Computational Urban Flow Predictions with Bayesian Inference: Validation with field data

Jorge Sousa*

jsousa@stanford.edu

Catherine Gorlé

gorle@stanford.edu

[a]*Stanford University, 473 Via Ortega, Stanford, 94305, USA*

Abstract

Urban areas are projected to expand at a rapid pace. In the context of supporting sustainable design of cities and buildings, computational fluid dynamics (CFD) can be used to provide detailed information on the urban flow field. However, the complexity and natural variability of atmospheric boundary layer flows can limit the predictive performance of CFD. In this paper, we present a validation study for a Bayesian inference method that estimates the inflow boundary conditions for Reynolds-averaged Navier-Stokes (RANS) simulations of urban flow by assimilating data from urban sensor measurements. The method employs the ensemble Kalman filter to iteratively estimate the probability density functions of the incoming wind and improve the subsequent RANS prediction. The measurements used in this study were obtained during a full-scale experimental campaign on Stanfords campus. Six sonic anemometers were deployed at roof and pedestrian level; a

*Corresponding author. Currently at: VITO, Boeretang 200, Mol, Belgium

subset of the sensors was used for data assimilation while the remaining ones were used for validation. The accuracy of the proposed inference method is compared to the conventional approach that defines the boundary conditions based on weather station data. The hit rates increased by a factor of two when using the inference method, and the predicted mean values were ∼20% more likely to be within the 95% confidence interval of the experimental data. An analysis of the impact of the number of sensors and their location indicates that the assimilation approach can consistently improve the predictions, as long as the inlet flow properties are identifiable from the sensor measurements.
Keywords: Urban flow physics, Bayesian Inference, Computational Predictions, Field Experiments, Data assimilation, Ensemble Kalman filter

1. Introduction

Cities accommodate more than 50% of the world population, and urbanized areas are projected to expand at a rapid pace [1]. The years to come present a window of opportunity to design urban environments that are optimized for sustainability and resiliency. The analysis and prediction of wind flow within urban canopies is an important aspect of this process: information on local wind characteristics can help to improve air quality [2], ensure pedestrian wind comfort [3], promote natural ventilation of buildings [4] and exploit available wind power [5]. In this regard, Computational Fluid Dynamics (CFD) offers a powerful tool [6], but the accuracy of these predictions remains a concern.

Several authors [7, 8, 9, 10] have shown that full scale urban measurements

present significant variability that cannot be encompassed by deterministic simulations, even when using higher fidelity CFD tools such as large-eddy simulations (LES) [11]. Recently, a number of studies have demonstrated that this variability can be quantified using uncertainty quantification (UQ) frameworks, both in Reynolds-averaged Navier-Stokes (RANS) simulations [12, 13, 14], and in LES [15]. These studies consider the inlet flow properties measured by a single sensor to be random variables. The inflow uncertainty is then propagated through several realizations of a surrogate model to compute quantities of interest (local wind properties or pollutant concentrations) with confidence intervals. A limitation of this approach is that these measurements, which are usually obtained from a weather station, are not typically available at or near the inlet plane of the computational domain; they might be remote, or on the downwind side of the urban canopy for the predominant incoming wind directions [8]. As a result, differences between the flow recorded at the weather station and the actual flow at the inlet boundary of the computational domain can result in inaccurate definitions of the random inflow parameters based on sensor data.

The objective of the present study is to validate RANS results, obtained with a previously proposed Bayesian inference technique to specify the inflow boundary conditions [16], with field measurement data. The method does not require input from a weather station, instead it assimilates data from urban wind sensors located within the computational domain to infer probability distributions for the inflow parameters. An ensemble Kalman filter is used for the inference, while the forward model is a spectral expansion surrogate model derived from 63 RANS simulations. Similar techniques

have been applied to identify the location of pollutant sources [17, 18, 10], to improve turbulence modeling [19], parameter estimation [20] and to infer inflow directions or wind profiles [21, 22]. These studies have shown that data assimilation approaches offer significant opportunities to improve the prediction of flow phenomena for a variety of applications.

To validate the proposed approach in a realistic urban environment we performed a field experiment on Stanford's campus. Fig. 1 shows some of the modeled buildings, along with the location of the nearest weather station and the wind sensors deployed within the urban area during the experiment. The Bayesian inference relies on input from the sensors in the assimilation network (AN); the accuracy of the resulting wind flow predictions is assessed with a set of independent sensors in the validation network (VN). The predictive capability of the proposed Bayesian inference approach will be compared to the traditional approach of using the nearest weather station to define deterministic inflow boundary conditions, and the influence of the number of sensors will be analyzed.

The remainder of the paper is organized in three parts. In section 2 we outline the experimental campaign. Subsequently, we describe the Bayesian inference methodology, along with the CFD solver and the surrogate model in section 3. Finally, we compare the accuracy of the two different approaches to define the boundary conditions and quantify the improvements associated with the proposed Bayesian methodology in section 4. The conclusions of the study are summarized in section 5.

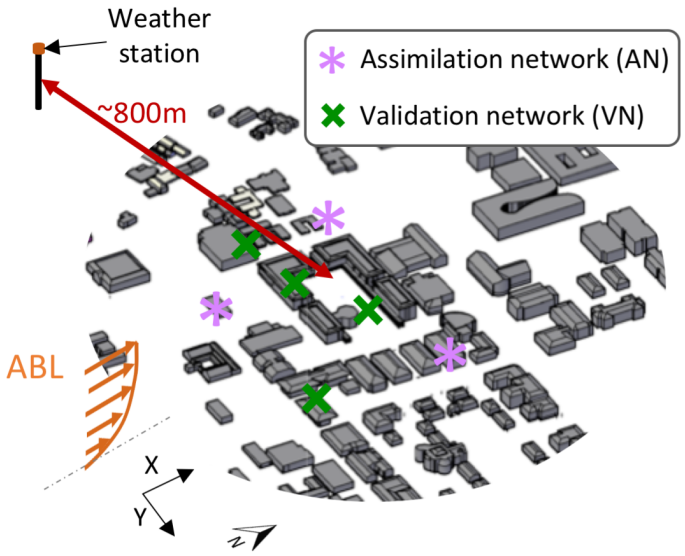

Figure 1: The area of interest on Stanford's campus, along with the sensors in the assimilation and validation networks

2. Field campaign on Stanford's campus

The experimental campaign was centered on the Science and Engineering Quad (SEQ, Lat/Long: 37.438/-122.175) of Stanford University, depicted in Fig. 2 a. The SEQ is a pedestrian zone surrounded on the south, west, and north side by the Huang, Y2E2, Shiram, and Spilker buildings, each about 20m tall. On the east side, the buildings are about 10m tall.

Six sonic anemometers, manufactured by Gill Instruments and distributed by Campbell Scientific, were deployed to measure the wind speed and direction with a stated accuracy of $\pm 2\%$ and $\pm 3^o$ respectively [23]. We evaluated the accuracy of the sensors in a wind tunnel instrumented with hot-wires and pitot tubes, measuring, on average, discrepancies below 0.2m/s for a range of velocities between 1 and 6m/s. Fig. 2 b. depicts one of the sensors de-

ployed at pedestrian level. Each station is composed of a sonic anemometer, a tripod, and a battery. Three of the sensors had a 7Ah battery and were connected to an electric outlet; the remaining three were powered by a 20W solar panel and a 12Ah battery for deployment flexibility. Finally, wireless data loggers (model CR300 from Campbell Scientific [24]) were connected to the campus wireless network for continuous data acquisition and status monitoring.

The sensor locations are indicated in Fig. 2 and summarized in Table 1. Their positioning was based on a previous design of experiments study [16], where we concluded that roof top sensors were optimal for data assimilation. Following these conclusions, we deployed the sensors in both ideal and non-ideal locations: three sonic anemometers (3, 5 and 6) were placed at roof level; one sonic anemometer (1) was placed on a second-floor terrace previously determined as non-ideal [16]; the last two sensors (2 and 4) were deployed at pedestrian level. In addition to these six sensors, we have access to hourly-averaged wind speeds and directions recorded at a weather station located at 10m above ground level, about 800m east of the center of the SEQ [25].

Figure 2: a) Locations of the sonic anemometers within the SEQ b) Example of an anemometer deployed during the experimental campaign.

