Proceedings of Building Simulation 2011:
12th Conference of International Building Performance Simulation Association, Sydney, 14-16 November.

STATISTICAL CALIBRATION OF CFD MODELLING FOR STREET CANYON FLOWS

Nina Glover [1], Serge Guillas[2], Liora Malki-Epshtein[1]
[1]Department of Civil, Environmental and Geomatic Engineering, UCL, London, UK
[2]Department of Statistical Science, UCL, London, UK

ABSTRACT

CFD simulations of complex outdoor environments present a significant modelling challenge. Simulations of airflow within an idealized street canyon are performed here. We test the model sensitivity to the empirical constants contained within the k-ε turbulence model and examine how a systematic variation of these values could produce improved prediction of the turbulent kinetic energy when compared against wind tunnel data. The Bayesian statistical calibration shows the range of values the constants should take. This results in improved CFD simulations in the region of flow inside the street canyon, which is normally very difficult to resolve accurately in CFD models.

INTRODUCTION

An in-depth understanding of the airflow processes within street canyons is important to fully understand the pollutant dispersion within these spaces as well as issues relating to pedestrian comfort and energy use. CFD has had a larger role to play in this process in recent years as computer power and availability of software has increased. The use of CFD modelling in the built environment is widespread for indoor applications, but simulations of the outdoor environment are still often carried out in atmospheric boundary layer wind tunnels (Blocken et al., 2011).

Blocken et al. (2011) present a detailed review of CFD simulations of the outdoor environment. The successful and systematic application of CFD for the outdoor environment is still hindered by lack of accuracy, high computational storage, time and costs, since the computational domains are very large and the boundary conditions are generally not well known. Even the most basic case, CFD simulation of an equilibrium atmospheric boundary layer in an empty domain, can show very large errors in simulation results. Attempts in validation of CFD models, such as simulations for idealized high-rise buildings surrounded by low-rise buildings with the standard k-ε model, can show, in some instances, underestimation of wind tunnel flows by up to a factor 4. Careful validation of CFD simulations with wind tunnel experiments requires a grid-sensitivity analysis, full consideration of the choice of boundary conditions and high-order discretization schemes, and a detailed comparison with the wind tunnel measurements.

We suggest a new method to help validate CFD models against wind tunnel data, one which improves results in the low speed regions within street canyons and provides much greater faith in the modelling process. Here we investigate how the tuning of the k-ε model constants can improve the prediction of the turbulent quantities within the street. We attempt the calibration for a particular case study of flow over a street canyon at aspect ratio of one and a Reynolds number of $5\text{x}10^4$. Three of the five model constants are calibrated against observations of turbulent kinetic energy collected in wind tunnel laboratory experiments (Kastner-Klein et al., 2001). These are $C_{\varepsilon 1}$, $C_{\varepsilon 2}$ and C_μ. The calibration process involves representing bias and computer model outputs as Gaussian processes, following the Bayesian calibration framework of Kennedy and O'Hagan (2001). Thus we can: (1) evaluate a small systematic bias of the CFD model, (2) narrow down the set of parameter values that provide the best match between the CFD model outputs and the observations, (3) quantify the uncertainty of the turbulent kinetic energy outputs resulting from both uncertainties in the CFD parameterization, the numerical code itself and measurement errors, (4) emulate the CFD model to get fast and improved predictions.

This paper starts with a brief introduction to the standard k-ε model followed by the description of the experimental and CFD set up. Then, an explanation of the Bayesian calibration is given. Finally, we discuss the results and the conclusions gained from them.

Turbulence Models

Turbulent flow is characterized by random fluctuations of velocity. It is possible to model turbulent flow within CFD without any adjustments to the Navier-Stokes equations. This type of simulation is known as direct numerical simulation (DNS) and is prohibitively computationally expensive. Turbulence models used within the CFD simulations enables the capture of the main features of the flow without having to explicitly model all the details of the turbulence, thus saving on computer costs. Large eddie simulations (LES) only models large features of the flow, but is either still out of reach for simulation of the outdoor environment, or when computational power is available, is very difficult to tune, which may make it imprecise, see Blocken et al. (2011) and references therein.

We can reduce the amount of computing power needed by focusing on the mean properties of the flow. This results in the Reynolds-Averaged Navier Stokes (RANS) equations. These equations contain corre-

lations of the fluctuating velocity components $\overline{u_i' u_j'}$ which are known as the Reynolds stresses. The turbulence model is a way of closing the RANS equations by approximating the Reynolds stress. The standard k-ε turbulence model is an example of a RANS model and is a popular choice in CFD modelling due to its robust nature and the fact that it has been well validated. However it has been noted that the standard k-ε model has problems predicting flow separation and underpredicts turbulent kinetic energy values within street canyons (Cheng et al., 2003; Lien et al., 2004; Solazzo et al., 2009). The practicality of the standard k-ε model means that it is still widely used in industry and research, thus creating a demand for improved performance from the model (Solazzo, 2008). This fact along with its well documented shortcomings provides the motivation behind the choice of the standard k-ε model for this study.

