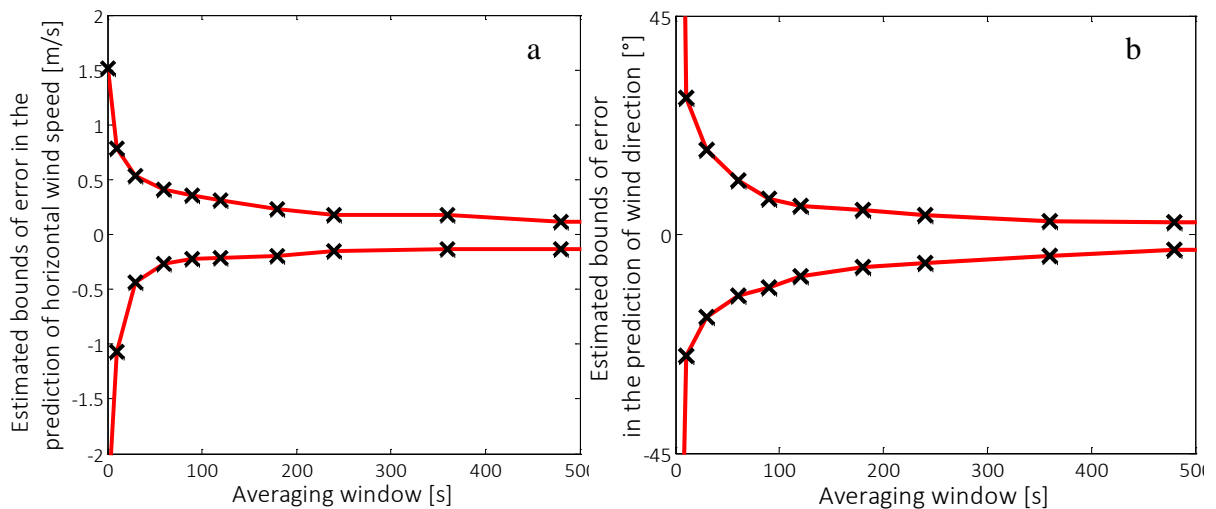
$$498 \quad \bar{\vartheta} = \text{atan2}(\overline{\sin \vartheta}, \overline{\cos \vartheta}) \quad (10)$$

499 Where  $\overline{\sin \vartheta}$  and  $\overline{\cos \vartheta}$  correspond to the average value of the sine and cosine of the wind direction  $\vartheta$ .

500 It is clear that the maximal fluctuations of flow quantities depend on the averaging window chosen.  
 501 Figure 10 presents the estimated bounds of error in the predictions of horizontal wind speeds and wind  
 502 directions with respect to the averaging window. If the averaging window increases, the estimated  
 503 ranges of error decrease.

504

505





506

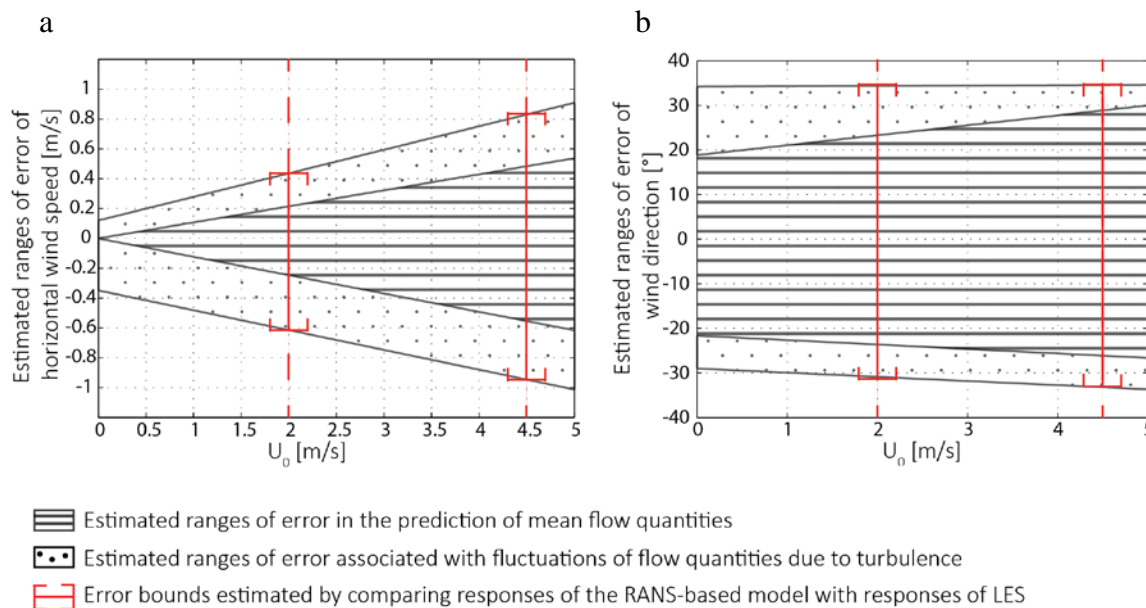
507 Figure 10: Estimated bounds of errors associated with the fluctuations of flow quantities due to  
508 turbulence with respect to the averaging window. (a) Bounds of error in the predictions of horizontal  
509 wind speeds. (b) Bounds of error in the predictions of wind directions;  $U_0=2\text{m/s}$

