Development of Methods for Uncertainty Quantification in CFD Applied to Wind Turbine Wake Prediction

Parekh, Jigar

DOI:
10.33612/diss.799510060

Development of Methods for Uncertainty Quantification in CFD Applied to Wind Turbine Wake Prediction

The work described in this thesis was performed in the research group of Computational & Numerical Mathematics of the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence at the University of Groningen, The Netherlands. It is part of the Computational Sciences for Energy Research program which is financed by the Netherlands Organization for Scientific Research (NWO).

university of groningen

Development of Methods for Uncertainty Quantification in CFD Applied to Wind Turbine Wake Prediction

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. J.M.A. Scherpen
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Tuesday 14 November 2023 at 12:45 hours

by

Jigar Parekh

born on 22 December 1990
in Jabalpur, India

Supervisors
Prof. R. W. C. P. Verstappen
Prof. C. A. Bertoglio

Assessment Committee
Prof. P. R. Onck
Prof. F. X. Trias
Prof. J. G. Peypouquet

Contents

1

INTRODUCTION

1.1. BACKGROUND

Climate change is one of the biggest challenges that the humanity is currently facing. The burning of fossil fuels, deforestation, and other human activities are releasing large amounts of greenhouse gases into the atmosphere, causing the earth's temperature to rise. This is leading to a range of negative impacts, including more frequent and intense natural disasters, displacement of communities, and negative impacts on food and water supplies, human health, and the economy. If we do not take immediate action to reduce our greenhouse gas emissions, the consequences could be catastrophic.

It is important for governments, communities, and individuals to work together to accelerate the transition to a low-carbon future. One of the most effective ways to address climate change is to transition to clean, renewable energy sources. The international community has also recognized the importance of renewable energy in addressing climate change. The Paris Agreement, a landmark international treaty signed by nearly 200 countries in 2015, calls for the phase-out of fossil fuels and the transition to clean, renewable energy sources [92]. According to the Intergovernmental Panel on Climate Change, increasing the use of renewable energy is essential to meet the goals of the Paris Agreement and to avoid the worst impacts of climate change [75].

Wind energy is particularly promising as it is a reliable and cost-effective source of electricity that does not emit any greenhouse gases. According to the International Energy Agency (IEA), wind energy could provide 18% of the world's electricity by 2025, up from just 4% in 2010 [47]. Therefore, by increasing the use of this clean, renewable energy source, we can mitigate the negative impacts of climate change. Due to the growing demand for wind energy around the world, many countries have set ambitious renewable energy targets and are looking to increase the amount of wind energy in their energy mix. This presents significant opportunities for the wind energy industry.

There are several challenges in harnessing wind energy. One of the major challenges is the variability of wind. The speed and direction of wind can vary significantly over time, which makes it difficult to predict the amount of energy that a wind turbine will produce.

This can make it difficult to integrate wind energy into the grid and to balance supply and demand. One way to address this challenge is through the use of advanced forecasting tools and energy storage technologies. Another major challenge is the cost of wind energy. While the cost of wind energy has decreased significantly in recent years, it is still generally more expensive than energy from fossil fuels. This can make it difficult for wind energy to compete with other energy sources. However, the IEA also notes that the cost of wind energy is expected to continue to decline as the technology improves and becomes more widespread [47]. Overall, harnessing wind energy is a complex and challenging process, but it also presents many opportunities to address some of the most pressing issues facing humanity today, such as climate change and energy security.

Wind turbine is one of the primary technologies used to generate electricity from wind energy. It consist of a rotor with blades that captures the wind's kinetic energy and converts it into rotational energy. This rotational energy is then used to drive a generator, which converts it into electricity. Wind turbines are typically divided into two categories: horizontal axis and vertical axis [76]. Horizontal axis wind turbines are the most common type, and they consist of a rotor with three blades that is mounted on top of a tower. The rotor is oriented in a horizontal plane and is connected to a gearbox, which increases the rotational speed of the rotor and drives the generator. Vertical axis wind turbines, on the other hand, have a rotor with blades that are oriented in a vertical plane. These turbines are less common and are typically used in small-scale applications.

Accurate prediction of the power output of a wind turbine is critical for the success of wind energy projects. There are a number of factors that can impact the power output of a wind turbine. To accurately predict the power output of a wind turbine, it is necessary to understand, characterize and model the physics (length and time scales) of all the relevant phenomena around the wind turbine. This calls for the use advanced modeling and simulation techniques such as computational fluid dynamics (CFD) [101].

1.2. WIND MODELING USING CFD

A schematic representation of the flow features behind a wind turbine is shown in Figure 1.1. The wind-turbine wake (region downstream of the turbine) generally consists of two regions, (i) the near-wake region that extends up to $2-4$ rotor diameters, and (ii) the far-wake region that is further downstream of the turbine [122]. The near-wake is a three-dimensional, complex, and heterogeneous flow that is directly influenced by the presence of the wind turbine. On the other hand, the far-wake zone is less impacted by the wind turbine. Therefore, it is possible to forecast the mean flow distribution in this region using global wind-turbine data like the thrust and power coefficient and incoming flow conditions [12].

Wind farms and turbines have been designed so far using manageable engineering technologies. Typically, straightforward wind shear models are used to describe the incoming wind, while measurements are used to determine turbulence characteristics. The upstream environment properties consisting of aerodynamic roughness, moisture, temperature gradients, etc., govern these characteristics. Moreover, the engineering wake models have been employed to maximize estimates of annual energy production while minimizing wake impacts in wind farm. The development of wind energy has been greatly aided by all of these technical tools, but as wind turbines and wind farms become larger, the fundamental

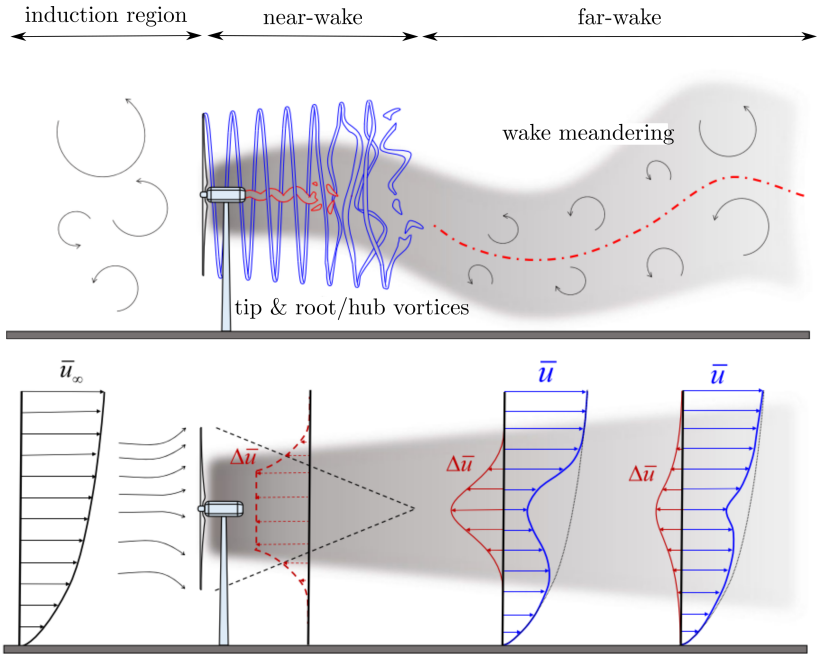


Figure 1.1 | Schematic representation of the instantaneous (top) and the time-averaged (bottom) flow features behind a wind turbine resulting from the interaction of the rotor and atmospheric boundary layer upstream, with $\overline{u}$ being the streamwise velocity and $\Delta\overline{u}$ its deficit (adapted from [90]).

assumptions in these models no longer hold true. For example, large wind turbines are more likely to operate outside the surface layer, particularly under stable atmospheric conditions. In these cases, the traditional logarithmic wind shear profile may not be suitable for accurately predicting the wind conditions experienced by the turbine, as it is based on the assumption that the wind behaves in a predictable and consistent manner within the surface layer. This can make it difficult to accurately predict the performance and behavior of the turbine under these conditions, as the characteristics of the wind are not fully understood.

Computational fluid dynamics (CFD) is a powerful tool for understanding and predicting the behavior of fluids, including air. In the context of wind energy, CFD can be used to optimize the design of wind turbines and wind farms, as well as to predict the performance of these systems under different operating conditions. By simulating the flow of air through a wind farm, engineers can predict how the wind turbines will interact with each other and determine the most efficient configuration. This can help to maximize the overall power output of the wind farm.

In computational fluid dynamics there are broadly three different approaches to simulating turbulent fluid flow as show in Figure 1.2.

- *Direct Numerical Simulation* (DNS) which involves solving the Navier-Stokes equations for the entire flow field, without any averaging or modeling of the turbulent eddies. DNS provides the most accurate and detailed simulation of fluid flow by resolving all length scales of the flow, including the smallest turbulent structures. However,

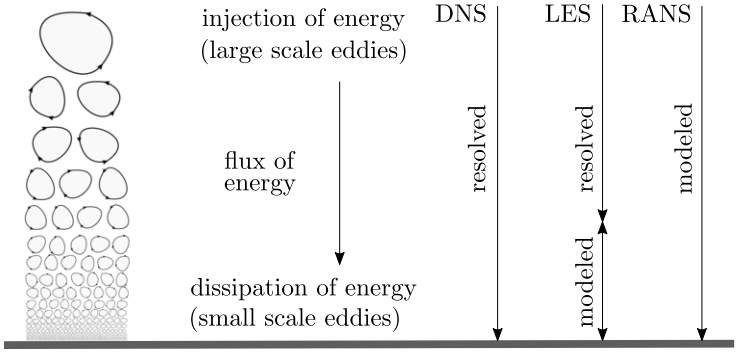


Figure 1.2 | A simplified representation of DNS, LES and RANS modeling approaches of a turbulent flow.

DNS is computationally expensive and is limited to relatively low Reynolds numbers due to the high resolution required to capture all the relevant scales of motion. It is commonly used in academic research and for fundamental studies of turbulence [17].

- *Large Eddy Simulation* (LES) is a hybrid CFD technique that combines resolved and modeled simulations. In LES, the larger turbulent structures are resolved directly (similar to DNS), while the smaller scales are modeled. The idea is to capture the energy-containing turbulent eddies and model the effect of the smaller, dissipative scales. LES provides a good compromise between accuracy and computational cost, allowing for simulations at higher Reynolds numbers than DNS. It is commonly used in engineering applications where the detailed flow features are important, such as aerodynamics and combustion [52, 99].

- *Reynolds-Averaged Navier-Stokes* (RANS) is a commonly used CFD technique that averages the Navier-Stokes equations over time, assuming that the flow quantities vary slowly and can be decomposed into mean and fluctuating components. The RANS equations are then solved, and additional turbulence models (e.g., $k-\epsilon$ or $k-\omega$ models) are employed to close the equations. RANS provides relatively fast and inexpensive simulations compared to DNS and LES. However, it is limited in capturing unsteady and highly three-dimensional flow features accurately. RANS is widely used in engineering practice for a variety of applications, including industrial aerodynamics, heat transfer, and fluid-structure interactions [3, 50, 65].

In the context of wind turbine simulations, LES and RANS approaches can be used to model the flow around the turbine and predict its performance. The choice of which approach to use will depend on the specific goals of the simulation, the level of accuracy required, and the available computational resources [90, 121, 142].

1.3. UNCERTAINTY QUANTIFICATION IN CFD

Uncertainty quantification (UQ) analysis, in general, provides a measure of the reliability of the simulation results. In any numerical simulation, there are various sources of un-

certainty, including input data, model parameters, and numerical methods. UQ aims to identify, quantify, and reduce these uncertainties to improve the accuracy and reliability of the simulation results. By quantifying uncertainty, a UQ analysis can provide valuable information about the confidence and limitations of the simulation results, which can help to make informed decisions. For example, in engineering applications, UQ can help to determine the safety margins and design criteria, while in scientific applications, it can help to evaluate the accuracy of the simulation results and guide further experimentation.

In the realm of computational science and engineering, uncertainty quantification (UQ) analysis generally involves making (mostly probabilistic) predictions of the quantities of interest (QoIs) under the presence of various uncertainties. Figure 1.3 presents the central idea of UQ analysis in four major steps: quantification, propagation, certification and calibration. In the following sub-sections we describe these steps in context of CFD followed by an example - uncertainty quantification analysis of stochastic Burgers' equation.

1.3.1. CHARACTERIZATION

The *characterization* step involves identifying and characterizing the potential sources of uncertainty in the system. Depending on the nature of the uncertainty, it can be generally categorized into *aleatory* and *epistemic* uncertainty. Aleatory uncertainties, also known as the stochastic uncertainties, are deemed irreducible and can be considered as the natural variability of the (stochastic) system. The epistemic uncertainties are reducible systematic uncertainties and are associated with the lack of knowledge. In other words, with an increase in knowledge, there is an improvement in the characterization of aleatory uncertainties and reduction in the epistemic uncertainties. Every source of uncertainty can be modeled as entirely aleatory, entirely epistemic or a combination of both [57, 98]. In general, the aleatory uncertainty is described using probability density functions, whereas the epistemic uncertainty is represented using approaches like interval analysis and evidence theory [81, 87].

Various sources of uncertainty (or error) appearing in CFD are listed in Table 1.1 [81, 124]. In addition to the general categorization of uncertainties into aleatory or epistemic, these sources of uncertainties can be broadly categorized into - *model*, *parametric* and *numerical* uncertainties. The model uncertainty arise due to the assumptions used to model a physical system using mathematical description, e.g. steady and/or laminar flow, incompressibility, turbulent eddy viscosity etc. Additionally, simplifications resulting in auxiliary models like the RANS turbulence models, further contribute towards the model uncertainties. The parametric uncertainties corresponds to the model's closure parameters that may vary based on the CFD application (typical in turbulence modeling), initial and/or boundary conditions, geometric representation etc. The numerical uncertainty is not only associated with the finite decimal representation (approximations) of numbers but also with the contribution of (reducible) errors from the spatial and temporal discretization of the partial differential equations used to model the physical system.

Out of all the above mentioned sources of uncertainty, it is mostly observed that the geometric uncertainty, the discretization error and the turbulence modeling are the largest contributors towards the uncertainty in a modern CFD simulation [124].

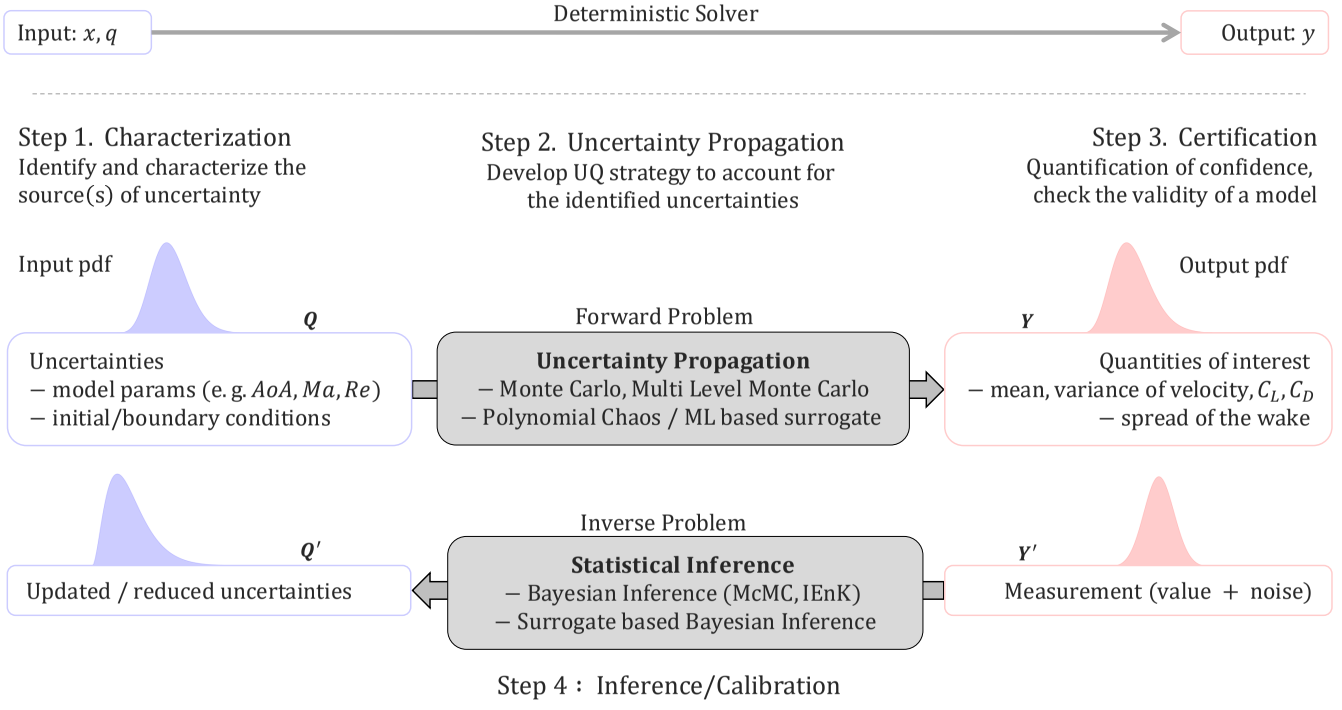


Figure 1.3 | Central idea of UQ analysis: quantification, propagation, certification and calibration. For a deterministic problem - x represents the input (independent) variables, q represents the model parameters, boundary conditions, etc., and y the output. In the context of UQ analysis - the uncertainty in q is represented using a probability density function (*pdf*) $\mathbf{Q}$ and the resulting distribution of the output as $\mathbf{Y}$. In the context of data assimilation - $\mathbf{Y}'$ represents the noisy measurement data and $\mathbf{Q}'$ the updated *pdf* of input random variables. The symbols *AoA, Re, Ma*, C_L and C_D represent angle of attack, Mach number, Reynolds number, coefficient of lift and coefficient of drag, respectively, which may be relevant for e.g. in the UQ analysis of an aerodynamic problem.

Sources	Examples
Physical Modeling (Assumptions in the PDEs)	Inviscid Flow Viscous Flow Incompressible Flow Transition / Turbulent Flow
Auxiliary Physical Models	Equation of State Thermodynamics properties Transport properties Turbulence Models
Boundary & Initial Conditions	Wall, e.g. roughness Open, e.g. far-field Free Surface
Discretization & Solution	Truncation error – spatial and temporal Iterative convergence – steady state, transient Geometry Representation
Round-off Error	Finite-precision arithmetic
Programming User Error	

Table 1.1 | Various sources of uncertainty and error in computational fluid dynamics (source - [81, 124]).

1.3.2. PROPAGATION AND CERTIFICATION

The *propagation* step quantifies the uncertainty in the output (QoIs) in the presence of uncertainty in the inputs and/or in the models (see section 1.3.1). Uncertainty quantification in CFD involves characterizing and propagating the uncertainty through the numerical simulation to evaluate the uncertainty (confidence) in the results [7, 35, 58, 105, 136]. Uncertainty propagation approaches can be broadly classified into *intrusive* and *non-intrusive* types. In the intrusive strategy the baseline mathematical model as well as the numerical solver (code) are both modified. This may result in an highly accurate and efficient method, although at a considerably high programming effort. The non-intrusive method is generally a sampling based technique that considers the mathematical model (and its code) as a black box model. This approach does not require modifications in the baseline solver but can be computationally expensive if the black box model is expensive to evaluate (e.g. CFD solve).

Consider a general stochastic differential equation:

$$\mathscr{L}(\boldsymbol{x}, t, \omega; \boldsymbol{v}(\boldsymbol{x}, t, \omega)) = S(\boldsymbol{x}, t, \omega), \tag{1.1}$$

where $\mathscr{L}$ is usually a nonlinear differential operator consisting of space and/or time derivatives (e.g. CFD model), $\boldsymbol{v}(\boldsymbol{x}, t, \omega)$ is the solution and $S(\boldsymbol{x}, t, \omega)$ is the source term. The random event ω represents the uncertainty in the system, introduced via uncertain parameters, the operator, the source term, the initial/boundary conditions, etc. The complete probability space is given by $(\Omega, \mathscr{A}, \mathscr{P})$, where Ω is the sample space such that $\omega \in \Omega$, $\mathscr{A} \subset 2^{\Omega}$ is the

σ-algebra on Ω and $\mathscr{P} : \mathscr{A} \mapsto [0,1]$ is the probability measure on $(\Omega, \mathscr{A})$.

Assuming that a baseline model (like a CFD code) exists that solves the deterministic form of equation (1.1), we now explore various common approaches that can be used for uncertainty propagation in the stochastic problem described above (1.1).

MONTE CARLO METHODS

Monte Carlo (MC) is the most straightforward sampling technique that can be used as non-intrusive uncertainty propagation approach. In the MC method the samples are randomly sampled from the input probability space i.e. from the joint probability density function (PDF) of the input random variables. Thereafter, these sample are evaluated using a black box model to compete the mean and the variance in the output QoIs. For the general stochastic problem (1.1), using the MC method, the mean $\mathbf{E}$ and variance $\mathbf{V}$ of the output $\boldsymbol{v}$ is approximates as

$$\begin{aligned} \mathbf{E}[\boldsymbol{v}(\boldsymbol{x},t,\omega)] &\approx \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{v}(\boldsymbol{x},t,\boldsymbol{\xi}_i(\omega)), \\ \mathbf{V}[\boldsymbol{v}(\boldsymbol{x},t,\omega)] &\approx \frac{1}{N-1}\sum_{i=1}^{N}\Big(\boldsymbol{v}(\boldsymbol{x},t,\boldsymbol{\xi}_i(\omega)) - \mathbf{E}[\boldsymbol{v}(\boldsymbol{x},t,\omega)]\Big)^2, \end{aligned} \tag{1.2}$$

where N is the number of samples, $\boldsymbol{\xi}_i(\omega)$ is the i^{th} MC sample of the vector of d independent random variables $\boldsymbol{\xi} = \{\xi_1, ..., \xi_d\}$, corresponding to d uncertain parameters and $\boldsymbol{v}(\boldsymbol{x},t,\boldsymbol{\xi}_i(\omega))$ is the output of the (deterministic) black box model at $\boldsymbol{\xi}_i(\omega)$. The error decreases with $1/\sqrt{N}$ and is independent of the number of input random variables. Besides its obvious advantages of easy implementation, non-intrusiveness and dimensionality independence, the Monte Carlo method generally requires a large number of samples. In the context of CFD, such a large number of runs is infeasible and therefore cheaper yet accurate alternatives shall be sought.

POLYNOMIAL CHAOS EXPANSION

Polynomial chaos (PC) expansion approach is based on the spectral decomposition of the random variable f, in terms of basis polynomials containing randomness ψ_i (known a priori) and the unknown deterministic expansion coefficients f_i, as

$$f(x,q) = \sum_{i=0}^{\infty} f_i(x)\psi_i(q). \tag{1.3}$$

The intrusive variant of PC expansion also known as *intrusive polynomial chaos* (IPC) approach reformulates (expands) the original model resulting in (coupled) set of equations for the PC coefficients (mode strengths) of the model output. One of the widely used IPC methods used for the propagation of uncertainty is the Galerkin polynomial chaos method [125] which provides the spectral representation of the stochastic solution and results into higher order approximations of the mean and variance. Galerkin polynomial chaos method is a non-statistical method where the uncertain parameter(s) and the solution become random variables (fields). For the general stochastic problem (1.1), the random variables

can be approximated using *polynomial chaos* (polynomial of random variables) as follow [35]

$$\boldsymbol{v}(\boldsymbol{x},t,\omega) \approx \sum_{i=0}^{P} \boldsymbol{v}_i(\boldsymbol{x},t)\psi_i(\boldsymbol{\xi}(\omega)). \tag{1.4}$$

It is worth noting that the expansion (1.4) is indeed the decomposition of a random variable into a deterministic component - the expansion coefficients $\boldsymbol{v}_i(\boldsymbol{x},t)$ and a stochastic component - the random basis functions (polynomial chaoses) $\psi_i(\boldsymbol{\xi}(\omega))$. Based on the dimension of $\boldsymbol{\xi}$ (which here is d) and the highest order n of the polynomials $\{\psi_i\}$, the infinite summation can be truncated to $P+1 = (d+n)!/(d!\,n!)$ terms. An important property of the basis $\{\psi_i\}$ is their orthogonality with respect to the joint PDF of the uncertain inputs, $\langle\psi_i\psi_j\rangle = \langle\psi_i^2\rangle\delta_{ij}$. Here, δ_{ij} is the Kronecker delta and $\langle\cdot,\cdot\rangle$ denotes the inner product in the Hilbert space of the variables $\boldsymbol{\xi}$, $\langle f(\boldsymbol{\xi})g(\boldsymbol{\xi})\rangle = \int f(\boldsymbol{\xi})g(\boldsymbol{\xi})w(\boldsymbol{\xi})d\boldsymbol{\xi}$. The weighting function $w(\boldsymbol{\xi})$ is indeed the probability density function of the uncertain parameters. Such polynomials already exist for some standard distributions which can be found in the Askey scheme [137]. For example, a Normal distribution leads to Hermite-chaos, while Legendre-chaos corresponds to a Uniform distribution. For other commonly used distributions or any arbitrary distribution, one can for example use the Gram-Schmidt algorithm [127] to construct the orthogonal polynomials. Substituting (1.4) in the general stochastic differential equation (1.1), we obtain

$$\mathscr{L}\left(\boldsymbol{x},t,\omega;\sum_{i=0}^{P}\boldsymbol{v}_i\psi_i\right) \approx S. \tag{1.5}$$

In order to ensure that the truncation error is orthogonal to the functional space spanned by the basis polynomials $\{\psi_i\}$, a Galerkin projection of the above equation is performed onto each polynomial $\{\psi_k\}$,

$$\left\langle \mathscr{L}\left(\boldsymbol{x},t,\omega;\sum_{i=0}^{P}\boldsymbol{v}_i\psi_i\right),\psi_k\right\rangle = \langle S,\psi_k\rangle, \qquad k=0,1,...,P. \tag{1.6}$$

After using the orthogonality property of the polynomials, we obtain a set of $P+1$ deterministic coupled equations for all the random modes of the solution $\{\boldsymbol{v}_0,\boldsymbol{v}_1,...,\boldsymbol{v}_k\}$.

The non-intrusive variant of PC expansion also known as *non-intrusive polynomial chaos* (NIPC) approach, unlike its intrusive counterpart, approximates the expansion coefficients $\boldsymbol{v}_i(\boldsymbol{x},t)$ instead of directly solving for them i.e. the projection in equation (1.6) is approximated using numerical quadrature, with quadrature points and respective weights based on the joint probability density function of the input uncertainty [44, 135]. This results into a decoupled system of equations (similar to the deterministic equation) for each mode of the polynomial expansion. The following linear system of equations is solved in order to obtain the expansion coefficients:

$$\begin{pmatrix} \psi_0(\xi_0) & \psi_1(\xi_0) & \cdots & \psi_P(\xi_0) \\ \psi_0(\xi_1) & \psi_1(\xi_1) & \cdots & \psi_P(\xi_1) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_0(\xi_P) & \psi_1(\xi_P) & \cdots & \psi_P(\xi_P) \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_0(\boldsymbol{x},t) \\ \boldsymbol{v}_1(\boldsymbol{x},t) \\ \vdots \\ \boldsymbol{v}_P(\boldsymbol{x},t) \end{pmatrix} = \begin{pmatrix} \boldsymbol{v}(\boldsymbol{x},t,\xi_0) \\ \boldsymbol{v}(\boldsymbol{x},t,\xi_1) \\ \vdots \\ \boldsymbol{v}(\boldsymbol{x},t,\xi_P) \end{pmatrix} \tag{1.7}$$

In contrast to the Monte Carlo method that samples randomly with equal sample weights, NIPC method can be seen as a *smarter* (PDF-based) sampling technique with non-uniform

weight distribution. For a polynomial chaos of order n with d random variables, the NIPC method requires $(n+1)^d$ deterministic solves to approximate all the expansion coefficients. Thus, the number of deterministic evaluations increase exponentially with the dimensionality of the input random space. Sparse-grid methods like Latin Hypercube sampling can significantly reduce the number of samples to be evaluated using the deterministic solver.

Computing lower order moments is straightforward once the expansion coefficients $\boldsymbol{v}_i(\boldsymbol{x}, t)$ are obtained using any of the variants discussed above. The mean and the variance of the solution are given by

$$\begin{aligned} \mathbf{E}[\boldsymbol{v}(\boldsymbol{x}, t, \omega)] &= \boldsymbol{v}_0(\boldsymbol{x}, t), \\ \mathbf{V}[\boldsymbol{v}(\boldsymbol{x}, t, \omega)] &= \sum_{i=1}^{P} \boldsymbol{v}_i^2(\boldsymbol{x}, t) \langle \psi_i^2 \rangle. \end{aligned} \tag{1.8}$$

Also, the probability distribution of the solution $\boldsymbol{v}(\boldsymbol{x}, t, \omega)$ can be obtained using for example kernel density estimation based on samples from the PC expansion.

Due to its properties like the high computational efficiency, exponential convergence, accurate representation of uncertainty etc., the polynomial chaos expansion is regarded as one of the most suitable alternatives to the Monte-Carlo methods [137].

SURROGATE BASED UNCERTAINTY PROPAGATION

In the realm of machine learning numerous methods are available to construct a surrogate (response surface) using data. One such meta-model is an artificial neural network (ANN) which generally involves regression of an arbitrary non-linear function f that maps input to the output [8, 37]. In context of the general stochastic problem (1.1), for a given input $\{\boldsymbol{x}, t, \boldsymbol{\xi}_i(\omega)\}$ we would like to approximate $\boldsymbol{v}$ with the function f based on the weights $\boldsymbol{w}$, such that

$$\boldsymbol{v}(\boldsymbol{x}, t, \boldsymbol{\xi}_i(\omega)) \approx f(\boldsymbol{x}, t, \boldsymbol{\xi}_i(\omega), \boldsymbol{w}). \tag{1.9}$$

Note that for a steady state CFD simulation with fixed spatial coordinates of the mesh, the above expression simplifies to $\boldsymbol{v}(\boldsymbol{\xi}_i(\omega)) \approx f(\boldsymbol{\xi}_i(\omega), \boldsymbol{w})$. An artificial neural network is typically a network of nodes organized into layers and connected to one another which apply the *activation* function g on the values gathered from the previous nodes. The activation functions introduce non-linearity in the network and essentially avails the approximation of arbitrary functions. Sigmoid, hyperbolic tangent and ReLU are a few most widely used activation functions. The output of the i^{th} node in the l^{th} layer represented as $a_{i,l}$ is expressed as:

$$a_{i,l} = g\left(\sum_{j=0}^{n_l - 1} w_{ij,l-1} a_{j,l-1} \right), \tag{1.10}$$

where n_l is the number of nodes per layer and $a_{0,l}$ is the bias for l^{th} layer and is set to 1 for all layers. All the degrees of freedom of the neural network (regular weights and biases) can simply be denoted by a single weight vector $\boldsymbol{w}$. Thus, the matrix representation of equation (1.10) is of the form $\boldsymbol{a}_l = \boldsymbol{g}(W_{l-1}\boldsymbol{a}_{l-1})$, with the activation function applied element-wise to the input vector. The weights are computed by solving an optimization problem where the loss function $L(\boldsymbol{v}, f(\boldsymbol{w}))$ is minimized usually under the presence of regularization to avoid an over-fitted model.

Once the surrogate model is constructed (and validated), we can, for instance, use a Quasi Monte Carlo (QMC) sampling to compute the statistics like mean and variance of the respective QoIs (see equation (1.2)). Note that techniques like QMC involve large number of samples and wouldn't be suitable for a direct UQ analysis using computationally expensive CFD runs. However, using an accurate surrogate model makes is drastically cheaper to evaluate these large number of samples.

The surrogate based uncertainty quantification outperforms Monte-Carlo methods by offering higher computational efficiency, better scalability to high dimensions, adaptability to complex response surfaces and seamless integration with optimization algorithms to perform optimization under uncertainty [34, 59, 114].

SENSITIVITY ANALYSIS

A natural step that follows uncertainty propagation is the so-called *sensitivity analysis*, in which the sources of uncertainty that have the most influence are identified while the sources with least effect can be considered to be fixed at their nominal values during the forward UQ problem [19, 41, 45, 72, 113, 124].

The sensitivity analysis can be carried out *locally* and/or *globally*. The local sensitivity analysis include computing partial derivatives using e.g. adjoint methods or finite-difference approximations. The global sensitivity analysis, on the other hand, include approaches like analysis of variance (ANOVA) decomposition which in turn leads to the so-called Sobol indices [106]. The first-order indices (also called the *main-effect*) measure the contribution of each individual input variable to the output variance independently of interactions or correlations with other variables. In other words, the first-order Sobol index measures the main effect of a variable while keeping other variables fixed. The *total-effect* measure the contribution of each input variable, including both its main effect and any interactions with other variables, there by accounting for all possible effects of a variable, whether they are independent or due to interactions with other variables.

In case of PCE the Sobol indices can be analytically determined once the coefficients of the expansion are approximated [113]. In the surrogate based approach, on the other hand, the sensitivities are computed numerically e.g. using a Quasi Monte Carlo sampling to approximate the contribution of inputs and their interactions [106].

1.3.3. CALIBRATION

The experimental and the measurement/field data can be used to (re)calibrate the model (e.g. RANS turbulence model) parameters to obtain an accurate (highly probable) ensemble of predictions under uncertainties. This process of propagating the information from QoIs (outputs) towards the inputs is seen as an *inverse or backward problem* and is commonly known as *calibration*. Bayesian calibration is one of most common techniques to solve the so-called inverse problem [54, 56, 94, 95]. In Bayesian calibration the uncertainty (aleatory or epistemic) in the parameters to be calibrated is treated in the probabilistic sense and are represented as random variables. Bayes' formula is used to obtain the *posterior* (calibrated) distribution of the random variables based on - (i) the assumed/modeled *prior* distribution of the parameters, and (ii) the statistical model that can describe the relationship between the computational model and the data. Note that the latter, in turn, results in a probability density function known as the *likelihood*. Denoting the data as z, the model as $\mathscr{M}$ and the

(random) model parameters as $\boldsymbol{\theta}$, a statistical model with an error ϵ can be expressed as [54]:

$$z_i = \mathscr{M}(\boldsymbol{x}, t, \boldsymbol{\theta}) + \epsilon_i, \tag{1.11}$$

where the subscript represents the i^{th} observation. The error term is usually a multi-variate normal distribution. Using Bayes rule the posterior distribution $\rho(\theta|\boldsymbol{z})$ of the parameters can be obtained as:

$$\rho(\boldsymbol{\theta}|\boldsymbol{z}) = \frac{\rho(\boldsymbol{z}|\boldsymbol{\theta})\rho(\boldsymbol{\theta})}{\rho(\boldsymbol{z})}, \tag{1.12}$$

where $\rho(\boldsymbol{z}|\boldsymbol{\theta})$ is the likelihood, $\rho(\boldsymbol{\theta})$ is the prior and $\rho(\boldsymbol{z})$ is the evidence. For a given sample, computing the likelihood involves the evaluation of the computational model. It is possible to obtain an explicit form (PDF) of the posterior for a linear computational model and simpler statistical error. For a non-linear computational model (e.g. CFD) an explicit form of likelihood is not known and one can only evaluate the posterior several times to approximate its form. Markov chain Monte Carlo (McMC) methods are widely used as a smart sampling strategy to sample from high probability region(s) of the posterior. Such methods, however, requires large number of evaluation of the likelihood and are therefore less useful when the non-linear model is, for example, a computationally expensive CFD solve. A feasible solution would be to use a surrogate which can not only accurately represent the computational model but is also cheap to evaluate. We have already discussed the use of PCE and machine learning models for the purpose of uncertainty propagation. Similarly, these models can also be used for calibration as well. Note that the calibration step can either be carried out after the first three steps in Figure 1.3 or as an independent step directly after the characterization step.

1.3.4. UQ ANALYSIS OF BURGERS' EQUATION

In this section we present a UQ analysis example which highlights all the steps listed in the previous section. For the sake of simplicity we consider a 1D viscous Burgers' equation:

$$u\frac{\partial u}{\partial x} = \mu\frac{\partial^2 u}{\partial x^2}, \quad x \in [-1,\ 1], \tag{1.13}$$

where u is the velocity with $u(\mp 1) = \pm 1$, and μ is the viscosity which is assumed to be random with a uniform distribution $\mathscr{U}[0.05,\ 0.15]$. The computational model $\mathscr{M}$ is based on finite differences and approximates the velocity for a given (deterministic) viscosity.

For the uncertainty propagation we use all the four methods described in section 1.3.2. In the Monte Carlo (MC) approach we employ 10,000 samples to obtain (a gradually converged) estimate of the mean and variance of the velocity. The intrusive and the non-intrusive polynomial chaos (IPC/NIPC) methods both use Legendre polynomials with highest polynomial degree $n = 2$. The IPC method therefore solves three coupled equations in the expansion coefficients using finite difference ($\geq 3 \times \mathscr{M}$) while the NIPC approach uses three deterministic solves for the respective quadrature points ($3 \times \mathscr{M}$) to approximate the integrals and in turn the expansion coefficients. The artificial neural network (ANN) consists of a single hidden layer with 8 neurons where each node have a hyperbolic tangent activation function. The input and the output layers have dimensions equal to the dimension of the input space (here 1) and the dimension of the output (here mesh size), respectively. We

used mean-squared-error loss function and *Adam* optimizer with a learning rate of $1e-2$ with 16 training and 4 validation samples. The ANN runs for 500 epochs (iterations) before convergence.

Figure 1.4 shows the comparison of all the four methods for characterization and forward propagation of uncertainty in the viscosity. The uncertainty in the velocity due to the randomness in the viscosity is denoted with the ± 2 standard deviations bound and is maximum near $x = \pm 0.2$. Any prediction (or realization) of the velocity falls well within these stochastic bounds and adhere to the deterministic boundary conditions. Due to its exponential convergence rate [137], both polynomial chaos expansion methods require significantly low computational resources to reach the same level of accuracy as the Monte Carlo method with large number of samples. In the IPC method we solve a system of only three coupled equations for expansion coefficients. Similarly, in the NIPC method we require only three deterministic solves to approximate expansion coefficients. Although the ANN based propagation of uncertainty requires a higher number of deterministic solves than the NIPC method, it is still significantly cheaper as compared to the MC method.

On the availability of observations (measurements) or high-fidelity data at certain locations in the domain, the calibration can be carried out using the surrogate based Bayesian inference approach. For the problem at hand, the uncertainty in the input (random viscosity) is updated using the (synthetic) noisy data followed by a forward propagation to compute (and update) uncertainty in the velocity. Note that the forward propagation reuses the previously constructed surrogate (ANN), thus allowing an inexpensive calibration. Figure 1.5 shows the prior, the noisy data and the posterior for both the random viscosity and random velocity. The calibration resulted in an overall reduction of uncertainty in the system.

1.4. UQ IN WIND ENERGY APPLICATIONS

Several wind energy studies have been performed to understand the effect of the input parameters on the outputs. The influence of the variation in wind direction on the power output of a wind farm was studied using LES simulations of the Horns Rev wind farm for a number of inflow angles [91]. This work showed that the power output of the wind farm is highly sensitive to the inflow wind direction. Since the probability distribution of the inputs are not considered in such analysis, they are usually not termed as uncertainty quantification despite the fact that they still deliver important results. A simple uncertainty quantification analysis is e.g. presented in the study of the wake effects over a normally distributed wind direction [32], where the measurement data is closely represented with the quantities averaged over several simulations as compared to that with only single inflow wind direction simulation. This study also asserts that, in addition to the inaccuracy of the simulation model, the disagreements between the data and the simulations can also be attributed to the significant uncertainty in the measurements.

One of the biggest advantage of using uncertainty quantification is to significantly increase the accuracy and confidence in numerical simulations. However, in the context of wind energy, such UQ analysis can incur a substantial computational cost if we look beyond the previously used simpler and inexpensive wake models. Particularly, for the complex RANS or LES models, analysis similar to the ones discussed above might not be feasible. We

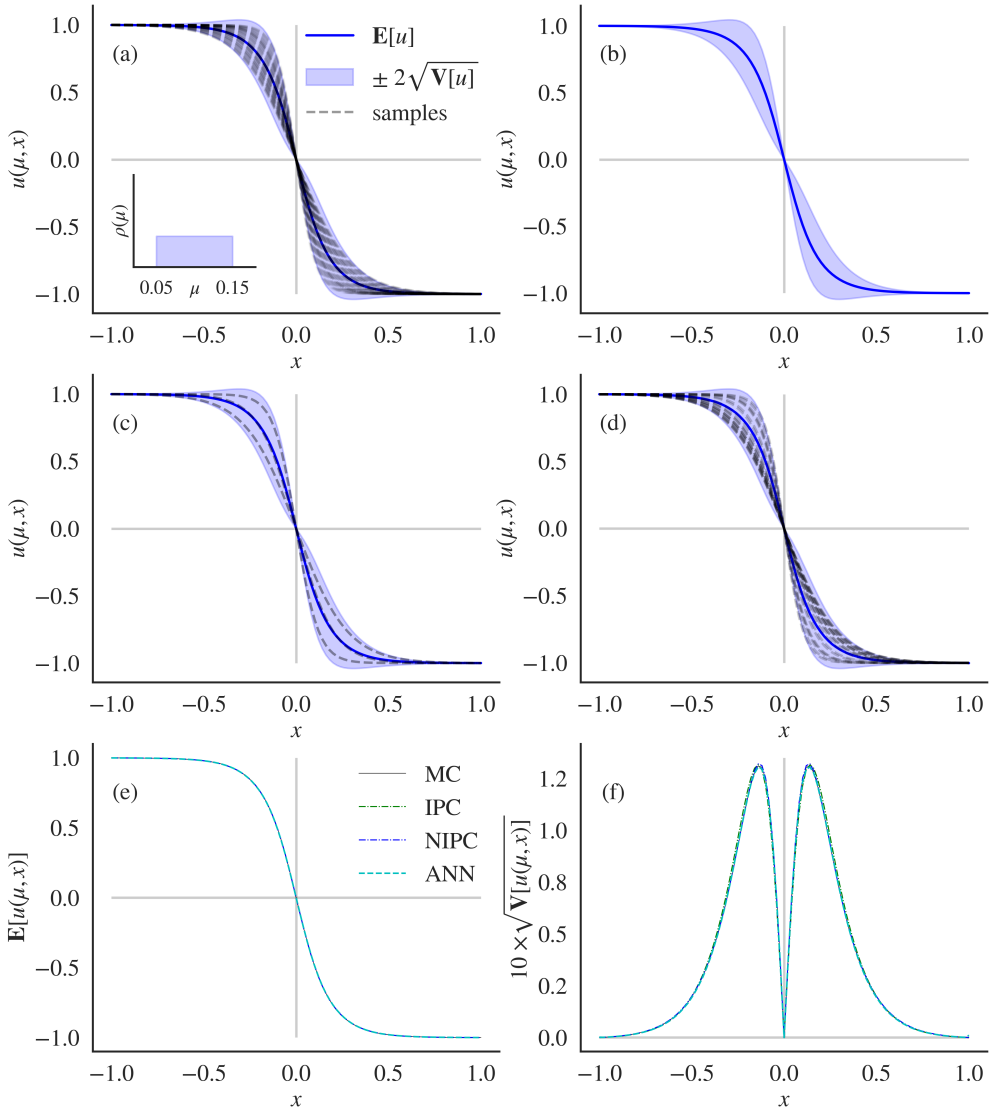


Figure 1.4 | Uncertainty propagation in stochastic Burgers' equation using (a) Monte Carlo (10000 samples), (b) Intrusive polynomial chaos (3 coupled equations), (c) Non-intrusive polynomial chaos (3 quadrature points), and (d) Artificial neural network (16 training samples). Comparison of mean and standard deviation in (e) and (f). The probability density function of the uncertain viscosity is given by $\rho(\mu)$.

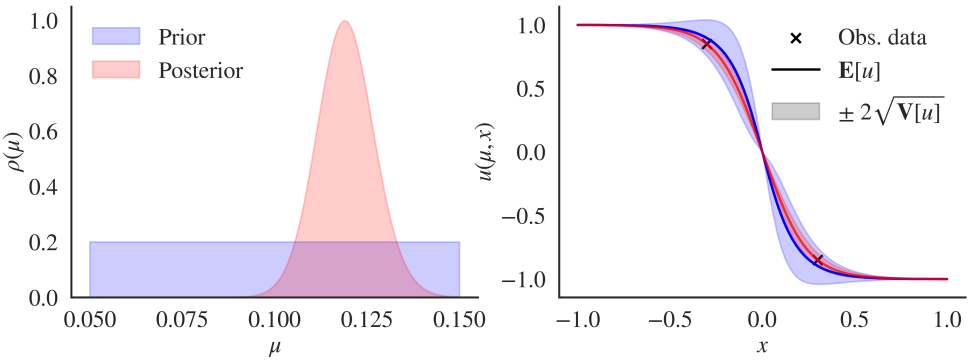


Figure 1.5 | Surrogate based Bayesian calibration of uncertainty in stochastic Burgers' equation.

continue our discussion on the recently employed advanced uncertainty quantification and propagation approaches in wind energy applications.

As discussed in section 1.3.2, several methods such as Monte-Carlo, polynomial chaos expansions and surrogate based approach can be used to propagate the uncertainties through a CFD model. The influence of the wind speed, the surface roughness and the air density on the wind energy potential was studied by [61] using the Monte-Carlo (MC) approach. The uncertainties considered were characterized as non-Gaussian. Moreover, model form uncertainty in the power performance is also included in their work. The MC method was also used to study the effect of correlated uncertainties on the wind farm's energy yield [24]. In [51], a normal distribution for the power curve with varying mean and fixed variance was used in along with MC method for uncertainty quantification analysis. Moreover, based on the histogram data, a joint probability distribution function for correlated wind speed and wind direction is presented in [27], which is in turn used by a random search algorithm to perform layout optimization. To overcome the computationally expensive MC method, one could use the advanced uncertainty propagation methods. Both non-intrusive and intrusive polynomial chaos expansion (PCE) methods have been used for UQ analysis in wind energy applications. A Weibull distribution and a uniform distribution for the wind speed and the wind direction, respectively, were used to construct the polynomial basis employed in a PCE method to compute the uncertainty in the AEP of the wind farm [80]. A similar approach was used in [84], where the analytic polynomial expression was used to optimize an offshore wind farm's layout. In their work, it was confirmed that the PCE is indeed a more efficient integration rule (w.r.t. the probability density function) than the expensive MC method. A common problem with PCE is the appearance of oscillations especially near the discontinuities. This problem was alleviated by using lower order polynomials which was in accordance with a course MC simulation [71, 97]. In their work, CFD simulations were used to study the influence of uncertain wind speed and angle of attack on the pressure and force coefficients of the turbine blades. The effect of randomness in the induction factor and ground roughness on the power output of the wind farm using only a fourth order polynomial chaos expansion provided results with higher accuracy than a MC approach with large number of samples [31]. A stochastic Galerkin (intrusive PCE) method was used in [29] to obtain a stochastic blade loading model which involved solving a single non-linear

coupled system of equations instead of running a large number of deterministic solves. The entire stochastic solution is computed at once and involves fairly simple post-processing steps to compute the statistics. In a continued work [30], the authors used a reduced order model based on a lower number of random variables to generate a turbulent wind field which in turn is used as an input for the stochastic model in [29]. Recently machine learning methods have gained popularity in UQ analysis of wind energy applications. Probabilistic neural network were used in [53] to quantify the model or epistemic uncertainty of the power curve estimation. The authors used a probabilistic loss function to estimate the aleatoric uncertainty in the data. In [73], the effect of geometric uncertainties on the wind turbine airfoil aerodynamics have been investigated. In that work, the authors developed an artificial neural network based reduced order model to predict the load coefficients and performance of wind turbine airfoils.