The standard *k*-ε model

The Reynolds stresses $\overline{u_i' u_j'}$ are related to the shear stress of the flow τ by the following equation:

$$\tau = \rho \overline{u_i' u_j'} \tag{1}$$

where ρ is the density of the fluid. We can find the value for τ, and hence $\overline{u_i' u_j'}$, by the Boussinesq hypothesis:

$$\tau = \mu_t \frac{du}{dy} \tag{2}$$

where μ_t is the turbulent viscosity and $\frac{du}{dy}$ is the mean velocity gradient. In the case of the standard k-ε turbulence model, turbulent viscosity is defined as

$$\mu_t = \rho C_\mu \frac{k^2}{\varepsilon} \tag{3}$$

where C_μ is a model constant.

By solving the following differential equations for the turbulent kinetic energy, k, and the turbulent dissipation, ε we can find a value for μ_t (Launder and Spalding, 1974):

$$\begin{aligned} \frac{D\varepsilon}{Dt} &= \frac{1}{\rho}\frac{\partial}{\partial x_k}\left[\frac{\mu_t}{\sigma_\varepsilon}\frac{\partial \varepsilon}{\partial x_k}\right] \\ &\quad + \frac{C_{\varepsilon 1}\mu_t}{\rho}\frac{\varepsilon}{k}\left(\frac{\partial U_i}{\partial x_k} + \frac{\partial U_k}{\partial x_i}\right)\frac{\partial U_i}{\partial x_k} \\ &\quad - \frac{C_{\varepsilon 2}\varepsilon^2}{k} \end{aligned} \tag{4}$$

$$\begin{aligned} \frac{Dk}{Dt} &= \frac{1}{\rho}\frac{\partial}{\partial x_k}\left[\frac{\mu_t}{\sigma_k}\frac{\partial k}{\partial x_k}\right] \\ &\quad + \frac{\mu_t}{\rho}\left(\frac{\partial U_i}{\partial x_k} + \frac{\partial U_k}{\partial x_i}\right)\frac{\partial U_i}{\partial x_k} - \varepsilon \end{aligned} \tag{5}$$

where σ_k, σ_ε, $C_{\varepsilon 1}$, $C_{\varepsilon 2}$ and C_μ are all empirical model constants. The default values for these constants in most commercial CFD softwares, including Ansys CFX tested here, are shown in Table 1.

Table 1: standard values used for model constants

C_μ	$C_{\varepsilon 1}$	$C_{\varepsilon 2}$	σ_ε	σ_k
0.09	1.44	1.92	1.3	1.0

These values were found through data fitting for a wide range of flows (Launder and Spalding, 1974) but ad-hoc adjustments are often made to them depending on the situation being modeled. When using the k-ε model to simulate the atmospheric boundary layer Richards and Hoxey (1993) chose a value of 0.013 for C_μ. This value was obtained through detailed field measurements and is based upon the following relationship:

$$C_\mu = \left(\frac{u_*^2}{k}\right)^2 \tag{6}$$

where u_* is the frictional velocity. Beljaars et al. (1987) and Hagen et al. (1981) used this relationship along with field data to justify the use of $C_\mu = 0.0324$ and $C_\mu = 0.026$, respectively, in their simulations. Hargreaves and Wright (2007), Richards and Hoxey (1993), Beljaars et al. (1987) and Solazzo et al. (2009) all alter the value for σ_ε. This is based on the types of boundary conditions used at the inlet and the restrictions these place on σ_ε . This is discussed in further detail in the section on statistical calibration.

Wind tunnel set up

The data used to validate the CFD model is from a wind tunnel experiment carried out by Kastner-Klein et al. (2001) in the wind tunnel at the University of Karlsruhe. The buildings on either side of the street were represented by two rectangular blocks. The wind direction was perpendicular to the street axis. The model of an atmospheric boundary layer was obtained through the use of small blocks placed on the wind tunnel floor. Measurements were taken using a laser velocimetry system. Velocity measurements were taken on a vertical cross section of the street at the center of the street length.

CFD SIMULATION

The numerical simulation was carried out using Ansys CFX, a commercial CFD software package. The geometry of the model is shown in Figure 1. The mesh used was an unstructured hexahedral mesh with a mesh refinement zone in-between and surrounding the two buildings in order to capture the detail in this area. A 2.5 dimensional simulation was performed. The geometry was extended a small distance in the y direction and symmetry boundaries applied. This reduces the CPU time which is especially useful when running the large number of models necessary for calibration.

It is important to ensure that we can accurately model boundary conditions within the wind tunnel. A simulation was first carried out modeling the empty domain (no streets present) to ensure the correct profiles for velocity and turbulence were being created and there

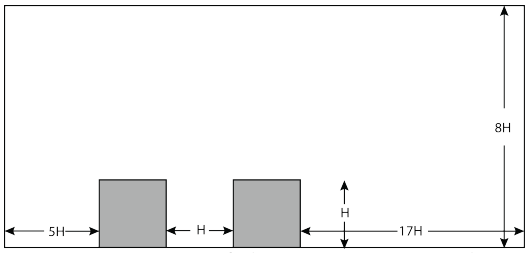

Figure 1: Geometry of the street canyon. The TKE profiles were taken in the middle of the canyon.