510 6.3) Estimated ranges of error for different inlet wind speeds

511 Bounds of modelling error  $\epsilon_{model}$  are computed by summing the estimated bounds of error of RANS-  
512 based models in the predictions of mean flow quantities (Section 6.1) with those associated with the  
513 fluctuations of flow quantities due to turbulence (Section 6.2).

514 Two sets of simulations with two inlet wind speeds  $U_0$  have been executed in order to evaluate the  
515 influence of the inlet wind speed on the estimated ranges of modelling error. In order to estimate ranges  
516 of error for other inlet wind speeds, a linear interpolation has been assumed. In future work, RANS-  
517 based models and LES will be compared with other inlet wind speeds in order to better define the  
518 relationship between inlet wind speed and ranges of modelling error.

519 Figure 11 presents the estimated ranges of error of RANS-based models in the predictions of mean flow  
520 quantities as well as the estimated ranges of error associated with the fluctuations of flow quantities due  
521 to turbulence with respect to the inlet wind speed  $U_0$ . Figure 12a presents the estimated ranges of errors  
522 in the predictions of horizontal wind speeds in the region of high wind speeds ( $U/U_0>1$ ). The estimated  
523 ranges of errors in the predictions of wind speeds in the region of low wind speeds ( $U/U_0<1$ ) are not  
524 represented in this figure. Figure 12b presents the estimated ranges of error in the predictions of wind  
525 directions in the region characterized by  $U/U_0>0.33$ .



526

527 Figure 11: Estimated ranges of modelling errors with respect to the inlet wind speed  $U_0$  (averaging  
528 window = 90s). (a) Modelling error associated with predictions of wind speeds in the region where  
529  $U/U_0>1$ . (b) Modelling error associated with predictions of wind directions in the region where  
530  $U/U_0>0.33$ .

531 **7) RESULTS**

532 7.1) Identification of inlet conditions

533 This Section presents the results of the model-based data interpretation framework using predictions of  
534 the model instances (Section 4), the measurement data (Section 5) as well as the knowledge of  
535 modelling errors (Section 6). In this framework, a population of model instances  $M(\theta_j)$  is generated  
536 through assigning parameter values  $\theta_j = [\vartheta_{inlet,j}, U_{ref,j}]$  to a model class  $M$ . Measurements are  
537 employed to falsify model instances that are not defined with plausible sets of parameter values.

538 Measurement errors are determined according to the precision of the sensor (Section 5). Moreover, if  
539 the measured value of horizontal wind speed is below the starting threshold of the sensor (0.4m/s),  
540 values of measurement errors are adjusted.

541 A study has been performed in order to evaluate errors associated with sensor-location uncertainties. In  
542 this study, one model instance has been used. For each sensor location, wind speeds and wind directions  
543 are predicted at 100 points located within a distance of 20cm from the sensor location. The maximal  
544 differences between predictions at those points and predictions at sensor locations are used to estimate  
545 these errors. The maximal range of differences for wind speeds is [-0.069m/s, 0.062m/s] and the  
546 maximal range of differences for wind directions is [-3.3°, 2.9°]. These errors have been added to the  
547 other sources of errors using elementary rules of interval arithmetic (Equation 3).

548 Model instances are candidates if the residual, defined as the difference between measured and predicted  
549 values of horizontal wind speed and wind direction, falls within the threshold bounds at each and every  
550 sensor location and for each and every compared quantity. In Figure 12, the final candidate model set  
551 is represented for two different time steps  $t_1$  and  $t_2$ . Green dots correspond to candidate models while  
552 red crosses correspond to falsified models. Wind direction measured at sensor location TX2 is compared  
553 with predictions of the model instances (Figure 12a and 12b). Horizontal wind speed measured at sensor  
554 location TX2 is compared with predictions of the model instances (Figure 12c and 12d). The  
555 comparison between measured and predicted flow quantities at the other two sensor locations (TX4 and  
556 TX1) are not represented in this figure. The purple-dashed lines represent the measured values of flow  
557 quantities at location TX2. The blue lines represent the threshold bounds computed at location TX2 in  
558 order to falsify model instances using measurements.

