Technische Universität München
Institut für Energietechnik

Professur für Thermofluiddynamik

Uncertainty Management in Thermoacoustic Instability Analysis

Shuai Guo

Vollständiger Abdruck der von der Fakultät für Maschinenwesen der Technischen Universität München zur Erlangung des akademischen Grades eines

DOKTOR – INGENIEURS

genehmigten Dissertation.

Vorsitzender:
Prof. Dr.-Ing. Dirk Weuster-Botz
Prüfer der Dissertation:
1. Prof. Wolfgang Polifke, Ph.D.
2. Prof. Phaedon-Stelios Koutsourelakis, Ph.D.

Die Dissertation wurde am 22.12.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 30.03.2021 angenommen.

Abstract

The present thesis focuses on developing an uncertainty management framework to quantify and subsequently mitigate the thermoacoustic modal instability risk. Acoustic solvers such as a Helmholtz solver or a network model are coupled with flame response models to calculate modal frequencies and growth rates. The associated flame model parameters and acoustic system parameters are considered as input uncertainty sources, leading to the generation of the modal instability risk.

Towards quantifying the modal instability risk, an efficient uncertainty propagation scheme is firstly proposed to quantify the impact of an uncertain flame impulse response (FIR) model on the calculation of the modal growth rate. This uncertainty quantification scheme is based on a novel dimensionality reduction methodology called Active Subspace approach, which successfully uncovered and exploited a one-dimensional representation of the original high-dimensional FIR model, leading to a 50-fold increase in computational efficiency compared with the conventional Monte Carlo simulation.

Later on, this one-dimensional manifold is recovered analytically via performing linearization analysis on the thermoacoustic governing equation. As a deeper understanding is obtained regarding the causal relationship between the variations of FIR model coefficients and the variations of modal growth rate, it is now possible to derive an analytical uncertainty propagation procedure to evaluate the impact of FIR model uncertainties on the thermoacoustic instability prediction, with dramatic efficiency improvement over the previous Active Subspace study, while avoiding any complex mathematical treatments.

In an effort to propagate uncertainties from both the acoustic system parameters and high-dimensional flame models (including FIR model), a general surrogate-based framework is proposed to efficiently perform uncertainty quantification (UQ) analysis in thermoacoustic instability prediction. The core of the framework are Gaussian Process (GP) surrogate models, through which an iterative procedure is developed to efficiently solve the thermoacoustic governing equation. Two case studies are conducted to demonstrate the effectiveness of the proposed UQ framework for both linear and nonlinear thermoacoustic instability analysis.

In addition to quantifying the modal instability risk, another important aspect of the uncertainty management framework lays in mitigating the risk by implementing the principle of robust design. Towards that end, Gaussian Process surrogate models are firstly trained to facilitate fast forward UQ embedded in robust design analysis. Subsequently, different tasks of robust design, ranging from fundamental risk analysis to control design and inverse tolerance design, are systematically explored, where detailed mathematical formulations and corresponding efficient solution strategies are provided for each of these tasks. The concept of a "*risk diagram*" is later

introduced, which displays the distribution of the modal instability risk over the entire parameter space and allows a convenient visualization of the connections between different robust design goals.

The accuracy of the GP surrogate models plays a crucial role in completing the above-mentioned robust design tasks. In practice, however, GP model uncertainty induced by the limited number of training samples may propagate downstream, thus rendering the risk calculation unreliable. To address this issue, the variation of the modal instability risk induced by the epistemic GP model uncertainty is firstly quantified by leveraging the Bayesian characteristic of the GP model. Subsequently, an active learning scheme is proposed to sequentially allocating training samples in the vicinity of the modal stability margin, thus significantly reduced the variation of risk calculation while maintaining the same computational budget for GP model training.

In summary, a comprehensive uncertainty management framework is developed in the present thesis for quantifying and mitigating the thermoacoustic modal instability risk. This framework provides practitioners with advanced dimensionality reduction and surrogate modeling strategies, which can efficiently propagate uncertainties from both acoustic system parameters and high-dimensional flame models to the modal instability risk calculation. In addition, this framework categorizes various aspects of robust design in combustor thermoacoustic analysis and offers dedicated surrogate-model-based solutions to mitigate instability risk for individual scenarios. Finally, this framework does not refrain from acknowledging the "imperfectness" of the surrogate model that is largely ignored in practice. Instead, procedures are proposed to effectively quantify and reduce the impact of surrogate model uncertainty on the variation of modal instability risk calculation. Overall, the developed uncertainty management framework lays a solid foundation for achieving reliable thermoacoustic prediction and design of the combustor.

Kurzfassung

Die vorliegende Arbeit befasst sich mit der Entwicklung eines Rahmenkonzeptes zum Unsicherheitsmanagement für die Quantifizierung und die anschließende Minderung der Risiken von thermoakustischen Instabilitäten. Akustische Lösungsmethoden wie Helmholz-Löser oder eine akustische Netzmodellierung werden mit Flame-Response-Modellen gekoppelt, um Frequenzen und Wachstumsraten zu berechnen. Die dazugehörigen Parameter für das Flammenmodell und akustische Systemparameter werden als unsichere Eingangsgrößen betrachtet, welche zur Erzeugung eines Risikos einer modalen Instabilität führen.

Zur Quantifizierung des modalen Instabilitätsrisikos wird zunächst ein effizientes Schema zur Fortpflanzung der Unsicherheiten vorgestellt, um den Einfluss eines unsicheren Flame-Impulse-Response-Modells (FIR) auf die Berechnung der modalen Wachstumsraten zu untersuchen. Dieses Schema zur Quantifizierung von Unsicherheiten basiert auf neuartigen Methoden zur Dimensionsreduktion, Active Subspace Ansatz genannt. Mittels dieser Methode wurde eine eindimensionale Betrachtungsweise des ursprünglich mehrdimensionalen FIR-Modells gefunden und ausgenutzt, was zu einer 50-fach größeren Recheneffizienz im Vergleich zu einer konventionellen Monte Carlo Simulation führt.

Eine eindimensionale Betrachtungsweise konnte im Späteren auch analytisch mittels einer Linearisierung der beschreibenden thermoakustischen Gleichungen abgeleitet werden. Nachdem ein tiefes Verständnis über die kausalen Zusammenhänge zwischen der Variation von FIR-Modellkoeffizienten und der Variationen der modalen Wachstumsrate erlangt werden konnte, war es möglich, eine analytische Prozedur zur Fortpflanzung von Unsicherheiten herzuleiten, um den Einfluss der FIR-Modellunsicherheiten auf die Vorhersage von thermoakustischen Instabilitäten untersuchen zu können. Dieses Vorgehen hat eine dramatische Verbesserung bezüglich der Effizienz im Vergleich zum Active Subspace Ansatz zur Folge.

Um die Fortpflanzung der Unsicherheiten sowohl der akustischen Systemparameter als auch des mehrdimensionalen Flammenmodells (inklusive FIR-Modell) beschreiben zu können, wird ein allgemeiner Ansatz zur Ersatzmodellierung vorgestellt um effizienter eine Analyse der Unsicherheiten bei der Vorhersage von thermoakustischen Instabilitäten durchführen zu können. Der Kern dieses Ansatzes sind Gaussian Process (GP) Ersatzmodelle, durch welche ein iteratives Verfahren entwickelt werden kann, um die thermoakustischen Gleichungen effizient lösen zu können. An zwei Fallbeispielen wird die Effizienz des entwickelten Frameworks zur Quantifizierung der Unsicherheiten demonstriert, sowohl für eine lineare als auch nichtlineare Analyse der thermoakustischen Instabilitäten.

Ein weiterer wichtiger Aspekt des entwickelten Konzepts zum Unsicherheitsmanagement, zusätzlich zur Quantifizierung des modalen Instabilitätsrisikos, ist die Minderung der Risiken durch das Prinzip eines robusten Designs. Zu diesem Zweck werden die GP-Ersatzmodelle zunächst trainiert, um eine schnelle Quantifizierung der Unsicherheiten innerhalb einer Designanalyse zu ermöglichen. Anschließend werden verschiedene Aspekte eines robusten Designs systematisch untersucht, angefangen von fundamentalen Risikoanalysen bis hin zu Reglerentwurf und inverser Toleranzauslegung. Für jeder dieser Aspekte werden detaillierte mathematische Formulierungen und korrespondierende effiziente Lösungsstrategien vorgestellt. Das Konzept eines „Risiko-Diagramms“ wird vorgestellt, welches die Verteilung des modalen Instabilitätsrisikos über den gesamten Parameterraum darstellt. Diese Darstellung erlaubt eine anschauliche Visualisierung der Zusammenhänge zwischen verschiedenen Designzielen eines robusten Designs.

Die Genauigkeit der GP-Ersatzmodelle spielt eine entscheidende Rolle bei der Ausführung der oben erwähnten Designaspekte. In der Praxis könnte jedoch eine Modellunsicherheit des GP-Modells bedingt durch eine limitierte Anzahl an Daten für das Modelltraining dazu führen, dass die Risikoanalyse unzuverlässig wird. Um dieses Problem genauer zu untersuchen, wird die Variation des modalen Instabilitätsrisikos, welche von epistemischen GP-Modellunsicherheiten hervorgerufen wird, zunächst quantifiziert durch Untersuchung der Bayes'schen Charakteristik des GP-Modells. Anschließend wird ein aktives Lernschema für das GP-Modell vorgestellt.

In der vorgelegten Arbeit konnte ein umfassendes Rahmenkonzept zum Unsicherheitsmanagement für die Quantifizierung und Minderung des thermoakustischen modalen Instabilitätsrisikos entwickelt werden. Dieses Konzept bietet Anwendern fortgeschrittene Strategien zur Dimensionsreduktion und der Ersatzmodellierung. Diese Strategien liefern effizient eine Aussage über die Fortpflanzung von Unsicherheiten - in den akustischen Systemparametern und des mehrdimensionalen Flammenmodells - auf die Berechnung des modalen Instabilitätsrisikos. Darüber hinaus kategorisiert dieses Rahmenkonzept verschiedene Aspekte für ein robustes Design bei thermoakustischen Analysen von Brennkammern und bietet spezielle Lösungen auf der Basis von Ersatzmodellen, um das Instabilitätsrisiko für einzelne Szenarien zu mindern. Schlussendlich erkennt diese Arbeit die „Unvollkommenheit“ der Ersatzmodelle an, was in der Praxis weitestgehend ignoriert wird. Stattdessen werden Verfahren vorgeschlagen, um effektiv den Einfluss von Unsicherheiten der Ersatzmodelle auf die Variation der Berechnung des modalen Instabilitätsrisikos zu quantifizieren und zu minimieren. Das entwickelte Rahmenwerk für das Unsicherheitsmanagement bietet eine solide Grundlage für eine zuverlässige thermoakustische Vorhersage und Auslegung von Brennkammern.

Acknowledgements

The work presented in this thesis was conducted at the Thermo-Fluid Dynamics Group of the Technical University of Munich. Financial support was provided by China Scholarship Council, project No. 201606830045.

First of all, I would like to thank my supervisor Prof. Wolfgang Polifke, Ph.D. for giving me the opportunity to do research on this interesting topic and guide me through the entire doctoral process. He introduced me to the world of uncertainty quantification, which still fascinates me and motivates me to keep learning and exploring more. He always encouraged me to look for alternative solutions when I was stuck and gave me valuable suggestions to let me see the big picture. From discovering new research ideas, to strategically design studies to implement ideas, to effectively communicate results to the community, he taught me how to go from a dependent idea implementer to an independent idea generator, which would benefit me significantly in my future career.

My thanks also go to Dr. Camilo F. Silva. More than often, he is the first person to hear my new research ideas and the first person to review my manuscripts. He always responded me with encouragement and insightful comments. His passion for research showed me the secret of doing excellent work. I would also like to thank Dr. Abdulla Ghani, who closely advised me in the beginning of my project. He showed me the nitty-gritty of doing research and gave me new ways of thinking when I was disoriented.

Special thanks also go to my colleagues at the Thermo-Fluid Dynamics Group, for their continued support and making my time at the chair an unforgettable experience. I also would like to thank Helga Bassett and Sigrid Schulz-Reichwald, who shielded me from tedious yet crucial administrative matters.

Last but not least, this dissertation cannot be done without the support of my parents Yu Guo and Yan Long, and my partner Ann. Thank you for loving me, believing in me, and giving me the strength to explore, to grow, and to innovate.

Contents

1 Introduction

1.1 Motivation

Gas turbines are widely used for power generation and propulsion applications. In recent years, increasingly stringent regulations regarding the emissions of NOx, soot and noise have led gas turbine manufacturers to develop cleaner and more efficient combustion technologies [1]. Among various new approaches to achieve pollution control, the concept of lean-premixed combustion demonstrates the most potential in practice [2]. For combustors running in a lean, premixed mode, fuel and air are premixed before entering the combustion chamber, forming a mixture with a lean fuel to air ratio, which leads to a lean combustion process. Since the flame temperature is comparably low in lean combustion, the thermal NOx formation can be strongly attenuated [3].

However, combustion instabilities, also known as thermoacoustic instabilities, have emerged as a serious issue in lean-premixed combustion systems [4]. Combustion instabilities are typically caused by the coupling of the fluctuating heat release of the combustion process with naturally occurring acoustic resonances [5]: flame heat release rate fluctuations add energy to the combustor acoustic field, thus generating acoustic pressure and velocity fluctuations that propagate in the combustor. Subsequently, the generated acoustic fluctuations may lead to oscillations of flow structure, fuel/air ratio oscillations, etc., which in turn result in further heat release rate fluctuations.

Lean-premixed combustor is especially prone to thermoacoustic instabilities, as the conventional liner cooling system is lacking for sufficient acoustic damping, as well as the flame is short and acoustically compact, which promotes coupling between fluctuating flame heat release and unsteady flow motion [6, 7]. Those instability phenomena could lead to intense pressure oscillations with large oscillation amplitudes, thus compromising the structural integrity of the combustion system and significantly limiting the operability of the gas turbine engine [8, 9]. As a result, predicting thermoacoustic instability constitutes a major concern for manufacturers when designing reliable combustion systems.

To numerically investigate the occurrence of thermoacoustic instability in the combustor, a common practice is to combine an acoustic solver with a flame response model to calculate the frequency, growth rate, and structure of the thermoacoustic mode in frequency domain. This is a typical "divide and conquer" approach [10], where the acoustic solver describes the propagation of the acoustic waves and the flame response model describes the response of the heat source to flow perturbations and acts as the source term in the wave propagation equation. In practice, acoustic network models [11–13] or finite-element/-volume based Helmholtz solvers [14, 15] are extensively used as the acoustic solver. In linear regime, various flame response models can be employed, ranging from a simple $n-\tau$ model [16] (n denotes the gain and τ denotes the

time delay) in frequency domain to a more sophisticated and realistic flame impulse response model [17, 18] in time domain. In nonlinear regime, a flame describing function [19] is usually employed to provide flame dynamic responses at various frequencies and velocity perturbation amplitudes.

In practice, however, obtaining a reliable calculation of the thermoacoustic eigenmodes is a non-trivial task. First of all, due to the nonlinear nature of the governing eigenvalue equation, thermoacoustic behavior of the combustor is known to be highly sensitive [20] to small variations in flame model parameters and acoustic boundary parameters such as the reflection coefficients at the combustor inlet R_{in} and at the outlet R_{out}. Secondly, in practice, uncertainties are always present in those parameters [21]: flame model parameters are usually identified from noisy experimental or computational data. For acoustic boundary parameters, an installation of an acoustic damping device or a change of the downstream turbine working condition could easily modify the value of R_{out}, for example. "Uncertainty" means that those parameters may not stay at some fixed nominal values, but rather display stochastic features, or simply not known. As a consequence, once the key parameters governing the thermoacoustic property of the combustor deviate from their nominal values, the performance of the combustor may deteriorate significantly. In extreme cases, a combustor estimated as stable using the nominal input parameters may become unstable affected by the stochastic parameter fluctuations.

Under this background, rigorously managing the uncertainties [22] present in the workflow of thermoacoustic instability calculations constitutes an essential step towards reliable combustor design. Specifically, competences need to be developed to efficiently propagate the input uncertainties to the output of interest (e.g., modal frequency and growth rate values), as well as to effectively mitigate the negative impact of variational output induced by the uncertain inputs. Towards that end, instead of settling for deterministic analyses, a framework of probabilistic system modeling has to be implemented, which provides means to account for the uncertain inputs from a statistical point of view and minimize their impact on the thermoacoustic performance of the combustor.

1.2 State-of-the-art

In terms of managing the uncertainties associated with thermoacoustic instability predictions, pioneer work has been primarily focused on forward uncertainty quantification (UQ). More specifically, the research objective is to calculate the modal instability risk P_f, i.e., the probability that the mode under investigation is unstable, induced by the uncertain flame model parameters and/or acoustic system parameters. Sampling-based Monte Carlo simulation [23] is the most popular approach adopted to perform the above-mentioned UQ task: firstly, a large number of realizations ($\sim o(10^4)$) are drawn according to the statistical distributions of the uncertain inputs. Then, for each input realization, its corresponding modal growth rate value is calculated by calling the thermoacoustic solver. Finally, P_f can be conveniently determined by dividing the number of realizations with positive growth rate values by the total number of realizations.

Obviously, calling potentially expensive thermoacoustic solvers at every Monte Carlo iteration would induce prohibitive computational cost for the associated UQ analysis. To address the

efficiency problem, *surrogate modeling* techniques have been explored recently. In this framework, only a small number of input samples and their corresponding modal growth rate values are collected to constitute a training dataset. Subsequently, machine learning techniques are employed to train a surrogate model based on the generated training dataset, which can accurately reproduce the solver output (within the region of training) given the same input, yet very fast to evaluate. Subsequently, instead of using the original thermoacoustic solver, the Monte Carlo procedure mentioned above can be directly applied to the trained surrogate model to derive P_f. Since the computational time for each iteration of Monte Carlo simulation is greatly reduced, the efficiency and affordability of the overall UQ analysis improves significantly.

In this context, a number of pioneer work have been done to accelerate P_f calculation by using various surrogate modeling techniques:

Linear regression

Linear regression is a simple technique yet proves to be useful in various situations [24]. It is easy to fit and convenient to interpret. It builds a linear model between the output and the inputs. Subsequently, a Monte Carlo procedure can be directly applied to the trained linear model and instantly obtain the probability density function (PDF) of the output. However, the limitation of the linear regression method is also prominent: a linear model may not be sufficient to capture the potentially highly nonlinear relationship between the modal eigenvalue and flame or acoustic parameters, which may further compromise the accuracy of the overall UQ analysis.

Ndiaye *et al.* [25] adopted the linear regression method to approximate a Helmholtz solver to facilitate affordable UQ analysis of a single swirled burner combustor. Flame gain n and time-delay τ were considered as the uncertain parameters, and their standard deviation values were set to be 10% of their respective nominal values. Two linear models were fitted to map from the uncertain flame model parameters to the modal growth rate of the first acoustic mode ($f_0 = 121Hz$). The first model was directly built upon n and τ, while the second one was built upon the real and imaginary part of the corresponding flame transfer function (FTF) evaluated at f_0. By applying a Monte Carlo procedure to the fitted linear models, improvements in the efficiency of the UQ analysis were observed in the case studies. An interesting observation is that the second model displayed higher approximation accuracy, which is attributed to the fact that it incorporated physical nonlinearity by using FTF values as the input parameters.

Adjoint-based approach

Adjoint methods provide a convenient way to calculate the sensitivity of the output with respect to all the input parameters [26]. It has an obvious advantage in computational efficiency over the conventional finite difference approach, as in adjoint methods, only one solve of the governing equations and one solve of the adjoint equations are sufficient to yield sensitivity to all the parameters [27]. Adjoint methods have been extensively used in thermoacoustic analysis for the purpose of sensitivity analysis [20, 28–31], stability analysis [32, 33], parametric study [34], optimization [35, 36], as well as passive control [37]. Recently, the potential of the adjoint methods have been further exploited to conduct UQ analysis. More specifically, the gradients calculated by the adjoint methods are leveraged to construct a local polynomial model via Taylor expansion, which facilitate instant uncertainty propagation, thus achieving several orders of magnitude increase in computational efficiency over the conventional Monte Carlo simulation [38]. Note that the adjoint-based UQ analysis assumes that each input sample is a small pertur-

bation around the nominal input values. As a result, adjoint method may only be able to handle small input uncertainty ranges where the perturbation assumption is valid. However, Silva *et al.* [34] recently proposed an adjoint-based, high-order perturbation method to efficiently calculate parametric thermoacoustic maps of typical combustion chambers. By leveraging on higher order sensitivity information, it may be possible for adjoint approaches to handle UQ analysis with large perturbations around the nominal input values.

Magri *et al.* [39] applied an adjoint-based approach to a network model of an annular combustor to obtain the sensitivity of modal growth rate with respect to uncertain flame parameters n and τ. Those sensitivity information was subsequently integrated into the Monte Carlo procedure, yielding highly efficient quantification of the uncertainty in modal growth rate predictions. The maximum standard deviation values investigated in [39] were set to be 10% of the nominal values of n and τ, respectively. For the maximum standard deviation values, the second-order adjoint method has to be employed to ensure prediction accuracy. Silva *et al.* [40] developed an in-house Helmholtz adjoint solver and focused on quantifying uncertainties in thermoacoustic stability analysis of a turbulent premixed swirled EM2C combustor [41, 42]. In that study, $\boldsymbol{\eta} = (n, \tau, r, \phi)$ are considered as uncertain parameters, where n and τ are the flame model parameters, r and ϕ are the magnitude and the phase of the reflection coefficient at the combustor outlet. Uniform distributions with a range of $0.9\boldsymbol{\eta}^0 \sim 1.1\boldsymbol{\eta}^0$ were assumed for those uncertain parameters. Consistent with [39], [40] also pointed out that a second-order adjoint method is necessary to accurately reproduce the results yielded by applying Monte Carlo procedure to the original Helmholtz solver. Mensa *et al.* [43] incorporated the adjoint perturbation theory into acoustic network and Helmholtz solvers and presented an analytical strategy to determine the stability margin and calculate the associated risk factor. Flame gain n and time delay τ were treated as uncertain parameters, with $\pm 10\%$ around their respective nominal values as the considered uncertain ranges. The effectiveness of the proposed strategy was demonstrated via case studies on a Rijke tube model and the same combustor from EM2C [41].