A data-driven approach can be adopted to incorporate the measurement or high-fidelity simulation data for the calibration of the computational models used in wind turbine wake modeling. In [115] the mixing length parameter of a RANS turbulence model was tuned using LES data to obtain an efficient wake model. Similarly, the Jensen wake model was calibrated using LES data and used in an optimization method to obtain optimal yaw settings which were later tested in a LES study [33]. In [118], a finite element model of a turbine blade was used to identify the parameters which had the most influence on the blade vibrations (based on analysis of variance). Thereafter, these parameters were calibrated using McMC during Bayesian inference. The finite element model was replaced with an inexpensive surrogate model to reduce the computational cost of McMC based calibration.

Previously, for CFD simulations in general, major efforts have been made to quantify the uncertainties in turbulence model via parametric approach, where the closure model parameters are either perturbed or directly injected with randomness [18, 74, 85]. The shortcomings of this parametric approach have been recognized by the turbulence modeling community and efforts are made towards the non-parametric approaches [82]. Commonly, uncertainties in the Reynolds stress are introduced by perturbing its eigenvalues [22, 38], eigenvectors [46] or tensor invariants [70, 132]. These low-dimensional global perturbation methods assume a uniform distribution of randomness throughout the computational domain which may result in an inaccurate estimation (mostly over-prediction) of the uncertainties in the quantities of interest. Considerable effort is required in the development of high-dimensional local perturbation methods for an accurate and detailed prediction of uncertainties in CFD simulations.

In context of wind energy applications, most of the widely used turbulence models for CFD simulations are inherently uncertain resulting in significant variance in the flow field solutions [16, 133]. Additionally, the uncertainties in the boundary conditions that mimic the real environmental conditions may also contribute to the inaccurate predictions [7]. Efforts have been made to study the influence of model form uncertainty in CFD simulations of wakes in a wind farm, e.g. in [20, 43]. However, further investigation with multiple sources of uncertainties (parametric and model form) to study their combined influence on the performance of the wind turbines has not been made so far.

1.5. RESEARCH OBJECTIVES

The overarching goal of this thesis is the development of methods for quantification and propagation of parametric and model form uncertainties in CFD simulations. These methods are used to study the non-linear effects of the uncertainties on the quantities of interest, particularly in context of wind farm simulations. The topics presented in this thesis can be abridged as follows:

- Proposal, motivation and testing of a new IPC based stochastic solver over increasing dimensionality of the randomness and/or the complexity of CFD simulations.
- Adaption of a deep learning method along with wake superposition principles to facilitate the development of a surrogate model used for the prediction of the flow field in a wind farm.
- Comparison of the IPC based solver and the surrogate based approach for the UQ analysis in the prediction of wake interactions in a wind farm under the presence of multiple sources and/or types of uncertainties.

1.6. THESIS STRUCTURE

The uncertainty quantification analysis of Burgers' equation in section 1.3.4 superficially reflects the idea of this thesis. In the upcoming chapters we investigate in detail the quantification and propagation of different forms of uncertainties in various complex CFD problems using the methods developed alongside.

In chapter 2, we present the formulation and implementation of a stochastic CFD solver which employs Generalized Polynomial Chaos (gPC) expansion to (a) quantify the uncertainties associated with the fluid flow simulations, and (b) study the non-linear propagation of these uncertainties. In particular, the parametric uncertainties - random laminar viscosity in the Poiseuille flow and random LES model parameter C_S in turbulent channel flow are characterized and propagated. A comparison with the non-intrusive counterpart is also discussed.

In chapter 3, we present a more holistic approach to treat uncertainties in a CFD simulation while still be able to use intrusive polynomial chaos method. We consider two distinct high-dimensional variants of the uncertainties, namely, the random eddy viscosity field and the random Reynolds stress tensor field. In contrast to the globally perturbed parametric uncertainties these random fields allow for local perturbations that can be characterized using the information from a baseline simulation. A stochastic RANS solver is developed and tested on two benchmark problems for RANS turbulence modeling - the flow over periodic hills and the flow in a square duct. A detailed investigation of the stochastic solver based on several influential hyper-parameters is also presented.

After establishing the capabilities and exploring the potential applications of the *physics-based* stochastic RANS solver in chapter 3, we make efforts towards developing a *data-driven* approach using one of the state of the art deep learning methods. In chapter 4, we use a 3D U-Net neural network (trained over single wind turbine data) combined with wake superposition principles to construct a surrogate model which is capable of predicting the flow field in a wind farm. This response surface is then used in a Surrogate-Based

Uncertainty Quantification (SBUQ) approach. Using both the approaches, the IPC solver and the SBUQ approach, a UQ analysis for the prediction of wake interactions in a wind farm with mixed (parametric and model form) uncertainties is carried out. The results obtained are compared and discussed in detail followed by an account on the advantages and disadvantages of both the approaches.

A summary and conclusion of the work presented in chapters 2, 3 and 4 is given in chapter 5, where we moreover provide some prospects and recommendations for future research.

2

INTRUSIVE POLYNOMIAL CHAOS FOR PARAMETRIC UNCERTAINTIES

The content of this chapter was published in J. Parekh and R. Verstappen. Intrusive Polynomial Chaos for CFD Using OpenFOAM. In Computational Science – ICCS 2020, pages 677–691, Cham, 2020. Springer International Publishing.

ABSTRACT: We present the formulation and implementation of a stochastic Computational Fluid Dynamics (CFD) solver based on the widely used finite volume library - OpenFOAM. The solver employs Generalized Polynomial Chaos (gPC) expansion to (a) quantify the uncertainties associated with the fluid flow simulations, and (b) study the non-linear propagation of these uncertainties. The aim is to accurately estimate the uncertainty in the result of a CFD simulation at a lower computational cost than the standard Monte Carlo (MC) method. The gPC approach is based on the spectral decomposition of the random variables in terms of basis polynomials containing randomness and the unknown deterministic expansion coefficients. As opposed to the mostly used non-intrusive approach, in this work, we use the intrusive variant of the gPC method in the sense that the deterministic equations are modified to directly solve for the (coupled) expansion coefficients. To this end, we have tested the intrusive gPC implementation for both the laminar and the turbulent flow problems in CFD. The results are in accordance with the analytical and the non-intrusive approaches. The stochastic solver thus developed, can serve as an alternative to perform uncertainty quantification, especially when the non-intrusive methods are significantly expensive, which is mostly true for a lot of stochastic CFD problems.

2.1. INTRODUCTION

In simulating a physical system with a model, uncertainties may arise from various sources [124], namely, initial and boundary conditions, material properties, model parameters, etc. These uncertainties may involve significant randomness or may only be approximately known. In order to enhance the predictive reliability, it is therefore important to quantify the associated uncertainties and study its non-linear propagation especially in CFD simulations.

In order to reflect the uncertainty in the numerical solution, we need efficient Uncertainty Quantification (UQ) methods. Broadly there exists two classes of UQ methods, the intrusive method, where the original deterministic model is replaced by its stochastic representation, and the non-intrusive method, where the original model itself is used without any modifications [66, 123]. Monte Carlo (MC) sampling is one of the simplest non-intrusive approaches. However, due to its requirement of a large number of samples, MC method is computationally expensive for application in CFD. As an alternative, we can use Generalized Polynomial Chaos (gPC) representations which has been proven to be much cheaper than MC [44, 124]. This approach is based on the spectral decomposition of the random variables in terms of basis polynomials containing randomness and the unknown deterministic expansion coefficients. In this paper, we focus mainly on the Intrusive Polynomial Chaos (IPC) method, where a reformulation of the original model is performed resulting in governing equations for the expansion coefficients of the model output [138].

As the model code, we use OpenFOAM [25], which is a C++ toolbox to develop numerical solvers, and pre-/ post-processing utilities to solve continuum mechanics problems including CFD. OpenFOAM (a) is a highly templated code, enabling the users to customize the default libraries as needed for their applications, and, (b) gives access to most of the tensor operations (divergence, gradient, laplacian etc.) directly at the top-level code. This avails enough flexibility to implement the IPC framework for uncertainty quantification in CFD. To obtain the inner products of polynomials we use a python library called chaospy [26], as a pre-processing step to the actual stochastic simulation.

First, the idea of generalized polynomial chaos is presented with a focus on the intrusive variant. A generic differential equation is used to explain the steps involved in IPC, leading to a simple expression for the mean and variance as a function of the expansion coefficients. Next, we present the set of deterministic governing equations followed by its stochastic formulation using IPC. In particular, a Large Eddy Simulation (LES) method is used to model turbulence, which includes an uncertain model parameter. Thereafter, we discuss the algorithm and implementation steps required for the new stochastic solver in OpenFOAM. The stochastic version of the Navier-Stokes equations has a similar structure to the original system. This allows reusing the existing deterministic solver with minimal changes necessary.

The stochastic solver developed so far is tested for various standard CFD problems. Here we present two cases, the plane Poiseuille flow with uncertain kinematic viscosity, and the turbulent channel flow with uncertain LES model parameter. The results are found to be in accordance with the non-intrusive gPC method.

2.2. GENERALIZED POLYNOMIAL CHAOS

The Generalized Polynomial Chaos approach is based on the spectral decomposition of the random variable(s) f, in terms of basis polynomials containing randomness ψ_i (known a priori) and the unknown deterministic expansion coefficients f_i, as $f(x,q) = \sum_{i=0}^{\infty} f_i(x)\psi_i(q)$. There are two methods to determine the expansion coefficients, namely, the Intrusive Polynomial Chaos (IPC) and the Non-intrusive Polynomial Chaos (NIPC). In IPC, a reformulation of the original model is performed resulting in governing equations for the PC mode strengths of the model output, while in NIPC, these coefficients are approximated using quadrature for numerical evaluation of the projection integrals.

The level of accuracy of these methods can be associated with the degree of gPC. To attain the same level of accuracy, particularly for a higher dimensional random space, IPC requires the solution of a much fewer number of equations that needed for NIPC. Moreover, for such a random space, the aliasing error resulting from the approximation of the exact gPC expansion in the NIPC method can become significant. This suggests that, for a multi-dimensional problem, the IPC method can deliver more accurate solutions at a much lower computational cost than the NIPC method [135].

Since the current work is based on IPC, we would introduce here the important features of the intrusive variant and we refer to the literature [35, 125, 137] for more details about NIPC and gPC in general.

INTRUSIVE POLYNOMIAL CHAOS

In order to demonstrate the application of IPC, we first consider a general stochastic differential equation

$$\mathscr{L}(\boldsymbol{x}, t, \omega; \boldsymbol{v}(\boldsymbol{x}, t, \omega)) = S(\boldsymbol{x}, t, \omega), \tag{2.1}$$

where $\mathscr{L}$ is usually a nonlinear differential operator consisting of space and/or time derivatives, $\boldsymbol{v}(\boldsymbol{x}, t, \omega)$ is the solution and $S(\boldsymbol{x}, t, \omega)$ is the source term. The random event ω represents the uncertainty in the system, introduced via uncertain parameters, the operator, source term, initial/boundary conditions, etc. The complete probability space is given by $(\Omega, \mathscr{A}, \mathscr{P})$, where Ω is the sample space such that $\omega \in \Omega$, $\mathscr{A} \subset 2^{\Omega}$ is the σ-algebra on Ω and $\mathscr{P} : \mathscr{A} \mapsto [0,1]$ is the probability measure on $(\Omega, \mathscr{A})$.

We now employ the Galerkin polynomial chaos method, which is an IPC method for the propagation of uncertainty [125]. It provides the spectral representation of the stochastic solution and results into higher order approximations of the mean and variance. Galerkin polynomial chaos method is a non-statistical method where the uncertain parameter(s) and the solution become random variables. These random variables are approximated using the polynomial chaos (polynomial of random variables) as follow [35]

$$\boldsymbol{v}(\boldsymbol{x}, t, \omega) \approx \sum_{i=0}^{P} \boldsymbol{v}_i(\boldsymbol{x}, t)\psi_i(\boldsymbol{\xi}(\omega)). \tag{2.2}$$

It is worth noting that the expansion (2.2) is indeed the decomposition of a random variable into a deterministic component, the expansion coefficients $\boldsymbol{v}_i(\boldsymbol{x}, t)$ and a stochastic component, the random basis functions (polynomial chaoses) $\psi_i(\boldsymbol{\xi}(\omega))$. Here, $\boldsymbol{\xi}(\omega)$ is the vector of d independent random variables $\{\xi_1, ..., \xi_d\}$, corresponding to d uncertain

parameters. Based on the dimension of $\boldsymbol{\xi}$ (which here is d) and the highest order n of the polynomials $\{\psi_i\}$, the infinite summation has been truncated to $P+1=(d+n)!/(d!\,n!)$ terms.

An important property of the basis $\{\psi_i\}$ is their orthogonality with respect to the probability density function (PDF) of the uncertain parameters, $\langle\psi_i\psi_j\rangle=\langle\psi_i^2\rangle\delta_{ij}$. Here, δ_{ij} is the Kronecker delta and $\langle\cdot,\cdot\rangle$ denotes the inner product in the Hilbert space of the variables $\boldsymbol{\xi}$, $\langle f(\boldsymbol{\xi})g(\boldsymbol{\xi})\rangle=\int f(\boldsymbol{\xi})g(\boldsymbol{\xi})w(\boldsymbol{\xi})d\boldsymbol{\xi}$. The weighting function $w(\boldsymbol{\xi})$ is the probability density function of the uncertain parameters. Such polynomials already exist for some standard distributions which can be found in the Askey scheme [137], for example, a Normal distribution leads to Hermite-chaos, while Legendre-chaos corresponds to a Uniform distribution. For other commonly used distributions or any arbitrary distribution, one can for example use Gram-Schmidt algorithm [127] to construct the orthogonal polynomials.

Substituting (2.2) in the general stochastic differential equation (2.1), we obtain

$$\mathscr{L}\left(\boldsymbol{x},t,\omega;\sum_{i=0}^{P}\boldsymbol{v}_i\psi_i\right)\approx S. \tag{2.3}$$

In order to ensure that the truncation error is orthogonal to the functional space spanned by the basis polynomials $\{\psi_i\}$, a Galerkin projection of the above equation is performed onto each polynomial $\{\psi_k\}$,

$$\left\langle\mathscr{L}\left(\boldsymbol{x},t,\omega;\sum_{i=0}^{P}\boldsymbol{v}_i\psi_i\right),\psi_k\right\rangle=\langle S,\psi_k\rangle,\quad k=0,1,...,P. \tag{2.4}$$

After using the orthogonality property of the polynomials, we obtain a set of $P+1$ deterministic coupled equations for all the random modes of the solution $\{\boldsymbol{v}_0,\boldsymbol{v}_1,...,\boldsymbol{v}_k\}$. Following the definition, the mean and the variance of the solution are given by

$$\mathbf{E}[\boldsymbol{v}]=\mu_{\boldsymbol{v}}=\boldsymbol{v}_0(\boldsymbol{x},t),\quad \mathbf{V}[\boldsymbol{v}]=\sigma^2_{\boldsymbol{v}}=\sum_{i=1}^{P}\boldsymbol{v}_i(\boldsymbol{x},t)^2\langle\psi_i^2\rangle. \tag{2.5}$$

As the coefficients $\boldsymbol{v}_i(\boldsymbol{x},t)$ are known, the probability distribution of the solution can be obtained.

2.3. GOVERNING EQUATIONS

We first discuss the governing equations in the deterministic setting. The Navier-Stokes equations for an incompressible flow is given by

$$\frac{\partial\boldsymbol{u}}{\partial t}+(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}=-\nabla p+\nabla\cdot(\nu\nabla\boldsymbol{u}),\quad \nabla\cdot\boldsymbol{u}=0, \tag{2.6}$$

where $\boldsymbol{u}$ is the velocity, p is the pressure and ν is the kinematic viscosity.

In Large Eddy Simulation, the reduction in the range of scales in a simulation is achieved by applying a spatial filter to the Navier-Stokes equations [99]. This results into

$$\frac{\partial\overline{\boldsymbol{u}}}{\partial t}+(\overline{\boldsymbol{u}}\cdot\nabla)\overline{\boldsymbol{u}}=-\nabla\overline{p}+\nabla\cdot(\nu\nabla\overline{\boldsymbol{u}})-\nabla\cdot\boldsymbol{\tau},\quad \nabla\cdot\overline{\boldsymbol{u}}=0 \tag{2.7}$$

where $\overline{\boldsymbol{u}}$ is the filtered velocity, $\overline{p}$ is the filtered pressure and $\boldsymbol{\tau} = \overline{\boldsymbol{u}\boldsymbol{u}} - \overline{\boldsymbol{u}}\,\overline{\boldsymbol{u}}$ is the so-called subgrid-scale (SGS) stress tensor. The subgrid-scale stress tensor represents the effect of the small (unresolved) scales on the resolved scales, and has to be modeled in order to close the filtered Navier-Stokes equations.

A popular class of SGS models is the eddy-viscosity models. In order to take account of the dissipation through the unresolved scales, the eddy-viscosity models, locally increases the viscosity by appending the molecular viscosity with the eddy viscosity. Mathematically, these models specify the anisotropic part of the subgrid-scale tensor as

$$\boldsymbol{\tau} - \frac{1}{3}\,\mathrm{tr}(\boldsymbol{\tau}) = -2\nu_t \overline{\boldsymbol{S}}, \tag{2.8}$$

where ν_t is the eddy-viscosity and $\overline{\boldsymbol{S}} = (\nabla\overline{\boldsymbol{u}} + (\nabla\overline{\boldsymbol{u}})^T)/2$ is the resolved strain tensor. Substituting into the filtered momentum equation (2.6), we obtain

$$\frac{\partial\overline{\boldsymbol{u}}}{\partial t} + (\overline{\boldsymbol{u}}\cdot\nabla)\overline{\boldsymbol{u}} = -\nabla\overline{p} + \nabla\cdot((\nu+\nu_t)\nabla\overline{\boldsymbol{u}}), \tag{2.9}$$

where the incompressibility constraint is used to simplify the equation. The pressure here is altered to include the trace term of equation (2.8).

Smagorinsky model [104] is one of the oldest and most popular eddy-viscosity SGS model. The eddy viscosity of the Smagorinsky model is expressed as

$$\nu_t = C_s^2\Delta^2|\overline{\boldsymbol{S}}|, \tag{2.10}$$

where C_s is the Smagorinsky coefficient, Δ is the LES filter width and $|\overline{\boldsymbol{S}}| = \sqrt{2\overline{\boldsymbol{S}}:\overline{\boldsymbol{S}}}$. It should be noted that the coefficient C_s must be known prior to the simulation and is usually adapted to improve the results [99]. For example, $C_s = 0.2$ is used for isotropic homogeneous turbulence, while a value of $C_s = 0.1$ is used in case of channel flow. Similar values ($C_s \simeq 0.1 - 0.12$) are realized from the shear flow studies based on experiments [83].

2.3.1. STOCHASTIC FORMULATION

Let us consider the Navier-Stokes Equations (2.6) with some uncertainty in the system. The sources of the uncertainty considered here are boundary conditions, material properties and model parameters. We employ the IPC method (see section 2.2) by presuming the dimensionality (d) and probability density function of the uncertain random variables $\{\xi_1, \xi_2, ...\xi_d\}$ to be known, allowing us to construct the finite set of orthogonal polynomial basis $\{\psi_i\}$.

In order to obtain a rather generic formulation, unless specified otherwise, we consider uncertainty in all sources listed above. Thus, the associated polynomial chaos expansion (PCE) for kinematic viscosity is given by

$$\nu \approx \sum_{i=0}^{P} \nu_i\psi_i(\boldsymbol{\xi}). \tag{2.11}$$

Note that the coefficients ν_i are assumed to be known. The dependence of the flow variables, i.e. velocity and pressure, on the stochastic variables is expressed by the following

PCEs

$$\boldsymbol{u}(\boldsymbol{x},t) \approx \sum_{i=0}^{P} \boldsymbol{u}_i(\boldsymbol{x},t)\psi_i(\boldsymbol{\xi}), \quad p(\boldsymbol{x},t) \approx \sum_{i=0}^{P} p_i(\boldsymbol{x},t)\psi_i(\boldsymbol{\xi}), \tag{2.12}$$

where $\boldsymbol{u}_i$ and p_i are the unknown polynomial chaos mode strengths of velocity and pressure fields, respectively. For deterministic boundary conditions, $\boldsymbol{u}_i$ and p_i are all zero for $i = 1,2,...,P$. In case of uncertain boundary conditions with a known (or modeled) probability density function, $\boldsymbol{u}_i$ and p_i, for $i = 0,1,...,P$, can be estimated.

Introducing these expansions in equations (2.6) and taking a Galerkin projection onto each polynomial $\{\psi_k\}$ while using the orthogonality of the polynomial chaos and finally dividing by $\langle\psi_k\psi_k\rangle$, results, for $k = 0,1,...,P$, into the following set of deterministic equations:

$$\frac{\partial \boldsymbol{u}_k}{\partial t} + \sum_{i=0}^{P}\sum_{j=0}^{P} (\boldsymbol{u}_i \cdot \nabla)\boldsymbol{u}_j M_{ijk} = -\nabla p_k + \sum_{i=0}^{P}\sum_{j=0}^{P} \nabla \cdot (\nu_i \nabla \boldsymbol{u}_j) M_{ijk}, \quad \nabla \cdot \boldsymbol{u}_k = 0, \tag{2.13}$$

where $M_{ijk} = \dfrac{\langle\psi_i\psi_j\psi_k\rangle}{\langle\psi_k\psi_k\rangle}$. Note that the original system of equations (2.6) is transformed into a system of $(P+1)$ divergence-free constraints on velocity modes and $(P+1)$ coupled equations in velocity and pressure modes. A detailed discussion on the solution procedure adopted for this large system of equations is deferred to section 2.4.

Similarly, for the filtered momentum equation (2.7) with Smagorinsky model for turbulence, applying the above steps, results in

$$\begin{aligned}\frac{\partial \overline{\boldsymbol{u}}_k}{\partial t} + \sum_{i=0}^{P}\sum_{j=0}^{P} (\overline{\boldsymbol{u}}_i \cdot \nabla)\overline{\boldsymbol{u}}_j M_{ijk} = -\nabla \overline{p}_k + \sum_{i=0}^{P}\sum_{j=0}^{P} \nabla \cdot (\nu_i \nabla \overline{\boldsymbol{u}}_j) M_{ijk} \\ + \sum_{i=0}^{P}\sum_{j=0}^{P}\sum_{l=0}^{P}\sum_{m=0}^{P} \nabla \cdot (C_{s_l} C_{s_m} \Delta^2 |\overline{\boldsymbol{S}}|_i \nabla \overline{\boldsymbol{u}}_j) M_{ijklm},\end{aligned} \tag{2.14}$$

where $M_{ijklm} = \dfrac{\langle\psi_i\psi_j\psi_k\psi_l\psi_m\rangle}{\langle\psi_k\psi_k\rangle}$. Note that $|\overline{\boldsymbol{S}}|^2 = 2\overline{\boldsymbol{S}} : \overline{\boldsymbol{S}}$, and applying the IPC steps to this identity - using polynomial chaos expansion and projecting on each basis polynomial, we obtain

$$\sum_{i=0}^{P}\sum_{j=0}^{P} |\overline{\boldsymbol{S}}|_i |\overline{\boldsymbol{S}}|_j M_{ijk} = 2\sum_{i=0}^{P}\sum_{j=0}^{P} \overline{\boldsymbol{S}}_i : \overline{\boldsymbol{S}}_j M_{ijk}, \tag{2.15}$$

where $\overline{\boldsymbol{S}}_i$ is the resolved strain tensor based on i^{th} velocity mode. The above corresponds to system of $(P+1)$ non-linear equations in the unknown expansion coefficients of $|\overline{\boldsymbol{S}}|$. This system is solved using Picard iterations with $|\overline{\boldsymbol{S}}|_i = \sqrt{2\overline{\boldsymbol{S}}_i : \overline{\boldsymbol{S}}_i}$ as the initial guess.

2.4. ALGORITHM AND IMPLEMENTATION

OpenFOAM uses the finite volume method (FVM) for the discretization of partial differential equations [28]. Among the various fluid dynamic solvers offered by OpenFOAM, we choose the solver called *pimpleFoam* [25], which allows the use of large time-steps to solve the incompressible Navier-Stokes equations (2.6). This solver is based on the PIMPLE algorithm for pressure-velocity coupling using Rhie and Chow type interpolation [49].

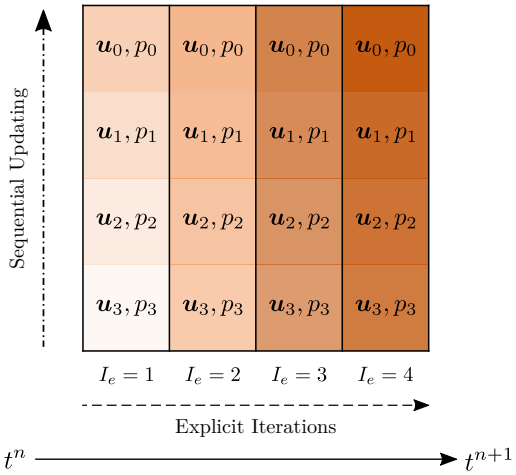


Figure 2.1 | Evolution of modes (for $P = 3$) between two consecutive time-steps via sequential updating using explicit iteration.

From the previous section, it can be realized that the system of governing equations (2.14) for the evolution of the velocity and pressure modes $\boldsymbol{u}_k, p_k$ for $k = 0, 1, ..., P$, has a structure similar to the original deterministic Navier-Stokes equations (2.6). Due to the coupling via convection and diffusion terms, the size of this new system is $P+1$ times its deterministic version. It can be observed that the divergence-free velocity constraints are decoupled and can be solved independently. Based on this observation, a fractional step projection scheme has been previously implemented [67]. In the first fractional step, the convection and diffusion terms are integrated followed by enforcing the divergence-free constraints in the second fractional step.

Our approach of the stochastic solver is based on the development of the existing deterministic solver (*pimpleFoam*) such that it can accommodate and solve $(P+1)$ coupled Navier-Stokes like systems in $\boldsymbol{u}_k, p_k$ for $k = 0, 1, ..., P$. We solve each of these systems sequentially, by using the initialized/updated velocity and pressure modes, and repeat until convergence. Figure 2.1, provides a graphical representation of this approach. Depending on the type of flow, the value of P and a few other parameters; it usually takes about $3-6$ explicit iterations (I_e) to realize convergence at every time-step. The default value of I_e is set to $P+1$. Following the conventions of OpenFOAM, we call this solver, *gPCPimpleFoam*. In contrast to the fractional step scheme, this approach admits better stability, stronger coupling and faster convergence; with an efficient data management and minimal changes in the exiting solver.

In Figure 2.2, we highlight the most important steps needed to develop *gPCPimpleFoam*, over the existing solver, *pimpleFoam*. In contrast to the deterministic solver, two nested loops are introduced. The first loop is over the explicit iterations (I_e), which updates the mode strengths between the two consecutive time-steps. For every explicit iteration, the second nested loop solves the Navier-Stokes like system $(\boldsymbol{u}_k, p_k)$ for each mode strength k, while employing the existing, however modified, structure of the PIMPLE scheme. The

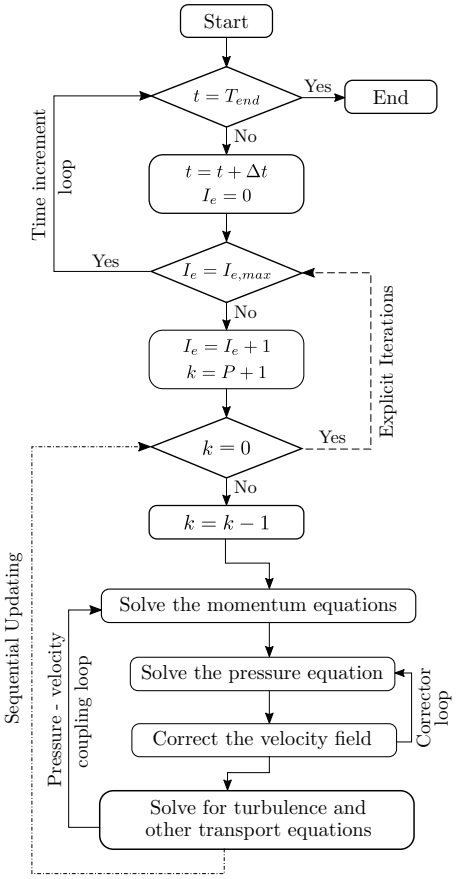


Figure 2.2 | Important steps of the algorithm implemented in the *gPCPimpleFoam* solver.

modifications are inevitable due to the summations in the convection and the diffusion terms of the stochastic equations. It is realized that the mean mode $(\boldsymbol{u}_0, p_0)$ changes slowly as compared to the other modes. Thus, in order to increase the stability, we start solving the last system $(\boldsymbol{u}_P, p_P)$ first, and updating all the modes before solving the first system (0^{th} mode) representing the mean.

In addition to the exiting modules, we require to either modify or create some completely new routines for turbulence, pre- and post- processing etc. Restricting the verbosity, we attempt to provide an overview of the major implementation steps: (a) Creation of new variables (vectors) for the list of mode strengths of all the uncertain parameters, flow variables and derived variables for post-processing, (b) For the solver to read these inner products (obtained using *chaospy* library [26]), only a small routine is added. Another similar lines of code are added to read in the values of d, n, I_e, etc, (c) Very subtle changes are needed in the transport model in order to read in the transport properties for all the mode strengths. To accommodate for reading and initializing the turbulence models, a few minor changes are made in the LES model library. Significant changes are required specially for the Smagorin-

sky model which allows automatic reading of all the mode strengths of C_s from the input file, and (d) To estimate the mean and the variance of flow variables and other derived quantities, a separate routine is added to the post-processing step of the solver. We use the IPC steps to calculate the derived quantities (like Reynolds Stresses) using the resolved expansion coefficients of the flow variables and the known coefficients of other parameters.

The order of accuracy and the convergence rate are governed by the large variety of space and time discretization schemes and iterative solvers offered by the OpenFOAM library.

2.5. TEST CASES

2.5.1. PLANE POISEUILLE FLOW

We first consider a 2D steady laminar flow in long rectangular channel (with a height of 2δ) in the absence of any external forces. For a given average inlet velocity u_{avg}, the fully developed flow has an analytical solution known as the Hagen-Poiseuille solution.

$$u(x,y) = \frac{3}{2}u_{avg}\left[1-\left(\frac{y}{\delta}\right)^2\right], \quad v(x,y) = 0, \quad \frac{\partial p}{\partial x} = -\frac{3\nu u_{avg}}{\delta^2}, \quad \frac{\partial p}{\partial y} = 0. \tag{2.16}$$

Therefore, the velocity field is independent of the viscosity and at $y = 0$, $u = u_{max} = \frac{3}{2}u_{avg}$. We assume the boundary condition to be deterministic and presume a known PCE for the uncertain kinematic viscosity, $\nu = \sum_{i=0}^{P} \nu_i \psi_i(\xi)$. Also, we consider a Gaussian random variable to model the viscosity, for which the associated polynomial chaoses $\psi_i(\xi)$, are the Hermite polynomials. Using PCE of pressure gradient and random viscosity, for $i = 0, 1, ..., P$, we obtain

$$\frac{\partial p_i}{\partial x} = -\frac{3u_{avg}}{\delta^2}\nu_i. \tag{2.17}$$

The use of polynomial chaos for incompressible laminar flow in a 2D channel has been previously investigated [67], and as a validation case, we carry out a similar study with the IPC solver developed using OpenFOAM.

A uniform velocity is used at the inlet with no-slip boundary conditions at the top and bottom walls, and the gradient of velocity is set to zero at the outlet. Note that the use of deterministic boundary condition implies, for $i = 1, ..., P$, the unknown mode strength and/or their derivatives are by default set to zero at the boundaries. The Reynolds number, $Re = 2\delta u_{avg}/\nu_0$, is set to 100. Fifth-order 1D Hermite polynomials are employed for all the PCEs, i.e. $P = 5$. The coefficient of variation ($CoV = \sigma/\mu$) for the uncertain viscosity is set to ~ 20 %, with $\nu_1/\nu_0 = 2 \times 10^{-1}$ and $\nu_2/\nu_0 = 8 \times 10^{-5}$. The remaining mode strengths of viscosity are assigned a value of zero. The simulation is performed in a domain with $L/\delta = 50$ and a 250×100 mesh with a near-wall grading. Second-order discretization schemes are used both in space and time, and time-step size $\Delta t = 10^{-2}\delta/u_{avg}$ is specified.

Figure 2.3 shows the profile of the mean and the standard deviation of velocity. The mean depicts the gradual transition in the flow along the channel length from a uniform inlet profile to a parabolic profile with maximum on the centerline. The uncertainty in velocity at inlet is zero, which is indeed the consequence of the deterministic boundary condition. In the developing region, a higher standard deviation is realized in the channel center as well as in the boundary layer with two lobes close to the walls. A significant

Figure 2.3 | Profiles of the mean and the standard deviation of velocity.

variation in the modes (and thus the standard deviation) is realized up to $10-12$ channel half-widths and further downstream, these modes become less significant. Figure 2.4 provides the axial velocity profile with confidence region ($\pm 2\sigma$) at different locations in the downstream direction. The uncertainties tend to zero in the fully developed region, which is in accordance with the theory where the velocity is independent of the viscosity (see equation (2.16)). Figure 2.5 shows the estimated ratios of the modes of pressure gradient with respect to the mean pressure gradient along the channel centerline. Clearly, after the recirculating regions near the channel inlet, these ratios gradually reach their constant values further downstream. For $x/\delta > 20$, the results are identical to analytical predictions (see equation (2.17)), characterizing the uncertainty in pressure due to the uncertainty in viscosity. Velocity mode strengths are shown in Figure 2.6. As evident, the results from the intrusive variant are in accordance with the non-intrusive counter-part, and due to the fast spectral convergence of the polynomial chaos representation, the magnitudes of the modes decrease as P increases.

2.5.2. TURBULENT CHANNEL FLOW

A turbulent channel flow is a theoretical representation of a flow driven by a constant pressure gradient between two parallel planes extending infinitely. A 3D schematic is shown in Figure 2.7. Since the computational domain has to be finite, in addition to channel width h, we fix the stream- and span- wise truncation lengths, l_x and l_z, respectively. The values of h, l_x and l_z are adopted from [13]. These values ensures that the computational domain is large enough to accommodate the turbulent structures in the flow.

In order to maintain an equivalent flow, instead of the pressure gradient, the bulk velocity can also be prescribed, $U_b = \frac{1}{h}\int_0^h \langle u \rangle dy$. The bulk Reynolds number is then defined as $Re_b = hU_b/\nu$. In context of turbulent channel flows, another characteristic velocity called the friction velocity is usually introduced in terms of the wall shear stress τ_w and the fluid density ρ, as $u_\tau = \sqrt{\tau_w/\rho}$. The friction Reynolds number is then defined as $Re_\tau = \delta u_\tau/\nu$, where $\delta = h/2$ is the channel half-width. Then the pressure gradient and the wall shear stress relates as $-\frac{d\bar{p}}{dx} = \frac{\tau_w}{\delta}$, where $\tilde{p}/\rho = \overline{p}$. Thus we have a choice between prescribing the bulk Reynolds number via bulk velocity and the friction Reynolds number via pressure gradient. Since we have to study the effect of the uncertain model parameter on the flow profile, we decide to fix the pressure-gradient and compute U_b through (stochastic) simulation.

The stochastic LES Smagorinsky model, as discussed in section 2.3, is employed to solve for the turbulence. Since the model parameter C_S may take a range of values, we assume it to be uncertain with a known PCE. We consider a Uniform random variable to model the parameter, for which the associated polynomial chaoses $\psi_i(\xi)$ are the Legendre polynomials. Table 2.1 summarizes the physical parameters used in the deterministic and stochastic

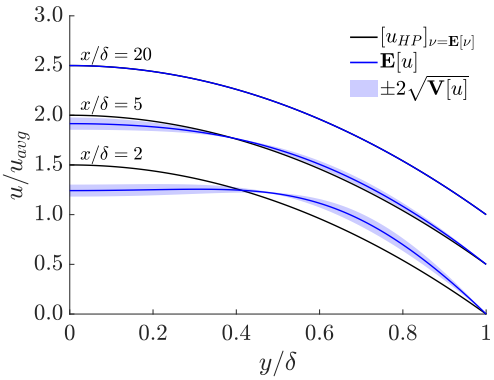


Figure 2.4 | Normalized axial velocity profiles at different cross-sections.

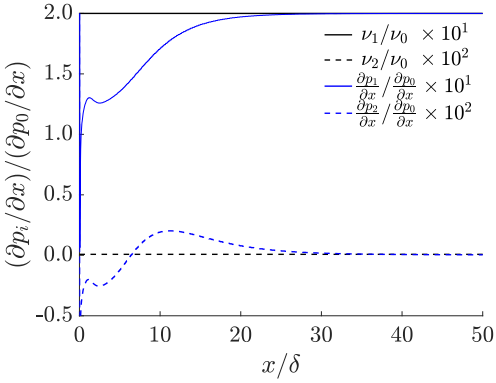


Figure 2.5 | Pressure gradient ratios along the centerline.

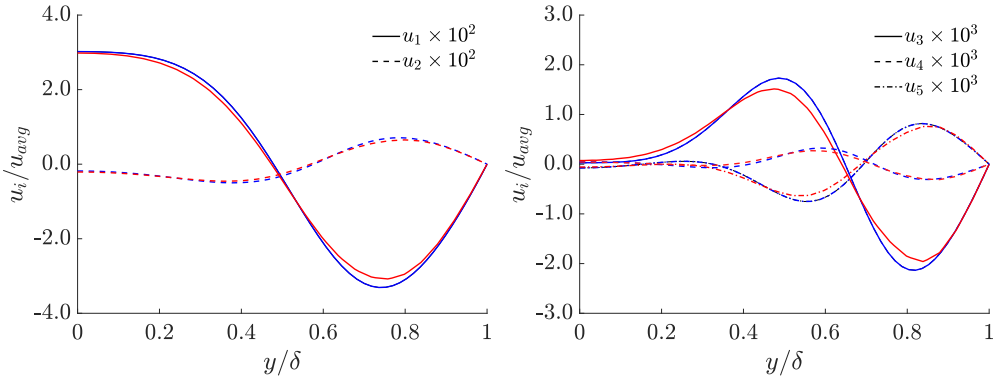


Figure 2.6 | Modes strengths of velocity at $x/\delta = 2$, for IPC (blue) and NIPC (red) methods.

simulations. 1D third-order Legendre polynomials are used for all the PCEs. The value of C_{S_1} is set equal to the standard deviation, while the remaining mode strengths of C_S are assigned a value of zero. Periodic boundary condition are applied in the stream- and span-wise directions, while no-slip boundary condition is used at the walls. The simulation results will be compared to the Direct Numerical Simulation (DNS) data from [79] at $Re_\tau = 395$. Based on the study of the effect of computational grid size from [13], we use a reasonably fine mesh with the details in Table 2.1. Note that $\Delta x^+ = \Delta x u_\tau/\nu$, $\Delta z^+ = \Delta z u_\tau/\nu$ and $y^+ = y u_\tau/\nu$, are calculated using value of u_τ, corresponding to the value of Re_τ in the DNS database. In order to capture the sharp gradients in the near-wall region, we specify a grading along y direction. We use the van Driest damping function to correct the behavior of the Smagorinsky model in the near-wall region [13].

Figure 2.8 presents the normalized time-averaged streamwise component of the velocity along with the mean and the confidence interval. The profile is compared with the DNS data, the deterministic solution (DET) at the mean value of C_S and also with the results from non-intrusive polynomial chaos. The stochastic mean from IPC is found to be close

Parameter	Value	Units
Kinematic viscosity (ν)	2×10^{-5}	$m^2 s^{-1}$
Pressure gradient ($-dp_0/dx$)	5×10^{-5}	ms^{-2}
Target Reynolds number (Re_τ)	395	-
Smagorinsky parameter (C_S)	$\mathcal{U}(0.075, 0.125)$	-

Mesh	Cells along x, y, z	Total cells	Δx^+	Δz^+	y^+
M1	$80 \times 100 \times 60$	480000	19.75	13.16	0.96

Table 2.1 | Details of the physical parameters and the computational mesh.

to DET and mean of NIPC, and deviates slightly from DNS in the same manner as the DET solution. In contrast to IPC, the NIPC approach under predicts the variance. The uncertainty in the LES model parameter is reflected in the solution in the regions close to the wall and the channel center. In Figure 2.9 the normalized square-root of the second order velocity moments and Reynolds shear stress are plotted together with their confidence regions. As evident, the stochastic mean of stresses are close to that of NIPC and DET solution. The deviation from DNS can mainly be attributed to the use of a relatively coarse mesh and the choice of LES model. Both the IPC and NIPC methods predicts high variance near the wall, with almost zero uncertainty in the channel center. This is expected as the Smagorinsky model parameter, when changed, usually affects significantly near the wall as compared to the channel center. While, the confidence region of intrusive method mostly overlaps with that of the non-intrusive counterpart, in some regions, NIPC still underestimates the uncertainty.

2.6. CONCLUSIONS

The IPC method, as it involves solving a lesser number of equations than NIPC, can be of great use when the deterministic simulation is already computationally expensive. To develop an IPC solver it is important to efficiently decouple the system of equations and ensure the overhead due to coupling is not significant.

In this work, intrusive polynomial chaos for CFD simulations using a popular finite-volume library OpenFOAM was presented. The aim was to use the existing deterministic solver for the incompressible Navier-Stokes equations, to develop a new stochastic solver for the quantification of the uncertainties involved and study its non-linear propagation.

To this end, we tested this solver for various standard CFD problems involving laminar and turbulent flows. The plane Poiseuille flow with uncertain kinematic viscosity was discussed first. Here, we realized a significant effect of the uncertainty in the re-circulation region of the flow. The results were also compared with the non-intrusive counterpart with same polynomial order. The results from IPC were found to be very close to NIPC, verifying its implementation using OpenFOAM. Next, we examined the turbulent channel flow with uncertain LES model parameter. The results were found to be in accordance with the non-intrusive gPC method. Deviations in the variance predicted by the two variants of gPC approaches can be attributed to the use of the two very different numerical methods to

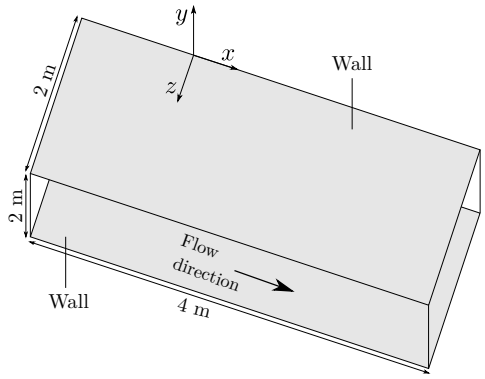


Figure 2.7 | Graphical representation of the turbulent channel flow.

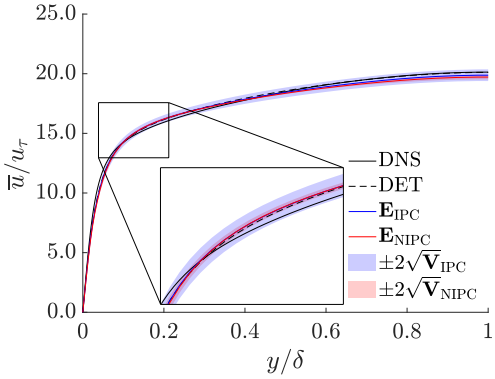


Figure 2.8 | Normalized time-averaged streamwise component of velocity.

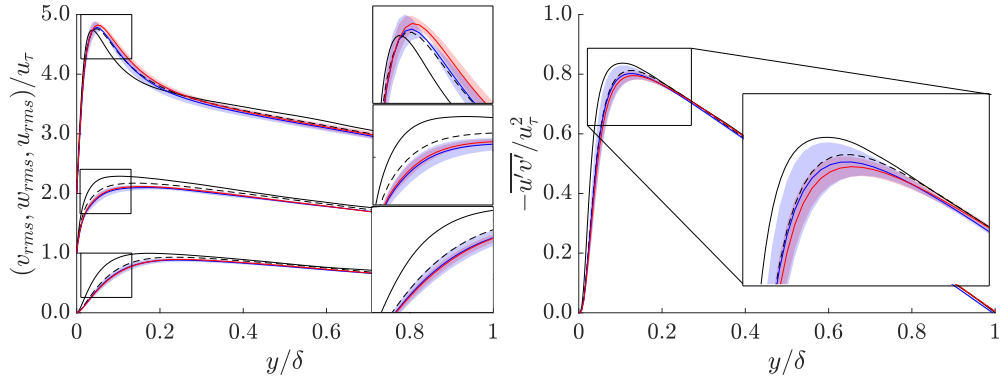


Figure 2.9 | Normalized square-root of the second order velocity moments (left) and normalized Reynolds shear stress (right). [Legend: Figure 2.8]

estimate the expansion coefficients in both variants.

Through this work, for UQ in CFD, we bring to light an alternate to the MC and the NIPC approaches. The promising results obtained from IPC method encourages to further pursue research in this direction. Pseudo-spectral methods along with early truncated expansions will be some next steps to reduce the computational cost.

3

QUANTIFICATION AND PROPAGATION OF MODEL-FORM UNCERTAINTIES

The content of this chapter was published in J. Parekh and R. Verstappen. Quantification and propagation of model-form uncertainties in RANS turbulence modeling via intrusive polynomial chaos. International Journal for Uncertainty Quantification., 13:1–29, 2023.

ABSTRACT: Undeterred by their inherent limitations, Reynolds-Averaged Navier-Stokes (RANS) based modeling is still considered the most recognized approach for several Computational Fluid Dynamics (CFD) applications. Recently, in the turbulence modeling community, quantification of model-form uncertainties in RANS has attracted significant interest. We present a stochastic RANS solver with an efficient implementation of the Intrusive Polynomial Chaos (IPC) method in OpenFOAM. The stochastic solver quantifies and propagates the uncertainties associated with the output of the RANS model (eddy viscosity or Reynolds-stress-tensor). Two distinct high-dimensional variants of the uncertainties are considered, namely, the random eddy viscosity field (REVF) and the random Reynolds stress tensor field (RRSTF). The randomness is introduced in the approximated eddy viscosity field and the Reynolds stress tensor, while asserting the realizability [100, 134]. The stochastic RANS solver has been tested on various benchmark problems for RANS turbulence modeling. In this study, we discuss two important problems where the stochastic RANS solver shows significantly better performance than the traditional Uncertainty Quantification (UQ) methods. The first problem analyzed is the flow over periodic hills with a REVF, while the second stochastic problem considered is the flow in a square duct with a RRSTF. Along with the comparison for three different RANS turbulence models, a detailed analysis of the stochastic solver based on various influential model parameters is also presented. The IPC based stochastic solver demonstrated the potential to be used in the UQ analysis of further complex CFD applications, especially when a large number of deterministic simulations is not feasible, e.g. windfarm CFD simulations.