559 Threshold bounds  $[T_{min,i,j,k}, T_{max,i,j,k}]$  are computed at each sensor location  $i \in \{TX1, TX2, TX4\}$ , for  
560 each model instance  $j \in \{1, \dots, 504\}$  and for each compared quantities  $k \in \{v_h, \vartheta\}$ . These bounds are  
561 defined by summing the bounds of modelling and measurement error. Ranges of modelling errors are  
562 determined according to the inlet wind speed at the height of the sensor, the averaging window chosen  
563 as well as the predicted amplification factors of wind speeds at sensor locations (Section 6). Model  
564 instances are not defined by the same inlet wind speed and they don't predict the same amplification  
565 factor of wind speed at sensor locations. Therefore, threshold bounds might be different for each model  
566 instance as shown in Figure 13.

567 Horizontal wind speed measured at location TX2 at time  $t_1$  differs largely from horizontal wind speed  
568 measured at time  $t_2$ , resulting in two sets of candidate models. This highlights the dynamic identification  
569 of inlet conditions in the proposed model-based data interpretation framework. Unlike single-model  
570 approaches, the proposed framework accommodates the time variability of atmospheric conditions.

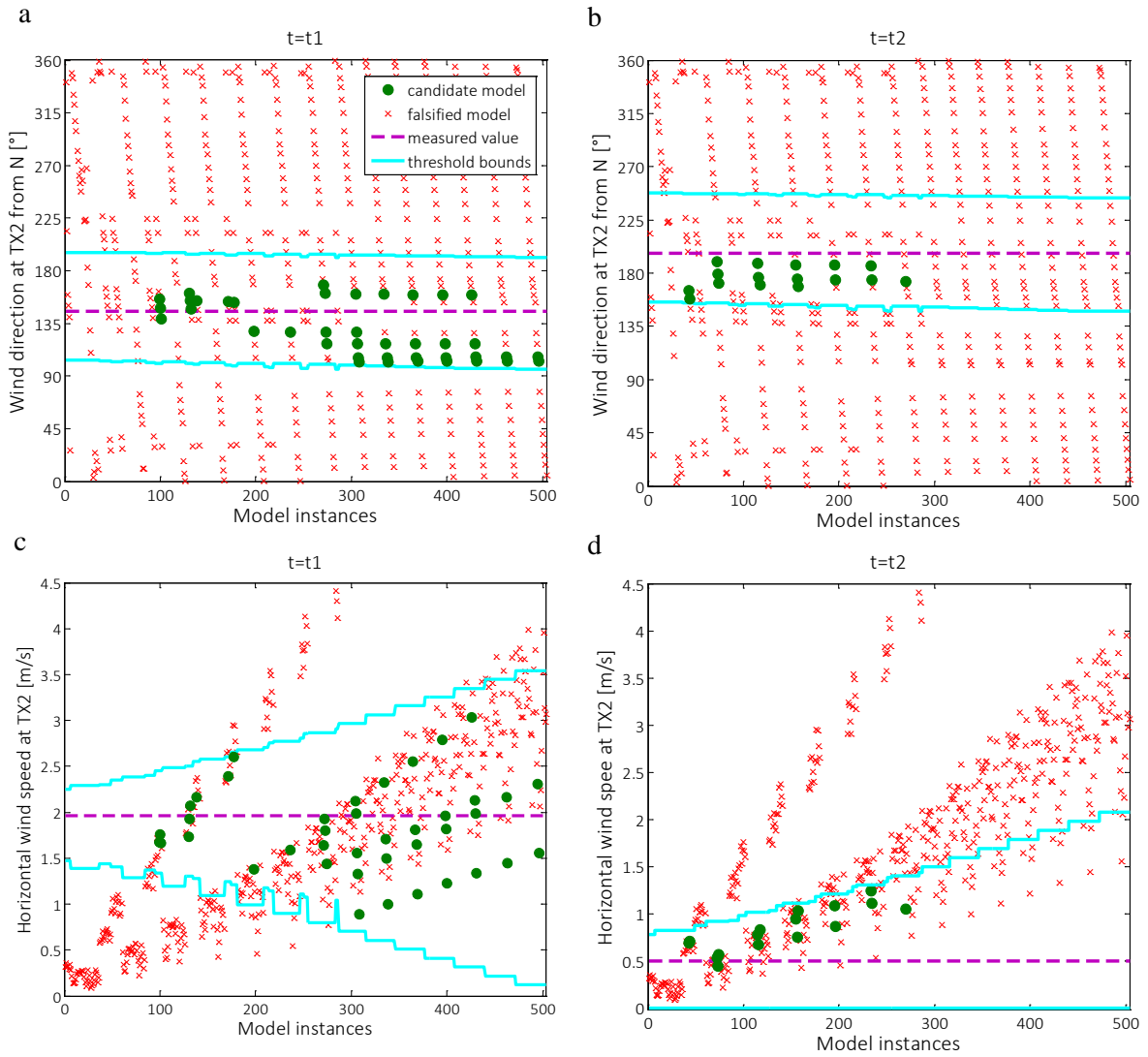
571

572

573

574

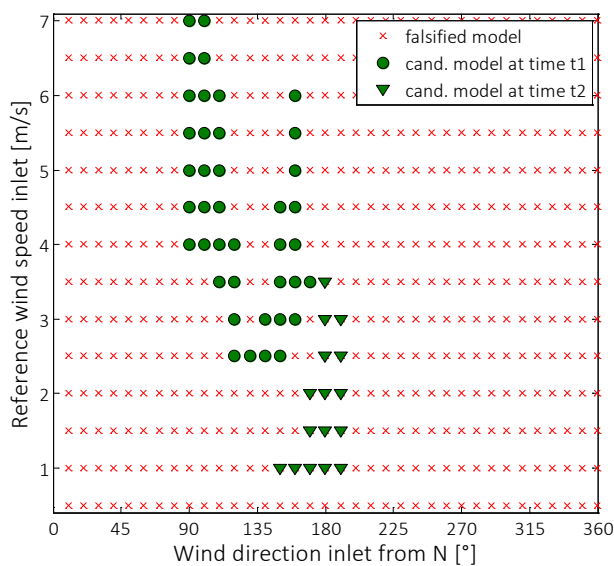
575  
 576  
 577  
 578  
 579



