Computers and Fluids 161 (2018) 211–224

Contents lists available at ScienceDirect

Computers and Fluids

journal homepage: www.elsevier.com/locate/compfluid

An efficient Bayesian uncertainty quantification approach with application to k-ω-γ transition modeling

Jincheng Zhang, Song Fu*

School of Aerospace Engineering, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 12 June 2017
Revised 3 November 2017
Accepted 15 November 2017
Available online 15 November 2017

Keywords:
Bayesian calibration
Uncertainty quantification
High dimensional model representation
Stochastic collocation
RANS transition model

ABSTRACT

An efficient Bayesian uncertainty quantification approach is proposed, which combines the adaptive high dimensional model representation technique (HDMR) and stochastic collocation (SC) method based generalized polynomial chaos (gPC) to construct the surrogate for sampling procedure in Bayesian calibration step. Specifically, the adaptive HDMR technique is used to decompose the original high dimensional problems into several lower-dimensional subproblems, which are subsequently solved with the gPC-based SC method. Then the Bayesian calibration and prediction are carried out with the so-constructed surrogate model. A new indicator based on the variance of the corresponding component function is employed to identify the important components of the HDMR, instead of the original one based on the impact on the output mean, as the input parameters that can be well informed in the inverse problem are the ones that the model output is sensitive to. Further, a rigorous convergence study of the approximate posterior to the true posterior is carried out for the proposed approach. Its applications to both a simple mathematical function and a complex fluid dynamic model, i.e. k-ω-γ transition model, are investigated, demonstrating both its efficiency and accuracy. In the application to k-ω-γ transition model, the results show not only a quantified uncertainty overlapping well with the experimental data, but also a great improvement of the match between the prediction mean and the experimental data, which may be due to the further account of the intermittency through the spread of the model parameters.

1. Introduction

Hypersonic boundary layer transition is not only of fundamental interest in fluid dynamics, it is also of great practical relevance in the design of many aerodynamic configurations at hypersonic speeds. Despite the rapid development of Direct numerical simulation (DNS) technique, Reynolds-averaged-Navier-Stokes (RANS) model and empirical e^N method are still the main tools for transition predictions in engineering applications, due to its affordability compared to DNS. A local-variable-based RANS model, namely k-ω-γ model, has been proposed recently, which can successfully simulate three-dimensional (3-D) high-speed aerodynamic flow transition with a reasonably wide range of Mach numbers [1,2]. However, flow transition to turbulence is a very complex process, involving receptivity process, linear modal growth, mode interaction, final breakdown to turbulence etc. [3], which cannot be correctly simulated by RANS model. Transition prediction with RANS model is highly unreliable and the aim of this work is to quantify the uncertainty of the quantity of interest (QoI) in the hypersonic boundary layer transition simulations with k-ω-γ model.

In the pioneering work of Kennedy and O'Hagan [4], the Bayesian calibration technique for a general computer model is presented and the uncertainties are classified into parameter uncertainty, model inadequacy, residual variability, parametric uncertainty, observation error and code uncertainty. The uncertainty of RANS predictions mainly comes from the first two sources, parameter uncertainty and model inadequacy. The former represents the uncertainty due to the lack of knowledge of the model parameters and the latter represents the discrepancy between true physical observation and model output at optimal model parameters. A number of studies have focused on the parameter uncertainty. Cheung et al. [5] have applied Bayesian uncertainty analysis to Spalart-Allmaras (SA) turbulence model for wall-bounded incompressible turbulent flow at variable pressure gradients. They employed three different stochastic models for inadequacy terms and compared them in terms of model plausibility and prediction of QoIs. Oliver and Moser [6] extended the work of Cheung et al. by considering four stochastic extensions of four eddy viscosity turbulence models. They proposed a more complex stochastic model to take account of the multi-scale structure of the boundary layer. In

* Corresponding author.
E-mail addresses: jc-zhang15@mails.tsinghua.edu.cn (J. Zhang), fs-dem@tsinghua.edu.cn (S. Fu).

https://doi.org/10.1016/j.compfluid.2017.11.007
0045-7930/

Ref. [7], Edeling et al. estimated the parameter variability within and across the scenarios (i.e. at different pressure gradients) under the Bayesian framework. Further, they [8] utilized Bayesian Model-Scenario Averaging approach to synthesize the results of 5 turbulence models at 14 scenarios, resulting a substantial improvement in both prediction mean and variance. Bayesian parameter estimation was also used for other flow configurations, e.g. Jet-in-Crossflow [9] etc. Besides, parameter uncertainty of RANS models has also been assessed in Ref. [10], with the input uncertainty imposed through a prior distribution, based on an extensive literature survey about the parameter dispersion.

Besides parameter uncertainty mentioned above, model-inadequacy is also a major source of uncertainty in RANS prediction and quantifying and reducing this uncertainty have raised the research interests recently in the field. Dow and Wang [11] employed the Bayesian approach to infer the turbulent viscosity from DNS data. In Ref. [12], Emory et al. proposed an approach to quantify the uncertainty directly through Reynolds stress. In Ref [13], Gorle and Iaccarino carried out uncertainty quantification of turbulent scalar flux models, taking account of the uncertainty directly through Reynolds stress. In Ref. [14], Duraisamy et al. proposed a data-driven approach for turbulence and transition modeling, which consists mainly of injecting the functional form of deficiencies inferred by experimental data into simulations to obtain more accurate predictions. A data-driven, physics-informed Bayesian approach has been proposed recently by Xiao et al. [15], taking account of the model-form uncertainty directly through Reynolds stress and an iterative ensemble Kalman method was used to incorporate the prior knowledge and the experimental data.

The RANS turbulence models have been the main focus of the previously mentioned work. The research of uncertainty quantification for RANS transition modeling is rather limited. An exception is the work of Pecnik et al. [16], in which they applied UQ for laminar-turbulent transition in turbo-machinery configurations, using the $\gamma - \tilde{Re}_{\theta t}$ model of Menter et al. [17], but only a forward uncertainty propagation is carried out and the input uncertainty is imposed through a prior distribution. Bayesian uncertainty analysis provides a rigorous approach to quantify the uncertainty arising from the mathematical modeling and simulation, and to incorporate the prior knowledge and experimental data systematically, through Bayesian data updating. Thus in this work, we apply the Bayesian framework to quantify the uncertainty arising from the k-ω-γ transition model in hypersonic transition simulations. We focus on the parameter uncertainty and the model inadequacy is simply termed as a multiplicative Gaussian random variable, as in the work [5]. Modeling the inadequacy terms requires more physical insight of transition process and is a RANS modeling issue rather than a UQ of an existed model. In this work we restrict ourselves to the latter issue and treat k-ω-γ model as a black box.

A key step in this framework is the Bayesian calibration. After identifying the prior distribution of the input parameters and constructing the stochastic model, the posterior distributions of the model parameters are obtained through Bayes' rule. This procedure usually requires a sampling method, e.g. Markov chain Monte-Carlo (MCMC) [18]. A large number of model evaluations, typically tens of thousands, are required in the sampling procedure, which are computationally expensive. A number of methods exist in the literature to reduce the computational cost while retaining the non-intrusiveness of the corresponding approach, e.g. [19,20]. In Ref. [21], Ma and Zabaras proposed an adaptive version of high dimensional model representation technique (HDMR) to decompose the original high dimension problem into lower dimension subproblems and solved them with the adaptive sparse grid collocation (ASGC) method [22] they proposed previously. The efficiency of this approach is demonstrated with some mathematical functions and also with a set of fluid-mechanic problems. In Ref. [23] Edeling et al. improved the original Simplex-stochastic collocation (SSC) [20] method and also combined it with the adaptive HDMR technique, resulting in an improved scalability. They applied this approach in a nozzle and an airfoil flow. These approaches have only been applied in the forward problem and their application to inverse problems hasn't been explored yet. Marzouk and Xiu [24] proposed a stochastic collocation approach to Bayesian inference in inverse problems and conducted a rigorous error analysis for the approximate posterior. Several examples were carried out to demonstrate the efficiency of the proposed method, including a Burgers' equation case and a genetic toggle switch case in biology.

In this paper we combine the adaptive high-dimensional stochastic model representation (HDMR) technique [21] with the stochastic collocation (SC) approach based on generalized polynomial chaos (gPC) [24], to construct the surrogate model. Then this surrogate model is used for Bayesian inference in the inverse problem. This idea is inspired by the work of Ma and Zabaras [21], in which they combined the HDMR technique with the adaptive sparse grid collocation (ASGC) [22] method to solve the forward problem. The proposed approach can be seen as an extension of the stochastic collocation approach proposed by Marzouk and Xiu [24], by integrating it into the high dimensional model representation framework.

The paper is organized as follows: the Bayesian uncertainty quantification framework is described in Section 2. In Section 3 the surrogate model construction approach is described, including the gPC-based stochastic collocation method and the HDMR technique. The algorithm is summarized in Section 4. In Section 5 we demonstrate the accuracy and efficiency of the proposed method through a simple mathematical function. A comparison between the proposed method and the exact model is given. After testing our approach with this simple mathematical function, we apply the approach to k-ω-γ transition modeling in hypersonic transition simulations in Section 6. The results are given in Section 6.4, including both the posterior distribution of the input parameters and the prediction mean with quantified uncertainty. Finally the conclusion is drawn in Section 7.

2. Bayesian uncertainty quantification framework

2.1. General review

This part provides a brief description of the UQ framework, following the work of [5,7]. The main steps are the specification of the flow class and quantity of interest (QoI), the collection of experimental data, the construction of the stochastic model, the Bayesian calibration, and validation and prediction. As is pointed out in Ref. [5], whether a model is considered valid or not depends on its ability to predict the QoIs to the required accuracy and precision, rather than to predict all aspects of the physical world. Thus the identification of the QoIs is a key issue and should be kept in mind during the whole uncertainty quantification process. In this work, we assume the QoIs are observable for the corresponding experiments and thus we use the observation of QoIs as the data to inform the model parameters.

In Bayesian framework, various forms of uncertainty, whether aleatoric or epistemic, are all represented through probability. Thus we can characterize the input parameter uncertainty by their probability density function (PDF). In the Bayesian calibration step, the posterior distributions of the parameters are obtained through Bayes' rule:

$$p(\mathbf{z}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{d})} \tag{1}$$

where random vector $\mathbf{z}$ represents the model parameter and $\mathbf{d}$ the experimental observation. Since the denominator doesn't depend on $\mathbf{z}$, we can omit it and obtain the un-normalized version of Bayes' rule:

$$p(\mathbf{z}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{z})p(\mathbf{z}) \tag{2}$$

Here the prior distribution p($\mathbf{z}$) can be specified based on prior knowledge and the calculation of the likelihood p($\mathbf{d}|\mathbf{z}$) requires the RANS model output and the experimental observation, combined by the constructed stochastic model.

After calibration, the prediction step can be done by propagating the PDF of the input parameters through the simulation code to obtain the PDF of the QoIs. The so-obtained PDF is actually the posterior distribution conditional on the experimental observation:

$$\begin{aligned} p(\tilde{\mathbf{d}}|\mathbf{d}) &= \int p(\tilde{\mathbf{d}}, \mathbf{z}|\mathbf{d})d\mathbf{z} = \int p(\tilde{\mathbf{d}}|\mathbf{d}, \mathbf{z})p(\mathbf{z}|\mathbf{d})d\mathbf{z} \\ &= \int p(\tilde{\mathbf{d}}|\mathbf{z})p(\mathbf{z}|\mathbf{d})d\mathbf{z} \end{aligned} \tag{3}$$

This derivation follows from Ref. [7] and the last step in Eq. (3) follows by assuming $\tilde{\mathbf{d}}$ and $\mathbf{d}$ are conditionally independent given $\mathbf{z}$.

