USE OF DETERMINISTIC SAMPLING FOR UNCERTAINTY QUANTIFICATION IN CFD

P. Hedberg[1] and P. Hessling
SP Technical Research Institute of Sweden, Measurement Technology, Box 857, SE-50115 Boras, Sweden
[1]current address: SSM Swedish Radiation Safety Authority, Solna strandväg 96, SE-171 16 Stockholm, Sweden
peter.hedberg@ssm.se
peter.hessling@sp.se

ABSTRACT

In measurement technology it is good engineering practice to estimate the uncertainties in the measurements. Estimates of the uncertainties in results from numerical simulations are much less common. This can partly be explained by the additional computational cost it involves. This cost is a function of which method is used to propagate the input and model uncertainty to the result. Some methods require a large number of repeated simulations. Other methods suffer from the "curse of dimensionality" and can realistically only treat a few uncertain parameters. A novel method, called deterministic sampling (DS), is proposed to quantify the uncertainty in a numerical flow simulation and does not display the previously mentioned shortcomings. The new method is efficient since relatively few simulations are required, and many parameters can be present. In this paper the method is exemplified by uncertainties originating from turbulence model constants. The method is based on the idea that a continuous probability density function can be replaced by an ensemble of discrete deterministic samples, provided the two representations have the same statistical moments. The method is first illustrated in a simple example, and later applied to the classic simulation of turbulent flow over a backward-facing step. The results obtained here are similar to what other researchers have found using Latin Hypercube Sampling for the same case, but at a lower computational cost. Both methods show that much of the experimental results fit within the calculated uncertainties, but not all of them. This indicates the presence of systematic uncertainties in the turbulence model, and/or systematic problems in the experimental setup, such as lack of two-dimensionality in the flow. Uncertainties from other origins, such as boundary conditions and sources, can be also included in the method if desired. The DS method is also not limited to fluid flow simulations and can for example be applied to systems analyses codes in thermal hydraulics.

KEYWORDS

deterministic sampling, uncertainty quantification, unscented transform, turbulence model constants

1. INTRODUCTION

Two experimental results can only be compared if they both have an uncertainty associated with them, often represented by mean values and uncertainty bars. If there is an overlap in the two uncertainty bars, the two measurements are said to be consistent, i.e. there can exist a value which can fit inside both uncertainty bars. It seems natural that the same should be done when comparing numerical results with experimental data, or results from two numerical simulations. Currently, there are several approaches available to evaluate uncertainty in simulated results, but many suffer from poor efficiency and are therefore not appealing. A new method, Deterministic Sampling (DS) [1], used for Uncertainty Quantification (UQ), is introduced in this paper to offer an efficient alternative. In numerical simulations, uncertainties are associated with a number of input parameters. The turbulence models and models which depict the transportation of heat, mass, and concentration, all contain parameters, or model constants, which values are often determined from experimental data. The uncertainties of the constants are as big as the experimental uncertainty they are based on. Other sources of uncertainties also stem from boundary conditions, geometry, material and fluid properties. On top of all this, we have the uncertainty due to inadequate mesh resolution of the computational domain, and the discretization of the equations. It is advisable that grid independent solutions will be reached before using the DS method. Code implementation errors can also introduce epistemic uncertainties which are difficult to quantify. The physical models also contain epistemic uncertainties since they cannot mimic all physical phenomena. Epistemic uncertainties can be handled by performing calibration of the numerical model. This is the subject of Inverse Uncertainty Quantification, or System Identification, which is not within the scope of the present paper. However, the DS method can favorably be applied for this task too [2].

If the model equations are linear, the method described in [3] can be used. The recommendation in [3] for situations where the parameters propagate non-linearly through equations is to apply the Monte Carlo (MC) method, [4,5]. This method requires several thousands of simulations with different, randomly generated, combinations of parameter values. Since most CFD simulations take hours, or sometimes days, this is an unrealistic approach. The system of transport equations in CFD is not linear with respect to the parameters, and therefore an alternative method is needed, which is the motivation for the present paper. A linear approximation of a non-linear function leads to a poor estimate of the mean output value [6]. Thankfully there are several alternative methods to the MC method, which are more efficient, i.e. requires fewer simulations, to determine how the uncertainties in the input parameters translate, or propagate, into an uncertainty of the output result. The low convergence rate of the MC method may be improved by partially controlling the distribution of samples deterministically using stratified sampling techniques, such as the Latin Hypercube Sampling (LHS) [7], a method born in the field of safety of nuclear reactors. Alternatively, large MC ensembles can be allowed for by substituting the complex model with a simple approximation, as in the Response Surface Methodology (RSM), also explained in [7]. Polynomial Chaos (PC) method is a non-sampling based method to determine the uncertainty, originally proposed by [8]. More recent work using PC is exemplified in [9, 10]. Yet another approach is the sensitivity based Adjoint Method (AM) [11]. Very rarely the selection of samples is calculated and optimized with deterministic rules. This paper will introduce the method of Deterministic Sampling (DS) [1], which has its roots in the field of signal processing and the Unscented Kalman Filter (UKF) [12]. The method will be used to estimate how the uncertainties, in some of the input parameters, propagate in the results in a CFD calculation. The main advantage of the DS method is that it requires relatively few samples, which translates into much less computational work compared with most other mentioned methods, for the equivalent accuracy. A known continuous probability distribution function of the input parameter, can be approximated by an ensemble with discrete samples, if they provide equally valid representation of the relevant statistics, such as the mean value and the higher statistical moments, i.e. the variance, the skewness, the flatness etc. In practice, the knowledge of the higher moments is often limited. They are assumed known when a continuous probability distribution is introduced for an input parameter. Many times the available information for the distribution is limited to its mean and its covariance or just its standard deviation. In such a case, it will suffice to have a few discrete samples in an ensemble to calculate the mean and the standard deviation, compared to the many thousands of samples that would be required using the MC-method. Section 2 begins by showing a simple example where the performance of the DS method is compared with the MC method for a non-linear model. In Section 3, three different sets of ensembles are prepared to be used in a CFD example which contains five independent uncertain parameters. Section 4 describes the CFD example, and Section 5 presents the results and analysis of the uncertain quantification in the simulations. Finally, in Section 6, we summarize our contributions in the conclusions.