Station	Level	GPS location	Sensor height [m]	Tripod height [m]
1	Terrace	37.428069/-122.174804	7.1	2.1
2	Pedestrian	37.428219/-122.173429	3.6	3.6
3	Roof	37.428223/-122.174186	18.2	3.2
4	Pedestrian	37.428473/-122.174415	2.7	2.7
5	Roof	37.428194/-122.176417	16	2.0
6	Roof	37.427175/-122.173329	20.2	2.2

Table 1: Summary of the locations and tripods heights of each of the 6 sonic anemometers.

In the present study we consider data recorded during three consecutive days, from the 10^{th} to the 12^{th} of October 2017. Fig. 3 presents the wind roses measured during the 72 hours of the measurement campaign at (a) the weather station, and (b) the different sonic anemometers. The size of each angular section in the wind roses represents the probability of occurrence of that direction, while radial sectors represent the probability of a certain velocity. One can observe the importance of natural wind variability in both the wind velocity and direction during the experiment. From the weather station and the roof-top sensors two predominant wind directions and velocities can be identified: North-West during the day, typically with higher velocities and variability, and South-West during the night, with lower velocities and lower wind direction variability.

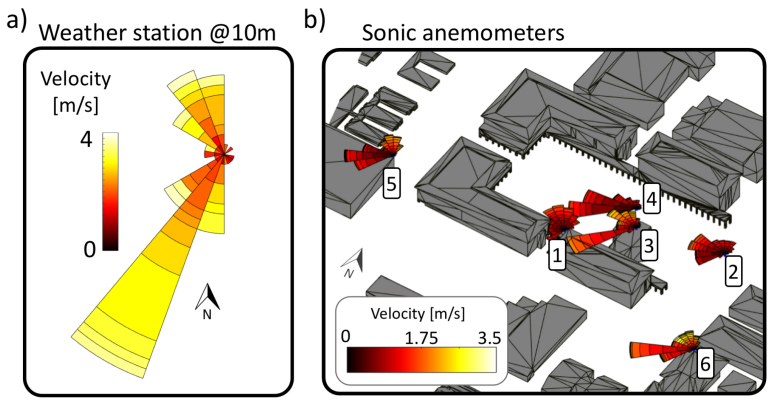

Figure 3: a) Wind rose measured by the weather station, during the 72 hours of the measurement campaign (October 10-12, 2017). b) Wind roses from each of the sonic anemometers deployed in the SEQ during the same time period.

3. Data Assimilation methodology

Fig. 4 depicts a flow chart of the Bayesian methodology. Initially, in step 1), we have no knowledge of the inflow parameters (inlet flow angle and inlet flow velocity) and assume a uniform distribution for both. In step 2) we propagate these distributions through the forward model, which is introduced in detail in section 3.1. Subsequently in step 3, we assimilate the experimental data measured by the sensors in the assimilation network. With the statistical properties of both the predicted and measured flow conditions in the assimilation network, we use the inverse ensemble Kalman filter [26, 27] to update the posterior distribution of the inlet parameters in step 4). The details of the inverse method are presented in section 3.2. Once convergence is achieved, we evaluate the accuracy of the predictions in step 5) by comparing the numerical and experimental data at the sensors in the validation network.

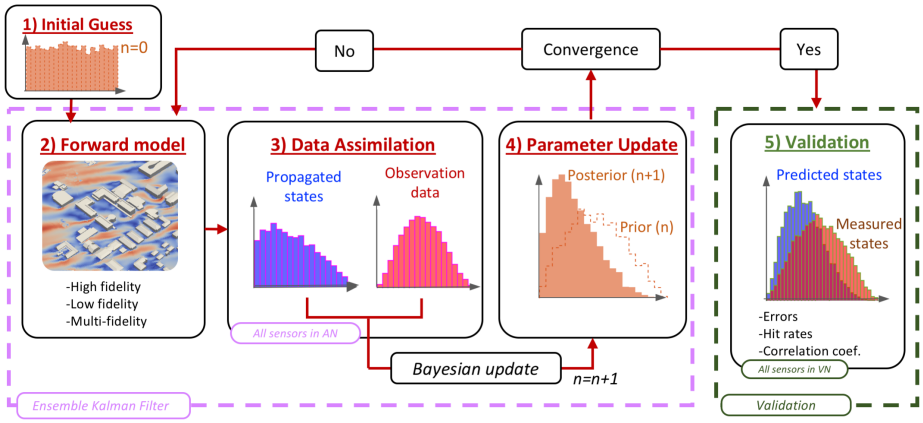

Figure 4: Flow chart of the iterative Bayesian inference method using the sensors from the assimilation network, and the subsequent validation process using the sensors from the validation network

3.1. Forward Model

The ensemble Kalman filter used for the Bayesian update requires a large number of forward model evaluations. Hence, to limit the computational cost, we used a surrogate model constructed from an initial set of RANS simulations. In the following section we first list the details of the RANS simulations and subsequently discuss the construction of the surrogate model.

3.1.1. Reynolds-averaged Navier-Stokes simulations

The computational set up of the RANS simulations is similar to the one presented in [16]. The RANS equations are solved with the OpenFoam libraries [28], and turbulence closure is obtained with the standard $k-\epsilon$ model [29]. Following the guideline that all buildings within a distance of 10 times the height of the buildings in the region of interest should be included [30], the computational domain comprises all the buildings within an 800m radius

from the Science and Engineering quad (SEQ). The domain is 2.6 by 2.6 km in the horizontal directions, and the top boundary is placed at $10H_{max}$ where H_{max}=75m corresponds to the highest building in the domain. The lateral boundaries are placed at least $12H_{max}$ away from the closest building to avoid interactions with the wakes and improve numerical convergence. This also ensured a blockage ratio below 3% as recommended by Franke et al. [30] and Tominaga et al. [31].

The computational grid is generated with snappyHexMesh [28], and consists of 22.5 million cells with a progressive grid refinement towards the SEQ to ensure a 1m meter resolution in the horizontal direction and a 0.5 m resolution in the vertical direction. A grid-sensitivity analysis performed with two additional grids, a coarser mesh with 15 million cells and a refined one with 48 million cells, demonstrated small differences between the three cases [16].

The inlet boundary condition prescribes neutral ABL profiles [32]:

$$U = \frac{u_*}{\kappa} log\Big(\frac{z+z_0}{z_0}\Big)\,, \quad k = \frac{u_*^2}{\sqrt{C_\mu}}\,, \quad \epsilon = \frac{u_*^3}{\kappa(z+z_0)}\,, \tag{1}$$

where U, k and ϵ represent the velocity, turbulence kinetic energy and dissipation rate. The height is defined by z, z_0 represents the roughness length, and u_* is the friction velocity. A spatially uniform roughness length equal to 0.3m is specified, since the campus is located in suburban terrain [33]. The von Karman constant κ is set to 0.41, and the turbulence model constant C_μ is 0.09. To ensure horizontal homogeneity of the inlet profiles [34, 35], a wall function based on the ABL roughness length [32] is applied along the ground. On the building walls, a standard smooth wall function is used.

The main difference between the simulations described in [16] and the current study is that the effect of vegetation is included in the model. To

account for the presence of trees, a drag term, S_{tree}, was added to the momentum equations [36, 37, 38]:

$$S_{tree,i} = -\frac{1}{2}\rho|U|\left(c_d LAD\right)U_i, \tag{2}$$

where c_d is the drag coefficient of a single leaf, and LAD (m^2/m^3) is the leaf area density representing the fraction of leaf area surface within the volume of the tree. A c_d value of 0.25 was selected [39, 37], and the LAD value of 0.8 was determined by computing an integral average of the empirical profile along the crown height obtained by Lalic and Mihailovic [40] for oak trees, which is the predominant tree type on campus.

Since the source term should only be active in locations where trees are present, the center of each tree and its geometrical properties (diameter D, trunk height h_{trunk}, and crown height h_{crown}) have to be identified in the model. The impact of the trunk is considered negligible, and each crown is assumed to have a cylindrical shape above the trunk height $h_{trunk} = 0.4h_{crown}$ [40]. The diameter and crown height of the trees inside the SEQ were measured individually and manually imposed in the computational domain. To avoid time-consuming individual measurements of the large number of trees on the remaining parts of the campus, an in-house algorithm that estimates the geometry of each tree and maps it into the numerical domain was developed. As shown in Fig. 5, the algorithm first loads the aerial image in Matlab and the user identifies the center of each tree graphically. Afterwards, for each tree center, the algorithm considers only the image data within a 35m radius. This radius is reduced iteratively and we compute the ratio of the standard deviation of the grey scale within the circle and its mean value.