2.2. Construction of the stochastic model

The stochastic model needs to be specified to calculate the likelihood function, and solve the inverse problem. Here we construct the stochastic model by accounting for the model inadequacy simply through a multiplicative Gaussian random variable:

$$\tilde{\mathbf{d}} = (1 + \boldsymbol{\eta})\mathscr{M}(\mathbf{x}, \mathbf{z}) \tag{4}$$

where $\boldsymbol{\eta}$ is a random vector with each component η_i as zero mean, independent and identically distributed Gaussian: i.e. $\eta_i \sim \mathscr{N}(0, \sigma^2)$. $\mathscr{M}(\mathbf{x}, \mathbf{z})$ is the output of QoIs from our simulation code, depending on the explanatory variable $\mathbf{x}$ (e.g. Mach number, Reynolds number etc.) and the model parameter $\mathbf{z}$. $\tilde{\mathbf{d}}$ represents the true process and can be related to the experimental observation $\mathbf{d}$ as:

$$\mathbf{d} = \tilde{\mathbf{d}} + \mathbf{e} \tag{5}$$

Here $\mathbf{e}$ represents the measurement error, which is modeled as a zero mean, independent and identically distributed Gaussian, i.e. $e_i \sim \mathscr{N}(0, \sigma_e^2)$. σ_e is determined from the corresponding experiments. Thus from Eqs. (4) and (5) we can relate the model output to the experimental observation, and we can obtain:

$$\mathbf{d}|\sigma, \mathbf{z} \sim \mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\lambda}) \tag{6}$$

where

$$\boldsymbol{\mu} = \mathscr{M}(\mathbf{x}, \mathbf{z}) \text{ and } \boldsymbol{\lambda} = \mathscr{M}^{\mathrm{T}}(\mathbf{x}, \mathbf{z})\sigma^2\mathscr{M}(\mathbf{x}, \mathbf{z}) + \sigma_e^2\boldsymbol{I} \tag{7}$$

Finally we denote $\boldsymbol{\theta} = (\mathbf{z}, \sigma)$ as the parameter to be calibrated and the likelihood can be written as:

$$p(\mathbf{d}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^{N_d}|\boldsymbol{\lambda}|}} \exp\left(-\frac{1}{2}\boldsymbol{\delta}^T\boldsymbol{\lambda}^{-1}\boldsymbol{\delta}\right) \tag{8}$$

where N_d is the dimension of the experimental observation, $|\boldsymbol{\lambda}|$ represents the determinant of $\boldsymbol{\lambda}$, and $\boldsymbol{\delta} = \boldsymbol{d} - \mathscr{M}(\mathbf{x}, \mathbf{z})$.

2.3. Bayesian calibration and posterior model check

After constructing the stochastic model, we can carry out the calibration process with the likelihood function calculated by Eq. (8) and also an appropriate prior proposed based on the prior knowledge of the model parameters. Here we use the uniform distribution for all the model parameters and parameter range is discussed in detail in Section 6.1. For sampling we employ an adaptive Metropolis-Hastings MCMC sampler [25], as implemented in the R [26] package MHadaptive [27]. The posterior density function is approximated by the simple kernel estimation. After calibration, a model check is carried out, which involves propagation of the parameter posterior through the model to obtain the distribution of the QoIs. The Bayesian prediction involves the same procedure, but with a new unobserved case specification. For simplicity, the samples obtained in the calibration step are reused in the model check step.

3. Surrogate model construction

With the prior distribution specified by the modeler and the likelihood function calculated by Eq. (8), we can recast Eq. (2) as:

$$p(\boldsymbol{\theta}|\mathbf{d}) \propto \frac{1}{\sqrt{(2\pi)^{N_d}|\boldsymbol{\lambda}|}} \exp\left(-\frac{1}{2}\boldsymbol{\delta}^T\boldsymbol{\lambda}^{-1}\boldsymbol{\delta}\right)p(\boldsymbol{\theta}) \tag{9}$$

Then an MCMC sampling is employed to obtain samples from the posterior. This procedure involves a large number of evaluations of $\mathscr{M}(\mathbf{x}, \mathbf{z})$, which is computationally prohibitive. Thus the construction of a surrogate model for the forward problem is usually needed. In this work, we employ the high dimensional model representation technique (HDMR) to decompose the original moderately high dimensional problems into several subproblems, which are solved by the stochastic collocation method based on generalized polynomial chaos (gPC). The resulting surrogate model is then used for the Bayesian inference in inverse problems. The novelties of this work are: (1) the combination of the gPC-based SC method and the HDMR technique, which is inspired by Ma and Zabaras [21]; (2) the application of the so-constructed surrogate model to Bayesian inverse problems, inspired by the work of Marzouk and Xiu [24].

3.1. Stochastic collocation method based on gPC

In this section we describe the stochastic collocation method based on generalized polynomial chaos, following the work of Marzouk and Xiu. The interested reader may refer to [24] for further details.

In the gPC-based stochastic collocation method, $\mathscr{M}(\mathbf{z})$ (we omit the explanatory variable $\mathbf{x}$ hereafter for brevity) is approximated by:

$$\mathscr{M}_{N_p}(\mathbf{z}) = \sum_{|\mathbf{i}|=0}^{N_p} \tilde{\boldsymbol{a}}_{\mathbf{i}}\Phi_{\mathbf{i}}(\mathbf{z}) \tag{10}$$

where $\mathbf{i} = (i_1, \ldots, i_{n_z})$ is the multi-index with $n_z = dim(\mathbf{z})$, $|\mathbf{i}| = i_1 + i_2 + \ldots + i_{n_z}$, and the orthogonal basis functions $\Phi_{\mathbf{i}}(\mathbf{z})$ is defined as:

$$\Phi_{\mathbf{i}}(\mathbf{z}) = \phi_{i_1}(z_1)\ldots\phi_{i_{n_z}}(z_{n_z}) \tag{11}$$

where $\phi_m(z_k)$ is the mth degree one-dimensional normalized orthogonal polynomial in the z_k direction, with the corresponding polynomial type determined by the distribution $p(z_k)$. And

$$\tilde{\boldsymbol{a}}_i = \sum_{m=1}^{Q} \mathscr{M}(\mathbf{z}^{(m)})\Phi_{\mathbf{i}}(\mathbf{z}^{(m)})\omega^{(m)} \tag{12}$$

where $\{\mathbf{z}^{(m)}\}$ are the nodal set at which we need to evaluate the model output and $\omega^{(m)}$ is the corresponding weights of the cubature rules. A detailed discussion of determining the polynomial type and the corresponding weight can be found in Ref. [28]. Here in our work the Smolyak algorithm [29] based on one-dimensional Clenshaw-Curtis quadrature rule is employed and the R package GPC [30] is used for constructing the polynomials and evaluating the corresponding $\omega^{(m)}$. It is to be noted that to evaluate the surrogate $\tilde{\mathscr{M}}_{N_p}(\mathbf{z})$, only Q code runs are required and the computational burden is greatly reduced.

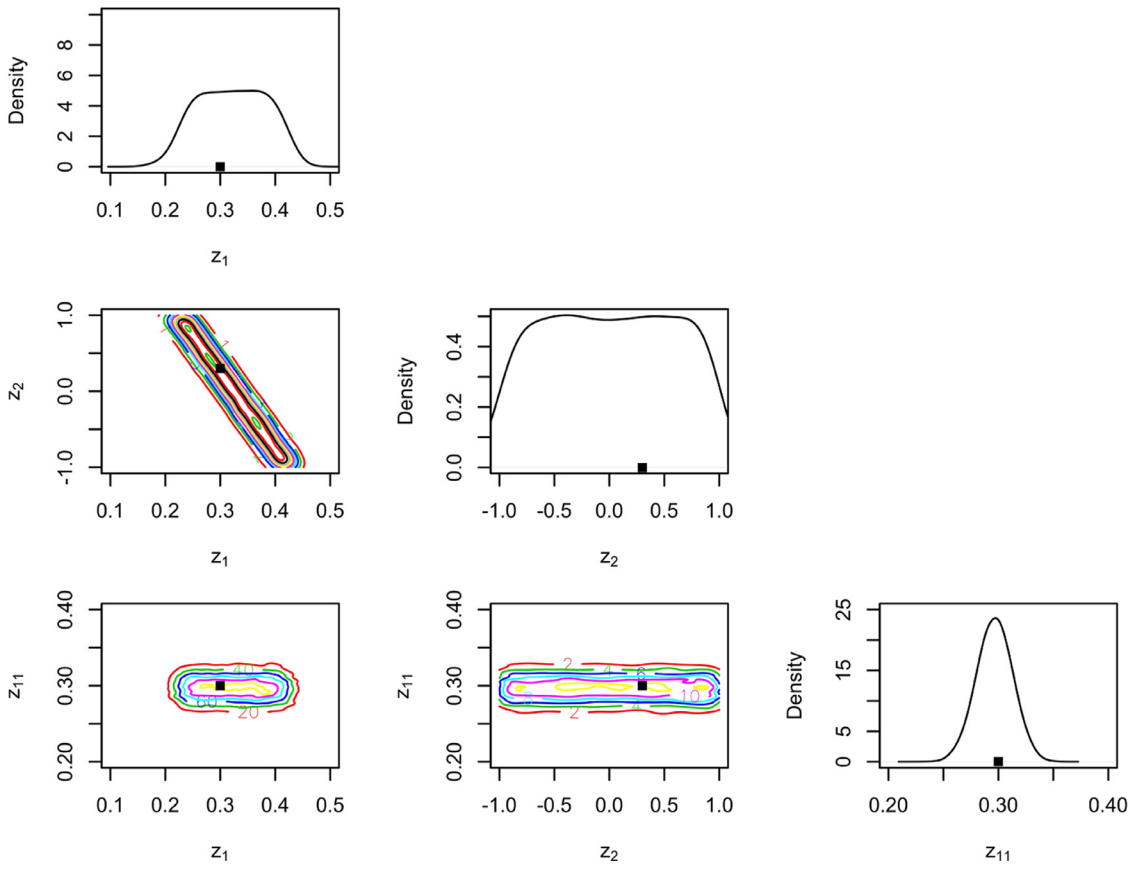


Fig. 1. The results by using the exact model. Only the posteriors of the first three most important dimensions are shown. The points in the figure represent the true values of the parameters.

Table 1
The number of code runs required, for sparse grid based on Smolyak algorithm [29], with various parameter dimensions N_z and sparse grid levels l.

	l = 2	l = 3	l = 4	l = 5	l = 6
$N_z = 2$	5	13	29	65	145
$N_z = 3$	7	25	69	177	441
$N_z = 4$	9	41	137	401	1105
$N_z = 5$	11	61	241	801	2433
$N_z = 6$	13	85	389	1457	4865

3.2. High dimensional model representation

The use of gPC-based stochastic collocation method can reduce the number of code runs greatly, provided that the input dimension is low. The so-called curse-of-dimensionality results in an exponential increase in the required number of runs with dimension, leading to a computational bottleneck for high dimensional problems. The number of the code runs required is shown in Table 1 for Smolyak sparse grid [29], as a function of parameter dimension and sparse grid level. As can be seen, even for a moderately high dimensional problem, e.g. $l = 4$ and $N_z = 6$, 389 code runs are needed, which demands great computational efforts. Thus we employ the high dimensional model representation technique (HDMR) to decompose the original moderately high dimensional problems into several subproblems with lower dimension to further reduce the number of code runs. The adaptive high dimensional stochastic model representation technique (HDMR) proposed by Ma and Zabaras is employed and a brief description of this technique is given here. The derivation below follows closely from Maand Zabaras and the interested reader may refer to Ref. [21] for further details.

The high dimensional stochastic model representation of $\mathscr{M}(\mathbf{z})$ is defined as:

$$\mathscr{M}(\mathbf{z}) = f_0 + \sum_{i=1}^{N_z} f_i(z_i) + \sum_{1\le i_1<i_2\le N_z} f_{i_1 i_2}(z_{i_1}, z_{i_2}) + \dots + \sum_{1\le i_1<\dots<i_s\le N_z} f_{i_1\dots i_s}(z_{i_1}, \dots, z_{i_s}) + \dots + f_{1,2,\dots,N_z}(z_1, \dots, z_{N_z}) \tag{13}$$

Here the zeroth-order term f_0 represents the mean effects, the 1st-order term $f_i(z_i)$ the effects of the individual input parameter z_i, the 2nd-order term $f_{i_1 i_2}(z_{i_1}, z_{i_2})$ the correlation effects of the input parameter z_{i_1} and z_{i_2}. And the higher order terms represent the corresponding correlation effects of the underlying multiple parameters. The Eq. (13) can be written in a more compact form, as in Refs. [21,31]:

$$\mathscr{M}(\mathbf{z}) = \sum_{\mathbf{u}\subset\mathscr{D}} f_{\mathbf{u}}(z_{\mathbf{u}}) \tag{14}$$

where $\mathscr{D} = \{1, 2, \dots, N_z\}$, $f_{\mathbf{u}}(z_{\mathbf{u}}) = f_{i_1\dots i_s}(z_{i_1}, \dots, z_{i_s})$ for $\mathbf{u} = \{i_1, i_2, \dots, i_s\}$ and by convention $f_\emptyset(z_\emptyset) = f_0$. Then the component functions are determined by minimizing the error functional:

$$\int \left[\mathscr{M}(\mathbf{z}) - \sum_{\mathbf{u}\subset\{0,1,\dots,s\}} f_{\mathbf{u}}(z_{\mathbf{u}})\right]^2 d\mu(\mathbf{z}) \tag{15}$$

where $0 \le s \le N_z$. The measure $d\mu(\mathbf{z})$ determines the corresponding projection operator:

$$P_{\mathbf{u}}\mathscr{M}(\mathbf{z}) = \int \mathscr{M}(\mathbf{z}) d\mu_{\mathscr{D}\setminus\mathbf{u}}(\mathbf{z}) \tag{16}$$

where $d\mu_{\mathscr{D}\setminus\mathbf{u}}(\mathbf{z}) = \prod_{i\in\mathscr{D}, i\notin\mathbf{u}} d\mu_i(z_i)$ with $d\mu_i(z_i)$ the marginal density of z_i. Thus minimizing functional (15) results in :

$$f_{\mathbf{u}}(z_{\mathbf{u}}) = P_{\mathbf{u}}\mathscr{M}(\mathbf{z}) - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}}(z_{\mathbf{v}}) \tag{17}$$

And it can be easily verified that the resulting component functions are orthogonal with the inner product defined by the measure $d\mu(\mathbf{z})$.