2. A SINGLE PARAMETER EXAMPLE USING THE DETERMINISTIC SAMPLING METHOD

Let h be a non-linear function $f(q)$ where q is an uncertain parameter:

$$h = f(q) = q^4 \quad (1)$$

Assume the parameter, q, has a normal probability density function (pdf) distribution with an expected value, $\langle q \rangle = 2$, and a standard deviation, $\sigma_q = 0.4$. A MC simulation will be performed for this case and will be compared with a DS simulation where initially only two samples are used. The fundamental idea of the Unscented Transform in the UKF is that it is easier to approximate a probability distribution than it is to approximate an arbitrary nonlinear function or transformation [12]. In the original Unscented Transform [12] for one parameter, two sigma points in an ensemble can be chosen as;

$$q_1 = \langle q \rangle - \sigma_q \quad \text{and} \quad q_2 = \langle q \rangle + \sigma_q \quad (2)$$

The mean value, and the standard deviation of these two points, is identical to the mean and standard deviation of the continuous normal pdf distribution. Propagating these two samples will give;

$$h_1 = (2 - 0.4)^4 \quad \text{and} \quad h_2 = (2 + 0.4)^4 \quad (3)$$

The output average and standard deviation from the ensemble, containing two samples, DS2;

$$\langle h_{DS2} \rangle = 0.5(h_1 + h_2) = 19.8656 \quad (4)$$

$$\sigma_{q_{DS2}} = \sqrt{0.5(h_1 - \langle h_{DS2} \rangle)^2 + 0.5(h_2 - \langle h_{DS2} \rangle)^2} = 13.312 \quad (5)$$

Figures 1 and 2 are illustrations of how many samples the MC method requires to reach their designated values of the statistical moments. As can be seen in both figure 1 and 2, it takes thousands of samples before the statistics finds their asymptotic value which are;

$$\langle h_{MC} \rangle = 19.87 \quad \text{and} \quad \sigma_{q_{MC}} = 15.43 \quad (6)$$

A surprisingly good agreement between the computed expected values from the DS and the MC method can be observed. However, the standard deviations differ.

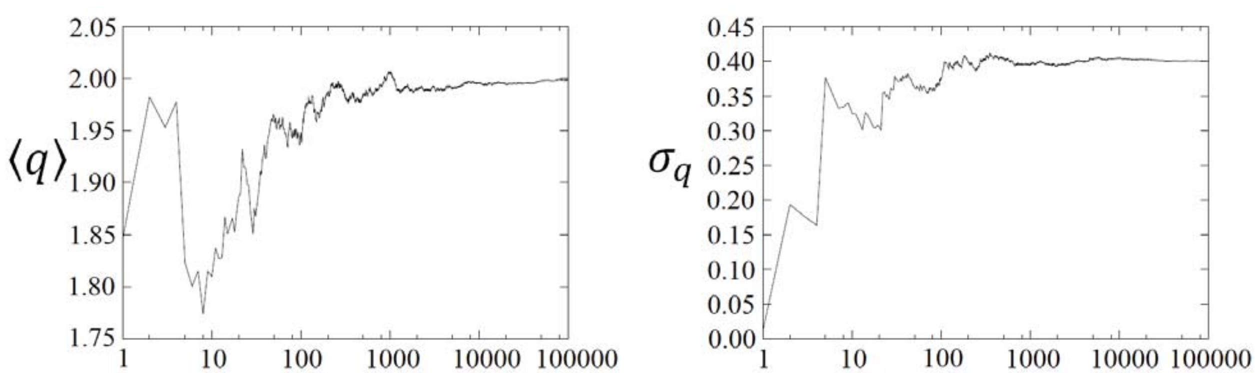

Figure 1. The expected value of the parameter q and the standard deviation as a function of number of random samples.

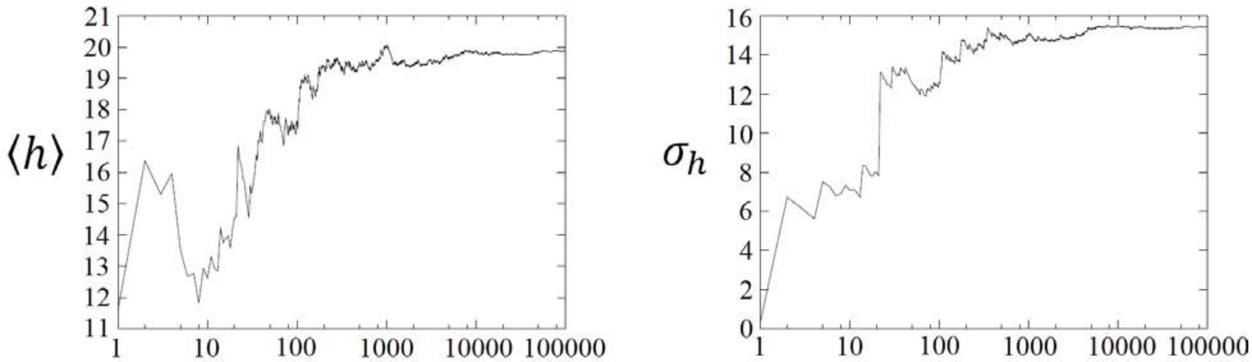

Figure 2. The expected value of the parameter h and the standard deviation as a function of number of random samples.