Polynomial Chaos Expansion

Polynomial Chaos Expansion (PCE) [44, 45] enjoys mathematical elegance and has been successfully applied to address different UQ problems in various domains [46–50]. Several open source software packages exist that extensively documented and implemented this method [51–54]. PCE approximates the output of interest as a truncated sum of multivariate polynomials orthonormal with respect to input probability density distributions [55]. Once the weighting coefficients of the polynomial basis terms are calculated, a surrogate model is automatically obtained, which maps from the input parameters to the output. Depending on how exactly the weighting coefficients are calculated, PCE can be categorized as intrusive type or non-intrusive type. For intrusive PCE, new governing equations for the coefficients are derived based on a Galerkin-projection reformulation of the original model equations [56]. Non-intrusive PCE, on the other hand, treats the underlying governing computational model as "black-box" and uses sampling-based approach to compute the coefficients. Several methods exist for that purpose. For example, projection method [57] can be used to take advantage of the orthogonality property of the polynomial basis. Using this method, the computation of the coefficients can then be cast as a numerical integration problem, which can be efficiently solved using quadrature rules. As another example, the least-square method [58] can be employed to derive the coefficients. Using this method, representative samples are drawn from the input distributions and their corresponding outputs are calculated via the employed computational model. Subsequently, a simple linear

regression is performed on the collected dataset to derive the weighting coefficients. Compared to the quadrature rules with fixed quadrature points, the least-square approach has the advantage that it allows an arbitrary number of points with flexible sampling locations.

Avdonin *et al.* [59] used non-intrusive PCE to efficiently calculate P_f. In a first case study, a turbulent swirl combustor modeled [41] by the Helmholtz equation was considered. A total of four parameters, i.e., n, τ, the magnitude and the phase of the outlet reflection coefficients, were treated as uncertain. For each parameter, a uniform distribution extending $\pm 10\%$ of the nominal value was adopted to describe the uncertainty statistically. Non-intrusive PCE managed to deliver highly accurate P_f calculation while consuming much less computational resources, compared to the conventional Monte Carlo simulation. For the second case study, the impact of uncertain CFD model parameters on the modal growth rate predictions was considered. The simulated configuration was a laminar slit burner (Kornilov flame [60]). The CFD simulations were used to derive the FTFs given the specific realizations of the input parameters. Subsequently, those identified FTFs were fed into an acoustic network model to compute the corresponding modal growth rate value. A total of three CFD model parameters, i.e., flow velocity ($u = 0.4 \pm 0.02$ m/s), burner plate temperature ($T = 398 \pm 50$ K) and equivalence ratio ($\phi = 0.8 \pm 0.1$), as well as two acoustic system parameters, i.e., the magnitude of the reflection coefficient at the combustor inlet $R_{in} = 0 \sim 0.4$ and outlet $R_{out} = -0.4 \sim 0$, were perceived as uncertain. For both case studies, a second order expansion is found to be necessary to accurately quantify the uncertainties in growth rate predictions.

Avdonin *et al.* [61] further investigated the impact of uncertain operating conditions on the dynamics of a premixed laminar flame by using non-intrusive PCE. In that study, flame dynamics were represented by a flame impulse response (FIR) model, which was identified via advanced system identification [17] on a broadband excited CFD simulations. The same uncertain parameters with the same variational ranges in [59] were considered here. The study demonstrated that by using non-intrusive PCE, it becomes affordable to quantify the variations in FIR model coefficients induced by the uncertain operating conditions. In addition, sensitivity analysis was conducted to understand which operating parameters contribute the most to the uncertainties presented in the FIR model coefficients. This is possible since PCE permits a convenient calculation of Sobol index [62], which is a widely employed index in global sensitivity analysis [63].

The above-mentioned studies only addressed UQ problems involving a small number of parameters (≤ 4). However, for practical thermoacoustic instability analysis, the flame model alone would introduce many uncertain model parameters. For example, for an annular combustor with multiple burners, if a simple $n-\tau$ model is adopted for each flame, the total number of flame model parameters will correspond to two times the number of burners, which could be significantly larger than the number of inputs considered in the previous studies. In surrogate modeling, it is a well-known fact that the cost of constructing an accurate surrogate model scales exponentially with the number of input variables. This is the famous "curse of dimensionality". As a result, a naive application of the surrogate modeling techniques to connect primary flame parameters and the modal growth rate would not be feasible in practices. To circumvent the efficiency problem, a dimensionality reduction treatment prior to training the surrogate models is often necessary. Under this framework, dimensionality reduction techniques are applied first to extract the key parameters in the input-output relationship under investigation. A successful dimensionality reduction is achieved if the number of the key parameters is smaller than

the number of prime parameters. Subsequently, surrogate models can be trained directly on the key parameters instead of the original high-dimensional inputs, thus potentially improving the training efficiency and reducing the associated computational cost.

In the field of thermoacoustic UQ analysis, the *Active Subspace* approach [64] has been employed to reduce the flame model dimensionality. This approach aims to identify active directions within the original input parameter space, such that along those directions the output varies the most. By projecting the input samples to the identified active directions, the corresponding *active variables* can be obtained, which manifest themselves as linear combinations of the original inputs. Those active variables capture the essences of the relationship between high-dimensional flame parameters and the corresponding modal growth rate values. As a result, subsequent surrogate model building only need to consider those active variables, thus achieving dimensionality reduction and accelerating the desired uncertainty propagation process.

Bauerheim *et al.* [65] adopted Active Subspace approach to calculate P_f of the azimuthal acoustic mode of two coupled annual cavities. In that study, an acoustic network was derived to describe the nonlinear dispersion relation for azimuthal modes. 19 flames were considered, where each flame is represented by an uncertain $n-\tau$ model. As a result, a total of 38 uncertain parameters entered into the UQ analysis. This number is significantly larger than that used in the previous studies. 5% and 10% of their mean values were assigned as the standard deviation values of n and τ, respectively. The case studies indicated that the Active Subspace approach managed to identify three active variables out of the original 38 inputs. Those active variables were subsequently exploited by fitting the modal growth rate response with polynomials, which were further fed into the Monte Carlo procedure to accelerate UQ analysis. In addition, physical interpretations of the identified active variables were investigated in [65], which unveils underlying phenomena controlling the thermoacoustic stability of the investigated configuration.

Magri *et al.* [39] continued the work of [65] by proposing a hybrid approach, where the potential of Active Subspace methodology was further augmented by adjoint method in conducting UQ analysis. In their work, adjoint method was employed for efficiently computing the gradients to facilitate active subspace identification. As a result, fewer thermoacoustic solver calls are needed in applying the Active Subspace approach and an overall more efficient and accurate UQ analysis was achieved.

Another method that has received attention recently in performing thermoacoustic UQ analysis is the intrusive PCE. Compared with its non-intrusive counterpart, the intrusive version requires projecting the governing equations under consideration onto the space spanned by the polynomial basis functions to form a new set of equations for the PC coefficients. Due to this design of new solvers/codes, intrusive PCE is not general for various systems and may require some overhead effort for specific applications. However, the advantage gain is that only a one-time solution of the reformulated governing equations is needed to calculate the PC coefficients for the model outputs. As a result, significant reduction in computational cost may be obtained as repetitive evaluations of the original model required by non-intrusive PCE can now be avoided.

Silva *et al.* [66] explored the feasibility of using intrusive PCE to quantify the uncertainty of combustion noise prediction of confined flames. In that study, a state-space model [67] was adopted to describe the thermoacoustic properties of a turbulent swirl premixed combustor [68,

69]. The corresponding stochastic state-space model was derived by substituting the polynomial basis expansions into all elements of the original state-space representation of the burner. The acoustic waves, gain and phase of the flame response, and acoustic reflection coefficients were considered as uncertain variables that entered into the intrusive PCE analysis. Thanks to its intrusive nature, only one solve of the stochastic state-space model is required by the PCE to determine the PDFs of the acoustic quantities at each node of the discretized domain. A comparison with the reference Monte Carlo simulation demonstrated the effectiveness of the developed intrusive PCE framework. In addition, [66] also discussed how to perform global sensitivity analysis given the computed PCE coefficients.

Despite the fact that remarkable progress has been made by the previous studies, critical gaps still exist that worth further considerations for practical uncertainty management in thermoacoustic instability analysis:

First of all, small parameter variational ranges. The variational ranges for the uncertain variables considered in the previous studies were generally small, i.e., 10% around the corresponding nominal values, within which various techniques summarized above were demonstrated to be effective. However, larger input parameter space is normally expected in practical thermoacoustic instability analysis, where the validity of the previously-reviewed approaches is unknown. This issue becomes even more prominent in robust design where mitigating modal instability risk is the main objective (discussed below), as larger parameter variational ranges have to be accommodated to allow optimization routines to fully explore various parameter combinations, thus increasing the change of locating the global optimum.

Second, non-flexible statistical descriptions of the uncertain parameters. Previous studies tended to consider a single probability distribution type for all the investigated uncertain parameters (e.g., uniform distribution, Gaussian distribution, etc.). In practice, however, it is not uncommon to have mixed distributions for different uncertain parameters, with potentially complex correlations between the parameters. In addition, the knowledge regarding the statistical properties of the uncertain parameters may be updated as the combustor design process evolves. Those two observations may pose serious challenge to the PCE methodology, despite its extraordinary performance reported in [59, 61, 66]. This is partially due to the fact that PCE demands the information of the PDFs of the considered uncertain parameters for selecting the optimum polynomial basis functions. It works the best when those PDFs are pre-defined and independent. Otherwise, complicated isoprobabilistic transformations [70] have to be employed to convert PDFs to their standard forms before submitted to PCE modeling.

Third, UQ analyses for sophisticated flame models are lacking. Previous studies focused exclusively on quantifying modal instability risk P_f induced by an uncertain $n-\tau$ model. The $n-\tau$ model is simple and easy to interpret, but possesses limited modeling capability in describing complex flame-acoustic interaction phenomenon usually encountered in practical thermoacoustic instability analysis. As a result, performing UQ analysis on $n-\tau$ model may only be of limited relevance for realistic applications.

One of the more sophisticated and realistic flame models is the flame impulse response (FIR) model [17]. It adopts a time-domain perspective to describe the unsteady response of a flame to acoustic or flow perturbations, which facilitates convenient physical interpretation of important features of the flame response [18]. Through z-transform [71], the FIR model can be easily

converted to the flame transfer function (FTF) in frequency domain, which can be subsequently integrated into the acoustic solvers to predict thermoacoustic behaviors of the combustor. FIR model is commonly derived by combining CFD simulations with advanced system identification procedures. The derived FIR model may be uncertain due to the statistical nature of the CFD simulations, low signal-to-noise ratio, as well as the limited length of CFD time series [72]. FIR model has already been used on several occasions, while the impact of its uncertainty on the prediction of the thermoacoustic eigenmodes is not fully investigated yet. This may attribute to the fact that UQ analysis with respect to FIR model constitutes a challenging high-dimensional problem. Active Subspace approach [39, 65] may shed lights on this problem, but further studies are necessary to confirm the expectation.

Fourth, a general UQ framework for considering combined uncertain acoustic parameters and uncertain high-dimensional flame models is lacking. In the paradigm of combining acoustic solvers with flame models to predict thermoacoustic instability, uncertainties may originate from both the acoustic side and the flame side, with the flame side potentially containing a high number of uncertain parameters. Previous studies were either dealt with uncertainties in both acoustic system parameters and simplistic $n-\tau$ models for flame dynamics [25, 40, 43, 59, 61, 66], or solely focused on tackling uncertainties from high-dimensional flame models [39, 65]. As a result, a more general UQ framework than the previously-reviewed methods is necessary to deliver effective thermoacoustic UQ analysis in realistic settings. Although the Active Subspace approach [39, 65] has demonstrated its capability in performing dimensionality reduction for high-dimensional flame models, a naive application of the approach to the mixed sources of uncertainty may not produce fruitful results. This may due to the fact that acoustic parameters and flame parameters are governed by different physical mechanisms. Consequently, a linear combination of the mixed parameters may not be sensible and may mislead the desired UQ analysis. In summary, currently available UQ techniques may be insufficient to address UQ problems with uncertainties coming from both acoustic parameters and high-dimensional flame models. Novel solutions are much needed for realizing affordable and reliable uncertainty management for thermoacoustic instability analysis in practical settings.

Fifth, UQ techniques for nonlinear thermoacoustic instability analysis are lacking. Previous studies focused exclusively on performing UQ investigations on linear thermoacoustic instability analysis, where the objective was to quantify the variations of modal frequencies and growth rates induced by the uncertain inputs. For nonlinear thermoacoustic instability predictions, Palies *et al.* [42] and Silva *et al.* [15] showed that an accurate prediction of limit cycle frequency and amplitude could be easily compromised by the uncertainties embedded in the system acoustic damping and nonlinear flame response model. However, to the author's knowledge, no rigorous UQ study has been done yet to quantify the variations of limit cycle frequency and amplitude induced by the uncertain acoustic flame model parameters. Consequently, it is unclear whether or not the previously-reviewed approaches are still effective when extending to performing UQ analysis in nonlinear settings. Considering the vital importance of obtaining reliable limit cycle predictions for practical combustor design, relevant strategies need to be developed to accurately and efficiently address uncertainty propagation problems in nonlinear thermoacoustic instability analysis.

Sixth, uncertainties contained in the surrogate models are not considered. The accuracy of a surrogate model relies heavily on the amount of training samples [73]. In practice, as the investigated parameter space becomes larger, a larger number of training samples is usually expected

to guarantee that the trained surrogate model is accurate. Meanwhile, as the fidelity of the thermoacoustic solver becomes higher, the induced computational cost to generate each training sample also becomes higher. As a consequence, it is not uncommon that only a small number of training samples can be afforded in many practical cases, thus leading to a potentially "inaccurate" surrogate model. This surrogate model uncertainties may propagate downstream and compromise the reliability of the modal instability risk calculations [74]. Unfortunately, the surrogate model uncertainty is largely ignored in previous studies, despite the fact that surrogate models have been widely adopted as the workhorse of thermoacoustic UQ analysis. In order to obtain robust modal instability risk predictions, relevant schemes need to be developed to quantify and reduce the surrogate model uncertainty.

Finally, investigations regarding mitigating the modal instability risk are lacking. Merely quantifying the modal thermoacoustic instability risk is not the end goal. Another important task would be to mitigate the modal instability risk, which can be perceived as a natural second step after risk quantification. This is in the domain of robust design [75], where the objective is to optimize the design parameters such that the performance of the system under investigation is insensitivity to various input uncertainties. Robust design task is also known as *optimization under uncertainty* [76], since an UQ module has to be integrated into the optimization routine, which computes the statistics of the outputs that can be fed into the objective and/or constraint functions. In the context of thermoacoustic instability analysis, however, UQ studies and optimization investigations were conducted separately and a synergy between these two aspects is lacking in previous works. For example, efforts have been made to suppress thermoacoustic instability by configuration optimizations in [77–80]. However, those studies only approached the optimization problems via a deterministic point of view. Consequently, the robustness of their optimized results may be challenged by the uncertainties present in the data or the model [81]. On the other hand, previously-reviewed studies in UQ analysis solely focused on developing efficient uncertainty propagation strategies to quantify the modal instability risk P_f, without any discussion of how to mitigate those risks. It is worth mentioning that an integration of UQ and optimization into the robust design framework is a non-trivial task, as it poses more strict requirements on the employed UQ schemes. This is mainly due to the following two reasons: first of all, robust design demands that the employed UQ scheme is very efficient with negligible computational cost, as a full UQ study has to be conducted at each iteration of the optimization routine. Secondly, robust design demands that the employed UQ scheme is valid over a relatively large parameter space, as the optimization routine needs to explore various parameter combinations to increase the chance of identifying the global optimum.

1.3 Scope

To fill the gap between what has already been achieved and what is required in practice, this thesis focuses on developing a comprehensive uncertainty management framework for thermoacoustic instability analysis in practical settings. This uncertainty management framework consists of a risk quantification part and a risk mitigation part. The risk quantification part addresses the issue of how to efficiently propagate uncertainties from uncertain flame and acoustic parameters to the prediction of the modal eigenvalues, while the risk mitigation part addresses the issue of how to eliminate the instability risk by implementing the principles of robust design.

The present thesis consists of the following deliverables:

1. Developing dimensionality reduction strategies to efficiently quantify the modal instability risk P_f induced by an uncertain high-dimensional FIR model [82, 83]. The FIR model is considered to be identified via applying advanced system identification routines on the broadband-excited CFD time series.

2. Developing online monitoring routines to assess the confidence in FIR model identification as longer CFD time series become available [83]. This is beneficial for practitioners as it helps to determine the required length of the CFD simulations to obtain FIR model identification with satisfactory confidence.

3. Developing a general surrogate-based framework for efficiently and accurately performing uncertainty propagation in thermoacoustic instability predictions [84]. This framework is designed to accommodate uncertain parameters with large variational ranges and flexible statistical descriptions. In addition, this framework is capable of considering uncertainties from both high-dimensional flame models and acoustic system parameters.

4. Performing UQ studies for nonlinear thermoacoustic instability predictions [84]. Here, the research goal is to quantify the variations of limit cycle frequency and amplitudes induced by an uncertain FDF dataset and other uncertain acoustic system parameters. This dataset describes flame frequency response (gain and phase) at different combinations of forcing frequency and velocity perturbation amplitude. Its uncertainty is caused by imperfect data collection process, i.e., noisy experimental measurements are only available at limited forcing frequencies and velocity amplitudes.

5. Identifying various risk mitigation tasks in thermoacoustic instability analysis, and providing mathematical formulations and surrogate-based solution strategies for each task [85].

6. Quantifying the variation of the modal instability risk prediction induced by the uncertain surrogate model [86]. Here, the uncertainty presented in the surrogate model is caused by the limited number of training data.

7. Developing effective surrogate model training scheme that can sequentially enrich the training dataset and significantly reduce the epistemic uncertainty of the surrogate model, thus achieving a more robust calculation of the modal instability risk[86].

To better highlight the gaps in thermoacoustic UQ analysis that the present thesis aims to address, Fig. 1.1 summarizes the established UQ methods, remaining issues, as well as the current effort to address those issues.

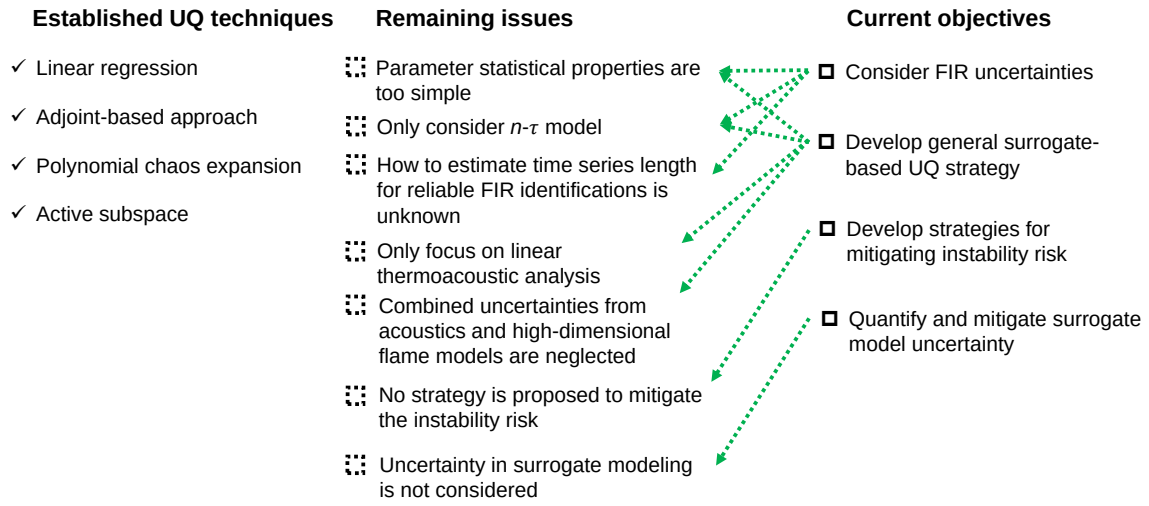


Figure 1.1: A summary of the established UQ methods, remaining issues and the current effort to address those issues in thermoacoustic instability analysis.

Constrained by the scope of the present thesis, only a limited number of topics are addressed in a larger field of uncertainty quantification in thermoacoustic instability analysis. Topics that are not discussed in this thesis include:

1. Inverse uncertainty propagation and global sensitivity analysis.

Inverse uncertainty propagation and global sensitivity analysis constitute two important tasks in the broader context of uncertainty quantification, in addition to forward uncertainty quantification and robust design. Inverse uncertainty propagation aims to derive the input uncertainty from observed noisy output data [87]. "Inverse" here represents that the information is flowing from the output to the input. Bayesian statistics [88] are commonly adopted for realizing inverse uncertainty propagation, where one's prior belief of the input parameter distributions is updated by learning from the observed output data, thus arriving at the posterior distribution of the input parameters. In the context of thermoacoustic instability analysis, Yu *et al.* [89, 90] employed a Bayesian method called *ensemble Kalman filter* [91] to derive the parameters of a reduced-order model that describes the dynamics of a premixed flame from high-fidelity experimental and/or simulation data. Through Bayesian statistics, the uncertainty of the parameter inference can be easily and robustly obtained.

Global sensitivity analysis (GSA) [63] aims to investigate which uncertainty source contributes the most to the variation of the output. Conceptually, GSA differs from adjoint-based local sensitivity analysis [38], since GSA assess the impact of inputs on the output by looking at the entire input space rather than at a point in that space. Variance decomposition approaches (e.g., Sobol indices) are frequently adopted to deliver GSA, where the total variance of the simulation output is partitioned into components attributable to each uncertainty source and its interaction

with other uncertainty sources [92]. In the context of thermoacoustic instability analysis, [61] and [66] used Sobol indices calculated by the PCE method to rank the parameters according to their contributions to the variance of the output of interest.