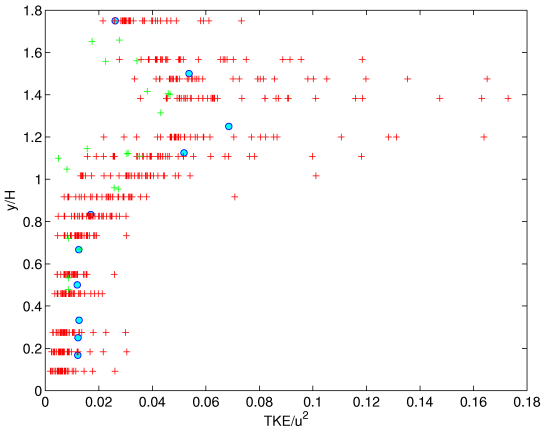

Figure 2: Normalized TKE data from wind tunnel (circles) and CFD simulation (crosses) against height normalized by the height h of the buildings. Red crosses correspond to the initial design and green crosses correspond to the additional design.

was little decay as we moved down the wind tunnel. From testing of the empty domain it was found that the following inlet boundary conditions provided the best match with the experimental data:

- Inlet Velocity Profile:

$$u = u_{ref}\left(\frac{z - d_0}{z_{ref} - d_0}\right)^{0.23} \tag{7}$$

  where z_{ref} = 0.48m is the boundary layer height, u_{ref} = 7m/s (velocity at z_{ref}) and d_0 = 0.002m is the displacement height. This profile is the power law and the value of 0.23 was found by Kastner-Klein et al. (2001).

- Turbulent Kinetic Energy profile:

$$k = \frac{u_*^2}{\sqrt{C_\mu}}\left(1 - \frac{z}{\delta}\right) \tag{8}$$

  where $u_* = 0.38$m/s and $\delta = 0.485$m is boundary layer height.

- Turbulent Dissipation profile:

$$\varepsilon = \frac{u_*^3}{\kappa(z + z_0)}\left(1 - \frac{z}{\delta}\right) \tag{9}$$

  where κ = 0.4 is the Von Karman constant and z_0 = 0.0008m is the roughness length.

These profiles have been applied in previous simulations of urban airflow such as Di-Sabatino et al. (2007). The outlet boundary was set to outflow with static pressure 0pa. The top of the wind tunnel was set to symmetry boundary. A sand grain roughness of $K_s = 0.024$m was applied to the floor of the domain. This value was calculated from the formula given by Blocken et al. (2007) of $K_s = 30z_0$. In order to carry out the calibration process the model is run forty times with varying values for the model constants. The design for these forty runs is described in the following section.

STATISTICAL CALIBRATION

The calibration of our CFD model consists of putting distributional assumptions (prior distributions or simply priors) on the calibration (also called tuning) parameters $C_{\varepsilon 1}, C_{\varepsilon 2}$ and C_μ before comparing with observations, and letting the information contained in the data update this a priori assumption to get as a result a posterior distribution of the calibration parameters. The advantage of such a Bayesian analysis (Kennedy and O'Hagan, 2001) over standard estimation of parameters (e.g. by minimizing the differences between observations and simulator outputs) lies mainly in the ability to retrieve a full description of the uncertainties about the parameters and consequently about the simulator outputs. Moreover, the possibility for the modelers to express their -uncertain- scientific beliefs in terms of priors on the parameters enables a natural integration of scientific knowledge and evidence given by measurements. It was decided to calibrate three out of the five k-ε model constants and not σ_ε and σ_k due to their strong interdependence on the other model constants. This arises from the fact that if equation (8) and (9) are to satisfy equation (4) then the following relationships must hold (Solazzo et al., 2009):

$$\sigma_k = \frac{\kappa^2}{\sqrt{C_\mu}}\frac{z + z_0}{z} \tag{10}$$

$$\sigma_\epsilon = \frac{\kappa^2}{\sqrt{C_\mu}((1 - \frac{z}{\delta})C_{\varepsilon 2} - C_{\varepsilon 1})} \tag{11}$$

We must choose a fixed value for σ_ε and σ_k i.e. they cannot vary with changing height, z. We therefore choose to use the relationship suggested by Richards & Hoxey (1993) to adjust the values of σ_ε and σ_k according to the values of C_μ, $C_{\varepsilon 1}$ and $C_{\varepsilon 2}$.

$$\sigma_k = \frac{\kappa^2}{\sqrt{C_\mu}} \tag{12}$$

$$\sigma_\epsilon = \frac{\kappa^2}{\sqrt{C_\mu}(C_{\varepsilon 2} - C_{\varepsilon 1})} \tag{13}$$

which also satisfy equation (4). For simplicity we now denote $C_{\varepsilon 1}$ and $C_{\varepsilon 2}$ as C_1 and C_2. The intervals chosen to be tested for the calibration constants were as

follows C_1 : 0.5 - 1.5, C_2: 1.5 - 2.5 and C_μ: 0 - 0.2. These were chosen based on the standard values suggested for the model constants and how these had been changed in the past. We put uniform priors on these parameters, allowing for equal initial probability of being at any location in these intervals. Further expert judgment, say via more preliminary experience on the sensitivity of CFD models to these parameters in various settings could have improved the priors by setting meaningful probability distributions on these.