By selecting only two samples in the DS-method only the first and second moment will agree with the normal pdf for q. For the MC example a continuous probability distribution function was assigned, where all the higher moments are assumed known. The third moment, the skewness, is zero for a normal pdf. The fourth moment normalized with the standard deviation to the power of four, sometimes called the flatness, is equal to three. An example of a DS ensemble that would satisfy the first-, second-, third-, and fourth moments could read;

$$q_1 = \langle q \rangle - \sqrt{3}\sigma_q \ , \ q_2 = \langle q \rangle, \text{ and } q_3 = \langle q \rangle + \sqrt{3}\sigma_q \tag{7}$$

Each sample has a weight, W, associate with them, see [12, 13], chosen in this example to be;

$$W_1 = 1/6, \ W_2 = 4/6, \text{and } W_3 = 1/6 \tag{8}$$

Then the four lower moments are satisfied,

$$\langle q \rangle = W_1 q_1 + W_2 q_2 + W_3 q_3 \tag{9}$$
$${\sigma_q}^2 = W_1(q_1 - \langle q \rangle)^2 + W_2(q_2 - \langle q \rangle)^2 + W_3(q_3 - \langle q \rangle)^2 \tag{10}$$
$$0 = W_1(q_1 - \langle q \rangle)^3 + W_2(q_2 - \langle q \rangle)^3 + W_3(q_3 - \langle q \rangle)^3/{\sigma_q}^3 \tag{11}$$
$$3 = W_1(q_1 - \langle q \rangle)^4 + W_2(q_2 - \langle q \rangle)^4 + W_3(q_3 - \langle q \rangle)^4/{\sigma_q}^4 \tag{12}$$

The sum of the weights equals one.

$$W_1 + W_2 + W_3 = 1 \tag{13}$$

By using these three samples we get;

$$\langle h_{DS3} \rangle = W_1 h_1 + W_2 h_2 + W_3 h_3 = W_1 {q_1}^4 + W_2 {q_2}^4 + W_3 {q_3}^4 = 19.91 \tag{14}$$

$$\sigma_{h_{DS3}} = \sqrt{W_1(h_1 - \langle h_{DS3} \rangle)^2 + W_2(h_2 - \langle h_{DS3} \rangle)^2 + W_3(h_3 - \langle h_{DS3} \rangle)^2} = 15.37 \tag{15}$$

Both the mean and the standard deviation agree well with the results obtained from thousands of samples, in the MC-method. This is achieved with just three deterministic samples. The results are summarized in table I.

Table I. Expected values and standard deviations

Method	DS2	DS3	MC
Expected value	19.87	19.91	19.87
Standard deviation	13.31	15.37	15.43

To be able to use the MC method, a continuous probability density function was assumed. From it, all the higher moments can be calculated. In practical cases, it is not unusual to just have knowledge about the variance, possibly the covariance, and the mean value of a parameter. Applying a continuous probability density function as done in the MC-method is pure guesswork, and the DS method will suffice. In addition, by including more samples in the ensemble of the DS method, arbitrarily high statistical moments can be reproduced, if they are known.

3. DETERMINISTIC SAMPLING ENSEMBLES FOR MULTIPLE PARAMETERS

The example in Section 4 will contain five uncertain parameters. They are the following turbulence model constants; c_μ, $C_{\varepsilon 2}$, σ_k, κ, and B, present in the standard k-ε-model [14]. These five parameters, see table II, were used in the study by [15] where the LHS method was applied to quantify the uncertainty of the turbulence model.

Table II. Uncertain turbulence model parameters

Parameter	Probability density function		Constants
c_μ (q_1)	Weibull	$g(x) = \frac{\alpha_w}{\beta_w}\left(\frac{x}{\beta_w}\right)^{\alpha_w - 1} e^{-(x/\beta_w)^{\alpha_w}}$	$\alpha_w = 45.54$ $\beta_w = 8.77 \times 10^{-2}$
$C_{\varepsilon 2}$ (q_2)	Beta	$g(x) = \frac{(x - A_1)^{p-1}(A_2 - x)^{q-1}}{\int_{A_1}^{A_2}(x - A_1)^{p-1}(A_2 - x)^{q-1}dx}$	$A_1 = 1.80$, $A_2 = 2.20$ $p = 1.20$, $q = 2.0$
σ_k (q_3)	Normal	$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x - \langle\sigma_k\rangle)^2}{2\sigma^2}}$	$\langle\sigma_k\rangle = 1.0$ $\sigma = 1.67 \times 10^{-2}$
κ (q_4)	Normal	$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x - \langle\kappa\rangle)^2}{2\sigma^2}}$	$\langle\kappa\rangle = 0.41$ $\sigma = 4.89 \times 10^{-3}$
B (q_5)	Normal	$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x - \langle B\rangle)^2}{2\sigma^2}}$	$\langle B\rangle = 5.2$ $\sigma = 0.10$

3.1 The Ensembles

Three different ensembles of samples will be prepared which are going to be used in the DS method. The first ensemble will be selected so the samples only satisfy the expected value and the standard deviation for all the parameters. In the second and third ensemble, the selected samples will in addition satisfy the third- and fourth statistical moment of the parameters. For the latter two ensembles, non-uniform weights will be associated with the samples.

3.2 The Binary Ensemble

The first ensemble will be constructed in accordance with the binary ensemble, described in [1]. For the five parameters the so-called excitation matrix becomes;

$$\hat{V}_{BIN} = \begin{pmatrix} +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ -1 & +1 & +1 & -1 & -1 & +1 & +1 & -1 \\ -1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 \end{pmatrix} \tag{17}$$

This is a table of canonical samples. The eight columns are the necessary samples for the five different parameters in the rows. A full binary table for five parameter would contain 32 samples, but the selected eight samples suffice to describe the expected zero means and unit standard deviations of the parameters [1]. The binary sample values need to be transformed, to satisfy the actual parameter statistics of the DS ensemble Σ. This is done using [1, eq.3.4];

$$\Sigma \equiv \langle q \rangle \otimes 1_{1\times m} + U^T S U \hat{V} \tag{18}$$

The vector with the five uncertain parameters are, $q = (q_1, q_2, q_3, q_4, q_5)^T$, and m = 8 is the number of samples. The matrix U was introduced in [1] to allow the parameters to have a dependency, in which case there would be non-zero entries in the off-diagonal elements in the covariance matrix for the parameters. In our example the parameters are assumed to be independent, i.e. $U = I$, where I is the identity matrix. S is the standard deviation scaling matrix, which allowed the ensemble to be written in a generic form. The eight samples in the ensemble, Σ, can be written explicitly. See table III. Table III constitutes the first ensemble. In the CFD example of Section 4, simulations will be performed using these eight combinations of parameter values as samples.