The tree radius is identified as the radius for which the minimum ratio is achieved. Subsequently, assuming a relation between crown diameter and tree height [41], the crown height is estimated using a linear fit of data from 30 trees on campus as shown in the center plot of Fig. 5. This geometrical information (D, h_{trunk}, h_{crown}), is used to identify the computational cells that overlap with a tree and will have the source term activated. As shown in the right plot of Fig. 5, this procedure was used to identify and map 550 trees on campus.

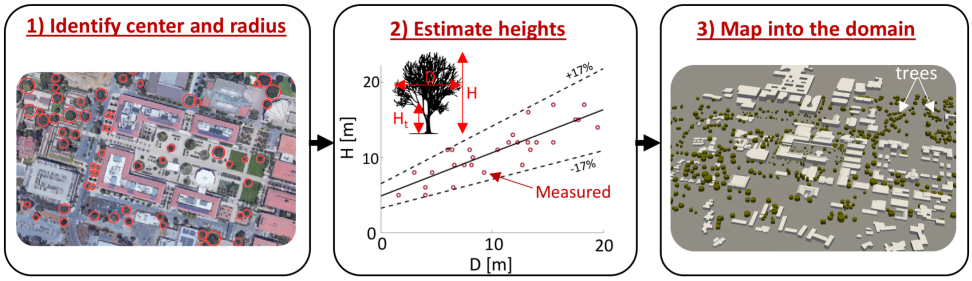

Figure 5: Algorithm to (1) identify trees from aerial images, (2) estimate the height of each tree based on its crown diameter, and (3) map the trees into the numerical domain

3.1.2. Polynomial chaos expansion (PCE) surrogate model

To limit the required computational cost, each realization of the forward model in the Bayesian inference algorithm uses a surrogate model $\hat{\boldsymbol{G}}(\boldsymbol{\alpha})$ for the RANS model $\boldsymbol{G}(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is the vector of random inflow parameters. A polynomial chaos expansion (PCE) [42] is constructed for each quantity of interest using a non-intrusive spectral method available in DAKOTA [43]:

$$\hat{\boldsymbol{G}}(\boldsymbol{\alpha}) = \sum_{k=0}^{P} \beta_k \Psi_k(\boldsymbol{\alpha}), \tag{3}$$

where β_k represents the polynomial coefficients and Ψ_k the basis functions of multivariate orthogonal polynomials. We use the Gauss-Patterson quadrature rule [44] to compute the coefficients from 63 simulations. This results in a Polynomial expansion truncated at P=47 [43]. Once the coefficients are determined, the PCE can be constructed and used as a surrogate model that can be sampled for all wind directions.

In the present study, the quantities of interest are the local flow angle, θ_{local}, and the local flow velocity, U_{local}, at the locations of the sensors in the validation network. For the high Reynolds numbers under consideration, the local flow angle and the reduced velocity $U_r = U_{local}(z)/U_{inlet}(z)$ are only a function of the inlet flow angle. Hence, the flow properties of interest can be computed using surrogate models that are only a function of the inlet flow angle:

$$U_{local} = \hat{\boldsymbol{G}}_{U_r}(\theta_{inlet})U_{inlet} \tag{4}$$

$$\theta_{local} = \hat{\boldsymbol{G}}_{\theta}(\theta_{inlet}) \tag{5}$$

Hence, we only ran the RANS model for different inlet flow angles, performing 63 simulations in total. Each RANS simulation was run on 36 cores (2.3-GHz Intel Xeon E5-2697V4), achieving full convergence after about 8 hours. All 63 simulations were run simultaneously on NCAR's Cheyenne supercomputer.

We quantified the difference between the RANS and the surrogate model to estimate the associated error. Fig. 6 depicts the probability distributions of the errors for the local flow angles and velocities at all sensor locations. The plots also show the Gaussian distributions fitted to estimate the error variance that will be used in the ensemble Kalman filter.

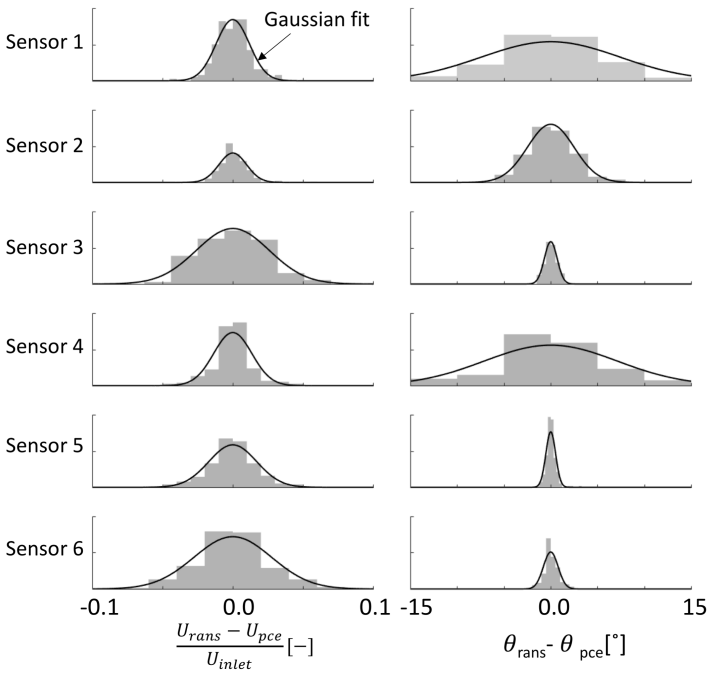

Figure 6: Estimates of the model error associated with the surrogate model based on the differences between the polynomial expansions and the RANS simulations

3.2. Bayesian Inference

The objective is to estimate the probability distributions of the two unknown inflow parameters, the inlet flow direction and velocity ($\boldsymbol{\alpha} = [\boldsymbol{\alpha}_\theta, \boldsymbol{\alpha}_U]$), by assimilating discrete observations of the physical state ($\boldsymbol{\psi}$) on the SEQ. In general, this corresponds to an optimization problem where we want to find the $\boldsymbol{\alpha}$ that minimizes the difference between the predicted and observed states. In Bayesian inference, this is equivalent to a maximum likelihood estimate [45] based on Bayes's theorem:

$$p(\boldsymbol{\alpha}|\boldsymbol{\psi}) = \frac{p(\boldsymbol{\psi}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})}{p(\boldsymbol{\psi})}, \tag{6}$$

where the posterior probability $p(\boldsymbol{\alpha}|\boldsymbol{\psi})$ is proportional to the prior p($\boldsymbol{\alpha}$) and the likelihood $p(\boldsymbol{\psi}|\boldsymbol{\alpha})$. The prior represents all prior knowledge of $\boldsymbol{\alpha}$; the likelihood represents the probability of the predicted state given the parameters $\boldsymbol{\alpha}$. Several approaches can be used to solve Eq. 6; in this work we use the ensemble Kalman filter, which is an approximate Bayesian inference technique. The method's advantages lie in its straightforward implementation, possible parallelization, and its gradient-free approach. A potential drawback is that the predictive performance of the formulated inverse problem may deteriorate if the assumption that the input parameters have a Gaussian distribution does not hold. For the present case, an analysis considering artificially generated non-Gaussian inputs (similar to the analysis presented in [16]), indicated that non-Guassian distributions can be recovered by the iterative process. Further analysis of the performance of the ensemble Kalman filter, including a comparison of the efficiency and accuracy of different approaches for solving the inverse problem, would be of interest for future research.

