COMPARISON OF POLYNOMIAL CHAOS EXPANSION METHODS FOR UNCERTAINTY QUANTIFICATION IN CFD SIMULATIONS

Miquel A.*, Muñoz-Cobo, J. L., Berna C. and Escrivá A.
*Author for correspondence
Instituto de Ingenieria Energética,
Universitat Politècnica de València,
Valencia, 46023,
Spain,
E-mail: armivey@upv.es

ABSTRACT

Computational Fluid Dynamics (CFD) computer codes have proven to be a powerful tool in the analysis of all kinds of fluid systems. However, there is still a lack of practical methods for determining the uncertainty of their results, as most current techniques require performing too many simulations to be affordable in industrial-scale situations. One of the most promising methods for uncertainty quantification in computational fluid dynamics is Polynomial Chaos Expansion, a name that includes a variety of techniques, all based on the same mathematical background: projecting the system's response into a basis of orthogonal polynomials. This paper discusses the main advantages and drawbacks of three of these techniques, namely random sampling, Gaussian quadrature and linear regression, in terms of reliability, ease of use and computational costs. All three techniques were applied to simulations of the turbulent mixing of two streams of water inside a Y-shaped channel, and the results compared with experimental data. Results show that, in this test case, quadrature method provides more reliable results than the other two techniques, with a lower computational cost. Due to its robustness and low number of simulations required, Polynomial Chaos Expansion via quadrature methods might be suitable for most industrial CFD simulations.

INTRODUCTION

Computational Fluid Dynamics (CFD) computer codes enable researchers and engineers to simulate fluid systems with great detail, and have become an important research and design tool in many fields. However, as CFD codes became more capable and popular, the question on how to determine the uncertainty in the simulations raised. The development of uncertainty quantification methods suitable for CFD simulations is an area of active research, since most current techniques require performing too many simulations to be affordable in industrial-scale situations. One recent attempt to alleviate this problem is the use of Polynomial Chaos Expansion (PCE) methods, which, if properly implemented, can produce accurate results from a limited number of simulations.

NOMENCLATURE

B	[any]	Orthogonal basis
d	[-]	Normalized distance between experimental points and the uncertainty band
F	[any]	Generic integral
FOM	[-]	Figure of merit
f, g	[any]	Arbitrary functions
I	[-]	Turbulence intensity
m	[-]	Number of simulations / Sample size
n	[-]	Number of uncertain parameters
N	[-]	Number of terms in the PC expansion
p_i	[any]	Probability distribution function of the i-th parameter
q	[-]	Order of the PC expansion
$\vec{r}$	[m]	Position vector
R	[any]	System's response
U	[m/s]	Mean velocity
u'	[m/s]	RMS of velocity fluctuations
$\boldsymbol{x}$	[any]	Vector of uncertain parameters
X	[any]	Generic region of integration
x_i	[any]	i-th uncertain parameter

Special characters

α_i	[-]	i-th coefficient of the PC expansion
δ_{ij}	[-]	Kronecker delta
λ	[-]	Parameter controlling the shape of the inlet velocity profile
μ	[any]	Mean system's response
ω_i	[any]	Weights of Gaussian quadrature
ϕ_i	[any]	i-th basis function
σ	[any]	Variance of the system's response

The PCE foundational paper can be found in [1], while a complete review of the different PCE techniques, along with some applications to CFD simulations, is presented in [2], [3] and [4].

PCE methods can be broadly classified into two distinct categories: (i) *intrusive*, which require modification of the CFD code, and (ii) *non-intrusive*, which allow using CFD software as a 'black box' with no need for modifications. Since modification of the CFD code is often difficult, or even impossible in the case of commercial software, intrusive methods are rarely used in practice. For this reason, the present paper is limited to the non-intrusive techniques.

POLYNOMIAL CHAOS EXPANSION

PCE methods are based on projecting the system's response into an orthogonal basis of polynomials. Once all the expansion's coefficients are known, the mean and variance of the response can be readily calculated.