The measure $d\mu(\mathbf{z})$ remains to be determined to obtain the component functions. To avoid the integral evaluation in Eq. (16), the measure $d\mu(\mathbf{z})$ is simply chosen as $\prod_{i=1}^{N_z}\delta(z_i - \bar{z}_i)$ in the CUT-HDMR approach, where $\{\bar{z}_i\}$ is the reference point. Thus the projection operator becomes:

$$P_{\mathbf{u}}\mathscr{M}(\mathbf{z}) = \mathscr{M}(\mathbf{z})|_{\mathbf{z}=\bar{\mathbf{z}}\setminus z_{\mathbf{u}}} \tag{18}$$

and the component functions of the CUT-HDMR can be obtained as:

$$\begin{aligned} &f_0 = \mathscr{M}(\bar{\mathbf{z}}),\ \ f_i(z_i) = \mathscr{M}(\bar{z}_1,\ldots,z_i,\ldots,\bar{z}_{N_z}) - f_0 \\ &f_{i,j}(z_i,z_j) = \mathscr{M}(\bar{z}_1,\ldots,z_i,\ldots,z_j,\ldots,\bar{z}_{N_z}) - f_i(z_i) - f_j(z_j) - f_0 \end{aligned} \tag{19}$$

Thus finally $\mathscr{M}(\mathbf{z})$ can be obtained from Eqs. (14) and (19) as:

$$\mathscr{M}(\mathbf{z}) = \sum_{\mathbf{u}\subset\mathscr{D}} f_{\mathbf{u}}(z_{\mathbf{u}}) = \sum_{\mathbf{u}\subset\mathscr{D}}\sum_{\mathbf{v}\subset\mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|}\mathscr{M}(z_{\mathbf{v}})|_{\mathbf{z}=\bar{\mathbf{z}}\setminus z_{\mathbf{v}}} \tag{20}$$

The integration of HDMR and gPC-based SC comes straightforward from Eqs. (20) and (10):

$$\mathscr{M}(\mathbf{z}) \approx \mathscr{M}_S(\mathbf{z}) = \sum_{\mathbf{u}\subset\mathscr{D}}\sum_{\mathbf{v}\subset\mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|}\tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}})|_{\mathbf{z}=\bar{\mathbf{z}}\setminus z_{\mathbf{v}}} \tag{21}$$

Instead of truncating the Eq. (21) to a certain order for all the dimensions, the adaptive HDMR proposed by Ma and Zabaras first identifies the important dimensions, then only the correlations of the important dimensions are considered. The weight of each dimension is defined as:

$$\eta_i = \frac{\|J_{\{i\}}\|_{L_2}}{\|J_{\emptyset}\|_{L_2}} \tag{22}$$

where

$$J_{\mathbf{u}} = \sum_{\mathbf{v}\subset\mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|} E[\tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}})|_{\mathbf{z}=\bar{\mathbf{z}}\setminus z_{\mathbf{v}}}] \tag{23}$$

and the L_2 norm is defined in the model output space. A predefined error threshold ϵ_1 is needed to identify the important dimension. Similar dimension adaptivity is extended to high order correlation terms, as:

$$\eta_{\mathbf{u}} = \frac{\|J_{\mathbf{u}}\|_{L_2}}{\|\sum_{\mathbf{v}\in\mathscr{S},|\mathbf{v}|<|\mathbf{u}|} J_{\mathbf{v}}\|_{L_2}} \tag{24}$$

where $\mathscr{S}$ is the set of index whose component function has already been computed. Only the term $f_{\mathbf{u}}(z_{\mathbf{u}})$ with a weight $\eta_{\mathbf{u}}$ greater than the threshold ϵ_1 is kept and we put all the corresponding index $\mathbf{u}$ into a set $\mathscr{T}$. Then for higher order terms, only component functions $f_{\mathbf{u}}(z_{\mathbf{u}})$ where $\mathbf{u}$ satisfies the admissibility relation (25), are calculated.

$$\mathbf{u}\in\mathscr{T} \text{ and } \mathbf{v}\subset\mathbf{u} \Rightarrow \mathbf{v}\in\mathscr{T} \tag{25}$$

Further a relative error ρ between two consecutive orders p and $p-1$ is defined as:

$$\rho = \frac{\|\sum_{\mathbf{v}\in\mathscr{T},|\mathbf{v}|\le p} J_{\mathbf{v}} - \sum_{\mathbf{v}\in\mathscr{T},|\mathbf{v}|\le p-1} J_{\mathbf{v}}\|_{L_2}}{\|\sum_{\mathbf{v}\in\mathscr{T},|\mathbf{v}|\le p-1} J_{\mathbf{v}}\|_{L_2}} \tag{26}$$

A threshold ϵ_2 is predefined to decide whether the adaptive HDMR construction process is converged.

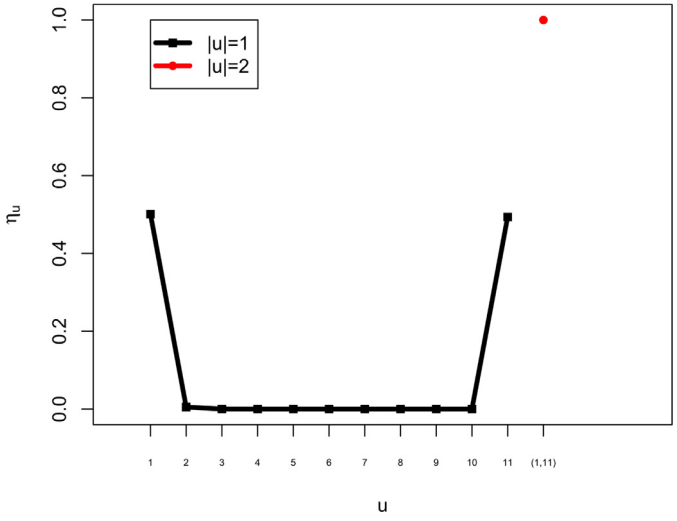


Fig. 2. The relative weights of each dimension, calculated with $l = 5$.

3.2.1. The weight of each dimension

The combination of the HDMR technique and gPC-based SC method can be applied in forward problems without modification, but in inverse problems the determination of the weight should be modified. One should note that the weight of each dimension determined by Eqs. (22) and (23) is an indicator of the impact of the corresponding input parameter on the output mean, thus a faster convergence and a smaller error in terms of the output mean can be achieved, as demonstrated in Ref. [21]. But for inverse problems, the input parameters that can be well informed are the ones the model output is sensitive to, as demonstrated in Ref. [7] for the inference of the model parameter in $k-\epsilon$ turbulence model. Thus in this work, for inverse problems, the relative weight of each dimension is determined by:

$$\eta_i = \frac{\|J_{\{i\}}\|_{L_2}}{\sum_{i=1}^{N_d}\|J_{\{i\}}\|_{L_2}} \tag{27}$$

where

$$J_{\mathbf{u}} = \mathbf{var}[f_{\mathbf{u}}(z_{\mathbf{u}})] = \mathbf{var}\left[\sum_{\mathbf{v}\subset\mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|}\tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}})|_{\mathbf{z}=\bar{\mathbf{z}}\setminus z_{\mathbf{v}}}\right] \tag{28}$$

Similarly the relative importance for the high order correlation terms is defined as:

$$\eta_{\mathbf{u}} = \frac{\|J_{\mathbf{u}}\|_{L_2}}{\sum_{|\mathbf{v}|=|\mathbf{u}|}\|J_{\mathbf{u}}\|_{L_2}} \tag{29}$$

where $J_{\mathbf{u}}$ is defined as in Eq. (28). The relative error ρ is expressed as:

$$\rho = \frac{\|\mathbf{var}[\sum_{\mathbf{v}\in\mathscr{T},|\mathbf{v}|\le p} f_{\mathbf{u}}(z_{\mathbf{u}})] - \mathbf{var}[\sum_{\mathbf{v}\in\mathscr{T},|\mathbf{v}|\le p-1} f_{\mathbf{u}}(z_{\mathbf{u}})]\|_{L_2}}{\|\mathbf{var}[\sum_{\mathbf{v}\in\mathscr{T},|\mathbf{v}|\le p} f_{\mathbf{u}}(z_{\mathbf{u}})]\|_{L_2}} \tag{30}$$

4. Algorithm summary

In this section, we summarize the Bayesian uncertainty quantification approach that we proposed above. The convergence study of the proposed approach can be found in Appendix A.

Remark. As the sparse nodal set is nested, every model evaluation in the previous level can be used in the subsequent calculations, rendering the current approach very efficient.

Algorithm 1 The proposed uncertainty quantification approach.

1: Identify the model input-output mapping $\mathscr{M}(\mathbf{z})$ and collect the corresponding experimental data $\mathbf{d}$. Specify the prior $p(\mathbf{z})$.
2: Set the set $\mathscr{A}_i = \emptyset$ as the set of the nodes where the model evaluation has been done and $\mathscr{B}_i = \emptyset$ as the set of corresponding model output. Set sparse grid level l, the gPC order $N_p = l - 1$.
3: **repeat**
4: Construct adaptively the HDMR surrogate model $\mathscr{M}_S^{N_p}(\mathbf{z})$ of $\mathscr{M}(\mathbf{z})$ with each subproblem solved with gPC-based SC of order N_p. In this procedure model evaluations are carried out only for the nodes that are not in $\mathscr{A}_i$. Add the evaluated nodal points z_e into $\mathscr{A}_i$ and store the corresponding $\mathscr{M}(z_e)$ into $\mathscr{B}_i$.
5: Construct the stochastic model to obtain the likelihood $p(\mathbf{d}|\mathbf{z})$ and combine with the prior $p(\mathbf{z})$ to obtain the posterior $p(\mathbf{z}|\mathbf{d})$ through Bayes' rule.
6: $l \leftarrow l + 1$, $N_p \leftarrow l - 1$
7: **until** the posterior converges.
8: The posterior is forward propagated through the surrogate model $\mathscr{M}_S(\mathbf{z})$ to obtain the model output with quantified uncertainty.