580 Figure 12: Comparison between measured and predicted flow quantities at location TX2. (a)  
 581 Comparison of wind direction at  $t_1=300$ s. (b) Comparison of wind direction at  $t_2=1000$ s. (c)  
 582 Comparison of horizontal wind speed at  $t_1=300$ s. (d) Comparison of horizontal wind speed at  $t_2=1000$ s.

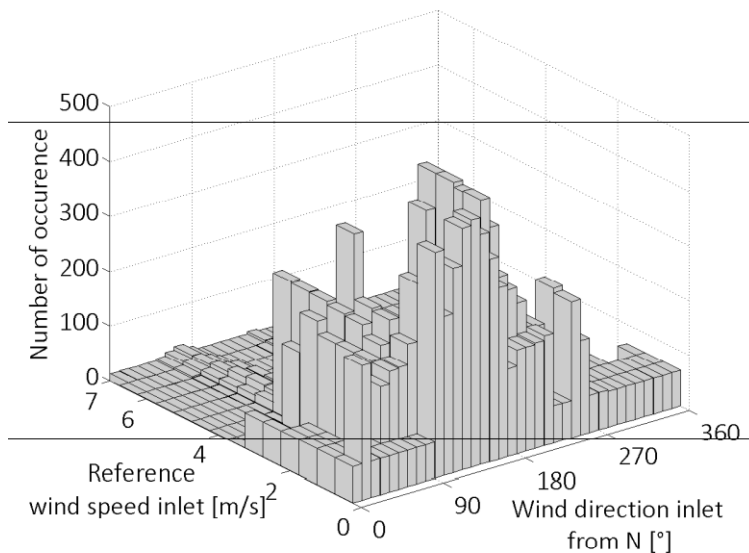
583 Figure 13 presents the final candidate model sets obtained after falsification using measurement taken  
 584 at locations TX1, TX2 and TX4 represented in the parameter space. Green dots correspond to the final  
 585 candidate model set obtained at time  $t_1$  while the green triangles correspond to the final candidate model  
 586 set obtained at time  $t_2$ . Red crosses correspond to the falsified models. It becomes clear that the initial  
 587 range of parameter values have been reduced through the use of measurement data taken in the UCL.  
 588 At time  $t_1$ , the number of candidate models is reduced from 504 to 40. At time  $t_2$ , the number of  
 589 candidate models is reduced from 504 to 16. The final candidate model set at time  $t_1$  is different from  
 590 the final candidate model set obtained at time  $t_2$ . These two final candidate model sets represent very

591 different boundary conditions existing at the selected time steps. It further describes typical time  
 592 variability of airflow in urban areas.