The complete set of inputs $x = (h, \theta)$ consists of parameters divided into two categories: the known parameters (normalized height h in $[0, 2]$) and the unknown calibration parameters $\theta = (C_1, C_2, C_\mu)$. We denote by $y^M(x)$ the empirical output of the computer model as a function of $x = (h, \theta)$ and $\eta(x)$ the expected output of the computer model as a function of $x = (h, \theta)$. The difference between $y^M(x)$ and $\eta(x)$ is the numerical intrinsic error. The computer code output $\eta(x)$ is an approximation of the reality $y^R(h)$. The notation used emphasizes that physical observations are only made at values of the observable parameter, h. To learn about the values of the calibration parameters, the CFD model is run at inputs x in a design (i.e. choice of values) D^M. Experimental data (i.e. TKE observations) $y^F(h)$ are collected at a number of inputs heights h in a design D^F from the center of the street canyon (see figure 1).

The design D^F (the normalized heights at which observations are collected) are given by Kastner-Klein et al. (2001). For our design of experiments D^M corresponding to the calibration parameters, we used a sequential design (Gramacy and Lee, 2009). We started with a maximin Latin Hypercube (LHS) design. With this design we try to cover as much space as possible in the three-dimensional space of the calibration parameters with only $n = 40$ runs. For the normalized height components of the computer design D^M, we first choose 15 irregularly spaced points that are close to the observations to maximize the amount of information obtained through these heights under the constraint of the computing time necessary to perform the Bayesian calibration. Note that using uneven heights is deliberate as it may help later on estimate parameters that describe the decays in the correlation (i.e. smoothness) across heights. The heights in D^F and D^M are different, but the methodology accommodates such variation. Figure 2 shows the CFD computed TKEs at these heights (red crosses) for the various choices of calibration parameters. This is the first step in our study.

The following equations constitute an extension of Kennedy and O'Hagan (2001) as they specify the intrinsic CFD model numerical error. They describe the relationships between the CFD model and the observations at the height design points, using bias $\delta(h)$, intrinsic CFD model numerical error ε_η and observation error ε (both assumed constant across heights h):

$$y^M(h, \theta) = \eta(h, \theta) + \varepsilon_\eta \quad (14)$$

$$y^R(h) = \eta(h, \theta^*) + \delta(h) \quad (15)$$

$$y^F(h) = y^R(h) + \varepsilon \quad (16)$$

Here, θ^* is used to represent the true (unknown) values of the calibration parameters. These equations suggest that even if the CFD model were run at the true values of the calibration parameters, it would still be a biased representation of reality. Thus the model can never perfectly match observations without some additional process of adjusting for model errors.

Because the simulator output $\eta(\cdot)$ is unknown except at the design points D^M, we assume that the unknown function follows a Gaussian stochastic process (GASP) distribution. That is, we model the N observed simulator responses $\eta(x), x \in R^p$ (here $p = 4$ since D^M is over a range of C_1, C_2, C_μ and h, values), as coming from a multivariate normal distribution with mean μ and $N \times N$ variance-covariance function Σ. Since we initially standardize the entire set of responses (CFD model and observed) by the mean and standard deviation of the CFD responses, $\mu = 0$ above and the variability in the simulator $(1/\lambda_\eta)$ below is approximately 1. Thus, we approximate the CFD model by specifying a distribution of functions that interpolate the response $\eta(x)$ in between the design points x in D^M. The random function is certain at the design points, and uncertain at untried points. To specify Σ according to the calibration parameters we use a product Gaussian variance-covariance. Thus, the $(i, j)^{th}$ element of Σ, $cov(\eta(x_i), \eta(x_j))$, is (conveniently using the notation θ_4 for h):

$$\frac{1}{\lambda_\eta} exp(-\sum_{k=1}^{4} \beta_k |\theta_{ik} - \theta_{jk}|^2).$$

The notation θ_{ik} denotes the i^{th} design point in D^M for θ_k. The hyperparameters λ_η (the precision of the GASP model), β_k (which we call "correlation hyperparameters") are to be estimated from the model output and the observations as described below. The unknown bias function $\delta(h)$ is also modeled as a GASP random function with mean 0 and correlation matrix with precision λ_δ and correlation parameter β_5. Finally, the random error and intrinsic error components are modeled as independent $\varepsilon \sim N(0, 1/\lambda_\epsilon)$ and $\varepsilon_\eta \sim N(0, 1/\lambda_{\varepsilon_\eta})$.

For the estimation of the calibration and hyperparameters, we employ the Markov Chain Monte Carlo (MCMC) algorithm. The chains are dependent random samples that ought to be distributed in the long run as the so-called posterior distributions of the parameters of interest, which are combinations of prior information about the values of these parameters and the information about the parameters provided by the data. We then retrieve the posterior distributions of

the various calibration parameters, which allows us to make inferences and quantify our uncertainty about the true values of these unknown quantities.