3.1. INTRODUCTION

The Reynolds-Averaged Navier-Stokes (RANS) based turbulence models are the most common choice for reasonably accurate and cost-effective simulation of several Computational Fluid Dynamics (CFD) applications involving turbulent flows. A lot of these predictions contain uncertainties due to the modeling of the Reynolds stresses which combines the mean flow with the small-scale turbulence [82]. The coefficients in a RANS turbulence model are often calibrated for certain types of flows with various modeling assumptions, and therefore a particular model can't be employed to the entire class of turbulent flows.

Previously, a large number of efforts have mostly focused on quantifying the uncertainties in RANS turbulence model via parametric approach, where the closure model parameters are either perturbed or directly injected with randomness [18, 74, 85]. Recently, the shortcomings of this parametric approach have been recognized by the turbulence modeling community and efforts are made towards the non-parametric approaches [82]. Commonly, uncertainties in the Reynolds stress are introduced by perturbing its eigenvalues [22, 38], eigenvectors [46] or tensor invariants [70, 132]. Methods involving perturbations of the closure model parameters, eigenvalues, eigenvectors and invariants of the Reynolds stress tensor can be classified as global perturbation method as the randomness is uniformly introduced throughout the computational domain. As a consequence, the resulting UQ problem turns into a low-dimensional forward problem that can be solved efficiently via traditional UQ methods [22, 70, 82]. A major drawback of these low-dimensional global methods is the inaccurate estimation (mostly over-prediction) of the uncertainties in the Quantities of Interest (QoIs), resulting from the assumption of globally uniform input randomness.

Recent studies on RANS model-form UQ have employed two different high-dimensional local perturbation methods. Here, the uncertainty is characterized via, (a) Random Eddy Viscosity Field (REVF) with randomness injected directly into the modeled eddy viscosity field using Gaussian random field with an assumed spatial covariance [60], or (b) Random Reynolds Stress Tensor Field (RRSTF) modeled using spatially correlated positive-definite random matrices ensuring realizability of the modeled Reynolds stress [9, 129, 134]. The REVF method considers uncertainties arising from the imprecise closure model parameters or from the model-form uncertainties. While the RRSTF approach accounts for the scenarios where the Boussinesq approximation fails or is inapplicable.

Several approaches have been developed to quantify and propagate uncertainties in stochastic simulations. These approached include, moment method, interval analysis method, sensitivity derivative method, Monte Carlo (MC) method, polynomial chaos method, and more recently data-driven methods [124, 128]. Conventional UQ methods like Monte Carlo (MC) and its variants require thousands of samples, which makes them impractical for use, especially for the computationally expensive deterministic CFD simulations. The Generalized Polynomial Chaos (gPC) method serves as a better alternative to MC method, as it has been proven to show exponential convergence at a significantly lower computational cost [44, 125]. Many researchers have implemented gPC for a range of CFD problems including laminar, turbulent, compressible and incompressible flows [44, 124, 125, 138]. In the gPC method, the random variable is represented in terms of a polynomial expansion. The two methods to determine the expansion coefficients are - the Intrusive Polynomial Chaos (IPC) method and the Non-intrusive Polynomial Chaos (NIPC)

method. In the case of the IPC method, the unknown expansion coefficients are computed (solved for) by projecting the resulting set of equations on orthogonal polynomials (basis functions). This requires modifications to the existing deterministic solver to account for the coupling of the expansion coefficients. On the other hand, the NIPC method avoids any modifications to the deterministic solver by employing various sampling or quadrature methods. However, to achieve an equivalent level of accuracy, specifically for a higher dimensional random space, NIPC requires a large number of deterministic solves than the number of equations to be solved in the IPC approach. Moreover, for such a random space, using NIPC method can result into a significant aliasing error from the approximation of the exact gPC expansion. Thus, for a multi-dimensional problem, the IPC method can deliver more accurate solutions at a much lower computational cost than the NIPC method [135].

In the current work, we primarily focus on the use of IPC method for quantification and propagation of RANS based model-form uncertainties. We present an efficient implementation of the coupled stochastic RANS solver based on the IPC method using an existing (deterministic) incompressible flow solver. As a model code for the simulations, we use OpenFOAM [25], which is basically a C++ toolbox used for developing numerical solvers, along with pre-/post-processing utilities to solve continuum mechanics problems. OpenFOAM (a) is a highly templated code, that allows the users to customize the default routines as needed for their applications, and, (b) provides access to frequently used tensor operations (gradient, divergence, laplacian, etc.) directly at the top-level code. This avails enough flexibility to implement the IPC framework for UQ analysis in CFD. The stochastic RANS solver numerically solves a coupled system of equations in order to estimate the moments (mean and variance) of the QoIs. The number of equations, and hence the convergence and computational cost depends on various solver parameters. The sensitivity of the predicted stochasticity w.r.t. these parameters is studied in detail. Techniques to overcome the so-called *curse of dimensionality*, comparison of stochasticity in different turbulence models and reduction of uncertainty via data assimilation are also presented in this work. The stochastic solver developed so far is capable of delivering an accurate prior for Bayesian approaches employed for data-assimilation and calibration process in turbulence modeling.

The next sections of the paper are structured as follows. Section 3.2 describes the gPC methods with emphasis on the intrusive variant. In section 3.3, we introduce the different RANS turbulence models. The stochastic formulation of the random eddy viscosity field and the random Reynolds stress tensor field is presented in section 3.4. In section 3.5, we discuss the algorithm and the implementation of the stochastic RANS solver based on an existing deterministic solver in OpenFOAM. The numerical results obtained for two different turbulent flow benchmark problems – flow over periodic hills with REVF and flow in a square duct with RRSTF are discussed in section 3.6. Section 3.7 concludes the paper.

3.2. GENERALIZED POLYNOMIAL CHAOS

The Generalized Polynomial Chaos approach is based on the spectral decomposition of the random variable(s) f, in terms of basis polynomials containing randomness ψ_i (known a priori) and the unknown deterministic expansion coefficients f_i, as $f(x,q) = \sum_{i=0}^{\infty} f_i(x)\psi_i(q)$. The two methods to determine the expansion coefficients are - the Intrusive Polynomial Chaos (IPC) method and the Non-intrusive Polynomial Chaos (NIPC) method. In the IPC

method, the original model is reformulated resulting in governing equations for the PC mode strengths (coefficients) of the model output. While in the NIPC method, these coefficients are approximated using quadrature for numerical evaluation of the projection integrals. The level of accuracy of these methods can be associated with the chosen degree of the polynomials in gPC.

3.2.1. INTRUSIVE POLYNOMIAL CHAOS

In order to demonstrate the application of IPC, we first consider a general stochastic differential equation

$$\mathscr{L}(\boldsymbol{x}, t, \omega; \boldsymbol{v}(\boldsymbol{x}, t, \omega)) = S(\boldsymbol{x}, t, \omega), \tag{3.1}$$

where $\mathscr{L}$ is usually a nonlinear differential operator consisting of space and/or time derivatives, $\boldsymbol{v}(\boldsymbol{x}, t, \omega)$ is the solution and $S(\boldsymbol{x}, t, \omega)$ is the source term. The random event ω represents the uncertainty in the system, introduced via uncertain parameters, the operator, source term, initial/boundary conditions, etc. The complete probability space is given by $(\Omega, \mathscr{A}, \mathscr{P})$, where Ω is the sample space such that $\omega \in \Omega$, $\mathscr{A} \subset 2^{\Omega}$ is the σ-algebra on Ω and $\mathscr{P} : \mathscr{A} \mapsto [0,1]$ is the probability measure on $(\Omega, \mathscr{A})$.

We now employ an IPC method for the propagation of uncertainty - the Galerkin polynomial chaos method [125]. It provides the spectral representation of the stochastic solution and results into higher order approximations of the mean and variance. Galerkin polynomial chaos method is a non-statistical method where the uncertain parameter(s) and the solution become random variables. These random variables are approximated using the polynomial chaos (polynomial of random variables) as follow [35]

$$\boldsymbol{v}(\boldsymbol{x}, t, \omega) \approx \sum_{i=0}^{P} \boldsymbol{v}_i(\boldsymbol{x}, t) \psi_i(\boldsymbol{\xi}(\omega)). \tag{3.2}$$

It is worth noting that the expansion (3.2) is indeed the decomposition of a random variable into a deterministic component, the expansion coefficients $\boldsymbol{v}_i(\boldsymbol{x}, t)$ and a stochastic component, the random basis functions (polynomial chaoses) $\psi_i(\boldsymbol{\xi}(\omega))$. Here, $\boldsymbol{\xi}(\omega)$ is the vector of d independent random variables $\{\xi_1, ..., \xi_d\}$, corresponding to d uncertain parameters. Based on the dimension of $\boldsymbol{\xi}$ (which here is d) and the highest order n of the polynomials $\{\psi_i\}$, the infinite summation has been truncated to $P+1 = (d+n)!/(d!\,n!)$ terms.

An important property of the basis $\{\psi_i\}$ is their orthogonality with respect to the probability density function (PDF) of the uncertain parameters, $\langle \psi_i \psi_j \rangle = \langle \psi_i^2 \rangle \delta_{ij}$. Here, δ_{ij} is the Kronecker delta and $\langle \cdot, \cdot \rangle$ denotes the inner product in the Hilbert space of the variables $\boldsymbol{\xi}$, $\langle f(\boldsymbol{\xi}) g(\boldsymbol{\xi}) \rangle = \int f(\boldsymbol{\xi}) g(\boldsymbol{\xi}) w(\boldsymbol{\xi}) d\boldsymbol{\xi}$. The weighting function $w(\boldsymbol{\xi})$ is the probability density function of the uncertain parameters. Such polynomials already exist for some standard distributions which can be found in the Askey scheme [137] . For example, a Normal distribution leads to Hermite-chaos, while Legendre-chaos corresponds to a Uniform distribution. For other commonly used distributions or any arbitrary distribution, one can for example use the Gram-Schmidt algorithm [127] to construct the orthogonal polynomials.

Substituting (3.2) in the general stochastic differential equation (3.1), we obtain

$$\mathscr{L}\left(\boldsymbol{x}, t, \omega; \sum_{i=0}^{P} \boldsymbol{v}_i \psi_i\right) \approx S. \tag{3.3}$$

In order to ensure that the truncation error is orthogonal to the functional space spanned by the basis polynomials $\{\psi_i\}$, a Galerkin projection of the above equation is performed onto each polynomial $\{\psi_k\}$,

$$\left\langle \mathscr{L}\left(\boldsymbol{x}, t, \omega; \sum_{i=0}^{P} \boldsymbol{v}_i \psi_i\right), \psi_k \right\rangle = \langle S, \psi_k \rangle, \quad k = 0, 1, ..., P. \tag{3.4}$$

After using the orthogonality property of the polynomials, we obtain a set of $P+1$ deterministic coupled equations for all the random modes of the solution $\{\boldsymbol{v}_0, \boldsymbol{v}_1, ..., \boldsymbol{v}_k\}$. Following the definition, the mean and the variance of the solution are given by

$$\mathbf{E}[\boldsymbol{v}] = \mu_{\boldsymbol{v}} = \boldsymbol{v}_0(\boldsymbol{x}, t), \quad \mathbf{V}[\boldsymbol{v}] = \sigma^2_{\boldsymbol{v}} = \sum_{i=1}^{P} \boldsymbol{v}_i^2(\boldsymbol{x}, t) \langle \psi_i^2 \rangle. \tag{3.5}$$

As the coefficients $\boldsymbol{v}_i(\boldsymbol{x}, t)$ are known, the probability distribution of the solution can be obtained.

3.2.2. NON-INTRUSIVE POLYNOMIAL CHAOS

Unlike the intrusive polynomial chaos, its non-intrusive counterpart approximates the polynomial expansion coefficients $\boldsymbol{v}_i(\boldsymbol{x}, t)$ instead of directly solving for them i.e. the projection in equation (3.4) is approximated using numerical quadrature, with quadrature points and respective weights based on the joint probability density function of the input uncertainty. This results into a decoupled system of equations (similar to the deterministic equation) for each mode of the polynomial expansion. For a polynomial chaos of order n with d random variables, the NIPC method requires $(n+1)^d$ deterministic solves to approximate all the expansion coefficients. As noted, the number of deterministic evaluations increase exponentially with the number of random dimensions. Sparse-grid methods like Latin Hypercube sampling can significantly reduce the number of samples to be evaluated using the deterministic solver. The following linear system of equations is solved in order to obtain the expansion coefficients:

$$\begin{pmatrix} \psi_0(\xi_0) & \psi_1(\xi_0) & \cdots & \psi_P(\xi_0) \\ \psi_0(\xi_1) & \psi_1(\xi_1) & \cdots & \psi_P(\xi_1) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_0(\xi_P) & \psi_1(\xi_P) & \cdots & \psi_P(\xi_P) \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_0(\boldsymbol{x}, t) \\ \boldsymbol{v}_1(\boldsymbol{x}, t) \\ \vdots \\ \boldsymbol{v}_P(\boldsymbol{x}, t) \end{pmatrix} = \begin{pmatrix} \boldsymbol{v}(\boldsymbol{x}, t, \xi_0) \\ \boldsymbol{v}(\boldsymbol{x}, t, \xi_1) \\ \vdots \\ \boldsymbol{v}(\boldsymbol{x}, t, \xi_P) \end{pmatrix} \tag{3.6}$$

As mentioned earlier, the stochastic solution is obtained as soon as we solve for the expansion coefficients using equation (3.5). Further discussion on NIPC is deferred to the extensive literature [44, 135].

3.2.3. GLOBAL SENSITIVITY ANALYSIS WITH SOBOL INDICES

In order to determine the relative influence of each random variable on the QoIs we can employ a global sensitivity analysis based on a variance-based approach called Sobol indices method [106]. Once the coefficients of the polynomial chaos expansion are determined, the computation of the Sobol indices is straight-forward. The total variance (D) in terms of the expansion coefficients can be written as:

$$D = \mathbf{V}[\boldsymbol{v}] = \sum_{i=1}^{P} \boldsymbol{v}_i^2(\boldsymbol{x}, t)\langle \psi_i^2 \rangle, \tag{3.7}$$

which can be decomposed as [113]:

$$D = \sum_{i=1}^{i=d} D_i + \sum_{1\le i<j\le d}^{i=d-1} D_{i,j} + \sum_{1\le i<j<k\le d}^{i=d-2} D_{i,j,k} + \quad \cdots \quad + D_{1,2,\cdots,d}, \tag{3.8}$$

where $(D_{i_1,\cdots,i_s})$ are partial variances given by

$$D_{i_1,\cdots,i_s} = \sum_{\beta\in\{i_1,\cdots,i_s\}} \boldsymbol{v}_\beta^2(\boldsymbol{x}, t)\langle \psi_\beta^2 \rangle, \quad 1 \le i_1 < \cdots < i_s \le d. \tag{3.9}$$

The Sobol indices $(S_{i_1,\cdots,i_s})$ are then defined as

$$S_{i_1,\cdots,i_s} = \frac{D_{i_1,\cdots,i_s}}{D}, \tag{3.10}$$

such that,

$$\sum_{i=1}^{i=d} S_i + \sum_{1\le i<j\le d}^{i=d-1} S_{i,j} + \sum_{1\le i<j<k\le d}^{i=d-2} S_{i,j,k} + \quad \cdots \quad + S_{1,2,\cdots,d} = 1 \tag{3.11}$$

Thus, the Sobol indices basically measures the combined sensitivity arising from the each random variable contribution (S_i) and from contribution due to their interactions $(S_{i,j}, S_{i,j,k}, ...)$. The combined effect of an uncertain variable with index i is therefore defined as the sum of the partial Sobol indices that includes the contribution from the i_{th} variable

$$S_{T_i} = \sum_{L_i} \frac{D_{i_1,\cdots,i_s}}{D}; \quad L_i = \{(i_1, \cdots, i_s) : \exists\, k, 1 \le k \le s, i_k = i\}. \tag{3.12}$$

Thus, the Sobol indices (S_{T_i}) can be used to estimate and compare the contribution of each uncertain parameter towards the uncertainty in QoIs.

3.3. DETERMINISTIC TURBULENCE MODELS

The Navier-Stokes equations for an incompressible flow are given by

$$\begin{aligned} \nabla \cdot \boldsymbol{u} &= 0, \\ \frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} &= -\nabla p + \nabla\cdot(\nu\nabla\boldsymbol{u}), \end{aligned} \tag{3.13}$$

where $\boldsymbol{u}$ is the velocity, p is the pressure and ν is the kinematic viscosity. Solving these equations for turbulent flows, in general, involves resolving a range of scales of motion which is very expensive to compute directly. For many practical or engineering problems, the information on how-the-turbulence-effects-the-mean-flow suffices, and the details of the entire range of the scales of motion is not always needed. For such scenarios, we can use the Reynolds-Averaged Navier-Stokes (RANS) approach, in which a system of mean flow equations is derived using Reynolds averaging. The instantaneous velocity $\boldsymbol{u}$ is decomposed

into the mean velocity $\overline{\boldsymbol{u}}$ and its fluctuation u', such that $\boldsymbol{u} = \overline{\boldsymbol{u}} + \boldsymbol{u}'$. Similarly, the pressure p is also expressed in terms of a mean pressure $\overline{p}$ and its fluctuation p'. Substituting these decompositions in the Navier-Stokes equations and applying an averaging procedure, results in the following mean flow equation[1],

$$\begin{aligned} \nabla \cdot \overline{\boldsymbol{u}} &= 0, \\ \frac{\partial \overline{\boldsymbol{u}}}{\partial t} + (\overline{\boldsymbol{u}} \cdot \nabla)\overline{\boldsymbol{u}} &= -\nabla \overline{p} + \nabla \cdot (\nu \nabla \overline{\boldsymbol{u}}) - \nabla \cdot \mathbf{R}, \end{aligned} \tag{3.14}$$

where $\mathbf{R} = \overline{\boldsymbol{u}'\boldsymbol{u}'}$ are the Reynolds stresses which couples the mean flow to the turbulence. A popular class of RANS models is the linear-eddy-viscosity models. These models employ the Boussinesq approximation, where Reynolds stresses are expressed linearly in terms of mean strain-rate,

$$-\mathbf{R} \approx 2\nu_t \overline{\boldsymbol{S}} - \frac{2}{3}\boldsymbol{I}k \tag{3.15}$$

where ν_t is the eddy viscosity, $\overline{\boldsymbol{S}} := (\nabla\overline{\boldsymbol{u}} + (\nabla\overline{\boldsymbol{u}})^T)/2$ is the mean strain-rate tensor, $\boldsymbol{I}$ is the second order identity tensor, and $k := \frac{1}{2}\,\mathrm{tr}(\mathbf{R})$ is the turbulent kinetic energy. Although the eddy viscosity model has shown a decent performance for a range of flows, it still suffers to accurately capture the flows with significant curvature, separation, impinging, anisotropy, etc. The eddy viscosity is computed after solving the equation(s) for the turbulent flow quantities like the turbulent kinetic energy k and the turbulent energy dissipation ϵ, or the specific dissipation ω. In the present work, we employ three different turbulence models, namely, the $k-\omega$ [126], the $k-\omega$ SST [78] and the Launder-Sharma $k-\epsilon$ [64] models. For the sake of brevity, we defer the discussion of these models to the available literature [64, 78, 126].

In the next section, we provide a formulation to introduce randomness in the RANS equations, resulting into a system of the stochastic RANS equations. Note that the eddy viscosity based RANS models as well as other RANS models can be studied in this framework.

3.4. STOCHASTIC FORMULATION

To quantify the model uncertainties in the RANS numerical simulations, we focus on the non-parametric approach which can be generally categorized into two methods [133] - (a) where the uncertainties are introduced into *model forms*, e.g., in the transport equations for the fields like k, ω, ϵ etc., and (b) where the uncertainties are inserted into the *model output*, e.g., in the eddy viscosity field or in the Reynolds stress tensor field. The latter being employed in the present work using the RANS models is discussed in section 3.3. The randomness in the eddy viscosity field can be used to quantify uncertainties resulting from the improper selection of the closure parameters in the RANS model which has a direct influence on the computed turbulent viscosity. While the randomness in the Reynolds stress tensor field is more suitable for quantifying the uncertainties where the Boussinesq approximation fails or is inapplicable. In the following sections, we will discuss and derive the steps involved in introducing uncertainties in these fields which will result into stochastic Navier-Stokes equations.

[1] Similar to the filtered LES equations in chapter 2, the overbar notation is now used for the RANS equation.

3.4.1. RANDOM EDDY VISCOSITY FIELD

The RANS turbulence modeling involves solving transport equations for turbulent quantities (k, ω, or ϵ), which contain closure coefficients that are usually determined by the data from experiments or direct numerical simulations (DNS) [17]. In a parametric approach of quantifying the uncertainties in RANS simulations, a joint probability distribution function of these closure coefficient (parameters) is propagated to obtain uncertain bounds for the velocity, pressure, eddy viscosity, or other quantities-of-interest (QoIs) [19, 74]. Thus, we obtain a globally randomized eddy viscosity field as reflection of the random closure coefficients. However, we would like to have a locally random-eddy-viscosity-field (REVF) considering that the Boussinesq approximation is only locally fallible. Thus, in the present work we use the method that locally randomizes the eddy viscosity field obtained from the deterministic solve [14]. We use a Gaussian random field with a prescribed covariance model to obtain a log-normal REVF,

$$\log \nu_t(\boldsymbol{x}, \omega) = \log \nu_t^{(det)}(\boldsymbol{x}) + Z(\boldsymbol{x}, \omega) \tag{3.16}$$

where ω represents the uncertainty in the system, $\nu_t^{(det)}$ is the eddy viscosity field obtained from a deterministic RANS solve and $Z(\boldsymbol{x}, \omega)$ is a Gaussian random field. The log-normal field guarantees the positivity of the REVF.

The RANS equations (and later the stochastic RANS) are defined in a bounded domain $D \subset \mathbb{R}^d$ ($d = 1, 2, 3$). A zero-mean Gaussian random field with a squared exponential kernel function can be characterized as,

$$\begin{aligned} \mathbf{E}[Z(\boldsymbol{x}, \cdot)] &= 0, \\ \mathbf{Cov}(Z(\boldsymbol{x}_1, \cdot), Z(\boldsymbol{x}_2, \cdot)) = C(\boldsymbol{x}_1, \boldsymbol{x}_2) &= \sigma_c^2(\boldsymbol{x}) \exp\left(-\frac{(x_1 - x_2)^2}{l_x^2} - \frac{(y_1 - y_2)^2}{l_y^2} - \frac{(z_1 - z_2)^2}{l_z^2} \right), \end{aligned} \tag{3.17}$$

where $C : \mathbb{R}^d \to \mathbb{R}_+$ with σ_c^2 being the marginal variance; l_x, l_y and l_z the correlation lengths along the x, y and z directions, respectively. This random field can be discretized using a truncated Karhunen-Loève (KL) decomposition,

$$Z(\boldsymbol{x}, \omega) = \sum_{i=1}^{N_{KL}} \sqrt{\lambda_i} \phi_i(\boldsymbol{x}) \xi_i, \quad \xi_i \sim \mathcal{N}(0, 1), \tag{3.18}$$

where, N_{KL} is the number of terms retained in the decomposition, λ_i and ϕ_i are the eigenvalues and the corresponding eigenfunctions for the ith mode, which is obtained by numerically solving the Fredholm integral [66],

$$\int_D C(\boldsymbol{x}_1, \boldsymbol{x}_2) \phi(\boldsymbol{x}_1) d\boldsymbol{x}_1 = \lambda \phi(\boldsymbol{x}_2). \tag{3.19}$$

A non-stationary and anisotropic kernel can be used to encode the structure of the flow field into the KL basis functions $\sqrt{\lambda_i}\phi_i(\boldsymbol{x})$. This can be achieved by using spatially varying and/or anisotropic correlation lengths [134].

Using the above KL decomposition, we can now express the REVF in terms of a polynomial chaos expansion,

$$\nu_t(\boldsymbol{x}, \omega) \approx \sum_{i=0}^{P} \nu_{t_i}(\boldsymbol{x}) \psi_i(\boldsymbol{\xi}), \tag{3.20}$$

where $\psi_i(\boldsymbol{\xi})$ is the multidimensional (N_{KL} dimensions) Hermite polynomials in the Gaussian set $\{\xi_i\}$ obtained from the KL decomposition. The value of P depends on number of dimensions $d = N_{KL}$ and the chosen polynomial order n as discussed in section 3.2. The expansion coefficients are given by [100],

$$\nu_{t_i}(\boldsymbol{x}) = \nu_t^{(det)} \frac{p!}{\langle \psi_i^2 \rangle} f_p(\boldsymbol{x}) \prod_{j=1}^{p} \sqrt{\lambda_{k(j)}} \phi_{k(j)}(\boldsymbol{x}) \tag{3.21}$$

where p is the order of the polynomial ψ_i, $\prod_{j=1}^{p} \sqrt{\lambda_{k(j)}} \phi_{k(j)}(\boldsymbol{x})$ is the product of those $\sqrt{\lambda_k}\phi_k(\boldsymbol{x})$, where k is an index on at least one of the ξ_k making up ψ_i, and

$$f_i(\boldsymbol{x}) = \frac{\sigma_Z(\boldsymbol{x})^i}{i!} \exp\left[\mu_Z(\boldsymbol{x}) + \frac{1}{2}\sigma_Z^2(\boldsymbol{x})\right] \tag{3.22}$$

where $\mu_Z(\boldsymbol{x})$ and $\sigma_Z^2(\boldsymbol{x})$ are the mean and variance of the Gaussian field. In order to obtain $\mathbf{E}[\nu_t(\boldsymbol{x},\omega)] = \nu_{t_0}(\boldsymbol{x}) = \nu_t^{(det)}$, we use a Gaussian field with $\mu_Z(\boldsymbol{x}) = \mathbf{E}[Z(\boldsymbol{x},\cdot)] = -\frac{1}{2}\sigma_Z^2(\boldsymbol{x})$. The stochastic form of RANS equations in terms of the REVF can be written as,

$$\begin{aligned} \nabla \cdot \overline{\boldsymbol{u}} &= 0, \\ \frac{\partial \overline{\boldsymbol{u}}}{\partial t} + (\overline{\boldsymbol{u}} \cdot \nabla)\overline{\boldsymbol{u}} &= -\nabla \overline{p}_* + \nabla \cdot \left((\nu + \nu_t(\boldsymbol{x},\omega))\nabla \overline{\boldsymbol{u}}\right) \end{aligned} \tag{3.23}$$

where $\overline{p}_* := \overline{p} - \frac{2}{3}k$. Upon substituting the polynomial chaos expansions of the velocity, the pressure, and the laminar and turbulent viscosity; taking the Galerkin projection and using the orthogonality of the polynomials, we obtain, for $k = 0, 1, ..., P$, the set of equations for the evolution of the modes (coefficients) of velocity,

$$\begin{aligned} \nabla \cdot \overline{\boldsymbol{u}}_k &= 0, \\ \frac{\partial \overline{\boldsymbol{u}}_k}{\partial t} + \sum_{i=0}^{P}\sum_{j=0}^{P} (\overline{\boldsymbol{u}}_i \cdot \nabla)\overline{\boldsymbol{u}}_j M_{ijk} &= -\nabla \overline{p}_{*_k} + \sum_{i=0}^{P}\sum_{j=0}^{P} \nabla \cdot ((\nu_i + \nu_{t_i}(\boldsymbol{x}))\nabla \overline{\boldsymbol{u}}_j) M_{ijk}, \end{aligned} \tag{3.24}$$

where $M_{ijk} = \dfrac{\langle \psi_i \psi_j \psi_k \rangle}{\langle \psi_k \psi_k \rangle}$. The original system of equations is transformed into a system of $(P+1)$ divergence-free constraints on velocity modes and $(P+1)$ coupled equations in velocity and pressure modes. Note that the steps involved in obtaining the above set of equations can be found in [85]. The solution procedure adopted for this large system of equations is deferred to section 3.5.

Although this stochastic model can be used to describe the uncertainties (with local variations) due to imprecise closure parameters in the deterministic RANS model, it inevitably suffers from the shortcomings of the Boussinesq approximation. This problem can be addressed by directly inserting randomness in the Reynolds stress tensor.

3.4.2. RANDOM REYNOLDS STRESS TENSOR FIELD

The idea of random Reynolds stress tensor emerges from the work by Soize [108], who was the first to derive the maximum entropy probability distribution of a symmetric positive

definite matrix with a given mean and standard deviation applied to a structural mechanics problem. Based on this work, Xiao et al. [134] recently proposed a random matrix approach as a promising alternative to the previously developed physics-based approach [22, 23]. In their work, Xiao et al. [134] introduced randomness in the Reynolds stress components using a Gaussian random field with a prescribed covariance function having spatially varying correlation lengths and variance, while guaranteeing the realizability. The following details on the construction of a random-Reynolds-stress-tensor-field (RRSTF) are mostly based on the original work from [108, 134] and some elements from our construction of REVF. We will use NIPC, as discussed in section 3.2, to compute the polynomial chaos expansion of a RRSTF.

RANDOM MATRIX FOR REYNOLDS STRESSES

Since the Reynolds stress tensor is a covariance of velocity fluctuations, it must be a symmetric positive semidefinite $\mathbb{M}_d^{+0}$. The Reynolds stresses are usually positive definite $\mathbb{M}_d^{+}$, except for some extreme cases, like the turbulence-free regions or that on the wall boundaries where the eigenvalues goes to zero. The Reynolds stress obtained from a deterministic simulation can be considered to be the mean of the random Reynolds tensor, i.e., $\mathbf{E}[\mathbf{R}] = \mathbf{R}^{(det)}$, where $\mathbf{R}, \mathbf{R}^{(det)} \in \mathbb{M}_d^{+}$. The randomness in the Reynolds stress can be introduced via a normalized positive definite random matrix $\mathbf{G}$ having the identity matrix $\mathbf{I}$ as mean. The dispersion parameter δ indicates the variance in $\mathbf{G}$ and is given by,

$$\delta = \sqrt{\frac{1}{d}\mathbf{E}[||\mathbf{G}-\mathbf{I}||_F^2]}, \tag{3.25}$$

where $||\cdot||_F$ is the Frobenius norm. The positive definiteness of $\mathbf{G}$ is guaranteed if the dispersion parameter is chosen in the range $0 < \delta < \sqrt{(d+1)/(d+5)}$ [107], which for $d = 3$ gives, $0 < \delta < 1/\sqrt{2}$. The matrix $\mathbf{G}$ is first decomposed (Cholesky factorization) as $\mathbf{G} = \mathbf{L}^T\mathbf{L}$, where $\mathbf{L}$ is the upper triangular matrix. The off-diagonal terms of $\mathbf{L}$ are given by,

$$\mathrm{L}_{ij} = \frac{\delta}{\sqrt{d+1}}\xi_{ij}, \quad \text{for} \quad i < j, \quad \xi_{ij} \sim \mathcal{N}(0,1). \tag{3.26}$$

While the diagonal terms are given by,

$$\mathrm{L}_{ii} = \frac{\delta}{\sqrt{d+1}}\sqrt{2g_i}, \quad \text{for} \quad i = 1,2,3, \tag{3.27}$$

where g_i is a gamma random variable, $g_i > 0$, with probability density function,

$$p(g_i) = \frac{g_i^{k_i-1}\exp(-g_i)}{\Gamma(k_i)}, \tag{3.28}$$

where $\Gamma(k_i)$ is a gamma distribution and $k_i = (d+1)/2\delta^2 + (1-i)/2$, is its shape parameter with scaling parameter set to 1. Thus, based on the shape parameter, the diagonal terms of $\mathbf{L}$ will take different values.

Once we obtain such random matrix $\mathbf{G}$, we can then construct a random Reynolds matrix as $\mathbf{R} = \mathbf{L}_R^T\mathbf{G}\mathbf{L}_R$, where $\mathbf{L}_R$ is a upper triangular matrix resulting from the decomposition (Cholesky factorization) of the mean Reynolds stress, $\mathbf{R}^{(det)} = \mathbf{L}_R^T\mathbf{L}_R$.

RANDOM MATRIX FIELD FOR REYNOLDS STRESSES

Using the procedure similar to that in the previous section, we now establish random Reynolds stresses with spatial correlation $\mathbf{R} = \mathbf{R}(\boldsymbol{x}, \omega)$ at different locations. The terms in the random upper triangular matrix $\mathbf{L} = \mathbf{L}(\boldsymbol{x}, \omega)$ are correlated as,

$$\begin{aligned} \mathbf{Cov}(\mathrm{L}_{ij}(\boldsymbol{x}_1, \cdot), \mathrm{L}_{ij}(\boldsymbol{x}_2, \cdot)) &= C(\boldsymbol{x}_1, \boldsymbol{x}_2), \quad i < j, \\ \mathbf{Cov}(\mathrm{L}^2_{ii}(\boldsymbol{x}_1, \cdot), \mathrm{L}^2_{ii}(\boldsymbol{x}_2, \cdot)) &= C(\boldsymbol{x}_1, \boldsymbol{x}_2), \quad i = j. \end{aligned} \tag{3.29}$$

Among various covariance models, the squared-exponential covariance function (see section 3.4.1) is used for the non-diagonal and the square of diagonal terms. The RRSTF is formed using three diagonal and three non-diagonal independent random fields. Similar to (3.26), the non-diagonal fields are given by,

$$\mathrm{L}_{ij}(\boldsymbol{x}, \omega) = \frac{\delta(\boldsymbol{x})}{\sqrt{d+1}} Z_{ij}(\boldsymbol{x}, \omega), \quad \text{for} \quad i < j, \quad Z_{ij}(\boldsymbol{x}, \omega) \sim \mathcal{N}(0, C). \tag{3.30}$$

While similar to (3.27), the diagonal fields are given by,

$$\mathrm{L}_{ii}(\boldsymbol{x}, \omega) = \frac{\delta(\boldsymbol{x})}{\sqrt{d+1}} \sqrt{2 g_i(\boldsymbol{x}, \omega)}, \quad \text{for} \quad i = 1, 2, 3, \tag{3.31}$$

where $g_i(\boldsymbol{x}, \omega)$ is a gamma random field, $g_i > 0$, with probability density function as,

$$p(g_i(\boldsymbol{x}, \cdot)) = \frac{g_i(\boldsymbol{x}, \cdot)^{k_i(\boldsymbol{x})-1} \exp(-g_i(\boldsymbol{x}, \cdot))}{\Gamma(k_i(\boldsymbol{x}))}, \tag{3.32}$$

where $\Gamma(k_i)$ is a gamma distribution and $k_i(\boldsymbol{x}) = (d+1)/2\delta(\boldsymbol{x})^2 + (1-i)/2$, is its shape parameter where the scaling parameter is set again to unity. The value of the dispersion parameter varies spatially, and in order to assure a positive definite Reynolds stress at each point it should be chosen in the range $0 < \delta(x) < \sqrt{(d+1)/(d+5)}$.

The KL decomposition of a non-Gaussian field is not straightforward and therefore the author in [134] suggests to use generalized polynomial chaos expansion to approximate the gamma random variables g_i at any location $\boldsymbol{x}$ as,

$$g(\boldsymbol{x}) \approx \sum_{\alpha=0}^{N_{PCE}} h_\alpha \mathscr{H}_\alpha(Z), \tag{3.33}$$

where $g(\boldsymbol{x})$ is a gamma random field, N_{PCE} is the PCE order and $\mathscr{H}_\alpha(Z)$ are the Hermite polynomials in Z with polynomial order α and coefficients h_α. These coefficients can be determined using the orthogonality condition with respect to the Gaussian measure [136]:

$$h_\alpha = \frac{\langle g \mathscr{H}_\alpha \rangle}{\langle \mathscr{H}^2_\alpha \rangle} = \frac{1}{\langle \mathscr{H}^2_\alpha \rangle} \int_{-\infty}^{\infty} F_g^{-1}[F_Z(z)] \mathscr{H}_\alpha(z) p_Z(z) dz, \tag{3.34}$$

where $F_Z(z)$ is the Cumulative Distribution Function (CDF) and $p_Z(z)$ is the PDF of the standard Gaussian random variable; F_g is the CDF of the random variable g and F_g^{-1} its inverse. The above integral is computed numerically at each spatial location $\boldsymbol{x}$ with $Z(\boldsymbol{x})$ as a standard Gaussian variable.

We now employ a non-intrusive polynomial chaos method to construct a PCE for the random matrix $\mathbf{G}(\boldsymbol{x},\omega)$, and with that for the Reynolds stress tensor $\mathbf{R}$ (at each spatial location),

$$\mathbf{G} = \mathbf{G}(\boldsymbol{x},\omega) = \sum_{i=0}^{P} \mathbf{G}_i(\boldsymbol{x})\psi_i(\boldsymbol{\xi}), \tag{3.35}$$

where $\psi_i(\boldsymbol{\xi})$ is the multidimensional (N_{KL} dimensions) Hermite polynomials in the Gaussian set $\{\xi_i\}$ obtained from the KL decomposition of the random field $Z(0,C)$. The value of P depends on number of dimensions $d = N_{KL}$ and the chosen polynomial order n as discussed in section 3.2. The expansion coefficients are given by,

$$\mathbf{G}_i = \frac{\langle \mathbf{G}\psi_i \rangle}{\langle \psi_i^2 \rangle}. \tag{3.36}$$

The numerator $\langle \mathbf{G}\psi_i \rangle$ is approximated using a quadrature rule which requires computing (sampling) $\mathbf{G}$ at different quadrature points. The polynomial chaos expansion of the RRSTF can then be simply obtained as,

$$\mathbf{R} = \mathbf{L}_R^T \mathbf{G}(\boldsymbol{x},\omega)\mathbf{L}_R = \sum_{i=0}^{P} \mathbf{L}_R^T \mathbf{G}_i(\boldsymbol{x})\mathbf{L}_R \psi_i(\boldsymbol{\xi}) = \sum_{i=0}^{P} \mathbf{R}_i(\boldsymbol{x})\psi_i(\boldsymbol{\xi}), \tag{3.37}$$

where $\mathbf{R}_i$ is the coefficient of the random Reynolds stress tensor field, $\mathbf{R}_i(\boldsymbol{x}) = \mathbf{L}_R^T \mathbf{G}_i(\boldsymbol{x})\mathbf{L}_R$, and $\mathbf{L}_R$ is a upper triangular matrix resulting from the decomposition (Cholesky factorization) of the mean Reynolds stress, $\mathbf{R}^{(det)} = \mathbf{L}_R^T \mathbf{L}_R$.

Using the RRSTF approach, the stochastic form of the RANS equations becomes,

$$\begin{aligned} \nabla \cdot \overline{\boldsymbol{u}} &= 0, \\ \frac{\partial \overline{\boldsymbol{u}}}{\partial t} + (\overline{\boldsymbol{u}} \cdot \nabla)\overline{\boldsymbol{u}} &= -\nabla \overline{p} + \nabla \cdot (\nu \nabla \overline{\boldsymbol{u}}) - \nabla \cdot \mathbf{R}(\boldsymbol{x},\omega). \end{aligned} \tag{3.38}$$

A set of equations governing the evolution of the modes of the velocity can be derived in a way similar to that for the REVF method: for $k = 0, 1, ..., P$,

$$\begin{aligned} \nabla \cdot \overline{\boldsymbol{u}}_k &= 0, \\ \frac{\partial \overline{\boldsymbol{u}}_k}{\partial t} + \sum_{i=0}^{P}\sum_{j=0}^{P} (\overline{\boldsymbol{u}}_i \cdot \nabla)\overline{\boldsymbol{u}}_j M_{ijk} &= -\nabla \overline{p}_k + \sum_{i=0}^{P}\sum_{j=0}^{P} \nabla \cdot (\nu_i \nabla \overline{\boldsymbol{u}}_j) M_{ijk} - \nabla \cdot \mathbf{R}_k(\boldsymbol{x}). \end{aligned} \tag{3.39}$$

Details of the derivation can be found in [85]. The solution procedure for this large system of equations is deferred to section 3.5.

This stochastic model directly uses the Reynolds stress obtained from the deterministic solve and thus abstains from using the Boussinesq approximation. Since all the six components of the Reynolds stress tensor field can have a different covariance, this model can also capture highly anisotropic flows.

3.5. IMPLEMENTATION

OpenFOAM uses the finite volume method (FVM) to discretize partial differential equations [28]. Among the various fluid dynamic solvers offered by OpenFOAM, we choose the solver

called *pimpleFoam* [25], which allows the use of large time-steps to solve the incompressible Navier-Stokes equations. This solver is based on the PIMPLE algorithm for pressure-velocity coupling using a Rhie and Chow type interpolation [49]. The code for the stochastic RANS solver is organized like any other solver in OpenFOAM and is publicly available at https://github.com/parallelwindfarms/UQFoam.

It may be stressed that the stochastic system of governing equations (3.24, 3.39) for the evolution of the velocity and pressure modes $\boldsymbol{u}_k, p_k$ for $k = 0, 1, ..., P$, has a structure similar to the original deterministic RANS equations (3.14). Due to the coupling via the convection and the diffusion terms, the size of this new system is $P+1$ times its deterministic version. It can be observed that the divergence-free velocity constraints are decoupled and can be solved independently. Based on this observation, a fractional step projection scheme has been previously implemented [67]. In the first fractional step, the convection and diffusion terms are integrated followed by enforcing the divergence-free constraints in the second fractional step.

Our approach of the stochastic solver is based on the development of the existing deterministic solver (*pimpleFoam*) such that it can accommodate and solve $(P+1)$ coupled Navier-Stokes like systems in $\boldsymbol{u}_k, p_k$ for $k = 0, 1, ..., P$. Figure 3.1 (top) provides a flow chart for the steps involved in quantifying and propagating model output-form uncertainty in RANS simulations. These steps can be realized as follows:

1. Obtain the eddy-viscosity or Reynolds stress tensor field from the deterministic RANS solver.

2. Introduce randomness in the eddy-viscosity or Reynolds stress tensor using a Gaussian random field (equation (3.17)) and express it in terms of a polynomial chaos expansion according to equation (3.20) or (3.37), respectively.

3. Propagate the random eddy-viscosity field (REVF) or random Reynolds stress tensor field (RRSTF) by numerically solving the stochastic RANS equations (3.24) or (3.39), respectively. In the stochastic RANS solver, the steps involved between two consecutive time steps are as follows:

   (a) For the first explicit iteration, start with sequentially updating the pressure and velocity modes, while using the updated values of the previous mode (hence the name *explicit* iteration).

   (b) Repeat the above step for the next explicit iteration using the updated modes in the previous explicit iteration.

   (c) Increase the number of explicit iterations (step (b)) until convergence.

4. Using the solution of stochastic RANS from step 3, obtain a PCE for velocity, pressure and other quantities of interest.

5. Compute the statistical moments of the quantities obtained in step 4 according to equation (3.5).

In Figure 3.1 (bottom), we highlight the most important steps needed to develop the stochastic RANS solver using *pimpleFoam*. In contrast to the deterministic solver, two nested loops are introduced. The first loop is over the explicit iterations (I_e), which updates

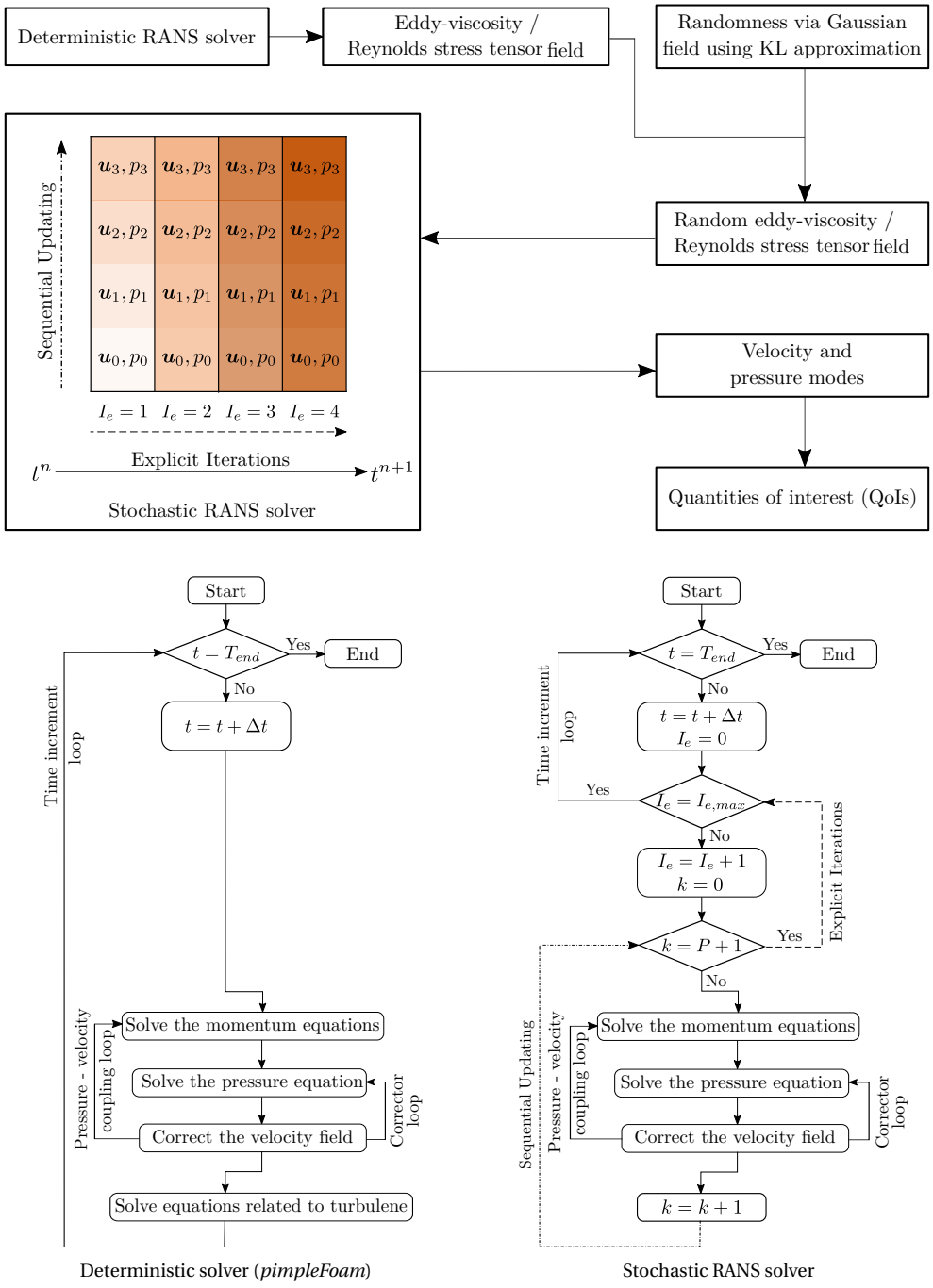


Figure 3.1 | (top) Flow chart for the steps involved in quantifying and propagating model output-form uncertainty in RANS simulations ($P = 3$), and (bottom) Graphical representation of the algorithm implemented in the deterministic RANS solver (left) and stochastic RANS solver (right).

the mode strengths between the two consecutive time-steps. For every explicit iteration, the second nested loop solves the Navier-Stokes like system $(\boldsymbol{u}_k, p_k)$ for each mode strength k, while employing the existing, however modified, structure of the PIMPLE scheme. The modifications are inevitable due to the summations in the convection and the diffusion terms of the stochastic equations. We solve each of these systems sequentially, by using the explicitly initialized or updated velocity and pressure modes from the previous time step, and repeat until convergence. Depending on the type of flow, the value of P and a few other parameters; it usually takes about $3-6$ explicit iterations (I_e) to converge at every time-step. Note that the stochastic RANS solver is developed using a transient deterministic solver and thus it can also be used as a stochastic URANS solver.