The ensemble Kalman Filter infers the statistical properties of unknown parameters $\boldsymbol{\alpha}$ that are inputs to a forward model $\hat{\boldsymbol{G}}(\boldsymbol{\alpha})$, based on available observations at discrete locations within the domain of interest [27]. Consider an ensemble of size J of the parameters, $\boldsymbol{\alpha}_{j=1,\ldots,J}$, characterized by mean values $\bar{\boldsymbol{\alpha}}$ and additive stochastic terms $\boldsymbol{\alpha}'_j$ with zero mean and known covariances $\boldsymbol{C}_{\alpha\alpha}$:

$$\boldsymbol{\alpha}_j = \bar{\boldsymbol{\alpha}} + \boldsymbol{\alpha}'_j \tag{7}$$

We can then construct the field vector $\boldsymbol{\psi}_j(\boldsymbol{x})$ by running J realizations

of the physical model $\hat{\boldsymbol{G}}(\boldsymbol{\alpha})$:

$$\boldsymbol{\psi}_j = \hat{\boldsymbol{G}}(\boldsymbol{\alpha}_j) + \boldsymbol{q}_j \tag{8}$$

where $\hat{\boldsymbol{G}}(\boldsymbol{\alpha})$ is our non-linear model, and $\boldsymbol{q}(\boldsymbol{x})$ is an additive stochastic term that is assumed to have a Gaussian distribution with zero mean and covariance $\boldsymbol{C}_{qq}$. $\boldsymbol{q}(\boldsymbol{x})$ represents modeling errors, which are estimated from the error between the PCE and the RANS model as shown in Fig. 6. The additional error associated with the RANS model is not considered in the current study; estimating and including this error, e.g using methods for RANS turbulence model-form uncertainty quantification [46, 47], is a subject of ongoing work.

The objective of the inverse ensemble Kalman filter is to find the statistical value of $\boldsymbol{\alpha}$ that minimizes the difference between the M experimental observations $\boldsymbol{y}_j$ and the field vector values at these locations $\hat{\boldsymbol{d}}_j$:

$$\boldsymbol{y}_j = \bar{\boldsymbol{y}} + \boldsymbol{\epsilon}_j, \tag{9}$$

$$\hat{\boldsymbol{d}}_j = \mathcal{M}[\boldsymbol{\psi}_j(x)]. \tag{10}$$

Here, $\boldsymbol{\epsilon}$ represents the variability of the measurements with a known covariance $\boldsymbol{C}_{\epsilon\epsilon}$, and $\mathcal{M}$ represents the measurement matrix used to extract the simulated field vector $\boldsymbol{\psi}_j$ at the locations of the observations. The measurement matrix has null values in the locations where measurements are not available.

The algorithm used for this parameter estimation is summarized in Algorithm 1. After providing an initial guess, an iterative process first predicts

$\hat{\boldsymbol{d}}_j$, and subsequently updates the parameter estimate using Eq. 13. This update equation computes the posterior parameter and field vector $(n+1)$ based on a weighted combination between the prior (n) and the error between the simulated $\hat{\boldsymbol{d}}^n$ and observed states $\boldsymbol{y}$, where the weight is given by the Kalman gain. When the initial guess is far from the true solution, the covariance matrices are poorly approximated by the ensemble. The iterative approach was proposed by Iglesias et al. [26] to address this issue. The iterative process always performs the specified maximum number of iterations n_{max}=10; typically the solution converges after 4 iterations. Subsequently, the validation process is initiated, comparing the solution to the measurements from the sensors in the validation network. When using the surrogate model this process takes approximately 100s, a single forward propagation takes about 16s in an Intel Xenon CPU E5-2609 v3 @ 1.90GHz (using 12 cores in parallel). Further details on the algorithm can be found in [27, 26] and an application to turbulence modeling has been presented in [19].

4. Results and discussion

In this section we compare the experimental data with (1) the flow predictions based on the boundary conditions from the weather station and (2) with the flow predictions based on the proposed Bayesian inference approach. The two methods can be summarized as follows:

1. Boundary conditions from the weather station: Considering a single hour of the measurement campaign, the hourly-averaged data provided by the weather station is used to run the surrogate model, computing

Algorithm 1 Ensemble Kalman filter for parameter estimation

ASSIMILATION DATA: Select the period and length of interest and draw J random samples from each sensor in the assimilation network: $\boldsymbol{y}_{1,...,J}$

INITIAL GUESS: Specify the prior distribution of the unknown parameter $p(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\alpha_1^0, ..., \alpha_J^0)$

ITERATIVE PROCEDURE:

for $n = 0$ to n_{max} **do**

1. **Prediction step**

1.1 Draw J realizations from the forward model and add the model error for each location x:

$$\boldsymbol{\psi}_j(x) = \hat{\boldsymbol{G}}(\boldsymbol{\alpha}_j) + \boldsymbol{q}_j(x) \tag{11}$$

1.2 Extract $\boldsymbol{\psi}_j(x)$ at the observation location:

$$\hat{\boldsymbol{d}}_j = \mathcal{M}[\boldsymbol{\psi}_j(x)] \tag{12}$$

1.3 Compute the covariance matrices $\boldsymbol{C}_{\alpha\boldsymbol{d}}$, $\boldsymbol{C}_{\psi\boldsymbol{d}}$, $\boldsymbol{C}_{dd}$:

$\boldsymbol{C}_{\boldsymbol{\alpha d}} = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{\alpha}_j(\boldsymbol{d}_j)^T - \bar{\boldsymbol{\alpha}}\bar{\boldsymbol{d}}^T$, where, $\bar{\boldsymbol{\alpha}} = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{\alpha}_j$, $\bar{\boldsymbol{d}} = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{d}_j$

$\boldsymbol{C}_{\boldsymbol{\psi d}} = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{\psi}_j(\boldsymbol{d}_j)^T - \bar{\boldsymbol{\psi}}\bar{\boldsymbol{d}}^T$, where, $\bar{\boldsymbol{\psi}} = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{\psi}_j$

$\boldsymbol{C}_{\boldsymbol{dd}} = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{d}_j(\boldsymbol{d}_j)^T - \bar{\boldsymbol{d}}\bar{\boldsymbol{d}}^T$

2. **Update step**

Update the parameter and flow field vectors:

$$\begin{pmatrix} \boldsymbol{\alpha}_j^{n+1} \\ \boldsymbol{\psi}_j^{n+1} \\ \hat{\boldsymbol{d}}_j^{n+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_j^{n} \\ \boldsymbol{\psi}_j^{n} \\ \hat{\boldsymbol{d}}_j^{n} \end{pmatrix} + \begin{pmatrix} \boldsymbol{C}_{\alpha d} \\ \boldsymbol{C}_{\psi d} \\ \boldsymbol{C}_{dd} \end{pmatrix} (\boldsymbol{C}_{dd} + \boldsymbol{C}_{\epsilon\epsilon})^{-1}(\boldsymbol{y}_j - \hat{\boldsymbol{d}}_j^n) \tag{13}$$

end for

VALIDATION STEP: Compare the flow field ψ with the observations from the validation network.

the deterministic flow field over the SEQ and providing the wind velocity at the sensor locations. This process is repeated for each of the 72 hours of the campaign.

2. Boundary conditions from Bayesian inference: Considering a single hour of the measurement campaign, we generate an initial ensemble of the inflow parameters with a uniform distribution and use the surrogate model to compute the corresponding wind velocities at the sensors in the assimilation network. The iterative ensemble Kalman filter is then used to infer the probability distributions for the inlet parameters (flow angle and velocity) and provide the stochastic flow field prediction. This process is also repeated for each of the 72 hours of the experimental campaign.

We will start by comparing the flow field predictions at two distinct time instances, considering one hour during the day, and one hour during the night. Subsequently, we consider the full 72 hour period and quantitatively compare statistics and compute validation metrics. Finally, we discuss the influence of the location and number of sensors used for data assimilation.

4.1. Comparison of flow field predictions at two time instances

In this section we quantitatively compare the flow field predictions obtained with the two different approaches for October 11th between 3 and 4pm, and for October 12th between 3 and 4 am. The data assimilation is performed using two sensors: sensor 2 at pedestrian level and sensor 6 at roof level. An analysis similar to that presented in Section 4.3 indicated that this combination of sensors provides optimal results. The initial guess

for the distributions of the flow angle and velocity magnitude is a uniform distribution between $[0 - 360]^o$ and $[0 - 10]$m/s, respectively. We used an ensemble size $J = 100$, following the analysis presented in [16], where we demonstrated that this value provides a sufficiently accurate estimate of the mean and standard deviation of the inlet conditions. We did not observe any deterioration in the final inferred distribution from using a uniform distribution as the initial guess; however, when a better initial guess is used, the iterative algorithm tends to converge in fewer iterations.