All the unknowns in the model (i.e. the calibration parameters and the hyperparameters) require specified prior distributions which represent uncertainty about the values of these parameters before any data is collected. The following choices are made for the priors:

- To represent vague prior information about the true calibration parameter values, we specify a uniform prior distribution over an interval twice as wide as the interval on which they were sampled for simulator runs.
- To model the correlation hyperparameters in Σ, we conservatively places most of its prior mass on values for the corresponding correlations near 1 (indicating an insignificant effect). Similarly, conservative priors were used for the hyperparameters associated with the bias function.
- Gamma prior distributions were used for each of the precision (i.e. inverse of the variance) hyperparameters λ_η, λ_δ and λ_ϵ. Specifically, we use priors $\lambda_\eta \sim GAM(10, 10)$ (with expectation 1 due to standardization of the responses), $\lambda_\delta \sim GAM(10, .03)$ (with expectation around 5% of standard deviation of the standardized responses), and $\lambda_\epsilon \sim GAM(10, .03)$ as well.

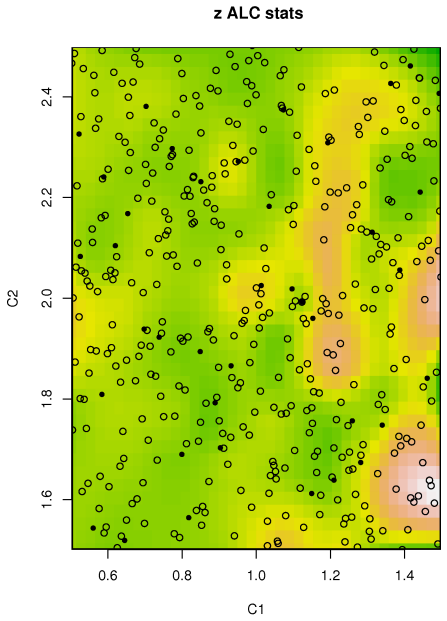

Figure 3: ALC scores for C_1 and C_2. Black dots are projections of the initial design, circles are proposed new design points from which we select the 20 most highly ranked according to ALC.

Because our choice of priors make the full conditional distributions of the unknowns difficult to sample from in the MCMC chain, we implement a Metropolis-Hastings algorithm to explore the multidimensional space of parameters. The algorithm makes use of a proposal distribution to draw a future value conditional on the current state. Then this move is either accepted or rejected according to a random toss with a probability of acceptance that depends upon the target distribution. This eventually yields draws from the target distribution (here the posterior). We used multiple chains to confirm the convergence towards a stationary posterior distribution (after an initial burn-in period), saving wall-clock time by running the chains in parallel.

In a second step, we target regions of higher uncertainty and reduce those. We apply the advanced Learning Cohn (ALC) which relies on building an emulator of the CFD model that enables the relatively fast computation of expected reductions in variance. We select a set of 20 additional points, see Gramacy and Lee (2009, section 3.3) for details. As a result, our final design D^M is of size $N = 620$. Figure 3 displays the ALC scores for projections of these expected reductions along the axis C_1, C_2.

DISCUSSION AND RESULT ANALYSIS

From the results taken from the CFD runs (shown in Figure 2 denoted by crosses) we can see the wide spread in TKE results produced by varying the k-ε model constants. This highlights the importance of choosing the most appropriate values for these constants and the need for calibration. The posterior

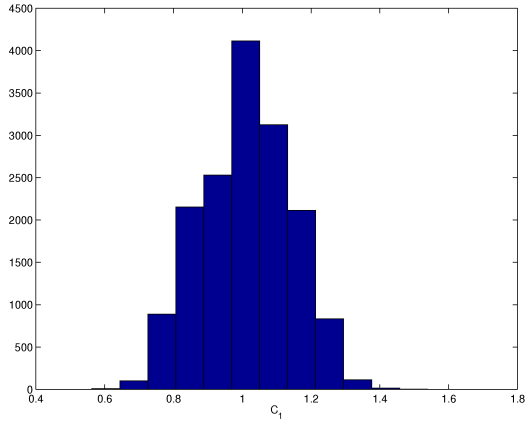

Figure 4: Histogram of all posterior draws for C_1.

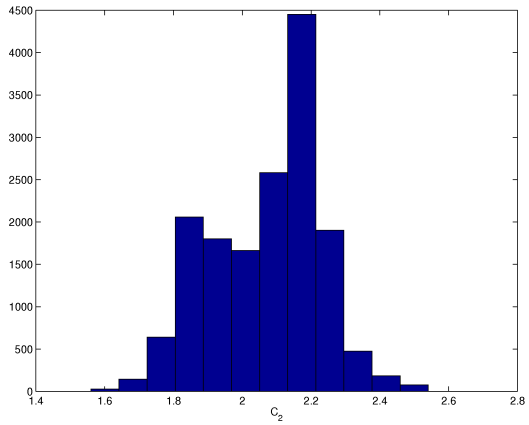

Figure 5: Histogram of all posterior draws for C_2.