In addition to the exiting modules, we require to either modify or create some completely new routines for pre- and post- processing. Restricting the verbosity, we attempt to provide an overview of the major implementation steps:

- New variables (vectors) are defined for the list of mode strengths of all the uncertain parameters, flow variables and derived variables for post-processing,
- A small routine is added to read these inner products (obtained using *chaospy* library [26]). Another similar lines of code are added to read in the values of d, n, I_e, etc, and
- A separate routine is added to the post-processing step to compute the moments of flow variables and other derived QoIs.

The order of accuracy and the convergence rate depend on the wide choice of space and time discretization schemes and iterative solvers offered by the OpenFOAM library. The discretization schemes and iterative solvers used to solve the stochastic RANS are the same as those used to solve the deterministic RANS.

3.6. NUMERICAL RESULTS

In order to study the performance of the stochastic RANS solver, we use two benchmark cases - (a) flow over periodic hill, and (b) fully developed turbulent flow in a square duct.

3.6.1. FLOW OVER PERIODIC HILLS

For the flow over periodic hills, a Reynolds number $Re = 2800$ along with the reference DNS data [10] is used. The recirculation zones, anisotropy, strong curvature in the mean flow and unbounded shear layer are some typical features of this problem that makes it difficult for the RANS models to make accurate flow predictions. For this case we only employ the REVF approach. In the Figure 3.2, the domain shape for the flow over periodic hills is represented. The coordinates in the x, y and z directions are aligned with streamwise, normal to wall and span-wise directions, respectively. The Reynolds number is based on the crest height H and the bulk velocity U_b, such that $Re = 2800$. For both the deterministic and stochastic solve, for all modes of velocity and pressure, we use periodic boundary conditions in streamwise direction, and no-slip boundary condition at the walls.

The mesh and REVF parameters are presented in Table 3.1. For the deterministic solve, we use the Launder-Sharma $k-\epsilon$ RANS turbulence model with finer mesh (grading) near

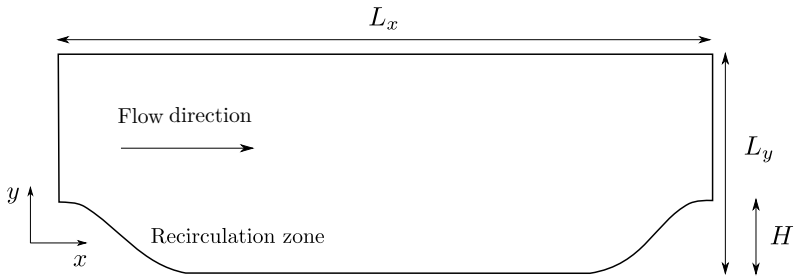


Figure 3.2 | Computational domain for turbulent flow over periodic hills.

the walls, such that $y^+ < 1$. This avoids the use of wall-functions which are one of the potential sources of uncertainty in RANS simulations. The stochastic simulation uses the same mesh as the deterministic simulation. The values of the correlation lengths l_x, l_y roughly represents the assumed (or observed) length of flow structures in x and y directions, respectively. A reasonable marginal variance (less than one) is chosen in order to avoid unrealistic realizations of eddy-viscosity. The number of terms retained in the KL decomposition N_{KL} is determined using a threshold s, such that,

$$N_{KL} = min\left\{k \in \mathbb{N} | \sum_{i=1}^{k} \lambda_i \geq (1-s) \times \sum_{i=1}^{+\infty} \lambda_i \right\}, \tag{3.40}$$

i.e. the retained modes capture a prescribed minimal fraction of the L_2-norm of the eddy viscosity field.

We use OpenTURNS [89] for numerically computing the eigenvalues. Note that the KL expansion is carried out on the same mesh used for the RANS solver. With a value of $1e-2$ for the threshold, we retain the first 18 modes of KL decomposition of the Gaussian field. We employ third order multi-dimensional Hermite polynomials for the polynomial chaos expansion, resulting in a system of $P+1 = (18+3)!/(18!3!) = 1330$ equations coupled in velocity modes. In order to reduce number of modes (i.e., to reduce the computation time) we use the hyperbolic truncation with the tuning parameter $q = 0.5$, which results in a system of 55 equations. A detailed analysis on the use of this truncation technique is provided in section 3.6.1.

Figure 3.3 shows the first six KL modes. The symmetry observed in the KL modes is due to the use of a uniform variance throughout the domain. However, one may use higher variance in the recirculation region as compared to other regions of the flow. The KL modes illustrate the uncertainty space of the REVF. Large characteristic scales are represented by lower modes, while the higher modes represent oscillations.

The mean along with the confidence intervals of $\pm$ 2 standard deviations for different quantities of interest is presented in Figure 3.4(a-c). Upon introducing randomness into the eddy viscosity field, the mean and the deterministic profiles coincide as expected. The mean velocities obtained from the stochastic solver are very close to those from the deterministic RANS solution. This is indeed anticipated since the REVF uses the modeled eddy viscosity field from the deterministic RANS simulation. However, a deviation from the

Parameter	Value
domain size ($L_x \times L_y$)	$9H \times 3.036H$
mesh ($n_x \times n_y$)	100×80
wall normal distance in y^+	≤ 0.6
correlation lengths (l_x, l_y)	$1.5H, 0.5H$
marginal variance (σ_c^2)	0.5
KL decomposition threshold (s)	$1e-2$
number of terms in KL decomposition (N_{KL})	18
degree of polynomial chaos expansion (n)	3
hyperbolic truncation set tuning parameter (q)	0.5

Table 3.1 | Mesh and computational parameters to generate a random eddy viscosity field for the flow over periodic hills.

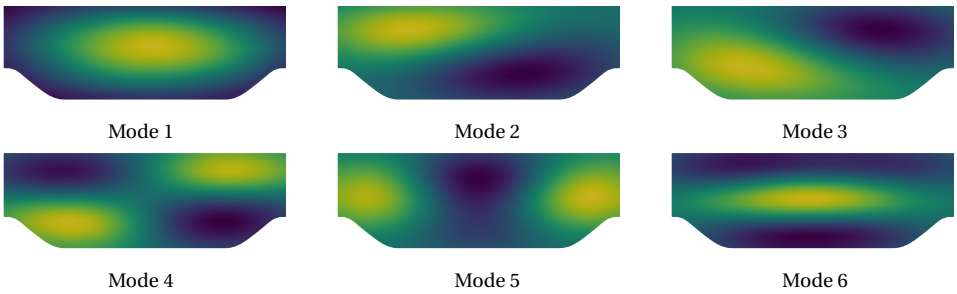


Figure 3.3 | First six KL expansion modes of the flow over periodic hill problem. The modes are shifted and normalized to the range of 0 (darkest) and 1 (lightest).

deterministic profile is expected for a higher level of uncertainty in the REVF. The velocities have high variance in the recirculation region starting from the flow separation until the reattachment point. As evident, both the deterministic and the stochastic mean velocities deviate significantly from the DNS solution, which can be associated to the already discussed limitation of the RANS turbulence models. Similar to the velocities, the deterministic and stochastic mean of the wall shear-stress are very close, with a large deviation from DNS. Note that the largest variance in the stochastic solution is around the predicted reattachment point ($x/H \approx 3.5$). It is worth mentioning that for both, the velocity and the wall shear-stress, the DNS data mostly lies within the uncertain bounds, which is a good indicative of improvement in the RANS modeling to further reduce the uncertainty. A RANS model with (i) mean velocity gradients closer to that of DNS profile and, (ii) DNS results lying well within its uncertain bounds – has a scope of improvement towards a better model upon further investigation towards mitigation of structural or parametric uncertainties.

The Karhunen-Loève expansion was used to discretize the Gaussian random field into a set of random variables $\boldsymbol{\xi}$. The relative importance and effect of these random variables on the QoIs can be studied using global sensitivity analysis via Sobol indices as discussed in section 3.2. Figure 3.5 presents the Sobol indices (S_{T_i}, S_i) of the random variables for the streamwise velocity and the wall shear-stress at different locations in the recirculation region, obtained from KL decomposition of the Gaussian field. The relative importance and

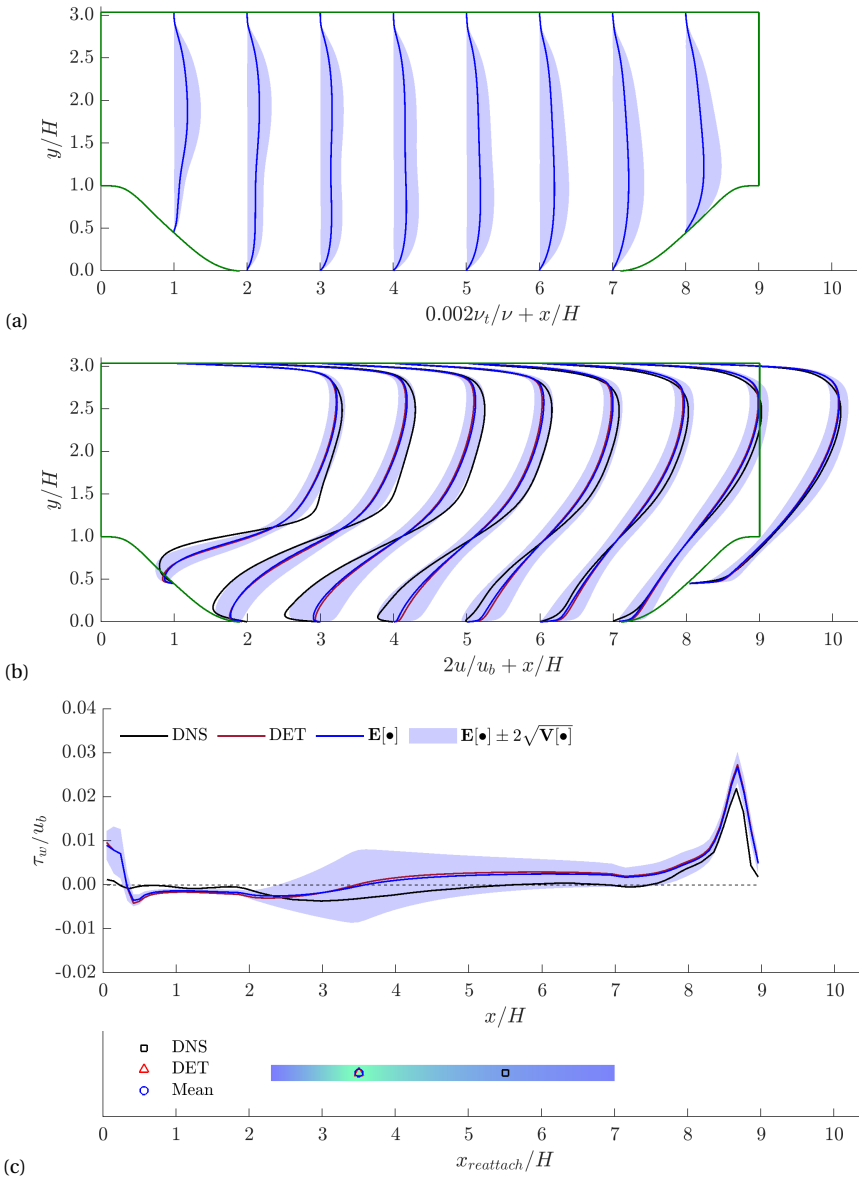


Figure 3.4 | Mean and variance of (a) turbulent viscosity, (b) streamwise velocity and (c) (top) wall shear-stress at different locations in $x-$direction for the flow over periodic hills. Compared with deterministic (DET) and DNS results. Legend in (c) (top) applies to (a) and (b) as well. Figu re (c) (bottom) shows reattachment location and bounds colored with variance.

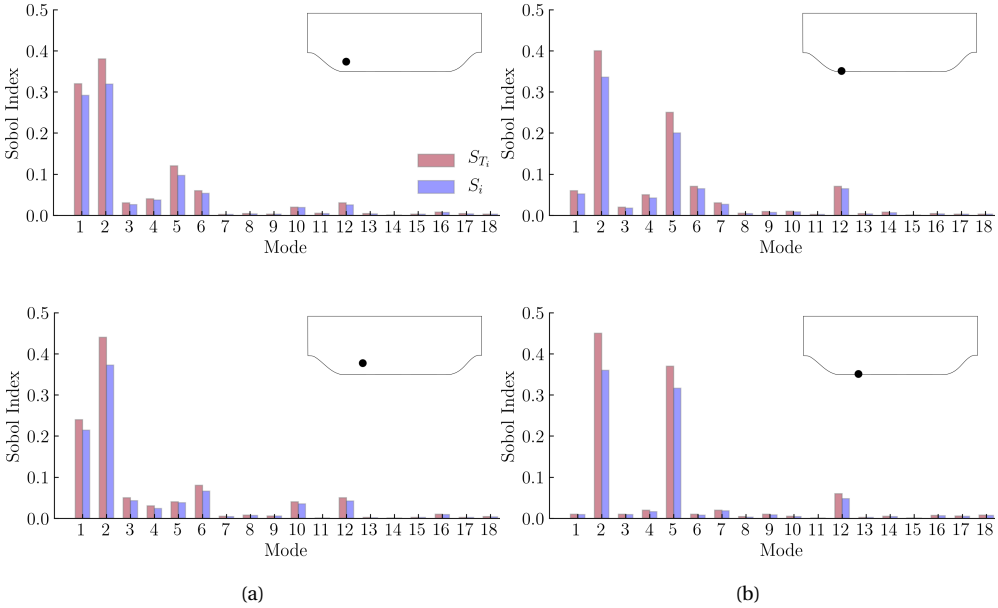


Figure 3.5 | Global sensitivity analysis using Sobol indices (S_{T_i}, S_i) for (a) streamwise velocity and (b) wall shear-stress at different locations in the recirculation region for the flow over periodic hills problem.

effect of these random variables on the QoIs can be seen in the figure. These modes are in the order of decreasing eigenvalues. The streamwise velocity shows significant dependence on the first two modes. For modes 3-6, the dependence is relatively smaller. The streamwise velocity is almost independent of the higher modes (< 1%) except mode 10 and 12. The wall shear-stress on the other hand predominantly depends on modes 2 and 5, with rather smaller dependence on remaining initial modes and almost independent of the higher modes (< 1%) except mode 12. Note that the difference in S_{T_i} and S_i is significant only for the first few modes, implying that the contribution from the interaction of these modes is indeed important and should be taken into account when using different methods to avoid the *curse of dimensionality* (discussed later in this section).

Before advancing further, the following remarks are in order:

1. The mesh sensitivity analysis shows that the stochasticity (mean and variance of QoIs) depends only marginally on the mesh size as compared to its dependence on the uncertainty in the turbulence model.

2. The polynomial degree is varied in the range $n \in \{1,2,3,4\}$. For polynomial degree of 1, there is no interaction terms between the polynomials, resulting in a lower variance in QoIs. For $n \geq 2$, the interaction terms populate drastically and thus a high polynomial degree implies high cardinality. Since the results for the stochastic simulation beyond $n = 3$ does not show any significant difference, a polynomial degree of three is chosen for the rest of the study.

3. The dependence of stochasticity of the solution on the marginal variance (σ_c^2) of the

Parameter	Case 1	Case 2	Case 3
correlation length along x-direction (l_x)	$1.00H$	$1.50H$	$2.00H$
correlation length along y-direction (l_y)	$0.33H$	$0.50H$	$0.67H$
marginal variance (σ_c^2)	0.5	0.5	0.5
number of terms in KL decomposition (N_{KL})	26	18	13

Table 3.2 | Random eddy viscosity field parameters for three different cases with varying correlation lengths for the flow over periodic hills problem.

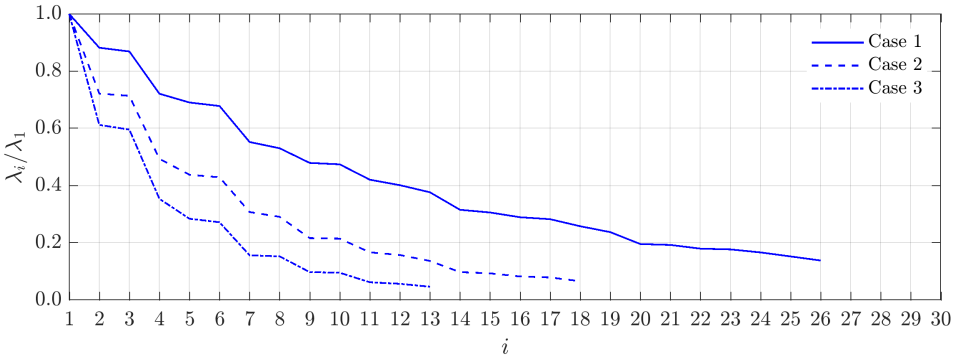


Figure 3.6 | Normalized eigenvalues of the covariance kernel for three different values of correlation length scales for the flow over periodic hills problem, as listed in Table 3.2.

random eddy viscosity field is examined for different values in range $[0.05, 1.0]$. The uncertainty in QoIs grows almost linearly with σ_c^2. However, a large value of marginal variance may lead to unrealistic samples of REVF resulting in non-physical solutions. In the present work we choose a value of $\sigma_c^2 = 0.5$ for the rest of the study.

VARYING CORRELATION LENGTH SCALES

We consider three different sets of correlation lengths in x and y -direction as described in Table 3.2. The remaining parameters are the same as in Table 3.1. The effect of varying the correlation length scales on the number of terms in KL approximation is depicted in Figure 3.6. It can be noted that, with increasing correlation length, the number of terms in the KL approximation reduces. This is due to the fact that as the correlation length increases, lesser number of eigenfunctions are required to represent the Gaussian field.

In Figure 3.7, we present a comparison of the solution from all the three cases. The mean may take near zero values and thus coefficient of variation (CoV) may be misleading. Moreover, the mean profiles for all the cases are extremely close to each other, therefore we directly compare the standard deviations at different locations in the domain. It can be realized that the standard deviation of the REVF is similar for all the three cases, while the standard deviation in velocity and wall shear-stress show significant variation. Larger correlation lengths results in a higher variance, especially near the reattachment location.

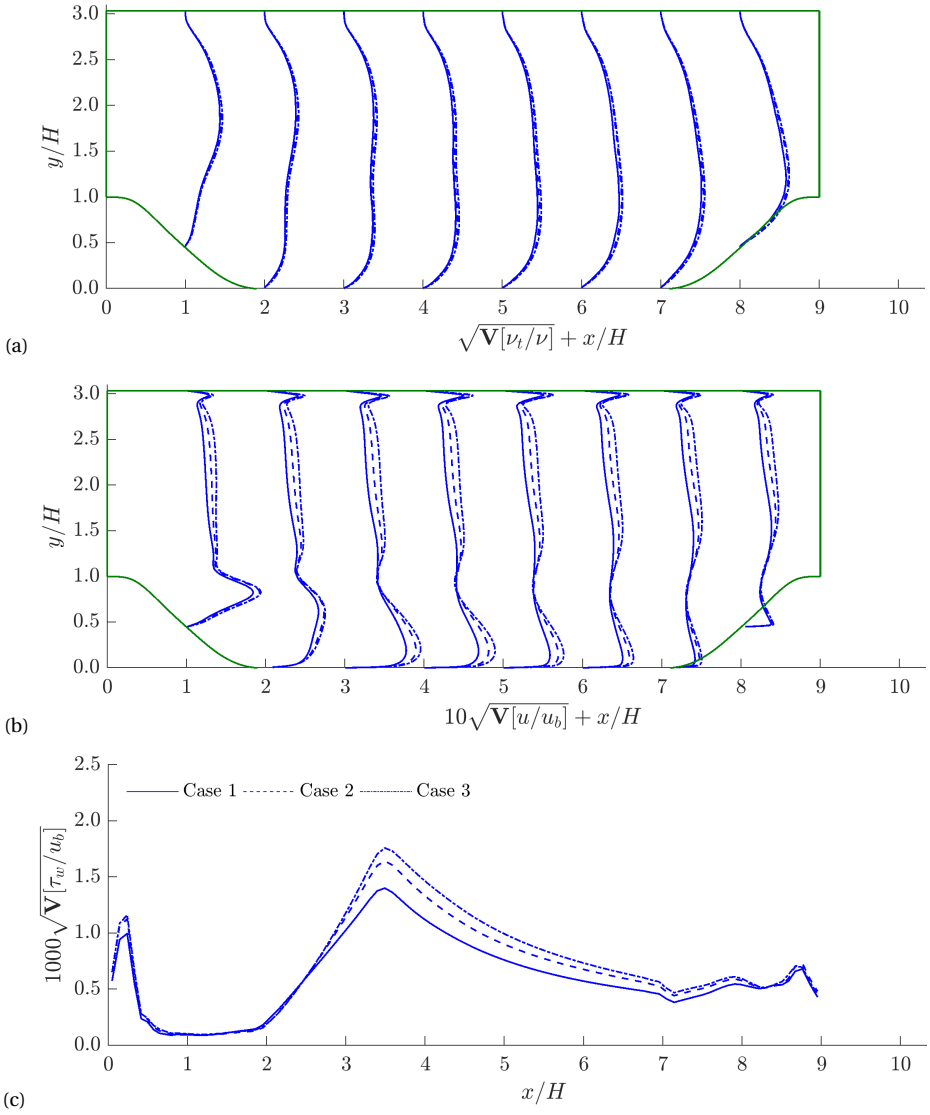


Figure 3.7 | Comparison of the standard deviations in (a) turbulent viscosity, (b) streamwise velocity, and (c) wall shear-stress with varying correlation lengths in x and y -directions as listed in Table 3.2. Legend in (c) applies to (a) and (b) as well.

TECHNIQUES TO OVERCOME THE CURSE OF DIMENSIONALITY

There are several schemes for the truncation of a polynomial chaos expansion. We already presented one of the most widely used method, which consists of upper-bounding the total degree of polynomials by simply choosing a maximal value n. The set of multi-indices is denoted by

$$\mathscr{S}^{d,n} = \{\boldsymbol{\alpha} \in \mathbb{N}^d : |\boldsymbol{\alpha}| \leq n\}, \tag{3.41}$$

where $\boldsymbol{\alpha} = \{\alpha_1, ..., \alpha_d\}$ is the multi-index. The cardinality of the truncated set $\mathscr{S}$ is then given by

$$|\mathscr{S}^{d,n}| = \frac{(d+n)!}{d!n!}. \tag{3.42}$$

Evidently, the above cardinality grows dramatically in both d and n. Thus, for high dimensional problems (large d, say $d > 10$) and/or highly non-linear problems (which require large n), such truncation method is not plausible. This issue is known as the *curse of dimensionality*.

The authors in [102] realized that a lot of systems only exhibit interactions of low degree polynomials and therefore the higher degree polynomial interactions can be ignored. Based on this observation they proposed a new truncation scheme called *hyperbolic truncation set*

$$\mathscr{S}_q^{d,n} = \{\boldsymbol{\alpha} \in \mathbb{N}^d : ||\boldsymbol{\alpha}||_q \leq n\}. \tag{3.43}$$

The q-norm is defined as

$$||\boldsymbol{\alpha}||_q = \left(\sum_{i=1}^{d} \alpha_i^q\right)^{1/q}, \tag{3.44}$$

where $q \in (0, 1\}$ is a tuning parameter. For $q = 1$, the hyperbolic index set is identical to the total degree index set. Decreasing q results in smaller set of polynomials and as $q \to 0$, the set is left with only the uni-variate polynomials.

Another approach to reduce the overall cardinality is to reduce the number of random dimensions (d) by increasing the KL decomposition threshold (s) defined in equation (3.40). In the present study, we use a combination of these approaches to study both low dimensional and low degree polynomial representations of our stochastic problem.

Table 3.3 lists all combinations considered for the KL decomposition threshold (s) and the hyperbolic truncation set tuning parameter (q). The remaining parameters are the same as in Table 3.1. As shown in Figure 3.8, for a fixed tuning parameter, the number of terms in the KL decomposition increases significantly as we decrease the threshold. While, as depicted in Figure 3.9, for a fixed threshold, the cardinality of the truncated set decreases drastically as we decrease the tuning parameter.

For all the cases in Table 3.3, a comparison is made in Figure 3.10. Obviously, the standard deviation of the REVF is close to each other for all the cases. The variance in velocity and wall shear-stress, is higher in the recirculation area and near the reattachment location. It can be remarked that increasing the KL decomposition threshold s, gradually reduces the cardinality of the truncated set, while under-predicting the variance. Whereas, reducing the tuning parameter q, drastically reduces the cardinality of the truncated set ($P + 1$), while practically predicting the same level of variance. Therefore, decreasing the tuning parameter, thereby ignoring the higher degree polynomial interaction results in an accurate prediction of the stochasticity at a significantly lower computational cost.

Parameter	C1	C2	C3	C4	C5	C6	C7
KL decomposition threshold ($s \times 100$)	1	2	5	1	2	5	5
number of KL modes (N_{KL})	18	13	6	18	13	6	7
hyperbolic truncation parameter (q)	0.50	0.50	0.50	0.75	0.75	0.75	1.00
cardinality of the truncated set ($P+1$)	55	40	19	208	118	34	120

Table 3.3 | Random eddy viscosity field parameters for different cases (C#) with varying KL decomposition threshold (s) and hyperbolic truncation set tuning parameter (q) for the flow over periodic hills problem.

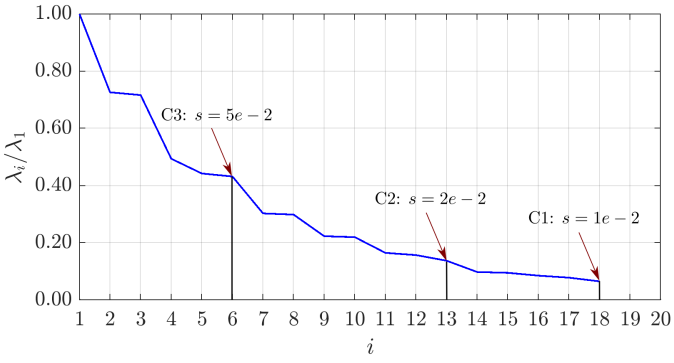


Figure 3.8 | Normalized eigenvalues of the covariance kernel for three different values of KL decomposition threshold while keeping the hyperbolic truncation parameter fixed, cases C1, C2 and C3 in Table 3.3.

COMPARISON BETWEEN TURBULENCE MODELS

In this section we compare stochasticity arising from three different turbulence models mentioned in section 3.3. Again, the uncertainty is introduced in the (deter
stic) eddy viscosity field and propagated through the stochastic solver to estimate the mean and variance in the velocity and wall shear-stress. The computational mesh and the physical parameters are the same for all three cases, as listed in Table 3.1.

Figure 3.11 (a) shows the mean and variance of the REVF for the three turbulence models. As evident, the mean is close to each other except in the bulk region, where the $k-\omega$ SST model estimates a lower viscosity. The amount of uncertainty introduced is the same for all three models, as shown in figure 3.11 (a). Figure 3.11 (b) and (c) compares the mean and the variance in the streamwise velocity and the wall shear-stress, for all three models, against the DNS results. For all cases, the maximum deviation of mean from DNS and the maximum uncertainty is found in the recirculation zone as well as near the reattachment location. Overall, the $k-\omega$ SST model predictions are more accurate as compared to the $k-\omega$ and the Launder-Sharma $k-\epsilon$ models. Note that, on average, the level of stochasticity is highest for the $k-\omega$ model and lowest for the $k-\omega$ SST model. A lower variance in the k-omega SST model predictions can be attributed to a lower sensitivity of its model parameters and/or a lower structural uncertainty in its formulation.

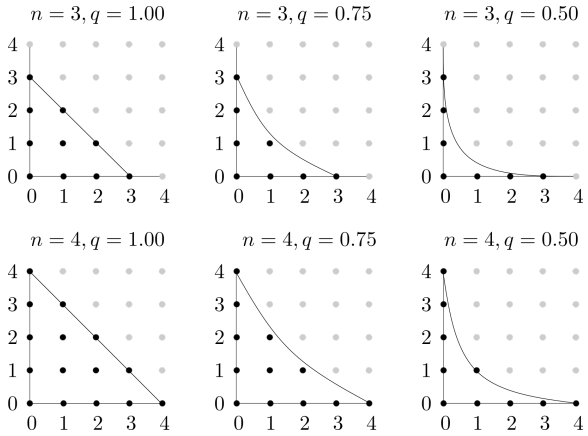


Figure 3.9 | Two-dimensional ($d = 2$) hyperbolic index set representation for varying degree (n) and tuning parameter (q). The dots in black indicate the truncated set.

REDUCING UNCERTAINTY VIA DATA ASSIMILATION

After introducing randomness into the system via the REVF (section 3.4.1), we now use an iterative data assimilation technique to update this uncertainty in the system. This will allow the inverse modeling of the REVF and improve the accuracy of the predictions made by the stochastic RANS solver. For this purpose, we consider the use of an Iterative Ensemble Kalman (IEnK) method [48], which is a Bayesian inference method to infer a posterior based on the prior and the observations (data).

The iterative ensemble Kalman method converts the steady state inverse problem into a dynamic data assimilation problem, updating the state $\mathbf{x}$ at every iteration until convergence [143]. Thus, at each iteration (or pseudo time-step), the (perturbed) observations are used to perform a single update of the state. In this work, the state of the system includes both the REVF and the velocity field as an "augmented-state", i.e., $\mathbf{x} \equiv [\nu_t, \overline{\boldsymbol{u}}]$. The samples from the prior augmented-state are updated after incorporating the uncorrelated observations for velocity ($\mathbf{y}$) at specific locations in the domain. At each iteration, the observations are chosen to be independent realizations from a Gaussian distribution with DNS data as the mean and standard deviation, $\sigma_{obs} = 0.1\times$ mean. Closer observation locations are considered in the regions of steep velocity gradients (recirculation zone and reattachment region), while lesser points are considered in the free shear region. The algorithm of the overall iterative procedure is as follows.

1. Characterize the uncertainty in eddy viscosity field using a log-normal Gaussian field and express it in terms of a polynomial chaos expansion $\nu_t(\boldsymbol{x},\omega) \approx \sum_{i=0}^{P} \nu_{t_i}(\boldsymbol{x})\psi_i(\boldsymbol{\xi})$.
2. Propagate the (initial or updated) REVF.
   (a) Using the stochastic RANS solver, obtain a PCE for velocity field,

$$\overline{\boldsymbol{u}}(\boldsymbol{x},\omega) \approx \sum_{i=0}^{P} \overline{\boldsymbol{u}}_i(\boldsymbol{x})\psi_i(\boldsymbol{\xi}). \tag{3.45}$$

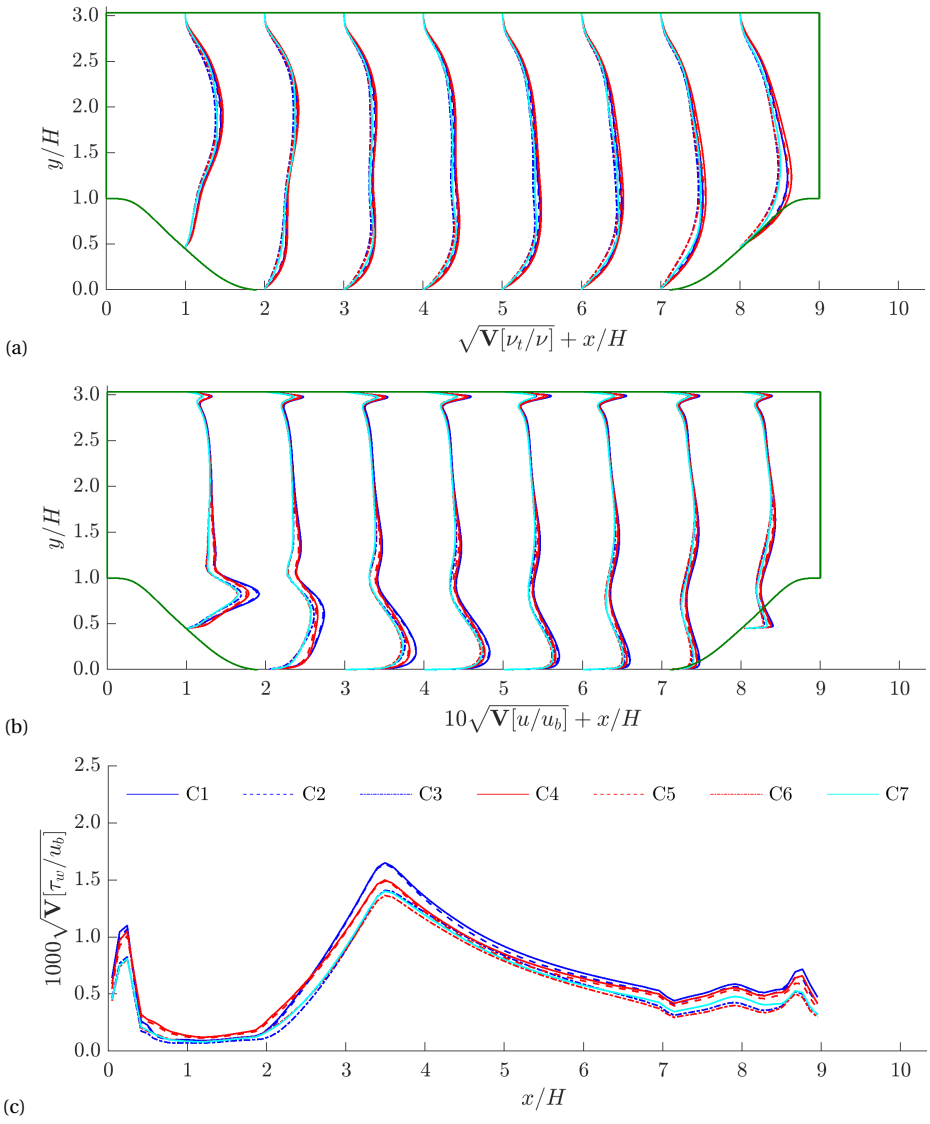


Figure 3.10 | Comparison of the standard deviations in (a) turbulent viscosity, (b) streamwise velocity, and (c) with varying varying KL decomposition threshold and hyperbolic truncation set tuning parameter as listed in Table 3.3. Legend in (c) applies to (a) and (b) as well.

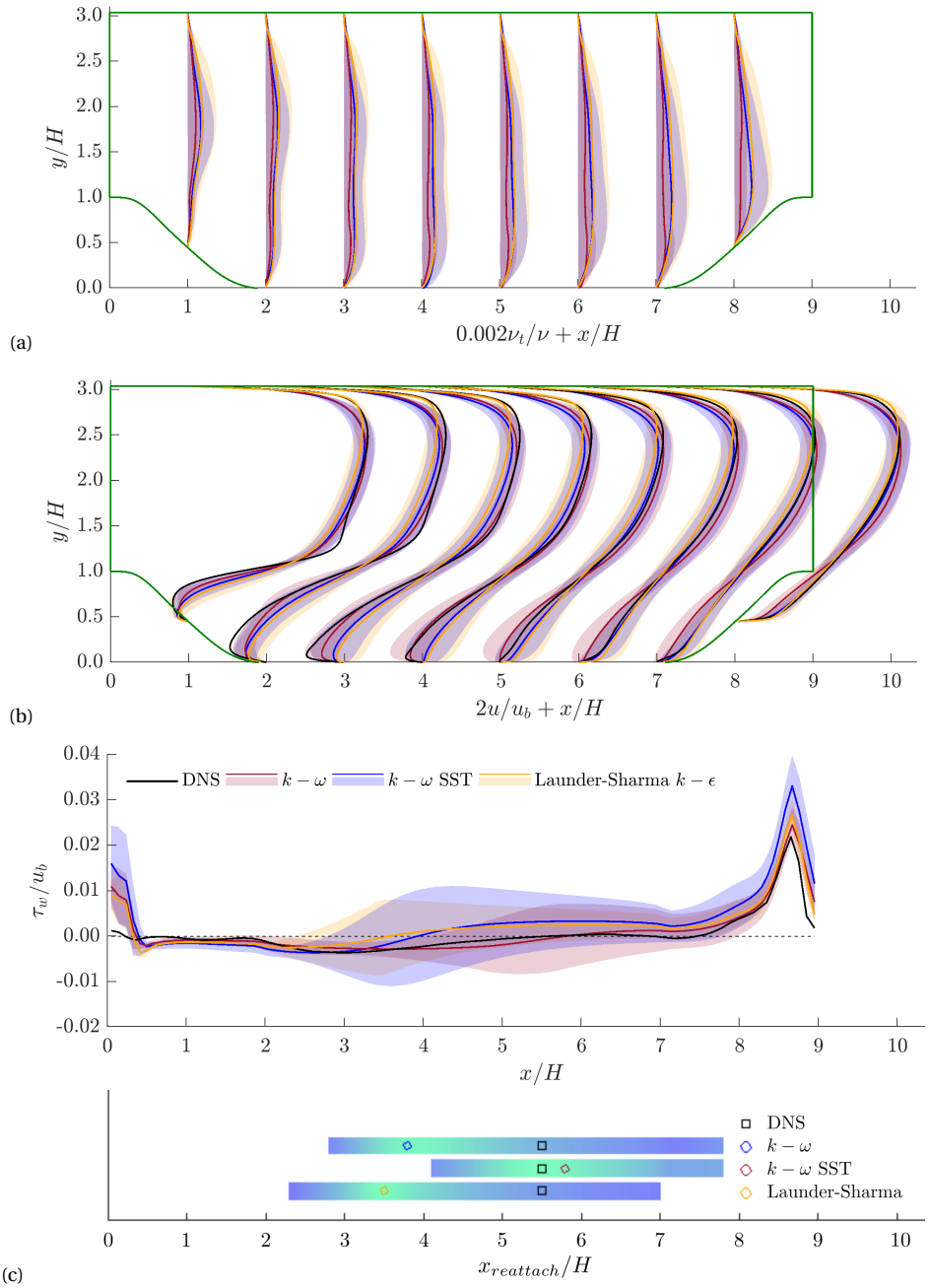


Figure 3.11 | Mean (solid line) and ±2× standard deviations (shaded region) of (a) turbulent viscosity, (b) streamwise velocity and (c) (top) wall shear-stress at different locations in $x-$direction for the flow over periodic hills using three different turbulent models. Compared with deterministic (DET) and DNS results. Legend in (c) (top) applies to (a) and (b) as well. Figure (c) (bottom) shows reattachment location and bounds colored with variance.

(b) Generate *prior* ensemble of the realizations of the state,

$$\{\mathbf{x}_{(j)}\}_{j=1}^{M} = [\{\nu_{t_{(j)}}\}_{j=1}^{M}, \{\overline{\boldsymbol{u}}_{(j)}\}_{j=1}^{M}], \tag{3.46}$$

by sampling the random variables, $\{\boldsymbol{\xi}_{(j)}\}_{j=1}^{M}$. Where,

$$\nu_{t_{(j)}}(\boldsymbol{x}) = \sum_{i=0}^{P} \nu_{t_i}(\boldsymbol{x})\psi_i(\boldsymbol{\xi}_{(j)}), \quad \overline{\boldsymbol{u}}_{(j)}(\boldsymbol{x}) = \sum_{i=0}^{P} \overline{\boldsymbol{u}}_i(\boldsymbol{x})\psi_i(\boldsymbol{\xi}_{(j)}) \quad j = 1, ..., M, \quad M \gg 1 \tag{3.47}$$

This step involves only polynomial evaluations and therefore large number of samples (M) can be generated as these computations doesn't require solving the deterministic governing equations.

3. Update the augmented-state ensemble $\{\mathbf{x}_{(j)}\}_{j=1}^{M}$, by employing the ensemble Kalman filtering method [48].

(a) Determine the mean and covariance of the ensemble:

$$\bar{\mathbf{x}} = \frac{1}{M}\sum_{j=1}^{M}\mathbf{x}_{(j)}, \quad \mathbf{P} = \frac{1}{M-1}\sum_{j=1}^{M}(\mathbf{x}_{(j)}\mathbf{x}_{(j)}^{T} - \bar{\mathbf{x}}\bar{\mathbf{x}}^{T}) \tag{3.48}$$

(b) Compute the Kalman gain matrix $\mathbf{K}$:

$$\mathbf{K} = \mathbf{P}\mathbf{H}^{T}(\mathbf{H}\mathbf{P}\mathbf{H}^{T} + \mathbf{D}) \tag{3.49}$$

where, $\mathbf{H}$ is the observation matrix and $\mathbf{D}$ is the covariance matrix of the error in the observations.

(c) Update each sample:

$$\mathbf{x}_{(j)} \rightarrow \mathbf{x}_{(j)} + \mathbf{K}(\mathbf{y}_{(j)} - \mathbf{H}\mathbf{x}_{(j)}) \tag{3.50}$$

where, $\mathbf{y}_{(j)}$ is an observation vector in the ensemble of observations $\{\mathbf{y}_{(j)}\}_{j=1}^{M}$, generated from the joint normal distribution $\mathcal{N}(\mathbf{y}, \mathbf{D})$.

4. Reconstruct the PCE modes of the REVF and the velocity field from the updated ensemble [69]:

$$\nu_{t_i}(\boldsymbol{x}, \omega) \approx \frac{1}{M}\sum_{j=1}^{M}\nu_{t_{(j)}}(\boldsymbol{x})\psi_i(\boldsymbol{\xi}_{(j)}), \quad \overline{\boldsymbol{u}}_i(\boldsymbol{x}, \omega) \approx \frac{1}{M}\sum_{j=1}^{M}\overline{\boldsymbol{u}}_{(j)}(\boldsymbol{x})\psi_i(\boldsymbol{\xi}_{(j)}) \tag{3.51}$$

5. Return to step 2 until the misfit between the velocity mean ($\bar{\boldsymbol{u}}_0$) from step 4 and step 2, is sufficiently low. This will result in an statistically converged ensemble.

The sampling error introduced due to the averaging procedure in (3.51) is small if a large number of samples are used in (3.47). As discussed earlier, the computing cost of these samples is minimal in comparison to solving the deterministic system.

The iterative ensemble Kalman method converges in about 12-17 iterations for the considered flow over periodic hills using the Launder-Sharma $k-\epsilon$ RANS model. However,

	NIPC	IPC	IPC (Table 3.3)	MC	MLMC
Cost / DET Sim	6.8×10^{10}	1330	19 – 208	3000*	200*

* A rough estimate from a recent work on a similar test case using MC and MLMC [60].

Table 3.4 | Computational cost of different UQ approaches w.r.t. a single deterministic simulation (DET Sim). For IPC and NIPC, $n = 3$ and $d = 18$.

this number may vary with the complexity of the flow, the quality of observations and the initialization (prior) of the state. The prior and the posterior obtained is presented in Figure 3.12. The posterior of the streamwise velocity and the wall shear-stress tends towards the DNS results. The uncertainty (variance) near the observation locations reduces significantly, while a higher level of uncertainty persists in the region where the observations are unavailable. Note that, in comparison to the prior, the reattachment location in the posterior is predicted closer to the DNS, with a lower uncertainty.

STABILITY, CONVERGENCE AND COMPUTATIONAL COST

The stability of the stochastic solver can be compared with that of the deterministic counterpart as the same discretization schemes were used for both. A central difference scheme in space and forward Euler in time are used in the present study.

With around three explicit iterations, the computational cost of the (decoupled) stochastic solver is about $3 \times (P + 1)$ times a deterministic solve. Table 3.4 shows a comparison between the computational cost of different UQ approaches. In comparison to the standard Monte Carlo (MC) method, the intrusive polynomial chaos is significantly cheaper. Furthermore, the cost of IPC with hyperbolic truncation is comparable (mostly lower) to multilevel Monte Carlo (MLMC) employed recently on a similar test case [60]. The NIPC approach for complete parametric space is not feasible. However, sparse-grid methods such as Latin hypercube sampling (LHS) may drastically reduce the number of samples needed for NIPC [44]. A comparison between IPC and NIPC for parametric uncertainties has been made in authors' previous work [85].

Since the first mode of the PCE is initialized with the deterministic solution, the convergence of the stochastic solver is relatively faster than the deterministic solver. When a Gaussian field with a higher marginal variance is used to represent the REVF, the stochastic solver shows either slow convergence or even divergence for a few cases. This can be attributed to non-physical realizations of the eddy viscosity field sampled from its high-variance PCE. Mostly, the higher modes fail to converge as fast as the lower modes.

Note that the comparison with the deterministic solve is made using the same number of processors for domain decomposition. A higher number of processors, may further decrease the computing time.

3.6.2. FLOW IN A SQUARE DUCT

For the flow in a square duct, a Reynolds number $Re = 1100$ along with the reference DNS data [88] is used. The most important feature of this flow is the secondary flow pattern due to the imbalance of normal stresses, which the isotropic eddy viscosity models fail to capture. For this reason we use the RRSTF approach. Figure 3.13 depicts a fully developed turbulent flow in a square duct. The computational domain is a two-dimensional top-right quadrant

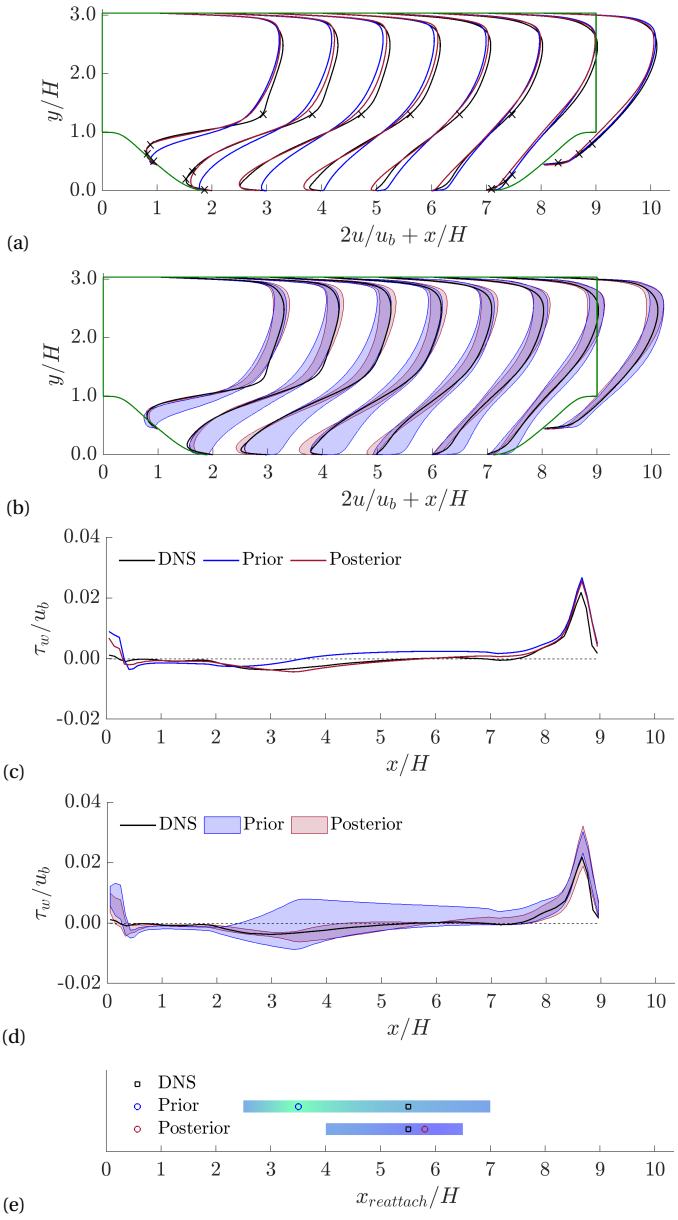


Figure 3.12 | Prior and posterior for mean and variance of (a, b) streamwise velocity and (c, d) wall shear-stress obtained from iterative ensemble Kalman method (iteration 15) using observations (×) at different locations in the domain for the flow over periodic hills. Legend in (c) and (d) applies to (a) and (b) as well, respectively. Figure (e) shows reattachment location and bounds colored with variance.