The current thesis only focuses on forward uncertainty propagation and risk mitigation aspect of uncertainty quantification. The input distributions are either assumed or directly obtained from upstream analysis procedures. Therefore, no inverse uncertainty propagation study is conducted to derive the input distributions. However, considering the fact that accurately and robustly characterizing the probability distributions of the uncertain input parameters lays the foundation for all the subsequent UQ analysis, further efforts should be put into developing effective schemes of inverse uncertainty propagation for thermoacoustic instability analysis. Nevertheless, it is worth emphasizing that the UQ strategies proposed in the framework of the current thesis are independent of specific input parameter distributions. Therefore, those strategies would remain valid when the considered input distributions are derived from advanced inverse uncertainty propagation procedures in future studies.

As for GSA, considering that the identification of important/unimportant parameters is not a primary goal of the present thesis, only [82] briefly touched upon this topic and no systematic GSA study is further performed. However, GSA may deliver several benefits: first of all, GSA could be leveraged to identify key parameters controlling the thermoacoustic instability; secondly, by ignoring those "unimportant" parameters identified by the GSA, reduced number of parameters will enter into the subsequent UQ analysis, thus a potential dimensionality reduction can be achieved; finally, the ranking results produced by the GSA could inform how to prioritize resources in reducing the uncertainties in different input sources, thus potentially achieving a maximum reduction in output uncertainty. Therefore, future studies are desired to further explore the potential of GSA in thermoacoustic instability analysis.

2. Solver verification and validation.

Verification and validation (V&V) aim to assess the credibility and establish the quality of the computational model [93–95]. Verification is concerned with the characterization of the numerical errors (e.g., discretization, iteration, and computer round-off) and demonstrating that a simulation code correctly solves the underlying mathematical model equations. Therefore, verification procedure tackles the numerical uncertainty, which is largely an exercise of computer science and mathematics and it usually involves comparisons of the simulation code with some exact analytical solutions [38]. Validation, on the other hand, tackles the model form uncertainty and it demonstrates that the results of the simulation code accurately reflect the actual physical process. Validation is largely an exercise of physics and engineering and usually involves comparisons of the simulation results with experimental measurements [38].

In practice, V&V is carried out before the standard UQ analyses that quantifies the variation of the model output induced by the uncertain model inputs, as the reliability of the underlying computational model needs to be fully assessed before it can be used to drive the subsequent UQ analyses. In the current thesis, the employed thermoacoustic solvers, i.e., an acoustic network model and a Helmholtz solver, have already been verified and validated in [96] and [40], respectively, although not conducted in a probabilistic framework. Since the current thesis mainly focuses on developing procedures to quantify and mitigate model output uncertainty directly induced by the model inputs, the underlying thermoacoustic solvers are simply assumed to

be perfect such that no solver uncertainty presents. As a consequence, the fidelity of the UQ analyses performed in the present thesis only approximates the fidelity of the employed thermoacoustic solvers. Future studies should investigate solver verification and validation issues from a probabilistic point of view to gain further insights regarding the solver numerical errors and discrepancies between solver outputs and physical reality. Subsequently, those quantified numerical errors and model bias should be integrated into the UQ procedures developed in the present thesis to achieve more robust and reliable thermoacoustic instability analysis and design.

3. Potentiality of other surrogate modeling techniques than Gaussian Process.

The present thesis mainly employs the Gaussian Process (GP) as the surrogate modeling technique, which is a variant of the supervised machine learning methods [97]. Other methods under this category include polynomial regression [98], polynomial chaos expansion (PCE), support vector machines (SVM) [99–101], decision trees [102], neural networks [103–105], etc. In the context of thermoacoustic instability analysis, in addition to the work reviewed previously where PCE were adopted [59, 61, 66], Jaensch *et al.* [106, 107] investigated the capability of neural networks for deducing nonlinear flame response models from a CFD simulations. An ensemble of neural networks were trained in that study to derive information regarding the modeling uncertainty. McCartney *et al.* [108] compared the capability of various machine learning approaches in interpolation and extrapolation of flame describing function, and performed preliminary UQ analyses to assess the impact of FDF approximation errors on the prediction of limit cycle amplitudes. In a broader context of UQ, PCE, SVM and neural networks have also been employed to facilitate efficient uncertainty propagation and risk calculation [49, 109, 110]. GP is chosen in the present thesis due to its accuracy, flexibility, easy-to-train, easy-to-interpret, relatively low computational cost, and most importantly, it naturally provides estimates of the prediction uncertainty, which is a rare property among other surrogate models [38]. Nevertheless, future studies are desired to thoroughly compare the strengths and weaknesses of GP with other surrogate modeling techniques in terms of aiding the uncertainty management in thermoacoustic instability analysis. Moreover, it is worth exploring the possibility of combining various techniques instead of relying on just one specific technique. This may help to obtain further gain in the accuracy and efficiency of surrogate modeling. For example, GP and PCE were combined in [111], where the PCE model is used as the general trend term in the GP formulations. This integration is possible thanks to the flexibility and extensibility of the GP approach. Case studies showed that the performance of the new approach is better than, or at least as good as the individual approaches for small training samples. Other examples of combining various surrogate modeling techniques can be found in [112–114].

4. Other uncertainty modeling techniques than probability theory.

Parameter uncertainties can be modeled by a variety of concepts. Conventional probability theory is the most popular approach to measure uncertainty, where each of the uncertain parameters is associated with a probability density function (PDF) [87]. In practice, however, it is not uncommon that assigning PDFs to uncertain parameters is not straight-forward. This issue may arise when the knowledge of a practitioner regarding the input parameters is limited: either a specific PDF form cannot be decided in the first place, or the form of PDF is determined but the parameters of this PDF cannot be decided [74]. Under this background, other concepts to represent uncertainty have been devised, including Bayesian hierarchical models, probability-boxes, evidence theory, fuzzy distributions, etc [115]. Compared with the probability theory,

other uncertainty representation approaches allows a clear separation of aleatory uncertainty (natural variability) and epistemic uncertainty (lack of knowledge) [116], therefore possessing better modeling capability in practical applications.

The current thesis adopts the conventional probability theory to describe input parameter uncertainties. Further studies are desired to investigate using other uncertainty formulations in uncertainty management of thermoacoustic instability analysis. Nevertheless, it is worth mentioning that the dimensionality reduction schemes [82, 83], as well as the surrogate-based techniques for uncertainty propagation [84] and risk mitigation [85] developed in the current thesis are independent of specific descriptions of the uncertain input parameters. As a result, the proposed procedures would still be valid when more sophisticated uncertainty measurement approaches are used in future studies.

5. Uncertainty management for annular combustors.

The current thesis only considers uncertainty presented in predicting thermoacoustic instabilities for longitudinal combustor configurations. In practice, annular combustors are widely adopted in modern gas turbines. However, as indicated in the work of [65], a prominent challenge in conducting UQ analysis for annular configurations is the associated high number of uncertain parameters. For example, although only a simple two-parameter $n-\tau$ model is adopted for each burner in [65], the total number of uncertain parameters under investigation accumulates to 38, as there were 19 burners in the considered annular configuration. As a result, when sophisticated flame models with more parameters are employed for each burner, this multiplication effect would be even more dramatic, thus posing a significant challenge for performing thermoacoustic UQ analysis in annular combustor. In addition, transverse modes exist alongside longitudinal modes in annular combustors [5]. Those two types of modes have different physics in the coupling between the acoustics and the flame response. As a result, distinct stochastic features of the eigenvalues of transverse and longitudinal modes could be observed under the same input parameter uncertainties. This could potentially lead to problems in performing thermoacoustic robust design for annular combustor, as it is non-trivial to optimize tunable system parameters to mitigate the instability risks for both transverse and longitudinal modes. Finally, due to a more complex geometry, solving eigenvalue equations for an annular combustor is potentially much more expensive than a longitudinal combustor, given the same acoustic modeling approaches (e.g., network model, Helmholtz solver, etc.). As a result, fewer samples may be afforded in surrogate model training, thus leading to a larger surrogate model uncertainty that may compromise the reliability of the UQ analysis. Despite the above-mentioned challenges for performing UQ in annular combustors, the uncertainty quantification and risk mitigation strategies developed in the current thesis could serve as foundations for designing more advanced approaches that specifically address the issues encountered in annular configurations.

6. Uncertainty propagation from CFD simulations to flame model identifications.

The flame models adopted in predicting thermoacoustic instability can be identified from high-fidelity CFD simulations. For example, the FIR model investigated in [82–84] is identified via applying advanced system identification procedures on broadband-excited velocity and heat release rate fluctuation time series obtained from CFD simulations [117]. During the FIR model identification process, due to low signal-to-noise ratios in turbulent combustion (low-quality training data), finite length of the CFD time series (limited training data), etc., the estimated

FIR model will inevitably contain uncertainties (characterized by the confidence intervals of the coefficients). The main objective of [82–84] is to quantify the modal instability risk induced by the uncertain FIR model, whose uncertainty is originated from the model identification process.

However, the derived uncertain FIR model considered in [82–84] is based on a single CFD simulation with deterministic values for boundary conditions, turbulent combustion models, as well as other CFD parameters. In reality, it is rarely the case that those CFD parameters could be uniquely identified. Consequently, any variation/uncertainty in those CFD parameters could lead to another set of FIR model [61] with different uncertainty descriptions. As a result, in addition to the identification process, the CFD data generation process also contributes to the uncertainties in the FIR model.

Uncertainties in CFD simulations have long been recognized in the community, as evident by the work of [118]. Those uncertainties usually originate from simplified geometries, initial and boundary conditions, as well as the employed physical models (e.g., turbulent models) [119]. Recently, focus has been shifted to evaluate the impact of uncertain turbulent models on the reliability of the CFD simulations [120–124], as turbulent modeling constitutes a major source of uncertainty in CFD simulations. Parametric uncertainties in turbulent combustion modeling has also been investigated in [125–129]. In the field of thermoacoustic instability analysis, rigorous UQ study to investigate the impact of CFD simulation uncertainty has not yet received enough attention, except the work of [61], where the uncertainties in boundary conditions (flow velocity, burner plate temperature, and equivalence ratio) were propagated to FIR model identification. Other works such as [130] only conducted sensitivity studies to investigate how the flame transfer function of a laminar premixed flame would change when the flame speed, the expansion angle of the burnt gases, the inlet air temperature, the inlet duct temperature, and the combustor wall temperature are varying.

There are two challenges associated with propagating uncertainties from CFD simulations to the flame model identification: first of all, each run of CFD simulation could induce very high computational cost. Consequently, the available computational budget would only permit acquiring a limited number of samples, which poses a serious challenge for applying any data-driven approaches for surrogate modeling and dimensionality reduction. As a result, novel UQ schemes working in the small data regime have to be developed and benchmarked in future studies. Secondly, a concurrent consideration of flame model uncertainties originated from both the CFD data generation process and the flame model identification process is not straight-forward. These two processes are coupled in a serial manner, with the CFD data generation happens first and the flame model identification happens later in the calculation chain. Uncertainty propagation through the interface of these two processes is especially challenging [61], and specific UQ formulations have to be developed in the future work.

1.4 Structure of the thesis

The main results of the current thesis were presented at conferences and in journals. Major publications that summarize the results of this dissertation are reproduced in the appendix. Throughout the manuscript the major publications are labelled. Table 1.1 explains the label and provides a short description of each paper. The first three studies summarized in Table 1.1 contribute

to the risk quantification aspect of the uncertainty management framework investigated in the current thesis, while the later two studies contribute to the risk mitigation aspect.

Table 1.1: Labels and short descriptions of the papers included in the present thesis.

Label	Paper Title	Description
ASME18	Quantification and propagation of uncertainties in identification of flame impulse response for thermoacoustic instability analysis	Calculating modal instability risk induced by uncertain FIR model via the Active Subspace approach
ISC18	Evaluating the impact of uncertainty in flame impulse response model on thermoacoustic instability prediction: A dimensionality reduction approach	Calculating modal instability risk induced by uncertain FIR model via an analytical approach
ISC20	A Gaussian-Process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability prediction	Developing a general surrogate-based framework that can address UQ problems with high-dimensional uncertain flame models
ASME19	Efficient robust design for thermoacoustic instability analysis: A Gaussian Process approach	Exploring various tasks of robust design in thermoacoustic instability analysis and developing efficient solution strategies for each task
ASME20	Reliable calculation of thermoacoustic instability risk using an imperfect surrogate model	Quantifying and reducing the impact of the epistemic surrogate model uncertainty on the risk calculation

In order to highlight the interconnection between publications and facilitate convenient comparisons of different aspects of the developed UQ strategies that are discussed dispersively among publications, the main content of the publications is re-organized into the following chapters. Chapter 2 provides essential background for the adopted acoustic solvers and flame models, with a special focus on discussing the origination of the flame model uncertainty. Chapter 3 depicts a bigger picture of uncertainty management, which puts the work conducted in the thesis into perspective. Chapter 4 and 5 discuss the surrogate modeling techniques and dimensionality reduction techniques developed in this thesis. The investigations of different aspects of those techniques are scattered in various publications. Chapter 4 and 5 aim to give an overview of those ingredients to form a more complete picture. Chapter 6 gives a summary on the outcome of each publication. The interconnection between the individual publications as well as how they contribute to the overall thesis are highlighted. Finally, Chapter 7 concludes the thesis and discuses potential directions for further investigation.

2 Thermoacoustic Framework

This chapter provides essential information for the thermoacoustic modeling techniques adopted in the current thesis. It starts with a presentation of the details of the combustor configurations, followed by introducing the employed acoustic solvers and the flame models. A summary of the thermoacoustic modeling details of the included publications is given in Table 2.1.

Table 2.1: Thermoacoustic models employed in the included publications

Publication	Combustor	Acoustic solver	Flame model
ASME18	BRS	Network model	FIR
ISC18	BRS	Network model	FIR
ASME19	BRS	Network model	$\tau - \sigma$ model
ASME20	BRS, EM2C	Network model Helmholtz solver	$\tau - \sigma$ model
ISC20	EM2C	Helmholtz solver	FIR, FDF

2.1 Combustor configuration

Two combustor configurations are considered in this thesis to perform UQ analysis. The first configuration is the BRS burner (Fig. 2.1), which represents a turbulent premixed swirl burner test rig. This configuration contains a plenum, followed by a duct with an axial swirl generator, as well as a combustion chamber. In the present thesis, a thermal power of 30 kW and an equivalence ratio of $\phi = 0.77$ of perfectly premixed methane-air mixture are considered. Previously, this configuration has been adopted to conduct stability map studies [131, 132], intrinsic thermoacoustic instability studies [133], and flame model identification studies, either via experimental investigations [134], or numerical investigations [117]. In the present thesis, four publications have adopted this configuration for different purposes: in *paper-ASME18* and *paper-ISC18*, BRS configuration is adopted to propagate uncertainties from an uncertain FIR model to the predictions of the modal growth rate. In *paper-ASME19*, the focus is on investigating various robust design tasks in mitigating thermacoustic instability risks. Finally, in *paper-ASME20*, BRS configuration is employed again to develop procedures to quantify and reduce the impact of uncertain surrogate models on the reliability of risk calculation. Instability risks of the first quarter-wave mode and the intrinsic thermoacoustic mode [133] are considered for all studies except *paper-ASME20*, where only the first quarter-wave mode is of interest.

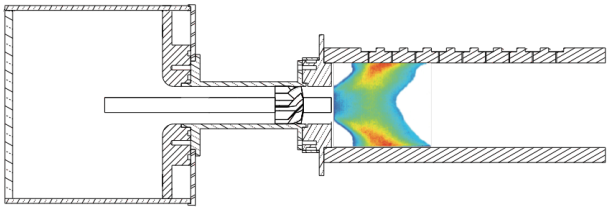


Figure 2.1: Sketch of the BRS burner.

Another configuration adopted in the present thesis is a turbulent swirled laboratory-scale test rig [42], which is depicted in Fig. 2.2. This combustor has been widely investigated in the context of thermoacoustic instabilities [15, 41, 42]. This axisymmetric configuration contains a cylindrical plenum, a convergent duct with a swirl generator, and a cylindrical combustion chamber.

In the present thesis, two works have adopted this configuration for different purposes: in *paper-ISC20*, the focus is to benchmark the effectiveness of the proposed surrogate-based UQ scheme in addressing high-dimensional linear/nonlinear thermoacoustic uncertainty propagation problems. There, configuration C11 [15] is chosen with a plenum length of $l_1 = 224mm$ and a combustion chamber length of $l_3 = 200mm$. The operation condition "A" is considered, which has a flame total power of $1.94kW$ and a mean flow velocity of $2.67m/s$.

In *paper-ASME20*, the focus is to benchmark the effectiveness of the employed active learning scheme to train a GP model with minimum epistemic uncertainty in terms of instability risk calculation. There, configuration C04 [15] is chosen with a plenum length of $l_1 = 124.8mm$ and a combustion chamber length of $l_3 = 400mm$. The operation condition "B" is considered, which has a flame total power of $3.03kW$ and a mean flow velocity of $4.16m/s$.

For both studies, the first longitudinal thermoacoustic mode of the test rig is the mode of interest. In *paper-ASME20*, its associated modal instability risk is calculated. In *paper-ISC20*, for the linear case study, its associated modal frequencies and growth rate values are calculated. For the nonlinear case, its associated limit cycle frequencies and amplitudes are calculated.

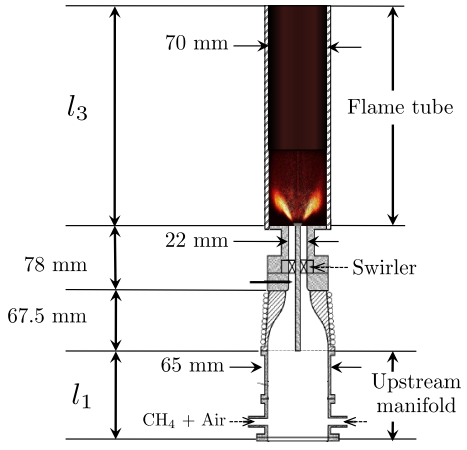


Figure 2.2: Sketch of the EM2C burner.

2.2 Acoustic solver

Thermoacoustic instability is a typical multi-physics/multi-scale phenomenon, encapsulating elements of fluid dynamics, combustion, as well as aeroacoustics. To perform effective thermoacoustic instability analysis, a common modeling approach employs a strategy of "divide and conquer". More specifically, an acoustic tool is adopted to model the acoustic wave propagation inside of the combustion system. In addition, a flame model is adopted to describe the response of flame heat release rate to the flow perturbations. Flame model can be inserted into the acoustic tool to constitute the source term. By solving the obtained thermoacoustic eigenvalue problem, the modal structure as well as the corresponding modal resonance frequency and growth rate values can be estimated. Section 2.2 overviews the acoustic tools employed in the current thesis, i.e., a low-order acoustic network model and an in-house Helmholtz solver. Section 2.3 focuses on introducing the employed flame models.

2.2.1 Network model

Network models are efficient tools to predict and interpret thermoacoustic eigenmodes of the combustion system [96]. They are based on the assumption of linear and time-harmonic acoustics and its analysis is usually performed in frequency domain. In this approach, the combustion system is represented as various acoustic elements connected with each other. Each of those elements has an associated acoustic transfer matrix T, which links the characteristic wave amplitudes f and g at the inlet plane and the outlet plane of the element:

$$\begin{bmatrix} f_{out} \\ g_{out} \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \begin{bmatrix} f_{in} \\ g_{in} \end{bmatrix}, \tag{2.1}$$

with f and g defined as:

$$f = \frac{1}{2}\left(\frac{p'}{\bar{\rho}\bar{c}} + u'\right), \quad g = \frac{1}{2}\left(\frac{p'}{\bar{\rho}\bar{c}} - u'\right). \tag{2.2}$$

Here, $\bar{\rho}$ denotes density and $\bar{c}$ denotes the speed of sound. The characteristic wave f is traveling in the downstream direction, while the characteristic wave g is traveling in the upstream direction.

Figure 2.3 depicts the low-order network model representing the BRS burner test rig introduce in Section 1. This network model is adopted in *paper-ASME18*, *paper-ISC18*, *paper-ASME19*, and *paper-ASME20*. In addition to two boundary elements at the inlet and the outlet, this network model contains three duct elements (i.e., plenum, swirler tube, combustion chamber), two area jumps (one connecting plenum and swirler tube and the other one connecting swirler tube and combustion chamber), as well as one flame element.

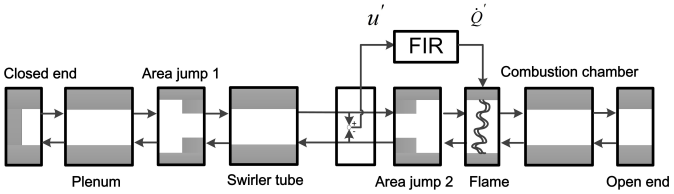


Figure 2.3: The acoustic network model for the BRS burner. Flow from left to right.