Table III. Samples in binary ensemble

sample	q_1	q_2	q_3	q_4	q_5
1	$\langle q_1\rangle + \sigma_1$	$\langle q_2\rangle + \sigma_2$	$\langle q_3\rangle + \sigma_3$	$\langle q_4\rangle - \sigma_4$	$\langle q_5\rangle - \sigma_5$
2	$\langle q_1\rangle - \sigma_1$	$\langle q_2\rangle + \sigma_2$	$\langle q_3\rangle + \sigma_3$	$\langle q_4\rangle + \sigma_4$	$\langle q_5\rangle - \sigma_5$
3	$\langle q_1\rangle + \sigma_1$	$\langle q_2\rangle - \sigma_2$	$\langle q_3\rangle + \sigma_3$	$\langle q_4\rangle + \sigma_4$	$\langle q_5\rangle + \sigma_5$
4	$\langle q_1\rangle - \sigma_1$	$\langle q_2\rangle - \sigma_2$	$\langle q_3\rangle + \sigma_3$	$\langle q_4\rangle - \sigma_4$	$\langle q_5\rangle + \sigma_5$
5	$\langle q_1\rangle + \sigma_1$	$\langle q_2\rangle + \sigma_2$	$\langle q_3\rangle - \sigma_3$	$\langle q_4\rangle - \sigma_4$	$\langle q_5\rangle + \sigma_5$
6	$\langle q_1\rangle - \sigma_1$	$\langle q_2\rangle + \sigma_2$	$\langle q_3\rangle - \sigma_3$	$\langle q_4\rangle + \sigma_4$	$\langle q_5\rangle + \sigma_5$
7	$\langle q_1\rangle + \sigma_1$	$\langle q_2\rangle - \sigma_2$	$\langle q_3\rangle - \sigma_3$	$\langle q_4\rangle + \sigma_4$	$\langle q_5\rangle - \sigma_5$
8	$\langle q_1\rangle - \sigma_1$	$\langle q_2\rangle - \sigma_2$	$\langle q_3\rangle - \sigma_3$	$\langle q_4\rangle - \sigma_4$	$\langle q_5\rangle - \sigma_5$

3.3 Ensemble with Weighted Samples for Higher Statistical Moments

Unlike the first ensemble in 3.2 which only satisfies the first and second statistical moments the second ensemble will be assembled to satisfy the first four statistical moments. In the example for one parameter (Section 2), an ensemble with three samples was used to satisfy the first four statistical moments. This ensemble can be reused for the three parameters q_3, q_4, and q_5, which all have normal distributions according to table II. The parameter q_3 will be given the following three samples;

$$q_{31} = \langle q_3 \rangle - \sqrt{3}\sigma_{q_3}\ ,\ \ q_{32} = \langle q_3 \rangle,\ \text{and}\ \ q_{33} = \langle q_3 \rangle + \sqrt{3}\sigma_{q_3} \tag{19}$$

Each sample has a weight associated with them,

$$W_{31} = 1/6,\ \ W_{32} = 4/6, \text{and}\ \ W_{33} = 1/6 \tag{20}$$

A similar ensemble construction can be used for the parameters four and five, q4 and q5. The fact that three well-chosen samples can satisfy all the statistical moments up to the fourth moment is a special case. Five equations need to be satisfied, and in a general case five samples will be needed. For the first parameter q_1 the requirements are;

$$\begin{aligned} 1 &= \Sigma_{i=1}^{5} W_{1i} \\ 0 &= \Sigma_{i=1}^{5} W_{1i}(q_{1i} - \langle q_1 \rangle) \\ 1 &= \Sigma_{i=1}^{5} W_{1i}(q_{1i} - \langle q_1 \rangle)^2 / variance[q_1] \\ skewness[q_1] &= \Sigma_{i=1}^{5} W_{1i}(q_{1i} - \langle q_1 \rangle)^3 / variance[q_1]^{3/2} \\ flatness[q_1] &= \Sigma_{i=1}^{5} W_{1i}(q_{1i} - \langle q_1 \rangle)^4 / variance[q_1]^2 \end{aligned} \tag{21}$$

This can be achieved in two ways. One, where the weights are assumed to be equal, $W_{1i} = 0.2$, and solve the non-linear system represented by equations 21 to find q_{1i}. However, it is not always possible to find a solution. With the present Weibull distribution, we cannot find a solution with real roots for all q_{1i}. The second way is to choose the samples q_{1i} first, and then find the weights, W_{1i}, in the set of equations 21. The system of equations is linear with respect to W_{1i}, meaning that a solution can always be found. The only possible way forward appears to be, to first select the samples, and then find the weights to satisfy the system of equations 21. Having introduced the weighted samples, it allows some freedom how the samples are chosen. The five samples could be chosen in a first attempt to be;

$$q_{11} = \langle q_1 \rangle - \sigma_1, q_{12} = \langle q_1 \rangle - \sigma_1/2,\ \ q_{13} = \langle q_1 \rangle, q_{14} = \langle q_1 \rangle + \sigma_1/2,\ \text{and}\ \ q_{15} = \langle q_1 \rangle + \sigma_1 \tag{22}$$

By calculating the statistical moments of the Weibull distribution, and inserting them in equation 21 together with the suggested samples in equation 22, the weights, W_{1i}, can be computed.

$$W_{11} = 3.6519,\ \ W_{12} = -11.2759, W_{13} = 14.9162,\ W_{14} = -8.6124, \text{and}\ \ W_{15} = 2.3202 \tag{23}$$

According to [12] the weights do not have to be in the range [0,1] as long as their sum equals one. However, even statistical moments of higher order, like the 6th or 8th moment for which we impose no requirement, could become negative by allowing negative weights. This is clearly prohibited and has potentially an influence on the results. One such problem, associated with negative weights, is identified in the results in Section 5, but works in all other cases. The remaining ensemble to be selected is for the second parameter, q2 see table II, which has an assumed beta pdf distribution in the work by [15]. An identical ensemble can be selected for the parameter

$$q_{21} = \langle q_2 \rangle - \sigma_2,\ \ q_{22} = \langle q_2 \rangle - \sigma_2/2,\ \ q_{23} = \langle q_2 \rangle,\ \ q_{24} = \langle q_2 \rangle + \sigma_2/2,\ \text{and}\ \ q_{25} = \langle q_2 \rangle + \sigma_2 \tag{24}$$

By solving a similar equation system to 21, but for the parameter q_2, the weights, W_{2i}, assume the following values;

$$W_{21} = 3.1388,\ \ W_{22} = -11.4132, W_{23} = 18.4069,\ W_{24} = -13.1294, \text{and}\ \ W_{25} = 3.9969 \tag{25}$$