Parameter	Value
domain size ($L_x \times L_y$)	$h \times h$
mesh ($n_x \times n_y$)	64×64
wall normal distance in y^+	≤ 0.3
correlation lengths (l_x, l_y)	$2h, 2h$
marginal variance (σ_c^2)	0.5
KL decomposition threshold (s)	$1e-2$
number of terms in KL decomposition (N_{KL})	6
degree of polynomial chaos expansion (n)	3
hyperbolic truncation set tuning parameter (q)	0.5
dispersion parameter (δ)	0.2

Table 3.5 | Mesh and computational parameters to generate a random Reynolds stress tensor field for the flow in a square duct.

of the flow domain, with $h = D/2$ as half-duct height. For both the deterministic and the stochastic simulation, we use a periodic boundary conditions in streamwise direction, a no-slip boundary condition at the walls and a symmetry boundary condition at the remaining two boundaries.

The mesh and RRSTF parameters are presented in Table 3.5. For the deterministic solve, we use the Launder-Sharma $k-\epsilon$ RANS turbulence model with finer mesh (grading) near the walls, such that $y^+ < 1$. This avoids the use of wall-functions which are one of the potential sources of uncertainty in RANS simulations. The stochastic simulation uses the same mesh as the deterministic simulation. The values of the correlation lengths l_x, l_y roughly represents the assumed (or observed) length of flow structures in x and y directions, respectively.

Figure 3.14 depicts the level of anisotropy plotted in barycentric coordinates within a barycentric triangle at location $(x/H, y/H) \approx (0.50, 0.23)$. The projection of the random Reynolds stresses on to the barycentric triangle is discussed in Appendix 3.8.1. The deterministic an-isotropy state and the sample mean state are at a certain distance which is proportional to the dispersion parameter. Also, for a larger dispersion parameter, the sample states lies further away from the deterministic state. The spread of the samples represents the level of uncertainty introduced in the Reynolds stress tensor field.

In Figure 3.15, the uncertainty in the velocity is presented alongside the DNS profiles at different x locations. As expected, the deterministic simulation predicts zero in-plane velocity, while the stochastic simulation predicts a non-zero mean velocity closer to DNS. The uncertain bound of ± 2 standard deviations, completely envelops the DNS solution at all three locations. A large variance in the stochastic solution can be attributed to – (i) arbitrarily chosen hyper-parameters for the generation of RRSTF, (ii) not considering any available data, and, (iii) the mean velocity being very sensitive to a small change in the Reynolds stresses.

The qualitative comparison of the stochastic mean and DNS fields is presented in Figure 3.16. Unlike the deterministic solver, the stochastic solver is able to capture the secondary flow perpendicular to the streamwise direction. The mesh density of the stochastic simulation is evidently lower than that of the DNS. Even with a coarse mesh the stochastic RANS

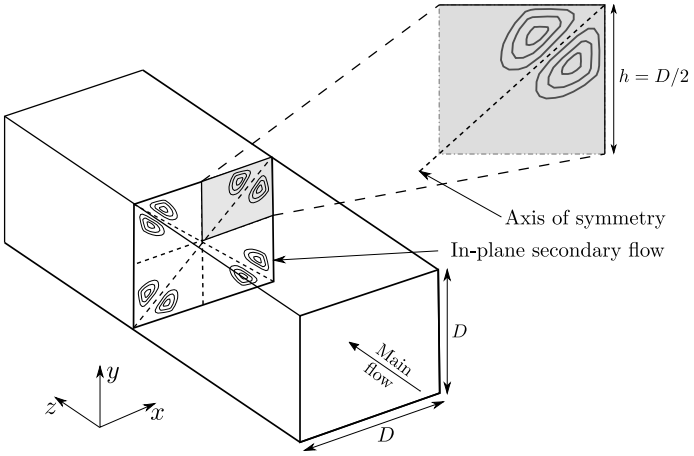


Figure 3.13 | Computational domain for turbulent flow in a square duct.

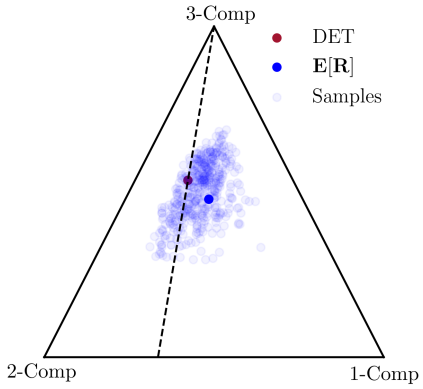


Figure 3.14 | Randomly sampled Reynolds stresses at location $(x/H, y/H) \approx (0.50, 0.23)$, projected onto the barycentric triangle (Appendix 3.8.1).

simulation is able to capture the rotating secondary flow pattern in accordance with DNS. However, the magnitude and direction of the field, depends on the parameters used in the construction of the random stress tensor, especially on the dispersion parameter.

Note that a detailed study of the parameters, similar to that conducted for the flow over periodic hill with random eddy viscosity, can also be carried out with random Reynolds stress tensor (not shown here for the sake of brevity).

3.7. CONCLUSIONS

In this work, we employed intrusive polynomial chaos to solve high-dimensional stochastic RANS problems using the popular finite-volume library OpenFOAM. In contrast to the notion of the IPC method being difficult to implement, the current solver provides an efficient

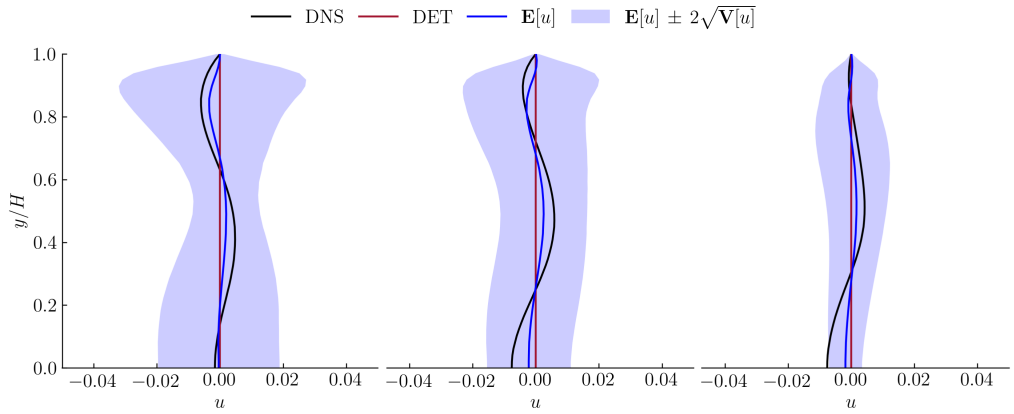


Figure 3.15 | Mean and variance of the u-component of velocity compared with the deterministic and DNS solutions, at locations $x/H = 0.25$ (left), $x/H = 0.50$ (center) and $x/H = 0.75$ (right).

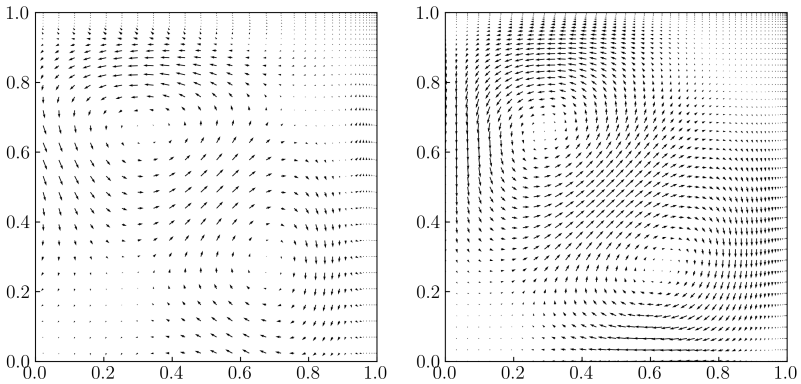


Figure 3.16 | Qualitative comparison of the secondary velocity field between stochastic mean flow (left) and DNS (right).

implementation which requires minimal top level changes in the existing deterministic code. We investigated various aspects of the stochastic solver including the effects of - the correlation length scales, the cardinality of the truncated polynomial expansion, the tuning parameter of the hyperbolic truncation set, different turbulence models, data-assimilation etc. The IPC based stochastic solver was tested with two high-dimensional random fields - the random eddy viscosity field and the random Reynolds stress tensor field. The computational cost for both the random field models was found to be of the same order. The results obtained from the stochastic simulations (QoIs) reveal that these random field models, upon proper parameter tuning, can certainly provide uncertainty bounds that envelops most of the realizable flow states. Tuning the hyper-parameters based on prior knowledge may allow for thinner uncertain bounds in QoIs. For a high variance random field, divergence was realized, where the higher modes of the polynomial expansion failed to converge after a few initial explicit iterations. As discussed in section 3.6.1, the IPC solver avails a significant im-

provement in the computational cost and the convergence as compared to the conventional UQ methods.

This work promotes further development in the direction of polynomial chaos method for uncertainty quantification in CFD. A more efficient implementation of the IPC solver can be explored with higher polynomial degree, different covariance kernel functions and an implicit algorithm to simultaneously solve for all the modes. As a future work, the IPC solver will be employed in the uncertainty quantification and its non-linear propagation in a CFD simulation of an array of wind turbines.

3.8. APPENDIX

3.8.1. PROJECTING REYNOLDS STRESS ON A BARYCENTRIC TRIANGLE

The Reynolds stress tensor can be decomposed into two parts, an isotropic component $2/3k\delta_{ij}$ and an anisotropic tensor which is given as the deviatoric part of the Reynolds stress normalized by the turbulent kinetic energy:

$$A_{ij} := \frac{R_{ij}}{2k} - \frac{\delta_{ij}}{3} \tag{3.52}$$

The eigen-decomposition of the anisotropy tensor results into eigenvalue tensor,

$$\mathrm{A} = \mathrm{V}\boldsymbol{\Lambda}\mathrm{V}^T, \tag{3.53}$$

where V is the matrix of orthogonal eigenvectors $[v_1, v_2, v_3]$ and $\boldsymbol{\Lambda}$ is the corresponding eigenvalue matrix $\mathrm{diag}[\lambda_1, \lambda_2, \lambda_3]$, such that $\mathrm{tr}[\boldsymbol{\Lambda}] = 0$ and $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Using linear relations, these eigenvalues can be mapped to unique coordinates (C_{1c}, C_{2c}, C_{3c}) in a Barycentric triangle, which in turn represents the state of anisotropy in a turbulent flow at a given point:

$$C_{1c} = \lambda_1 - \lambda_2, \quad C_{2c} = 2(\lambda_2 - \lambda_3), \quad C_{3c} = 3\lambda_3 + 1. \tag{3.54}$$

Note that, $C_{1c} + C_{2c} + C_{3c} = 1$. When any of these coordinates equals to one, the anisotropy is said to achieve a limiting state. In a Barycentric map, these limiting states are represented by the three vertices that form an equilateral triangle. Thus, C_{1c}, C_{2c} and C_{3c} represents 1-component, 2-component and 3-component turbulence, respectively. In Cartesian coordinates, if the location of these vertices is expressed as $(x_{1c}, y_{1c}), (x_{2c}, y_{2c})$ and (x_{3c}, y_{3c}), then any arbitrary turbulent state (x, y) can be represented by the following linear combination of the limiting states:

$$x = C_{1c}x_{1c} + C_{2c}x_{2c} + C_{3c}x_{3c}, \quad y = C_{1c}y_{1c} + C_{2c}y_{2c} + C_{3c}y_{3c}. \tag{3.55}$$

All the realizable states lie within or on this triangle. The anisotropy state of the samples generated using the polynomial chaos expansion of the Reynolds stress tensor can therefore be represented and analyzed using the Barycentric triangle as shown in Figure 3.14.

4

UNCERTAINTY QUANTIFICATION ANALYSIS IN WIND-FARMS

The content of this chapter is was published in J. Parekh and R. Verstappen. Uncertainty quantification analysis for simulation of wakes in wind-farms using a stochastic RANS solver, compared with a deep learning approach. Computers & Fluids, 257:105867, 2023.

ABSTRACT: Quantification of uncertainties in Reynolds-Averaged Navier–Stokes (RANS) simulations has gained a considerable interest in turbulence modeling. We present two different approaches for the quantification and propagation of model-form and operational uncertainties in context of wind turbine RANS simulations. The first approach is based on a stochastic RANS solver in OpenFOAM using intrusive polynomial chaos method [86]. Here the uncertainties are propagated through a single (large) simulation for the coupled coefficients of the polynomial expansion. The second approach is a surrogate based uncertainty quantification (SBUQ) method. The surrogate model comprises of a 3D U-Net neural network (trained over a single wind turbine) combined with a wake superposition model in order to the prediction of flow field in an array of wind turbines. The above-mentioned approaches are applied for uncertainty quantification analysis in RANS simulations of two turbulent engineering flow problems - (i) a wake past a single wind turbine, and (ii) wake interactions and power losses in a an array of wind turbines. The results show that the uncertain RANS solutions from the two approaches are able to reasonably capture the reference high-fidelity solution. We also discuss comparisons between the two approaches including computational cost, applicability, generality etc. The two methods can be further explored and applied to engineering applications where it is critical to compute the turbulent RANS solution in presence of various sources of uncertainties.

4.1. INTRODUCTION

The energy consumption by the developed and the developing countries is expected to increase by about 28% from 2015 to 2040 [47]. In the quest of alleviating global warming and limiting the carbon release, a significant investment is directed towards sustainable renewable energy sources. In the last few decades, the wind energy contributions via large wind farms has proven to be one of the most reliable sustainable energy sources. This demands more attention towards the development of further efficient wind farms using computational fluid dynamics (CFD) simulations. These simulations are based on fast and cost-effective numerical solvers that can provide details of the flow field in a wind farm, particularly the wake interactions and their effect on the power output of individual wind turbines. However, the estimated power at each wind turbine in a wind farm is highly dependent on the accurate prediction of the turbulence in the flow [4, 36, 112]. Most of the widely used turbulence models for CFD simulations are inherently uncertain resulting in significant variance in the flow field solutions [16, 133]. Additionally, the uncertainties in the boundary conditions that mimic the real environmental conditions may also contribute to the inaccurate predictions [7]. This motivates the efforts towards the quantification and propagation of model-form uncertainties in turbulence models and the uncertainties associated to the operating conditions in wind farm simulations.

Based on the fidelity and accuracy levels, various approaches are available for the prediction of wind turbine wakes. A high-order and high-fidelity direct numerical simulation (DNS), a large eddy simulation (LES) [17, 77], a relatively cheaper and reasonably accurate Reynolds-averaged Navier-Stokes (RANS) model [62, 90] and low-order engineering wake models [4, 36], are some of the most discussed approaches. Although expensive, the CFD approaches (DNS, LES, RANS) offer a high level of accuracy as compared to the computationally cheap wake-models. For industrial applications, especially for the simulation of wakes behind wind turbines, a non-scale-resolving approach like the RANS turbulence model is preferred over scale-resolving and computationally expensive high-order models like DNS and LES [101]. However, RANS models are often called out in research articles for inaccurate predictions of the wake flow field and the power production in a wind farm [11, 90, 93].

The central idea of the RANS turbulence modeling is based on the closure of the Reynolds stresses using the Boussinesq's hypothesis that assumes a linear relationship between the stresses and the mean rate-of-strain tensor with an eddy-viscosity as a proportionality constant [17]. The shortcomings of this hypothesis have been widely discussed in the literature, particularly for flows with streamline curvature, separation, strong pressure gradients and significant anisotropy [139] etc. Recent efforts have been made to overcome some of these deficiencies in RANS modeling particularly in the simulation of the wind turbine wakes. These include the traditional approach of improving the modeling of turbulent viscosity [119, 120] and the data-driven approaches for RANS closures [110, 111]. Most often it is not possible to determine beforehand if such flow features will be observed in a configuration. Therefore, RANS predictions are deemed inaccurate in the sense that they posses a structural or model-form uncertainty [15]. Based on more sophisticated constitutive relations than the Boussinesq's hypothesis, efforts have been made to overcome the limitations of eddy-viscosity models. These include nonlinear eddy-viscosity models [109] and Reynolds stress transport models (RSTM) [63]. However, despite of their mathematical superiority,

these models are not as robust as the linear eddy-viscosity models and their usage is limited to a small set of turbulent flow problems.

In light of the above discussion, it is essential to quantify the structural uncertainty in RANS turbulence models. Previous efforts towards quantifying the model-form uncertainties in RANS using a non-parametric approach include introducing perturbations in the eigenspace of the anisotropy tensor, thereby resulting in a perturbed Reynolds stress tensor [22]. More specifically, the global perturbations in turbulent kinetic energy (TKE), eigenvalues, and eigenvectors of an anisotropy tensor relates to the variations in its amplitude, shape, and orientation, respectively. Although this method has been applied for various engineering problems [39, 40, 96], the assumption of such globally-uniform low-dimensional input uncertainty results in an inaccurate estimation of the uncertainties. More recently, a high-dimensional local perturbation approach have been employed by directly injecting the randomness into the Reynolds stress tensor field using spatially correlated positive-definite random matrices while guaranteeing the realizability [134]. The resulting random Reynolds stress tensor field (RRSTF) accounts for the structural uncertainty of the RANS model. In authors' previous work, a generalized (intrusive) polynomial chaos based stochastic solver was developed to propagate uncertainties in RANS models using the random matrix approach for eddy-viscosity field or Reynolds stress tensor field [86]. This method has been studied using a number of benchmark turbulent flow problems and has shown potential for uncertainty quantification and propagation in engineering applications. For more details about the stochastic solver, the reader is referred to [86].

In context of wind turbine simulations the inlet operating conditions define (model) the atmospheric boundary layer profile. The predefined inlet profile helps avoiding the use of periodic boundary condition which in turn require large domain size in the streamwise direction, thus making the RANS simulations computationally less expensive. The operating conditions are determined from various measurements and are prone to uncertainties. In particular, the influence of the reference velocity and the turbulent intensity majorly governs the atmospheric boundary layer profile and may have variance due to imprecise measurements. Thus, it is important to quantify and propagate these uncertainties to study their effect on the overall performance of wind turbine/farms. Therefore, in addition to the randomness in the Reynolds stress tensor, in the present work we also include uncertain inlet conditions that (indirectly) contribute to the overall model-form uncertainty in the RANS simulations.

In wind energy applications, eigenspace perturbation of the Reynolds stress tensor was used to quantify and propagate model-form uncertainties in RANS simulations of a single wake behind a wind turbine [43] as well as for the wakes and power losses in wind farms [20]. In the present work, we consider the RANS model-form uncertainty in the Reynolds stress tensor (via the random matrices approach) as well as the uncertainties in the operational conditions for RANS simulations of a wake behind a wind turbine and for the wake interactions in a wind farm. Moreover, we use the previously developed stochastic solver [86] in order to propagate the different sources of uncertainties in a single simulation.

With the exponentially increasing applications of deep learning, we explored a potential application of neural networks towards making predictions for a flow field in a wind farm using only data from single wind turbine simulations. Recently, a U-Net neural network was developed for the prediction of 2D flow fields around an airfoil with an error of less than 3% on unseen data [116]. We adapt this neural network to device a 3D U-Net model

for predictions of flow fields (wakes) near a wind turbine. Moreover, we use the trained neural network along with a wake superposition model [36] in order to construct a surrogate that can (per wind turbine) predict the flow field in a wind farm. This idea is motivated from a recent work [117], where the authors used a fully-connected neural network, which in the present study is replaced with a more efficient and less expensive convolutional neural network. For a given sampling strategy, the above surrogate model avails the use of a surrogate based uncertainty quantification (SBUQ) approach.

The next sections of the paper are structured as follows. Section 4.2 describes the fundamentals of a RANS turbulence modeling and atmospheric boundary layer modeling. In section 4.3, we introduce the two uncertainty quantification approaches - (i) stochastic RANS using intrusive polynomial chaos, and (ii) surrogate based uncertainty propagation. The uncertainty quantification and propagation analysis for the numerical simulation of a wind turbine and a wind farm is presented in section 4.4. Section 4.5 summarizes the conclusions.

4.2. TURBULENCE MODELING

4.2.1. RANS TURBULENCE MODELS

The Reynolds-Averaged Navier-Stokes equations for an incompressible turbulent flow with a momentum source is given by

$$\nabla \cdot \overline{\boldsymbol{u}} = 0, \tag{4.1a}$$

$$\frac{\partial \overline{\boldsymbol{u}}}{\partial t} + (\overline{\boldsymbol{u}} \cdot \nabla)\overline{\boldsymbol{u}} = -\nabla \overline{p} + \nabla \cdot (\nu \nabla \overline{\boldsymbol{u}}) - \nabla \cdot \mathbf{R} + \boldsymbol{f}, \tag{4.1b}$$

where $\boldsymbol{u}$ is the velocity, p is the pressure, ν is the kinematic viscosity, $\mathbf{R} = \overline{\boldsymbol{u}'\boldsymbol{u}'}$ is the Reynolds-stress-tensor which couples the mean flow with the turbulence, and $\boldsymbol{f}$ is the forcing term. In the present study $\boldsymbol{f}$ models the effect of a wind turbine on the flow field. The overbar and the prime represents the mean and the fluctuating components of the respective variables. The most common class of RANS models is the linear-eddy-viscosity model which employs the Boussinesq's hypothesis. This hypothesis assumes a linear relation between Reynolds stresses and the mean strain-rate,

$$-\mathbf{R} \approx 2\nu_t \overline{\boldsymbol{S}} - \frac{2}{3}\boldsymbol{I}k, \tag{4.2}$$

where ν_t is the eddy viscosity, $\overline{\boldsymbol{S}} = (\nabla \overline{\boldsymbol{u}} + (\nabla \overline{\boldsymbol{u}})^T)/2$ is the mean strain-rate tensor, $\boldsymbol{I}$ is the second order identity tensor, and $k := \frac{1}{2}\mathrm{tr}(\mathbf{R})$ is the turbulent kinetic energy. The eddy viscosity is computed after solving the equation(s) for the turbulent flow quantities like the turbulent kinetic energy k and the turbulent energy dissipation ϵ, or the specific dissipation ω.

Despite their popularity, the eddy viscosity models suffers to accurately capture the flows with significant separation, curvature, anisotropy, impinging, etc. The Reynolds-stress-transport models (RSTMs) overcome many of these issues by directly solving the

model equations for the Reynolds stresses. In the present work, we employ two linear-eddy-viscosity models, namely the $k-\omega$ SST model and the realizable $k-\epsilon$ model, as well as the Launder-Reece-Rodi (LRR) Reynolds stress transport model. For the sake of brevity, we defer the discussion of these models to the available literature [64, 78, 126].

4.2.2. ATMOSPHERIC BOUNDARY LAYER MODELING

A neutral atmospheric boundary layer (ABL) modeling involves inlet boundary conditions providing log-law type ground-normal inflow boundary conditions for wind velocity and turbulence quantities [42, 141]. The expression for the ground-normal profiles of - streamwise flow speed u, turbulent kinetic energy k, turbulent kinetic energy dissipation rate ϵ and specific dissipation rate ω are given by:

$$u = \frac{u^*}{\kappa} ln\left(\frac{z+z_0}{z_0}\right), \quad k = \frac{(u^*)^2}{\sqrt{C_\mu}}, \quad \epsilon = \frac{(u^*)^3}{\kappa(z+z_0)}, \quad \omega = \frac{u^*}{\kappa\sqrt{C_\mu}(z+z_0)}, \tag{4.3}$$

with u^* as the friction velocity given by

$$u^* = \frac{u_{ref}\kappa}{ln\left(\frac{z_{ref}+z_0}{z_0}\right)},$$

where z_{ref} is the reference height, u_{ref} is the reference mean streamwise speed at z_{ref}, z_0 is the aerodynamic roughness length, C_μ is te empirical model constant and κ is the von Kármán constant. Note that for a given turbulent intensity I and mean streamwise speed $\bar{u}$, one can compute the turbulent kinetic energy k using the relation $I = \sqrt{2k/3}/\bar{u}$ and thereby compute the friction velocity u^*. Subsequently, one can compute the roughness length for a given reference height and reference mean streamwise speed.

4.3. UNCERTAINTY QUANTIFICATION APPROACHES

The characterization of model uncertainties in RANS simulations can be broadly classified into two methods [133] - (i) where the uncertainties are injected into the *model form*, e.g., in the transport equations for the fields like k, ϵ, ω etc., and (ii) where the uncertainties are introduced in the *model output*, e.g., in the eddy viscosity field or in the Reynolds stress tensor field. In the present work we employ the latter form of characterization using a random Reynolds stress tensor field (RRSTF) which is more desirable for quantifying the uncertainties in flows where the Boussinesq approximation fails or is inapplicable.

In RANS wind turbine simulations, the operating conditions used to model the atmospheric boundary layer and atmospheric wall functions may be highly uncertain. This may lead to a significant variance in the RANS computed flow fields. Therefore, it is important to quantify these sources of uncertainties and study their effect on various quantities of interest (QoIs). Following the discussion of the atmospheric boundary layer modeling in section 4.2.2, in the present study, we consider the uncertainties in the hub height velocity u_h and the turbulent intensity I_h at hub height.

In the following sections, we will discuss two different approaches used in this work to collectively quantify and propagate the above-mentioned uncertainties through the linear-eddy-viscosity based RANS models discussed in section 4.2.1.

4.3.1. STOCHASTIC RANS USING INTRUSIVE POLYNOMIAL CHAOS

Introducing randomness in the Reynolds stress, the stochastic form of RANS equations becomes:

$$\nabla \cdot \overline{\boldsymbol{u}} = 0, \tag{4.4a}$$

$$\frac{\partial \overline{\boldsymbol{u}}}{\partial t} + (\overline{\boldsymbol{u}} \cdot \nabla)\overline{\boldsymbol{u}} = -\nabla \overline{p} + \nabla \cdot (\nu \nabla \overline{\boldsymbol{u}}) - \nabla \cdot \mathbf{R}(\boldsymbol{x}, \omega) + \boldsymbol{f}. \tag{4.4b}$$

where $\mathbf{R}(\boldsymbol{x}, \omega)$ represents a random Reynolds stress tensor field (RRSTF).

In this section we present the stochastic RANS solver developed recently by the authors [86]. Here we made use of the finite-volume library OpenFOAM [25]. The steps involved are shown in Figure 4.1 (top). Deferring the full elaboration to [86], we provide a brief summary of the major steps involved in the algorithm used to solve the above stochastic model.

- The characterization of the RRSTF is performed using the random matrix approach proposed by Xiao et al. [134]. Here, while guaranteeing the realizability, the randomness is introduced directly in the Reynolds stress components using a Gaussian random field $\mathbf{G}(\boldsymbol{x})$ such that,

$$\mathbf{R}(\boldsymbol{x}) = \mathbf{L}_R^T \mathbf{G}(\boldsymbol{x}) \mathbf{L}_R, \tag{4.5}$$

  where $\mathbf{L}_R$ is the upper triangular matrix resulting from the decomposition (Cholesky factorization) of the mean Reynolds stress $\mathbf{R}^{(det)} = \mathbf{L}_R^T \mathbf{L}_R$. The variance in $\mathbf{G}(\boldsymbol{x})$ is governed by a dispersion parameter δ, which for a three-dimensional flow takes a value in the range $0 < \delta < 1/\sqrt{2}$. The covariance kernel of $\mathbf{G}(\boldsymbol{x})$ may have spatially varying correlation lengths $\{l_x, l_y, l_z\}$ and marginal variance σ_c^2. After the construction of RRSTF, it is represented using a finite (N_{KL}) set of random variables using Karhunen-Loève (KL) decomposition.

- The propagation of the uncertainties is carried out using generalized polynomial chaos approach [137]. In particular, we use the intrusive polynomial chaos (IPC) method, where the deterministic RANS equations are reformulated resulting in a set of governing equations for the polynomial chaos mode strengths (also called modes or coefficients) of the output [86]. The above steps result in the following set of equations governing the evolution of the modes of velocity and pressure, for $k = 0, 1, ..., P$,

$$\begin{aligned} \nabla \cdot \overline{\boldsymbol{u}}_k &= 0, \\ \frac{\partial \overline{\boldsymbol{u}}_k}{\partial t} + \sum_{i=0}^{P} \sum_{j=0}^{P} (\overline{\boldsymbol{u}}_i \cdot \nabla)\overline{\boldsymbol{u}}_j M_{ijk} &= -\nabla \overline{p}_k + \sum_{i=0}^{P} \sum_{j=0}^{P} \nabla \cdot (\nu_i \nabla \overline{\boldsymbol{u}}_j) M_{ijk} - \nabla \cdot \mathbf{R}_k(\boldsymbol{x}) + \boldsymbol{f}_k, \end{aligned} \tag{4.6}$$

  where, for a random variable $\boldsymbol{v}$, the term $\boldsymbol{v}_i$ is the i^{th} mode in the polynomial chaos expansion $\boldsymbol{v}(\boldsymbol{x}, t, \omega) = \sum_{i=0}^{P} \boldsymbol{v}_i(\boldsymbol{x}, t)\psi_i(\boldsymbol{\xi}(\omega))$ where ψ_i is the multivariate polynomial based on the joint probability density function of all the input random variables $\boldsymbol{\xi}$, $P + 1 = (d + n)!/(d!\, n!)$ with d being number of input random variables and n being

the highest polynomial order of the polynomials $\{\psi_i\}$, and $M_{ijk} = \langle \psi_i \psi_j \psi_k \rangle / \langle \psi_k \psi_k \rangle$. The expansion coefficients of the Reynolds stress $\mathbf{R}_i(\boldsymbol{x}) = \mathbf{L}_R^T \mathbf{G}_i(\boldsymbol{x}) \mathbf{L}_R$, with $\mathbf{G}_i(\boldsymbol{x})$ being the expansion coefficient of $\mathbf{G}(\boldsymbol{x})$, are approximated using a quadrature-based non-intrusive polynomial approach.

In RANS wind turbine simulations, the uncertainties in the inlet conditions u_h, I_h is directly accounted in the modes via the atmospheric boundary layer inlet and atmospheric wall functions boundary conditions. The solver framework is based on a deterministic solver in OpenFOAM called *pimpleFOAM* and is again depicted in Figure 4.1 (bottom). In order to significantly reduce the computational cost and attain the same level of accuracy a hyperbolic truncation set with a tuning parameter (q) which controls the number of terms in the polynomial chaos expansion was used in [86]. Once the modes of a random variable are computed, its mean and variance are computed from

$$\mathbf{E}[\boldsymbol{v}] = \boldsymbol{v}_0(\boldsymbol{x}, t), \qquad \mathbf{V}[\boldsymbol{v}] = \sum_{i=1}^{P} \boldsymbol{v}_i^2(\boldsymbol{x}, t) \langle \psi_i^2 \rangle. \tag{4.7}$$

4.3.2. A SURROGATE BASED UNCERTAINTY PROPAGATION APPROACH

In contrast to the intrusive polynomial chaos methods, the surrogate based uncertainty quantification (SBUQ) is a sampling based approach. The steps involved in SBUQ are presented in Figure 4.2 (top). A three-dimensional (3D) U-Net deep learning model is trained to predict the flow field for a single wind turbine with uncertainties as the inputs and QoIs as the outputs. Then, the 3D U-Net model is used in conjugation with a wake-superposition model in order to predict the flow field for an array of wind turbines.

The inspiration of employing a 3D U-Net model arises from the work of Thuerey et al. [116]. They investigated the accuracy of a 2D U-Net model for the inference of RANS airfoil flow solutions. In the present study, we extend this model for 3D flow configurations in context of wind turbine simulations. A schematic of the neural network is shown in Figure 4.2 (bottom). For the sake of brevity, we only provide an overview of the adaptation of the neural network architecture used in this work and defer - (i) the elaborate discussion of the U-Net model to [116], and (ii) the fundamentals of artificial neural network (auto-encoders in particular) to [5, 8]. For a given input x, we want to approximate the output y of a true function $\hat{f}$ using a representation f such that $y \approx f(x, \boldsymbol{w})$, where $\boldsymbol{w}$ are the degrees of freedom. In the present study f is represented using a U-Net neural network which avails a non-linear mapping between the inputs and the outputs. A U-Net neural network is a special case of an encoder-decoder architecture and usually has a bowtie structure that translates spatial information into features that are extracted using convolutional layers. In the encoding part of the network, using the convolutions and pooling with strides, the size of the data is regularly down-sampled by a factor of two. This allows the gradual growth of the number of feature channels representing increasingly large-scale and abstract features. In the decoding part of the architecture, using de-convolutions and de-pooling with strides, the spatial resolution regularly increases while the number of feature channels decreases. In order to ensure a better reconstruction of the solution, low-level input information is passed to the output channels using skip-connections. These connections concatenate the corresponding encoding-decoding channels, thereby doubling the number of output

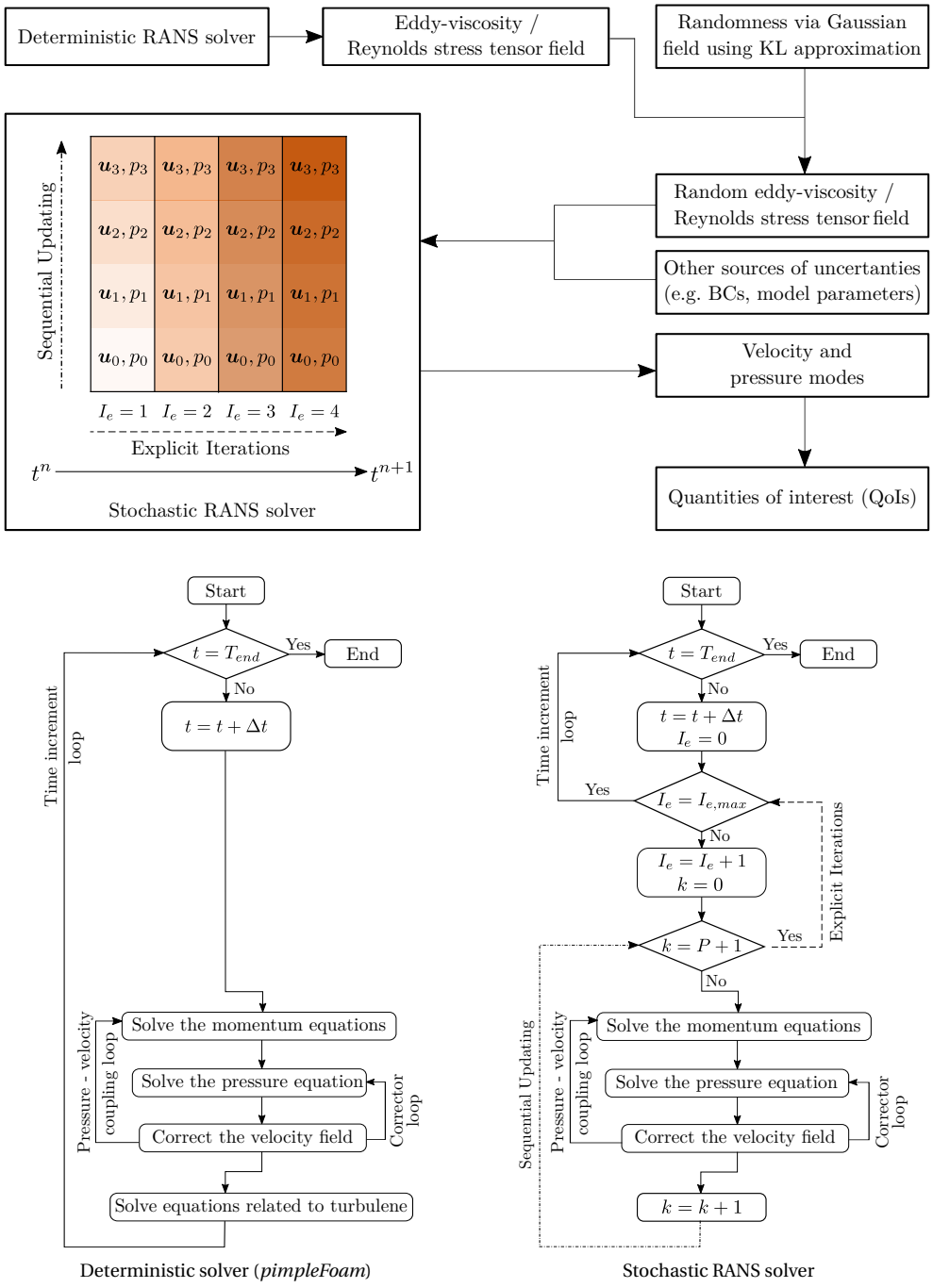


Figure 4.1 | (top) Flow chart of the steps involved in quantifying and propagating model output-form and other sources of uncertainties in RANS simulations (P = 3), and (bottom) Graphical representation of the algorithm implemented in the deterministic RANS solver (left) and stochastic RANS solver (right) [86].

channels at each decoding level. For flow around a wind turbine, the inputs of the U-Net network consist of all (six) components of Reynolds stress tensor field (RSTF) and the inlet conditions (u_h, I_h) spanned over the flow domain which is covered by a $128 \times 128 \times 128$ mesh. The output of the network consists of the velocity (u) and turbulent intensity (I) fields. Every block of the U-Net network consists of an activation function (mostly non-linear), a (de)convolutional layer and a batch normalization layer. As depicted in Figure 4.2 (bottom), in the encoder part we use 7 blocks to convert an input of size $128^3 \times 8$ into an array of 512, while in the decoder part we use another 7 blocks to reconstruct the target output of size $128^3 \times 2$. These blocks mostly use the (de)convolution kernels of size 4^3 except that the two inner layers of the encoder and decoder part use a kernel of size 2^3. This results in a neural network with approximately 46.7 *million* weights. We use the leaky ReLU and regular ReLU activation functions in the encoding and decoding parts, respectively, except in the first encoder block where a linear activation function is used.

After training the surrogate (U-Net neural network) over a single wind turbine, we predict the solution for an array of wind turbines by using the surrogate along with a wake superposition model as depicted in Figure 4.2 (bottom). A popular wake superposition model is the sum of squares (SOS) model which can be written as [36],

$$\left(1 - \frac{u_{i+1}}{u_{in}}\right)^2 = \sum_{j=1}^{n_{up}} \left(1 - \frac{u_{i+1,j}}{u_j}\right)^2 = \sum_{j=1}^{n_{up}} \left(\Delta u_{i+1,j}\right)^2, \tag{4.8}$$

where, u_i is the velocity at turbine i, u_{ij} is velocity at turbine i explicitly in the presence of the wake of turbine j and n_{up} is the number of wind turbines upstream. The subscript in denotes the inflow. Thus, the SOS model states that in an array of wind turbine configuration, the square of the overall velocity deficit normalized by the inflow condition is equal to the sum of the squares of the local normalized velocity deficits of all the upstream wind turbines. to model energy conservation, we assume that the increase of the turbulent kinetic energy (TKE) at a turbine can be written as the sum of the local added TKEs of all the upstream wind turbines,

$$k_{i+1} - k_{in} = \sum_{j=1}^{n_{up}} \left(k_{i+1,j} - k_j\right) = \sum_{j=1}^{n_{up}} \Delta k_{i+1,j}, \tag{4.9}$$

where, k_i is the TKE at turbine i, k_{ij} is TKE at turbine i explicitly in the presence of the wake of turbine j. Finally, the turbulent intensity can then be computed using $I = \sqrt{2k/3}/u_{in}$.

The following steps summarize the overall process of using a U-Net neural network and a wake superposition model to predict the flow field for an array of wind turbines. For each sample in the input:

- Step 1: Assign the initial and the boundary conditions of the flow field. At the first wind turbine $(i = 1)$, extract the hub height velocity u_h and turbulent intensity I_h.
- Step 2: Use the trained surrogate model (U-Net) to predict the velocity and the turbulent intensity fields for the i^{th} wind turbine.
- Step 3: At the $(i+1)^{th}$ wind turbine, compute the velocity deficit and the added TKE fields for all the upstream turbines and approximate the hub height velocity and turbulent intensity using the respective wake superposition models (4.8, 4.9).

Repeat the process (Steps 2 and 3) sequentially to compute the wake field for the subsequent wind turbines. For a given sampling strategy, the above discussed combined model (3D U-Net with wake-superposition) can now be used as a surrogate in the SBUQ approach to compute the statistic of the uncertain flow field in an array of wind turbines.

4.4. NUMERICAL RESULTS

Using the two approaches discussed in section 4.3, we present an uncertainty quantification analysis for the RANS simulations of wakes behind a single wind turbine and wake interactions in an array of wind turbines in the next sections.

4.4.1. WAKE BEHIND A SINGLE WIND TURBINE

SIMULATION SETUP

For the deterministic RANS simulation of the wake behind a single wind turbine, we consider the Vestas V80-2MW turbine, which has a hub height z_h of $70m$, a turbine diameter D of $80m$ and a thrust coefficient C_T of 0.8. As can be seen in Figure 4.3, the size of the computational domain is $30D \times 10D \times 4.3D$ (in the streamwise (x), spanwise (y) and vertical (z) directions, respectively). The wind turbine is location $5D$ downstream of the inlet. A Cartesian mesh with higher cell density near the wind turbine in streamwise and spanwise directions is used to capture steep gradients in the flow. In the present work, we use the actuator disk approach to model the effect of turbine blades as a body force [130]. The turbine induced force is distributed uniformly across the cells in the rotor disk region and is represented as a source term in the momentum equations (4.1) [93],

$$\boldsymbol{f} = -\frac{1}{2} u_0^2 A C_T, \tag{4.10}$$

where, u_0 is the mean upstream undisturbed velocity and A is the rotor area. The boundary conditions at the inlet and the top surface are set in order to simulate an atmospheric boundary layer profile [42, 141] such that the hub height velocity u_h and turbulent intensity I_h are $8m/s$ and 5.4%, respectively. At the outlet a constant pressure is applied with zero gradient conditions for both the velocity and the turbulence variables (depending on the model used). The bottom surface is modeled as a rough ground i.e. no-slip conditions for the velocity, zero gradient for the pressure and wall functions for the turbulence variables (depending on the model used). In the spanwise direction a zero gradient boundary condition is applied to all variables. Based on the analysis of different linear-eddy-viscosity RANS models in [43], we choose the $k-\omega$ SST model [78, 126] to simulate the wake behind a single wind turbine. We also use the LRR Reynolds stress transport model [64] to make comparisons with the two-equation model and to verify the expected better performance of the RSTM.

The stochastic simulation uses the same mesh as the deterministic simulation. The parameters used to characterize the random hub height variables (u_h, I_h) and the random Reynolds stress tensor field (RRSTF) are presented in Table 4.1. The hub height variables are uniformly distributed with a coefficient of variation (CoV) of about 4% and 5%, respectively. Such levels of variations in the measurement data are often reported in the literature [6]. The parameters chosen for the characterization of RRSTF are based on the previous work by

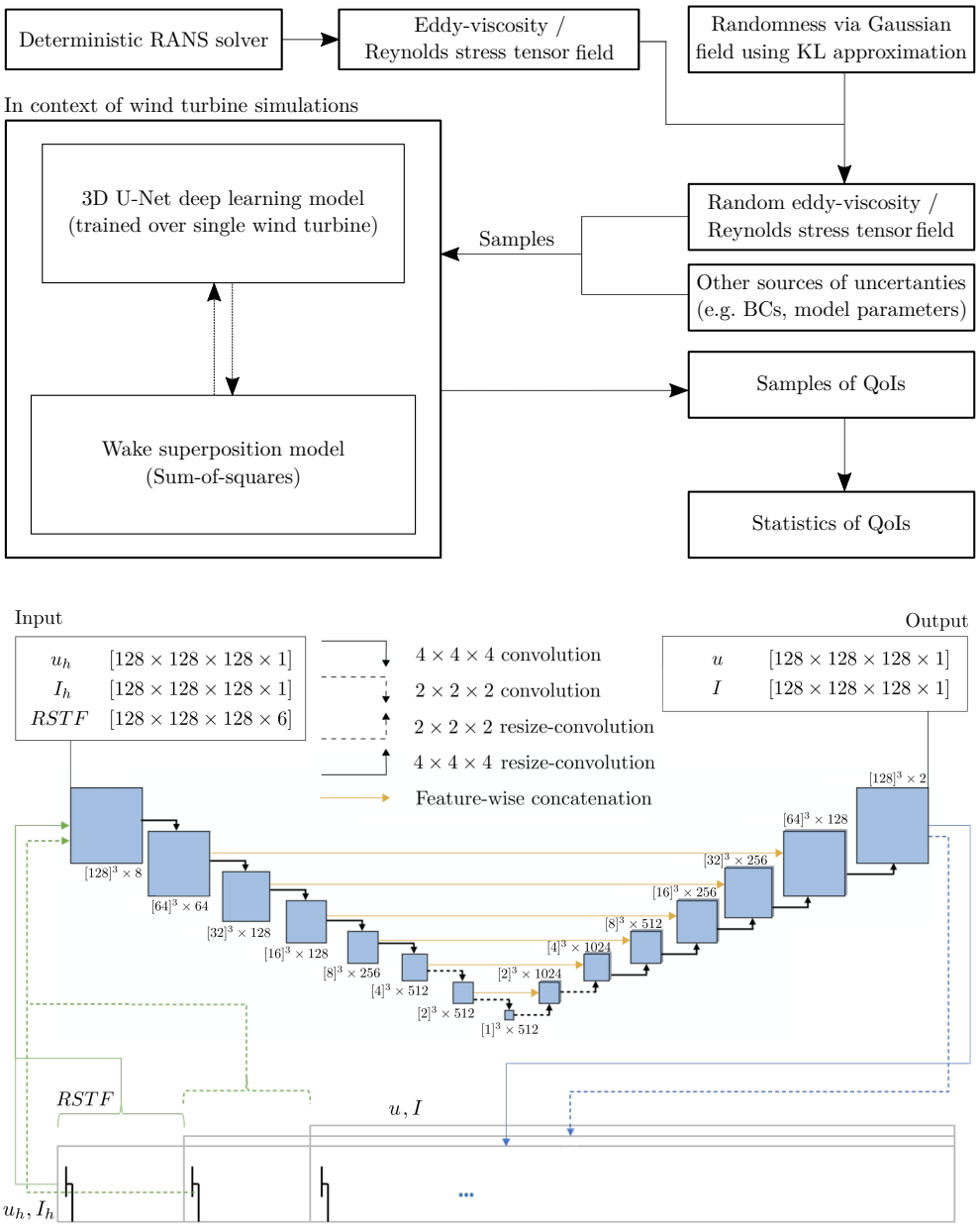


Figure 4.2 | (top) Flow chart of the steps involved in SBUQ - surrogate based quantification and propagation of model output-form and operational uncertainties in RANS simulations, and (bottom) Graphical representation of the surrogate model i.e. a 3D U-Net (adapted from [116]) combined with a wake superposition model, where $[a]^3$ implies a 3D array of shape $a \times a \times a$.