At the inlet and outlet, the boundary conditions are described by the associated reflection coefficients, which can be written as the follows:

$$R_{in} = \frac{f_{in}}{g_{in}}, \quad R_{out} = \frac{g_{out}}{f_{out}}. \tag{2.3}$$

For the duct element, the associated acoustic transfer matrix can be expressed as:

$$\begin{bmatrix} f_{out} \\ g_{out} \end{bmatrix} = \begin{bmatrix} e^{-sl/c} & 0 \\ 0 & e^{sl/c} \end{bmatrix} \begin{bmatrix} f_{in} \\ g_{in} \end{bmatrix}, \tag{2.4}$$

where eigenfrequency $s = \sigma + \boldsymbol{j}\omega$, l is the duct length, and c is the sound speed. The axial swirler is assumed to be acoustically transparent, thus the swirler tube is treated as a pure duct. For area jump element, the associated transfer matrix can be written as:

$$\begin{bmatrix} f_{out} \\ g_{out} \end{bmatrix} = \begin{bmatrix} 1+\alpha & 1-\alpha \\ 1-\alpha & 1+\alpha \end{bmatrix} \begin{bmatrix} f_{in} \\ g_{in} \end{bmatrix}, \tag{2.5}$$

with α being the area ratio. Any losses occurring in area change are not considered here. For the combined elements of flame and area jump, linearizing the quasi-1D mass, momentum, and entropy equations in the low Mach number regime would yield [9]:

$$p'_d = p'_u, \tag{2.6}$$

$$A_d u'_d = A_u u'_u + \frac{(\gamma-1)}{\gamma \bar{p}} \dot{Q}', \tag{2.7}$$

where γ denotes the ratio of specific heats, A denotes the cross-sectional area, and subscripts u and d represent upstream and downstream, respectively. Here, the heat release rate fluctuations $\dot{Q}'$ can be related to the velocity fluctuations via a flame model $\mathcal{F}$ (Eq. (2.16)). Therefore, the transfer matrix for the combined elements of flame and area jump can be expressed as:

$$\begin{bmatrix} f_{out} \\ g_{out} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(\xi+\alpha+\alpha\theta\mathcal{F}) & \frac{1}{2}(\xi-\alpha-\alpha\theta\mathcal{F}) \\ \frac{1}{2}(\xi-\alpha-\alpha\theta\mathcal{F}) & \frac{1}{2}(\xi+\alpha+\alpha\theta\mathcal{F}) \end{bmatrix} \begin{bmatrix} f_{in} \\ g_{in} \end{bmatrix}, \tag{2.8}$$

where $\alpha = A_u/A_d$, and $\xi = \rho_u c_u/\rho_d c_d$ is the ratio of specific acoustic impedances. By assembling the individual transfer matrices discussed above, a homogeneous system of equations for f and g can be obtained [96]:

$$\underbrace{\begin{bmatrix} 1 & -R_{in} & 0 & 0 \\ 0 & 0 & -R_{out} & 1 \\ T_{11} & T_{12} & -1 & 0 \\ T_{21} & T_{22} & 0 & -1 \end{bmatrix}}_{\mathscr{M}} \begin{bmatrix} f_{in} \\ g_{in} \\ f_{out} \\ g_{out} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{2.9}$$

where the subscripts in and out denote the cross-sectional planes at the combustor inlet and the outlet. By letting the determinant of matrix $\mathscr{M}$ be equal to zero, the corresponding characteristic equation can be obtained:

$$T_{22} - R_{out}T_{12} + R_{in}T_{21} - R_{in}R_{out}T_{11} = 0, \tag{2.10}$$

through which the frequencies and the growth rates of the thermoacoustic modes can be computed.

It is worth mentioning that the adopted network model (Fig. 2.3) is a rather simple yet practical model, such that a conventional Monte Carlo simulation is entirely feasible to serve as the benchmark results for UQ analysis. Nevertheless, the UQ procedures developed in the entire thesis are non-intrusive in nature, thus independent of the specific employed acoustic solver.

2.2.2 Helmholtz solver

Helmholtz solvers are extensively used for investigating the thermoacoustic instabilities [14, 15, 135]. This approach of obtaining thermoacoustic eigenmodes starts with a linearization of the Euler equation and an assumption of low Mach number flows, followed by employing the modal transformations for the acoustic pressure and heat release rate:

$$p'(\boldsymbol{x}, t) = \hat{p}(\boldsymbol{x}) \exp(-st), \tag{2.11}$$

$$\dot{q}'(\boldsymbol{x}, t) = \hat{\dot{q}}(\boldsymbol{x}) \exp(-st), \tag{2.12}$$

which yields an inhomogeneous Helmholtz equation for the acoustic pressure perturbation [9]:

$$\underbrace{\nabla(\bar{c}^2 \nabla \hat{p})}_{\mathscr{A}\hat{p}} - s^2 \hat{p} = \underbrace{s(\gamma - 1)\hat{\dot{q}}(\boldsymbol{x})}_{\mathscr{H}(s)\hat{p}}. \tag{2.13}$$

Here, $\dot{q} = \dot{Q}/V_f$, where $\dot{Q}$ denotes the global heat release rate and V_f is the volume of the flame region. The flame heat release constitutes the source term in Eq. (2.13), which can be linked to

the velocity fluctuation $\hat{u}_{ref}$ via a flame response model (Eq. (2.16)), and further linked to $\hat{p}_{ref}$ through the momentum conservation $\hat{u}_{ref} = \nabla\hat{p}_{ref}/(s\bar{\rho})$, with $\bar{\rho}$ being the mean-flow density. By assuming that the flame is acoustically compact, the gain and phase can be considered to be uniform within the flame region and zero elsewhere. This is a valid assumption as Silva *et al* [15] showed that the influence of the flame spatial distribution on the eigenvalue prediction is negligible in compact flames. Robin conditions are used at the boundaries:

$$\underbrace{\nabla\hat{p} + \beta\hat{p}}_{\mathscr{B}(s)\hat{p}} = 0, \qquad \text{where} \qquad \beta = s\frac{1}{\bar{c}}\frac{(1-R)}{(1+R)}, \tag{2.14}$$

where R is the reflection coefficient at a given boundary.

Combining Eq. (2.13) and Eq. (2.14), the nonlinear eigenvalue equation can be obtained as:

$$[\mathscr{A} - \mathscr{H}(s) + \mathscr{B}(s)]\,\hat{p} = s^2\hat{p}, \tag{2.15}$$

where $\hat{p}$ is the eigenvector describing the mode structure of a given eigenmode of the combustion system, and $\mathscr{A}$, $\mathscr{H}(s)$ and $\mathscr{B}(s)$ are the matrices containing the finite-volume discretization of the corresponding terms indicated in Eq. (2.13) and Eq. (2.14). Modal frequency and growth rate can be obtained by solving the nonlinear eigenvalue problem Eq. (2.15) via an iterative scheme based on the secant algorithm [136]. This Helmholtz solver is adopted in *paper-ISC20* and *paper-ASME20*. More numerical implementation details of the solver can be found in [40].

2.3 Flame response model

Flame response model describes the flame heat release rate response $\dot{Q}'$ to the velocity perturbations u'_{ref} at a reference position upstream of the flame [17] in frequency domain:

$$\frac{\dot{Q}'}{\bar{\dot{Q}}} = \frac{u'_{ref}}{\bar{u}_{ref}}\mathscr{F}, \tag{2.16}$$

where $\mathscr{F}$ denotes the flame response model. Here fluctuations $\dot{Q}'$ and u'_{ref} are normalized with respect to their mean value $\bar{\dot{Q}}$ and $\bar{u}_{ref}$, respectively.

In general, flame response model $\mathscr{F}$ takes in a complex frequency $s = \sigma + \boldsymbol{j}\omega$, where σ and ω denote the modal growth rate and frequency, respectively, and outputs flame gain value G and phase value ϕ. By integrating into the acoustic governing equations and acting as a source term, flame response model facilitates a calculation of thermoacoustic eigenmodes in frequency domain.

The current thesis employs three different types of flame response models: a flame impulse response (FIR) model for linear thermoacoustic instability analysis; a distributed time lag model

($\tau - \sigma$ model) with a presumed functional form proposed by Komarek *et al.* [134], also for linear thermoacoustic instability analysis; a flame describing function (FDF) model for nonlinear thermoacoustic instability analysis.

2.3.1 FIR model

Compared with the simple frequency-independent $n - \tau$ model that is widely adopted in the thermoacoustic community, the flame impulse response model represents a more sophisticated and realistic flame model that works in linear regime. In general, impulse response model completely describes the dynamic properties of a linear, time-invariant, single-input-single-output system [96]. Its model structure has the following form:

$$y[k] = G(q)u[k] + e[k], \tag{2.17}$$

where $u[k]$, $y[k]$, and $e[k]$ denote the discrete input, output, and noise term, respectively. For the $G(q)$ term, it can be written as follows:

$$G(q) = h_0 + h_1 q^{-1} + h_2 q^{-2} + ... + h_{(L-1)} q^{-(L-1)}, \tag{2.18}$$

with $q^{-n}x[k] = x[k-n]$ acting as the backward-shift operator, L denotes the "model order", i.e., the order of the polynomial, and h_k's are the impulse response model coefficients. When using the above-mentioned impulse response model to describe flame dynamics, the corresponding flame impulse response (FIR) model can be written as:

$$\mathscr{F} = \sum_{k=0}^{L-1} h_k e^{-\boldsymbol{j}(k+1)\Delta t(\omega - \boldsymbol{j}\sigma)}, \tag{2.19}$$

where h_k denotes the FIR model coefficient that needs to be identified and Δt represents the sampling interval. FIR model facilitates convenient physical interpretations of the relevant flow-flame interaction mechanisms [137]. In addition, identification procedures have already been established to efficiently derive FIR model from CFD time series data [17]: First of all, CFD simulation of the combustor under investigation is acoustically excited using a broadband signal. Then, the global heat release rate fluctuation signal $\dot{Q}'$ and the velocity fluctuation signals u'_{ref} at a reference position upstream of the flame are recorded. Finally, advanced system identification procedures are applied to those collected signals to derive the FIR model. For a more comprehensive overview of the applications of system identification to aero/thermoacoustics, readers can refer to the work of [17, 72, 138].

In practice, the identified FIR model may be uncertain due to both data quality issue and data quantity issue. Regarding the data quality issue, the employed CFD simulation, e.g., large eddy simulation (LES), is not deterministic. As a consequence, each run of the CFD simulation would generally not yield the exact same results. The stochasticity of the CFD simulations may introduce uncertainties to the identified FIR model. Also, when turbulent combustions are considered, a relatively high level of combustion noise is frequently encountered in practice. Meanwhile, the broadband forcing signal cannot have a too high amplitude (as a rule of thumb, < 10%

of the mean velocity value) to prevent nonlinear effects. As a result, the signal-to-noise ratio in a turbulent combustion environment is rather low, thus contributing uncertainties to the FIR model identification. Regarding the data quantity issue, obviously, only limited length of CFD time series can be obtained in practice due to the constrain of the available computational budget, which also induce uncertainties in FIR model identification. In summary, due to the noisy $\dot{Q}'$ and u'_{ref} signals, and limited time series length, the identified FIR model coefficients h_k's will be uncertain. Fortunately, system identification procedures automatically provide estimates of the uncertainty of h_k's, and attach a multivariate Gaussian distribution to describe the statistical property of those coefficients. Detailed identification procedure as well as the derivation of the identification uncertainty are discussed in [83].

In *paper-ASME18* and *paper-ISC18*, the adopted FIR model is identified from real LES simulations of the BRS burner (Fig. 2.1). The global heat release rate fluctuation signal $\dot{Q}'$ and velocity fluctuation signal u'_{ref} were produced by Tay-Wo-Chong *et al.* [139]. By using the algorithm outlined in [140], the nominal values as well as the covariance matrix (an indicator of uncertainty) of h_k's can be obtained. Figure 2.4 displays the identified FIR model with the associated 95% confidence interval. A total of 16 coefficients are employed for this FIR model. For both publications, the focus is on developing novel dimensionality reduction strategies to efficiently propagate uncertainties from the identified FIR model coefficients to the prediction of the modal growth rate values. More details regarding the proposed dimensionality reduction strategies are discussed in Chapter 3.

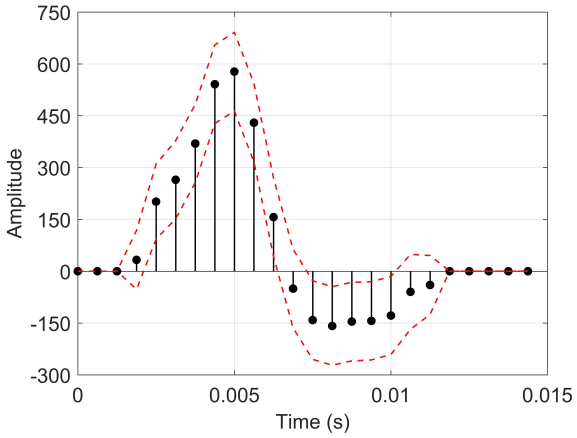


Figure 2.4: Impulse response model identified from LES data. Each discrete stem represents one coefficient h_k. Upper and lower dot lines are the 95% confidence interval.

Paper-ISC20 has also performed an UQ study on an uncertain FIR model. In that study, the employed FIR model is not identified from real LES simulations. Instead, surrogate $\dot{Q}'$ and u'_{ref} signals are generated by inserting a reference FIR model [15] and a colored combustion noise model derived from a real LES study [141] into a time-domain thermoacoustic network model [142]. Subsequently, an uncertain FIR model can be identified by applying the system identification technique on the generated artificial data. This "surrogate data" approach is beneficial as it offers a flexible way to design an FIR model with a desired uncertainty level to assess the effectiveness of the proposed UQ strategy. Figure 2.5 displays the identified FIR model

with the associated 95% confidence interval. A total of 65 coefficients are employed for this FIR model, since the focus of the *paper-ISC20* is to develop a general framework to handle high-dimensional thermoacoustic UQ problems. More details are discussed in Chapter 3.

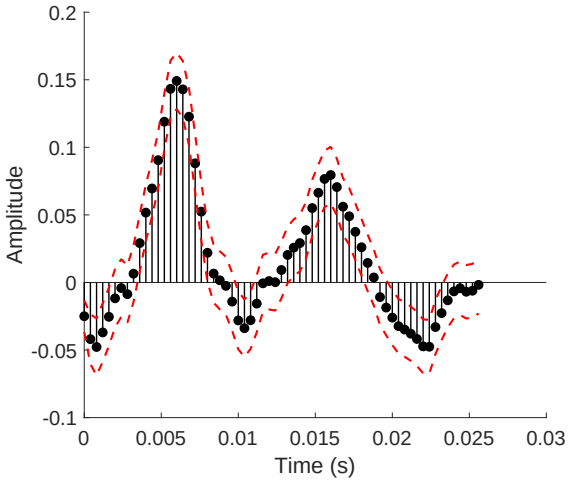


Figure 2.5: The FIR model employed in *paper-ISC20*. Each discrete stem represents one coefficient h_k; upper and lower bounds constitute the 95% confidence interval.

2.3.2 $\tau - \sigma$ model

Paper-ASME19 and *paper-ASME20* have employed a distributed time lag model proposed by Komarek *et al.* [134] and further validated by Oberleithner *et al.* [143]. This flame response model is essentially an FIR model designed specially for swirling flame dynamics. A sketch of this model is shown in Fig. 2.6. Based on the parametrization of this model, we name it $\tau - \sigma$ *model*.

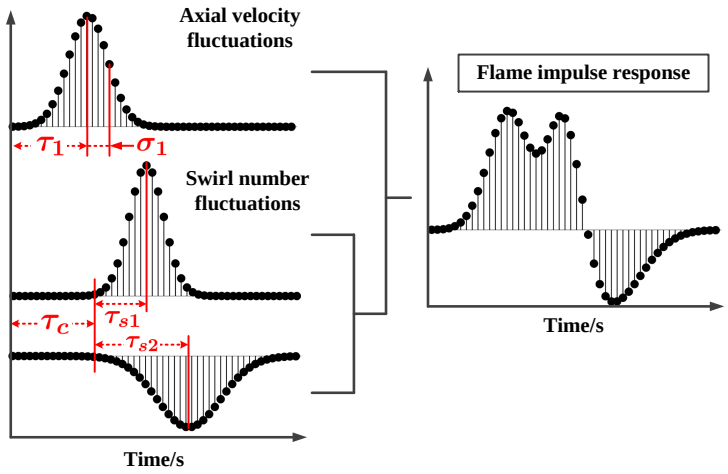


Figure 2.6: Sketch of the $\tau - \sigma$ model adopted in *paper-ASME19* and *paper-ASME20*. This flame model considers flame response to both axial velocity fluctuations and swirl number fluctuations.

The $\tau - \sigma$ model treats the shape of the flame impulse response as a decomposition of three Gaussian distributions, where one positive (+) Gaussian representing the flame response to axial velocity fluctuations, and one positive (+) alongside a negative (−) Gaussian representing the flame response to swirl number fluctuations. This flame model determines the coefficients of the overall flame impulse response by using 5 parameters: τ_c, τ_1, σ_1, τ_{s1} and τ_{s2}. Notice that this parametrization is slightly different from what is adopted in [134]: first of all, τ_c, which denotes the time for an inertial wave traveling from the swirler to the flame base [144], are now explicitly expressed and can be used as a tuning parameter for the purpose of risk mitigation [134]. Secondly, this new parametrization ensures a continuous shape of the flame impulse response under the variational model parameters. For other parameters besides τ_c, τ_1 and σ_1 represent the mean and standard deviation of the Gaussian distribution corresponding to the flame response to the axial velocity fluctuations. τ_{s1} and τ_{s2} denote the characteristic time lags for the flame response to the swirl number fluctuations. By assuming that the standard deviations of the positive/negative Gaussian distributions are one third of τ_{s1} and τ_{s2}, respectively, one can relate the FIR model coefficients h_k's to the model parameters as follows:

$$h_k = \frac{\Delta t}{\sigma_1\sqrt{2\pi}} e^{-\frac{(k\Delta t-\tau_1)^2}{2\sigma_1^2}} + \frac{\Delta t}{\sigma_2\sqrt{2\pi}} e^{-\frac{(k\Delta t-\tau_2)^2}{2\sigma_2^2}} - \frac{\Delta t}{\sigma_3\sqrt{2\pi}} e^{-\frac{(k\Delta t-\tau_3)^2}{2\sigma_3^2}}, \quad k \in [1,2,...N], \tag{2.20}$$

where N is the number of FIR coefficients, and

$$\begin{aligned} &\tau_2 = \tau_c + \tau_{s1}, \quad \tau_3 = \tau_c + \tau_{s2}, \\ &\sigma_2 = \tau_{s1}/3, \quad \sigma_3 = \tau_{s2}/3, \end{aligned} \tag{2.21}$$

In practice, τ_c, τ_1, σ_1, τ_{s1} and τ_{s2} can be calibrated from CFD time series. However, due to the exact reasons discussed above when identifying FIR model using system identification tech-

niques, data quality and data quantity issues will lead to an imperfect calibration of those parameters, thus inducing parametric uncertainties. In *paper-ASME19*, a bootstrap analysis is applied to surrogate time series of flame heat release rate and velocity fluctuations to derived the mean and covariance of each model parameter.

2.3.3 FDF model

The above-mentioned FIR model and $\tau - \sigma$ model are only valid in the linear regime. When the velocity perturbations are with high amplitudes, nonlinear effects would dominant the flame dynamic behaviors. Under this circumstance, the flame response to velocity perturbation would not only be a function of forcing frequencies, but also a function of velocity amplitudes. This observation motivates an use of a flame describing function (FDF) to describe the nonlinear flame dynamics, which contains a dataset of flame gain G and phase ϕ values at discrete forcing frequencies and amplitudes [19]:

$$\frac{\dot{Q}'}{\bar{\dot{Q}}} = \frac{u'_{ref}}{\bar{u}_{ref}} G(\omega, |\hat{u}|) \exp\left(\boldsymbol{j}\phi(\omega, |\hat{u}|)\right) \tag{2.22}$$

By integrating the FDF model with an acoustic solver, nonlinear features of thermoacoustic instabilities can be predicted, such as the limit cycle frequency and amplitude of the mode under investigation [15].

To numerically derive the FDF model, the conventional way is to force the CFD simulations using harmonic signals at a grid of forcing frequencies and amplitudes. Subsequently, Fourier analysis can be adopted to post-process the response signals of heat release rate and velocity fluctuations at each specific combination of forcing frequency and amplitude, thus obtaining the corresponding flame gain and phase value. For more recent and advanced FDF identification approaches (e.g., neural networks), readers are referred to [106, 107].

Following the conventional way of identifying FDF model, the obtained FDF dataset may be uncertain since in practice measurements have measurement error/uncertainties and could only be made at limited combinations of forcing frequencies and amplitudes. Meanwhile, to ensure the accuracy when using FDF model to predict limit cycles, it is preferred to have a fine grid of FDF values at various frequencies and amplitudes, which can be interpolated from the limited noisy measurements of the FDF data. Obviously, interpolating from uncertain FDF data will make the whole FDF dataset uncertain, which will further propagate downstream to compromise the reliability of limit cycle predictions. One of the main goals of *paper-ISC20* is to investigate this uncertainty propagation problem in nonlinear thermoacoustic analysis. More details regarding the strategies to deliver this UQ analysis is discussed in Chapter 4.

3 Uncertainty Management

In the context of model-based predictive analytics and design, *uncertainty management* can be broadly defined as a management framework to identify, quantify, and mitigate uncertainties associated with models, numerical algorithms, and predicted quantities of interest, which facilitates reliably understanding and predicting physical phenomena and making informed decisions and designs. Uncertainty management expands on the well-known concept of *uncertainty quantification*: in addition to characterizing various sources of uncertainty and propagating the characterized uncertainties through the predictive models to the outputs of interest, which are the main goals of uncertainty quantification, uncertainty management serves an extra purpose of mitigating the negative impact of the input uncertainties on the reliability of the output prediction. As a result, uncertainty management can be considered as a combination of the uncertainty quantification and the uncertainty mitigation.

This chapter serves as an introduction to the practice of uncertainty management and provides context for a better understanding of the publications included in the current thesis. The remainder of this chapter is organized as follows. To begin with, sources of uncertainty and their characterizations in computational science and engineering are reviewed. Later on, major steps in practical uncertainty management are described. Finally, challenges associated with uncertainty management are briefly discussed. Table 3.1 overviews the uncertainty analyses performed in the publications included in the present thesis.

Table 3.1: Summary of the uncertainty analyses considered in the included publications.

Publication	Uncertainty source	Uncertainty type	Task	Monte Carlo scheme
ASME18	Input data	Epistemic	Quantification	LHS
ISC18	Input data	Epistemic	Quantification	LHS
ISC20	Input data	Epistemic Aleatory	Quantification	LHS
ASME19	Input data	Epistemic Aleatory	Mitigation	Halton sequence
ASME20	Input data Model form	Epistemic Aleatory	Mitigation	Random sampling

3.1 Characterization of uncertainty sources

Thanks to the rapid development in high-performance computing and algorithmic advances, computer simulations are able to reproduce the process of complex systems in greater details,

thus gradually becoming an indispensable part in analysis and design in many engineering domains. This point can be best seen in the rapid development of *digital twin* technology [145], where computer simulations play a crucial role to create digital replicas of the real-world physical processes and systems and achieve improved productivity in manufacturing and increased reliability of the products. Computational models, which describe the behaviors of the complex natural processes and usually display itself as a set of governing equations, are at the core of the simulations. Those models are solved using numerical schemes (e.g., finite difference, finite element, etc.) in the simulations, from which valuable insights of the characteristics of the system under investigation can be obtained.