593  
 594 Figure 13: Final candidate model set at time  $t_1=300s$  (green dots) and at time  $t_2=1000s$  (green triangles)  
 595 represented in the parameter space

596 Figure 14 presents the number of occurrence of each model instance over time. Each bin represents one  
 597 model instance. The height of the bin represents the number of occurrences. Model instances with high  
 598 occurrences represent predominant inlet conditions.



599  
 600 Figure 14: Number of occurrences of each model instance over time

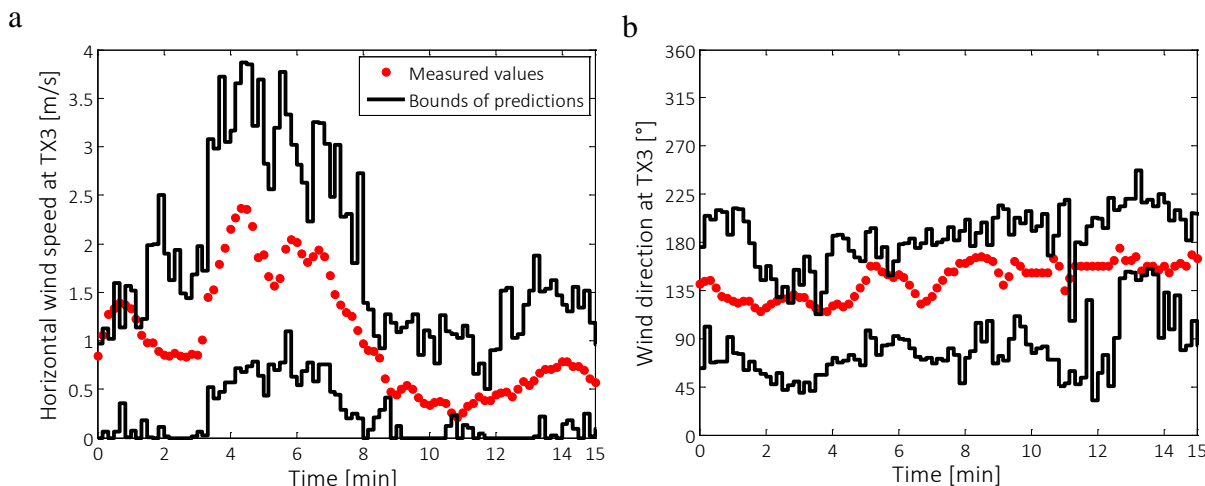
601 *7.2) Predictions of flow quantities where there are no-sensors*

602 At each time step (every 10 seconds), the final candidate model set is employed to predict flow  
 603 quantities where no measurements are carried out. Modelling uncertainty is combined to the distribution  
 604 of predictions obtained with the final candidate set as explained in Section 3.3. Ranges of predictions  
 605 are calculated using the uncertainty combination and a confidence level of 95%.

606 For validation of the proposed framework, moving-average time series of horizontal wind speeds and  
 607 wind directions are computed at locations TX3 using the measurement dataset and an averaging window  
 608 of 90s. At each time step, measured values of horizontal wind speeds and wind directions are compared  
 609 with ranges of predictions obtained at locations TX3. It is emphasized that measurement data obtained  
 610 at location TX3 was not used to identify candidate models and was used only to test whether the  
 611 predictions of flow quantities are reliable.

612 Figure 15a presents ranges of predicted horizontal wind speeds as well as the moving-average time series  
 613 of horizontal wind speeds measured at location TX3. Figure 15b presents ranges of predicted  
 614 wind directions as well as the moving-average time series of wind directions measured at location TX3.  
 615 Only 15 minutes of the measurement dataset are represented in these figures. If we consider the whole  
 616 measurement campaign (2 hours), horizontal wind speeds measured at location TX3 fall within ranges  
 617 of predictions 96% of the time. Wind directions fall within ranges of predictions 93% of the time. This  
 618 demonstrates that predictions of flow quantities using the model-based data-interpretation framework  
 619 at location TX3 are fairly reliable.

620 Other sources of modelling error such as geometry simplification need to be considered in the model  
 621 falsification approach in order to improve the reliability of identification and subsequent predictions.  
 622 Moreover, the modelling errors have been estimated by comparing responses of a RANS-based model  
 623 with those of LES. Although LES is more accurate than RANS-based in regions of flow separation and  
 624 recirculation, LES is not an exact representation of reality. Therefore, only lower bound estimates of  
 625 modelling errors are determined in this paper.