5. Application to a simple mathematical function

Here we apply the proposed approach to a simple mathematical function to demonstrate its accuracy and efficiency. The proposed approach is expected to have great advantage in strongly anisotropic cases, thus we propose the following function, inspired by the Example 3 in Ref. [21] :

$$\mathscr{M}(\mathbf{z}) = \left(\frac{1 + \alpha_{11} z_{11}}{1 + \sum_{i=1}^{10} \alpha_i z_i}, \frac{1 + \alpha_{11} z_{11}}{1 - \sum_{i=1}^{10} \alpha_i z_i} \right) \tag{31}$$

where $\alpha_i = \frac{0.1}{10^{i-1}}$ for $i = 1, \ldots, 10$ and $\alpha_{11} = 0.1$. The prior distribution of each model parameter is assumed as:

$$z_i = \mathscr{U}(-1, 1) \text{ for i=1,...,11} \tag{32}$$

The model inadequacy term is neglected and the posterior is expressed as:

$$p(\mathbf{z}|\mathbf{d}) \propto \frac{1}{\sqrt{(2\pi)^{N_d} |\boldsymbol{\lambda}|}} \exp\left(-\frac{1}{2} \boldsymbol{\delta}^T \boldsymbol{\lambda}^{-1} \boldsymbol{\delta} \right) p(\mathbf{z}) \tag{33}$$

where the dimension of the experimental observation $N_d = 2$ and $\boldsymbol{\lambda} = diag(\sigma_e^2)$ where $\sigma_e = 0.005$. Here the experimental observation is artificially generated by:

$$\mathbf{d} = \mathscr{M}(\mathbf{z}_{exact}) + \mathbf{e} \tag{34}$$

where $z_{exact_i} = 0.3$ for $i = 1, .., 11$ and $e_i \sim \mathscr{N}(0, \sigma_e)$ for $i = 1, 2$. Five experimental observations so-generated are used in the results presented in this paper. To obtain the posterior distribution, 3×10^5 MCMC samples are generated, which are enough to guarantee the convergence of the MCMC chains. The "burn-in" length is set to 10^4. First the posterior is obtained by using the exact model $\mathscr{M}(\mathbf{z})$ and the results are shown in Fig. 1. Here we only show the posterior of the parameter z_1, z_2, z_{11}, as all the other parameters have little impact on the model output and thus haven't been well informed. As can be seen, z_1 and z_{11} are well informed and z_2 remains almost the same as its prior. The results by using the surrogate model constructed following Algorithm 1 are shown in Figs. 2 and 3, with sparse grid level from 2 to 5. The weights of each dimension calculated with different grid levels are almost the same and only the weights calculated with $l = 5$ are given in Fig. 2. As can be seen, z_1 and z_{11} are the first two most important dimensions and account for 99.5% of the total variance of the first order terms, thus only f_0, $f_1(z_1)$, $f_{11}(z_{11})$, $f_{(1,\,11)}(z_1, z_{11})$ are included in the HDMR of $\mathscr{M}(\mathbf{z})$. Then each subproblem is solved with a fix-order gPC-SC method, and the results are shown in Fig. 3. As can be seen, $l = 3$ is sufficient to capture the corresponding posterior. The Maximum A Posteriori (MAP) values are very close to the exact values of the parameters. However, a discrepancy of the posterior of z_1 is observed between the results of exact model and surrogate model. This is due to the neglect of the other model parameters, i.e. z_2, which are strongly correlated with z_1, as can be seen from Fig. 1. Further inclusion of terms in the HDMR will result in a better comparison with the posterior obtained with the exact model, as can be seen from Fig. 4.

To be noted, here the reference values used in CUT-HDMR $\bar{\mathbf{z}}$ is zero, which is different from its exact value 0.3. This is done on purpose to demonstrate the ability of the present approach even when the nominal values of the model parameters aren't optimal, as is often encountered in practical situations. It is also worth noting that if the weight of each dimension is determined by the impact of the corresponding input parameter on the output mean, z_{11} will be neglected, which is in fact one of the most important dimensions. After successfully testing the proposed approach with this simple mathematical function, we apply it to a complex and time-consuming computer model, i.e., k-$\omega - \gamma$ transition model, expecting to improve the prediction mean and also obtain the corresponding quantified uncertainty.

6. Application to *k-ω-γ* transition modeling

6.1. Transition model formulation

In this section, the k-ω-γ transition model proposed by Fu and Wang is briefly described. The interested reader may refer to Ref. [1] for further details.

This model is based on a k-ω-γ three-equation eddy-viscosity concept with k representing the fluctuating kinetic energy, ω the specific dissipation rate and γ the intermittency factor. The equations are formulated as:

$$\frac{\partial(\rho k)}{\partial t} + \frac{\partial(\rho u_j k)}{\partial x_j} = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_{eff}}{\sigma_k} \right) \frac{\partial k}{\partial x_j} \right] + P_k(\mu_{eff}) - \epsilon \tag{35}$$

$$\frac{\partial(\rho \omega)}{\partial t} + \frac{\partial(\rho u_j \omega)}{\partial x_j} = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_{eff}}{\sigma_\omega} \right) \frac{\partial \omega}{\partial x_j} \right] + P_\omega - D_\omega + Cd_\omega \tag{36}$$

$$\frac{\partial(\rho \gamma)}{\partial t} + \frac{\partial(\rho u_j \gamma)}{\partial x_j} = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_{eff}}{\sigma_\gamma} \right) \frac{\partial \gamma}{\partial x_j} \right] + P_\gamma(F_{onset}) - \epsilon_\gamma \tag{37}$$

Here the effective viscosity μ_{eff} is the weighted sum of the non-turbulent part and the turbulent part, i.e. :

$$\mu_{eff} = (1 - \gamma)\mu_{nt} + \gamma \mu_t \tag{38}$$

The non-turbulent part μ_{nt} can be further expressed as:

$$\mu_{nt} = C_6 \times \bar{\rho} k \tau_{nt} \tag{39}$$

Here τ_{nt} represents the characteristic timescale in the flow transition, as associated with different instabilities including 1st Mack mode, 2nd Mack mode etc. In this work we consider only the 1st Mack mode and 2nd Mack mode and the timescale is formulated as:

$$\tau_{nt} = \tau_{nt1} + \tau_{nt2} \times \frac{1}{2}[1 + sgn(M_{rel} - 1)] \tag{40}$$

where $\tau_{nt1} = C_4 \times \xi_{eff}^{1.5} / [(2E_u)^{0.5} \nu]^{0.5}$, $\tau_{nt2} = C_5 \times 2\xi_{eff} / U(y_s)$, and $M_{rel} = (U - c_r)/a$. The effective length scale ξ_{eff} is set as:

$$\xi_{eff} = min(C_1 \times \xi, C_2 \times l_t, C_3 \times l_B) \tag{41}$$

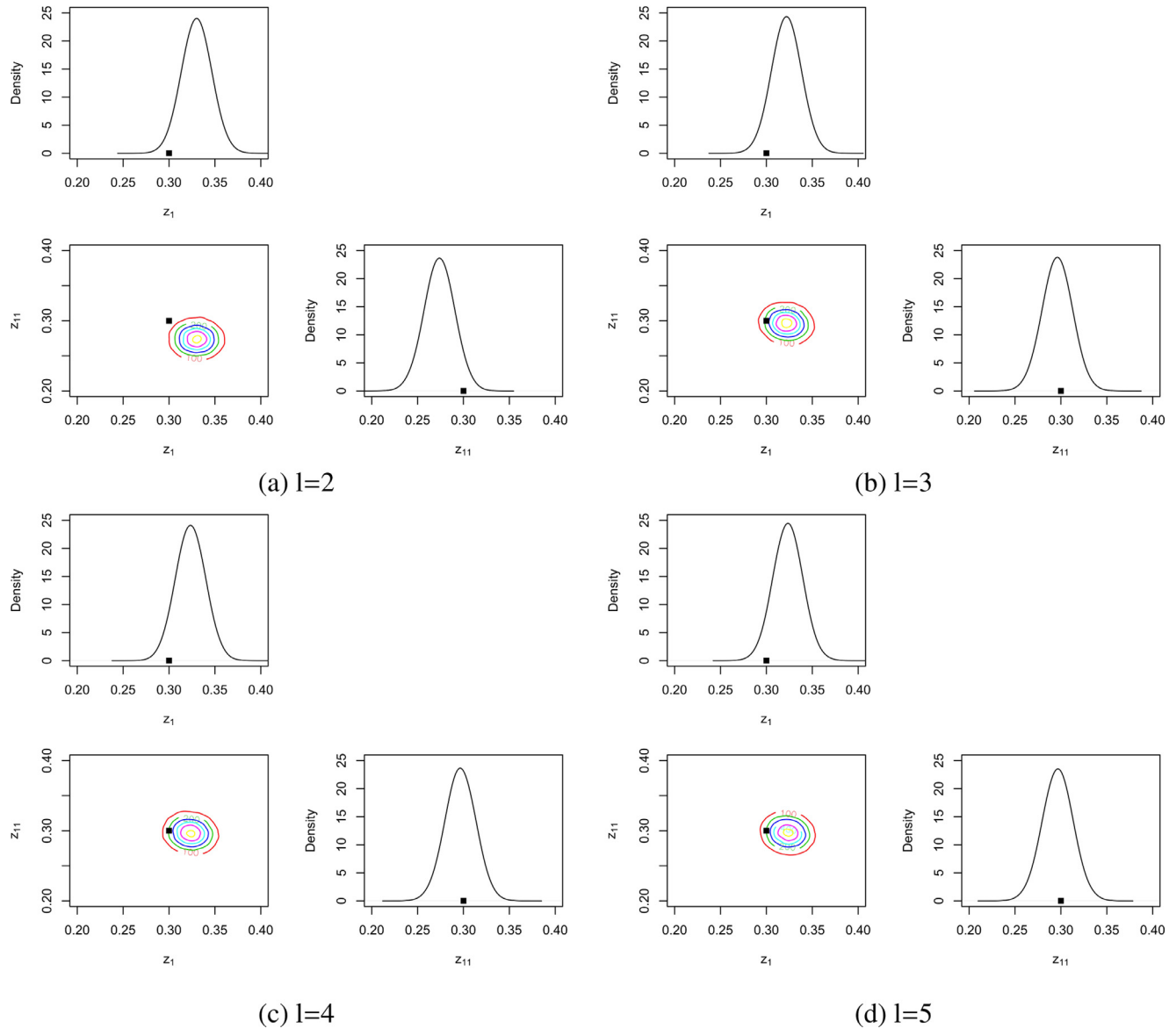


Fig. 3. The results by using the surrogate model constructed following Algorithm 1, z_1, z_{11} included as important dimensions. The points in the figure represent the true values of the parameters.

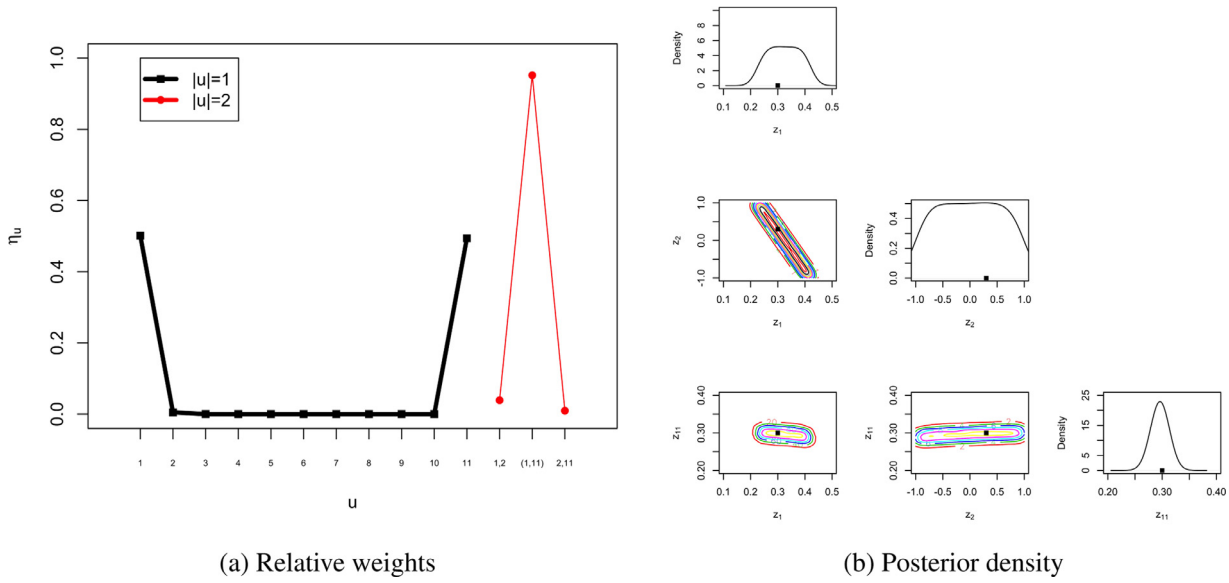
where the boundary layer length scale $\xi = d^2\Omega/(2E_u)^{0.5}$, the turbulence length scale $l_t = k^{0.5}/\omega$, the length scale bound to avoid stagnation point anomaly $l_B = k^{0.5}/(C_6|S|)$. Moreover, the intermittency production and dissipation term are modeled as:

$$P_\gamma = C_9 \times \rho F_{onset}[-ln(1-\gamma)]^{0.5}\left(1 + C_8 \times \sqrt{\frac{k}{2E_u}}\right)\frac{d}{\nu}|\nabla E_u|,$$
$$\epsilon_\gamma = \gamma P_\gamma \tag{42}$$

where the function F_{onset} is given by:

$$F_{onset} = 1.0 - \exp\left(-C_7 \times \frac{\xi_{eff} k^{0.5}|\nabla k|}{\nu|\nabla E_u|}\right) \tag{43}$$

As can be seen, there are totally 9 parameters in this transition model, without counting the parameters of the underlying $k-\omega$ turbulence model. Before carrying out the Bayesian calibration, we can argue that there is a degree of redundancy between C_1, C_5 and C_6 in determining the characteristic timescale τ_{nt}, as the 2nd Mach mode is dominant in this model for hypersonic flows. Thus we fix C_1 and C_6 to their nominal values in the subsequent simulations. Furthermore, we also fix C_4 to its nominal value, which was calibrated against the incompressible flow cases. To calibrate the model parameters (C_2, C_3, C_5, C_7, C_8, C_9), denoted as $\mathbf{z}$, we should at first specify their prior distributions. Here in this work, we perturb the parameters' nominal value by $\pm 10\%$ to obtain the prior range and use the uniform distributions for all the parameters. The range of perturbation ($\pm 10\%$) is determined by trial and error, in order to guarantee the convergence of the calculations. However, it is observed that transition can't be triggered in certain parameter configurations, so we further narrow the prior range of the corresponding parameters in order to put the design points in the narrowed stochastic domain, thus improving the accuracy of approximating the true response surface in the domain we are interested in. The final prior range is reported in Table 2.