In practice, however, performing reliable simulations are non-trivial tasks: various sources of uncertainty exist in the simulation process, and their stochastic nature could easily deviate the simulation outputs from the true physical process. The common uncertainty sources encountered in computer simulations can be broadly classified into the following three categories [87]:

1. Input data uncertainty

    - Geometry uncertainty, induced by e.g., manufacturing tolerance [22], geometry simplification to facilitate simulation, etc.
    - Model parameter uncertainty, e.g., empirical parameters in turbulence models [122], parameters in flame models used in thermoacoustic instability analysis [25], etc. Those parameters are usually calibrated from noisy observations, therefore containing uncertainty.
    - Boundary conditions, e.g., combustor wall temperatures in combustion simulations [130], etc.
    - Initial conditions, e.g., the initial pressure, temperature, velocity fields may not be precisely known in unsteady CFD simulations.

2. Model form uncertainty

    Model form uncertainty is also known as *model bias* or *model discrepancy* [146]. This uncertainty source comes from the fact that the computational model is only an approximation to the true physical process, and this approximation may either result from our lack of knowledge - certain physical phenomena are not properly modelled in the simulation code, or it may result from a limited computational budget, e.g., only 2D simulation is performed instead of 3D simulation.

3. Numerical error

    Discretization error, round-off error, coding error/bugs introduce numerical uncertainties in a computer simulation [147, 148].

It is a common practice to classify various sources of uncertainty into two categories, i.e., aleatory and epistemic uncertainty [149, 150].

Aleatory uncertainty arises due to inherent variation or randomness. In principle, aleatory uncertainty cannot be reduced by additional physical or experimental knowledge. Generally, aleatory

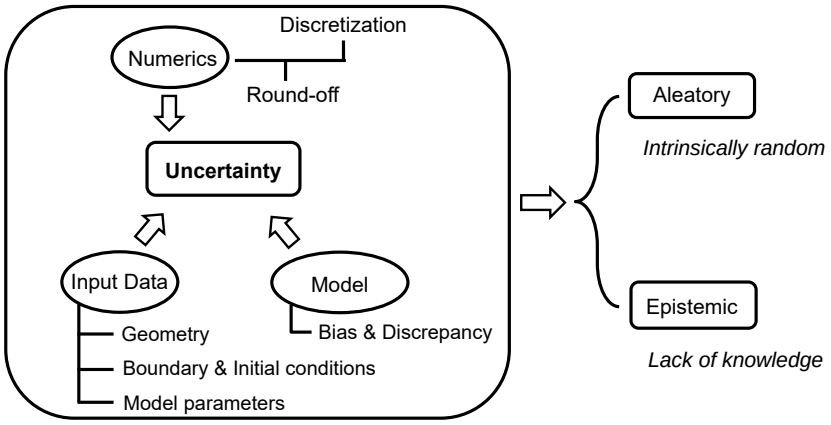


Figure 3.1: Summary of the uncertainty sources and their characterizations encountered in computational science and engineering.

uncertainty is defined in a probabilistic framework, i.e., the uncertain variables are treated as a random vector, where a joint probability density function (PDF) is assigned to describe their statistical properties. Among various sources of uncertainty mentioned above, geometry uncertainty induced by manufacturing tolerance belongs to this category.

Epistemic uncertainty arises due to a lack of knowledge, e.g., simplified model assumptions, missing physics, etc. In principle, epistemic uncertainty can be reduced if sufficient knowledge is gained via experiments, higher fidelity solvers, etc. Epistemic uncertainty is usually tackled by probability-box analysis [151], Dempster-Shafer theory [152], etc. By using Bayesian reasoning [153], it is also possible to model epistemic uncertainty using a probability distribution, since probability represents a degree of belief in Bayesian paradigm. Among various sources of uncertainty mentioned above, model parameter and form uncertainty, boundary and initial condition uncertainty belong to this category. For numerical errors, they can be converted into epistemic uncertainties as discussed in [94].

A summary of the uncertainty sources and their characterizations discussed above is given in Fig. 3.1.

3.2 Uncertainty management in practice

Various uncertainties merge with each other throughout the entire process of computer simulations. As a result, properly managing those uncertainties becomes a necessity to obtain meaningful and reliable analysis results. Under this background, the practice of uncertainty management has received increasing attention in recent years, bringing synergy between probability theory, statistics, computational mathematics, as well as disciplinary sciences [22]. In the following, the overall framework of uncertainty management is firstly introduced. Later on, two main objectives of uncertainty management, i.e., forward uncertainty quantification and risk mitigation, that the current thesis is focused on, are described in detail, along with their applications in the relevant publications.

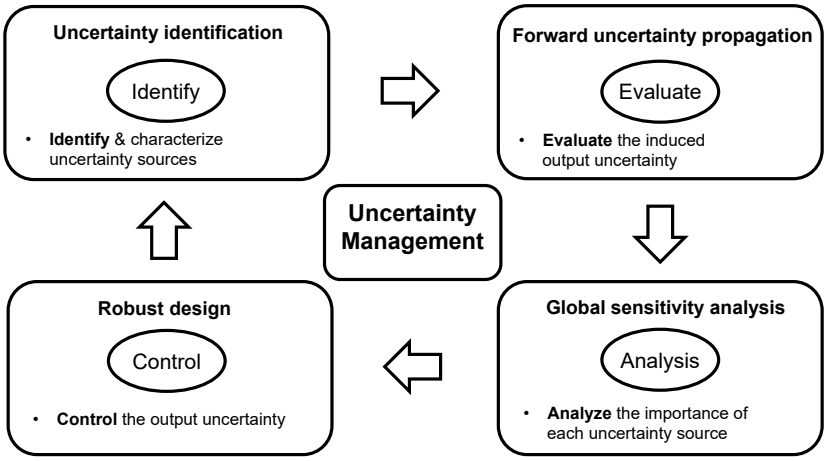


Figure 3.2: Summary of the main steps involved in uncertainty management.

3.2.1 Uncertainty management framework

The main steps of practical uncertainty management are summarized in Fig. 3.2. Detailed descriptions of each step are given in the following.

Uncertainty identification

Uncertainty identification constitutes the first step towards uncertainty management in computational science and engineering. In this step, the parameters which are uncertain are firstly determined, followed by classifications of their uncertainty nature (i.e., aleatory or epistemic). Finally, proper probabilistic models are assigned to quantify the uncertainty of the investigated parameters. Depending on the uncertainty nature, different identification treatments are proposed for aleatory and epistemic uncertainties. Specifically, for aleatory uncertainties, joint probability density functions are determined to describe their statistical distributions. In case when only marginal distributions of the parameters are known, copula techniques [154] can be employed to describe their dependency. For epistemic uncertainties, probability-box, Dempster-Shafer evidence theory, fuzzy theory, etc. [115] are all available methods to characterize the lack-of-knowledge.

Uncertainty identification can be either based on the available data or the expert judgement. In case when data is scarce, Bayesian statistics [155] may be used to aggregate expert judgement (prior information) and scarce data to achieve the goal.

Forward uncertainty propagation

This step propagates all the input data uncertainties (defined in the previous step) to the outputs of interests via running simulations. As a result, one can assess the impact of uncertain input on the reliability of output prediction. More technical details is discussed in Section 3.2.2.

Global sensitivity analysis

This step aims to rank the "importance" of the uncertain parameters in terms of driving the

output variations. Global sensitivity analysis apportions the total variance of the simulation output to different uncertain parameters and their interactions. The outcome of this sensitivity analysis is usually summarized in the form of the so-called Sobol indices [63]. Parameters with larger Sobol index values contribute more to the variation of the simulation output, while parameters with smaller Sobol index values basically play no role. Obtaining this sensitivity information is useful for achieving dimensionality reduction (e.g., by fixing the values of non-contributive parameters to their respective nominal values), as well as prioritizing resources to potentially realize maximum reduction in output uncertainty (e.g., by reducing the uncertainties of the most contributive parameters).

Robust design

This step mitigates the adverse effects of various uncertain input sources on the output prediction. The goal of this step could be either reducing the output variation (i.e., reducing the standard deviation value of the output distribution) or reducing the probability that the output value exceeds a certain threshold, namely, the failure risk. In general, those goals can be achieved by either directly reducing the uncertainty level of the input parameters, or optimizing the system to make it less sensitive to the changes in the inputs. The outcome of the robust design analysis would usually inform how much reduction of uncertainties in the input sources is required. This information could be fed into the *Uncertainty identification* step to help determine the amount of resources to allocate to better characterize the input uncertainty sources, thus forming a circular process as shown in Fig. 3.2.

The current thesis focuses on reducing the failure risk, i.e., thermoacoustic instability risk, by optimizing the thermoacoustic system under investigation to enhance its robustness with respect to various input uncertainties. More technical details is discussed in Section 3.2.3.

3.2.2 Forward uncertainty quantification

The primary goal of forward uncertainty quantification is to assess the variation of the output under the influence of various input uncertainties [156–158]. “Forward” here means that uncertainty information is flowing from inputs, through the computational model, towards the output. Depending on one’s objective, the focus could be on simply estimating the mean and variance of the output, or the probability of the output value exceeding a certain threshold (e.g., modal growth rate value larger than zero), or the entire probability density function (PDF) of the output.

The workhorse of forward uncertainty quantification is the sampling-based Monte Carlo method [159]. It involves drawing random samples from input probability distributions and employing the computational model to calculate the corresponding response of each sample. Relevant statistics of the output can then be inferred based on the obtained ensemble of results. Monte Carlo method has a very straight-forward implementation and its non-intrusive nature (i.e., no need to modify the underlying computation model) has earned its widespread usage.

However, the main drawback of the Monte Carlo method is the slow convergence [87]. For example, when using the Monte Carlo approach to estimate the mean of a random variable based on an ensemble of its realizations, the root mean squared error of the mean estimation is $\sigma/\sqrt{N}$,

where σ denotes the standard deviation of the random variable and N represents the number of samples. This observation states that the estimation error of the mean converges at the order of $N^{-1/2}$. In other words, one more decimal digit of accuracy would require a 100-fold increase in computation. Apparently, this observation poses a serious challenge for computationally expensive models. On the other hand, the dimension of the input $\boldsymbol{x}$ does not appear in $\sigma/\sqrt{N}$, meaning that Monte Carlo procedure does not suffer from the *curse of dimensionality*. This fact makes Monte Carlo a competitive method for high-dimensional uncertainty propagation tasks.

In practice, Monte Carlo simulation can be implemented in the following way:

Step 1: Generate samples $\boldsymbol{x}_i, i = 1, ..., N$ from input distributions;

Step 2: For each sample $\boldsymbol{x}_i$, run the model to obtain the corresponding output $y_i = f(\boldsymbol{x}_i)$;

Step 3: Based on the ensemble of outputs $y_i, i = 1, ..., N$, estimate the statistics of interests, e.g., mean, standard deviation, quantile, etc.;

Step 4: Visualize the results by using histogram, density plot and boxplot, etc.

In the current thesis, Monte Carlo method is used in all of the included publications to perform forward uncertainty quantification analysis. The main objective is to calculate the thermoacoustic instability risk P_f, which is defined as the probability that a thermoacoustic mode is unstable. In practice, P_f can be simply computed as the portion of samples with positive modal growth rate values.

In the following, an example is given to demonstrate the application of forward uncertainty quantification in the current thesis. In this example, a network model (Fig. 2.3) combined with a flame impulse response model (Fig. 2.4) are employed as the thermoacoustic model. Coefficients of the impulse response model are considered as the uncertain parameters. The Monte Carlo approach is adopted to propagate flame model uncertainties to the growth rate prediction of the quarter wave mode of the combustor. A visual demonstration of the workflow is shown in Fig. 3.3, where the Monte Carlo steps outlined previously are highlighted with the corresponding numbers.

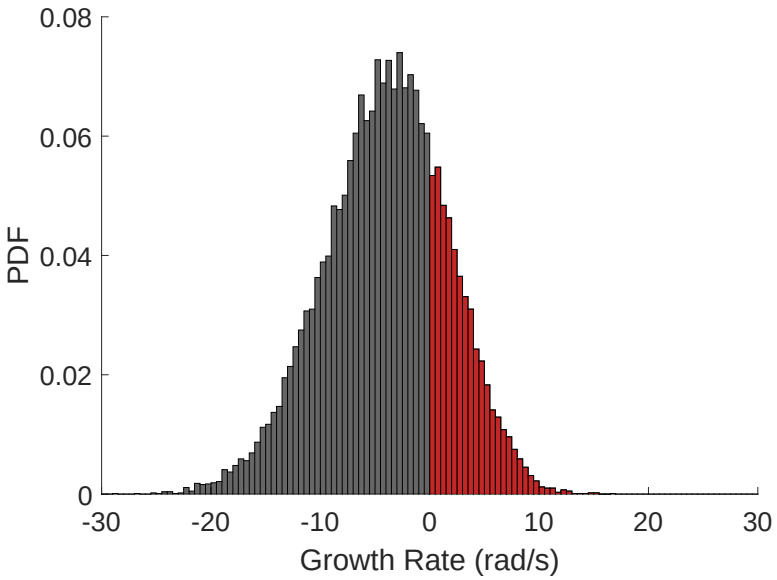


Figure 3.4: Estimated distribution of the modal growth rate induced by the uncertain flame impulse response model. The estimation is based on Monte Carlo approach with 20000 coefficient samples.

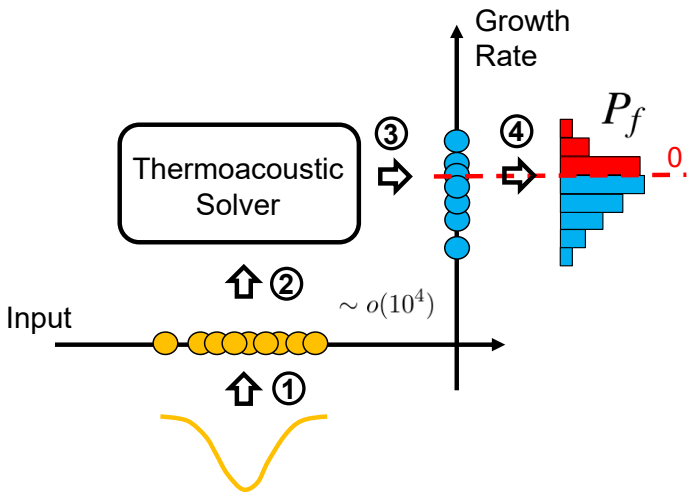


Figure 3.3: A visual display of the workflow of the Monte Carlo approach.

Results of the estimated distribution of the modal growth rate is presented in Fig. 3.4, which are based on 20000 coefficient samples and an equal number of network model calculations. Based on the portion of the samples with positive growth rate values, the instability risk P_f is calculated as 24.1%.

3.2.3 Sampling schemes

As mentioned above, vanilla Monte Carlo method is "sample-hungry", i.e., it requires a large number of samples ($\sim o(10^4)$) in order to reach a reliable statistical inference. To reduce the required number of samples while maintaining the inference accuracy, other advanced sampling schemes that having better space-filling properties [160] than the simple random sampling are preferred in practice.

There are two popular types of space-filling strategies:

1. Latin Hypercube sampling (LHS) [161]

   To generate M samples for N input variables, this method first divides the range of each input variable into M equally probable intervals. Subsequently, M sample points are placed such that each of the M intervals in each dimension are only sampled once. To make this idea concrete, consider generating 4 LHS samples for 2 input variables x_1 and x_2. Figure 3.5 provides one possible way to allocate samples, where there is only one sample in each row and column. Figure 3.5 is also called Latin square design. When extending this idea to higher dimensions, a "square" concept in 2D space becomes a "hypercube" in multi-dimensions space, thus earning its name of Latin Hypercube Sampling.

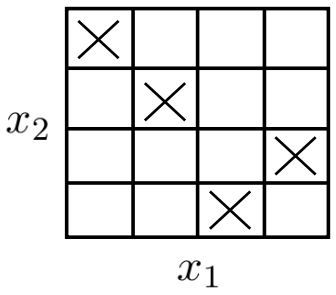


Figure 3.5: A Latin Hypercube design for two random variables.

2. Low-discrepancy sampling [162]

   Another approach to generating space-filling designs is by using quasi-random sequences. Those sequences are seemingly random points (thus "quasi") designed to fill the space as evenly as mathematically possible. Generally, the uniformity of a set of points is measured by discrepancy: sampling points forming clumps and leaving voids have high discrepancy values, while sampling points that leave no large gaps possess low discrepancy values. Hence, quasi-random sequences are also called low-discrepancy sequences. Due to their extraordinary space-filling capability, low-discrepancy sampling method can often achieve a high accuracy with far fewer samples, thus effectively decreasing the computational cost when compared with the vanilla Monte Carlo method. Two popular sampling methods that fall under this category are Halton sequences and Sobol sequences [38].

In the following example, sample designs generated by naive random sampling, Latin Hypercube sampling, Halton sequences, and Sobol sequences are compared. To allow convenient

visualization, only two random variables are considered, each of which follows independent a uniform distribution within the range of [0 1]. Here it can be seen that naive random sampling forms clusters and leave large gaps within the parameter space. In contrast, LHS and in particular low-discrepancy sequences display much better space-filling property.

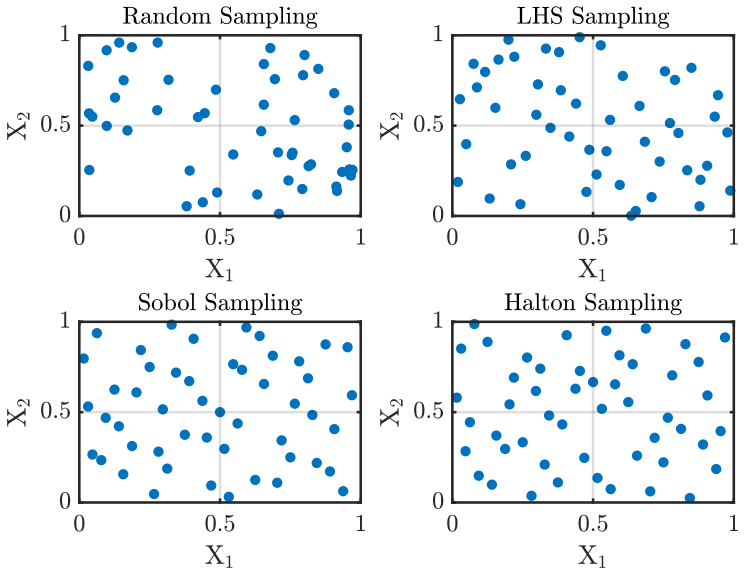


Figure 3.6: Samples generated by various sampling scheme.

In practice, as a rule of thumb, low-discrepancy sequences are the preferred choice when the input parameter space is lower than 8 dimensions [38]. For higher dimensions, low-discrepancy sequences tend to generate samples with strong correlations, with large gaps of unfilled space. Meanwhile, LHS yields more consistent results across different dimensions. As a result, LHS is more favored for high input numbers.

Another consideration in practice is sample enrichment, i.e., adding new samples to the existing designs [163]. In reality, it is rarely the case that a practitioner could foretell the required sample size for achieving a satisfactory Monte Carlo simulation. As a result, it is imperative that the practitioner could flexibly generate additional samples based on the analysis results obtained from the initial sample design. For Halton and Sobol sampling method, sample enrichment corresponds to selecting the next elements of the respective sequences; for LHS, sample enrichment can be achieved via nested Latin hypercube design [164], so that the enriched sample set forms a (pseudo-) Latin Hypercube sampling.

3.2.4 Risk mitigation

Risk mitigation aims to eliminate the system failure risk (e.g., thermoacoustic instability risk) given the input uncertainties [165, 166]. One way to achieve this goal is by implementing the principles of robust design, i.e., carefully selecting design parameters of the system such that the system output is insensitivity to various input uncertainties [167].

In the context of thermoacoustic instability analysis, risk mitigation serves as a natural second step after forward uncertainty quantification, where the instability risk is quantified. A visual comparison of those two types of analyses is shown in Fig. 3.7.

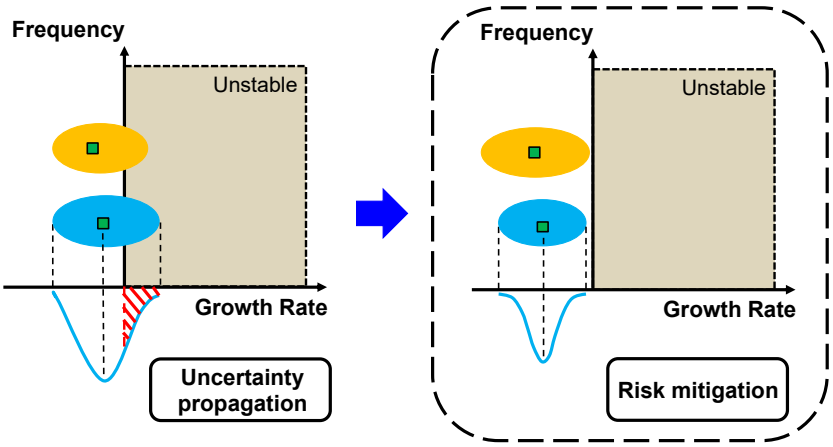


Figure 3.7: Uncertainty propagation aims to quantify the modal instability risk, while the goal of risk mitigation is to eliminate the modal instability risk.

In the current thesis, risk mitigation processes are formulated as an integration of optimization analysis and UQ analysis, as illustrated in Fig. 3.8. The skeleton of the overall risk mitigation methodology is an optimization routine, where a global optimization algorithm is adopted to search for the design parameters which optimize the objective function while conforming to constraint functions. At its core, an UQ analysis is performed at each optimization iteration to determine the instability risk given the most recent design parameters and uncertain input parameters. The calculated risk values are then used to construct objective and constraint functions, thus closing the loop for the overall optimization routine. Therefore, in the overall uncertainty management framework, accurate uncertainty quantification constitutes the cornerstone for a reliable risk mitigation.

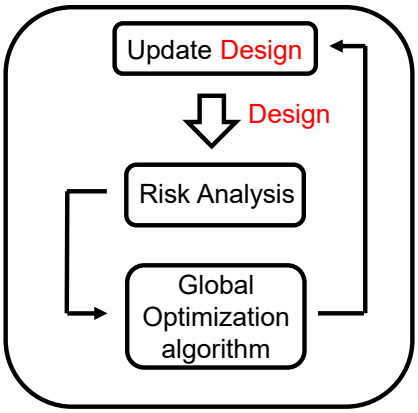


Figure 3.8: An illustration of the methodology for risk mitigation.