626 Figure 15: Ranges of predictions as well as the moving-average time series of flow quantities at  
 627 location TX3 (averaging window of 90 seconds)

## 628 8) DISCUSSION

629 This paper proposed a model-based data interpretation framework for the assessment of airflow in urban  
 630 areas. In this framework, predictions obtained with CFD simulations are integrated with data obtained  
 631 with measurements as well as with knowledge of uncertainties in order to improve the accuracy of  
 632 airflow predictions in urban areas. A systematic approach for the estimation of modelling error  
 633 associated with RANS-based models has been incorporated to this framework. The results show that  
 634 predictions in terms of mean wind speeds and mean wind directions are more accurate in regions of  
 635 high wind speeds (high amplification factor of wind speeds) than in regions of low wind speeds. This  
 636 is in agreement with recent work by Blocken et al. (2011) on the performance of RANS-based models  
 637 for airflow modelling using CFD simulations. The framework accommodates the time variability of  
 638 atmospheric conditions by 1) identifying different sets of boundary conditions at each time step and by

639 2) adding additional source of errors associated with the fluctuations of flow quantities during the  
640 computation of thresholds bounds.

641 There are (of course) limitations. In the present study, only inlet conditions have been identified using  
642 measurements taken in the UCL. Other parameters that may influence the airflow around buildings,  
643 such as the roughness of the surrounding buildings or the inertial resistance of trees, have not been  
644 considered. In further studies, more parameters will be identified that have varying values with respect  
645 to time (e.g. inlet conditions) as well as constant values with respect to time (e.g. roughness of  
646 buildings). A grid-based sampling has been used for the generation of model instances in this work. In  
647 this approach, the size of the initial candidate model set increases exponentially with the number of  
648 parameters requiring identification. More efficient techniques may be used in order to decrease the  
649 complexity of sampling in high-dimensional parameter space. Moreover, if many parameter values need  
650 to be identified, care is needed in order to avoid parameter compensation at sensor locations.

651 CFD simulations have been performed on small-size buildings. Only the “BubbleZERO” and its near-  
652 surroundings are explicitly modelled. Therefore the computational domain is relatively small  
653 (length×width×height=220m×140m×40m). Sensors used in this study are located close to each other  
654 and therefore, the spatial variability of airflow is not well-pronounced. In future work, the framework  
655 will be tested and validated in a case study with larger computational domain and taller buildings. More  
656 sensors will be employed and they will be located further away from each other.

657 In this paper, only modelling errors in the predictions of mean flow quantities using RANS-based  
658 models as well as errors associated with fluctuations of flow quantities have been acknowledged. The  
659 modelling errors have been defined by comparing responses of a RANS-based model with those of LES  
660 around an isolated building with a cubical shape. Although LES is clearly more accurate than RANS-  
661 based models in the predictions of mean flow characteristics, LES does not perfectly simulate the  
662 airflow behaviour around buildings (Lim et al., 2009; Tominaga et al., 2008a).

663 Moreover, in other building configurations, ranges of errors as well as their relationship with the  
664 amplification factor of wind speeds or the averaging window may differ from those obtained in the  
665 present study. Furthermore, other sources of errors need to be considered in order to predict more  
666 reliable ranges of predictions at locations where there are no sensors. Sources include simplifications  
667 of urban shapes, idealisation of boundary conditions with a logarithmic profile that has been derived  
668 for neutral conditions as well as the assumption that thermal processes, such as convection, are  
669 negligible. Also, errors due to values of parameters that are difficult to quantify and that have not been  
670 identified, such as the roughness of the surrounding buildings or the inertial resistance of trees, influence  
671 reliability. Finally, the proposed methodology used to estimate spatial variations of ranges of modelling  
672 errors depending on the predicted amplification factor of wind speeds should help define optimal sensor  
673 configurations in further studies.

## 674 **9) CONCLUSIONS**

675 This work has led to the following conclusions:

- 676 1) The model-falsification methodology has much potential for interpreting sensor measurements  
677 to improve the accuracy of airflow simulations around buildings.
- 678 2) Adapting the error-domain model falsification approach to represent the dynamic behaviour of  
679 airflow has successfully led to narrowing down several hundred parameter-value sets to a few  
680 possible inlet conditions for the selected case-study. Thus the case-study illustrates an approach  
681 for identifying time-varying inlet conditions and predicting wind characteristics at locations  
682 where there are no sensors.