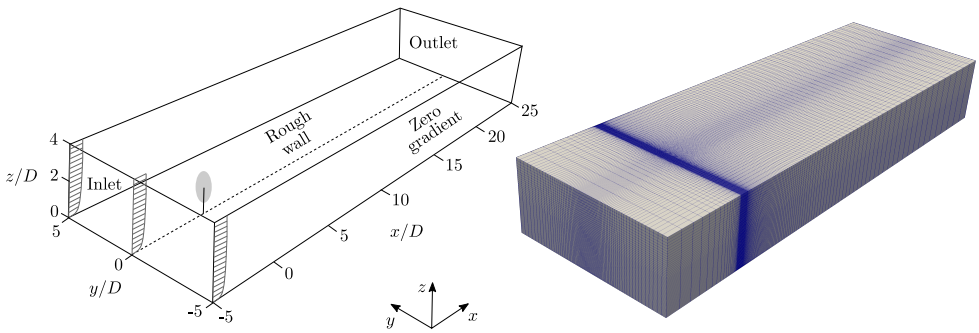


Figure 4.3 | Computational domain for the wind turbine simulation. The turbine is situated $5D$ distance from the inlet, while the outlet is $25D$ from the turbine to capture the turbulent wake (left). The computational mesh has fine cells across the actuator disk to capture the momentum and pressure jump, and smaller cells near the rough wall to capture the steep gradients (right).

the authors [86]. The values of the correlation lengths l_x, l_y, l_z roughly represent the length of flow structures in x, y and z directions, respectively, observed typically from the baseline (deterministic) RANS simulation. In a recent study on turbulence scales of a wind turbine wake [119], it was found that at relatively low turbulence levels the turbulence length scale in the wake region is about $D/4$ and it gradually increases for higher levels of turbulent intensity. The present work considers moderate levels of turbulence and thus the streamwise correlation length l_x is varied in the range $[D/4, 4D]$ while l_y and l_z are fixed at D and $D/2$, respectively. The predicted uncertainty bounds showed no significant difference beyond $l_x = 2D$. Thus, we continue to use $l_x = 2D$ for the remaining study, which in turn allows lower number of KL modes while capturing most of the variance in RSTF. The dispersion parameter is chosen such that the (field averaged) CoV of the random field is close to the CoV of the hub height variables. This is particularly useful to later comment on the relative contribution of each input random variable towards the variance in QoIs for similar CoVs. Moreover, choosing very high levels of perturbation may increase the chances of including non-physical realizations. The remaining hyper-parameters are based on the findings in authors' previous work [86]. The threshold s controls the number of KL decomposition modes while capturing most of the random field's variance, whereas the turning parameter q controls the number of higher order polynomial interactions to be considered for a given degree of PCE n. In the construction of the RRSTF we use OpenTURNS [89] to obtain the random field $\mathbf{G}(x)$ and subsequently compute the modes of the KL decomposition.

Figure 4.4 depicts the level of anisotropy plotted in barycentric coordinates within a barycentric triangle at location $(x/D, y/D, z/D) = (2.0, 0.0, z_h/D)$ for both turbulence models used. The projection of the Reynolds stresses onto the barycentric triangle is discussed in Appendix 4.6.1. The uncertainty introduced in the Reynolds stress tensor field is represented by the spread of the samples. The distance between the deterministic anisotropy state and the sample mean state is proportional to the dispersion parameter. As the dispersion param-

Parameter	Distribution or Value
hub height velocity (u_h [m/s])	uniform [7.5, 8.5]
hub height turbulent intensity (I_h [%])	uniform [5.0, 6.0]
Reynolds stress tensor field (RSTF, [m^2/s^2])	
correlation lengths (l_x, l_y, l_z)	$2.0D, 1.0D, 0.5D$
marginal variance (σ_c^2)	0.5
dispersion parameter (δ)	0.5
KL decomposition threshold (s)	$1e-2$
degree of polynomial chaos expansion (n)	3
hyperbolic truncation set tuning parameter (q)	0.5

Table 4.1 | Parameters used to characterize the random variables and the random fields for the simulation of a wake behind a wind turbine.

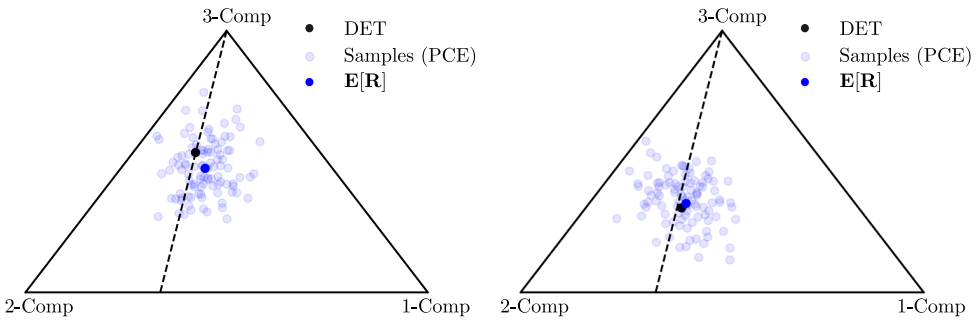


Figure 4.4 | Randomly sampled Reynolds stresses at location $(x/D, y/D, z/D) = (2.0, 0.0, z_h/D)$, projected onto the barycentric triangle for the uncertain $k-\omega$ SST (left), and LRR (right) turbulence models, see 4.6.1.

eter increases, the sample states (and the mean) lies further away from the deterministic state.

For visualizing the anisotropy in the wake region, we map the barycentric triangle, for the sake of consistency with other contour plots, into a *Viridis* colormap using the method proposed in [21]. As shown in Figure 4.5, the colors blue, yellow and green are assigned to the one, two and three component turbulence states, respectively, while the colors are interpolated for all the states in between these extremities. Using this mapping, we present in Figure 4.6 the Reynolds stresses anisotropy tensor representation for the deterministic RANS simulation and a sample from the PCE of RRSTF for both turbulence models at various locations downstream. It can be seen that, in the near wake region, the $k-\omega$ SST model underestimates the anisotropy of the shear layer as compared to the LRR model. Also, in comparison to the RSTM, the two-equation model tends towards the isotropic state (three component) rather quickly as we move further downstream. Note that the PCE sample for the two-equation model shows a higher level of anisotropy than the deterministic one. Based on the above discussion, we conclude that the level of uncertainty introduced in the Reynolds stress tensor field allows for a (locally) significant increase in the anisotropy and is sufficient for model-form uncertainty quantification analysis of the two RANS models.

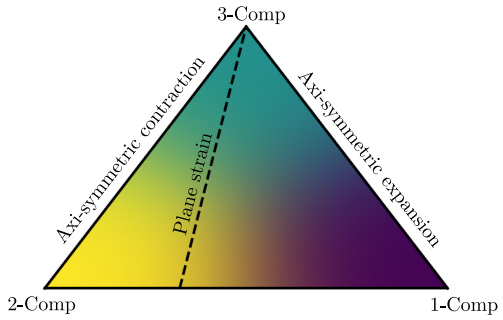


Figure 4.5 | Viridis colormap representation of the Barycentric triangle used to visualize the anisotropy in a flow region [21].

SIMULATION RESULTS

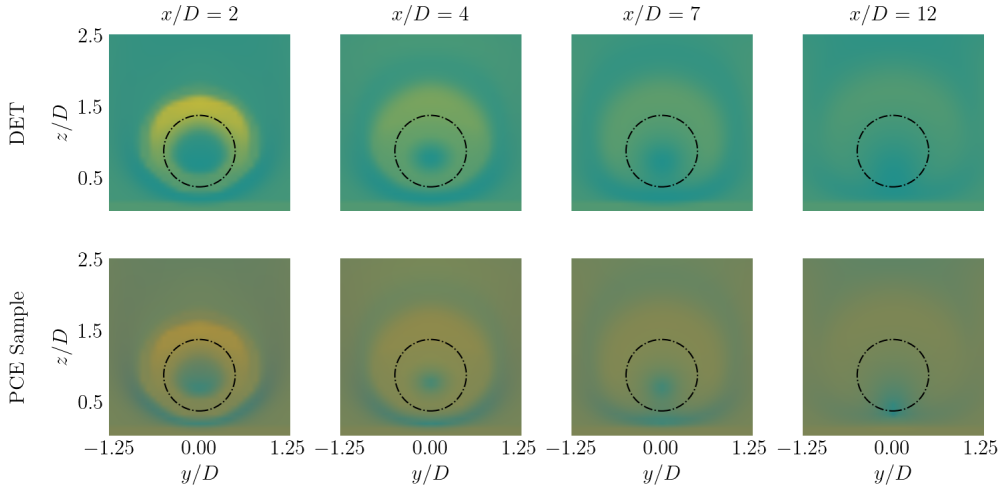
In this section, we present the uncertainty quantification results obtained using both the approaches discussed in section 4.3. The input uncertainties (the uncertain hub height variables u_h, I_h and the RRSTF) are characterized based on the hyper-parameters listed in Table 4.1. We consider the normalized velocity deficit and turbulent intensity fields as the QoIs. We compare the results obtained from the analysis of both turbulence models with a reference LES data [43]. Note that the LES solution corresponds to the mean of hub height variables in the present work. The LES framework in [43] consists of a pseudo spectral code with a computational domain of size $40D \times 10D \times 4.44D$ which is slightly larger than the one used for our RANS simulations (present work). Note, the LES domain is larger since it uses a fringe zone. The LES uses a precursor simulation for a fully-developed boundary layer flow under neutral conditions driven by an imposed pressure gradient and modeled the turbine-induced forces using a non-rotational actuator disk model. For further details of the LES see [2, 140].

In order to assess the effect of the uncertain inlet conditions alone (u_h, I_h), a forward propagation of these uncertainties through the stochastic RANS solver (using deterministic RSTF of the baseline) is performed (Appendix 4.6.2). The results obtained from this study show a significant contribution of the uncertain operating parameters towards the variance in the QoIs. Next we study the combined effect of the uncertain inlet conditions and the structural uncertainties in RANS.

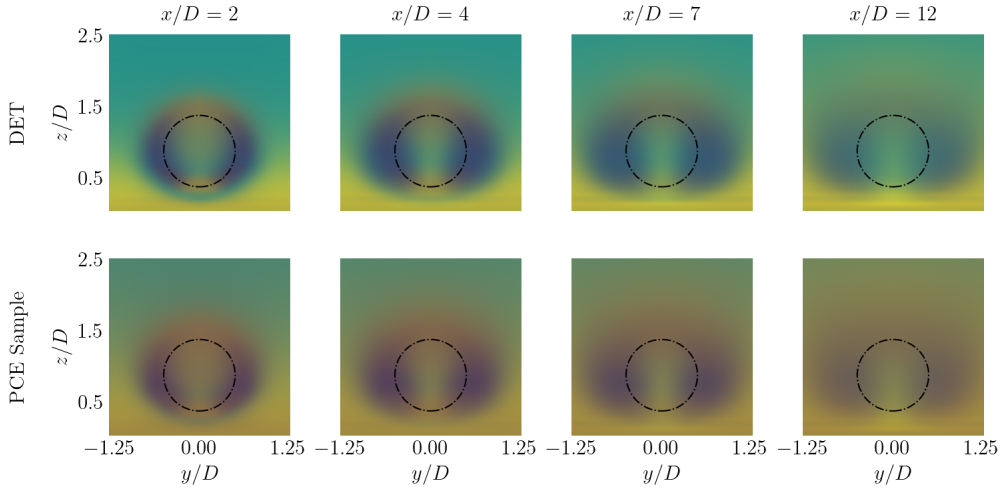
Following the discussion of the SBUQ approach in section 4.3.2, we now discuss the training, the validation and the test results obtained using the data-driven approach.

- After the characterization of the input uncertainties (section 4.4.1), we use a Quasi-Monte Carlo (QMC) sampling technique [45, 68] to sample 1300 independent samples of $\{u_h, I_h,$ RSTF$\}$. The RSTFs are sampled from the PCE of RRSTF. Therefore, we essentially sample from the joint probability distribution of $\{u_h, I_h, \xi_1, ..., \xi_{N_{KL}}\}$ which indeed represents the input random space for the stochastic RANS solver as well. For each sample we use a CFD solver to obtain the velocity and the turbulent intensity fields, see Appendix 4.6.4 for more details.

- From all the samples we extract the hub height velocity and turbulent intensity, the



(a)



(b)

Figure 4.6 | Visualization of the Reynolds stresses anisotropy tensor for the deterministic RANS simulation and a sample from the PCE of Reynolds-stress tensor using the Barycentric Viridis colormap map (Figure 4.5) at different locations in the streamwise direction for the uncertain (a) $k-\omega$ SST, and (b) LRR RANS turbulence models.

RSTF, the velocity and the turbulent intensity fields from $1D$ upstream to $6D$ downstream the turbine. A 128^3 mesh is used in the region of the extracted quantities which is the same as the input size of the U-Net neural network. Thus we avoid interpolating these fields to match the U-Net input. Note that a training sample here consists of - (i) the inputs as u_h, I_h and a RSTF spanned throughout the input space of 128^3, and (ii) the outputs as the velocity and the turbulent intensity fields of size 128^3.

- The training, the validation and the test set of the 3D U-Net neural network consist of 1000, 200 and 100 samples, respectively. We implement the network in *TensorFlow* [1] and train for 1000 epochs (iterations) using cross validation. We use the *Adam* optimizer [55] with a mean absolute error (MAE) loss function and an exponential decay schedule for the leaning rate initialized with a value of $1e-3$. We also monitor the field averaged relative L1 error between the true and the predicted fields.

- The training and validation history of the U-Net model trained over $k-\omega$ SST RANS simulations data of a single wind turbine using the steps mentioned above is shown in Figure 4.7 (a). As observed, the convergence is obtained after 400 epochs, with the field-averaged relative L1 error dropping to about 0.96% . With the use of the exponential decay schedule for the learning rate the loss (and the L1 error) reduces steadily within the first 250 epochs. The validation loss is only slightly higher than the training loss. In Figure 4.7 (b) we present the regression plots for the training, the validation and the testing set of the surrogate model. The high correlation between the predicted solution and the true values indicate that the surrogate model has good convergence and properly represents (fits) the dataset. To assess the performance of the surrogate model trained over $k-\omega$ SST RANS simulations, the baseline true and predicted solutions are presented in Figure 4.8. Several samples from the test set are shown in Figure 4.20. The surrogate model has high accuracy for the samples close to baseline and only suffers slightly in a few edge cases (far away from baseline). The discrepancies in the wake region (< 1.5D) over the entire test set has a maximum local error of about 5.3% and 3.1% for velocity and turbulent intensity, respectively. This can be associated to the steep pressure gradient across the actuator disk which disturbs the inflow. Overall, the surrogate severs as a reasonable approximator of the QoIs and is deemed suitable for the surrogate based uncertainty quantification approach.

- To obtain the statistics (mean and variance) of the QoIs with a reasonable accuracy we use 10,000 QMC samples of inputs. Note that we now use the surrogate constructed in the above steps for fast and accurate predictions of the QoIs for such large number of samples.

Similar to the $k-\omega$ SST model, we also train the 3D U-Net model over the data from LRR RANS simulations to predict the QoIs and in turn their statistics. What follows is a comparison between the solutions obtained from the stochastic RANS solver and the surrogate based approach for both the turbulence models. Figure 4.9 shows the normalized velocity deficit and the turbulent intensity averaged over the rotor area along the streamwise direction. As observed in the plots and as expected, the stochastic mean for both the turbulence models is in accordance with the reference LES solution. The uncertain bound of $\pm 2\times$ standard deviations reflects the effect of the overall input uncertainties on the QoIs.

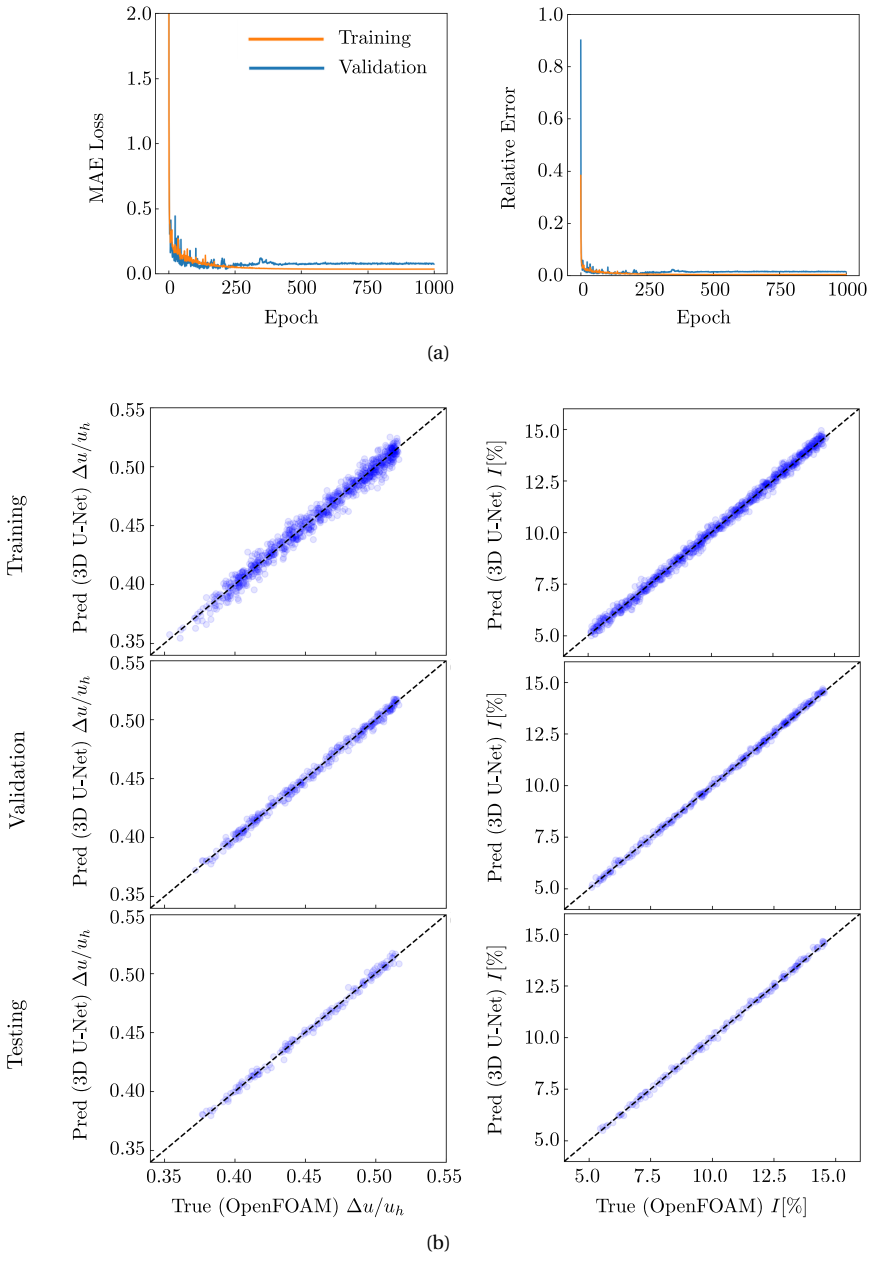


Figure 4.7 | (a) History of the training and validation losses (left) and relative errors (right) over epochs for the 3D U-Net surrogate model trained with the data from a single wind turbine simulations using the $k-\omega$ SST RANS turbulence model, and (b) Predicted versus true values of velocity deficit and turbulent intensity $2D$ distance downstream the center of the rotor, plotted for training, validation and testing sets of the surrogate model.

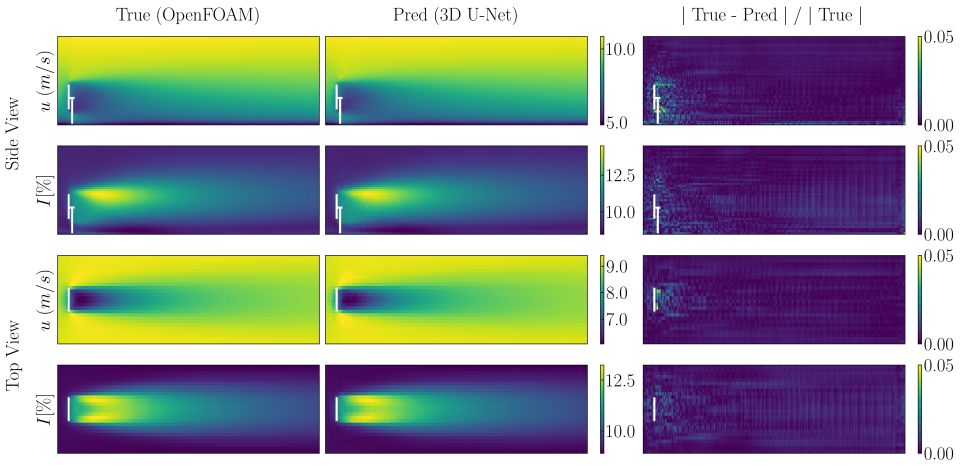
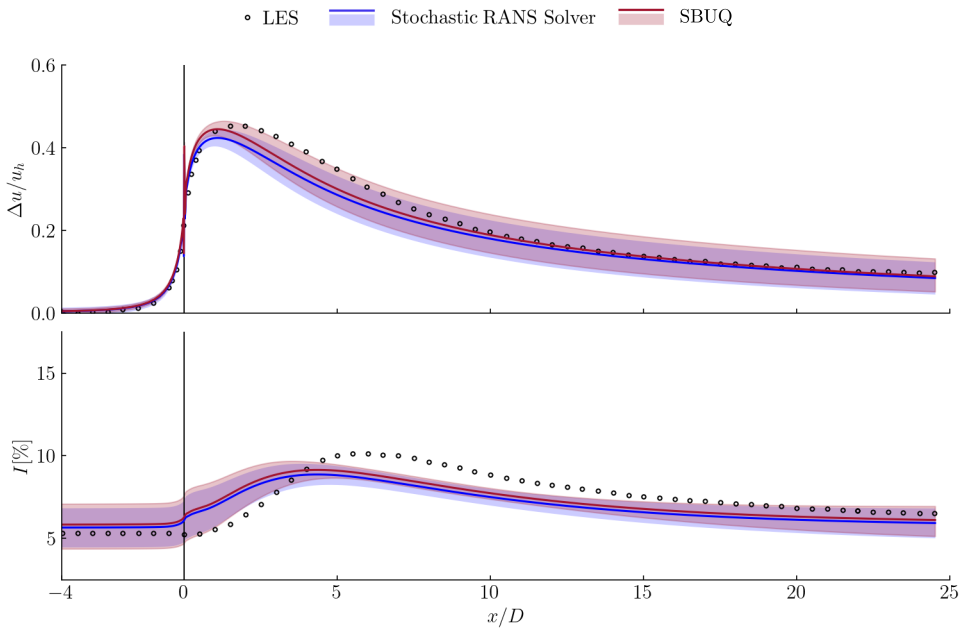


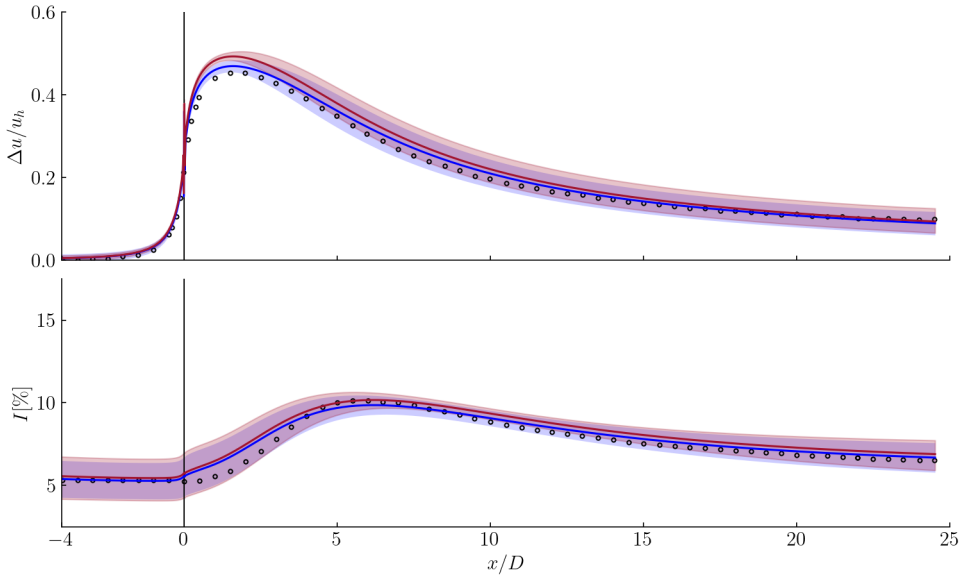
Figure 4.8 | Comparison of the true (OpenFOAM) solution and the predicted solution for baseline using the 3D U-Net surrogate model trained with the data from a single wind turbine simulations using the $k-\omega$ SST RANS turbulence model.

For the $k-\omega$ SST model these bounds contain the LES solution except for the near wake region. This can be associated with the inherent shortcomings of the linear eddy viscosity model as well as with the use of actuator disk to model the effect of turbine blades. The uncertain bounds of the LRR model captures the LES solution throughout the streamwise direction. Figure 4.10 shows the normalized velocity deficit and turbulence intensity at four different downstream locations. As observed again, the stochastic mean for both turbulence models is in conformity with the reference LES solution. Considering both RANS models, the velocity deficit - (i) shows only very small effect of the input uncertainties in the near wake region ($x/D = 2$), while the variance in velocity deficit grows gradually and peaks after $x/D = 4$, (ii) bounds are able to contain the LES solution except for the near wake region which can again be associated with the inherent shortcomings of the linear eddy viscosity model and the use of actuator disk model. The $k-\omega$ SST model under-predicts the turbulent intensity, although its bounds capture the LES solution at least in the near wake region. On the other hand, the LRR model shows a good prediction of the turbulent intensity and its bounds capture the LES profile throughout all the downstream location. In both the results discussed above, the mean profiles obtained from both uncertainty quantification approaches are mostly close to each other, with a slightly overpredicted velocity deficit in case of the surrogate based approach in the wake region. The variance in the solution, especially in the wake regions is overpredicted by the surrogate based method which may be attributed to the noise in the training data and/or the inherent uncertainty of the neural network.

Thus, from the above discussion, it is clear that in the near wake region, due to the adverse pressure-gradient resulting from the turbine, the linear eddy viscosity model fails to perform as good as RSTM. Overall, the three input uncertainties (u_h, I_h, RSTF) collectively result in a slightly higher variance in the stochastic solution ($\Delta u/u_h, I$) of the two-equation model as compared to RSTM. This makes the latter, although computationally expensive,

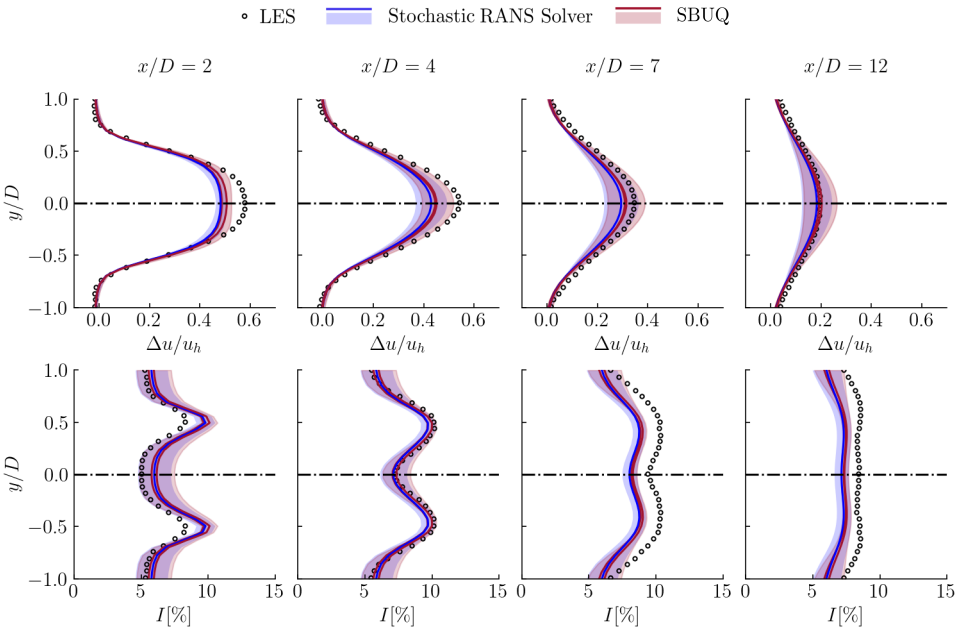


(a)

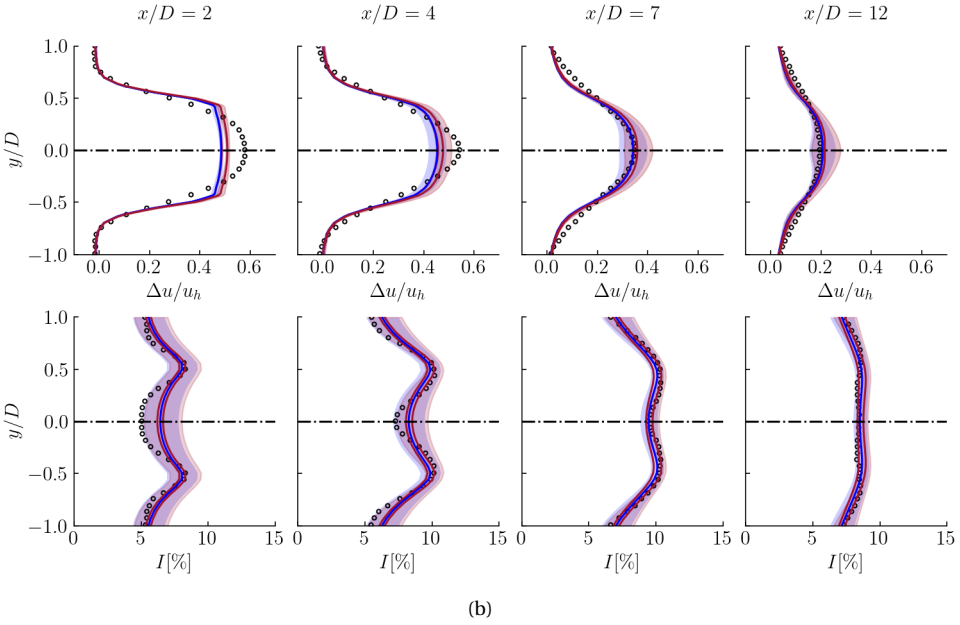


(b)

Figure 4.9 | Mean and variance of the normalized velocity deficit and turbulence intensity, in streamwise direction and averaged over the rotor area for the uncertain (a) $k-\omega$ SST, and (b) LRR RANS turbulence models. The results from the stochastic RANS solver (blue) are compared with those from the SBUQ approach (red). The line represents the mean profile and the shaded area represents the bound of ±2× standard deviations.



(a)



(b)

Figure 4.10 | Mean and variance of the normalized velocity deficit and turbulence intensity, at different downstream locations for the uncertain (a) $k-\omega$ SST, and (b) LRR RANS turbulence models. The results from the stochastic RANS solver (blue) are compared with those from the SBUQ approach (red). The line represents the mean profile and the shaded area represents the bound of $\pm 2\times$ standard deviations.

more accurate and slightly less sensitive to the input uncertainties. However, the linear eddy viscosity model, apart from its shortcomings in some regions of the flow, performs reasonably well in matching the LES solution. We also gather that the surrogate model developed using the RANS simulations over a single wind turbine may not only be used to accurately predict the QoIs but also be considered as an efficient alternate to the full-size stochastic RANS solver for uncertainty quantification analysis.

4.4.2. WAKE INTERACTIONS FOR AN ARRAY OF WIND TURBINES

SIMULATION SETUP

For the deterministic RANS simulation of the interactions of the wakes in an array of wind turbines, we consider a single row of six Vestas V80-2MW turbines. Note that such an arrangement is usually part of a large wind farm (e.g. the first six wind turbines of a row in the Horns Rev windfarm ($\theta_{wind} = 270°$) [131] which consists of 8×10 Vestas wind turbines). Each of these turbines has a hub height z_h of $70m$, a turbine diameter D of $80m$ and a thrust coefficient C_T of 0.8. As can be seen in Figure 4.11, the size of the computational domain is $65D \times 10D \times 4.3D$ in the streamwise (x), spanwise (y) and vertical (z) directions, respectively. The first wind turbine is location $5D$ downstream of the inlet and the spacing between the wind turbines is $7D$. Similar to the single wind turbine case - (i) a Cartesian mesh with higher cell density near the wind turbine in streamwise and spanwise directions is used to capture steep gradients in the flow, and (ii) for each wind turbine, an actuator disk approach is used to model the effect of turbine blades (see equation 4.10), and (iii) the boundary conditions are set to match the atmospheric boundary layer profile at inlet ($u_h = 8m/s, I_h = 5.8\%$), pressure outlet and zero gradient for all other variables at the outlet, rough ground using no-slip condition and wall-functions at the bottom surface, and farfield conditions using zero gradient for all variables in the spanwise direction. Based on the analysis performed using the different linear-eddy-viscosity RANS models in [20], we choose the realizable $k-\epsilon$ model [78, 126] to simulate the wake interaction for an array of wind turbines. Similar to the wake prediction of a single wind turbine, we also use the LRR Reynold stress transport model [64] to make comparisons with the two-equation model and to verify the expected better performance of RSTM.

The stochastic simulation uses the same mesh as the deterministic simulation. The parameters used to characterize the random hub height variables (u_h, I_h) at the first wind turbine and the random Reynolds stress tensor field (RRSTF) for the entire domain are essentially the same as discussed in section 4.4.1 and are presented in Table 4.1, with an exception of I_h which is now set to have a uniform distribution with bounds as [5.3%, 6.3%], allowing the mean to be 5.8% as required.

SIMULATION RESULTS

In this section, similar to the single wind turbine simulation analysis (section 4.4.1), we present the uncertainty quantification results obtained using both the approaches discussed in section 4.3. As earlier, we consider the normalized velocity deficit and turbulent intensity fields as the QoIs. We compare the results obtained from the analysis of both turbulence models with reference LES data from [20]. Note that the LES corresponds to the mean of hub height variables in the present work. The LES uses a computational domain of

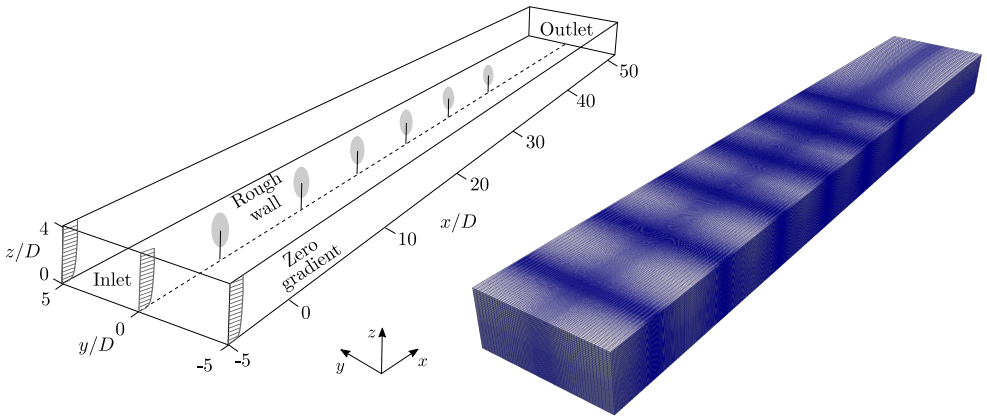


Figure 4.11 | Computational domain for the simulation of an array of wind turbines. The first turbine is situated at $5D$ distance from the inlet and the spacing between the wind turbines is $7D$. The outlet is at $60D$ from the first turbine (left). The computational mesh has fine cells across the actuator disks to capture the momentum and pressure jump, and smaller cells near the rough wall to capture the steep gradients (right).

size $60D \times 10D \times 4.44D$ including two columns of six wind turbines each with the lateral spacing of $5D$ and streamwise spacing of $7D$ (same as RANS in present work). The LES results averaged over the two columns are considered for comparison with RANS. For further details of the LES see[2, 140].

Following the discussion of the SBUQ approach in section 4.3.2, we now discuss the training, the validation and the test results obtained by combining the deep learning surrogate with the wake superposition model to predict the QoIs in an array of wind turbines.

- We aim to train a U-Net model over single wind turbine simulation data and use it along with wake-superposition model to predict the flow field in an array of wind turbines. This requires training the U-Net model for a range of possible (scalar or field) values that the hub height variables and the RSTF may attain around the turbines. For this purpose we use the baseline (deterministic) RANS simulation for an array of wind turbines and extract the hub height values of velocity and turbulent intensity, and the RSTF for each wind turbine. Thereafter, the random variables $(u_h,\ I_h)$ are characterized using the same distribution type as listed in Table 4.1, with the bounds determined from the their extracted values. Similarly, using the extracted RSTFs as the realizations of a random field, the RRSTF is constructed and characterized by the set of hyper-parameters listed in Table 4.1. Note that the mesh size in the region of the extracted field is already 128^3 which is the same as the input size of the U-Net neural network. Thus we avoid interpolating these fields to match the size of the U-Net input.

- Similar to the single wind turbine analysis in section 4.4.1, we use a Quasi-Monte Carlo (QMC) sampling technique [45, 68] to sample 1300 independent samples of $\{u_h, I_h, \text{RRSTF}\}$. For each sample we use a CFD solver to obtain the velocity and the turbulent intensity fields. We defer the discussion of the forward CFD solver used to Appendix 4.6.4. Note that a training sample consists of - (i) the inputs as u_h, I_h and a

RSTF spanned throughout the input space of 128^3, and (ii) the outputs as the velocity and the turbulent intensity fields of size 128^3.

- The training, the validation and the test set of the 3D U-Net neural network consists of 1000, 200 and 100 samples respectively. We implement the network in *TensorFlow* [1] and train for 1000 epochs (iterations) using cross validation. We use the *Adam* optimizer [55] with a mean absolute error (MAE) loss function and an exponential decay schedule for the leaning rate initialized with a value of $1e-3$. We also monitor the field averaged relative L1 error between the true and the predicted fields.

- The training and validation history is shown in Figure 4.12 (a). As observed, the convergence is obtained after 750 epochs, with the field-averaged relative L1 error dropping to about 1.8%. With the use of the exponential decay schedule for the learning rate the loss (and the L1 error) reduces steadily within the first 500 epochs. As expected, the validation loss is higher than the training loss. In Figure 4.12 (b) we present the regression plots for the training, the validation and the testing set of the surrogate model. The apparent correlation between the predicted and the true values imply that the surrogate model has acceptable convergence and represents (fits) the dataset with reasonable accuracy. Note that in comparison to the U-Net training in section 4.4.1, the number of the epochs required to attain convergence is larger and the field-averaged error is significantly higher, while the correlation between predicted and true values are slightly lower. This is associated to the larger bounds and/or variance of hub height variables, the RSTF and the QoIs used to train the current U-Net model. These larger bounds and/or variance in the random variables and fields are in turn directly associated to the uncertainty characterization procedure used in the first step. To assess the performance of the surrogate model trained over realizable $k-\epsilon$ RANS simulations, we present a test case using the inflow conditions and RSTF for the first wind turbine, see Figure 4.13. Several samples from the test set are shown in Figure 4.21. The surrogate model is accurate for the samples with higher levels of turbulent intensity, whereas for lower turbulent intensities the predictions are more erroneous. The discrepancies in the wake region ($< 1.5D$) over the entire test set has a maximum local error of about 6.8% and 4.7% for velocity and turbulent intensity, respectively. This can be associated to the steep pressure gradient across the actuator disk that disturbs the inflow as well as with the high variance in the training data as discussed above. Overall, the surrogate severs as a reasonable approximator of the QoIs and is deemed suitable for the SBUQ approach.

- After training the 3D U-Net surrogate model for a single wind turbine predictions, we now use the neural network in conjugation with a wake superposition model (section 4.3.2) to sequentially predict the wake fields in an array of wind turbines. To sample the inputs, we use the same random space $\{u_h, I_h, \xi_1, ..., \xi_{N_{KL}}\}$ as was created during the uncertainty quantification analysis using the stochastic solver approach. To assess the quality of the coupled surrogate model the baseline case with the true and predicted solutions is presented in Figure 4.14 while the test samples presented in Figure 4.22. As can be observed, the wake regions of all wind turbines have moderate discrepancies with a significantly high error near the first wind turbine. The maximum field-averaged relative L1 error in velocity and turbulent intensity over the entire test

set is about 4.8% and 3.7%, respectively. However, in the wake regions, for some test cases the maximum local error for velocity and the turbulent intensity are as high as 12% and 9%, respectively. Such high local errors can be correlated to the empirical wake superposition model which may admit obvious errors in the wake predictions. Nevertheless, the hub height velocity and turbulent intensity predictions are not significantly affected by these larger error regions as most of these discrepancies are concentrated in the wake region up to $3-4D$ distance downstream of the wind turbines. Overall, the predicted solution is in accordance with the true solution while having a reasonable accuracy. The SBUQ approach (3D U-Net + wake superposition) is therefore deemed suitable for uncertainty quantification in an array of wind turbines.

- To obtain the statistics (mean and variance) of the QoIs with a reasonable accuracy we use 10,000 QMC samples. Note that we now use the surrogate constructed in the above steps (3D U-Net + wake superposition) for fast and accurate predictions of the QoIs for such large number of samples.

Similar to the realizable $k-\epsilon$ model, we also train 3D U-Net model over the data from LRR RANS simulations and in turn combine it with the wake superposition model to predict the QoIs and perform uncertainty quantification analysis in an array of wind turbines. What follows is a comparison between the solutions obtained from the stochastic RANS solver and the surrogate based approach for both the turbulence models. The streamwise variation in the normalized velocity deficit and turbulence intensity averaged over the rotor area is shown in Figure 4.15. As evident, for both turbulence models, the stochastic mean is very close to the deterministic simulation and is in accordance with the LES solution. Due to the inadequacies of the linear-eddy-viscosity models, the two-equation model struggles to such an extent that even its bounds are not able to capture the LES solution in the near wake regions. On the other hand, RSTM shows a very close prediction to LES except for the velocity deficit in the wake region of the second wind turbine. For the realizable $k-\epsilon$ model, the variance in the wake regions increases slightly over the wind turbines. While for the LRR model, the variance in the wake regions grows significantly in the streamwise direction. In the single wind turbine study (Figure 4.9), we observed that the variance was larger in regions of higher turbulent intensity and gradually decreased (or plateaued) downstream as the wake recovered. In contrast, for an array of wind turbines, the wake region for every wind turbine consists of high shear and turbulence levels, resulting in a larger variance. This observation is also consistent with the findings in [20]. In Figure 4.16, we present the hub-height spanwise profiles of the normalized velocity deficit and turbulence intensity at $5D$ downstream of each wind turbine. Except behind the first wind turbine, both the turbulence models are able to capture the LES solution fairly well. The variance in the realizable $k-\epsilon$ solution is almost the same behind each wind turbine, while the variance in the LRR solution grows not only downstream but also laterally. This might be associated with the higher levels of anisotropy inherent in the RSTM, which on perturbation and propagation, results in a higher overall variance in the QoIs. In the results discussed above, the mean profiles obtained from both uncertainty quantification approaches are very close, with a slightly overpredicted turbulent intensity in case of the surrogate based approach. Similar to the single wind turbine study (section 4.4.1), the variance in the solution, especially in the wake regions is overpredicted by the surrogate based method. As mentioned earlier, the added variance may be attributed to the noise in the training data and/or the inherent uncertainty

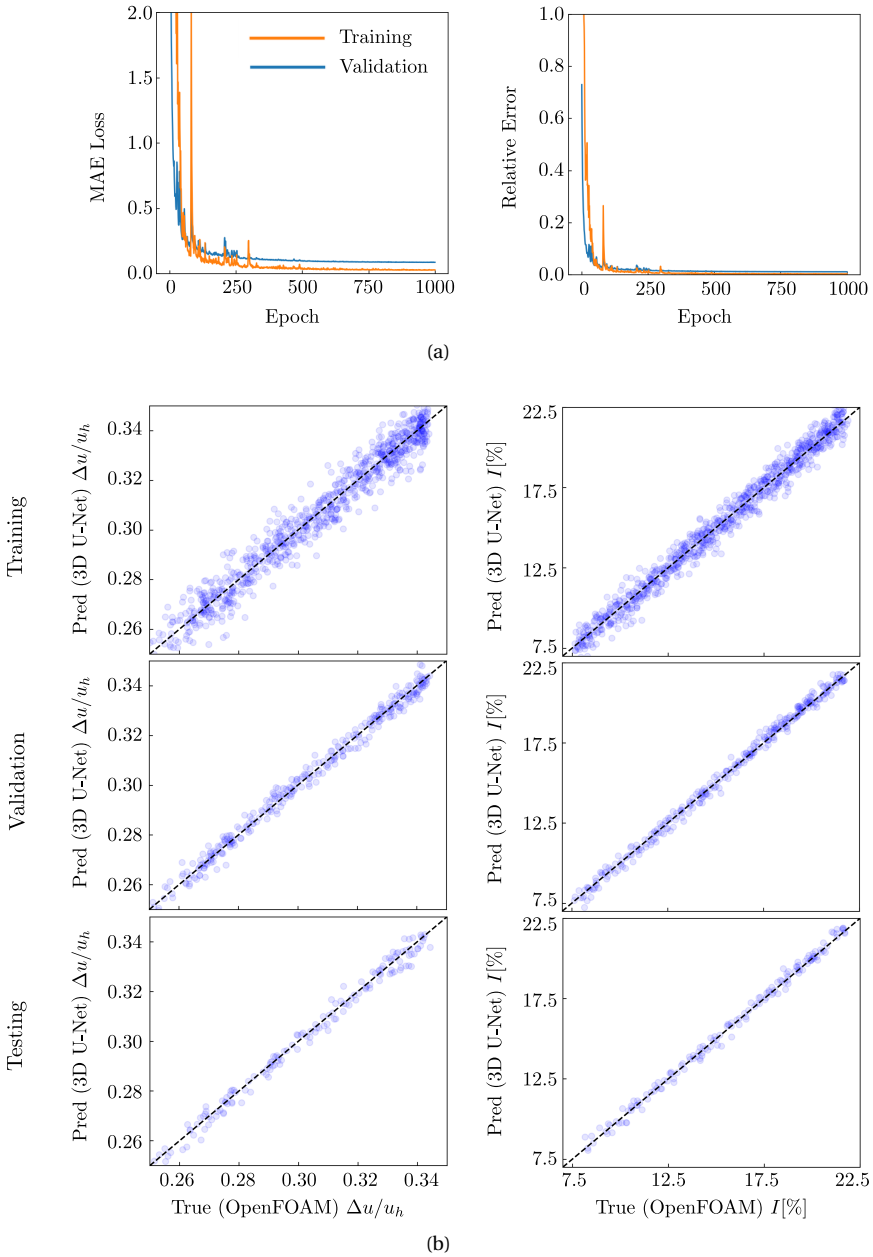


Figure 4.12 | (a) History of the training and validation losses (left) and relative errors (right) over epochs for the 3D U-Net surrogate model trained with the data from a single wind turbine simulations using the realizable $k-\epsilon$ RANS turbulence model, and (b) Predicted versus true values of velocity deficit and turbulent intensity at $2D$ distance downstream the center of the rotor, plotted for training, validation and testing sets of the surrogate model.