All the samples, q_{ij} , and associated weights, W_{ij} , for the five parameters, q_i, have now been calculated to represent their individual marginal pdf's. All the nineteen samples can be lumped into a grand ensemble. Five of the samples are identical, namely the zero or center sample. The zero or center sample is the sample that represents the average parameter value, i.e., q_{13}, q_{23}, q_{32}, q_{42}, and q_{52}.Thus the number of samples is reduced to fifteen by adding up the five weights for the zero sample. The sum of all weights from the five parameters equal five, however, the sum should be one. This can be adjusted by reducing the weight for the zeroth sample by four. The zero sample weight will then become;

$$W_{13} + W_{23} + W_{32} + W_{42} + W_{52} - 4 = 14.9162 + 18.4069 + \frac{4}{6} + \frac{4}{6} + \frac{4}{6} - 4 = 31.3231 \tag{26}$$

This is the second total ensemble and it now satisfies all the marginal pdf's, as well as the requirement on the sum of weights to be one.

3.4 Ensemble with Annealed Weighted Samples

Non-uniform weights of the samples provide a lot of freedom to select the samples. As pointed out earlier, using negative weights might produce unrealistic negative values for higher, even, statistical moments. This can be remedied by finding samples such that their weights are always positive. This is not a trivial matter and will lend itself to further development. The samples for q_3, q_4, and q_5 all have positive weights associated with them. These samples can be kept as is. To avoid negative weights in the ensemble for the parameters, q_1, and q_2, the samples must be modified. Finding an analytical solution on how the samples should be modified has proven to be difficult. As mentioned above, choosing the weights, and then solving the non-linear equations for q_{ij} does not always produce a solution with real roots for all samples. A trial and error procedure has been adopted to randomly generate the samples in the ensemble, whereafter the weights are analytically calculated. If the weights are all positive, and are more evenly distributed than in the previous attempt, the new best ensemble replaces the old best ensemble. This process is called *annealing*.

The Weibull distribution, with its first four statistical moments, can also be represented by the following deterministic samples;

$$q_{11} = 0.0809,\ q_{12} = 0.0868,\ q_{13} = 0.0869,\ q_{14} = 0.0870, \text{ and } q_{15} = 0.0907 \tag{27}$$

and their associated weights,

$$W_{11} = 0.1150,\ W_{12} = 0.1153, W_{13} = 0.5393, W_{14} = 0.1155, \text{ and } W_{15} = 0.1150 \tag{28}$$

for the first parameter, q_1 (c_μ) in table II.

With the same process, an ensemble can be found to represent the statistical moments of q_2 $(C_{\varepsilon 2})$. It will read;

$$q_{21} = 1.64, q_{22} = 1.9388,\ q_{23} = 1.9444, q_{24} = 1.9498, \text{ and } q_{25} = 2.3318 \tag{29}$$

with their associated weights,

$$W_{21} = 0.0956,\ W_{22} = 0.2724, W_{23} = 0.2716, W_{24} = 0.2706, \text{ and } W_{25} = 0.0898 \tag{30}$$

Again, the sum of all weights from five parameters equals five. As before the weight for the zero sample will be reduced by four. The zero sample weight will in this case become;

$$W_{32} + W_{42} + W_{52} - 4 = \frac{4}{6} + \frac{4}{6} + \frac{4}{6} - 4 = -2 \tag{31}$$

This particular weight is negative, but for the individual marginal distributions there are no negative weights. This will eliminate the risk of producing negative higher even statistical moments. The -4 is introduced to not count the center sample more than once. This is the third total ensemble, containing 17 samples, which satisfies all the marginal pdf's, and also $W = 1$. The next sections will describe and compare the simulation results and their uncertainty, for a well-known fluid problem, using the three ensembles discussed in this section.

4. BACKWARD FACING STEP EXAMPLE

The flow over a backward-facing step in figure 3 is a generic problem and has been studied numerically by many in the CFD community. An example where uncertainty quantification is performed on this problem is the work by [15]. They used the LHS method with 400 samples.

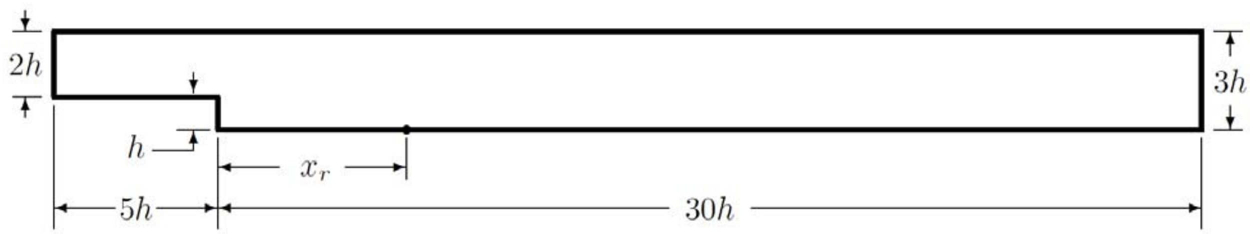

Figure 3. Backward-facing step geometry

In the present three ensembles, 8, 15, and 17 samples are used. To be able to make a direct comparison with their results, this study will use the same software, OpenFOAM [16], the same turbulence model [14], alter the same parameters in it, and use an identical computational grid. The difference will be in the method used to propagate the uncertainty of the turbulence model constants through the system of transport equations. The result from the experiment by Kim [17] is the reference to judge if the numerical results are consistent with the measured data. As previously mentioned, the numerical result, and the experimental results, are consistent if there is an overlap of the two uncertainty bars. Otherwise they are inconsistent. For example, the mean value of numerical results can be outside the uncertainty bar of the experiment, but still be consistent as long as there is an overlap of the two uncertainty bars. The seven constants in the turbulence model [14] will be viewed as uncertain. They are; c_μ, σ_k, σ_ε, $C_{\varepsilon1}$, $C_{\varepsilon2}$, κ, and B. Five of them are listed in table II, and are assumed to be independent. According to [18] and [15], σ_ε and $C_{\varepsilon1}$ are functions and are dependent on some of the five independent parameters.