In the current thesis, three risk mitigation scenarios, i.e., ideal control design, realistic control design, and tolerance design, are identified for the thermoacoustic instability analysis. In the following, an example is given to consolidate the concepts of those risk mitigation tasks and illustrate how they are connected with each other. In this example, a network model (Section 2.2.1) combined with the flame impulse response model proposed by Komarek *et al.* [134] (Section 2.3.2) are employed as the thermoacoustic model. Flame model parameters $\tau_1, \sigma_1, \tau_{s1}, \tau_{s2}$ are considered as the uncertain parameters. For τ_c and R_{out} (the magnitude of the reflection coefficient at the combustor outlet), their treatment depends on the requirement of the individual design task. When τ_c and R_{out} are fixed to their nominal values, the uncertainty embedded in $\tau_1, \sigma_1, \tau_{s1}, \tau_{s2}$ would result in instability risks for both quarter wave mode of the combustor (*cavity mode*) [133] and the intrinsic thermoacoustic mode (*ITA mode*) [133]. The following design tasks aims to mitigate the instability risk for both modes by optimizing τ_c and the trade-off between the allowable uncertainty level of τ_c and R_{out}.

1. Ideal control design (Q1)

   "Using τ_c as a control factor, what is the minimum modification of τ_c to eliminate the instability risk of both cavity and ITA mode simultaneously?"

   This task treats τ_c as the design variable and it seeks for the minimum modification of τ_c from its nominal value to eliminate the instability risk of both cavity and ITA modes. In practice, τ_c can be modified by adjusting the distance between the swirler and the flame base. For this task, R_{out} is assumed to be fixed to its nominal value.

   This optimization problem can be expressed as:

$$\begin{aligned} \min_{\tau_c} f(\tau_c) = \quad & (\tau_c - \tau_c^0)^2 \\ \text{subject to}: \quad & P_f^{(I)}(\tau_c) \leq 0.1\% \\ & P_f^{(C)}(\tau_c) \leq 0.1\% \\ & \tau_c^{min} \leq \tau_c \leq \tau_c^{max} \end{aligned} \tag{3.1}$$

   In the first line, the optimization objective is given, which is to minimize the modification of τ_c. In the following two lines, $P_f^{(I)}$ and $P_f^{(C)}$ represent the instability risk of cavity mode and ITA mode, respectively. Note that the goal of eliminating instability risk is formulated as constraints instead of objectives. The benefit of this treatment is that the well-posedness of the optimization problem is ensured, i.e., only a single optimum exist. The final line states the searching range for τ_c. Overall, Eq. (3.1) constitutes a single-variable, nonlinear constrained optimization problem. Efficient global optimization algorithms, like *pattern search* algorithm [168], are available to solve Eq. (3.1).

2. Realistic control design (Q2)

   "In practice, τ_c cannot be perfectly controlled. Meanwhile, $|R_{out}|$ is also uncertain. How would these affect the decision made previously"

   Compared to the previous task, this task goes one step further towards reality in the following two aspects: first of all, this task considers the fact that τ_c can not be perfectly controlled.

Instead, it assumes a Gaussian distribution for τ_c, i.e., $\tau_c \sim \mathcal{N}(\bar{\tau_c}, (0.05\tau_c^0)^2)$. Here, $\bar{\tau_c}$ is assigned to the mean and 5% of τ_c^0 is assigned to the standard deviation. This treatment is more realistic since τ_c can be easily affected by the convective velocity in the swirler tube. Secondly, this task drops the assumption that $|R_{out}|$ value is perfectly known. This is also more realistic since the turbine operating conditions, which largely dictate the $|R_{out}|$ value, are inherently uncertain.

The corresponding mathematical formulations can be expressed as:

$$\begin{aligned} \min_{\bar{\tau_c}} f(\bar{\tau_c}) = \quad & (\bar{\tau_c} - \tau_c^0)^2 \\ \text{subject to}: \quad & P_f^{(I)}(\tau_c, |R_{out}|) \leq 0.1\% \\ & P_f^{(C)}(\tau_c, |R_{out}|) \leq 0.1\% \\ & \tau_c \sim \mathcal{N}(\bar{\tau_c}, (0.05\tau_c^0)^2) \\ & |R_{out}| \sim \mathcal{U}(R_L^*, R_U^*) \end{aligned} \tag{3.2}$$

Here the goal is to optimize the mean value $\bar{\tau_c}$ under the constraint that instability risk of both modes are sufficiently small. The optimization objective is given in the first line, which is to minimize the modification of τ_c. The following two lines states the optimization constraints that $P_f^{(I)}$ and $P_f^{(C)}$ are sufficiently small. Line 4 mimics the situation that τ_c cannot be perfectly controlled and a Gaussian distribution is used to represent its statistical fluctuation. The final line states that $|R_{out}|$ follows a uniform distribution within the known lower bound R_L^* and the known upper bound R_U^*. As a result, in addition to the uncertain $\tau_1, \sigma_1, \tau_{s1}, \tau_{s2}$, uncertainties embedded in τ_c and $|R_{out}|$ also have to taken into account when calculating the modal instability risks. Same as Q1, Eq. (3.2) constitutes a single-variable, nonlinear constrained optimization problem. Efficient global optimization algorithms like pattern search algorithm [168] could be adopted to solve Eq. (3.2).

3. Tolerance design (Q3)

"Given that the instability risk of both modes are sufficiently small, what are the maximum allowable uncertain ranges for τ_c and $|R_{out}|$?"

Q3 differs from Q1 and Q2 in that it approaches the risk mitigation from the perspective of an "inverse problem": it seeks the maximum permissible uncertain ranges of the parameters while conforming to the constraints on the modal instability risks. This is beneficial in practice since it enhances the understanding of the trade-off between the turbine and combustor operational uncertainties.

The mathematical formulations for this task can be written as follows:

$$
\begin{aligned}
\max_{\sigma_{\tau_c}} f(\sigma_{\tau_c}) = \quad & \frac{\sigma_{\tau_c}}{\tau_c^0} \\
\min_{R_L} g(R_L) = \quad & \frac{R_L}{|R_{out}|^0} \\
\text{subject to}: \quad & P_f^{(I)}(\tau_c, |R_{out}|) \leq 0.1\% \\
& P_f^{(C)}(\tau_c, |R_{out}|) \leq 0.1\% \\
& \tau_c \sim \mathcal{N}(\bar{\tau_c}^{opt}, (\sigma_{\tau_c})^2) \\
& |R_{out}| \sim \mathcal{U}(R_L, R_U^*)
\end{aligned}
\tag{3.3}
$$

This task treats the σ_{τ_c} (standard deviation of τ_c) and R_L (lower bound of $|R_{out}|$) as the design variables, as they serve as the indicators for the variational ranges of τ_c and $|R_{out}|$. The corresponding optimization objectives are given in the first two lines of Eq. (3.3), where σ_{τ_c} is maximized (equivalent to maximizing the uncertain range of τ_c) while R_L is minimized (equivalent to maximizing the uncertain range of $|R_{out}|$). The next two lines are the usual constraints that the modal instability risks are sufficiently small. The last two lines of Eq. (3.3) describe the statistical distributions of τ_c and $|R_{out}|$. For τ_c, a Gaussian distribution is considered with the mean being τ_c^{opt}, which is the optimum τ_c value obtained from Q2, and the standard deviation being σ_{τ_c}. For $|R_{out}|$, an uniform distribution is considered with lower bound R_L and known upper bound R_U^*. Unlike Q1 and Q2, Q3 constitutes a multi-objective, multi-variate, nonlinear constrained optimization problems. As a result, multi-objective global optimization algorithms [169] are required to determine the *Pareto front* [170], which allows a convenient visualization of the trade-offs between those two objectives.

3.3 Challenges and solutions

This section aims to discuss the major challenges commonly encountered in implementing the uncertainty management framework and briefly overview the dedicated strategies to alleviate those challenges that are developed in the current thesis. Detailed descriptions of those strategies are given in the next two chapters.

3.3.1 High computational cost

Due to the employment of Monte Carlo approach, practical implementation of uncertainty management framework usually involves repetitively running the simulation code, with different realizations of uncertain sources as the inputs. As already discussed in Section 3.2.1, Monte Carlo approach possesses low convergence rate, thus a large number of simulation runs are required to obtain a reliable estimation of the modal instability risk. Obviously, the resulting high computational cost would be unbearable for practical usage.

As discussed in Section 3.2.2, space-filling sampling schemes, including Latin hypercube sampling and low-discrepancy sequences, are proposed to reduce the required sample number while

maintaining the inference accuracy. Compared with the naive random sampling scheme, those space-filling sampling schemes can achieve higher convergence rate, thus leading to a more efficient uncertainty quantification.

Additionally, variance reduction sampling strategies [38] are investigated specifically for calculating the the probability of the output exceeding a certain threshold, such as calculating the modal instability risk given input uncertainties. Importance sampling [171] and subset sampling [172] are two popular choices under this category: importance sampling introduces an auxiliary sampling density that outweighs the instability region. By drawing samples from this auxiliary density instead of the input probability density, more samples can be allocated in the importance region (e.g., instability region), therefore reducing the variance of instability risk calculation and improving the UQ efficiency. For subset sampling, it calculates a series of risk values with intermediate stability margins. Those intermediate risk values can be then combined to yield the desired result via conditional probability. Since each intermediate risk calculation problem is simpler to solve, the overall convergence rate of the subset sampling scheme is much higher than direct Monte Carlo approach, thus reducing the required number of samples.

Despite the fact that advanced sampling schemes have led to improvement on the efficiency of performing UQ analysis, they still struggle to meet the needs of practical uncertainty management, especially for the risk mitigation tasks, where an UQ analysis has to be performed at each optimization iteration. Under this background, another methodology called "surrogate modeling" [173] has been heavily investigated in recent years and has showed its potential in delivering accurate and efficient uncertainty management in realistic settings.

Surrogate modeling is strongly connected to the disciplines of machine learning [174]. It aims to "learn" a relationship between the inputs and outputs of the computer simulation, and embed this relationship into a surrogate model, which is cheap to evaluate. Afterwards, Monte Carlo can be directly applied to this cheap-to-run surrogate model, thus potentially saving a considerable amount of computational budget. As reviewed in Section 1.2, surrogate modeling techniques like linear regression, polynomial chaos expansion, etc., have already been employed in accelerating the uncertainty propagation in thermoacoustic instability analysis. The current thesis has employed *Gaussian Processes* [175] as the surrogate model to achieve efficient thermoacoustic uncertainty management. Technical details of Gaussian Process and its applications in uncertainty quantification and risk mitigation are discussed in Chapter 4.

3.3.2 Curse of dimensionality

The *curse of dimensionality* refers to the phenomenon that the computational cost to fit an accurate surrogate model grows exponentially when the number of uncertain parameters increases [176]. In thermoacoustic instability analysis, sophisticated flame response models are usually high-dimensional, such as a flame impulse response (FIR) model with $15 \sim 50$ model coefficients, or a flame describing function (FDF) with tens to hundreds of discrete flame response data points at different combinations of velocity amplitude and forcing frequency. High dimensionality prohibits a direct construction of the surrogate model to map from inputs to the output of interest. Therefore, dimensionality reduction techniques [177, 178] are required to extract only the essential elements from the original high-dimensional inputs, thus facilitating afford-

able surrogate model constructions.

Broadly speaking, dimensionality reduction techniques can be data-driven or physics-informed. Data-driven approaches are "black-box" in nature, i.e., they require no insights regarding the physical system under investigation. By probing the system using carefully selected samples, data-driven approaches can potentially uncover a low-dimensional representation of the original input-output relationship, thus achieving dimensionality reduction. Physics-informed approaches, on the other hand, relies heavily on a solid understanding of the underlying physical process. Domain knowledge injected by the physics-informed approaches may significantly help distinguish the redundancy and the essentials of the input-output mapping, thus realizing dimensionality reduction.

In summary, data-driven approaches are general applicable to different problems but may require a significant amount of training samples for each application. Physics-informed approaches are much more efficient as domain knowledge replaces the role of data in discovering the low-dimensional manifold, but may only be valid for the specific type of problem under investigation, thus not general applicable.

The current thesis adopted both a data-drive approach and a physics-informed approach to circumvent the high-dimensional uncertainty quantification problems commonly encountered in thermoacoustic instability analysis. Technical details of those approaches as well as their comparisons are summarized in Chapter 5.

4 Surrogate Modeling

As discussed in Section 3.3, one possible way to alleviate the high computational cost associated with Monte Carlo simulations is the use of surrogate models. Surrogate modeling technique has its root in machine learning domain. By carefully selecting a small number of training samples and obtaining their corresponding outputs, surrogate modeling technique aims to construct a cheap-to-run yet accurate surrogate model that can capture the essence of the input-output relationship of a potentially expensive computational model. Subsequently, Monte Carlo simulations can be directly applied to the trained surrogate model, thus achieving accurate UQ with negligible cost. A visual illustration of the surrogate-based forward uncertainty quantification is shown in Fig. 4.1.

Various surrogate modeling techniques with their applications in thermoacoustic instability analysis were reviewed in Section 1.2. In this section, the focus is on the *Gaussian Process* (GP) [175] models, which is the main surrogate modeling technique adopted in the current thesis. GP stands out among other surrogate modeling techniques [179] as it is flexible, accurate, and most importantly, it provides estimates of the prediction uncertainty at testing locations [180]. This key feature facilitates a quantification of the modal instability risk calculation [74], as well as developing active learning scheme to efficiently train the GP model [181].

In the context of thermoacoustic instability analysis, Schneider *et al.* [182] proposed a real-time modeling approach based on GP to estimate the pressure pulsation amplitudes of an annular combustor. The accuracy of their strategy was further verified on the measured data for various operating conditions. Chattopadhyay *et al.* [183, 184] employed GP models to predict combustor instability at untried operational conditions and successfully benchmarked their results with

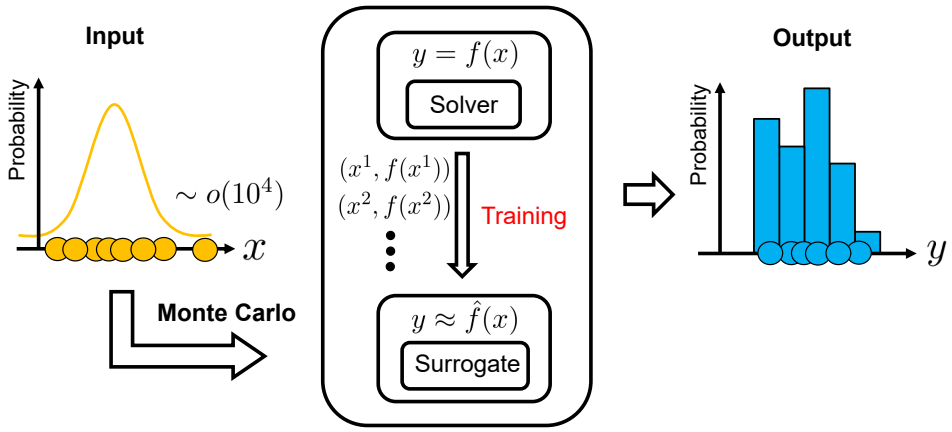


Figure 4.1: Surrogate models, which are built upon a small number of carefully collected training data, are capable of enabling highly efficient Monte Carlo simulations.

Table 4.1: Summary of the technical details of the employed GP models.

Publication	Parameter number	Training sample number	Training scheme	Reference solver
ASME19	6	102	Space-filling	Network model
ASME20	6	193	U-learning	Network model
	3	53		Helmholtz solver
ISC20	5	150	EPE learning	Helmholtz solver

the experimental data. In addition to the thermoacoustic investigations, GP has also been widely adopted in other domains. An overview of employing GP in the context of surrogate modeling can be found in [73]. Applications of GP modeling specifically in robust design can be found in [185, 186]. Recently, the basic form of GP has been further enriched to possess more advanced features. For example, gradient-enhanced GP [187–189], which can assimilate the gradient information of the training samples to further improve the GP accuracy; PCE-enhanced GP [58, 111], which combines the respective modeling strength of a GP model and a polynomial chaos expansion model, therefore leading to a more accurate and reliable surrogate model; Multi-fidelity GP [190, 191], which aggregates training data produced by solvers with different fidelities, leading to a surrogate model with the same accuracy as the high-fidelity solver, while demanding significantly reduced computational effort; Multivariate GP [192], which trains a single GP model to estimate multiple correlated outputs. By transferring the knowledge across related outputs, Multivariate GP has the potential to achieve improved prediction quality.

In the following, the fundamentals of GP modeling are briefly reviewed, followed by introducing three different GP model training schemes. Details of the GP models employed by the included publications are summarized in Table 4.1.

4.1 GP fundamentals

For supervised machine learning, the goal is to train a surrogate model based on the collected training dataset $\mathscr{D} = \{(\boldsymbol{x}^i, f(\boldsymbol{x}^i)), i = 1 : N\}$, and predict the function output $\boldsymbol{f}^* \in \mathbb{R}^{N^* \times 1}$ at the test dataset $\boldsymbol{X}^* \in \mathbb{R}^{N^* \times D}$, where N^* denotes the number of testing locations and D denotes the dimension of input vector $\boldsymbol{x}$. Common techniques to construct the surrogate model is by assuming a parametric representation for the function f and focus on identifying the underlying surrogate model parameters $\boldsymbol{\theta}$. In contrast, Gaussian Process take a different approach: it defines a Gaussian prior over the function f and uses training dataset $\mathscr{D}$ to update the assumed prior to arrive at the posterior over the function f.

4.1.1 GP formulations

In practice, the Gaussian prior is defined over the function values at an arbitrary set of points $\boldsymbol{x}_1, ..., \boldsymbol{x}_N$, such that each $f(\boldsymbol{x}_i)$ is a random variable and their ensemble $f(\boldsymbol{x}_1), ..., f(\boldsymbol{x}_N)$ follows a multivariate Gaussian distribution. $f(\boldsymbol{x}_1), ..., f(\boldsymbol{x}_N)$ constitutes a realization of the *Gaussian Process*, which is defined as an infinite collection of random variables, where any finite number

of them follows a multivariate Gaussian distribution [175]. This connection between Gaussian processes and joint Gaussian random variables significantly enhances the mathematical tractability and makes GP a very popular surrogate modeling technique.

Due to the assumption of a GP prior, the distribution of the ensemble $\boldsymbol{f} = (f(\boldsymbol{x}_1, ..., f(\boldsymbol{x}_N))$ is fully characterized by a mean function $\mu(\boldsymbol{x})$ and a covariance matrix $\boldsymbol{K}$, i.e.,

$$\boldsymbol{f} \sim \mathcal{N}(\mu(\boldsymbol{x}), \boldsymbol{K}), \tag{4.1}$$

$$\mu(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})], \tag{4.2}$$

$$K_{ij} = \mathbb{E}[(f(\boldsymbol{x}_i) - \mu(\boldsymbol{x}_i))(f(\boldsymbol{x}_j) - \mu(\boldsymbol{x}_j))], \tag{4.3}$$

where K_{ij} denotes the elements of the covariance matrix $\boldsymbol{K}$. In practice, the mean $\mu(\boldsymbol{x})$ is simply set as some unknown constant m, since the GP is flexible enough to model the mean arbitrarily well. For the covariance matrix $\boldsymbol{K}$, a kernel function $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is employed to define its elements K_{ij}, which governs the spatial correlation between inputs $\boldsymbol{x}$ at different locations.

Suppose one has observed a training dataset $\mathcal{D} = \{(\boldsymbol{x}^i, f(\boldsymbol{x}^i)), i = 1 : N\}$ and aims to predict $\boldsymbol{f}^*$ at the test dataset $\boldsymbol{X}^*$. Firstly, notice that by the definition of Gaussian process, the joint distribution of $\boldsymbol{f}$ and $\boldsymbol{f}^*$ has the following form:

$$\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{f}^* \end{pmatrix} \sim \mathcal{N}\left(m, \begin{pmatrix} \boldsymbol{K} & \boldsymbol{K}^* \\ \boldsymbol{K}^{*T} & \boldsymbol{K}^{**} \end{pmatrix}\right), \tag{4.4}$$

where $\boldsymbol{K} = \kappa(\boldsymbol{X}, \boldsymbol{X}) \in \mathbb{R}^{N \times N}$ represents the covariance matrix between the training samples, $\boldsymbol{K}^* = \kappa(\boldsymbol{X}, \boldsymbol{X}^*) \in \mathbb{R}^{N \times N^*}$ represents the covariance matrix between the training samples and the testing samples, $\boldsymbol{K}^{**} = \kappa(\boldsymbol{X}^*, \boldsymbol{X}^*) \in \mathbb{R}^{N^* \times N^*}$ represents the covariance matrix between the testing samples. By resorting to the standard formulas for conditioning Gaussian distributions [97], the distribution of $\boldsymbol{f}^*$ given the observed training dataset can be written as:

$$\begin{aligned} \boldsymbol{f}^* &\sim \mathcal{N}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \\ \boldsymbol{\mu}^* &= m + \boldsymbol{K}^{*T}\boldsymbol{K}^{-1}(\boldsymbol{y} - \mathbf{1}m) \\ \boldsymbol{\Sigma}^* &= \boldsymbol{K}^{**} - \boldsymbol{K}^{*T}\boldsymbol{K}^{-1}\boldsymbol{K}^*, \end{aligned} \tag{4.5}$$

where $\boldsymbol{\mu}^*$ can be used as the prediction mean and $\boldsymbol{\Sigma}^*$ describes the corresponding prediction uncertainty. $\mathbf{1}$ here denotes a vector of ones of dimension D.

The kernel function $\kappa(\boldsymbol{x}^i, \boldsymbol{x}^j)$ constitutes a crucial ingredient for a GP predictor, as it is a measure of the "similarity" between the function outputs at two different predicting locations. To be eligible as a kernel function, $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ must be symmetric and positive definite. For mathematical convenience, the kernel function is often assumed to be *stationary*, meaning that the magnitude

of the kernel function solely depends on the shift between the inputs:

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \kappa(\boldsymbol{x} - \boldsymbol{x}'). \tag{4.6}$$

In addition, the multi-dimensional kernel function $\kappa(\boldsymbol{x} - \boldsymbol{x}')$ is often assumed to be constructed as a tensor product of one-dimensional kernel functions:

$$\kappa(\boldsymbol{x} - \boldsymbol{x}') = \prod_{i=1}^{D} \kappa_i(x_i - x_i'). \tag{4.7}$$

This treatment has a direct physical implication as it allows tailored correlation structures for individual input parameter, which may convey very different physical meaning. Popular choices for the kernel function include cubic kernel, exponential kernel, Gaussian kernel, as well as Matérn kernel. A comprehensive discussion and comparison of various kernel functions can be found in [175].