683 3) Modelling errors need to be recognized and quantified in order to perform reliable predictions  
684 of airflow characteristics at locations where there are no-sensors. Ranges of modelling errors  
685 depend on the predicted amplification factor of wind speeds, the wind speed at the inlet as well  
686 as the time period employed to average measurement data.

## 687 **ACKNOWLEDGMENT**

688 This work was established at the Singapore-ETH Centre for Global Environmental Sustainability  
689 (SEC), co-funded by the Singapore National Research Foundation (NRF) and ETH Zurich. The first  
690 author is a registered PhD student at EPFL. The authors would like to thank Prof. C. Sekhar, Prof. M.  
691 Santamouris, Dr. J-A. Goulet and M. Papadopoulou for their support. We also thank the anonymous  
692 reviewers for their comments.

## 693 **REFERENCES**

- 694 Al-Sallal, K.A., Al-Rais, L., 2011. Outdoor airflow analysis and potential for passive cooling in the  
695 traditional urban context of Dubai. *Renewable Energy* 36 (9), 2494-2501.
- 696 Allegrini, J., Dorer, V., Carmeliet, J., 2013. Wind tunnel measurements of buoyant flows in street  
697 canyons. *Building and Environment* 59 315-326.
- 698 Beranek, W., 1979. *Beperken van windhinder om gebouwen, deel 2*. Deventer: Stichting Bouwresearch  
699 no. 90. Kluwer Technische Boeken BV,
- 700 Beven, K., 2008. *Environmental modelling: An uncertain future?* Taylor & Francis,
- 701 Beven, K., 2006. A manifesto for the equifinality thesis. *Journal of hydrology* 320 (1), 18-36.
- 702 Blocken, B., Carmeliet, J., 2008. Pedestrian wind conditions at outdoor platforms in a high-rise  
703 apartment building: generic sub-configuration validation, wind comfort assessment and  
704 uncertainty issues. *Wind and Structures* 11 (1), 51-70.
- 705 Blocken, B., Persoon, J., 2009. Pedestrian wind comfort around a large football stadium in an urban  
706 environment: CFD simulation, validation and application of the new Dutch wind nuisance  
707 standard. *Journal of Wind Engineering and Industrial Aerodynamics* 97 (5), 255-270.
- 708 Blocken, B., Stathopoulos, T., Carmeliet, J., 2007. CFD simulation of the atmospheric boundary layer:  
709 wall function problems. *Atmospheric environment* 41 (2), 238-252.
- 710 Blocken, B., Stathopoulos, T., Carmeliet, J., Hensen, J.L., 2011. Application of computational fluid  
711 dynamics in building performance simulation for the outdoor environment: an overview.  
712 *Journal of Building Performance Simulation* 4 (2), 157-184.
- 713 Cheng, Y., Lien, F., Yee, E., Sinclair, R., 2003. A comparison of large Eddy simulations with a standard  
714  $k-\epsilon$  Reynolds-averaged Navier-Stokes model for the prediction of a fully developed turbulent  
715 flow over a matrix of cubes. *Journal of Wind Engineering and Industrial Aerodynamics* 91 (11),  
716 1301-1328.
- 717 Defraeye, T., Blocken, B., Carmeliet, J., 2011. Convective heat transfer coefficients for exterior  
718 building surfaces: Existing correlations and CFD modelling. *Energy Conversion and*  
719 *Management* 52 (1), 512-522.
- 720 Franke, J., 2006. Recommendations of the COST action C14 on the use of CFD in predicting pedestrian  
721 wind environment. In: *The fourth international symposium on computational wind engineering*,  
722 *Yokohama, Japan*, pp. 529-532.
- 723 Franke, J., Hellsten, A., Schlunzen, K.H., Carissimo, B., 2011. The COST 732 Best Practice Guideline  
724 for CFD simulation of flows in the urban environment: a summary. *International Journal of*  
725 *Environment and Pollution* 44 (1), 419-427.
- 726 Germano, M., Piomelli, U., Moin, P., Cabot, W.H., 1991. A dynamic subgrid-scale eddy viscosity  
727 model. *Physics of Fluids A: Fluid Dynamics* 3 1760.
- 728 Ghiaus, C., Allard, F., 2005. *Natural ventilation in the urban environment: assessment and design*.  
729 Earthscan,
- 730 Goulet, J.-A., Coutu, S., Smith, I.F.C., 2013. Model falsification diagnosis and sensor placement for  
731 leak detection in pressurized pipe networks. *Advanced Engineering Informatics* 27 (2), 261-  
732 269.