Fig. 7a. presents the probability density functions inferred by the Bayesian approach for the wind speed and direction, together with the corresponding hourly-averaged value measured by the weather station during the day (between 3-4pm) and the night (between 3-4am). During the day (top figure) the mean values of the inferred distributions are similar to the hourly-averaged measurement, but the Bayesian inference reveals significant variability in the inlet wind conditions. During the night (bottom figure), the mean values measured by the weather station are outside the inferred distributions, and the Bayesian inference results show a much smaller variability in the inferred conditions. The contour plots in Fig. 7 b. and c. depict the velocity ratio, i.e. the local horizontal velocity normalized by the horizontal velocity measured at the weather station at 5m height. Fig 7 b. presents the deterministic result based on the boundary conditions defined by the hourly-averaged weather station while 7 c. presents the mean result based on the inferred pdfs for the inlet conditions. Fig. 7 d. shows the difference between both cases. During the day time there are some locations, such as shear layers and narrow street canyons, where the velocity ratios differ by about 30%. In

the night time case, the overall difference in the flow field is more significant, primarily due to the difference in the incoming flow angle.

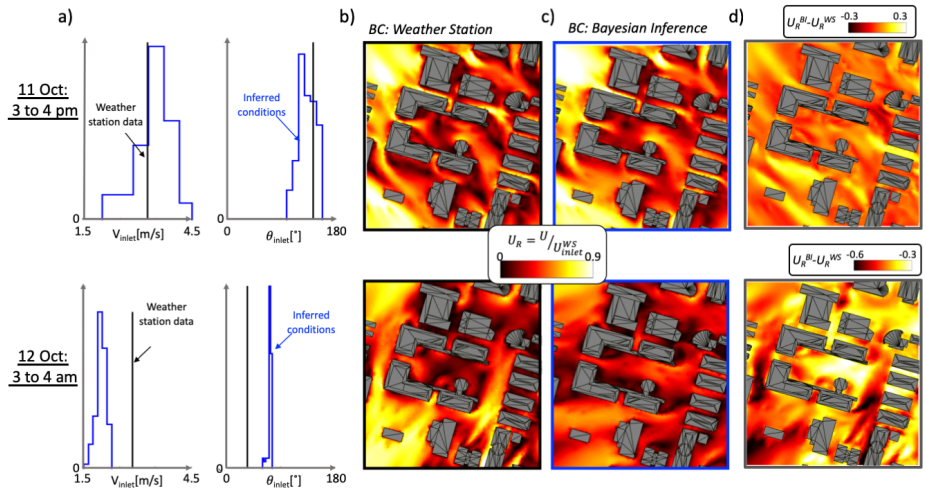

Figure 7: a) Probability density functions of the inlet velocity and flow angle measured by the weather station and estimated using Bayesian inference; contour plots of the velocity ratio based on b) weather station inflow conditions and c) Bayesian inference inflow conditions; d) the difference between contour plots b) and c).

These preliminary observations indicate that in certain wind conditions the mean values measured by the weather station are very similar to the inferred ones, while in other conditions significant discrepancies can occur.

4.2. Comparison and validation of the flow prediction over 3 days

To compare the performance of each methodology for a variety of wind conditions we first compare the statistics of the predictions over all 72 hours of the measurement campaign. Subsequently we evaluate each hour individually to identify time periods during which the uncertainties are magnified.

4.2.1. Comparison of statistics

Figure 8 depicts three wind roses, representing a) the boundary conditions measured by the weather station, and b) the inferred inlet conditions. To perform the inference analysis, the ensemble Kalman filter was ran 72 times for each hour of the experimental campaign. The ensembles from each hour were then combined in a single wind-rose. We present wind roses for the ensemble averaged hourly results to facilitate comparison to the weather station data, and for all 100 samples in the ensemble used for each hour. From this analysis we can identify a higher mismatch in the flow angle for south-west wind directions: in these conditions the weather station consistently recorded a higher wind velocity.

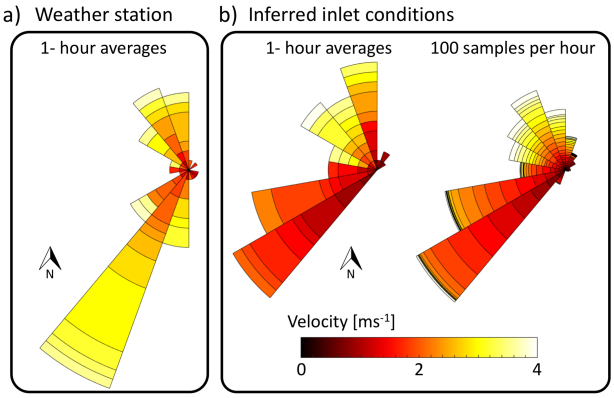

Figure 8: a) Wind rose measured by the weather station during the experimental campaign; b) Inferred wind rose: left - represents the hourly average data; right - shows the data from the 100 samples per hour used in for Bayesian inference. [10 Oct: 00:00 - 12 Oct: 24:00]

To investigate the accuracy of both approaches, we compare the predicted results to the experimental data. Typically in wind engineering, one is in-

terested in predicting the probability density functions (PDF) of the wind velocity to evaluate for example, pedestrian comfort or wind power availability. Hence, Fig. 9 compares the experimental distributions to the ones predicted by the Bayesian inference (with 100 samples per hour, over 72 hours) and with the ones based on the weather station boundary conditions (hourly-averaged data over 72 hours). Overall visual comparison of these distributions indicates that the predictions with the Bayesian approach reproduce the statistical characteristics of the experimental data with a higher accuracy than the conventional approach. Some differences occur, for example at sensor 1 there is a consistent underestimation of the velocity magnitude and a larger variability in the flow angle. At sensor 5 there is a slight overestimation of the magnitude, but the PDF of the flow angle is accurately reproduced, showing two distinct predominant wind directions.

4.2.2. Comparison of time-series and quantitative metrics

In order to compare the performance of each methodology under different wind conditions, we consider each of the 72 hours individually. Fig. 10 plots, for each sensor, the hourly-averaged values of the velocity components (U_x and U_y) and the corresponding 95% Confidence Interval (CI) of the experimental data and the predictions obtained from Bayesian inference. Since the weather station only provides mean values, 95% CI can not be computed for this approach. The Bayesian based predictions accurately reproduce the mean and the 95% CI of the experimental data: 84% and 77% of the mean predicted values are within the 95% CI of the experimental data for U_x and U_y, respectively. When using the weather station, these values drop to 69% and 49%. The bottom plots depict the cumulative errors in the mean values

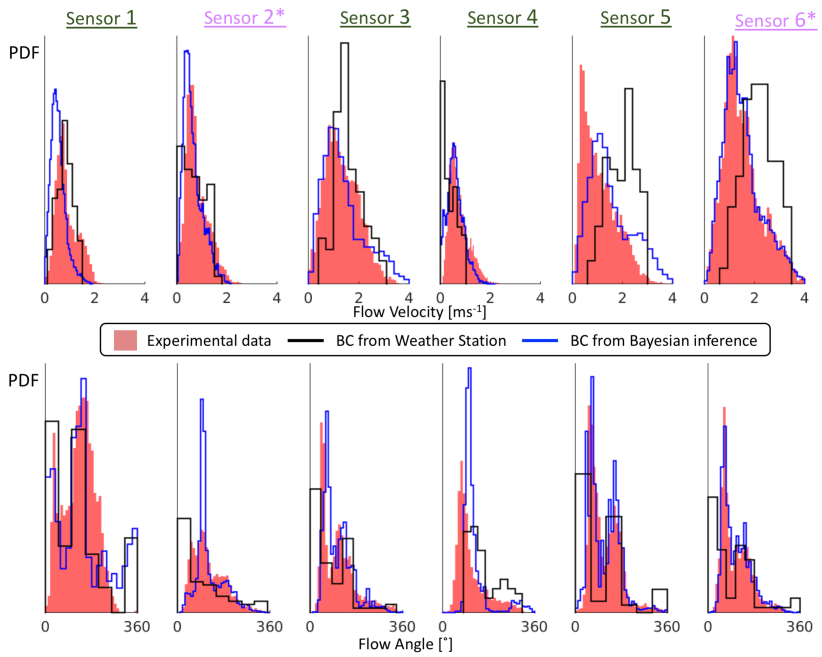

Figure 9: Experimental, weather station, and Bayesian-based predictions of the probability density functions of wind magnitude and direction at each sensor, combining all 72 hours of data. Sensors 2 and 6 where used in the assimilation network.

of the 6 sensors. During all periods of the experimental campaign, Bayesian inference outperforms the conventional approach, and the error is relatively constant over the three days. In contrast, when using the weather station data, the predicted errors are higher during the night when the flow direction is predominantly from the South-West. This corresponds to the direction for which the weather station is not on the upwind side of the urban canopy.