Let $R(\boldsymbol{x})$ be the system's response, depending on $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ uncertain parameters of the CFD simulation, such as the velocity at the inlet or some fluid properties. All uncertain parameters are assumed to be independent, with probability distribution functions $p_i(x_i)$. Also, let $B = \{\phi_i(\boldsymbol{x})\}_{i=0}^{\infty}$ be an orthogonal basis in the L^2 space, with inner product

$$\langle fg \rangle = \int_X f(\boldsymbol{x}) g(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} \tag{1}$$

The orthogonality condition implies that

$$\langle \phi_i \phi_j \rangle = \langle \phi_i^2 \rangle \delta_{ij} \tag{2}$$

where δ_{ij} is the Kronecker delta. Hence, the system's response can be expressed as a linear combination of the basis functions [4]:

$$R(\boldsymbol{x}) = \sum_{i=0}^{\infty} \alpha_i \phi_i(\boldsymbol{x}) \tag{3}$$

In practice, this infinite series must be truncated at some order q. The total number of terms in the expansion is then given by the number of uncertain parameters, n, and the expansion order, q, as shown in [4]:

$$N = \frac{(n+q)!}{n!q!} \tag{4}$$

Thus, the expansion is written as [4]

$$R(\boldsymbol{x}) \approx \sum_{i=0}^{N-1} \alpha_i \phi_i(\boldsymbol{x}) \tag{5}$$

There are several methods for calculating the α_i coefficients: (i) spectral projection via random sampling, (ii) spectral projection via quadrature and (iii) linear regression. The details of these three methods are explained below.

The basis function must be selected such that they satisfy the orthogonality property (2). In the case of of uncertain parameters with uniform distributions, multidimensional Legendre polynomials must be used [5]. The uncertain parameters must be properly scaled in the range $[0, 1]$.

Once the expansion's coefficients are known, the mean and variance of the system's response can be obtained as follows. Using the orthogonality property (2) and choosing the basis functions such that $\phi_0 = 1$, we have that [3]

$$\mu = \alpha_0 \qquad \text{and} \qquad \sigma^2 = \sum_{i=1}^{N-1} \alpha_i^2 \langle \phi_i^2 \rangle \tag{6}$$

Spectral projection via random sampling

In this method, the α_i coefficients are calculated using [4]:

$$\alpha_i = \frac{\langle R\phi_i \rangle}{\langle \phi_i^2 \rangle} \tag{7}$$

In the above expression, the term $\langle \phi_i^2 \rangle$ can be readily obtained evaluating the following integral:

$$\langle \phi_i^2 \rangle = \int_X \phi_i^2(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} \tag{8}$$

The numerator in (7) is the expectation value of the product of $R(\boldsymbol{x})$ and $\phi_i(\boldsymbol{x})$. Hence, it can be approximated from a random sample of $\boldsymbol{x}$. If the sample size is m, then

$$\langle R\phi_i \rangle \approx \frac{1}{m} \sum_{j=1}^{m} R(\boldsymbol{x}_j) \phi_i(\boldsymbol{x}_j) \tag{9}$$

The main drawback of this method, as stressed in [4], is the large number of simulation required to obtain converged results. Although low-discrepancy sampling techniques, such as Latin Hypercube Sampling, can speed up convergence, the number of simulations needed is still too high for most practical applications. On the other hand, the quantity of simulations is independent from the number n of uncertain parameters. This can be an important advantage in problems where n is large, since in all other methods, namely quadrature and stochastic collocation, the number of simulations increases rapidly with n.