$$\sigma_\varepsilon = \frac{\kappa^2}{c_\mu^{0.5}(C_{\varepsilon2} - C_{\varepsilon1})} \tag{32}$$

$$C_{\varepsilon1} = \frac{1}{(P/\varepsilon)^*} C_{\varepsilon2} + \frac{(P/\varepsilon)^* - 1}{(P/\varepsilon)^*} \quad \text{where} \quad (P/\varepsilon)^* \approx 1.8 \tag{33}$$

In many implementations of the k-ε model a constant E, is used in the wall function instead of B.

$$E = e^{\kappa B} \tag{34}$$

The step height, h, in Figure 3 is resolved with 21 equally spaced control volumes with an aspect ratio of one. The Reynolds number in the present case is, 1.32×10^5, based on the outlet channel height and the inlet velocity. The inlet turbulence intensity is set to 5%, and a uniform inlet velocity profile is assumed. The inlet dissipation rate of kinetic energy is set to $\varepsilon = c_\mu^{3/4} k^{3/2}/l$, where the turbulent length-scale is estimated to be $l = 0.1 \times 2h$.

5. RESULTS

Each ensemble will require as many simulations as there are samples in it. The result from a simulation is then multiplied by the weight the sample was associated with. Results will be presented for the velocity, U, the turbulent kinetic energy, k, the pressure coefficient on the walls, C_p, and the reattachment length, x_r. The mean and the standard deviation for these quantities can be calculated from;

$$\langle R \rangle = \sum_{i=1}^{N} W_i R_i \quad \text{, and } \sigma_R = \sqrt{\sum_{i=1}^{N} W_i (R_i - \langle R \rangle)^2} \tag{35}$$

where R can be any of U, k, C_p, or x_r, and N is the number of samples in the ensemble. The results can be directly compared with the ones produced by [15], since the case, software, and mesh size are the same. No MC simulation has been attempted since the number of samples required would be discouragingly many. Not only would the CPU usage be large, but it would also be difficult to check if all performed simulations are correct.

5.1 The velocity profiles

The average streamwise velocity component was measured at several stations downstream of the step by [17]. Some of them are shown in the figure 4. The computed non-dimensional velocity profiles using the three different ensembles presented as $E[U/U_0] \pm 3\sigma/U_0$, where U_0 is the velocity at the center of the inlet. The results for the three ensembles are indistinguishable. The uncertainty in the hot-wire measurements quoted in, [17] as being 1%. At the locations $x/h = 5.33$ and 6.22 the simulated results are consistent with the measured velocity profiles as they lie in the interval $\pm 3\sigma$. At the other locations there is only a partial overlap, or no overlap, of the two uncertainties. Clearly, the uncertainty of the simulated results does not encompass the measured results. Several explanations are possible. Systematic uncertainties can be present in either the measurements and/or in the turbulence model. Physical experiments often have difficulties to produce the intended two-dimensional flow field and this is a systematic uncertainty. The assumed uncertainty of the turbulence model constants might also be too optimistic. The uncertainty of the inlet conditions does play a role, but is not included in this study. Many would blame the turbulence model and switch to a more elaborate model. However, all turbulence models have yet failed to work under all circumstances. Regardless of which turbulence model is used, there will be discrepancies with the reference data. A logical step forward, from an engineering point of view, is to calibrate the turbulence model for the case at hand. This means giving up on the idea of ever finding a correct model, but settle for a model, with calibrated constants, which will work, i.e. give results with an acceptable tolerance level, for a particular case. A better turbulence model will just imply that the adjustment to the constants needs to be smaller.

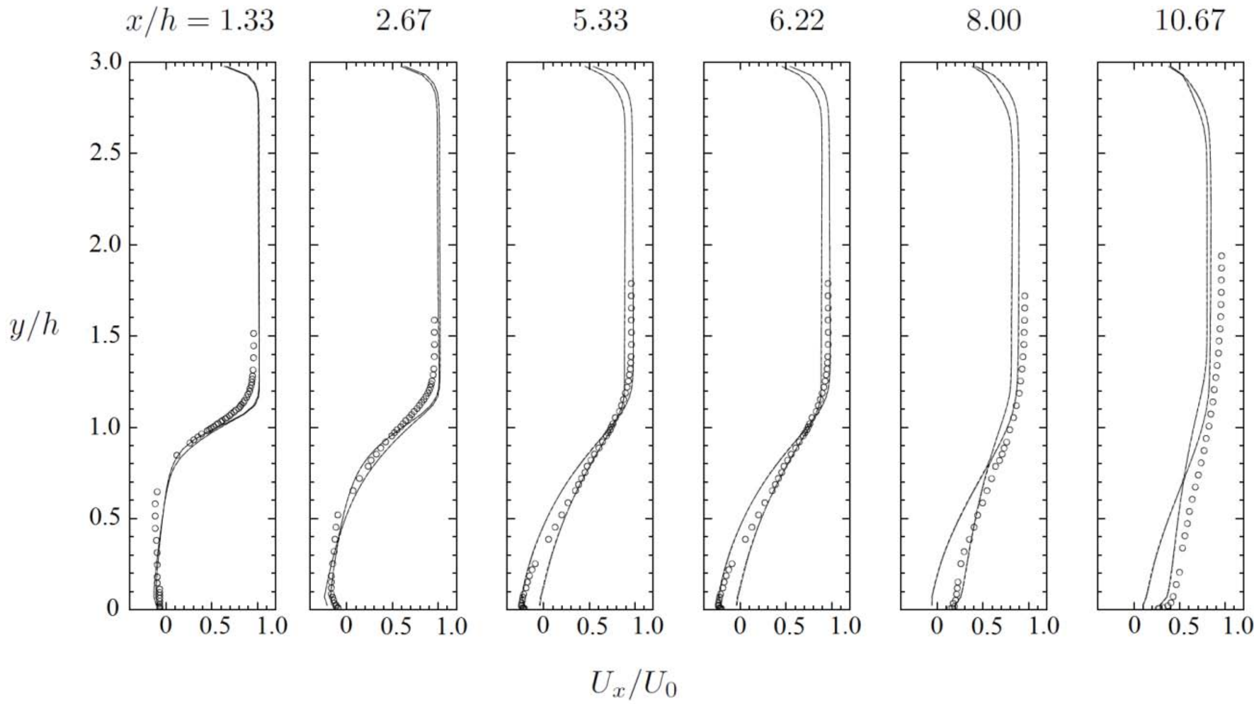

Figure 4. Streamwise velocity component, $\pm 3\sigma$, maximum and minimum. Results using the binary ensemble (solid lines), the 2nd ensemble (dashed lines), and the 3rd ensemble (dotted lines). Measurements [17] (circles). The results from the three ensembles are indistinguishable.