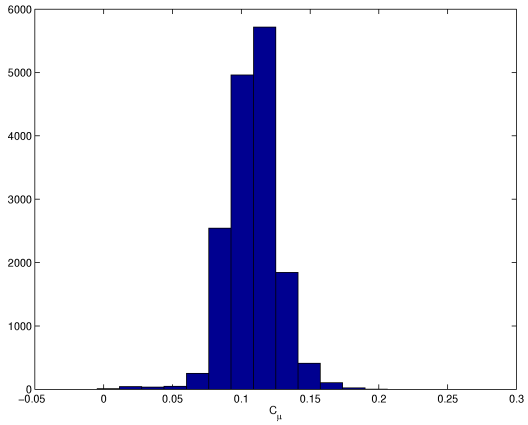

Figure 6: Histogram of all posterior draws for C_μ.

for the numerical intrinsic error ε_η turns out to stay very small, from 0.9 to 1.6 $\times 10^{-3}$. This indicates that within CFX the errors due to the intrinsic numerical codes, unrelated to which parameters are used in the model, are very small, which is reassuring. Figures 4, 5, and 6 display the histograms of posterior draws for the calibration parameters, C_1, C_2 and C_μ. From the distribution of the histograms we can infer information about the preferred values for each parameter and the uncertainty related to this parameter. In particular, we notice that the initial non-calibrated value of 1.44 for C_1 on Figure 4 is not deemed a good value for this CFD model. This is a major finding for future runs in this context. Furthermore, the spread for C_2 is wide, which indicates that the likely values can vary in the original interval 1.5-2.5 though with more probability between 1.8 and 2.3. As for C_μ, the tuning procedure reduced the prior interval of 0-0.2 to 0.05-0.17 with a very large probability in the center of this interval as the histogram shows a sharp peak there.

Based on these histograms, values of $C_1 = 1$, $C_2 = 2.1$, $C_\mu = 0.12$ were selected as being likely values to give the most improved model performance. Using equations (12) and (13) the value of σ_ϵ= 0.42 and σ_k= 0.462 were found. The CFD model was run with the new choice of calibration parameters. The results for the modified CFD model can be seen in Figures 7(b) and 8. Qualitative comparison of the CFD flow patterns against that from the wind tunnel shows the improvement produced by the modification of the parameters. The size of the flow separation and vortex produced above the upstream building is reduced (Figure 7(b)). This causes the center of vortex contained within the street canyon to move down towards the center of the street. This is in line with the results of Kastner-Klein et al. (2004) and previous research on flow within regular street canyons. A quantitative comparison is shown in Figure 8 which shows the TKE profiles taken from the center of the street canyon. From this we can see that TKE predictions are improved by the modified CFD model. The common problem of under prediction of TKE within the street is still present but less so compared with the previous

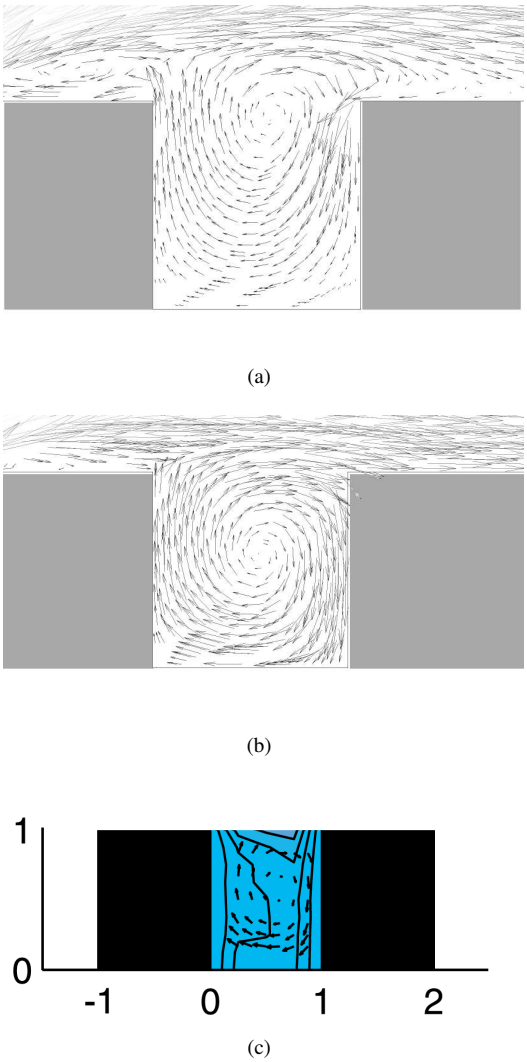

Figure 7: Vector plot showing velocities for (a) unmodified CFD model (b) CFD model with modified constants (c) results from wind tunnel experiment Kastner-Klein et al. (2004)

model. Although this was not part of the Bayesian calibration process it is used to demonstrate how more informed choices of the calibration parameters can improve both TKE and flow patterns.