Spectral projection via quadrature

In this case, the terms $\langle R\phi_i \rangle$ are obtained using numerical methods to evaluate the following integral

$$\langle R\phi_i \rangle = \int_X R(\boldsymbol{x}) \phi_i(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} \tag{10}$$

An efficient method to calculate the above integral is Gaussian quadrature [6]. This method provides an optimal evaluation of integrals in the form

$$F = \int_X f(\boldsymbol{x}) p(\boldsymbol{x}) d(\boldsymbol{x}) \tag{11}$$

by evaluating $f(\boldsymbol{x})$ at m points, and expressing the value of the integral as

$$F \simeq \sum_{j=1}^{m} \omega_j f(\boldsymbol{x}_j) \tag{12}$$

where ω_j are weights to be calculated. The evaluation points $\boldsymbol{x}_j$ are the roots of the orthogonal polynomials in X with weighting function $p(\boldsymbol{x})$. Conveniently, the values of ω_j and $\boldsymbol{x}_j$ for most usual cases can be easily found in literature [7], so the evaluation of the integral is straightforward. Gaussian quadrature is optimal in the sense that equation (12) is exact if $f(\boldsymbol{x})$ is a polynomial of order up to $2m-1$. Hence, the more $f(\boldsymbol{x})$ resembles a polynomial of order $2m-1$ or below, the more accurate the evaluation of the integral is.

Using the properties of Gaussian quadrature, it can be shown that the minimum number of evaluation points (i. e., simulations) needed to obtain good accuracy is [4]

$$m = (q+1)^n \tag{13}$$

As shown in the equation above, the minimum number of simulations grows exponentially with n. For this reason, Gaussian quadrature is a convenient method for problems with a reduced number of uncertain parameters, but impractical if the number is moderately large. If dealing with high-dimensionality problems, alternative quadrature methods based on sparse arrays, such as Smolyak's [8], significantly reduce the amount of simulations required. However, applications of these techniques to CFD simulations are still scarce [9].

Linear regression

In this approach, the expansion coefficients are calculated from m simulations, solving the following system

$$\begin{pmatrix} \phi_0(\boldsymbol{x}_1) & \phi_1(\boldsymbol{x}_1) & \cdots & \phi_{N-1}(\boldsymbol{x}_1) \\ \phi_0(\boldsymbol{x}_2) & \phi_1(\boldsymbol{x}_2) & \cdots & \phi_{N-1}(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\boldsymbol{x}_m) & \phi_1(\boldsymbol{x}_m) & \cdots & \phi_{N-1}(\boldsymbol{x}_m) \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_0 \\ \vdots \\ \alpha_{N-1} \end{pmatrix} = \begin{pmatrix} R(\boldsymbol{x}_1) \\ R(\boldsymbol{x}_2) \\ \vdots \\ R(\boldsymbol{x}_m) \end{pmatrix} \tag{14}$$

At least $m = (n+q)!/(n!q!)$ simulations are needed to solve the equations. However, some authors recommend using twice this number in order to increase the robustness of the results [10]. In that case, the system will be overdetermined and can be solved using linear regression methods.

Simulation points can be selected in several ways, such as (i) random or Latin Hypercube Sampling [3], (ii) regular sampling [11] or (iii) according to the roots of some orthogonal polynomials, as in Gaussian quadrature [12].

EXPERIMENTAL SETUP

The experimental data was obtained from the work by Badillo et al.[11], performed at the GEMIX facility. Two identical streams of water were injected thorugh the legs of a Y-shaped channel (Figure 1). Several sensors located along the channel measured the velocity, turbulence kinetic energy and concentration profiles at various planes normal to the stream flow.

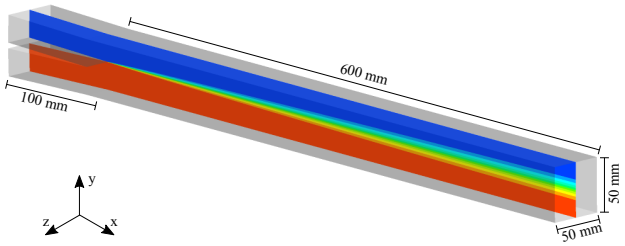

Figure 1. Geometry of the GEMIX experiment.

Water was injected at 23°C, at a rate of 1 kg/s in each leg. Before entering the channel, water passed through honeycombs and a series of grids to reduce turbulence and favor a uniform velocity profile. The angle between the two legs was 3°. The velocity and turbulent kinetic energy profiles were measured using Particle Image Velocimetry, while concentration was measured using Laser Induced Fluorescence. The experimental measurements were performed along 2 vertical profiles, located at the central plane of the channel, at $x = 70$ mm and $x = 450$ mm.