To perform further quantitative comparisons we combined the data from the 6 sensors over the 72 hours to compute correlation coefficients ($\rho_{e,p}$), hit rates (HR), and the percentage of hours that the predicted mean values are

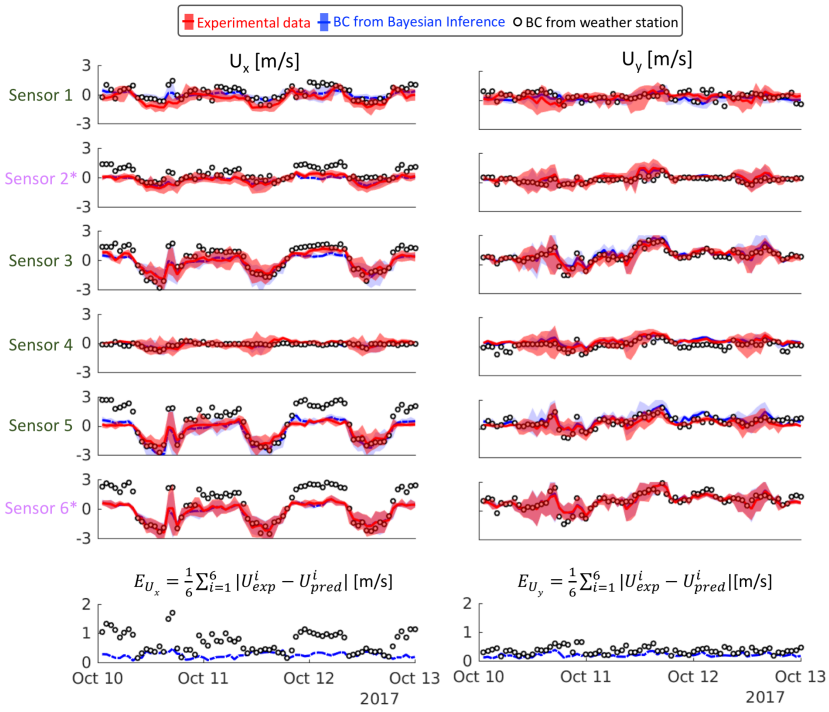

Figure 10: Time evolution of the hourly-averaged wind velocity components and their 95% confidence intervals (shaded regions) for the experimental data and the Bayesian inference based predictions, compared to the deterministic weather station predictions.

within the $95\%CI$ of the experimental data ($\bar{p}_{\in CI}$) as follows:

$$\rho_{e,p} = \frac{cov(e,p)}{\sigma_e \sigma_p} \tag{14}$$

$$HR = \frac{1}{72}\sum_{1}^{72} \begin{cases} 1, & \text{if } \frac{|e-p|}{e} < D \text{ or } |e-p| < W \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

$$\bar{p}_{\in CI} = \frac{1}{72}\sum_{1}^{72} \begin{cases} 1, & \text{if } e_{5\%CI} < \bar{p} < e_{95\%CI} \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

In these equations, e is the experimental data and p the predicted results. The metrics are computed for the two horizontal velocity components. The hit rate is calculated using the thresholds D = 0.2 for the relative error and W = 0.2m/s for the absolute error. The latter corresponds to the sensor measurement uncertainty observed during the wind tunnel calibration. The results for both cases are summarized in Table 2, showing that for all metrics the Bayesian inference approach outperforms the weather station based approach. On average the hit rates are improved by about 0.28, while the mean values of the Bayesian based predictions are about 22% more likely to be within the 95%CI interval of the experimental data.

Station	HR_{U_x}	HR_{U_y}	ρ_{U_x}	ρ_{U_y}	$p_{\in CI_{U_x}}$	$p_{\in CI_{U_y}}$
Weather station	0.23	0.37	0.83	0.73	0.69	0.49
Bayesian approach	0.54	0.64	0.89	0.89	0.84	0.77

Table 2: Comparison of hit rates, correlation coefficients, and the percentage of time the hourly-averaged prediction is within the experimental confidence interval when using the weather station or the Bayesian inference method.

To conclude this comparison, Fig. 11 presents a scatter plot of the predictions versus the experiments, where each point represents one-hour averaged data from any of the six sensors. A point on the $x = y$ line implies a perfect prediction. Fig. 11a compares the experimental results with the predictions when using the weather station, while Fig. 11b shows the result obtained with the Bayesian inference approach. In these two plots the $*$ symbol represents the results from the two sensors in the assimilation network, while the $\times$ symbol embodies the data from the 4 sensors in the validation network. When using boundary conditions based on the weather station, a consider-

able spread can be observed in the data. This corroborates the low hit rate values. In contrast, when using the Bayesian inference approach, the results lay considerable closer to the experimental data, resulting in improved hit rates. As expected, the assimilation sensors show an almost perfect match between the experimental and predicted data.

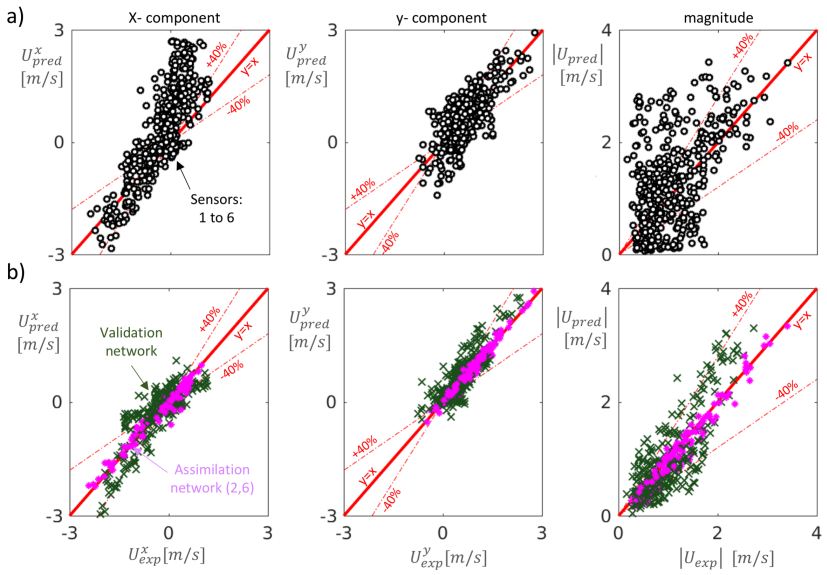

Figure 11: Scatter plots for the x- and y-velocity components and the velocity magnitude when using boundary conditions based on: a) weather station data; b) Bayesian inference. [10 Oct, 00:00 - 13 Oct, 00:00]

4.3. Influence of location and number of sensors used for assimilation

To quantify the impact of the sensor locations we present the results when using only one sensor in the assimilation network, with the remaining five sensors in the validation network. Table 3 summarizes the hit rates and correlation coefficients for each velocity component. Each row represents a different assimilation setup, with the first column indicating the assimilated

	HR_{U_x}	HR_{U_y}	ρ_{U_x}	ρ_{U_y}	$p_{\in CI_{U_x}}$	$p_{\in CI_{U_y}}$
Weather station	0.23	0.37	0.83	0.73	0.69	0.49
Assimilation sensor						
1	0.42	0.25	0.54	-0.27	0.46	0.66
2	0.39	0.46	0.80	0.84	0.68	0.67
3	0.52	0.55	0.87	0.87	0.79	0.74
4	0.38	0.49	0.59	0.83	0.78	0.63
5	0.55	0.48	0.88	0.82	0.72	0.78
6	0.53	0.61	0.89	0.86	0.86	0.77

Table 3: Comparison of correlation coefficients and hit rates, using single sensor for data assimilation compared to the results when using the weather station to define the boundary conditions.

sensor. For comparison, the top row represents the coefficients when relying on the weather station to define the inlet conditions. The results clearly show that the Bayesian inference outperforms the conventional method in all cases, except when assimilating data from sensor 1. The scatter plot for this case is shown in Fig. 12a. As discussed in [16], this sensor is located close to a narrow street canyon where the flow is guided by the buildings and inference problem becomes ill-posed. The scatter plot also identifies over-fitting, where the data from the assimilation network perfectly fits the experimental data but predictions for the validation network are inaccurate.