(a) Relative weights (b) Posterior density

Fig. 4. The results by using the surrogate model constructed following Algorithm 1, z_1, z_2, z_{11} included as important dimensions. The component functions f_0, $f_1(z_1)$, $f_2(z_2)$, $f_{11}(z_{11})$, $f_{(1,\,11)}(z_1, z_{11})$ are included in the HDMR of $\mathscr{M}(\mathbf{z})$ (sparse grid level $l = 5$). The points in the figure represent the true values of the parameters.

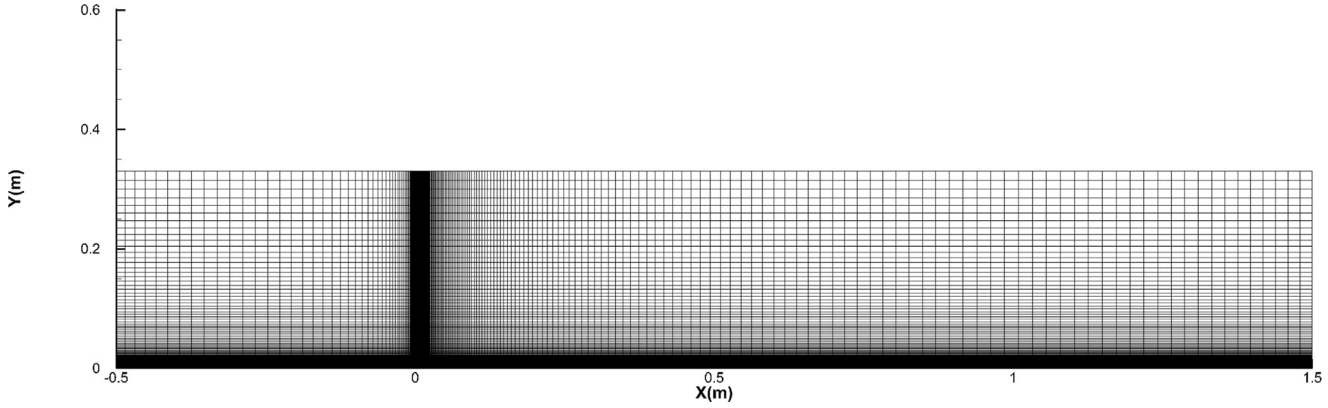
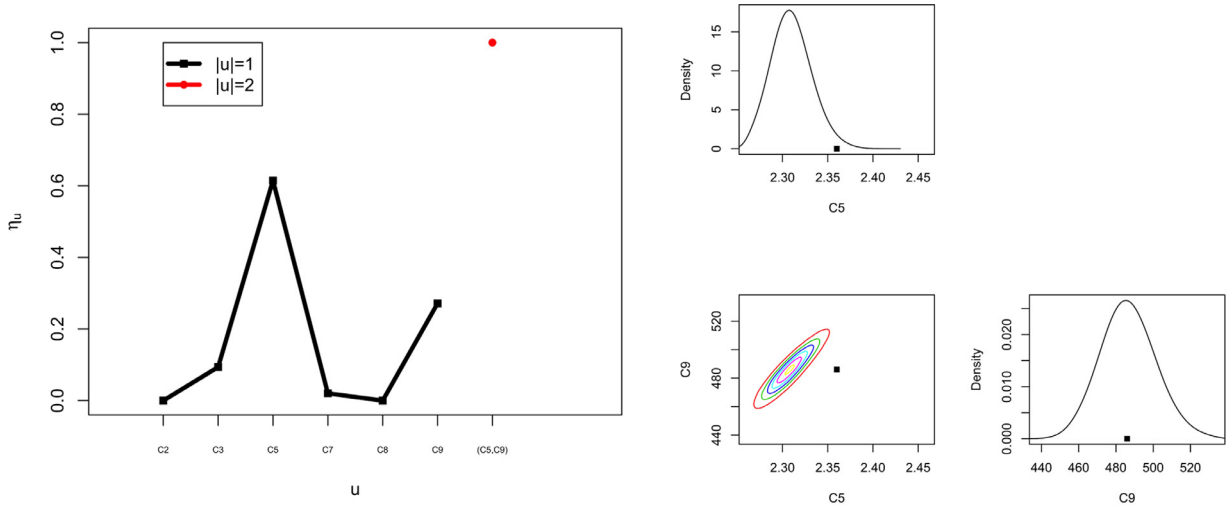


Fig. 5. The computational domain and the final mesh.



(a) Relative weights (b) Posterior density

Fig. 6. The results including both the relative weights and the posteriors, with the sparse grid level $l = 2$. The points in figure (b) represent the nominal values of the parameters.

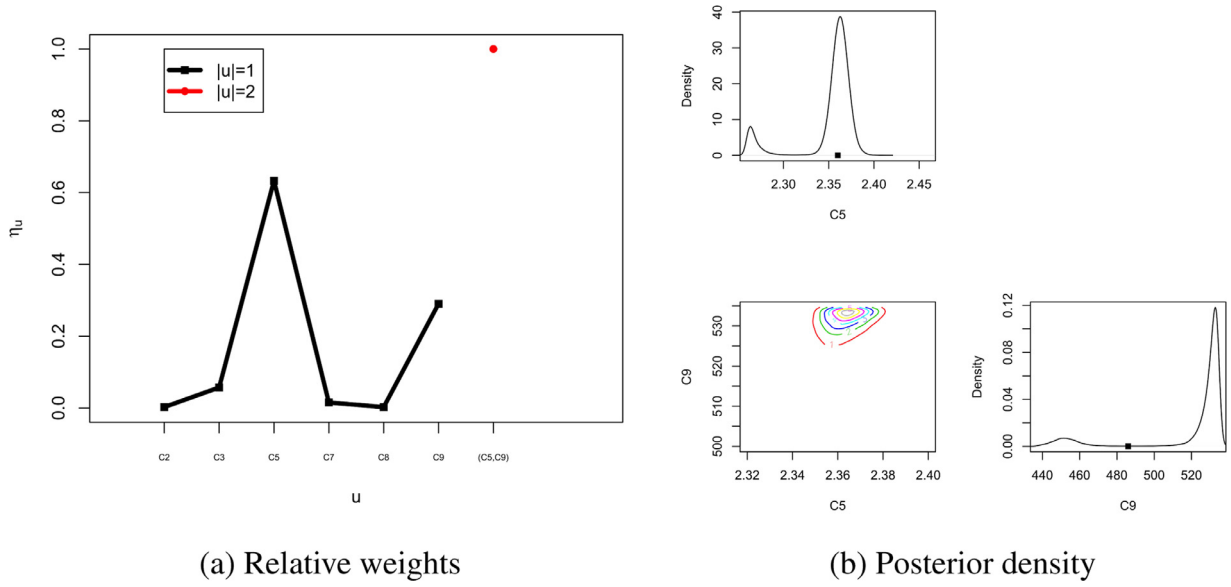


Fig. 7. The results including both the relative weights and the posteriors, with the sparse grid level $l = 3$. The points in figure (b) represent the nominal values of the parameters.

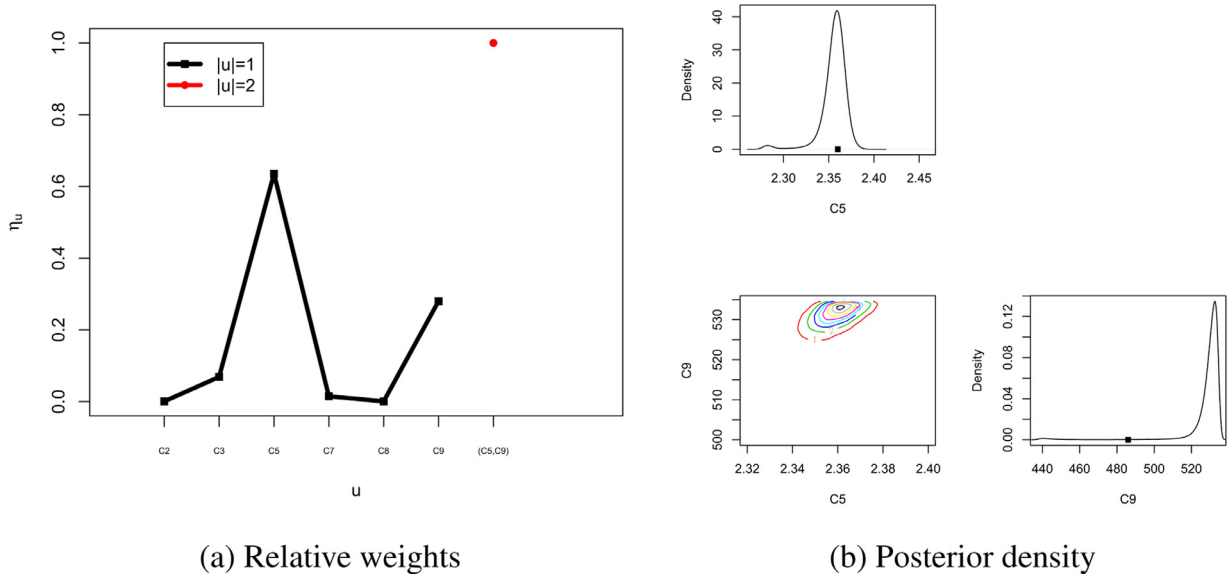


Fig. 8. The results including both the relative weights and the posteriors, with the sparse grid level $l = 4$. The points in figure (b) represent the nominal values of the parameters.

Table 2
The prior range for model parameters and the hyper-parameter σ.

Coefficient	Left boundary	Right boundary
C_2	6300	7700
C_3	0.54	0.66
C_5	2.26	2.46
C_7	3.3	3.5
C_8	0.063	0.077
C_9	437.4	534.6
σ	0	0.5

6.2. Numerical methods and grid generation

All the numerical simulations are carried out in our in-house three-dimensional compressible Navier-Stokes solver. The compressible Navier-Stokes equations are solved with Roe's implicit, finite volume, upwind algorithm. By means of the monotone upstream-centered schemes for conservation laws interpolation of the primitive variables, the quantity in the inviscid fluxes is obtained. The viscous flux terms are calculated by a second-order central difference scheme. Lower Upper Symmetric Gauss Seidel (LU-SGS) scheme is used for temporal integration. The no-slip, constant wall temperature conditions are imposed. The mesh independence study has been carried out and the final mesh used in this work is shown in Fig. 5. The grid dimension is 401×231 and the

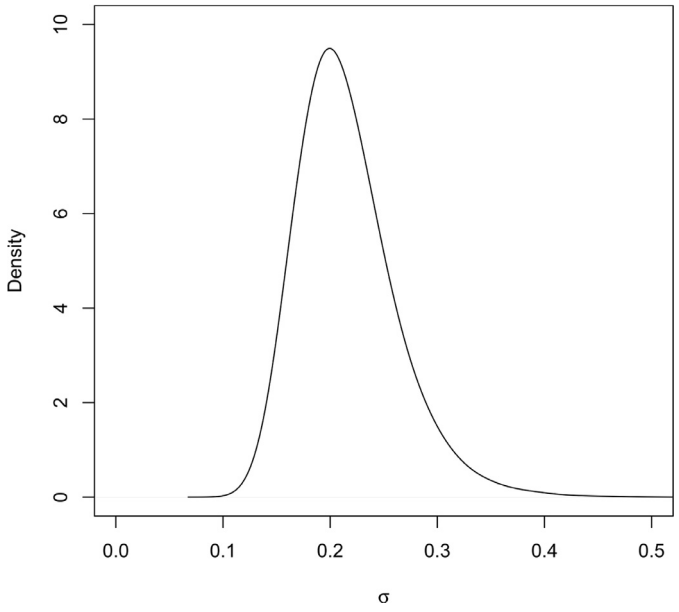


Fig. 9. The posterior of σ, with the sparse grid level $l = 4$.