In the current thesis, the Gaussian kernel function [193] is adopted for GP model building. This kernel function can be written as the follows:

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp\left[-\sum_{k=1}^{D} \theta_k (x_k - x_k')^2 \right], \tag{4.8}$$

where σ^2 denotes the process variance and θ_k represents the hyperparameter that controls the spatial correlation strength between the predicting locations within dimension k [173]. Figure 4.2 illustrates how the choice of θ affects the correlation. It can be seen that a low θ value signifies a low rate of decay of correlation in the corresponding dimension, thus suggesting that the function f is rather "inactive" along that dimension. On the contrary, a high θ value signifies a high rate of decay of correlation in the corresponding dimension, thus suggesting that the function f is rather "active" along that dimension. By inspecting the elements of $\boldsymbol{\theta}$, it is possible to rank the importance of the input variables, which could be helpful in terms of performing sensitivity analysis or dimensionality reduction.

4.1.2 Hyperparameter estimation

In practice, the kernel parameters σ^2 and $\boldsymbol{\theta}$ as well as the constant mean function m are estimated from the training data. Towards that end, an empirical Bayes approach [97] is adopted to find the parameters such that the likelihood of obtaining the observations is maximized. For any given parameter set $(m, \sigma^2, \boldsymbol{\theta})$, the likelihood function L can be obtained from Eq. (4.1):

$$L(\boldsymbol{y}|\boldsymbol{X}, m, \sigma^2, \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{D/2} |\boldsymbol{R}|^{1/2}} \exp\left[-\frac{1}{2\sigma^2} (\boldsymbol{y} - \mathbf{1}m)^T \boldsymbol{R}^{-1} (\boldsymbol{y} - \mathbf{1}m) \right], \tag{4.9}$$

where $R_{ij} = \exp[-\sum_{k=1}^{D} \theta_k (x_k^i - x_k^j)^2]$, which is also known as the *correlation matrix*. In practice, the logarithm of Eq. (4.9) is maximized instead to avoid round-off error:

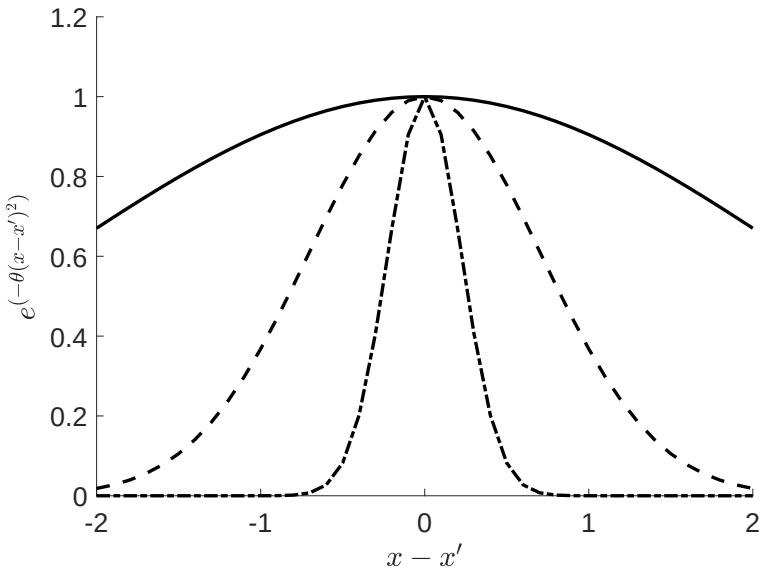


Figure 4.2: Correlation strength varies with different values of θ.

$$\log(L(\boldsymbol{y}|\boldsymbol{X}, m, \sigma^2, \boldsymbol{\theta})) = -\frac{D}{2}\ln(2\pi) - \frac{D}{2}\ln(\sigma^2) - \frac{1}{2}\ln(|\boldsymbol{R}|) - \frac{1}{2\sigma^2}(\boldsymbol{y} - \mathbf{1}m)^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathbf{1}m). \tag{4.10}$$

Analytical expressions exist for the optimum values of m and σ^2:

$$\widehat{m} = (\mathbf{1}^T \boldsymbol{R}^{-1}\mathbf{1})^{-1}\mathbf{1}^T \boldsymbol{R}^{-1}\boldsymbol{y} \tag{4.11}$$

$$\widehat{\sigma^2} = \frac{1}{D}(\boldsymbol{y} - \mathbf{1}\widehat{m})^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathbf{1}\widehat{m}), \tag{4.12}$$

which are obtained by setting the derivatives of Eq. (4.10) with respect to m and σ^2 to zeros.

For $\boldsymbol{\theta}$, its estimation requires solving an auxiliary optimization problem:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left[-\frac{D}{2}\ln(\widehat{\sigma^2}) - \frac{1}{2}\ln(|\boldsymbol{R}|) \right]. \tag{4.13}$$

The optimization problem expressed in Eq. (4.13) can be solved either via local search methods (e.g., gradient-based approaches) or global search methods (e.g., evolutionary algorithms). Local optimization methods tend to converge faster and require only fewer objective function evaluations, but may only yield local minima that lead to an inaccurate GP model. Global optimization methods involve a significantly higher computational cost, but may be more robust against local minima and able to locate the global minima. In the current thesis, a multi-start

global optimization approach [194] is adopted to estimate the optimum $\boldsymbol{\theta}$. This approach initially generates an array of feasible values of $\boldsymbol{\theta}$. Subsequently, local optimizers are run starting from each of the initial $\boldsymbol{\theta}$ value. Finally, the $\boldsymbol{\theta}$ value that yields the minimum objective function value is considered to be the optimum.

It is worth mentioning that numerical instability issues usually emerge when inverting the $\boldsymbol{R}$ to calculate the likelihood function L. This issue is especially prominent when the distances between the training samples $\boldsymbol{x}_i$ are small. To circumvent this problem, a *nugget* term ϑ is conventionally added to the main diagonal of $\boldsymbol{R}$ [195], such that

$$R_{ii} = 1 + \vartheta. \tag{4.14}$$

An example of training a GP model to approximate the function $y = x\sin(x)$ is illustrated in Fig. 4.3. For that case, a total of 5 samples are chosen at $x = 0, 2, 4, 6, 8$, respectively. The Gaussian kernel ((4.8)) is adopted for the model training. The associated hyperparameters σ^2 and θ are estimated via employing a hybrid Genetic Algorithm [193], where the final solution of the Genetic Algorithm [196] is used as a starting point for an Interior point local optimization approach [197]. In addition, a nugget $\vartheta = 10^{-10}$ is used in (4.14) to avoid numerical instability issues in matrix conversion.

Three main characteristics of the GP model can be observed in Fig. 4.3: first of all, GP model perfectly interpolates the training data, i.e., GP predictions exactly pass through the corresponding function values at the training samples with zero uncertainty. Second, the prediction uncertainty estimated by the GP model increases as it moves further away from the training data. Finally, GP model is not reliable when performing extrapolation, which is evident in the region of $8 \sim 10$. As a consequence, if possible, GP model should only be used for performing interpolation.

4.2 GP training schemes

The prediction quality of the GP model relies heavily on how the training samples are allocated in the input parameter space. In practice, it is difficult to predict how many samples are required to train a satisfactory GP model, as the sample number depends on the investigated parameter number and their ranges, as well as the complexity of the function that the GP model aims to approximate. Considering the potentially high computational cost associated with obtaining the training dataset, it is often preferable to adopt a sequential sampling scheme for GP model training, i.e., new training samples are sequentially added to the current training dataset. In the current thesis, three different sequential sampling schemes are employed to train the GP model: Section 4.2.1 discusses a training scheme that sequentially enrich the training dataset by filling the input parameter space. Section 4.2.2 introduces a training scheme that actively learns the "landscape" of the underlying function by sequentially allocating new samples to the location where maximum GP prediction error is expected. Finally, Section 4.2.3 explains another active learning scheme that is specifically designed for achieving reliable risk calculations.

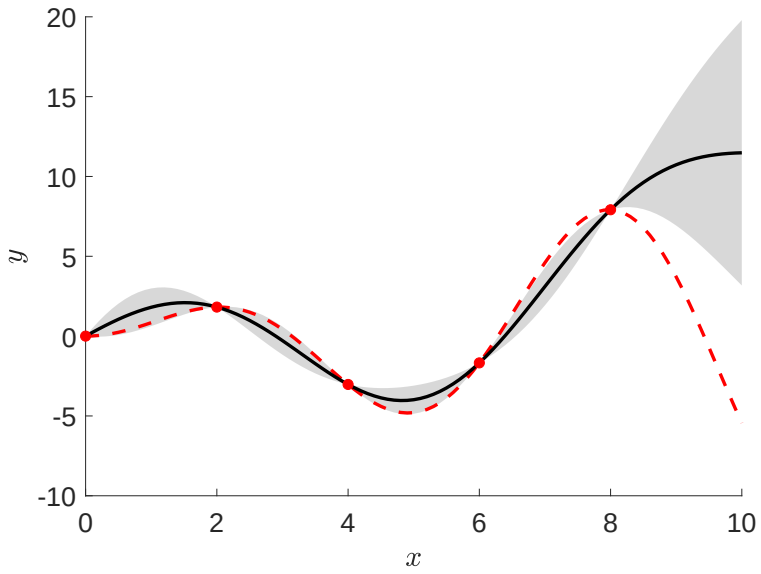


Figure 4.3: A GP model is trained to approximate the one-dimensional function $y = x\sin(x)$. The red dashed line is obtained via directly evaluating the function, the red dots are the training samples, the black line is the GP predictions, and the shaded area is 95% confidence interval estimated by the GP model.

4.2.1 Space-filling scheme

A naive way to distribute training samples is to allocate them as evenly as possible in the input parameter space, so that the obtained training samples are representative of all the regions in the parameter space. Since this space-filling scheme distributes computational effort evenly over the entire parameter space, oversampling may happen in the regions where f displays rather "regular" behavior, while undersampling may happen in the regions where f displays rather "irregular" behavior. As a result, a potentially large number of training samples may be required to train a GP model with sufficient accuracy.

In practice, low-discrepancy sequences are usually employed to ensure that training samples can be added sequentially. In *paper-ASME19*, the Halton sequence is adopted to generate training samples. As introduced in Section 3.2.2, the Halton sequence possesses excellent space-filling property and its construction naturally allows sample enrichment. To determine when to terminate the sample enrichment process, the leave-one-out cross-validation (LOOCV) error [98] of the GP model is monitored during the training process. LOOCV error estimates the generalization error of the GP model, which can be written as:

$$\text{LOOCV} = \frac{1}{D}\sum_{i=1}^{D}\left(y^i - \widehat{f}^{(-i)}\right), \tag{4.15}$$

where D is the total number of training samples, y^i denotes the known function output at $\boldsymbol{x}^i$, and $\widehat{f}^{(-i)}$ denotes the prediction at $\boldsymbol{x}^i$ via the GP model constructed upon all the training samples

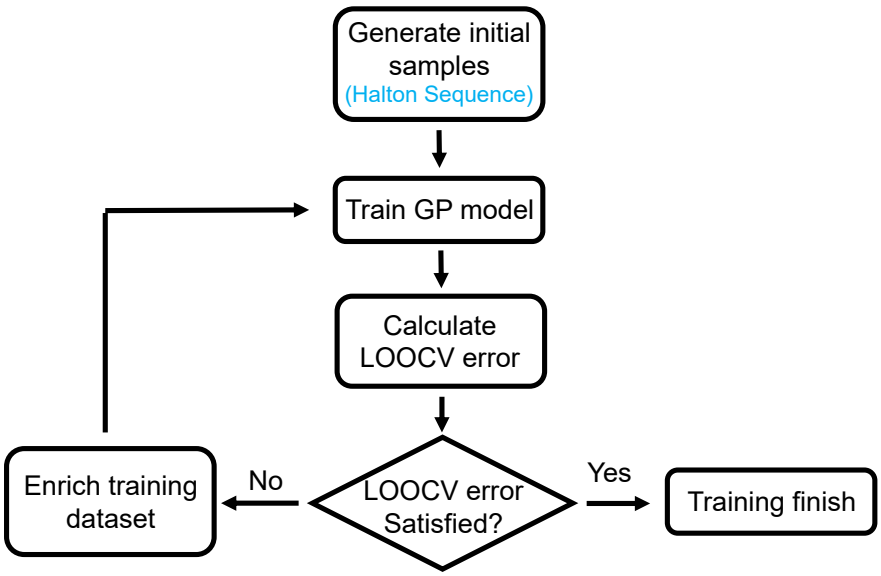


Figure 4.4: Workflow of employing space-filling scheme to sequentially train a GP model.

except $(\boldsymbol{x}^i, y^i)$. A saturated LOOCV error indicates that the GP model training has converged.

The workflow for using the space-filling scheme to train GP model is given as follows. A visual illustration is given in Fig. 4.4.

Step 1: Use Halton sequence to determine the initial sampling plan and calculate the corresponding function outputs of those initial samples. As a rule of thumb, Loeppky *et al*. [198] suggested 10 samples for each input variable are required ($= 10D$) when using GP to approximate the target function. Empirical experience shows that a third of $10D$ is a good candidate for the initial sample size [85].

Step 2: Train the GP model based on the generated training samples.

Step 3: Calculate the LOOCV error to judge the prediction accuracy of the newly trained GP model.

Step 4: If the calculated LOOCV error is above the pre-defined threshold, add one more sample to the current training dataset using Halton sequence.

Step 5: Repeat **Step 2** to **Step 4** until the LOOCV error satisfies the required prediction accuracy.

Paper-ASME19 follows the above-mentioned workflow to train GP models to approximate an acoustic network model. For 6 input variables, a total of 102 samples are used to reach the target prediction accuracy. The GP models trained in *paper-ASME19* are used to tackle various robust

thermoacoustic design tasks. More details are provided in Section 3.2.3.

4.2.2 EPE-learning scheme

Compared with the naive space-filling scheme, a more efficient way to distribute training samples is by allocating the new sample to the location where GP has the largest prediction error [199]. This is possible since GP model naturally estimates the prediction variance (Eq. (4.5)), which serves as an indicator of the expected prediction error. This sampling scheme is in direct contrast to the naive space-filling scheme in that it actively searches for the regions inside of the parameter space where the GP model has not yet reached a satisfactory prediction accuracy. As a result, more samples will be allocated to the regions where f displays complex behaviors, while fewer samples will be allocated to the flat regions of f.

The key ingredient for performing the active GP training is the learning function, which selects new samples based on pre-defined learning criterion. For the current learning scheme, the employed learning criterion is to select the next sample that maximizes *expected prediction error* (EPE) [200]. EPE is derived from the bias-variance decomposition that is well-known in the field of machine learning. The EPE criterion balances the local exploitation and the global exploration of the input parameter space: local exploitation tends to allocate samples in regions with large prediction errors, while global exploration tends to allocate samples in unexplored regions associated with large prediction variance [173]. By assimilating information from both prediction bias and prediction uncertainty, EPE criterion is able to gather very informative samples and has the potential to significantly improve the GP model training efficiency. Effectiveness of the EPE criterion has been benchmarked for functions with diverse characteristics [200]. Here, a brief derivation of the EPE expression is given in the following. Thorough discussions regarding the properties of EPE can be found in Liu *et al* [200].

Consider the prediction error in a squared form:

$$\mathscr{L}(\boldsymbol{x}) = (y(\boldsymbol{x}) - \mathscr{GP}(\boldsymbol{x}))^2, \tag{4.16}$$

where $y(\boldsymbol{x})$ is the observed function output at $\boldsymbol{x}$ and $\mathscr{GP}(\boldsymbol{x})$ is the GP prediction at the same location. The expected value of prediction error can be written as:

$$\begin{aligned}\mathbb{E}[\mathscr{L}(\boldsymbol{x})] =& (\mathbb{E}[y(\boldsymbol{x})] - \mathbb{E}[\mathscr{GP}(\boldsymbol{x})])^2 \\ &+ \mathbb{E}[(\mathscr{GP}(\boldsymbol{x}) - \mathbb{E}[\mathscr{GP}(\boldsymbol{x})])^2] \\ &+ \mathbb{E}[(y(\boldsymbol{x}) - \mathbb{E}[y(\boldsymbol{x})])^2],\end{aligned} \tag{4.17}$$

where the first term on the right-hand side is the *bias* term, which represents the average difference between the GP prediction and the actual function output; the second term is the variance of the GP prediction (Eq. (4.5)); the third term is the noise of the data. Since the current thesis only considers deterministic thermoacoustic models, i.e., each run of the thermoacoustic models with the same inputs will yield the exact same output, this term equals zero. Therefore, Eq.

(4.17) can be re-organized to obtain the desired EPE expression:

$$\mathrm{EPE}(\boldsymbol{x}) = \underbrace{(y(\boldsymbol{x}) - \mathscr{GP}(\boldsymbol{x}))^2}_{\mathrm{bias}^2} + \underbrace{\mathrm{var}(\boldsymbol{x})}_{\mathrm{variance}} . \tag{4.18}$$

In practice, to evaluate EPE value at an arbitrary location $\boldsymbol{x}$, the variance term in Eq. (4.18) can be directly obtained from Eq. (4.5). However, for the bias term, since the true function output $y(\boldsymbol{x})$ is unknown, it has to be estimated from the current training data and the trained GP model. In the current thesis, LOOCV is adopted to estimate the bias term with the following steps:

Step 1: the cross-validation errors at all training sample locations $\boldsymbol{x}^i$ are estimated as

$$e^2_{CV}(\boldsymbol{x}^i) = (y(\boldsymbol{x}^i) - \widehat{f}^{-i}(\boldsymbol{x}^i))^2, \quad i = 1, 2, ..., N, \tag{4.19}$$

Step 2: For an arbitrary location $\boldsymbol{x}$, its closest (in terms of Euclidean distance) training sample is identified and the corresponding cross-validation error is assigned to $e^2_{CV}(\boldsymbol{x})$:

$$e^2_{CV}(\boldsymbol{x}) = e^2_{CV}(\boldsymbol{x}^i), \quad \min_i |\boldsymbol{x} - \boldsymbol{x}^i| \quad i = 1, 2, ..., N \tag{4.20}$$

By replacing the bias term with the cross-validation error, the final EPE expression can be written as:

$$\mathrm{EPE}(\boldsymbol{x}) = e^2_{CV}(\boldsymbol{x}) + \mathrm{var}(\boldsymbol{x}). \tag{4.21}$$

To avoid the potentially high computational cost associated with optimizing Eq. (4.21) over continuous $\boldsymbol{x}$, in practice, a pool of pre-generated candidate samples ($\sim o(10^4)$) is usually employed, so that the EPE criterion only has to be applied to those discrete candidate samples and the sample with the maximum EPE value is chosen as the new enrichment sample. The candidate samples are usually generated via space-filling sampling scheme (e.g., Latin Hypercube sampling) to ensure that they are representative of the function over the entire parameter space. The workflow for using the EPE-learning scheme to train the GP model is given as follows. A visual illustration of the workflow is given in Fig. 4.5.

Step 1: Generate L samples $\boldsymbol{X}_U$ as a pool of candidate samples for sample enrichment.

Step 2: Generate m initial training samples using space-filling sampling scheme. Employ thermoacoustic solver to calculate their corresponding modal eigenvalue.

Step 3: Construct the initial GP model using the collected training dataset.

Step 4: Based on the current GP model, calculate the EPE values of the candidate samples $\boldsymbol{X}_U$ using Eq. (4.21), identify the sample with the largest EPE value, enrich the current training dataset with this sample, employ thermoacoustic solver to calculate its corresponding modal eigenvalue.

Step 5: Update the GP model with the enriched training dataset.

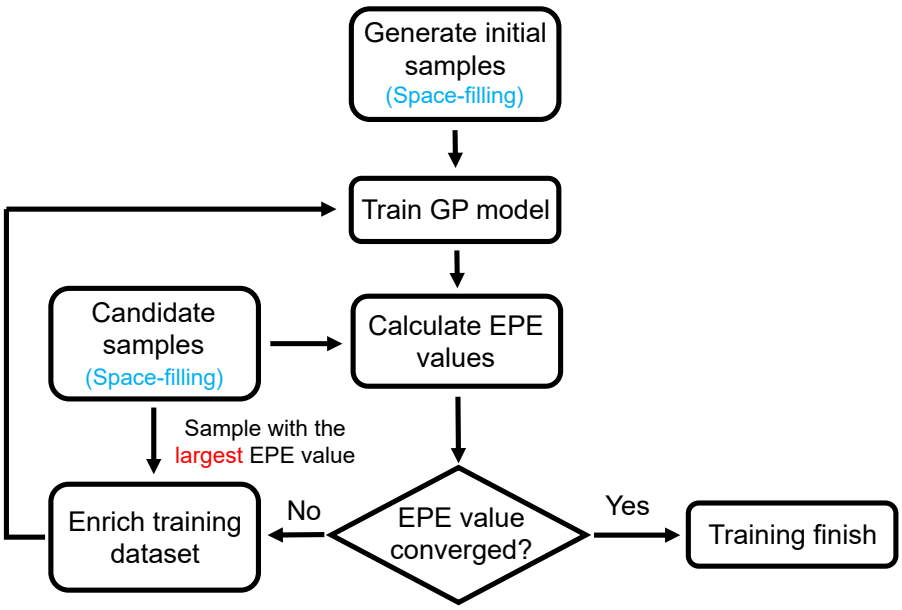


Figure 4.5: Workflow of employing EPE-learning scheme to sequentially train a GP model.

Step 6: Repeat **Step 4** - **Step 5** until the current largest EPE value drops below 5% of the largest EPE value recorded in the first iteration.

Paper-ISC20 follows the above-mentioned workflow to train GP models to approximate a 3D Helmholtz solver. For 5 input variables, a total of 150 samples are used to reach the target prediction accuracy. The GP models trained in *paper-ISC20* are used to construct *surrogate equations* to achieve dimensionality reduction for performing thermoacoustic UQ problems involving high-dimensional flame models. More details are provided in Section 5.3.