The assimilation test case that performs best is the one using sensor 6, with an overall improvement of about 25% relative to the traditional approach. Fig. 12b shows the corresponding scatter plots. Even if more than 57% of the data is within 20% of the experimental data, the plot shows a con-

sistent over estimation of the wind velocity. This indicates a possible bias in the numerical or experimental fields at the location of sensor 6, highlighting that ideally a multi-sensor assimilation is preferred to minimize the impact of localized sources of uncertainty.

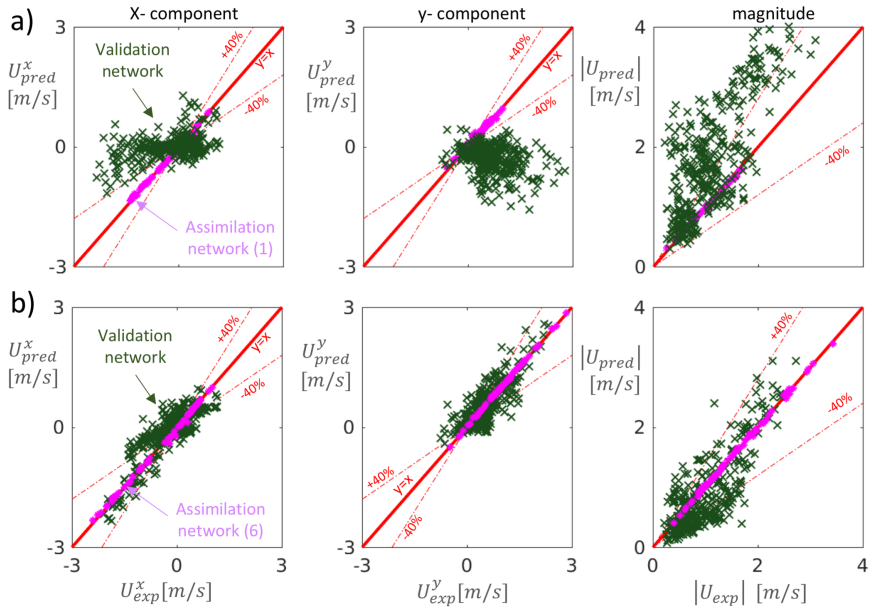

Figure 12: Scatter plots for hourly-averaged data for the value of the two velocity components and the velocity magnitude; a) the solution when performing data assimilation from sensor 1; b) the solution when performing data assimilation from sensor 6. [10 Oct,00:00 - 13 Oct, 00:00]

Finally, comparison of Fig. 12 to Fig 11 demonstrates that multi-sensor assimilation helps to improve the predictions and avoids over-fitting of the experimental data. In general, it can be concluded that optimal locations to deploy the sensors are on roof tops as well as open spaces (far from buildings and narrow street canyons) at pedestrian level. In these locations the inlet parameters are identifiable from the local flow properties, which ensures that

the inverse problem is well-posed. It also appears to be advantageous to rely on sensors at different heights; this can minimize uncertainties associated with the vertical profile of the atmospheric boundary layer. Given the set-up of the inference problem, which relies on the sensors only to define the incoming wind direction and magnitude at a reference height, it is plausible that these recommendations are generally applicable to a variety of urban morphologies; this will be further investigated in future work.

5. Conclusions

In the present work we validate a novel, computationally efficient approach to perform data-driven wind flow predictions in a full-scale urban environment. The method uses Bayesian inference to assimilate data from sensors deployed within the urban canopy and infer the probability distributions for the inflow boundary conditions. To minimize the computational time we employ an ensemble Kalman filter, using a surrogate forward model constructed from a prior set of RANS simulations.

To assess the performance of the proposed approach, we conducted a field measurement campaign on Stanford's campus during the summer of 2017. Six sonic anemometers were deployed during three consecutive days. A subset of the deployed sonic anemometers were used for data assimilation, while the remaining sensors were used for validation. This strategy allowed us to test the sensitivity to the number and location of the sensors used for data assimilation. We compared the accuracy of the predictions obtained using Bayesian inference to predictions that relied on a nearby weather station to define the boundary conditions. We show that the Bayesian inference

approach consistently outperforms the conventional method for all evaluated metrics, and that the method is reliable over a variety of wind conditions. In contrast, when using the weather station, the error is amplified for one particular wind direction. Nonetheless, the results demonstrate that the success of the Bayesian inference approach is dependent on the location of the sensors. Based on the current analysis, we recommend that sensors used for assimilation are placed at roof level or at pedestrian level in open spaces. To further improve accuracy, the assimilation should be based on multiple sensors at different heights.

In summary, the proposed Bayesian inference approach was shown to significantly improve urban wind flow predictions for the test case considered. The results indicate that data assimilation provides a promising route for improving the predictive capabilities of RANS simulations of urban flow. To generalize the conclusions of this study and investigate further opportunities for improvements in the predictions, we should consider validation in other urban morphologies, compare the performance of different inference methods, and investigate approaches to represent the error of the RANS model.

6. Acknowledgements

The authors are thankful for the support provided by Stanford staff members and by Yunjae Hwang and Giacomo Lamberti during the deployment of the experimental campaign. We also acknowledge the high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation (NSF), along with support from NSF Career

award #1749610.

References

[1] K. C. Seto, B. Güneralp, L. R. Hutyra, Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools, Proceedings of the National Academy of Sciences 109 (40) (2012) 16083–16088. doi:10.1073/pnas.1211658109.

[2] V. Boppana, Z.-T. Xie, I. P. Castro, Large-eddy simulation of dispersion from surface sources in arrays of obstacles, Boundary-layer meteorology 135 (3) (2010) 433–454. doi:10.1007/s10546-010-9489-9.

[3] B. Blocken, T. Stathopoulos, J. Van Beeck, Pedestrian-level wind conditions around buildings: Review of wind-tunnel and cfd techniques and their accuracy for wind comfort assessment, Building and Environment 100 (2016) 50–81. doi:10.1016/j.buildenv.2016.02.004.

[4] R. Ramponi, B. Blocken, Cfd simulation of cross-ventilation for a generic isolated building: impact of computational parameters, Building and Environment 53 (2012) 34–48. doi:10.1016/j.buildenv.2012.01.004.

[5] A.-S. Yang, Y.-M. Su, C.-Y. Wen, Y.-H. Juan, W.-S. Wang, C.-H. Cheng, Estimation of wind power generation in dense urban area, Applied energy 171 (2016) 213–230. doi:10.1016/j.apenergy.2016.03.007.

[6] B. Blocken, 50 years of computational wind engineering: past, present and future, Journal of Wind Engineering and Industrial Aerodynamics 129 (2014) 69–102. doi:10.1016/j.jweia.2014.03.008.

[7] P. Klein, B. Leitl, M. Schatzmann, Concentration fluctuations in a downtown urban area. part ii: analysis of joint urban 2003 wind-tunnel measurements, Environmental Fluid Mechanics 11 (1) (2011) 43–60.

[8] M. Schatzmann, B. Leitl, Issues with validation of urban flow and dispersion cfd models, Journal of Wind Engineering and Industrial Aerodynamics 99 (4) (2011) 169–186. doi:10.1016/j.jweia.2011.01.005.

[9] D. Wise, V. Boppana, K. Li, H. Poh, Effects of minor changes in the mean inlet wind direction on urban flow simulations, Sustainable cities and society 37 (2018) 492–500. doi:10.1016/j.scs.2017.11.041.

[10] D. D. Lucas, A. Gowardhan, P. Cameron-Smith, R. L. Baskett, Impact of meteorological inflow uncertainty on tracer transport and source estimation in urban atmospheres, Atmospheric Environment 143 (2016) 120–132. doi:10.1016/j.atmosenv.2016.08.019.