SIMULATION CONDITIONS

Simulations were performed using the Shear Stress Transport (SST) turbulence model of ANSYS CFX 15.0. Although for turbulent mixing processes, Large Eddy Simulation (LES) turbulence models might be more suitable than Reynolds Averaged Navier-Stokes (RANS) models, such as the SST, the high computational cost of LES makes RANS more suitable for industrial simulations.

Simulations were carried out in accordance with best practice guidelines [13] in order to ensure robust results . In particular, independence from the mesh, numerical precision and convergence criteria used was verified.

The mesh was composed of 210000 rectangular prisms, and was refined close to the walls and horizontal mid-plane. A mass flow rate of 1 kg/s was imposed at each leg's inlet, while a 0 Pa relative pressure was set at the outlet. The convergence criteria were normalized root mean square residuals below 10^{-5} for the conservation equations and global mass, momentum and energy imbalances below 0.1%. All the remaining options were set as default.

UNCERTAINTY QUANTIFICATION

Two uncertain parameters were considered: turbulence intensity at the inlet (I) and a parameter controlling the shape of the

inlet velocity profile (λ). Turbulence intensity is defined as usual

$$I = \frac{u'}{U} \tag{15}$$

where U is the mean velocity and u' the root mean square of velocity fluctuations. The definition of the inlet velocity profile was taken from [11]:

$$u(\vec{r}) = \left(\frac{\lambda}{u_d(\vec{r})} - \frac{1-\lambda}{u_u} \right)^{-1} \tag{16}$$

being u_d and u_u the fully developed and uniform velocity profiles. The developed profile was obtained from a previous CFD simulation, allowing the flow to develop until a stable regime was attained. The above expression shows that, for $\lambda = 0$, we have a uniform velocity profile at the inlet, while for $\lambda = 1$, the profile is fully developed. Hence, λ determines the shape of the velocity profile at the inlet.

The range of the uncertain parameters was $[1\%, 10\%]$ for the turbulence intensity and $[0, 1]$ for λ. A uniform distribution was assigned to each uncertain parameter.

RESULTS AND DISCUSSION

Due to space limitations, we will limit the discussion of the results to the velocity profile at $x = 450$ mm. However, similar results were obtained for the rest of variables and locations.

The results produced by each of the three PCE methods (i. e., random sampling, quadrature and linear regression) are shown in Figures 2–4, together with the number of simulations used. All PC expansions were truncated at first order.

Random sampling

This method required a large amount of simulations (90 in the example of Figure 2) to obtain reasonable results. Even with 90 simulations, convergence was not achieved, and results were highly dependent on the sampled values of the uncertain parameters. Moreover, as displayed in Figure 2, the uncertainty band was too wide for this method to be useful in practical engineering problems.

Gaussian quadrature

Gaussian quadrature required only 4 simulations to produce a realistic uncertainty band. As Figure 3 shows, the uncertainty band is significantly thinner than that obtained with the random sampling method, but still contains most of the experimental points.

Linear regression

The linear regression technique was applied to a regular grid of 4 points. Results (Figure 4) show that the fraction of experimental points inside the uncertainty band is smaller than when Gaussian quadrature was used, despite the two bands being of approximately equal width.

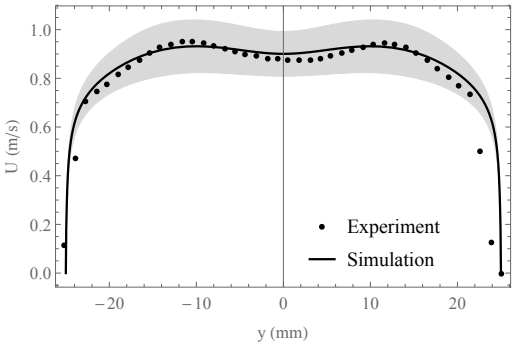

Figure 2. Velocity profile calculated via random sampling from 90 simulations. Shaded area indicates $\pm 2\sigma$ uncertainty band.