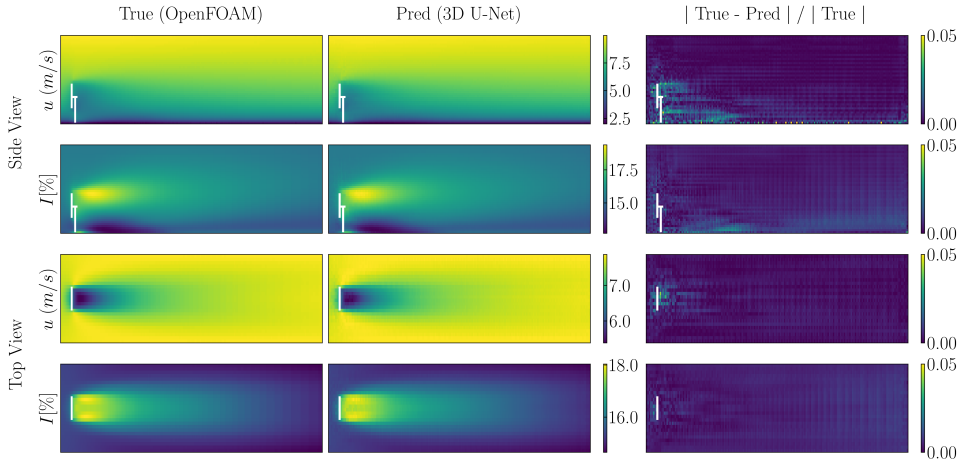


Figure 4.13 | Comparison of the true (OpenFOAM) solution and the predicted solution for baseline using the 3D U-Net surrogate model trained with the data from a single wind turbine simulations using the realizable $k-\epsilon$ RANS turbulence model. The baseline here represents the configuration with the inflow conditions and RSTF for the first wind turbine in baseline.

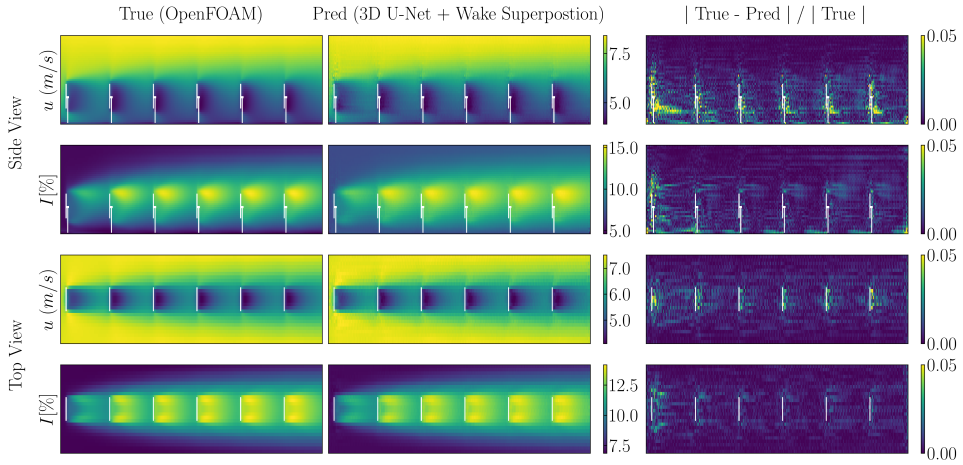


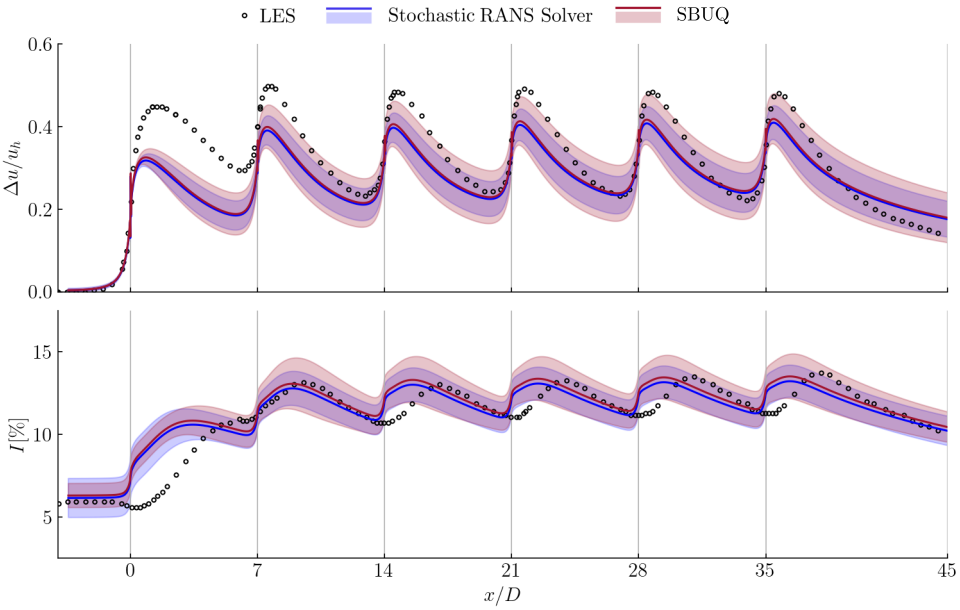
Figure 4.14 | Comparison of the true (OpenFOAM) solution and the solution predicted using the SBUQ surrogate (3D U-Net trained with the data from a single wind turbine simulations using the realizable $k-\epsilon$ RANS turbulence model + wake superposition) for a test sample.

of the neural network. The discrepancy w.r.t. the LES solution can be associated with the inherent shortcomings of the linear-eddy-viscosity model and with the use of actuator disk model as discussed earlier. As before in the single wind turbine study (section 4.4.1), we also assert that the 3D U-Net model trained using the RANS simulations over a single wind turbine and coupled with the wake superposition model may not only be used to accurately predict the QoIs but also be considered as an efficient alternate to the full-size stochastic RANS solver for uncertainty quantification analysis for an array of wind turbines.

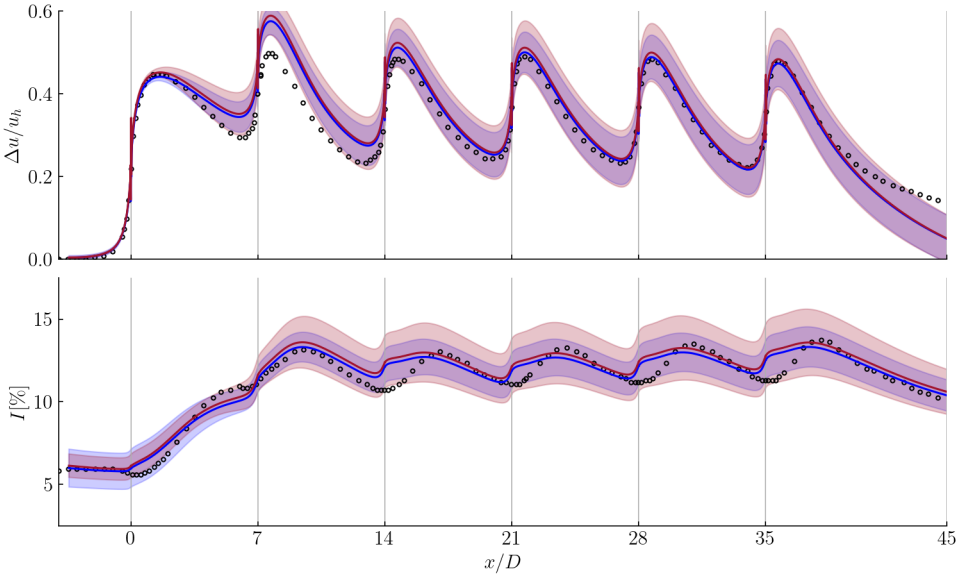
The Gaussian random field used in the characterization of the RRSTF was discretized using a finite set of random variables $\boldsymbol{\xi}$ obtained from its Karhunen-Loève decomposition (see section 4.3.1). Including u_h and I_h, the overall input random vector comprises of $N_{KL}+2$ random variables $\{u_h, I_h, \xi_1, ..., \xi_{KL}\}$. In Figure 4.17 we present the Sobol indices for the normalized velocity deficit and the turbulent intensity at hub height and $5D$ distance downstream of each wind turbine. Sobol indices quantify the relative contribution of the input uncertainties towards the variance in the QoIs. The details on the computation of these indices can be found in Appendix 4.6.3. Note that the higher modes of RRSTF have insignificant contribution, and thus for the sake of clarity, we present only the contributions from the first eight modes of the RRSTF. In the Figure 4.17, the mode index 1 and 2 refer to the random variables u_h and I_h, respectively, while the remaining eight indices refer random variables representing the RRSTF. The overall contribution from the RRSTF is considered to be the sum of the contributions of all RRSTF modes. As can be observed, the normalized velocity deficit and the turbulent intensity are almost insensitive to the uncertainty in the hub height velocity. The contributions of the uncertain hub height turbulent intensity and RRSTF are significant and comparable. In the variance of the normalized velocity deficit, the contribution of the uncertain hub height turbulent intensity gradually decreases, while the contribution of the RRSTF slowly increases in the streamwise direction. On the other hand, in the variance of the turbulent intensity, the contribution of the uncertain hub height turbulent intensity dominates for the first wind turbine, drops significantly for the second wind turbine and remains almost constant for the remaining wind turbines downstream. Here, the contribution of the RRSTF varies accordingly and is least for the first wind turbine. Note that among the all RRSTF modes, the contribution from the first two modes is always the largest.

In Figure 4.18, we present the normalized power output for each wind turbine with a probability distribution based on statistics obtained from both the approaches. As can be seen, the mean power for both approaches is in agreement with the LES and the LRR model, with the exception of the second and the sixth wind turbine. However, the LES solution at all turbines is captured by the high probability region of the power probability distributions. Overall, the surrogate based approach shows a reasonable agreement with the stochastic solver and can be considered an alternate technique to the intrusive method when an accurate surrogate model can be developed.

Table 4.2 lists the comparison between the two uncertainty quantification approaches present above. The accuracy of the stochastic RANS solver is generally comparable to traditional methods like Monte-Carlo or its variants like Multi-level Monte-Carlo [60]. However, the accuracy may also depend on the variation in hyper-parameters as shown in [86]. The SBUQ approach is reasonably accurate (as discussed previously) although it may be affected by the inherent uncertainties in the neural network which shall be quantified. The computational cost for the uncertainty quantification analysis of the wake interactions in an array of



(a)



(b)

Figure 4.15 | Mean and variance of the normalized velocity deficit and turbulence intensity, in streamwise direction and averaged over the rotor area for the uncertain (a) realizable $k-\epsilon$, and (b) LRR RANS turbulence models. The results from the stochastic RANS solver (blue) are compared with those from the SBUQ approach (red). The line represents the mean profile and the shaded area represents the bound of $\pm 2\times$ standard deviations.

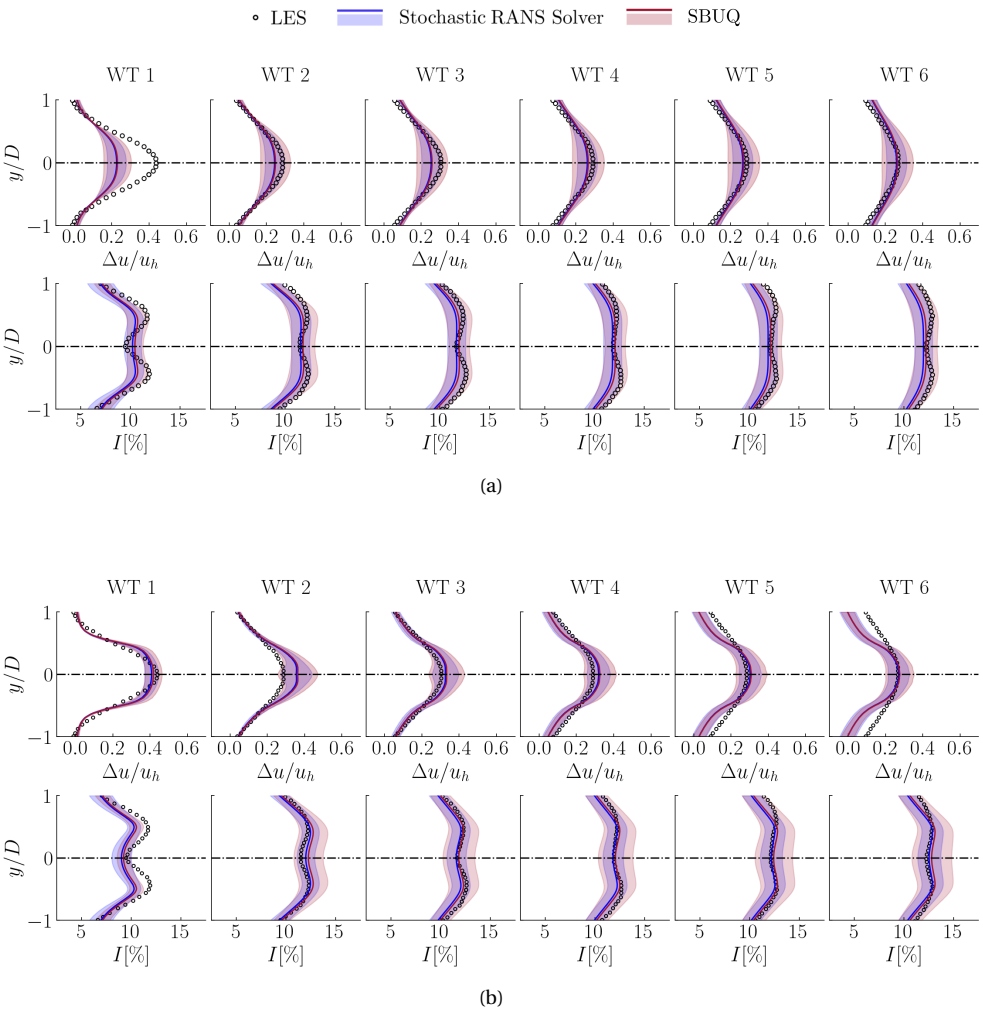


Figure 4.16 | Mean and variance of the normalized velocity deficit and the turbulence intensity, at $5D$ behind each wind turbine (WT) for the uncertain (a) realizable $k-\epsilon$, and (b) LRR RANS turbulence models. The results from the stochastic RANS solver (blue) are compared with those from the SBUQ approach (red). The line represents the mean profile and the shaded area represents the bound of $\pm 2\times$ standard deviations.

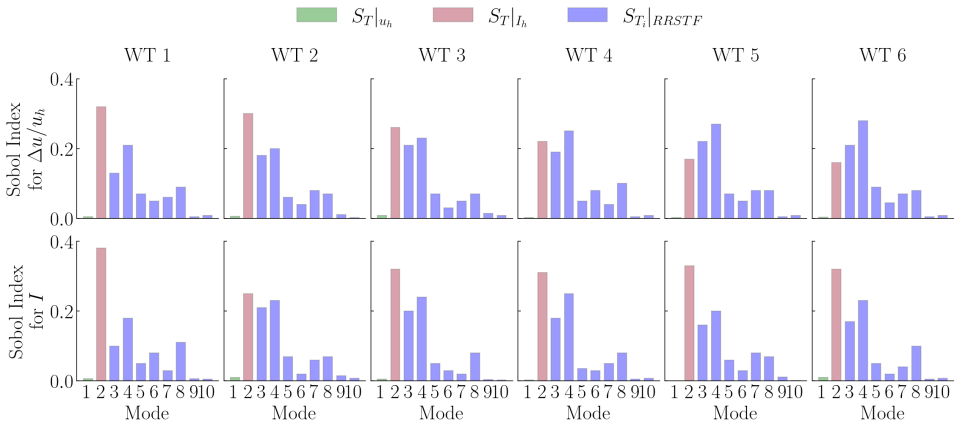


Figure 4.17 | Sobol indices of all the random inputs $\{u_h, I_h, \xi_1, ..., \xi_{KL}\}$ for the normalized velocity deficit and the turbulence intensity at hub height $5D$ behind each wind turbine (WT) for the uncertain realizable $k-\epsilon$ RANS turbulence model. The higher modes of RRSTF have insignificant contribution, and thus, only the contributions from the first eight modes of the RRSTF are shown.

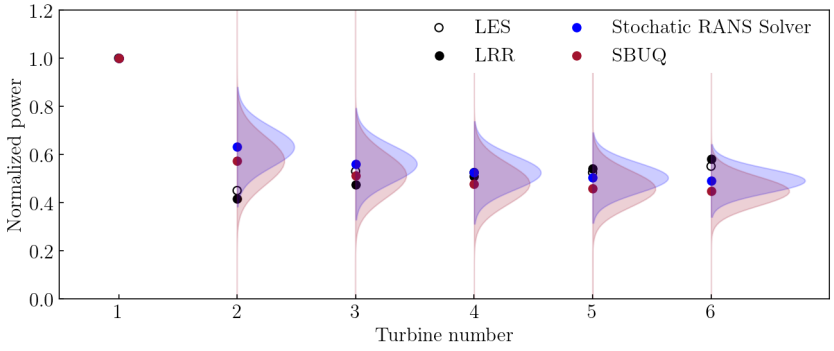


Figure 4.18 | Normalized power distribution at each wind turbine, for the uncertain realizable $k-\epsilon$ RANS turbulence model. The results from the stochastic RANS solver (blue) are compared with those from the SBUQ approach (red). LES [20] and LRR data are used as a reference to asses the accuracy of the baseline RANS solution.

	Stochastic RANS solver	SBUQ
Accuracy	Comparable to traditional (expensive) methods. May also depend on hyper-parameters [86].	Reasonably accurate but may be affected by the inherent uncertainty of the neural network.
Computational cost (× 1 single WT sim.)		
PCE of RRSTF	0.5	0.5
Simulation	630	-
Training	-	1000
Statistics	-	1
Overall comment	Relatively less expensive as compared to sampling-based methods [60, 86].	Sampling and computing statistics using surrogate is significantly cheaper.
Convergence	Slow convergence of higher PCE modes.	Slow convergence for samples further from mean of inputs.
General applicability	Applicable to a wide range of engineering applications.	Restricted to use in closely related turbulent flows.
Possible improvements	An implicit algorithm to simultaneously solve for all modes.	Use of dimensionality reduction techniques for RSTF.

Table 4.2 | Overall comparison of the Stochastic RANS solver and the SBUQ approach. The figures for the computational cost are specifically for the uncertainty quantification analysis of the wake interactions in an array of wind turbines.

wind turbines is around 40% lower for the stochastic solver. Using early-truncation techniques the computational cost can be further reduced [86]. We observed slower convergence rates for the higher modes in the simulations using the stochastic solver. On the other hand, in the training phase of U-Net, slower convergence rates were observed for the samples in the regions away from the input mean. The stochastic RANS solver has capabilities to accommodate for various sources of uncertainties (parametric, model-form, operating or boundary conditions etc.) and can be employed for the uncertainty quantification analysis of numerous engineering simulations with significant input uncertainties. Whereas, the SBUQ approach is restricted in the sense that it can only be used for similar (or related) flow fields (as seen in the current study). Since the U-Net model only relies on data, it is very likely to predict non-physical solutions for inputs outside the training data-set. In terms of possible improvements, the stochastic solver may use an implicit algorithm in order to simultaneously solve all PCE modes. While the SBUQ approach can make use of dimensionality reduction techniques such as auto-encoders [5] to avail a low dimensional representation of the Reynolds-stress tensor field.

4.5. CONCLUSIONS

In this work we present two different approaches for the uncertainty quantification and propagation in RANS turbulent flow simulations. Particularly, we focus on the RANS simulations of - (i) a wake behind a wind turbine, and (ii) wake interactions and power losses in an array of wind turbines, acting under the model-form and the operational uncertainties.

The first approach for the characterization and the forward propagation of uncertainties

is based on the previously developed stochastic RANS solver that is based on an intrusive polynomial chaos method [86]. The second approach, called the surrogate-based uncertainty quantification (SBUQ), is based on a 3D U-Net deep learning model (trained over a single wind turbine) in conjugation with a wake superposition model, to sequentially predict the flow field for each wind turbine in an array of wind turbines.

We consider different RANS turbulence models namely, the $k-\omega$ SST and the realizable $k-\epsilon$ linear-eddy viscosity models [78, 126] as well as the LRR Reynolds-stress transport model [64]. The results from the uncertainty analysis of the above mentioned models are compare with reference (high-fidelity) LES data at mean values of the input random variables/fields [20, 43].

The uncertain bounds of the linear-eddy viscosity RANS models are able to fairly capture the high-fidelity solution except in regions very close to or in the wake immediately behind the wind turbine. The uncertain bounds of LRR model on the other hand, are able to precisely capture the LES solution over the wind turbines asserting its overall higher accuracy. However, it is also observed that the variance in the solutions from two-equation models remain nearly uniform in the streamwise direction, whereas it grows steadily for the Reynolds-stress transport model.

In contrast to the previous wind turbine/farm studies based only on the eigenspace perturbation of the Reynolds stress tensor [20, 43], in this work, we realize the importance of the effect of randomness in the operating conditions (via inlet boundary layer) and its interaction with the uncertain Reynolds stress tensor. The sensitivity analysis (in terms of Sobol indices) for similar levels of coefficient of variation in the inputs (u_h, I_h, RSTF) reveal that the contribution of inlet conditions towards the variance in the QoIs is as significant as that from the Reynolds stress tensor. In particular, the normalized velocity deficit and the turbulent intensity fields are insensitive to variation in hub-height velocity (u_h) and are highly dependent on the variation in the hub-height turbulent intensity (I_h) and the Reynolds-stress tensor field (RSTF).

The surrogate constructed using the U-Net and the wake-superposition model is able to predict the flow field in an array of wind turbines with high accuracy and significantly low (prediction) cost. Thus it is considered suitable for a surrogate-based uncertainty quantification (SBUQ) approach. The results (mostly statistics) obtained from the SBUQ approach are in accordance with the results from the stochastic solver. Furthermore, in Table 4.2, we discuss the overall comparison of both approaches in detail. Based on the engineering application, one approach will have more advantages over the other and it will be crucial to decide between the two. Nevertheless, they both have demonstrated the potential to be applied for the analysis of a RANS turbulent flow simulation under various uncertainties.

In the lights of the above findings, it would be of great interest for the turbulence modeling community to develop strategies for stochastic RANS simulation under multiple sources of uncertainties. Future research aspects may include the implementation of an implicit algorithm to simultaneously solve all PCE modes in the stochastic RANS solver, lower-dimension representation of RSTF using dimensionality reduction techniques and uncertainty quantification analysis for more complex unsteady RANS (URANS) or LES engineering simulations.

4.6. APPENDIX

4.6.1. PROJECTING REYNOLDS STRESS ON A BARYCENTRIC TRIANGLE

The Reynolds stress tensor can be decomposed into two parts, an isotropic component $2/3k\delta_{ij}$ and an anisotropic tensor which is given as the deviatoric part of the Reynolds stress normalized by the turbulent kinetic energy:

$$A_{ij} := \frac{R_{ij}}{2k} - \frac{\delta_{ij}}{3} \tag{4.11}$$

The eigen-decomposition of the anisotropy tensor results into an eigenvalue tensor,

$$\mathrm{A} = \mathrm{V}\mathbf{\Lambda}\mathrm{V}^T, \tag{4.12}$$

where V is the matrix of orthogonal eigenvectors $[v_1, v_2, v_3]$ and $\mathbf{\Lambda}$ is the corresponding eigenvalue matrix $\mathrm{diag}[\lambda_1, \lambda_2, \lambda_3]$, such that $\mathrm{tr}[\mathbf{\Lambda}] = 0$ and $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Using linear relations, these eigenvalues can be mapped to unique coordinates (C_{1c}, C_{2c}, C_{3c}) in a Barycentric triangle, which in turn represents the state of anisotropy in a turbulent flow at a given point:

$$C_{1c} = \lambda_1 - \lambda_2, \quad C_{2c} = 2(\lambda_2 - \lambda_3), \quad C_{3c} = 3\lambda_3 + 1. \tag{4.13}$$

Note that, $C_{1c} + C_{2c} + C_{3c} = 1$. If any of these coordinates equals one, the anisotropy is said to achieve a limiting state. In a Barycentric map, these limiting states are represented by the three vertices of an equilateral triangle. Thus, C_{1c}, C_{2c} and C_{3c} represents 1-component, 2-component and 3-component turbulence, respectively. In Cartesian coordinates, if the location of these vertices is expressed as $(x_{1c}, y_{1c}), (x_{2c}, y_{2c})$ and (x_{3c}, y_{3c}), then any arbitrary turbulent state (x, y) can be represented by the following linear combination of the limiting states:

$$x = C_{1c}x_{1c} + C_{2c}x_{2c} + C_{3c}x_{3c}, \quad y = C_{1c}y_{1c} + C_{2c}y_{2c} + C_{3c}y_{3c}. \tag{4.14}$$

All the realizable states lie within or on this triangle. The anisotropy state of the samples generated using the polynomial chaos expansion of the Reynolds stress tensor can therefore be represented and analyzed using the Barycentric triangle as shown in Figure 4.4.

4.6.2. UNCERTAINTY PROPAGATION OF RANDOM INLET CONDITIONS

Figure 4.19 shows the normalized velocity deficit and the turbulent intensity under the uncertain inlet conditions, averaged over the rotor area along the streamwise direction for a wake behind a wind turbine simulated using the $k-\omega$ SST RANS model. The input random variables are listed in Table 4.1. As observed in the plots and as expected, the stochastic mean is in accordance with the reference LES solution. The uncertain bound of $\pm 2\times$ standard deviations reflects the effect of the uncertain inlet conditions on the QoIs. These bounds contain the LES solution except for the near wake region. This can be associated with the inherent shortcomings of the linear eddy viscosity model and with the use of actuator disk to model the effect of turbine blades. The results clearly indicate a significant contribution of the uncertain operating parameters towards the variance in the QoIs.

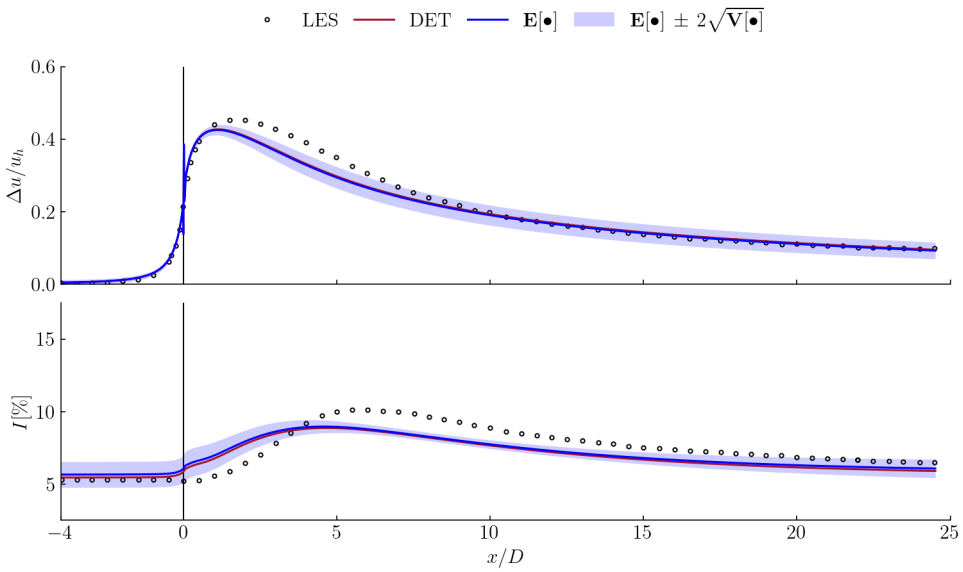


Figure 4.19 | Mean and variance of the normalized velocity deficit and turbulence intensity for the uncertain inlet operating conditions using $k-\omega$ SST at different downstream locations.

4.6.3. GLOBAL SENSITIVITY ANALYSIS WITH SOBOL INDICES

In order to determine the relative influence of each random variable on the QoIs we can employ a global sensitivity analysis based on a variance-based approach called Sobol indices method [106]. Once the coefficients of the polynomial chaos expansion are determined, the computation of the Sobol indices is straight-forward. The total variance (D) in terms of the expansion coefficients can be written as:

$$D = \mathbf{V}[\boldsymbol{v}] = \sum_{i=1}^{P} \boldsymbol{v}_i^2(\boldsymbol{x}, t)\langle\psi_i^2\rangle, \tag{4.15}$$

which can be decomposed as [113]:

$$D = \sum_{i=1}^{i=d} D_i + \sum_{1\le i<j\le d}^{i=d-1} D_{i,j} + \sum_{1\le i<j<k\le d}^{i=d-2} D_{i,j,k} + \cdots + D_{1,2,\cdots,d}, \tag{4.16}$$

where $(D_{i_1,\cdots,i_s})$ are partial variances given by

$$D_{i_1,\cdots,i_s} = \sum_{\beta\in\{i_1,\cdots,i_s\}} \boldsymbol{v}_\beta^2(\boldsymbol{x}, t)\langle\psi_\beta^2\rangle, \quad 1 \le i_1 < \cdots < i_s \le d. \tag{4.17}$$

The Sobol indices $(S_{i_1,\cdots,i_s})$ are then defined as

$$S_{i_1,\cdots,i_s} = \frac{D_{i_1,\cdots,i_s}}{D}, \tag{4.18}$$

such that,

$$\sum_{i=1}^{i=d} S_i + \sum_{1\le i<j\le d}^{i=d-1} S_{i,j} + \sum_{1\le i<j<k\le d}^{i=d-2} S_{i,j,k} + \quad \cdots \quad + S_{1,2,\cdots,d} = 1 \tag{4.19}$$

Thus, the Sobol indices basically measure the combined sensitivity arising from the contribution of each random variable (S_i) and from contributions due to their interactions ($S_{i,j}, S_{i,j,k}, ...$). The combined effect of an uncertain variable with index i is therefore defined as the sum of the partial Sobol indices that include the contribution from the i_{th} variable

$$S_{T_i} = \sum_{L_i} \frac{D_{i_1,\cdots,i_s}}{D}; \quad L_i = \{(i_1,\cdots,i_s) : \exists\, k, 1 \le k \le s, i_k = i\}. \tag{4.20}$$

Thus, the Sobol indices (S_{T_i}) can be used to estimate and compare the contribution of each uncertain parameter to the uncertainty in QoIs.

4.6.4. FORWARD CFD SOLVER BASED ON A PERTURBED RST

Based on a precursor RANS simulation, we obtained a deterministic Reynolds stress tensor field $\mathbf{R}^{(det)}$. For a given perturbed RSTF $\mathbf{R}^{(pert)}$, we compute the perturbation in the RSTF as

$$\Delta\mathbf{R}^* = \mathbf{R}^{(pert)} - \mathbf{R}^{(det)} \tag{4.21}$$

where $\Delta\mathbf{R}^*$ accounts for the model (output) form uncertainty. In order to predict the flow field based on the perturbed RSTF, based on the approach mentioned in [96], the divergence of the perturbation is added to the RANS momentum equations as a forcing term

$$\frac{\partial \overline{\boldsymbol{u}}}{\partial t} + (\overline{\boldsymbol{u}}\cdot\nabla)\overline{\boldsymbol{u}} = -\nabla\overline{p} + \nabla\cdot(\nu\nabla\overline{\boldsymbol{u}}) - \nabla\cdot(\mathbf{R} + \Delta\mathbf{R}^*) + \boldsymbol{f}. \tag{4.22}$$

The above was implemented in the *pimpleFoam* solver in OpenFOAM. Note that the new forward solver, in addition to the usual parameters and fields, also requires the perturbed RSTF in order to predict the flow fields. The perturbation 4.21 is computed internally by the solver.

4.7. ANNEXURE

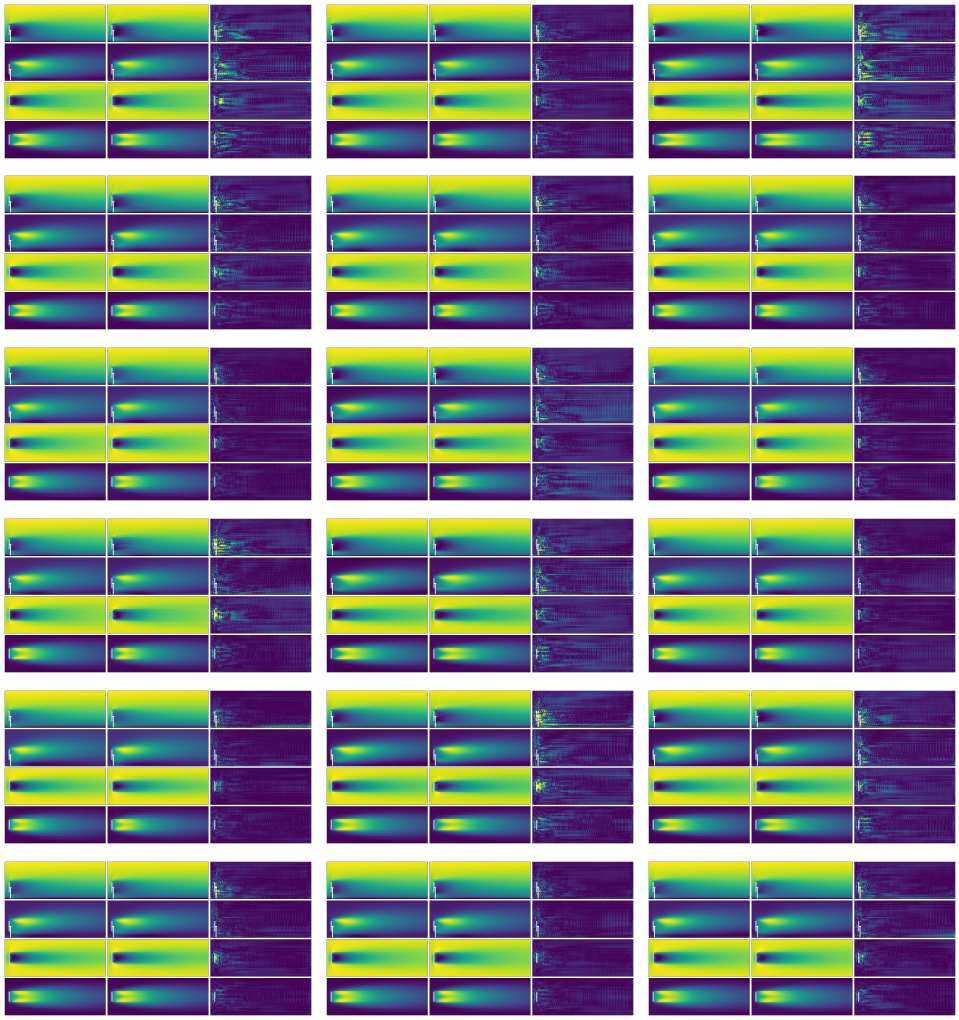


Figure 4.20 | The outputs of a subset of 3D U-Net's test set trained with the data from a single wind turbine simulations using the $k-\omega$ SST RANS turbulence model. As presented in Figure 4.8, each triplet represents one data point and consists of four rows and three columns. The top two rows are the side view of the velocity and turbulent intensity, respectively. The bottom two rows are the top view of the velocity and turbulent intensity, respectively. The columns (left to right) shows the true solution, surrogate prediction and the relative absolute error (in the range [0, 0.05]), respectively.

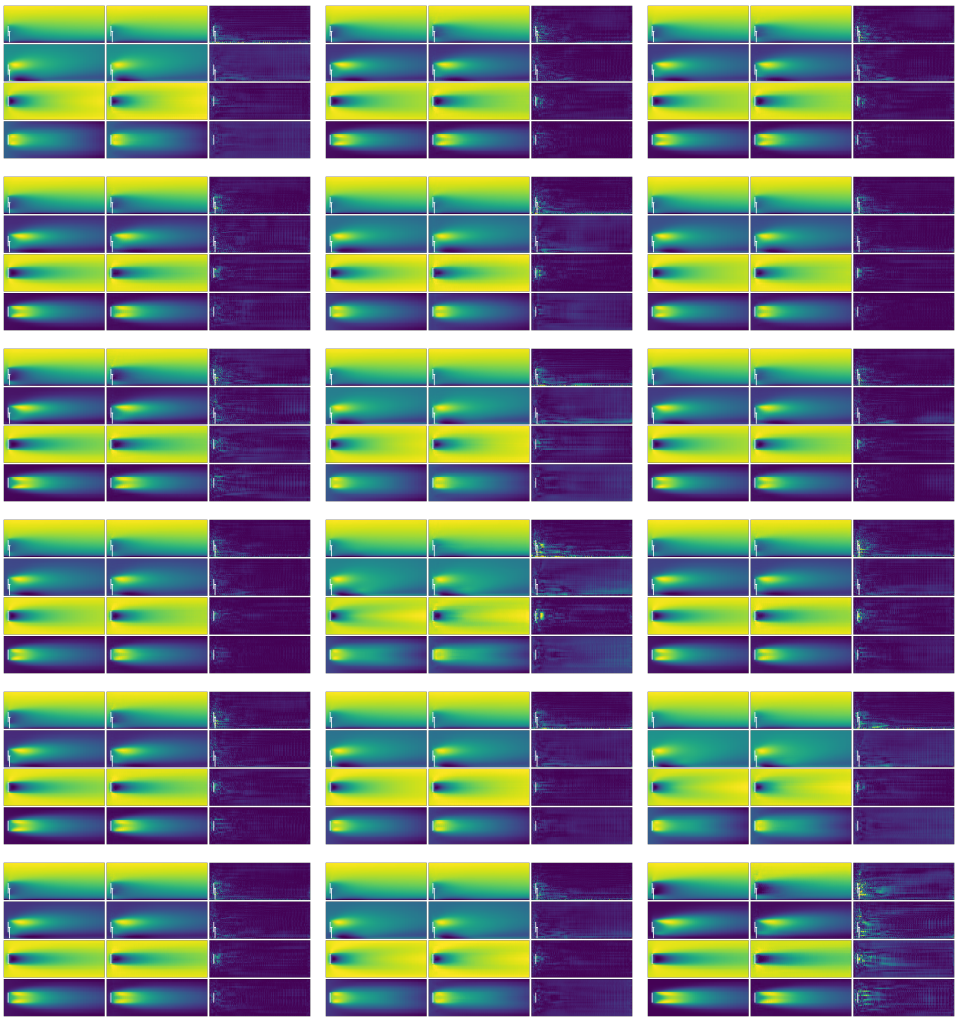


Figure 4.21 | The outputs of a subset of 3D U-Net's test set trained with the data from a single wind turbine simulations using the realizable $k-\epsilon$ RANS turbulence model. As presented Figure 4.13, each triplet represents one data point and consists of four rows and three columns. The top two rows are the side view of the velocity and turbulent intensity, respectively. The bottom two rows are the top view of the velocity and turbulent intensity, respectively. The columns (left to right) shows the true solution, surrogate prediction and the relative absolute error (in the range [0, 0.05]), respectively.

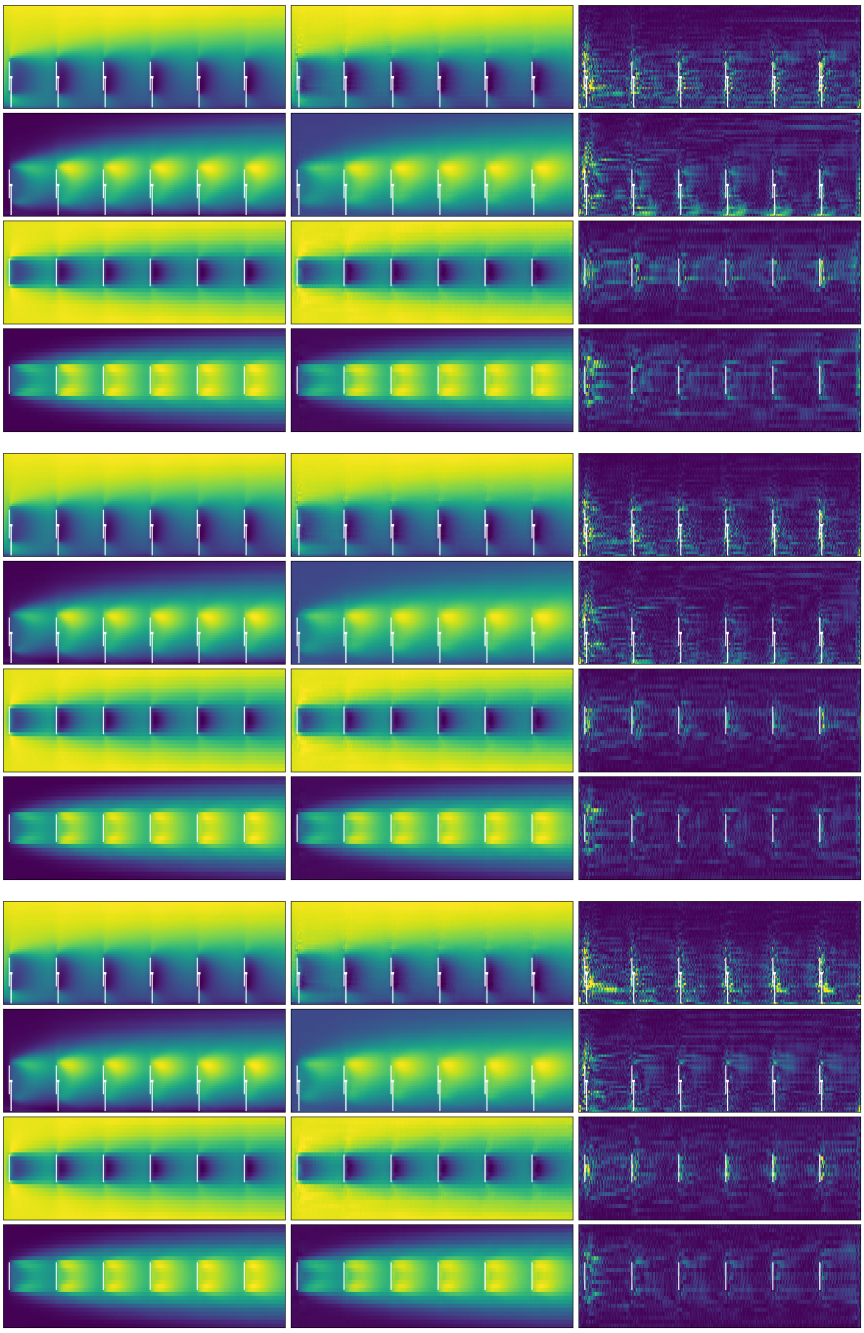


Figure 4.22 | Wake interaction predictions for a subset of SBUQ's (3D U-Net + Wake Superposition) test set trained with the data from a single wind turbine simulations using the realizable $k-\epsilon$ RANS turbulence model. As presented Figure 4.14, each triplet represents one data point and consists of four rows and three columns. The top two rows are the side view of the velocity and turbulent intensity, respectively. The bottom two rows are the top view of the velocity and turbulent intensity, respectively. The columns (left to right) shows the true solution, surrogate prediction and the relative absolute error (in the range [0, 0.05]), respectively.

5

CONCLUSIONS AND PROSPECTS

5.1. CONCLUSIONS

In the context of the CFD 2030 vision [103], Verification, Validation, and Uncertainty Quantification (VVUQ) is expected to play a critical role in the development of next-generation Computational Fluid Dynamics (CFD) simulations that are more accurate, reliable, and efficient. The use of VVUQ will enable CFD simulations to be used with greater confidence in applications such as aerospace, energy, and environmental engineering.

As motivated and stated in chapter 1, the goal of this thesis was to develop methods for quantification and propagation of different forms of uncertainties in complex CFD simulations, and apply them to study the non-linear effects of the uncertainties on the quantities of interest, particularly in the context of wake predictions in wind farms. Majorly, two such methods were presented - (i) the Intrusive Polynomial Chaos (IPC) based stochastic RANS solver, and (ii) the Surrogate Based Uncertainty Quantification (SBUQ) approach using a deep learning model. We will now summarize the most crucial results, observations and conclusions which were encountered during the development of these methods.

INTRUSIVE POLYNOMIAL CHAOS FOR PARAMETRIC UNCERTAINTIES

In Chapter 2, we introduced the implementation and presented the application of a solver for incompressible Navier-Stokes equations that leverages the use of Generalized Polynomial Chaos (gPC) expansion to effectively address the uncertainties associated with fluid flow simulations. The solver was specifically used to characterize *parametric* uncertainties and study their non-linear propagation.

In order to develop an efficient Intrusive Polynomial Chaos (IPC) solver, we discussed the important steps to decouple the system of equations while ensuring the overhead due to coupling is not significant.

To test the IPC based solver, we presented two studies for random laminar viscosity in Poiseuille flow and random Large Eddy Simulation (LES) model parameter C_S in turbulent channel flow. In the plane Poiseuille flow with uncertain viscosity, we realized a significant effect of the uncertainty in the re-circulation region of the flow. The results were also

compared with the Non-intrusive Polynomial Chaos (NIPC) approach with same polynomial order and were found to be very close, verifying our implementation in OpenFOAM. In case of the turbulent channel flow with uncertain LES model parameter, the results were found to be in accordance with NIPC approach with some deviations which were attributed to the error in the NIPC *approximation* of the expansion coefficients. Through this work, using the IPC based solver, we brought to light an alternative to the mostly expensive non-intrusive approaches for UQ analysis in CFD.

QUANTIFICATION AND PROPAGATION OF MODEL-FORM UNCERTAINTIES

The promising results obtained from IPC method discussed in chapter 2 encouraged us to further pursue research in this direction. Despite its capabilities, further development of the IPC based solver for parametric uncertainties may be restricted as it requires significant modifications in the deterministic solver (including templated base classes) to obtain the solution for the coupled equations in the expansion coefficients. In order to overcome this problem, in chapter 3, we presented a more holistic approach to treat uncertainties in a CFD simulation while still be able to use intrusive polynomial chaos method.