[11] C. García-Sánchez, J. van Beeck, C. Gorlé, Predictive large eddy simulations for urban flows: Challenges and opportunities, Building and Environment 139 (2018) 146–156. doi:10.1016/j.buildenv.2018.05.007.

[12] C. García-Sanchez, D. Philips, C. Gorlé, Quantifying inflow uncertainties for cfd simulations of the flow in downtown oklahoma city, Building and Environment 78 (2014) 118–129. doi:10.1016/j.buildenv.2014.04.013.

[13] C. Gorlé, C. Garcia-Sanchez, G. Iaccarino, Quantifying inflow and rans turbulence model form uncertainties for wind engineering flows, Journal of Wind Engineering and Industrial Aerodynamics 144 (2015) 202–212. doi:10.1016/j.jweia.2015.03.025.

[14] C. García-Sánchez, G. Van Tendeloo, C. Gorlé, Quantifying inflow uncertainties in rans simulations of urban pollutant dispersion, Atmospheric Environment 161 (2017) 263–273.

[15] L. Margheri, P. Sagaut, A hybrid anchored-anova–pod/kriging method for uncertainty quantification in unsteady high-fidelity cfd simulations, Journal of Computational Physics 324 (2016) 137–173. doi:10.1016/j.jcp.2016.07.036.

[16] J. Sousa, C. García-Sánchez, C. Gorlé, Improving urban flow predictions through data assimilation, Building and Environment 132 (2018) 282–290. doi:10.1016/j.buildenv.2018.01.032.

[17] V. Mons, L. Margheri, J.-C. Chassaing, P. Sagaut, Data assimilation-based reconstruction of urban pollutant release characteristics, Journal of wind engineering and industrial aerodynamics 169 (2017) 223–250.

[18] F. K. Chow, B. Kosovi, S. Chan, Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations, Journal of applied meteorology and climatology 47 (6) (2008) 1553–1572. doi:10.1175/2007JAMC1733.1.

[19] H. Xiao, J.-L. Wu, J.-X. Wang, R. Sun, C. Roy, Quantifying and reducing model-form uncertainties in reynolds-averaged navierstokes simula-

tions: A data-driven, physics-informed bayesian approach, Journal of Computational Physics 324 (2016) 115–136.

[20] J. D. Christopher, N. T. Wimer, C. Lapointe, T. R. Hayden, I. Grooms, G. B. Rieker, P. E. Hamlington, Parameter estimation for complex thermal-fluid flows using approximate bayesian computation, Physical Review Fluids 3 (10) (2018) 104602. doi:10.1103/PhysRevFluids.3.104602.

[21] X. Li, F. Xue, Bayesian inversion of inflow direction and speed in urban dispersion simulations, Building and Environmentdoi:10.1016/j.buildenv.2018.08.042.

[22] K. W. Ayotte, Optimization of upstream profiles in modelled flow over complex terrain, Boundary-Layer Meteorology 83 (2) (1997) 285–309. doi:10.1023/A:1000229609412.

[23] WindSonic1 and WindSonic4 Two-Dimensional Sonic Anemometers, Tech. rep., Campbell Scientific, Inc (01 2016).
URL https://s.campbellsci.com/documents/us/manuals/windsonic.pdf

[24] Cr300 series dataloggers, https://www.campbellsci.com/cr300, accessed: 2018-08-1 (2018).

[25] http://www.westernwx.com/stanford/stanfordweather.html, accessed: 2018-09-15 (2018).

[26] M. A. Iglesias, K. J. Law, A. M. Stuart, Ensemble kalman methods for inverse problems, Inverse Problems 29 (4) (2013) 045001.

[27] G. Evensen, Data assimilation: the ensemble Kalman filter, Springer Science & Business Media, 2009.

[28] C. Greenshields, Openfoam user guide, CFD Direct Ltd.

[29] D. Wilcox, Turbulence modeling for cfd, Inc. La Canada, California.

[30] J. Franke, Best practice guideline for the CFD simulation of flows in the urban environment, Meteorological Inst., 2007.

[31] Y. Tominaga, A. Mochida, R. Yoshie, H. Kataoka, T. Nozu, M. Yoshikawa, T. Shirasawa, Aij guidelines for practical applications of cfd to pedestrian wind environment around buildings, Journal of wind engineering and industrial aerodynamics 96 (10-11) (2008) 1749–1761. doi:10.1016/j.jweia.2008.02.058.

[32] P. Richards, R. Hoxey, Appropriate boundary conditions for computational wind engineering models using the k-ε turbulence model (1993) 145–153doi:10.1016/B978-0-444-81688-7.50018-8.

[33] J. Wieringa, Updating the davenport roughness classification, Journal of Wind Engineering and Industrial Aerodynamics 41 (1-3) (1992) 357–368. doi:10.1016/0167-6105(92)90434-C.

[34] B. Blocken, T. Stathopoulos, J. Carmeliet, Cfd simulation of the atmospheric boundary layer: wall function problems, Atmospheric environment 41 (2) (2007) 238–252. doi:10.1016/j.atmosenv.2006.08.019.

[35] A. Parente, C. Gorlé, J. van Beeck, C. Benocci, A comprehensive modelling approach for the neutral atmospheric boundary layer:

consistent inflow conditions, wall function and turbulence model, Boundary-layer meteorology 140 (3) (2011) 411–428. doi:10.1007/s10546-011-9621-5.

[36] G. Kang, J.-J. Kim, D.-J. Kim, W. Choi, S.-J. Park, Development of a computational fluid dynamics model with tree drag parameterizations: Application to pedestrian wind comfort in an urban area, Building and Environment 124 (2017) 209–218. doi:10.1016/j.buildenv.2017.08.008.

[37] A. M. Endalew, M. Hertog, M. Delele, K. Baetens, T. Persoons, M. Baelmans, H. Ramon, B. Nicolaï, P. Verboven, Cfd modelling and wind tunnel validation of airflow through plant canopies using 3d canopy architecture, International Journal of Heat and Fluid Flow 30 (2) (2009) 356–368. doi:10.1016/j.ijheatfluidflow.2008.12.007.

[38] S. Vranckx, P. Vos, B. Maiheu, S. Janssen, Impact of trees on pollutant dispersion in street canyons: A numerical study of the annual average effects in antwerp, belgium, Science of the Total Environment 532 (2015) 474–483. doi:10.1016/j.scitotenv.2015.06.032.

[39] G. G. Katul, L. Mahrt, D. Poggi, C. Sanz, One-and two-equation models for canopy turbulence, Boundary-Layer Meteorology 113 (1) (2004) 81–109.

[40] B. Lalic, D. T. Mihailovic, An empirical relation describing leaf-area density inside the forest for environmental modeling, Journal of Applied

Meteorology 43 (4) (2004) 641–645. doi:10.1175/1520-0450(2004)043<0641:AERDLD>2.0.CO;2.

[41] N. K. Verma, D. W. Lamb, N. Reid, B. Wilson, Comparison of canopy volume measurements of scattered eucalypt farm trees derived from high spatial resolution imagery and lidar, Remote Sensing 8 (5) (2016) 388. doi:10.3390/rs8050388.

[42] N. Wiener, The homogeneous chaos, American Journal of Mathematics 60 (4) (1938) 897–936. doi:10.2307/2371268.

[43] M. S. Eldred, A. A. Giunta, B. G. van Bloemen Waanders, S. F. Wojtkiewicz, W. E. Hart, M. P. Alleva, Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis, Tech. rep., Tech. Rep. SAND2006-6337, Sandia National Laboratories (2006).

[44] T. c. Patterson, The optimum addition of points to quadrature formulae, Mathematics of Computation 22 (104) (1968) 847–856. doi:10.1090/S0025-5718-68-99866-9.

[45] M. Asch, M. Bocquet, M. Nodet, Data assimilation: methods, algorithms, and applications, Vol. 11, SIAM, 2016.

[46] M. Emory, J. Larsson, G. Iaccarino, Modeling of structural uncertainties in reynolds-averaged navier-stokes closures, Physics of Fluids 25 (11) (2013) 110822.

[47] C. Gorlé, S. Zéoli, J. Larsson, M. Emory, G. Iaccarino, Epistemic uncertainty quantification for reynolds-averaged navier-stokes modeling of separated flows over streamlined surfaces, Physics of Fluids In press.