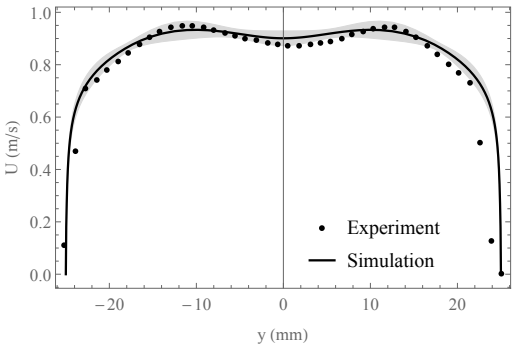

Figure 3. Velocity profile obtained using Gaussian quadrature with 4 simulations. Shaded area indicates $\pm 2\sigma$ uncertainty band.

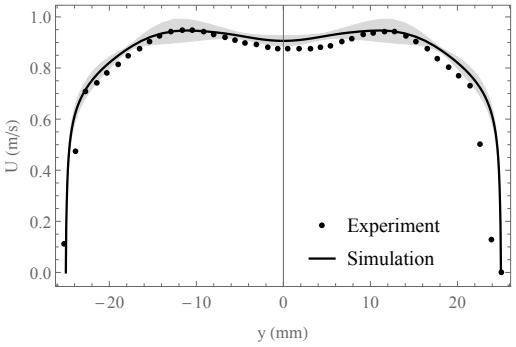

Figure 4. Velocity profile obtained with the linear regression method using 4 simulations. Shaded area indicates $\pm 2\sigma$ uncertainty band.

Comparison of the methods

In order to facilitate a quantitative comparison of the three PCE methods discussed above, two issues were considered. First, how close the experimental points were to the uncertainty band; second, how wide the uncertainty band was. The ideal

method would be one that produces a narrow uncertainty band, still containing all the experimental points inside. To address the first issue, the following figure of merit (FOM) was defined:

$$FOM = 1 - d \tag{17}$$

where d is the average distance between the experimental points and the $\pm 2\sigma$ band, normalized such that $d \in [0,1]$. If one point lies inside the band, $d = 0$ for that point. Hence, $FOM = 1$ if all points are inside the band, and $FOM \to 0$ as the points move away from it. The width of the uncertainty band was measured using its area, normalized in the range $[0,1]$.

Figures 5 and 6 display the FOM and area of the uncertainty band obtained with each PCE method. The random sampling technique achieved a higher FOM than the other two methods, but at the cost of an uncertainty band several times wider. The excessive width of the uncertainty band renders random sampling useless for most engineering applications, unless a large quantity of simulations is used. With respect to the other two PCE methods, Gaussian quadrature provided a higher FOM than linear regression, with approximately the same uncertainty area. The advantages of Gaussian quadrature over linear regression are better appreciated when few simulation are used.

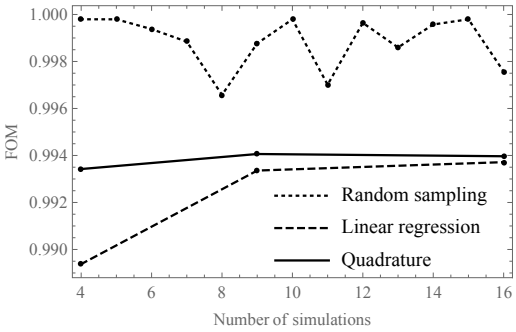

Figure 5. Figure of merit obtained with each PCE method.

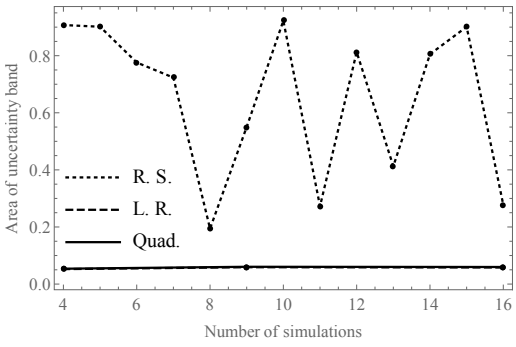

Figure 6. Normalized area of the $\pm 2\sigma$ uncertainty band produced by each PCE method.