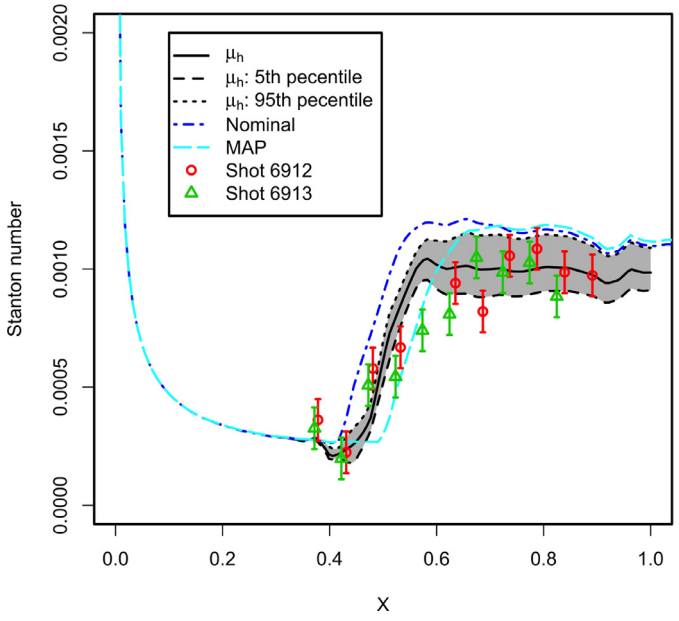


Fig. 10. The posterior distribution of $\mathscr{M}_S(\mathbf{C})$, the non-dimensional heat flux coefficient along the streamwise direction without including the model inadequacy term $\boldsymbol{\eta}$. The experimental data (Shot 6912 and 6913) is plotted with the 3σ error bar.

near wall grid spacing is 5×10^{-7}m, making $y^+ = 0.1$ at the end of the plate.

6.3. Specification of the flow class and QoIs

In this work, we are interested in the hypersonic boundary layer transition over a flat plate with a relatively low inlet turbulent intensity. The QoIs are the non-dimensional heat flux (Stanton number) along the streamwise direction, characterized by a steep rise across the transition region, as can be seen in Fig. 10. The input-output mapping can thus be denoted as $\mathscr{M}(\mathbf{C})$ with $\mathbf{C} = (C_2, C_3, C_5, C_7, C_8, C_9)$. The experimental data from Mee [32] (Shot 6912 and Shot 6913) is used in the current investigation to demonstrate the ability of the present approach to quantify the uncertainty efficiently and rigorously. The flow condition corresponding to this experimental setup is reported in Table 3. Here in this work only experimental data near the transition region is used and the measurement error is estimated as zero-mean gaussian with $\sigma_e = 3 \times 10^{-5}$. More experimental data over different working conditions should be included to inform the model and validate the model through cross validation. Further, the parameter variability across scenarios is believed to be large, thus will result in a large uncertainty in the prediction of QoIs. Whether to accept or reject the underlying model is up to the decision makers' requirement for the prediction precision, which is out of the scope of this paper.

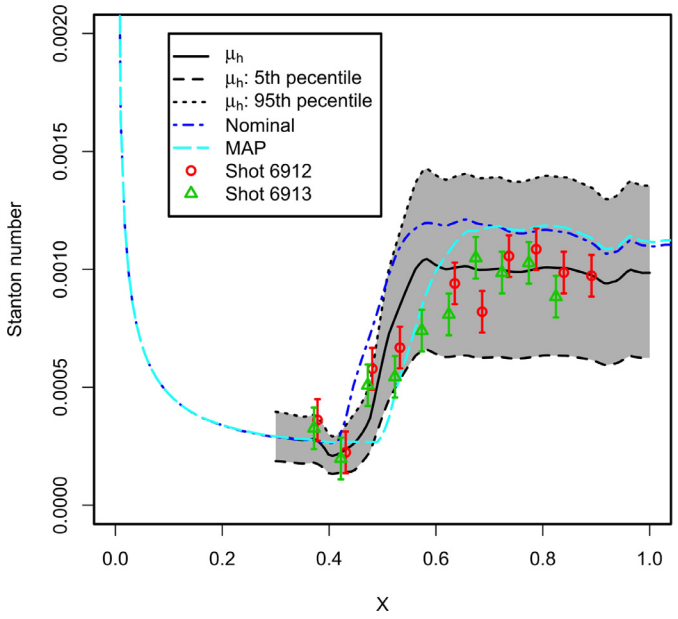


Fig. 11. The posterior distribution of $\mathscr{M}_S(\mathbf{C})(1 + \boldsymbol{\eta})$, the non-dimensional heat flux coefficient along the streamwise direction including the model inadequacy term $\boldsymbol{\eta}$. The experimental data (Shot 6912 and 6913) is plotted with the 3σ error bar.

Table 3
The flow condition corresponding to the experimental setup (Shot 6912 and Shot 6913 [32]).

Ma_∞	T_∞	P_∞	Re_∞/m	T_w
6.1	800 K	12.1 kPa	4.9×10^6	300 K

6.4. Results

The stochastic model is constructed following Section 2.2, with the prior of the model parameters described in Section 6.1. The prior for the hyper-parameter σ is also chosen as uniform distribution and the corresponding range is also reported in Table 2. 5×10^5 MCMC samples are generated to estimate the posterior distributions, with a burn-in length of 10^4. The results are shown in Figs. 6–8, including both the relative weights and the posteriors, with the sparse grid level from 2 to 4. As can be seen, C_5 and C_9 are the first two most important dimensions and account for over 90% of the total variance of the first order terms, thus only f_0, $f_5(C_5)$, $f_9(C_9)$, $f_{(5,\,9)}(C_5,\,C_9)$ are included in the HDMR of $\mathscr{M}(\mathbf{C})$. The posteriors are almost the same for $l = 3$ and $l = 4$, indicating gPC-SC method with grid level $l = 4$ is sufficient in this case. The model parameters are well informed, especially C_5. And the correlation between C_5 and C_9 is weak, which may indicate C_5 and C_9 capture

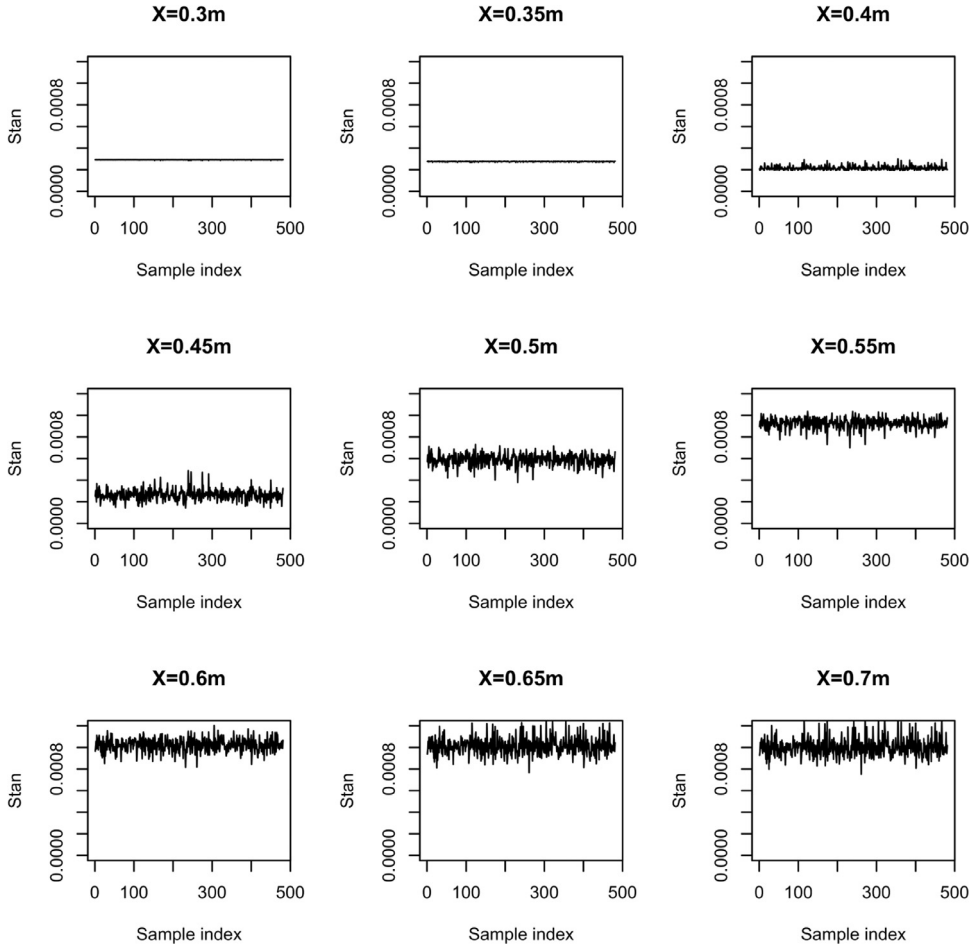


Fig. 12. The details of the MCMC samples of Stanton number at different streamwise locations, without the inadequacy term $\boldsymbol{\eta}$. The samples are obtained by extracting the original MCMC samples with un interval of 1000.

the different aspects of the transition process. This is in accordance with the idea behind the design of the transition model. Specifically, C_5 has a great impact on the growth of the 2nd mode, which determines the onset of transition through F_{onset}, while C_9 is the multiplicative coefficient of the production term P_γ, which determines the growth rate of the intermittency coefficient γ, hence has a great impact on the length of the transition zone. The posterior distribution of σ is given in Fig. 9. The inadequacy term is quite large and it may be due to the inability of the transition model and the underlying SST turbulence model to fit the experimental data.

The posterior model check is done by propagating the parameter uncertainty through the surrogate, which is constructed in the same way as in the Bayesian calibration step. The results are shown in Figs. 10 and 11, with and without including the model inadequacy term $\boldsymbol{\eta}$ respectively. The results for the nominal values and Maximum A Posteriori (MAP) values are also shown for comparison. As can be seen, the match between the prediction mean and the experimental data is much better than the nominal one, indicating a successful calibration. The 90% confidence interval overlaps quite well with the experimental data, indicating the output uncertainty can be well captured by the parameter uncertainty.

It is also interesting to note that the prediction mean matches with experimental data better than the MAP calculation. This is because the intermittency of transition to turbulence is also accounted for by the spread of the model parameters. Specifically, the intermittency factor γ of the transition model is based on the physical observation that at a fixed location of the transitional zone, the flow remains laminar for a portion of time and turbulent for another portion of time. Thus the intermittency factor is defined as the proportion of time that the flow remains turbulent. The intermittency factor values from 0 to 1 and is directly modeled in the k-ω-γ transition model, but it is further accounted for by the spread of model parameters, i.e. at a fixed location, the flow remains laminar for a certain parameter setup and it becomes turbulent for another set of parameters. Thus the use of probability distribution for model parameters somehow mimics the chaotic or pseudo-stochastic behavior of flow transition. Similarly, in Ref. [33], Serino et al. combined the e^N method with uncertainty quantification to capture the intermittency factor γ as the probability of transition, by assuming a probability distribution for the frequency and the propagation angle of the oblique waves. It is also worth noting that the experimental data for heat flux is obtained also by averaging the measurement over a period of time, during which the flow switches between laminar and turbulent frequently. Fig. 12 gives the details of the MCMC samples of Stanton number at different streamwise locations. The samples are obtained by subsampling the original MCMC chain with an interval of 1000. The spread of the Stanton number is much greater after transition onset and the intermittency-like behavior becomes more clear if we also include the inadequacy term, as is shown in Fig. 13.

The results for other quantities can also be obtained, with quantified uncertainty. Here in Fig. 14 we give the friction drag coefficient along the streamwise direction, without including $\boldsymbol{\eta}$. There is no experimental data available thus whether the results for C_f is consistent or not is hard to determine.