5.2 The turbulent kinetic energy profiles

The kinetic energy profiles shown in figure 5, the numerical results show only a slight difference for the three ensembles, at stations downstream of the step. The binary ensemble produces the smallest uncertainty. The measured data of the Reynolds' stress components have a 10% to 20% uncertainty associated with them. Only two of the three normal Reynolds stresses are measured, $(u')^2$, and , $(v')^2$. In order to make a comparison of the turbulent kinetic energies, the third component, $(w')^2$, has to be estimated. Its value is conjectured to be somewhere in between the value of the other two components. This makes the uncertainty large for the kinetic energy derived from the experiments, see figure 5. In most areas the measured and computed results are consistent, i.e. there is an overlap of the two uncertainties.

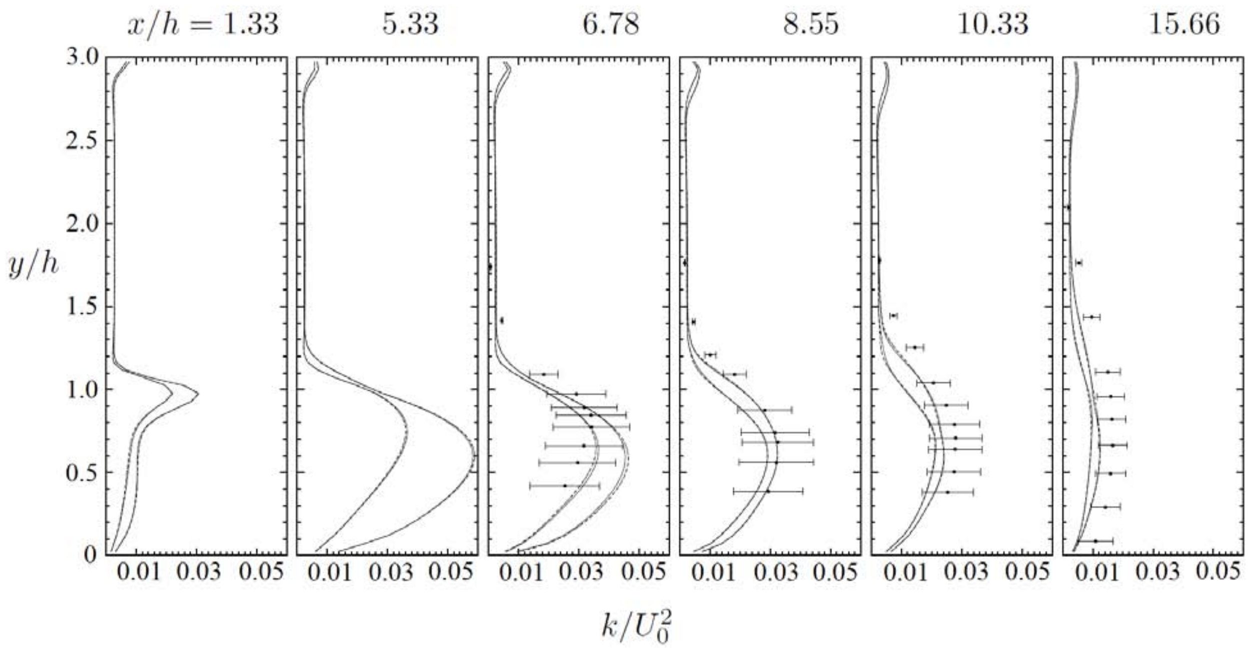

Figure 5. Turbulent kinetic energy, $\pm 3\sigma$, maximum and minimum. Results using binary ensemble (solid lines), 2nd ensemble (dashed lines), and 3rd ensemble (dotted lines). Measurements [17] (uncertainty bars). Results from the three ensembles are for the most part indistinguishable.

5.3 The reattachement length

The reattachment length, x_r, is illustrated in figure 3. In the measurements [17], the length was observed to be $x_r/h = 7 \pm 1$, using tufts. The simulations, using the binary ensemble, produced a reattachment length, $(x_r \pm 3\sigma)/h = 7.06 \pm 1.79$. The third ensemble gave a value, $(x_r \pm 3\sigma)/h$, of 7.05 ± 1.86. The results from the measurements and the simulations are thus consistent in both cases. The ensemble that satisfies higher moments indicates a larger uncertainty in the results. Interestingly, the second ensemble produced a non-physical negative variance for x_r. Using the negative weights associated with this ensemble, the average value became, $x_r/h = 7.66$, which is larger than the largest value observed in the 15 samples, which was $x_r/h = 7.65$. This clearly demonstrates that it can be problematic to use negative weights, even if it was tolerated in the original UT-formulation [12].

5.4. The pressure coefficient on walls

The pressure coefficient is defined as;

$$C_p = 2(P - P_0)/\rho U_0^2 \tag{36}$$

where P_0, is the pressure at the center of the inlet. The distribution of the pressure coefficient on the lower and upper wall, after the step, is shown in figure 6. The uncertainty of the measurement of C_p is quoted to be 1%. The largest uncertainties in the present simulations appear around the reattachment zone. The results from second and third ensemble are very similar, and the uncertainty is larger than for the binary ensemble that just satisfies the first and second statistical moment of the input parameters. At most locations the experimental results are consistent with the results from the theoretical model.