The final stage in the calibration process is to produce an emulator of the CFD model. This extracts all the information contained within the CFD runs regarding the calibration parameters and their uncertainties and uses it to produce predictions of the TKE values at specified heights. The results of the initial emulation can be seen in Figure 9 denoted by the green line. By adding in the bias found for the CFD model we get our final prediction (blue line). We can now see how the sequential design has improved the emulation and predictions. Using these 20 additional points results in overall tighter confidence regions for the prediction. Indeed, Figure 9 displays larger confidence bands in the initial design-based predictions than when 20 additional well-chosen points are included in the analy-

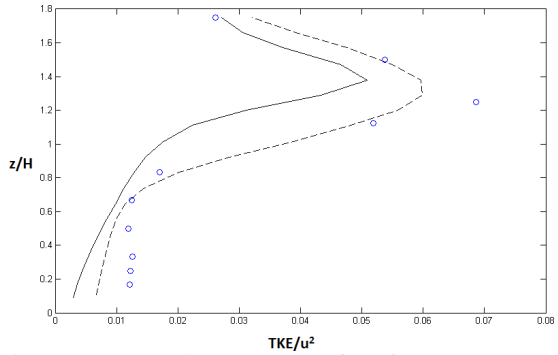

Figure 8: Normalized TKE profiles from CFD model with un-modified constants (black line), modified constants (dashed line) and wind tunnel data (circles)

sis. This occurs in particular at the height of 1.3 where TKE is maximum, and at the height of 0.4 where there is little variation and levels are low. The larger and widening confidence width at the bottom heights of around 0.2, with the full design compared to the initial one, does not contradict this point. To the contrary, it highlights the fact that correlations and levels of uncertainties in the predictions are better estimated using the 20 additional points. Indeed, one expects widening confidence widths when predictions are made at the borders of intervals as there is no possibility of borrowing strength from unavailable information across the border.

In these figures, the biases are small, as shown by the difference between the emulator of the CFD model (in green) and the prediction using the bias (in blue). It is slightly larger within the street canyon (normalized heights less than 1). This is as expected as it has been noted that the standard k-ε model has problems predicting the TKE values within the street. Here we put a small prior assumption on the size of the height related bias, thus constraining this bias to be small. We wanted the statistical calibration to find strong evidence in the data to change the opinion that CFD models can be very accurate by just tuning our calibration parameters. This belief cannot be held in general, and is specific to the numerical model and the application: for some atmospheric models pushing the calibration parameters to extreme values in order to match observations is unrealistic and acknowledging a potentially large bias is a better strategy (Guillas et al., 2009).

From this process we can see that making informed choices about the standard k-ε model constants can drastically improve the CFD model. However the best predictions of the TKE values were found by emulating the CFD model using the full Bayesian framework which enables the quantification of uncertainties regarding the parameters and the numerical model, not by merely choosing some combination of parameter values.

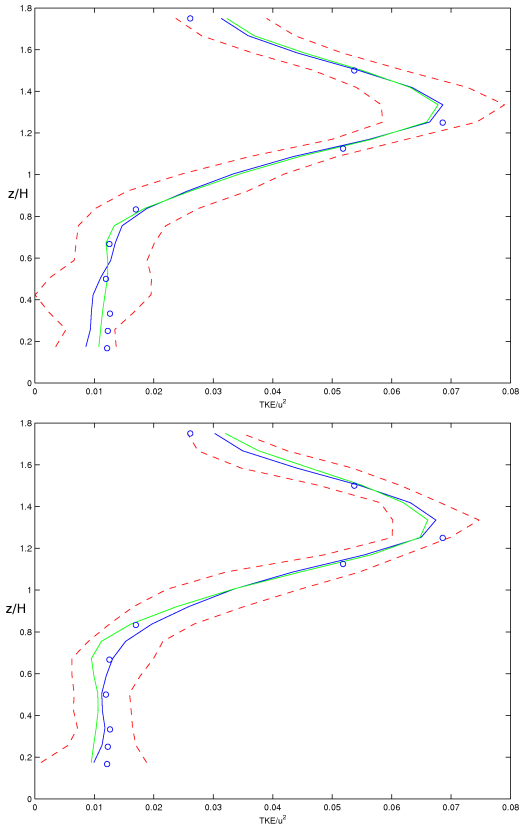

Figure 9: Emulator (green line) and prediction with estimated bias added (blue line, with red lines corresponding to the 95% confidence interval) and observations (circles), using the initial design (top panel) or the full design (bottom panel).

CONCLUSION

The focus of this paper was on the constants contained within the standard k-ε model, which are often in industry left unchanged from the original values suggested by Launder and Spalding (1972) seen in Table 1. By varying these constants we were able to asses their influence on the CFD models capability to simulate flow within and above a street canyon. The wide spread of results shown in Figure 2 suggest these values have a significant impact on the turbulence values and can be adjusted to improve model predictions as seen by the use of modified parameters shown in Figures 7 and 8. Although these values have been changed in the past, the approach used here is unique as it involves the use of advanced statistical methods which have not been applied to this problem before.