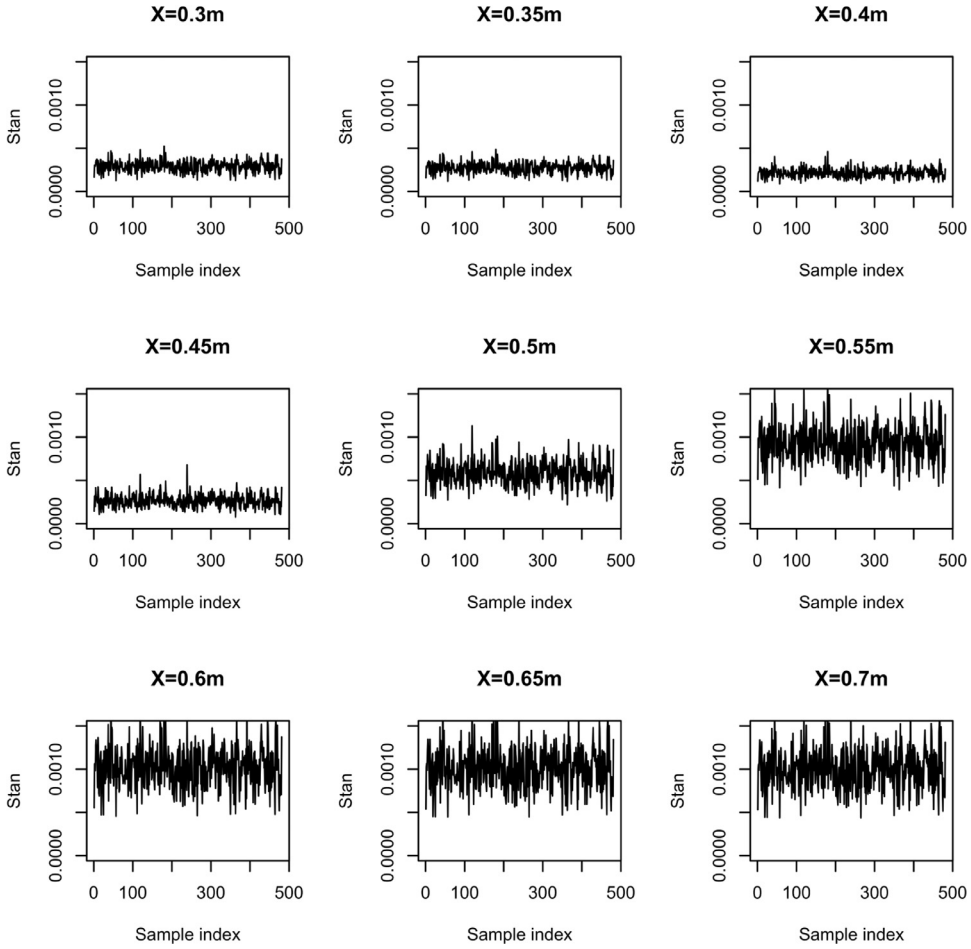


Fig. 13. The details of the MCMC samples of Stanton number at different streamwise locations, with the inadequacy term η. The samples are obtained by extracting the original MCMC samples with un interval of 1000.

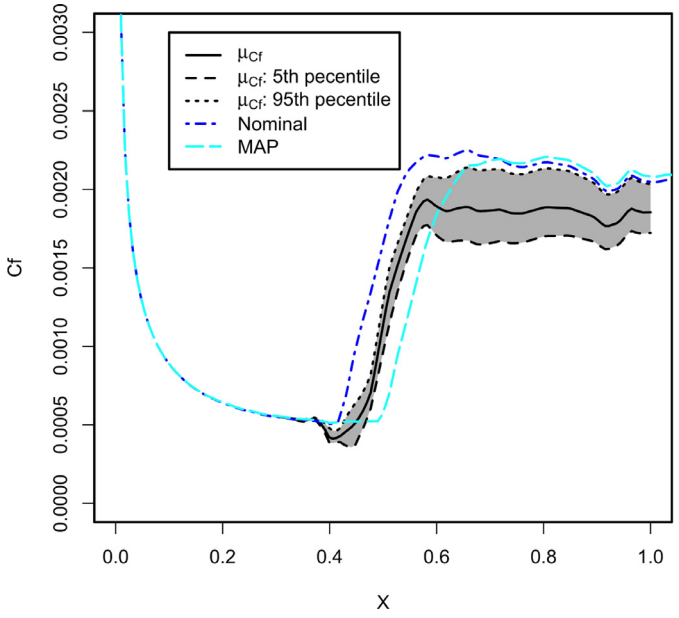


Fig. 14. The posterior friction drag coefficient along the streamwise direction, without the model inadequacy term η.

7. Conclusion

In this paper, an efficient Bayesian uncertainty quantification approach is proposed, and its efficiency and accuracy are demonstrated with application to both a simple mathematical function and a complex fluid dynamic model, i.e. k-ω-γ transition model. The uncertainty quantification framework follows closely from Chueng et al. [5]. To avoid the formidable number of model evaluations in the sampling procedure, a surrogate model construction method is proposed, combining the powerful adaptive high dimensional stochastic model representation (HDMR) technique [21] and stochastic collocation (SC) method based on generalized polynomial chaos (gPC) [24]. Specifically, the exact model is approximated by high dimensional model representation, with each component function adaptively chosen based on the relative importance, then each subproblem is solved by gPC-based SC method. In this work, the importance indicator is defined based on the variance of the corresponding output, rather than the impact on the output mean. This is reasonable because the input parameters that can be well informed are the ones the model output is sensitive to. Further, a rigorous convergence study of the approximate posterior to the true posterior is carried out in terms of Kullback–Leibler divergence, with increasing gPC order and HDMR truncation order.

The application of the proposed method to the k-ω-γ transition model is carried out and the results show not only a quantified uncertainty overlapping with the experimental data, but also a great improvement of the match between the prediction mean and the experimental data. The prediction mean results are even

better than the Maximum A Posteriori (MAP) estimates, which can be explained by the further account of the intermittency effect of the transition process through the spread of the input parameters. In terms of efficiency, only 65 code runs are done in the calibration procedure, comparing the 389 code runs for gPC and ten of thousands code runs for direct model evaluations.

The current investigation only includes one experimental set-up, thus can be used only in a limited number of flow configurations. However, the sensitivity analysis for the model parameters can still provide valuable information for the modeler to appreciate the underlying model. Further work will involve the inclusion of more experimental data or direct numerical simulation (DNS) data for model updating. A more physics-based stochastic model construction is also needed to well capture the model inadequacy.

Appendix A. Convergence study

The convergence of the proposed approach is proved below, following the work of Refs. [21,24]. In Theorem 1, the convergence of the surrogate model to the true model output is proved. Then the convergence of posterior obtained with the surrogate model is proved in terms of the Kullback–Leibler divergence in Theorem 2, following the similar procedure of the proof of Lemma 4.2. in Ref. [24].

Theorem 1. *Let $S = S_{N_t,N_s}$ the set of indices that the corresponding component functions are included in the high dimensional model representation, where N_t, N_s are the corresponding superposition and truncation dimensions. Let $\tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}})|_{\mathbf{z}=\bar{\mathbf{z}}\setminus z_{\mathbf{v}}}$ the gPC-SC approximation with the same polynomial order and the same sparse grid level for all the subproblems. Then the surrogate model is constructed as: $\mathscr{M}_S(\mathbf{z}) = \sum_{\mathbf{u}\in S}\sum_{\mathbf{v}\subset\mathbf{u}}(-1)^{|\mathbf{u}|-|\mathbf{v}|}\tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}})|_{\mathbf{z}=\bar{\mathbf{z}}\setminus z_{\mathbf{v}}}$. If the L^2 error of the gPC approximation of all the subproblems are denoted as ϵ and the trunction error denoted as ϵ_t, i.e.:*

$$\|\mathscr{M}(z_{\mathbf{v}})|_{\mathbf{z}=\bar{\mathbf{z}}\setminus z_{\mathbf{v}}} - \tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}})|_{\mathbf{z}=\bar{\mathbf{z}}\setminus z_{\mathbf{v}}}\|_{L^2_{p(\mathbf{z})}} \le \epsilon \tag{A.1}$$

$$|\mathscr{M}(\mathbf{z}) - \mathscr{M}_{N_t,N_s}| \le \epsilon_t \tag{A.2}$$

where $\mathscr{M}_{N_t,N_s} = \sum_{\mathbf{u}\in S}\sum_{\mathbf{v}\subset\mathbf{u}}(-1)^{|\mathbf{u}|-|\mathbf{v}|}\mathscr{M}(z_{\mathbf{v}})|_{\mathbf{z}=\bar{\mathbf{z}}\setminus z_{\mathbf{v}}}$, then:

$$\|\mathscr{M}(\mathbf{z}) - \mathscr{M}_S(\mathbf{z})\|_{L^2_{p(\mathbf{z})}} \le \epsilon_t + c(N_s,N_t)\epsilon \tag{A.3}$$

for all $\mathscr{M}(\mathbf{z}) \in L^2$space with the inner product induced by $p(\mathbf{z})$.

Proof.

$$\begin{aligned}
&\|\mathscr{M}(\mathbf{z}) - \mathscr{M}_S(\mathbf{z})\|^2_{L^2_{p(\mathbf{z})}} \\
&= \int[(\mathscr{M}(\mathbf{z}) - \mathscr{M}_{N_t,N_s})^2 + (\mathscr{M}_{N_t,N_s} - \mathscr{M}_S)^2 \\
&\quad + 2(\mathscr{M}(\mathbf{z}) - \mathscr{M}_{N_t,N_s})(\mathscr{M}_{N_t,N_s} - \mathscr{M}_S(\mathbf{z}))]p(\mathbf{z})d\mathbf{z}
\end{aligned} \tag{A.4}$$

Also:

$$\int(\mathscr{M}(\mathbf{z}) - \mathscr{M}_{N_t,N_s})^2 p(\mathbf{z})d\mathbf{z} \le \epsilon_t^2 \tag{A.5}$$

$$\begin{aligned}
&\int(\mathscr{M}_{N_t,N_s} - \mathscr{M}_S)^2 p(\mathbf{z})d\mathbf{z} \\
&= \int\left[\sum_{\mathbf{u}\in S}\sum_{\mathbf{v}\subset\mathbf{u}}(-1)^{|\mathbf{u}|-|\mathbf{v}|}(\mathscr{M}(z_{\mathbf{v}}) - \tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}}))\right]^2 p(\mathbf{z})d\mathbf{z} \\
&\le \int\left(\sum_{\mathbf{u}\in S}\sum_{\mathbf{v}\subset\mathbf{u}}|\mathscr{M}(z_{\mathbf{v}}) - \tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}})|\right)^2 p(\mathbf{z})d\mathbf{z} \\
&\le \int C_1^2 \max_{i,j}(|\mathscr{M}(z_{\mathbf{v}_i}) - \tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}_i})||\mathscr{M}(z_{\mathbf{v}_j}) - \tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}_j})|)p(\mathbf{z})d\mathbf{z} \\
&\le C_1^2 \max_{i,j}\|\mathscr{M}(z_{\mathbf{v}_i}) - \tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}_i})\|_{L^2_{\mathbf{z}}}\|\mathscr{M}(z_{\mathbf{v}_j}) - \tilde{\mathscr{M}}_{N_p}(z_{\mathbf{v}_j})\|_{L^2_{\mathbf{z}}} \\
&\le C_1^2\epsilon^2
\end{aligned} \tag{A.6}$$

$$\begin{aligned}
&\int(\mathscr{M}(\mathbf{z}) - \mathscr{M}_{N_t,N_s})(\mathscr{M}_{N_t,N_s} - \mathscr{M}_S(\mathbf{z}))p(\mathbf{z})d\mathbf{z} \\
&\le \|\mathscr{M}(\mathbf{z}) - \mathscr{M}_{N_t,N_s}\|_{L^2_{p(\mathbf{z})}}\|\mathscr{M}_{N_t,N_s} - \mathscr{M}_S(\mathbf{z})\|_{L^2_{p(\mathbf{z})}} \\
&\le C_1\epsilon\epsilon_t
\end{aligned} \tag{A.7}$$

where $C_1 = \sum_{k=1}^{N_s}\binom{N_t}{k}\sum_{j=1}^{k}\binom{k}{j} \le 2^{N_s+1}N_t^{N_s}$.

Thus from Eqs. (A.4)–(A.7) we obtain:

$$\|\mathscr{M}(\mathbf{z}) - \mathscr{M}_S(\mathbf{z})\|_{L^2_{p(\mathbf{z})}} \le \epsilon_t + C_1\epsilon \tag{A.8}$$

□

Remark. The theorem is formulated for the scalar model output. The application to a vector model output can be written as: $\|\mathscr{M}_i(\mathbf{z}) - \mathscr{M}_{S_i}(\mathbf{z})\|_{L^2_{p(\mathbf{z})}} \le \epsilon_t + c(N_s,N_t)\epsilon$.