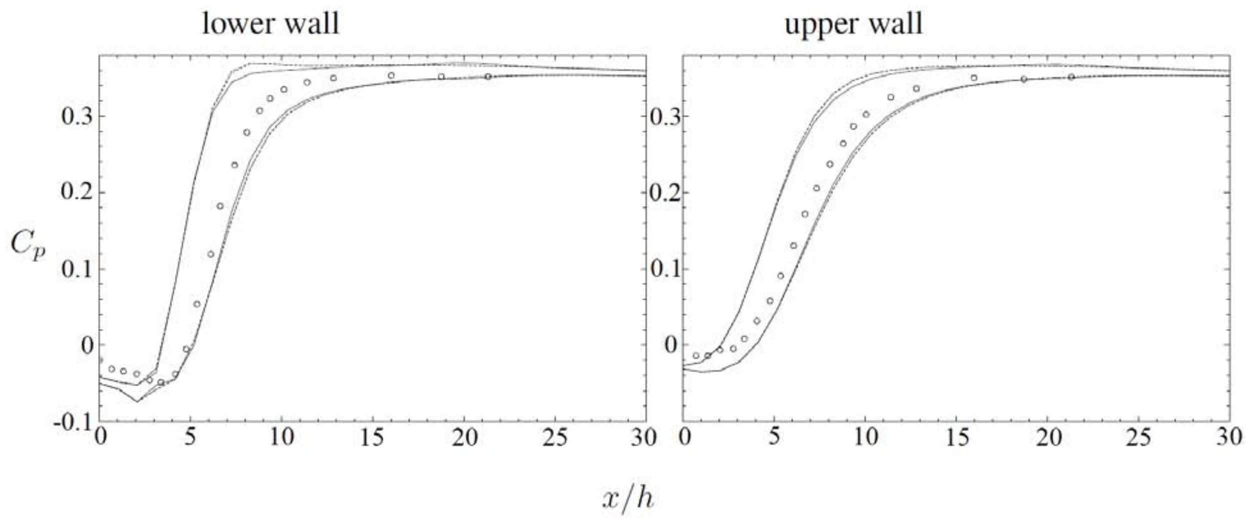

Figure 6. Pressure coefficient distribution, $\pm 3\sigma$, on walls after the step. Results using binary ensemble (solid lines), 2nd ensemble (dashed lines), and 3rd ensemble (dotted lines). Measurements [17] (circles). The results from 2nd and 3rd ensemble are indistinguishable.

6. CONCLUSIONS

This paper has presented a new method to propagate uncertain input parameters to results produced with CFD. The parameters have been chosen to be the turbulence model constants of the standard k-ε model [14]. Five of the parameters are independent, and have an assumed known continuous pdf. The remaining turbulence model constants are assumed to be functions of the five parameters. Instead of using the continuous pdf, these parameters have been approximated with a discrete number of samples into an ensemble that can produce sufficiently high statistical moments. Based on the results of this study, it appears to be sufficient to satisfy the first four statistical moments. Satisfying yet higher statistical moments did not lead to increased uncertainty. It appears that ensembles containing samples that are associated with negative weights, can give reasonable results. However, this should be treated with care, since it can potentially produce unrealistic, non-bounded, average values and non-physical negative statistical moments of even order. Eight samples are enough to represent the mean and the covariance for the five parameters. The skewness and flatness can be represented using an extra seven to nine samples. This produces a more reliable uncertainty quantification of the results, but at the expense of an increased computational cost. Still the DS method is very lean. The LHS approach by [15] required 400 samples, which translates into 400 individual simulations.

ACKNOWLEDGMENTS

Financial support for the National Metrology Program from VINNOVA, The Swedish Governmental Agency for Innovation Systems, is gratefully acknowledged.

REFERENCES

1. J.P. Hessling, "Deterministic Sampling for propagating model covariance". *SIAM/ASA Journal on Uncertainty Quantification*, **1**, pp. 297-318, (2013)
2. J.P. Hessling, "Identification of Complex Models", *SIAM/ASA J. Uncertainty Quantification*, **2**(1), pp. 717–744, (2014)
3. JCGM 100:2008. *Evaluation of measurement data - Guide to the expression of uncertainty in measurement*, Joint Committee for Guides in Metrology.
4. N. Metropolis, S. Ulam. "The Monte Carlo Method". *Journal of the American Statistical Association*, **44,** pp. 335-341 (1949)
5. Rubenstein RY, Kroese DP. *Simulation and the Monte Carlo Method*, 2nd Ed. New York: John Wiley & Sons Inc., (2007)
6. J.P. Hessling, T. Svensson, "Propagation of uncertainty by sampling on confidence boundaries". *Int J. Uncertainty Quantification*, **3**, pp. 421-444, (2013)
7. J. Helton, F. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems". *Reliability Engineering and System Safety*, **81**,pp. 23-69, (2003)
8. N. Wiener, The Homogeneous Chaos. *American Journal of Mathematics*, **60**, pp. 897-936, (1938)
9. J. A. S. Witteveen, "Efficient and Robust Uncertainty Quantification for Computational Fluid Dynamics and Fluid-Structure Interaction". PhD Dissertation, Stanford University, (2009)
10. M.P. Pettersson, "Uncertainty Quantification and Numerical Methods for Conservation Laws". PhD Dissertation. Stanford University, (2012)
11. Q. Wang, "Uncertainty Quantification for Unsteady Fluid Flow Using Adjoint-based Approaches". PhD Dissertation. Stanford University, (2009)
12. S. Julier, J. Uhlmann, "Unscented filtering and nonlinear estimation". *Proceedings IEEE,* **92** ,pp. 401-422, (2004)
13. L. Angrisani et al., "Unscented transform: A powerful tool for measurement uncertainty evaluation", *IEEE Transactions on instrumentation and measurement*, **55**, pp. 737-743, (2006)
14. W.P. Jones and B.E. Launder, "The prediction of laminarization with a two-equation model of turbulence", *International Journal of Heat and Mass Transfer*, 15, (1972)
15. M.C. Dunn, B. Shotorban, and A. Frendi, "Uncertainty Quantification of Turbulence Model Coefficients via Latin Hypercube Sampling Method". *Journal of Fluids Engineering- Transactions of The ASME*, **133**:041402, (2011)
16. OpenFOAM, The open source CFD toolbox, User Guide, version 1.7.1, 25th August (2010)
17. J. Kim, "Investigation of separation and reattachment of a turbulent shear layer: Flow over a backward-facing step". PhD Dissertation. Stanford University, (1978)
18. P.D.A. Platteeuw, "Application of the Probabilistic Collocation method to uncertainty in turbulence models", Master of Science Thesis, Delft University of Technology, (2008)