733 Goulet, J.-A., Michel, C., Smith, I.F.C., 2012. Hybrid probabilities and error-domain structural  
734 identification using ambient vibration monitoring. *Mechanical Systems and Signal Processing*  
735 37 (1), 199-212.

736 Goulet, J.-A., Smith, I.F., 2011. Extended uniform distribution accounting for uncertainty of  
737 uncertainty. In: *International Conference on Vulnerability and Risk Analysis and*  
738 *Management/Fifth International Symposium on Uncertainty Modeling and Analysis*,  
739 Maryland, USA, pp. 78-85.

740 Guo, L., Maghirang, R.G., 2012. Numerical simulation of airflow and particle collection by vegetative  
741 barriers.

742 Jiang, Y., Chen, Q., 2002. Effect of fluctuating wind direction on cross natural ventilation in buildings  
743 from large eddy simulation. *Building and Environment* 37 (4), 379-386.

744 Launder, B.E., Spalding, D., 1974. The numerical computation of turbulent flows. *Computer methods*  
745 *in applied mechanics and engineering* 3 (2), 269-289.

746 Lim, H.C., Thomas, T., Castro, I.P., 2009. Flow around a cube in a turbulent boundary layer: LES and  
747 experiment. *Journal of Wind Engineering and Industrial Aerodynamics* 97 (2), 96-109.

748 Meng, Y., Hibi, K., 1998. Turbulent measurements of the flow field around a high-rise building. *Journal*  
749 *of Wind Engineering* 76 55-64.

750 Mochida, A., Lun, I.Y., 2008. Prediction of wind environment and thermal comfort at pedestrian level  
751 in urban area. *Journal of Wind Engineering and Industrial Aerodynamics* 96 (10), 1498-1527.

752 Moore, R.E., 1966. *Interval analysis*. Prentice-Hall Englewood Cliffs,

753 Murakami, S., 2006. Environmental design of outdoor climate based on CFD. *Fluid dynamics research*  
754 38 (2), 108-126.

755 Patankar, S.V., Spalding, D.B., 1972. A calculation procedure for heat, mass and momentum transfer  
756 in three-dimensional parabolic flows. *International Journal of Heat and Mass Transfer* 15 (10),  
757 1787-1806.

758 Ramponi, R., Blocken, B., 2012. CFD simulation of cross-ventilation flow for different isolated  
759 building configurations: validation with wind tunnel measurements and analysis of physical  
760 and numerical diffusion effects. *Journal of Wind Engineering and Industrial Aerodynamics* 104  
761 408-418.

762 Schatzmann, M., Leitl, B., 2011. Issues with validation of urban flow and dispersion CFD models.  
763 *Journal of Wind Engineering and Industrial Aerodynamics* 99 (4), 169-186.

764 Shih, T.-H., Liou, W.W., Shabbir, A., Yang, Z., Zhu, J., 1995. A new  $k-\epsilon$  eddy viscosity model for high  
765 reynolds number turbulent flows. *Computers & Fluids* 24 (3), 227-238.

766 Tominaga, Y., Mochida, A., Murakami, S., Sawaki, S., 2008a. Comparison of various revised  $k-\epsilon$   
767 models and LES applied to flow around a high-rise building model with 1: 1: 2 shape placed  
768 within the surface boundary layer. *Journal of Wind Engineering and Industrial Aerodynamics*  
769 96 (4), 389-411.

770 Tominaga, Y., Mochida, A., Yoshie, R., Kataoka, H., Nozu, T., Yoshikawa, M., Shirasawa, T., 2008b.  
771 AIJ guidelines for practical applications of CFD to pedestrian wind environment around  
772 buildings. *Journal of Wind Engineering and Industrial Aerodynamics* 96 (10), 1749-1761.

773 Van Hooff, T., Blocken, B., 2010. Coupled urban wind flow and indoor natural ventilation modelling  
774 on a high-resolution grid: A case study for the Amsterdam ArenA stadium. *Environmental*  
775 *Modelling & Software* 25 (1), 51-65.

776 Wieringa, J., 1992. Updating the Davenport roughness classification. *Journal of Wind Engineering and*  
777 *Industrial Aerodynamics* 41 (1), 357-368.

778 Yoshie, R., Mochida, A., Tominaga, Y., Kataoka, H., Harimoto, K., Nozu, T., Shirasawa, T., 2007.  
779 Cooperative project for CFD prediction of pedestrian wind environment in the Architectural  
780 Institute of Japan. *Journal of Wind Engineering and Industrial Aerodynamics* 95 (9), 1551-  
781 1578.

782 This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0  
783 International License