Theorem 2. *Assume the stochastic model is constructed as Eq. (6), then $L(\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^{N_d}|\boldsymbol{\lambda}|}}\exp(-\frac{1}{2}\boldsymbol{\delta}^T\boldsymbol{\lambda}^{-1}\boldsymbol{\delta})$ and posterior $p(\boldsymbol{\theta}|\mathbf{d}) = \frac{L(\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}$ with $\boldsymbol{\delta} = \mathbf{d} - \mathscr{M}(\mathbf{z})$, $\boldsymbol{\lambda}$ is given by Eq. (7), and $\boldsymbol{\theta} = (\mathbf{z},\sigma)$. If the surrogate model $\mathscr{M}_S(\mathbf{z})$ converge to $\mathscr{M}(\mathbf{z})$ in the L^2 sense, i.e.*

$$\|\mathscr{M}_i(\mathbf{z}) - \mathscr{M}_{S_i}(\mathbf{z})\|_{L^2_{\mathbf{pz}}} \to 0\ ,\ \epsilon \to 0\ ,\ \epsilon_t \to 0 \text{ for } 1 \le i \le N_d \tag{A.9}$$

Then:

$$D(p_S^{\mathbf{d}}(\boldsymbol{\theta})\|p^{\mathbf{d}}(\boldsymbol{\theta})) \to 0\ ,\ \epsilon \to 0\ ,\ \epsilon_t \to 0 \tag{A.10}$$

where N_d is the dimension of the model output, the superscript $\mathbf{d}$ represents the posterior conditioned on $\mathbf{d}$, the subscript S represents the replacement of the true model by the surrogate model, and the Kullback–Leibler divergence (KLD) is defined as:

$$D(p_1\|p_2) = \int p_1(\mathbf{z})log\left(\frac{p_1(\mathbf{z})}{p_2(\mathbf{z})}\right)d\mathbf{z} \tag{A.11}$$

Proof. Let

$$\gamma = \int L(\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}\ ,\ \gamma_S = \int L_S(\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{A.12}$$

then $\gamma > 0$, $\gamma_S > 0$. By following the definition of $L(\boldsymbol{\theta})$ and $L_S(\boldsymbol{\theta})$, and utilizing the fact that e^{-x} is Lipshitz continuous for $x \ge 0$, i.e. $|e^{-x} - e^{-y}| \le \Lambda|x-y|$ for all x, $y > 0$, where Λ is a positive constant. Then we have:

$$\begin{aligned}
&|\gamma_S - \gamma| \\
&= \left|\int(L(\boldsymbol{\theta}) - L_S(\boldsymbol{\theta}))p(\boldsymbol{\theta})d\boldsymbol{\theta}\right| \\
&= \left|\int\frac{1}{\sqrt{(2\pi)^{N_d}|\boldsymbol{\lambda}|}}(e^{-\frac{1}{2}\boldsymbol{\delta}^T\boldsymbol{\lambda}^{-1}\boldsymbol{\delta}} - e^{-\frac{1}{2}\boldsymbol{\delta}_S^T\boldsymbol{\lambda}^{-1}\boldsymbol{\delta}_S})p(\boldsymbol{\theta})d\boldsymbol{\theta}\right| \\
&= \left|\int\prod_{i=1}^{N_d}\frac{1}{\sqrt{2\pi\sigma_t^2}}(e^{-\frac{1}{2\sigma_t^2}(d_i-\mathscr{M}_i(\mathbf{z}))^2} - e^{-\frac{1}{2\sigma_t^2}(d_i-\mathscr{M}_{S_i}(\mathbf{z}))^2})p(\boldsymbol{\theta})d\boldsymbol{\theta}\right| \\
&\le \prod_{i=1}^{N_d}\int\frac{1}{\sqrt{2\pi\sigma_t^2}}|e^{-\frac{1}{2\sigma_t^2}(d_i-\mathscr{M}_i(\mathbf{z}))^2} - e^{-\frac{1}{2\sigma_t^2}(d_i-\mathscr{M}_{S_i}(\mathbf{z}))^2}|p(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&\le \prod_{i=1}^{N_d}\int\frac{1}{\sqrt{2\pi\sigma_t^2}}\frac{\Lambda}{2\sigma_t^2}|(d_i-\mathscr{M}_i(\mathbf{z}))^2 - (d_i-\mathscr{M}_{S_i}(\mathbf{z}))^2|p(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&\le \prod_{i=1}^{N_d}\frac{1}{\sqrt{2\pi\sigma_t^2}}\frac{\Lambda}{2\sigma_t^2}\|\mathscr{M}_i(\mathbf{z}) - \mathscr{M}_{S_i}(\mathbf{z})\|_{L^2_{p(\mathbf{z})}} \\
&\quad\times\|2d_i - \mathscr{M}_i(\mathbf{z}) - \mathscr{M}_{S_i}(\mathbf{z})\|_{L^2_{p(\mathbf{z})}}
\end{aligned}$$

$$\leq C_1 \prod_{i=1}^{N_d} \|\mathscr{M}_i(\mathbf{z}) - \mathscr{M}_{S_i}(\mathbf{z})\|_{L^2_{p(\mathbf{z})}} \tag{A.13}$$

where we have assumed $err_i = \exp[-\frac{1}{2\sigma_t^2}(d_i - \mathscr{M}_i(\mathbf{z}))^2] - \exp[-\frac{1}{2\sigma_t^2}(d_i - \mathscr{M}_{S_i}(\mathbf{z}))^2]$ is mutually independent for each i and the Hölder's inequality has been used.

Also we can derive:

$$\log \frac{p_S^d(\boldsymbol{\theta})}{p^d(\boldsymbol{\theta})} = -\frac{1}{2\sigma_t^2}\sum_{i=1}^{N_d}((d_i - \mathscr{M}_{S_i}(\mathbf{z}))^2 - (d_i - \mathscr{M}_i(\mathbf{z}))^2) + \log\frac{\gamma}{\gamma_S} \tag{A.14}$$

Therefore,

$$\begin{aligned} D(p_S^d \| p^d) &= \int p_S^d \log\left(\frac{p_S^d}{p^d}\right) d\boldsymbol{\theta} \\ &= \frac{1}{\gamma_S}\sum_{i=1}^{N_d}\int -\frac{1}{2\sigma_t^2}((d_i - \mathscr{M}_{S_i}(\mathbf{z}))^2 - (d_i - \mathscr{M}_i(\mathbf{z}))^2) \\ &\quad \times L_S(\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} + \log\frac{\gamma}{\gamma_S} \end{aligned} \tag{A.15}$$

Considering both γ and γ_S are positive constants and $L_S(\boldsymbol{\theta})$ is bounded, i.e. $0 < L_S(\boldsymbol{\theta}) \leq C_2$, we can derive:

$$\begin{aligned} D(p_S^d \| p^d) &\leq \frac{C_2}{\gamma_S}\frac{1}{2\sigma_e^2}\sum_{i=1}^{N_d}\int |(d_i - \mathscr{M}_{S_i}(\mathbf{z}))^2 - (d_i - \mathscr{M}_i(\mathbf{z}))^2| p(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\quad + \left|\log\frac{\gamma}{\gamma_S}\right| \\ &\leq \frac{C_3}{\gamma_S}\frac{1}{2\sigma_e^2}\sum_{i=1}^{N_d}\|\mathscr{M}_i(\mathbf{z}) - \mathscr{M}_{S_i}(\mathbf{z})\|_{L^2_{p(\mathbf{z})}} + \left|\log\frac{\gamma}{\gamma_S}\right| \end{aligned} \tag{A.16}$$

The first term converges by following Eq. (A.9) and the second term by following Eq. (A.13), thus the posterior obtained from the surrogate model converges to the true posterior in the sense of KLD convergence, i.e.:

$$D(p_S^{\mathbf{d}}(\boldsymbol{\theta}) \| p^{\mathbf{d}}(\boldsymbol{\theta})) \to 0 \ , \ \epsilon \to 0 \ , \ \epsilon_t \to 0 \tag{A.17}$$

□

References

[1] Fu S, Wang L. Rans modeling of high-speed aerodynamic flow transition with consideration of stability theory. Prog Aerosp Sci 2013;58:36–59.
[2] Wang L, Fu S. Development of an intermittency equation for the modeling of the supersonic/hypersonic boundary layer flow transition. Flow Turbul Combust 2011;87(1):165–87.
[3] Zhong X, Wang X. Direct numerical simulation on the receptivity, instability, and transition of hypersonic boundary layers. Annu Rev Fluid Mech 2012;44:527–61.
[4] Kennedy MC, O'Hagan A. Bayesian calibration of computer models. J R Stat Soc 2001;63(3):425–64.
[5] Cheung SH, Oliver TA, Prudencio EE, Prudhomme S, Moser RD. Bayesian uncertainty analysis with applications to turbulence modeling. Reliab Eng Syst Saf 2011;96(9):1137–49.
[6] Oliver TA, Moser RD. Bayesian uncertainty quantification applied to RANSturbulence models. In: Journal of physics: conference series, 318. IOP Publishing; 2011. p. 042032.
[7] Edeling W, Cinnella P, Dwight RP, Bijl H. Bayesian estimates of parameter variability in the k–ε turbulence model. J Comput Phys 2014;258:73–94.
[8] Edeling W, Cinnella P, Dwight RP. Predictive RANS simulations via Bayesian model-scenario averaging. J Comput Phys 2014;275:65–91.
[9] Ray J, Lefantzi S, Arunajatesan S, Dechant L. Bayesian parameter estimation of ak-ε model for accurate jet-in-crossflow simulations. AIAA J 2016:2432–48.
[10] Margheri L, Meldi M, Salvetti MV, Sagaut P. Epistemic uncertainties in RANS model free coefficients. Comput Fluids 2014;102:315–35.
[11] Dow E, Wang Q. Quantification of structural uncertainties in the $k-\omega$ turbulence model. AIAA Paper 2011;1762:2011.
[12] Emory M, Larsson J, Iaccarino G. Modeling of structural uncertainties in Reynolds-averaged Navier-Stokes closures. Phys Fluids 2013;25(11):110822.
[13] Gorlé C, Iaccarino G. A framework for epistemic uncertainty quantification of turbulent scalar flux models for Reynolds-averaged Navier-Stokes simulations. Phys Fluids 2013;25(5):055105.
[14] Duraisamy K, Zhang ZJ, Singh AP. New approaches in turbulence and transition modeling using data-driven techniques. In: 53rd AIAA aerospace sciences meeting; 2015. p. 1284.
[15] Xiao H, Wu J-L, Wang J-X, Sun R, Roy C. Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations: a data-driven, physics-informed bayesian approach. J Comput Phys 2016;324:115–36.
[16] Pecnik R, Witteveen J, Iaccarino G. Uncertainty quantification for laminar-turbulent transition prediction in RANS turbomachinery applications. In: 49th AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition; 2011. p. 660.
[17] Menter FR, Langtry RB, Likki S, Suzen Y, Huang P, Völker S. A correlation-based transition model using local variables—part I: model formulation. J Turbomach 2006;128(3):413–22.
[18] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 1970;57(1):97–109.
[19] Xiu D, Hesthaven JS. High-order collocation methods for differential equations with random inputs. Siam J Sci Comput 2005;27(3):1118–39.
[20] Witteveen JAS, Iaccarino G. Simplex stochastic collocation with ENO–type stencil selection for robust uncertainty quantification. J Comput Phys 2013;239(8):1–21.
[21] Ma X, Zabaras N. An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. J Comput Phys 2010;229(10):3884–915.
[22] Ma X, Zabaras N. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. J Comput Phys 2009;228(8):3084–113.
[23] Edeling W, Dwight RP, Cinnella P. Simplex-stochastic collocation method with improved scalability. J Comput Phys 2016;310:301–28.
[24] Marzouk Y, Xiu D. A stochastic collocation approach to Bayesian inference in inverse problems. Commun Comput Phys 2009;6(4):826–47.
[25] Haario H, Saksman E, Tamminen J. An adaptive metropolis algorithm. Bernoulli 2001;7(2):223–42.
[26] R Core Team. R: alanguage and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria; 2017. URL https://www.R-project.org/.
[27] Chivers C.. MHadaptive: general Markov chain Monte Carlo for Bayesian inference using adaptive Metropolis-Hastings sampling; 2012. R package version 1.1–8; URL https://CRAN.R-project.org/package=MHadaptive.
[28] Maître OPL, Knio OM. Spectral methods for uncertainty quantification. Springer Netherlands; 2010.
[29] Smoljak SA. Quadrature and interpolation formulae on tensor products of certain function classes.. Dokl Akad Nauk SSSR 1963;4(5):240–3.
[30] Zuniga MM, Ko J GPC: generalized polynomial chaos; 2014. R package version 0.1; URL https://CRAN.R-project.org/package=GPC.
[31] Griebel M, Holtz M. Dimension-wise integration of high-dimensional functions with applications to finance. J Complex 2010;26(5):455–89.
[32] Mee DJ. Boundary-layer transition measurements in hypervelocity flows in a shock tunnel. AIAA J 2002;40(8):1542–8.
[33] Serino G, Marxen O, Pinna F, Rambaud P, Magin T. Transition prediction for oblique breakdown in supersonic boundary layers with uncertain disturbance spectrum. In: 42nd AIAA fluid dynamics conference and exhibit; 2012. p. 2973.