CONCLUSIONS

From all the three PCE methods discussed in this paper, Gaussian quadrature provided the best results from an engineering standpoint, since with as few as 4 simulations, it produced a relatively narrow uncertainty band containing most of the experimental points. Moreover, the solid mathematical principles upon which Gaussian quadrature stands, guarantee reliable results as long as the system's response can be accurately approximated by a polynomial of sufficient order. This condition is rather general, and can be easily satisfied in most practical situations.

On the other hand, the criteria for selecting the simulation points in the linear regression method, using either regular grids or random sampling, have no solid mathematical justification, and can inadvertently lead to erroneous results if the simulation points do not capture the full behavior of the system's response.

Finally, the random sampling method requires too many simulations to obtain accurate results, being of little use for industrial applications, where CFD simulations are usually very time-consuming. The only situation where randon sampling might offer some advantage is in problems with a large amount of uncertain parameters, as in the other two PCE methods, the number of simulations rapidly grows with n.

The results of this work reveal the ability of PCE via Gaussian quadrature to provide an accurate estimation of the uncertainty in CFD simulations, while keeping the computational costs sufficiently low to be feasible even in complex problems. However, it is important to remark that no uncertainty quantification method is able to include the uncertainty associated with the hypothesis and simplifications of the CFD model [13]. This limitation should be considered in cases where model uncertainty dominates over other sources of uncertainty, such as input data and discretization and numerical errors.

REFERENCES

[1] Wiener, N., The homogeneous chaos, *American Journal of Mathematics*, Vol. 60, No. 4, 1938, pp. 897-936

[2] Eldred, M.S., and Burkardt, J., Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification, *AIAA paper*, Vol. 976, No. 2009, 2009, pp. 1–20

[3] Hosder, S., and Walters, R.W., Non-intrusive polynomial chaos methods for uncertainty quantification in fluid dynamics, *8th AIAA Aerospace Sciences Meeting*, 2010, pp. 4–7

[4] Najm, H.N., Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, *Annual Review of Fluid Mechanics*, Vol. 41, 2009, pp. 33–52

[5] Xiu, D., and Karniadakis, G.E., Modeling uncertainty in flow simulations via generalized polynomial chaos, *Journal of Computational Physics*, Vol. 187, No. 1, 2003, pp. 137 - 167

[6] Weisstein, E.W., The CRC encyclopedia of mathematics, *CRC Press/Chapman Hall*, 3rd edition, 2009

[7] Abramowitz, M., and Stegun, A., Handbook of mathematical functions with formulas, graphs and mathematical tables, *Dover*, 1972

[8] Smolyak, S.A., Quadrature and interpolation formulas for tensor products of certain classes of functions, *Dokl. Akad. Nauk SSSR*, Vol. 148, No. 5, 1972, pp. 1042–1045

[9] Congedo, P.M, Corre, C. and Marnez, J. M., Shape optimization of an airfoil in a BZT flow with multiple-source uncertainties, *Computer Methods in Applied Mechanics and Engineering*, Vol. 200, No. 1, 2011, pp. 216–232

[10] Hosder, S., Walters, R.W., and Balch, M., Efficient sampling for non-intrusive polynomial chaos applications with multiple uncertain input variables, *Proceedings of the 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Vol. 1939, 2007

[11] Badillo, A., Kapulla, R., and Niceno, B., Uncertainty Quantification in CFD Simulations of Isokinetic Turbulent Mixing Layers, *Proceedings of the NURETH-15 Congress*, Paper 559, 2013

[12] D'Ammaro, A., and Montomoli, F., Uncertainty quantification and film cooling, *Computers & Fluids*, Vol. 71, 2013, pp. 320–326

[13] V 20 Committee and others, Standard for verification and validation in computational fluid dynamics and heat transfer, *American Society of Mechanical Engineers*, 2009