Instead of global low-dimensional random parameters/variables, we represented the uncertainties in terms of locally-variable high-dimensional random fields characterized using the information from a deterministic baseline simulation.

This leads to the development of a stochastic RANS solver primarily based on the *model output form* uncertainties. In contrast to the previous IPC based solver for parametric uncertainties, the stochastic RANS solver provides an efficient implementation which required minimal changes in the top-level deterministic code – promoting the re-usability and compatibility with the newer versions of OpenFOAM.

The new IPC based stochastic RANS solver was tested on two benchmark problems for RANS turbulence modeling – the flow over periodic hills and the flow in a square duct using the random eddy viscosity field and the random Reynolds stress tensor field, respectively. A detailed account on the influence of the hyper-parameters of the random fields and that of the stochastic solver was given. Larger correlation lengths resulted in a higher variance, especially near the reattachment location. On the one hand, increasing the Karhunen-Loève decomposition threshold gradually reduced the cardinality of the truncated set, however under-predicted the variance in the output. On the other hand, decreasing the hyperbolic truncation set tuning parameter drastically reduced the cardinality of the truncated set while practically predicting the same level of variance.

A comparison between three widely used RANS turbulence models ($k-\omega$, $k-\omega$ SST and Launder-Sharma $k-\epsilon$) revealed that $k-\omega$ SST model predictions are more accurate with the lowest level of stochasticity implying lower sensitivity of its model parameters and/or lower structural uncertainty in its formulation.

Upon solving for the expansion coefficients, the measurement data was assimilated directly using the response surface (PCE) resulting in a significant reduction in the overall uncertainty levels.

The use cases of the stochastic solver confirmed that – (i) upon proper parameter tuning, the random field models can certainly provide uncertainty bounds that envelop most of the realizable flow states, and (ii) tuning the hyper-parameters based on prior knowledge may allow for thinner uncertain bounds in quantities of interest (QoIs). It shall also be pointed

out that divergence was realized for a high-variance random field, where the higher modes of the polynomial expansion failed to converge after a few initial explicit iterations.

Overall, with minimal changes in the deterministic solver, this work added to the potential of intrusive polynomial chaos based approach for an accurate and efficient UQ analysis of complex CFD simulations.

UNCERTAINTY QUANTIFICATION ANALYSIS IN WIND-FARMS

The IPC based stochastic RANS solver developed so far can be considered a *physics-based* UQ analysis method, where we solve for the coupled expansion coefficients using the Navier-Stokes like systems of equations. In order to assess the performance of this solver with a pure *data-driven* approach for UQ analysis, we employed a Surrogate Based Uncertainty Quantification (SBUQ) method.

In the lights of exponentially increasing applications of deep learning, we constructed a surrogate model using a 3D U-Net neural network (trained over single wind turbine data) combined with wake superposition principles, that is capable of predicting the flow field in a wind farm. This surrogate model was further used to compute the desired statistics of the QoIs over an array of wind turbines.

In chapter 4, a UQ analysis for the prediction of wake interactions in a wind farm with parametric and model form uncertainties was carried out using both the IPC solver and the SBUQ approach.

Two different linear-eddy viscosity based RANS turbulence models namely, the $k-\omega$ SST and the realizable $k-\epsilon$ as well as the LRR Reynolds-stress transport model were considered for investigation. The results from the uncertainty analysis of the above mentioned models are compared with reference (high-fidelity) LES data at mean values of the input random variables or random fields.

The uncertain bounds of the linear-eddy viscosity RANS models were found to capture the high-fidelity solution reasonably well except in wake regions immediately behind the wind turbine. The uncertain bounds of LRR model were able to precisely capture the LES solution over the wind turbines asserting its overall higher accuracy. However, it is also observed that the uncertainty levels in the outputs of the two-equation model remain nearly uniform in the streamwise direction, whereas it grows gradually for the stress transport model.

In this work, we particularly realized the importance of the effect of randomness in the operating conditions and its interaction with the model form uncertainty. The sensitivity analysis revealed that the effect of the variability in the inlet conditions towards the variance in the QoIs was as significant as that of the structural variability of the turbulence model. Interestingly, the normalized velocity deficit and the turbulent intensity fields were found to be insensitive towards the variation in hub-height velocity, but highly dependent on the variation in the hub-height turbulent intensity as well as in the Reynolds-stress tensor field.

The surrogate model was able to predict the flow field in an array of wind turbines with high accuracy and significantly low (prediction) cost and was therefore considered suitable for the SBUQ approach. The results obtained from the SBUQ approach were found to be very close and in accordance with that from the high-fidelity stochastic solver. Both the approaches developed so far, demonstrated the potential to be applied for the analysis of a RANS turbulent flow simulation under various uncertainties.

Overall, this study successfully delivered two highly useful approaches for an accurate and efficient uncertainty quantification and propagation analysis, targeted especially towards the expensive CFD simulations. It was therefore asserted that, based on the engineering application and the time constraints, one approach may have more advantages over the other and a decision shall be taken accordingly.

5.2. PROSPECTS

This work promotes further development in the direction of polynomial chaos method for uncertainty quantification in CFD. A more efficient implementation of the IPC solver can be explored with different covariance kernel functions, and an implicit algorithm to simultaneously solve for all the modes while avoiding the memory issues.

As mentioned earlier, VVUQ is expected to play a critical role in the development of next-generation CFD simulations. This would require handling multiple sources and/or forms of uncertainties in a CFD simulation (see Table 1.1). In this regard, the IPC based stochastic solver can be readily used with minimal tweaking and/or modifications to incorporate e.g. model discrepancy, discretization error uncertainty, etc.

The SBUQ approach can be further enhanced using Graph Neural Networks (GNNs) instead of the currently used Convolutional Neural Networks (CNNs). GNNs can handle unstructured data, capture local as well as global information and offer better interpretbility as compared to CNNs.

The use of low dimensional representation of eddy-viscosity or Reynolds-stress tensor field using dimensionality reduction techniques can significantly speed up the training and prediction time of the SBUQ approach.

The IPC solver and the SBUQ approach can be applied for the uncertainty quantification analysis of further complex unsteady RANS (URANS) or LES engineering simulations. Moreover, these UQ analysis approaches can be extended with e.g. calibration using Bayesian inference, reliability analysis to estimate the failure probability, optimization to obtain robust design etc.

TECHNICAL OUTPUTS

Publications

- J. Parekh and R. Verstappen. Uncertainty quantification analysis for simulation of wakes in wind-farms using a stochastic RANS solver, compared with a deep learning approach. Computers & Fluids, 257:105867, 2023.
- J. Parekh and R. Verstappen. Quantification and propagation of model-form uncertainties in RANS turbulence modeling via intrusive polynomial chaos. International Journal for Uncertainty Quantification., 13:1–29, 2023.
- J. Parekh and R. Verstappen. Intrusive Polynomial Chaos for CFD Using OpenFOAM. In Computational Science – ICCS 2020, pages 677–691, Cham, 2020. Springer International Publishing.

Oral presentations

- Development of Methods for Quantification and Propagation of Uncertainties in CFD Simulations of Wakes in Windfarm, German Aerospace Center (DLR), Braunschweig, Germany. 2022
- Intrusive Polynomial Chaos for Uncertainty Quantification in CFD using OpenFOAM, 15th OpenFoam Workshop, Arlington, USA. 2020
- Intrusive Polynomial Chaos for CFD Using OpenFOAM, UNcErtainty QUantIficatiOn for ComputationAl modeLs (UNEQUIvOCAL) ICCS, Amsterdam, The Netherlands. 2020

Poster presentations

- Intrusive Polynomial Chaos for Uncertainty Quantification in CFD Using OpenFOAM, Frontiers of Uncertainty Quantification in Fluid Dynamics (FrontUQ), Pisa, Italy. 2019
- Uncertainty Quantification in Wind-Farm Simulations, Applied Computational Science symposium (ACOS), Eindhoven, The Netherlands. 2018

BIBLIOGRAPHY

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] Mahdi Abkar and Fernando Porté-Agel. Mean and turbulent kinetic energy budgets inside and above very large wind farms under conventionally-neutral condition. *Renewable Energy*, 70:142–152, oct 2014.

[3] Giancarlo Alfonsi. Reynolds-Averaged Navier–Stokes Equations for Turbulence Modeling. *Applied Mechanics Reviews*, 62(4), 06 2009.

[4] Cristina L. Archer, Ahmadreza Vasel-Be-Hagh, Chi Yan, Sicheng Wu, Yang Pan, Joseph F. Brodie, and A. Eoghan Maguire. Review and evaluation of wake loss models for wind energy applications. *Applied Energy*, 226:1187–1207, sep 2018.

[5] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, nov 2017.

[6] R. J. Barthelmie, K. Hansen, S. T. Frandsen, O. Rathmann, J. G. Schepers, W. Schlez, J. Phillips, K. Rados, A. Zervos, E. S. Politis, and P. K. Chaviaropoulos. Modelling and measuring flow and wind turbine wakes in large wind farms offshore. *Wind Energy*, 12(5):431–444, 2009.

[7] Hester Bijl, Didier Lucor, Siddhartha Mishra, and Christoph Schwab. *Uncertainty Quantification in Computational Fluid Dynamics*, volume 92. 2013.

[8] Christopher M. Bishop. *Pattern Recogniton and Machine Learning*. Springer-Verlag New York, 2006.

[9] Bernardo P. Brener, Matheus A. Cruz, Roney L. Thompson, and Rodrigo P. Anjos. Conditioning and accurate solutions of Reynolds average Navier-Stokes equations with data-driven turbulence closures. *Journal of Fluid Mechanics*, 915:110, 2021.

[10] M. Breuer, N. Peller, Ch Rapp, and M. Manhart. Flow over periodic hills - Numerical and experimental study in a wide range of Reynolds numbers. *Computers and Fluids*, 38(2):433–457, feb 2009.

[11] D. Cabezón, E. Migoya, and A. Crespo. Comparison of turbulence models for the computational fluid dynamics simulation of wind turbine wakes in the atmospheric boundary layer. *Wind Energy*, 14(7):909–921, oct 2011.

[12] A. Crespo, J. Hernández, and S. Frandsen. Survey of modelling methods for wind turbine wakes and wind farms. *Wind Energy*, 2(1):1–24, 1999.

[13] Egene DeVilliers. *The Potential of Large Eddy Simulation for the Modeling of Wall Bounded Flows Eugene de Villiers*. PhD thesis, Imperial College of Science, Technology and Medicine, 2006.

[14] Eric Dow and Qiqi Wang. Quantification of structural uncertainties in the k-ω turbulence model. *Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, 2011.

[15] David Draper. Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):45–70, jan 1995.

[16] Karthik Duraisamy, Gianluca Iaccarino, and Heng Xiao. Turbulence modeling in the age of data. *Annual Review of Fluid Mechanics*, 51(1):357–377, 2019.

[17] Franz Durst. *Turbulent Flows*. Cambridge University Press, Cambridge, 2022.

[18] W. N. Edeling, P. Cinnella, and R. P. Dwight. Predictive RANS simulations via bayesian model-scenario averaging. *Journal of Computational Physics*, 275:65–91, 2014.

[19] W. N. Edeling, P. Cinnella, R. P. Dwight, and H. Bijl. Bayesian estimates of parameter variability in the k-ϵ turbulence model. *Journal of Computational Physics*, 258:73–94, feb 2014.

[20] Ali Eidi, Reza Ghiassi, Xiang Yang, and Mahdi Abkar. Model-form uncertainty quantification in RANS simulations of wakes and power losses in wind farms. *Renewable Energy*, 179:2212–2223, aug 2021.

[21] By M Emory and G Iaccarino. Visualizing turbulence anisotropy in the spatial domain with componentality contours. *Center for Turbulence Research Annual Research Briefs*, pages 123–137, 2014.

[22] Michael Emory, Johan Larsson, and Gianluca Iaccarino. Modeling of structural uncertainties in Reynolds-averaged Navier-Stokes closures. *Physics of Fluids*, 25(11):110822, aug 2013.

[23] Michael Emory, Rene Pecnik, and Gianluca Iaccarino. Modeling Structural Uncertainties in Reynolds-Averaged Computations of Shock/Boundary Layer Interactions. In *49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, Reston, Virigina, jan 2011. American Institute of Aeronautics and Astronautics.

[24] C.M. Engelen. *The nonlinear effect of combining uncertainties on the energy yield of an offshore wind farm*. PhD thesis, 2015.

[25] ESI-OpenCFD. OpenFOAM (v1806) - The Open Source CFD Toolbox, 2018.

[26] Jonathan Feinberg and Hans Petter Langtangen. Chaospy: An open source tool for designing methods of uncertainty quantification. *Journal of Computational Science*, 11:46–57, 2015.

[27] Ju Feng and Wen Zhong Shen. Modelling Wind for Wind Farm Layout Optimization Using Joint Distribution of Wind Speed and Wind Direction. *Energies 2015, Vol. 8, Pages 3075-3092*, 8(4):3075–3092, apr 2015.

[28] Joel H. Ferziger and Milovan Perić. *Computational Methods for Fluid Dynamics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.

[29] Manuel Fluck and Curran Crawford. A stochastic aerodynamic model for stationary blades in unsteady 3D wind fields. *Journal of Physics: Conference Series*, 753(8):082009, sep 2016.

[30] Manuel Fluck and Curran Crawford. An engineering model for 3-D turbulent wind inflow based on a limited set of random variables. *Wind Energy Science*, 2(2):507–520, jul 2017.

[31] Daniel Foti, Xiaolei Yang, and Fotis Sotiropoulos. Uncertainty quantification of infinite aligned wind farm performance using non-intrusive polynomial chaos and a distributed roughness model. *Wind Energy*, 20(6):945–958, jun 2017.

[32] M. Gaumond, P. E. Réthoré, S. Ott, A. Peña, A. Bechmann, and K. S. Hansen. Evaluation of the wind direction uncertainty and its impact on wake modeling at the Horns Rev offshore wind farm. *Wind Energy*, 17(8):1169–1178, aug 2014.

[33] P. M.O. Gebraad, F. W. Teeuwisse, J. W. Van Wingerden, P. A. Fleming, S. D. Ruben, J. R. Marden, and L. Y. Pao. A data-driven model for wind plant power optimization by yaw control. *Proceedings of the American Control Conference*, pages 3128–3134, 2014.

[34] Roger Ghanem, Houman Owhadi, and David Higdon. *Handbook of uncertainty quantification*. 2017.

[35] Roger G. Ghanem and Pol D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer New York, New York, NY, 1991.

[36] Tuhfe Göçmen, Paul Van Der Laan, Pierre Elouan Réthoré, Alfredo Peña Diaz, Gunner Chr Larsen, and Søren Ott. Wind turbine wake models developed at the technical university of Denmark: A review. *Renewable and Sustainable Energy Reviews*, 60:752–769, jul 2016.

[37] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[38] C. Gorlé and G. Iaccarino. A framework for epistemic uncertainty quantification of turbulent scalar flux models for Reynolds-averaged Navier-Stokes simulations. *Physics of Fluids*, 25(5):55105, may 2013.

[39] C. Gorlé, S. Zeoli, M. Emory, J. Larsson, and G. Iaccarino. Epistemic uncertainty quantification for Reynolds-averaged Navier-Stokes modeling of separated flows over streamlined surfaces. *Physics of Fluids*, 31(3):035101, mar 2019.

[40] Catherine Gorlé, Clara Garcia-Sanchez, and Gianluca Iaccarino. Quantifying inflow and RANS turbulence model form uncertainties for wind engineering flows. *Journal of Wind Engineering and Industrial Aerodynamics*, 144:202–212, sep 2015.

[41] Daoru Han and Serhat Hosder. Inherent and epistemic uncertainty analysis for computational fluid dynamics simulations of synthetic jet actuators. Technical Report 6, 2014.

[42] D. M. Hargreaves and N. G. Wright. On the use of the k-ϵlunate model in commercial CFD software to model the neutral atmospheric boundary layer. *Journal of Wind Engineering and Industrial Aerodynamics*, 95(5):355–369, may 2007.

[43] Simon D. Hornshøj-Møller, Peter D. Nielsen, Pourya Forooghi, and Mahdi Abkar. Quantifying structural uncertainties in Reynolds-averaged Navier–Stokes simulations of wind turbine wakes. *Renewable Energy*, 164:1550–1558, 2021.

[44] Serhat Hosder and Robert W Walters. Non-Intrusive Polynomial Chaos Methods for Stochastic CFD-Theory and Applications. In *RTO Meeting Proceedings*, number 2007, pages 1–18. Computational Uncertainty in Military Vehicle Design, 2007.

[45] Tianfeng Hou, Dirk Nuyens, Staf Roels, and Hans Janssen. Quasi-Monte Carlo based uncertainty analysis: Sampling efficiency and error estimation in engineering applications. *Reliability Engineering and System Safety*, 191:106549, nov 2019.

[46] Gianluca Iaccarino, Aashwin Ananda Mishra, and Saman Ghili. Eigenspace perturbations for uncertainty estimation of single-point turbulence closures. *Physical Review Fluids*, 2(2):024605, feb 2017.

[47] IEA. International energy outlook. *U.S. Energy Information Administration, Office of Energy Analysis, US Department of Energy: Washington, DC, USA*, page 20585, 2017.

[48] Marco A. Iglesias, Kody J.H. Law, and Andrew M. Stuart. Ensemble kalman methods for inverse problems. *Inverse Problems*, 29(4):45001, 2013.

[49] R. I. Issa. Solution of the implicitly discretised fluid flow equations by operator-splitting. *Journal of Computational Physics*, 62(1):40–65, jan 1986.

[50] Javier Jiménez. Computers and turbulence. *European Journal of Mechanics - B/Fluids*, 79:1–11, 2020.

[51] Tongdan Jin and Zhigang Tian. Uncertainty analysis for wind energy production with dynamic power curves. *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2010*, pages 745–750, 2010.

[52] Volker John. Large eddy simulation of turbulent incompressible flows - analytical and numerical results for a class of les models. In *Lecture Notes in Computational Science and Engineering*, 2003.

[53] Farzad Karami, Nasser Kehtarnavaz, and Mario Rotea. Probabilistic Neural Network to Quantify Uncertainty of Wind Power Estimation. *Proceedings of the 2022 IEEE Dallas Circuits and Systems Conference, DCAS 2022*, 2022.

[54] Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 63(3):425–464, jan 2001.

[55] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, dec 2015.

[56] Peter K. Kitanidis. Parameter Uncertainty in Estimation of Spatial Functions: Bayesian Analysis. *Water Resources Research*, 22(4):499–507, apr 1986.

[57] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2):105–112, mar 2009.

[58] O. M. Knio and O. P. Le Maître. Uncertainty propagation in CFD using polynomial chaos decomposition. *Fluid Dynamics Research*, 38(9):616–640, 2006.

[59] Katiana Kontolati, Dimitrios Loukrezis, Dimitrios G. Giovanis, Lohit Vandanapu, and Michael D. Shields. A survey of unsupervised learning methods for high-dimensional uncertainty quantification in black-box-type problems. *Journal of Computational Physics*, 464, 2022.

[60] Prashant Kumar, Martin Schmelzer, and Richard P. Dwight. Stochastic turbulence modeling in RANS simulations via multilevel Monte Carlo. *Computers and Fluids*, 201:104420, 2020.

[61] Soon Duck Kwon. Uncertainty analysis of wind energy potential assessment. *Applied Energy*, 87(3):856–865, mar 2010.

[62] Ryan J Langdon, Paul D Yousefi, Caroline L Relton, and Matthew J Suderman. *Efficient turbulence modeling for CFD wake simulations*. PhD thesis, 2014.

[63] B. E. Launder, G. J. Reece, and W. Rodi. Progress in the development of a Reynolds-stress turbulence closure. *Journal of Fluid Mechanics*, 68(3):537–566, 1975.

[64] B. E. Launder and B. I. Sharma. Application of the energy-dissipation model of turbulence to the calculation of flow near a spinning disc. *Letters in Heat and Mass Transfer*, 1(2):131–137, nov 1974.

[65] B.E. Launder and D.B. Spalding. The numerical computation of turbulent flows. *Computer Methods in Applied Mechanics and Engineering*, 3(2):269–289, 1974.

[66] O. P. Le Maître and O. M. Knio. *Introduction: Uncertainty Quantification and Propagation*. Springer, 2010.

[67] Olivier P. Le Matre, Omar M. Knio, Habib N. Najm, and Roger G. Ghanem. A stochastic projection method for fluid flow. I. Basic formulation. *Journal of Computational Physics*, 173(2):481–511, 2001.

[68] Gunther Leobacher and Friedrich Pillichshammer. *Introduction to Quasi-Monte Carlo Integration and Applications*. Compact Textbooks in Mathematics. Springer International Publishing, Cham, 2014.

[69] Jia Li and Dongbin Xiu. A generalized polynomial chaos based ensemble Kalman filter with high accuracy. *Journal of Computational Physics*, 228(15):5454–5469, 2009.

[70] J. Ling and J. Templeton. Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty. *Physics of Fluids*, 27(8):85103, aug 2015.

[71] Zhi Yi Liu, Xiao Dong Wang, and Shun Kang. Stochastic performance evaluation of horizontal axis wind turbine blades using non-deterministic CFD simulations. *Energy*, 73:126–136, aug 2014.

[72] Didier Lucor, Johan Meyers, and Pierre Sagaut. Sensitivity analysis of large-eddy simulations to subgrid-scale-model parametric uncertainty using polynomial chaos. *Journal of Fluid Mechanics*, 585:255–279, 2007.

[73] Koushik Marepally, Yong Su Jung, James Baeder, and Ganesh Vijayakumar. Uncertainty quantification of wind turbine airfoil aerodynamics with geometric uncertainty. *Journal of Physics: Conference Series*, 2265(4), jun 2022.

[74] L. Margheri, M. Meldi, M. V. Salvetti, and P. Sagaut. Epistemic uncertainties in RANS model free coefficients. *Computers and Fluids*, 102:315–335, 2014.

[75] V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield. IPCC, 2018: Summary for Policymakers. Technical report, 2018.

[76] S. Mathew and G.S. Philip. Wind Turbines: Evolution, Basic Principles, and Classifications. *Comprehensive Renewable Energy*, pages 104–123, jan 2012.

[77] D. Mehta, A. H. van Zuijlen, B. Koren, J. G. Holierhoek, and H. Bijl. Large Eddy Simulation of wind farm aerodynamics: A review. *Journal of Wind Engineering and Industrial Aerodynamics*, 133:1–17, oct 2014.

[78] F. R. Menter. Two-equation eddy-viscosity turbulence models for engineering applications. *AIAA Journal*, 32(8):1598–1605, may 1994.

[79] Robert D. Moser, John Kim, and Nagi N. Mansour. Direct numerical simulation of turbulent channel flow up to Reτ=590. *Physics of Fluids*, 11(4):943–945, 1999.

[80] J. P. Murcia, P. E. Réthoré, A. Natarajan, and J. D. Sørensen. How Many Model Evaluations Are Required To Predict The AEP Of A Wind Power Plant? *Journal of Physics: Conference Series*, 625(1):012030, jun 2015.

[81] William L. Oberkamp, Jon C. Helton, and Kari Sentz. Mathematical representation of uncertainty. *19th AIAA Applied Aerodynamics Conference*, 2001.

[82] Todd A Oliver, Ernesto Prudencio, Serge Prudhomme, and Robert D Moser. Uncertainty Quantification for RANS Turbulence Model Predictions. *APS*, 62:2005–2005, 2005.

[83] John O'Neil and Charles Meneveau. Subgrid-scale stresses and their modelling in a turbulent plane wake. *Journal of Fluid Mechanics*, 349:253–293, oct 1997.

[84] A. S. Padrón, A. P.J. Stanley, J. J. Thomas, J. J. Alonso, and A. Ning. Polynomial chaos for the computation of annual energy production in wind farm layout optimization. *Journal of Physics: Conference Series*, 753(3):032021, sep 2016.

[85] J. Parekh and R. Verstappen. Intrusive Polynomial Chaos for CFD Using OpenFOAM. In *Computational Science – ICCS 2020*, pages 677–691, Cham, 2020. Springer International Publishing.

[86] J. Parekh and R. Verstappen. Quantification and propagation of model-form uncertainties in RANS turbulence modeling via intrusive polynomial chaos. *International Journal for Uncertainty Quantification.*, 13:1–29, 2023.

[87] Enrico Zio Pedroni and Nicola. Methods for representing uncertainty. *Cahiers Georges Sorel*, 5(1):77–88, 1987.

[88] Alfredo Pinelli, Markus Uhlmann, Atsushi Sekimoto, and Genta Kawahara. Erratum: Reynolds number dependence of mean flow structure in square duct turbulence. *Journal of Fluid Mechanics*, 653:537, 2010.

[89] Anne-Laure Popelin and Anne Dutfoy. Open TURNS, an Open Source Uncertainty Engineering Software. In *EDP Sciences*, page 05408, 2014.

[90] Fernando Porté-Agel, Majid Bastankhah, and Sina Shamsoddin. Wind-Turbine and Wind-Farm Flows: A Review. *Boundary-Layer Meteorology*, 174(1):1–59, jan 2020.

[91] Fernando Porte-Agel, Yu Ting Wu, and Chang Hung Chen. A numerical study of the effects ofwind direction on turbine wakes and power losses in a largewind farm. *Energies*, 6(10):5297–5313, oct 2013.

[92] United Nations Environment Programme. Paris agreement, 2015.

[93] J. M. Prospathopoulos, E. S. Politis, K. G. Rados, and P. K. Chaviaropoulos. Evaluation of the effects of turbulence model enhancements on wind turbine wake predictions. *Wind Energy*, 14(2):285–300, 2011.

[94] J. Ray, Z. Hou, M. Huang, K. Sargsyan, and L. Swiler. Bayesian calibration of the community land model using surrogates. *SIAM-ASA Journal on Uncertainty Quantification*, 3(1):199–233, mar 2015.

[95] Jaideep Ray, Sophia Lefantzi, Srinivasan Arunajatesan, and Lawrence Dechant. Bayesian calibration of a RANS model with a complex response surface-a case study with jet-in-crossflow configuration. *45th AIAA Fluid Dynamics Conference*, may 2015.

[96] Luis F.Cremades Rey, Denis F. Hinz, and Mahdi Abkar. Reynolds stress perturbation for epistemic uncertainty quantification of RANS models implemented in OpenFOAM. *Fluids*, 4(2):113, jun 2019.

[97] Jennifer M. Rinker. Calculating the sensitivity of wind turbine loads to wind inputs using response surfaces. *Journal of Physics: Conference Series*, 753(3):032057, sep 2016.

[98] Christopher J. Roy and William L. Oberkampf. A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. *Computer Methods in Applied Mechanics and Engineering*, 200(25-28):2131–2144, jun 2011.

[99] Pierre Sagaut. *Large Eddy Simulation for Incompressible Flows – An Introduction*, volume 73. Springer, 2001.

[100] Shigehiro Sakamoto and Roger Ghanem. Polynomial Chaos Decomposition for the Simulation of Non-Gaussian Nonstationary Stochastic Processes. *Journal of Engineering Mechanics*, 128(2):190–201, feb 2002.

[101] B. Sanderse, S. P. Van Der Pijl, and B. Koren. Review of computational fluid dynamics for wind turbine wake aerodynamics. *Wind Energy*, 14(7):799–819, 2011.

[102] Jie Shen and Li Lian Wang. Sparse spectral approximations of high-dimensional problems based on hyperbolic cross. *SIAM Journal on Numerical Analysis*, 48(3):1087–1109, 2010.

[103] Jeffrey P. Slotnick, Abdollah Khodadoust, Juan J. Alonso, David L. Darmofal, William Gropp, Elizabeth A. Lurie, and Dimitri J. Mavriplis. CFD vision 2030 study: A path to revolutionary computational aerosciences. 2014.

[104] J. Smagorinsky. General Circulation Experiments With the Primitive Equations. *Monthly Weather Review*, 91(3):99–164, mar 1963.

[105] Ralph C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, USA, 2013.

[106] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[107] C. Soize. Nonparametric model of random uncertainties for reduced matrix models in structural dynamics. *Probabilistic Engineering Mechanics*, 15(3):277–294, jul 2000.

[108] C. Soize. Random matrix theory for modeling uncertainties in computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, 194(12-16):1333–1366, apr 2005.

[109] Charles G. Speziale. On nonlinear K-l and K-ϵ models of turbulence. *Journal of Fluid Mechanics*, 178:459–475, 1987.

[110] Julia Steiner, Richard P. Dwight, and Axelle Viré. Data-driven RANS closures for wind turbine wakes under neutral conditions. *Computers and Fluids*, 233:105213, jan 2022.

[111] Julia Steiner, Axelle Viré, and Richard P. Dwight. Classifying Regions of High Model Error Within a Data-Driven RANS Closure: Application to Wind Turbine Wakes. *Flow, Turbulence and Combustion*, 109(3):545–570, sep 2022.

[112] Richard J.A.M. Stevens and Charles Meneveau. Flow Structure and Turbulence in Wind Farms. *Annual Review of Fluid Mechanics*, 49:311–339, jan 2017.

[113] Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, 93(7):964–979, 2008.

[114] Bruno Sudret, Stefano Marelli, and Joe Wiart. Surrogate models for uncertainty quantification: An overview. In *2017 11th European Conference on Antennas and Propagation (EUCAP)*, pages 793–797, 2017.

[115] Salar Taghizadeh, Freddie D Witherden, Sharath S Girimaji, G V Iungo, F Viola, U Ciri, M A Rotea, and S Leonardi. Data-driven RANS for simulations of large wind farms. *Journal of Physics: Conference Series*, 625(1):012025, jun 2015.

[116] Nils Thuerey, Konstantin Weißenow, Lukas Prantl, and Xiangyu Hu. Deep learning methods for reynolds-averaged navier–stokes simulations of airfoil flows. *AIAA Journal*, 58(1):25–36, jan 2020.

[117] Zilong Ti, Xiao Wei Deng, and Hongxing Yang. Wake modeling of wind turbines using machine learning. *Applied Energy*, 257, 2019.

[118] Kendra L. Van Buren, Mark G. Mollineaux, François M. Hemez, and Sezer Atamturktur. Simulating the dynamics of wind turbine blades: part II, model validation and uncertainty quantification. *Wind Energy*, 16(5):741–758, jul 2013.

[119] M. P. Van Der Laan and S. J. Andersen. The turbulence scales of a wind turbine wake: A revisit of extended k-epsilon models. *Journal of Physics: Conference Series*, 1037(7):072001, jun 2018.

[120] M. Paul Van Der Laan, Niels N. Srensen, Pierre Elouan Réthoré, Jakob Mann, Mark C. Kelly, Niels Troldborg, J. Gerard Schepers, and Ewan Machefaux. An improved k-ϵ model applied to a wind turbine wake in atmospheric turbulence. *Wind Energy*, 18(5):889–907, may 2015.

[121] Brian J. Vanderwende, Branko Kosović, Julie K. Lundquist, and Jeffrey D. Mirocha. Simulating effects of a wind-turbine array using LES and RANS. *Journal of Advances in Modeling Earth Systems*, 8(3):1376–1390, sep 2016.

[122] L. J. Vermeer, J. N. Sørensen, and A. Crespo. Wind turbine wake aerodynamics. *Progress in Aerospace Sciences*, 39(6-7):467–510, aug 2003.

[123] Luc Vinet and Alexei Zhedanov. *A 'missing' family of classical orthogonal polynomials*, volume 44. Society for Industrial and Applied Mathematics, 2011.

[124] Robert W Walters and Luc Huyse. Uncertainty Analysis for Fluid Mechanics with Applications. Technical report, ICASE NASA Langley Research Center Hampton, 2002.

[125] Norbert Wiener. The Homogeneous Chaos. *American Journal of Mathematics*, 60(4):897, 1938.

[126] David C. Wilcox. Reassessment of the scale-determining equation for advanced turbulence models. *AIAA Journal*, 26(11):1299–1310, may 1988.

[127] Jeroen A.S. Witteveen and Hester Bijl. Modeling arbitrary uncertainties using gram-schmidt polynomial chaos. In *Collection of Technical Papers - 44th AIAA Aerospace Sciences Meeting*, volume 14, pages 10695–10711, Reston, Virigina, jan 2006. American Institute of Aeronautics and Astronautics.

[128] Jin-Long Wu, Heng Xiao, Eric G Paterson, Christopher J Roy, and K Todd Lowe. *Predictive Turbulence Modeling with Bayesian Inference and Physics-Informed Machine Learning*. PhD thesis, Virginia Polytechnic Institute and State University, 2018.

[129] Jinlong Wu, Heng Xiao, Rui Sun, and Qiqi Wang. Reynolds-averaged Navier-Stokes equations with explicit data-driven Reynolds stress closure can be ill-conditioned. *Journal of Fluid Mechanics*, 869:553–586, 2019.

[130] Yu Ting Wu and Fernando Porté-Agel. Large-Eddy Simulation of Wind-Turbine Wakes: Evaluation of Turbine Parametrisations. *Boundary-Layer Meteorology*, 138(3):345–366, dec 2011.

[131] Yu Ting Wu and Fernando Porté-Agel. Modeling turbine wakes and power losses within a wind farm using LES: An application to the Horns Rev offshore wind farm. *Renewable Energy*, 75:945–955, mar 2015.

[132] H. Xiao, J. L. Wu, J. X. Wang, R. Sun, and C. J. Roy. Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations: A data-driven, physics-informed Bayesian approach. *Journal of Computational Physics*, 324:115–136, 2016.

[133] Heng Xiao and Paola Cinnella. Quantification of model uncertainty in RANS simulations: A review, 2019.

[134] Heng Xiao, Jian Xun Wang, and Roger G. Ghanem. A random matrix approach for quantifying model-form uncertainties in turbulence modeling. *Computer Methods in Applied Mechanics and Engineering*, 313:941–965, 2017.

[135] Dongbin Xiu. Efficient collocational approach for parametric uncertainty analysis. *Communications in Computational Physics*, 2(2):293–309, 2007.

[136] Dongbin Xiu. Fast numerical methods for stochastic computations: A review. *Communications in Computational Physics*, 5(2-4):242–272, 2009.

[137] Dongbin Xiu and George Em Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal of Scientific Computing*, 24(2):619–644, 2003.

[138] Dongbin Xiu and George Em Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics*, 187(1):137–167, 2003.

[139] X. I.A. Yang, J. Sadique, R. Mittal, and C. Meneveau. Integral wall model for large eddy simulations of wall-bounded turbulent flows. *Physics of Fluids*, 27(2), feb 2015.

[140] Xiang I.A. Yang and Mahdi Abkar. A hierarchical random additive model for passive scalars in wall-bounded flows at high Reynolds numbers. *Journal of Fluid Mechanics*, 842:354–380, may 2018.

[141] Yi Yang, Ming Gu, Suqin Chen, and Xinyang Jin. New inflow boundary conditions for modelling the neutral equilibrium atmospheric boundary layer in computational wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, 97(2):88–95, feb 2009.

[142] Jincheng Zhang and Xiaowei Zhao. Quantification of parameter uncertainty in wind farm wake modeling. *Energy*, 196, 2020.

[143] Xin Lei Zhang, Carlos Michelén-Ströfer, and Heng Xiao. Regularized ensemble Kalman methods for inverse problems. *Journal of Computational Physics*, 416:109517, 2020.

SUMMARY

The CFD 2030 vision aims to improve the accuracy, reliability, and efficiency of Computational Fluid Dynamics (CFD) simulations in fields such as aerospace, energy, and environmental engineering. To achieve this, the use of Verification, Validation, and Uncertainty Quantification (VVUQ) is critical. Uncertainty Quantification (UQ) helps to quantify and propagate different forms of uncertainties in complex CFD simulations, which can be used to study their non-linear effects on quantities of interest.

This thesis focuses on developing methods for quantifying and propagating uncertainties in CFD simulations and applying them to wind farm wake predictions. Two methods were presented: (i) the Intrusive Polynomial Chaos (IPC) based stochastic solver, and (ii) the Surrogate Based Uncertainty Quantification (SBUQ) approach using a deep learning model.

In Chapter 2, the IPC solver was introduced, which uses the Generalized Polynomial Chaos (gPC) expansion to address the uncertainties associated with fluid flow simulations. The solver was specifically used to characterize parametric uncertainties and study their non-linear propagation. The IPC solver was tested on two studies for random laminar viscosity in Poiseuille flow and random Large Eddy Simulation (LES) model parameter C_S in turbulent channel flow. The results showed that the IPC solver is a promising alternative to non-intrusive and expensive approaches for UQ analysis in CFD.

Chapter 3 presents a more holistic approach for treating uncertainties in a CFD simulation, where uncertainties are represented in terms of locally-variable high-dimensional random fields characterized using the information from a baseline simulation. This leads to the development of a stochastic RANS solver primarily based on the model output form uncertainties. The new IPC based stochastic RANS solver was tested on two benchmark problems for RANS turbulence modeling – the flow over periodic hills and the flow in a square duct using the random eddy viscosity field and the random Reynolds stress tensor field, respectively. A detailed account of the influence of the hyper-parameters of the stochastic solver and that of the random fields was given. The use cases of the stochastic solver confirmed that the random field models can certainly provide uncertainty bounds that envelops most of the realizable flow states. Tuning the hyper-parameters based on prior knowledge may allow for thinner uncertain bounds in quantities of interest (QoIs). However, it should be noted that divergence was realized for a high-variance random field.

In chapter 4, we proposed a SBUQ method, particularly for the UQ analysis of wind farm wake predictions. We constructed a surrogate model using a 3D U-Net neural network (trained over single wind turbine data) combined with wake superposition principles, capable of predicting the flow field in a wind farm. This surrogate model was further used to compute the desired statistics of the QoIs over an array of wind turbines. A UQ analysis for the prediction of wake interactions in a wind farm with parametric and model form uncertainties was carried out using both the IPC solver and the SBUQ approach. In this work, we particularly realized the importance of the effect of randomness in the operating conditions and its interaction with the model form uncertainty.

Both the above-mentioned approaches demonstrated the potential to be applied for the analysis of a RANS turbulent flow simulation under various uncertainties. The efforts in this thesis added to the potential of intrusive polynomial chaos based approaches for accurate and efficient UQ analysis of complex CFD simulations. The IPC based stochastic RANS solver provides an efficient implementation with minimal changes in the deterministic code, promoting reusability and compatibility with newer versions of OpenFOAM. The SBUQ approach used the data from single wind turbine simulations to predict the flow field in an array of wind turbines with a reasonable accuracy, which was further used in the SBUQ approach to make fast predictions.

The results presented in this thesis have important implications for the development of next-generation CFD simulations, which can be used with greater confidence in applications such as aerospace, energy, and environmental engineering.

SAMENVATTING

De Computational Fluid Dynamics CFD 2030-visie beoogt de nauwkeurigheid, betrouwbaarheid en efficiëntie van numerieke stromingsleer te verbeteren op toepassingsgebieden zoals lucht- en ruimtevaart, energie en milieutechniek. Volgens deze visie zijn Verificatie, Validatie en Uncertainty Quantification (VVUQ) van cruciaal belang. Onzekerheidskwantificering (Uncertainty Quantification, UQ) kan worden gebruikt om verschillende vormen van onzekerheden in complexe numerieke stromingssimulaties te kwantificeren en te propageren. Aldus kunnen niet-lineaire effecten van onzekerheden op relevante grootheden worden bestudeerd.

Deze dissertatie gaat over het ontwikkelen van methoden voor het kwantificeren en propageren van onzekerheden in numeriek simulaties van stromingen en het toepassen daarvan op het voorspellen van zoggen in windparken. Twee methoden werden gepresenteerd: (i) een stochastische methode gebaseerd op Intrusive Polynomial Chaos (IPC), en (ii) een Surrogate Based Uncertainty Quantification (SBUQ) benadering die gebruik maakt van een deep learning model.

De IPC-methode wordt geïntroduceerd in Hoofdstuk 2. Om onzekerheden in numerieke simulaties van stromingen te beschrijven wordt de Generalized Polynomial Chaos (gPC) expansie gebruikt. Deze IPC-methode wordt toegepast om parametrische onzekerheden te karakteriseren en hun niet-lineaire propagatie te bestuderen. De IPC-methode is getest op een laminaire kanaalstroming (Poiseuille profiel) met een onzekere viscositeit en een Large Eddy Simulatie (LES) van een turbulente kanaalstroming met een onzekere modelparameter, de Smagorinsky constante CS. De resultaten tonen aan dat de IPC methode een veelbelovend alternatief is voor niet-intrusieve en dure UQ-analyses in numerieke stromingsleer.

Hoofdstuk 3 presenteert een meer holistische benadering van onzekerheden in numerieke stromingsleer, waarbij onzekerheden worden voorgesteld door lokaal-variabele, hoogdimensionale random velden die worden gekarakteriseerd met behulp van een referentiesimulatie. Dit heeft geleid tot de ontwikkeling van een stochastische RANS-methode die hoofdzakelijk gebaseerd is op de onzekerheden in de modeluitvoer. De nieuwe op IPC gebaseerde stochastische RANS-methode is getest op twee benchmarkproblemen voor RANS-turbulentiemodellering – de stroming over periodieke heuvels en de stroming in een kanaal met een vierkante doorsnede met respectievelijk een random eddy-viscositeitsveld

en een random Reynolds-spanningstensorveld. De invloed van de hyperparameters van de stochastische methode en die van de random velden is beschrijven in detail. De twee benchmarkproblemen waarop de stochastische methode is toegepast bevestigen dat de random veldmodellen onzekerheidsgrenzen kunnen opleveren die de meeste realiseerbare stromingstoestanden omhullen. Door de hyperparameters af te stemmen op basis van voorkennis kunnen de onzekerheidsgrenzen van de relevante grootheden (QoIs) scherper worden. Er moet echter worden opgemerkt dat de methode divergeerde voor random velden met een hoge variantie.

In hoofdstuk 4 wordt een SBUQ-methode voorgesteld, met name voor de UQ-analyse van voorspellingen van zoggen in windparken. Daartoe is een surrogaatmodel geconstrueerd met behulp van een 3D U-Net neuraal netwerk (getraind op basis van data van afzonderlijke windturbines) in combinatie met een superpositieprincipe voor zoggen. Op basis van dit model kan het stromingsveld in een windpark worden voorspeld. Dit surrogaatmodel is verder gebruikt om de gewenste statistieken van de QoIs over een reeks windturbines te berekenen. Een UQ-analyse van de voorspelling van zoginteracties in een windpark met parametrische en modelonzekerheden is uitgevoerd met behulp van zowel de IPC-methode als de SBUQ- benadering. In deze studie hebben wij vooral het belang van het effect van onzekerheden in de bedrijfsomstandigheden en de interactie daarvan met onzekerheden in het simulatiemodel onderkend.

Beide bovengenoemde aanpakken toonden aan dat zij kunnen worden toegepast voor de analyse van een turbulente stromingssimulaties op basis van RANS onder verschillende onzekerheden. Daarmee is het potentieel van een op IPC gebaseerde aanpak voor een nauwkeurige en efficiënte UQ-analyse van complexe CFD-simulaties vergroot. De op IPC gebaseerde stochastische RANS simulatiemethode biedt een efficiënte implementatie met minimale wijzigingen in de deterministische code, wat de herbruikbaarheid en compatibiliteit met nieuwere versies van OpenFOAM bevordert. De SBUQ methode benodigt data van enkele windturbine simulaties om het stromingsveld in een array van windturbines met een redelijke nauwkeurigheid te voorspellen. De in dit proefschrift gepresenteerde resultaten hebben belangrijke implicaties voor de ontwikkeling van de volgende generatie numeriek stromingssimulaties, die met meer vertrouwen kunnen worden gebruikt in toepassingen als lucht- en ruimtevaart, energie en milieutechniek.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor Roel Verstappen. You were always interested in the progress I made and offered complete freedom to pursue my ideas even if they did not fully overlap with your initial plans for this project. You were always available in the office and even online (especially during pandemic times) for any questions I had. I really liked your advice on taking a step back and reflecting on my academic and personal goals every now and then. I would also like to thank Cristóbal Bertoglio, my second supervisor for all the ideas and discussions from his own PhD thesis, which certainly helped me to better understand my project. I enjoyed my back to back TA'ing for the NM1 course with you. Fred Wubs, thank you for always taking an interest in my work and offering me my first ever teaching duty. I would like to extend my thanks to Arthur Veldman for some amazing questions and discussions after my talks. I also thank you for sharing your life and professional experiences from your inspiring career.

Larissa, thank you for being one of the best people I have ever come across. I really enjoyed our several sports adventures including squash, canoeing, swimming and padel. Looking forward to more in Braunschweig. Ronald, your will power and stoic mindset inspires me. Thank you for bearing with me when our canoe got stuck in the rocks quite a few times. Also, a huge thanks to you for introducing me to squash which turned into my PhD stress buster. Jeremias, thanks for your prayers during my shoulder injury – it definitely helped. Reidmen, thanks for the countless discussions about fitness and for accompanying me for a coffee at BB. It was fun organizing the Groningen quiz with you. Thank you Johan and Pablo for collaborating and sharing your Python and GitHub tips and tricks. Thank you all the other colleagues for the nice conversations we had often during lunch or coffee breaks: Christian, David, Erik, Georgia, Hugo, Julie, Maurits, Miriam, Paolo, Sven and Theresa.

Felipe, thanks for sharing you passion for coffee, even in the midst of canoeing in Sweden. Thank you Bhupesh for the long discussions about life in general. Ankush, thanks for being my swimming partner and all those late night Kebab treats. Thank you Atul for brainstorming the idea of 3D UNet with me. Hina, thanks for being such an amazing friend and a true motivator especially during the last phase of my PhD. Michiel, I enjoyed our short yet intense squash routine as well as the chilly winter mornings tennis sessions. I would like to thank Namita and Israfil, for your super delicious Indian food every time I visited you.

Also, a lot of love for your joyful dog Simba, who helped me get over my fear of dogs.

Finally, thanks to my family, for always being supportive through out my PhD journey. Thank you mom for being interested in my work and even making an effort to understand it. Thanks to my brother for helping me out in difficult situations and also for taking care of our family over all these years. Also, thanks to my grand mother, uncles and aunts who made it possible for me to pursue higher education outside India at the first place. Aunt Nayna, thanks for all those weekend calls you made to care for my well being.

Curriculum Vitae

Jigar Parekh was born on December 22, 1990 in Jabalpur, India. He completed his primary and secondary education in Jabalpur. From 2010 to 2014 he was a student at the Indian Institute of Technology (BHU) Varanasi, where he obtained his bachelor's degree in chemical engineering. From 2014 to 2017, he pursued master's in simulation sciences at Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen University, Germany. Thereafter, from 2017 to 2018, he worked as a research associate at Nanyang Technological University (NTU), Singapore.

From 2018-2023, Jigar was a PhD student at the University of Groningen, The Netherlands under the supervision of prof. dr. ir. R.W.C.P. Verstappen. His research topic was the development of methods for uncertainty quantification in CFD applied to wind turbine wake prediction. Currently, he is working as a researcher at German Aerospace Center (DLR) Braunschweig.