Through the use of Bayesian calibration we were able to gain a greater understanding of the uncertainties relating to model constants and quantify the bias of the CFD model itself. Although the bias is slightly larger within the street canyon it is still small enough to assume that the standard k-ε model is capable of accurately reproducing flow within a street canyon. We were also able to use statistical methods to emulate the CFD model, providing a much improved predic-

tion of the turbulent kinetic energy values within the street canyon by integrating over the range of possible values for the parameters. To our knowledge, it is the first study to quantify uncertainties directly relating to the k-ε model constants, see Najm (2009) for a review of direct computation of numerical uncertainties in CFD outputs. Probabilistic statements can thus be made about critical thresholds for turbulence and flow speeds that could not be stated before.

This paper demonstrates the Bayesian calibration process of a CFD model for one particular case study. A suggestion for future work would be to perform the same process for flows with different Reynolds numbers thus testing the sensitivity of the k-ε model constants to changes in the Reynolds number. In this case we chose the k-ε model as its limitations have been well documented by previous research making it an ideal candidate for calibration. However the calibration process can be used on any independent CFD input parameter therefore Bayesian calibration posesses a wide range of possibilities for CFD modeling in the future including investigating the uncertainties inherent within other turbulence models.

REFERENCES

Beljaars, A. C. M., Walmsley, J. L., and Taylor, P. A. 1987. A mixed spectral finite-difference model for neutrally stratified boundary-layer flow over roughness changes and topography. *Bound.-Layer Meteor.*, 38(3):273–303.

Blocken, B., Stathopoulos, T., and Carmeliet, J. 2007. CFD simulation of the atmospheric boundary layer: wall function problems. *Atmos. Environ.*, 41(2):238–252.

Blocken, B., Stathopoulos, T., Carmeliet, J., and Hensen, J. L. M. 2011. Application of computational fluid dynamics in building performance simulation for the outdoor environment: an overview. *J. Building Perform. Simul.*, 4(2):157–184.

Cheng, Y., Lien, F. S., Yee, E., and Sinclair, R. 2003. A comparison of large eddy simulations with a standard k-ε Reynolds-averaged Navier-Stokes model for the prediction of a fully developed turbulent flow over a matrix of cubes. *J. Wind Eng. Ind. Aerodyn.*, 91(11):1301–1328.

Di-Sabatino, S., Buccolieri, R., Pulvirenti, B., and Britter, R. 2007. Simulations of pollutant dispersion within idealised urban-type geometries with CFD and integral models. *Atmos. Environ.*, 41(37):8316–8329.

Gramacy, R. B. and Lee, H. K. H. 2009. Adaptive Design and Analysis of Supercomputer Experiments. *Technometrics*, 51(2):130–145.

Guillas, S., Rougier, J., Maute, A., Richmond, A. D., and Linkletter, C. D. 2009. Bayesian calibration of the thermosphere-ionosphere electrodynamics general circulation model (TIE-GCM). *Geosci. Model Dev.*, 2:137–144.

Hagen, L., Skidmore, E., Millerand, P., and Kipp, J. 1981. Simulation of effect of wind barriers on airflow. *Trans. ASAE*, 24:1002–1008.

Hargreaves, D. and Wright, N. 2007. On the use of the k-ε model in commercial CFD software to model the neutral atmospheric boundary layer. *J. Wind Eng. Ind. Aerodyn.*, 95(5):355–369.

Kastner-Klein, P., Berkowicz, R., and Britter, R. 2004. The influence of street architecture on flow and dispersion in street canyons. *Meteorology and Atmospheric Physics*, 87(1):121–131.

Kastner-Klein, P., Fedorovich, E., and Rotach, M. W. 2001. A wind tunnel study of organised and turbulent air motions in urban street canyons. *J. Wind Eng. Ind. Aerodyn.*, 89(9):849–861.

Kennedy, M. C. and O'Hagan, A. 2001. Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(3):425–464.

Launder, B. E. and Spalding, D. B. 1972. *Lectures in Mathematical Models of Turbulence*. Academic Press, London.

Launder, B. E. and Spalding, D. B. 1974. The numerical computation of turbulent flows. *Comput. Meth. Appl. Mech. Eng.*, 3(2):269–289.

Lien, F. S., Yee, E., and Cheng, Y. 2004. Simulation of mean flow and turbulence over a 2D building array using high-resolution CFD and a distributed drag force approach. *J. Wind Eng. Ind. Aerodyn.*, 92(2):117–158.

Najm, H. N. 2009. Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annu. Rev. Fluid Mech.*, 41:35–52.

Richards, P. and Hoxey, R. 1993. Appropriate boundary conditions for computational wind engineering models using the k-ε turbulence model. *J. Wind Eng. Ind. Aerodyn.*, 46-47:145–153.

Solazzo, E. 2008. *Modelling dispersion of traffic-related pollutants in urban canyons and intersections*. PhD thesis, University of Birmingham.

Solazzo, E., Cai, X., and Vardoulakis, S. 2009. Improved parameterisation for the numerical modelling of air pollution within an urban street canyon. *Environ. Modell. Softw.*, 24(3):381–388.