4.2.3 U-learning scheme

Computing thermoacoustic modal instability risk P_f constitutes one of the major goals in the current thesis. P_f is defined as the probability that a thermoacoustic mode is unstable, under the influence of various input uncertain parameters. In surrogate-model-based Monte Carlo simulation, P_f can be expressed as:

$$P_f = \frac{1}{n_{MC}} \sum_{i=1}^{n_{MC}} I(\widehat{f}(\boldsymbol{x}^i)), \tag{4.22}$$

where n_{MC} denotes the total number of Monte Carlo samples, $\boldsymbol{x}^i$ represents the ith Monte Carlo sample, and $\widehat{f}(\boldsymbol{x}^i)$ is the GP prediction of the modal growth rate value at $\boldsymbol{x}^i$. $I(x)$ is an

indicator function that equals zero if $x \leq 0$ and equals one if $x > 0$. In Eq. (4.22), GP prediction $\widehat{f}(\boldsymbol{x}^i)$ is uncertain according to Eq. (4.5). As a consequence, those uncertain GP predictions will propagate through Eq. (4.22), resulting in an uncertain P_f. However, note that only the sign of $\widehat{f}(\boldsymbol{x}^i)$ contribute to the P_f calculation in Eq. (4.5). Therefore, not all uncertain $\widehat{f}(\boldsymbol{x}^i)$ will lead to a change in P_f calculation, only the samples that are close to the stability margin, i.e., samples with $\widehat{f}(\boldsymbol{x}^i)$ values close to zero. This observation suggests that when constructing the GP model, instead of improving the GP prediction accuracy everywhere inside of the parameter space, which is the goal of the space-filling scheme and the EPE-learning scheme, efforts should be put into making GP model particularly accurate in the vicinity of the stability margin.

Towards that end, the U-learning scheme [201] is adopted to sequentially allocate training samples in the vicinity of the stability margin. The workflow of the U-learning scheme is the same as the EPE learning scheme, except a U learning function is used instead of the EPE learning function. The U-learning function is expressed as follows:

$$U(\boldsymbol{x}) = \frac{|\mu(\boldsymbol{x})|}{\sigma(\boldsymbol{x})}, \tag{4.23}$$

where $\mu(\boldsymbol{x})$ denotes the mean of the GP prediction at $\boldsymbol{x}$ and $\sigma(\boldsymbol{x})$ denotes the corresponding standard deviation of the prediction.

$U(\boldsymbol{x})$ is closely linked to the concept of probability of misclassification $P_m(\boldsymbol{x})$. For a Monte Carlo sample $\boldsymbol{x}^i$, misclassification happens when the true function output at $\boldsymbol{x}^i$ is negative whereas the corresponding GP prediction is positive, or vice versa. For the first situation, the probability of misclassification can be written as:

$$P_m(\boldsymbol{x}) = \Phi\left[-\frac{\mu(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\right], \tag{4.24}$$

and for the second situation:

$$P_m(\boldsymbol{x}) = \Phi\left[\frac{\mu(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\right]. \tag{4.25}$$

Combining Eq. (4.24)-(4.25), the probability of misclassification can be derived as:

$$P_m(\boldsymbol{x}) = \Phi\left[-\frac{|\mu(\boldsymbol{x})|}{\sigma(\boldsymbol{x})}\right] = \Phi[-U(\boldsymbol{x})]. \tag{4.26}$$

At each learning iteration, the candidate sample with the minimum U value is selected to enrich the current training dataset, as a smaller $U(\boldsymbol{x})$ value indicates a higher probability of misclassification according to Eq. (4.26). In addition, as can be seen in Eq. (4.23), samples that are either close to the stability margin or with large prediction uncertainties tend to be selected by the employed learning scheme. Therefore, Eq. (4.23) also balances between local exploitation and global exploration, as demonstrated by the EPE criterion in Eq. (4.21). A visual illustration of the U-learning scheme is demonstrated in Fig. 4.6.

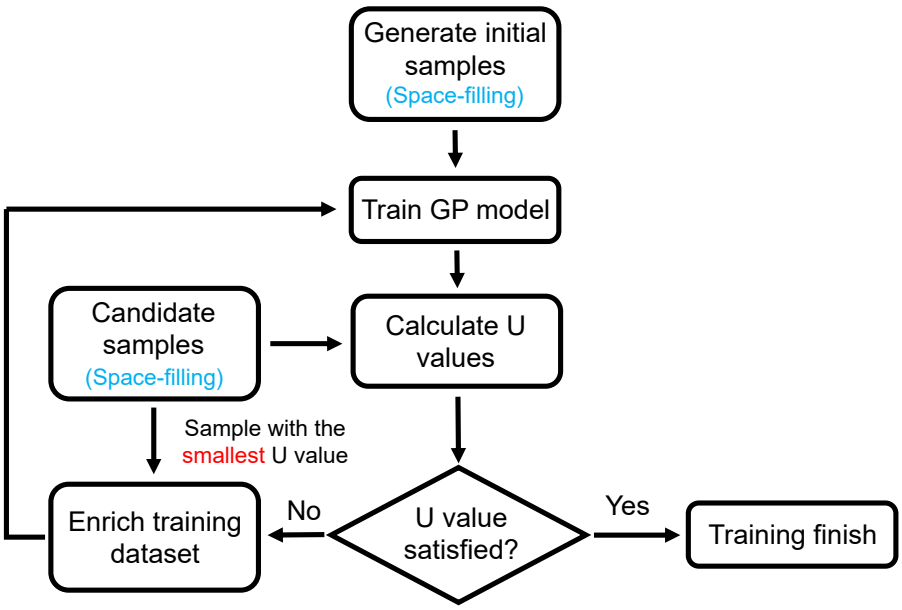


Figure 4.6: Workflow of employing the U-learning scheme to sequentially train a GP model.

Paper-ASME20 employs the U-learning scheme to train GP models to approximate an acoustic network model and a 3D Helmholtz solver, respectively. For the network model case, 6 input variables are considered and a total of 193 samples are used to reach the target prediction accuracy. For the Helmholtz solver case, 3 input variables are considered and a total of 53 samples are used to reach the target prediction accuracy. The capability of the U-learning scheme is further compared with the naive space-filling scheme in the network model case study. There, two GP models are trained respectively via those two learning schemes, and subsequently employed to derive the PDF of the P_f of the cavity mode. Exact same number of training samples (193) are used by both learning schemes. The predicted PDFs are compared in Fig. 4.7. It can be seen that the U-learning scheme is able to make better use of the computational resources and significantly improves the accuracy and robustness of the P_f calculation, with prediction mean much closer to the reference risk value and prediction variance much smaller than the naive space-filling training scheme.

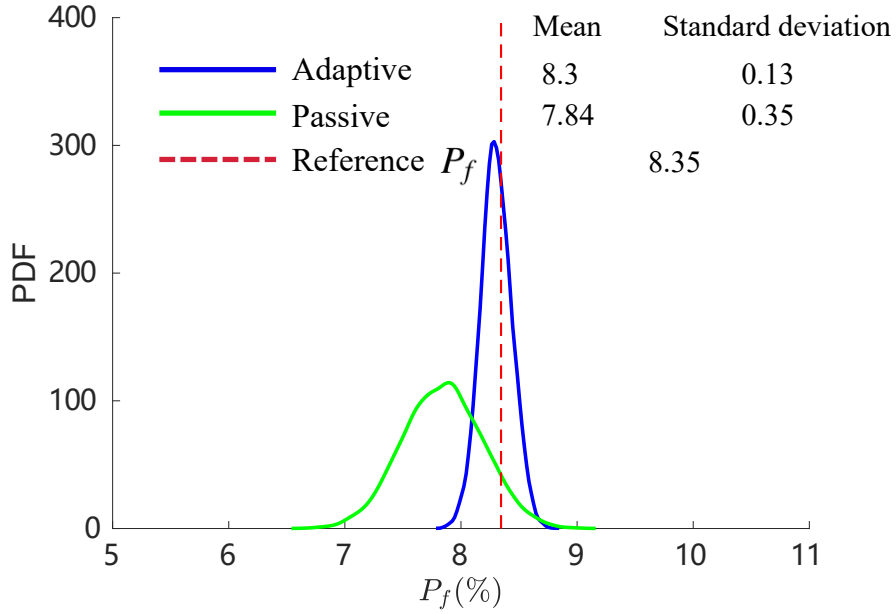


Figure 4.7: Given the same number of training samples, the U-learning scheme produces a more accurate and less uncertain P_f calculation.

5 Dimensionality Reduction

As higher fidelity of thermoacoustic instability analysis is actively pursued in the community, more sophisticated and realistic flame response models are being employed, featuring a high number of model parameters. Performing uncertainty management for those flame models is challenging since the usual surrogate modeling techniques cannot be directly adopted without modification: due to the curse of dimensionality [87], the computational cost would be prohibitive to directly construct a surrogate model to map from the flame model parameters to the output of interests (e.g, modal frequency and growth rate). Under this circumstance, it is necessary to employ dimensionality reduction techniques beforehand to extract the essentials of the investigated input-output relationship, reduce the problem dimensionality, and pave the way for subsequent surrogate-based UQ analysis.

In the current thesis, three different strategies have been developed to deliver efficient high-dimensional UQ analysis for thermoacoustic instability analysis. This chapter is organized as follows. Section 5.1 overviews the Active Subspace approach employed in *paper-ASME18*, Section 5.2 introduces an analytical dimensionality reduction strategy developed in *paper-ISC18*. The connections between the Active Subspace approach and the current analytical approach are highlighted. Finally, Section 5.3 discusses a surrogate-based iterative scheme, which handles high-dimensional thermoacoustic UQ problems in a novel way. The first two sections focus on propagating uncertainties from the flame impulse response (FIR) model to the modal growth rate predictions, while the final one considers high-dimensional flame models in general. For all sections, theory is presented first, followed by illustrating the results obtained via case studies. Table 5.1 provides an overview of the dimensionality reduction techniques adopted in the current thesis.

Table 5.1: Summary of the employed dimensionality reduction techniques.

Publication	Dimensionality reduction technique	Flame model	Input dimension	Sample number	Reference solver
ASME18	Active Subspace	FIR	16	400	Network model
ISC18	Analytical	FIR	16	6	Network model
ISC20	Surrogate-based	FIR FDF	68 523	150	Helmholtz solver

5.1 Active Subspace

5.1.1 Theory

In general, dimensionality reduction approaches can be subset-based or subspace-based [202]. Subset-based approaches seek and preserve an important subset of the input parameters based on sensitivity analysis. In contrast, subspace-based approaches seek a low dimensional space, which captures the essence of the input-output relations. By projecting the full dimension system into the low dimensional space, dimensionality reduction is achieved while all the original input parameters are preserved. The *Active Subspace* approach belongs to this category [64] and has been successfully applied in various engineering problems [203–206].

In theory, the Active Subspace approach identifies directions in the high-dimensional input parameter space, such that the output of interest varies the most along these directions. In practice, those directions can be detected by evaluating the gradients of the function at a set of input samples. To exploit those identified directions, firstly, the inputs are projected onto these directions to obtain the so-called "active variables", which constitute a low-dimensional portray of the original high-dimensional system. Subsequently, surrogate models can be directly built upon those active variables to approximate the system under investigation. More details on its mathematical derivations can be found in Constantine *et al.* [207].

Figure 5.1 provides an example of how Active Subspace approach is used to identify a low-dimensional structure of the target function $y = e^{0.1x_1+0.2x_2}$. In this example, the investigated function has two input parameters, therefore the input dimensionality is two. By inspecting the contour of the function, it can be observed that along direction $\overrightarrow{l}$ the output varies the most, while in the orthogonal direction of $\overrightarrow{l}$ the output stays fixed. Active Subspace approach identifies this direction $\overrightarrow{l}$ in the input parameter space, and uses this direction to construct an active variable AV, which appears as a linear combination of the inputs. This active variable AV captures the essences of the investigated input-output relation, since when the value of the active variable is fixed, meaning that x_1 and x_2 can only vary in the orthogonal direction of $\overrightarrow{l}$, output y will also stay the same. Subsequently, one can train a surrogate model only based on the identified active variable. Therefore, in this case, the problem dimensionality shrinks from 2 to 1.

In the following, details on practical implementation of the Active Subspace approach are given. Here, $q = f(\mathbf{x})$ denotes the investigated input-output relation, where $q \in \mathbb{R}$ represents the scalar quantity of interest, and $\mathbf{x} \in \mathbb{R}^m$ represents a $m-$dimensional input vector.

Step 1: Normalize the input parameters to be centered at zero with equal ranges.

Step 2: Draw M samples of $\mathbf{x}$ according to the distribution of $\mathbf{x}$.

Step 3: For each sample $\mathbf{x}^{(i)}, i = 1,2,...,M$, compute the gradient $\bigtriangledown f(\mathbf{x}^{(i)})$. For a simulation code equipped with an adjoint solver, computing gradients is straight-forward. When an adjoint solver is not available, approximated gradients can be obtained via local linear regression.

Step 4: Compute the covariance matrix of the gradient vector and perform eigenvalue decom-

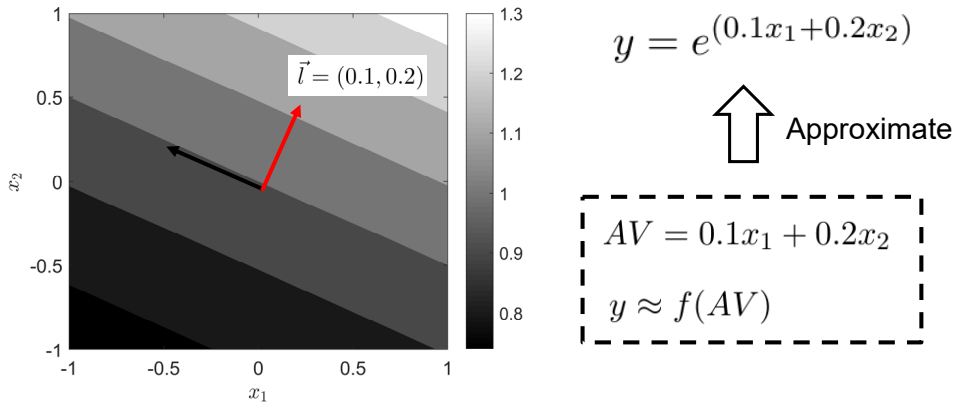


Figure 5.1: Active Subspace approach is employed to identify a low-dimensional representation of the investigated toy function. Contour of the function is shown on the left.

position:

$$C = \frac{1}{M}\sum_{i=1}^{M} \nabla f(\mathbf{x}^{(i)}) \nabla f^T(\mathbf{x}^{(i)}) = \mathbf{W}\Lambda\mathbf{W}^T, \tag{5.1}$$

where $\mathbf{W}$ is the matrix of eigenvector and $\Lambda = \text{diag}(\lambda_1, ..., \lambda_m)$ is the diagonal matrix of eigenvalues ordered in decreasing order.

Step 5: Plot eigenvalues $\lambda_1, ..., \lambda_m$ and look for "gaps". A prominent gap in the eigenvalues indicates a separation between active and inactive subspaces. The dimension n of the active subspaces is chosen, which equals the number of eigenvalues before the most prominent gap happens.

Step 6: A corresponding partition of the eigenvector matrix can be made $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2]$, where $\mathbf{W}_1$ only contains n eigenvectors corresponding to the first n eigenvalues. The active variables can then be defined as $\mathbf{H} = \mathbf{W}_1\mathbf{x}$.

Step 7: Subsequently, a surrogate model can be built $\hat{q} \approx \hat{f}(\mathbf{H})$, which takes in active variable $\mathbf{H}$ and approximates $f : \mathbb{R}^m \rightarrow \mathbb{R}$.

Step 8: To quickly evaluate the response q^* of any given $\mathbf{x}^*$, firstly, its corresponding active variable values are calculated $\mathbf{H}^* = \mathbf{W}_1\mathbf{x}^*$. Subsequently, the trained surrogate model is employed calculate the approximated response $\hat{q}^* \approx \hat{f}(\mathbf{H}^*)$.

5.1.2 Case study

In *paper-ASME18*, the Active Subspace approach is adopted to discover the low-dimensional representation of the relationship between FIR model coefficients and modal growth rate values. The network model and the FIR model introduced in Chapter 2 are considered in the current case study. Here, the focus is on quantifying the variation of the growth rate value of the combustor quarter wave mode induced by the uncertain FIR model. Under the nominal values of the FIR

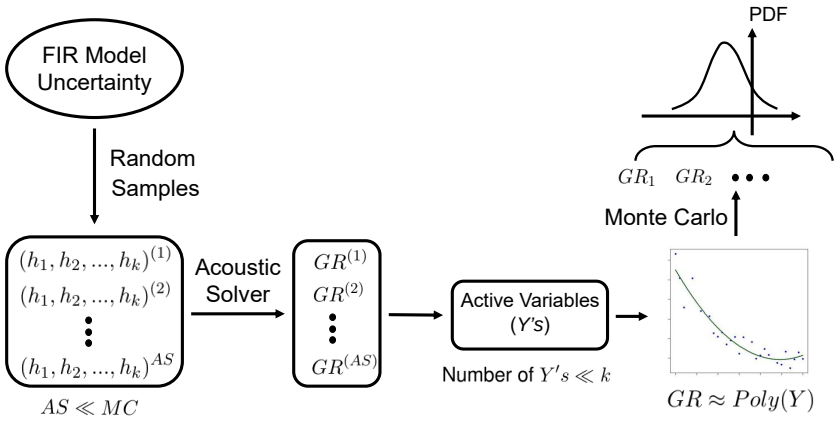


Figure 5.2: Application of the Active Subspace approach to derive the growth rate uncertainty induced by the uncertain FIR model. In the figure, *AS* denotes the number of training samples required by the Active Subspace approach, *MC* denotes the number of Monte Carlo samples, *GR* is short for growth rate.

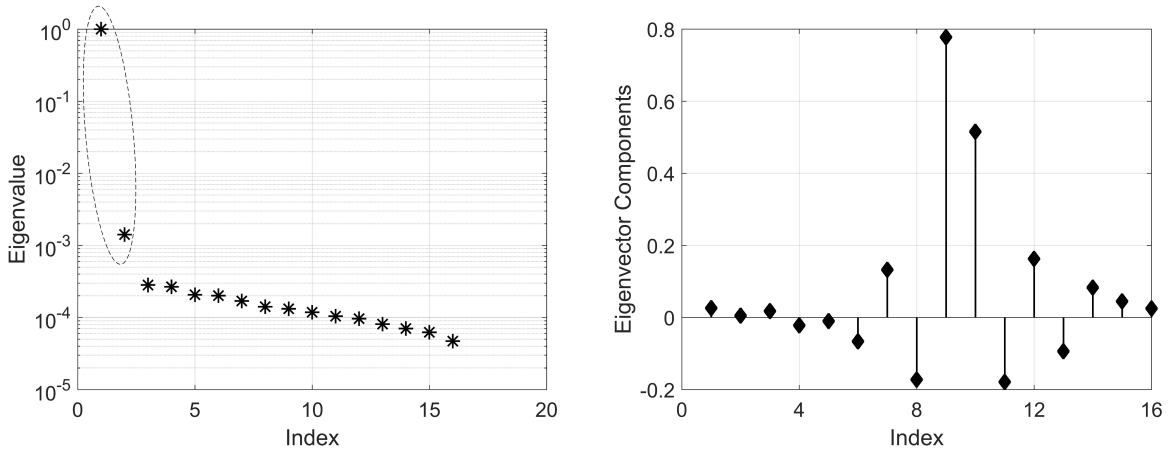
model coefficients, this mode has a frequency of 434.2 Hz and a growth rate value of -4 rad/s. Since this mode is only marginally stable, it may become unstable when taking into account the FIR model uncertainty.

This FIR model has 16 coefficients, making it eligible to be considered as a high-dimensional UQ problem. Directly building a surrogate model to link those FIR model coefficients to the modal growth rate value would induce prohibitive computational cost. Therefore, it is beneficial to perform dimensionality reduction beforehand to alleviate the difficulty encountered by the surrogate model training process. Towards that end, Active Subspace approach is employed to find linear combinations of FIR model coefficients, such that along those linear combinations the growth rate values vary the most.

Paper-ASME18 has adopted the procedures outlined in the previous section to perform Active Subspace identification. Since direct calculations of the gradient of the growth rate value against the FIR model coefficients is unavailable at the time, local linear regressions are used instead to approximate the required gradients. An visual illustration of the overall workflow of the Active-subspace-based Monte Carlo is demonstrated in Fig. 5.2.

Figure 5.3a displays the calculated eigenvalues in descending order. It is clear to see that there exist a prominent gap between the first and second largest eigenvalues. This result suggests that only the first eigenvector (shown in Fig. 5.3b) needs to be retained to construct a single active variable. As a result, significant dimensionality reduction is achieved which shrinks the current UQ problem from 16-dimensional to just one-dimensional.

To visualize the relation between the growth rate value and the corresponding active variable value for each sample, sufficient summary plot is depicted in Fig. 5.4. It can be seen clearly that a strong univariate relationship is presented, which further confirms that the identified one-dimensional approximation is sufficiently accurate to represent the original input-output mapping.



(a) Eigenvalues (b) Eigenvector

Figure 5.3: (a) The calculated eigenvalues in descending order. A prominent gap exists between the first and second eigenvalues, which signals a one-dimensional active subspace exist. (b) The components of the first eigenvector, which will serve as the linear combination coefficients to form a single active variable.

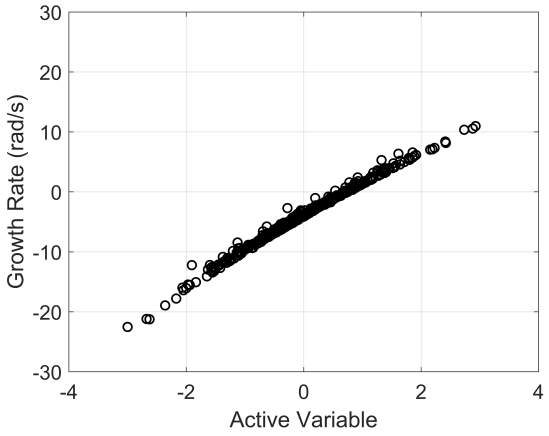


Figure 5.4: Sufficient summary plot to show the relationship between the modal growth rate value and active variable value for each sample.

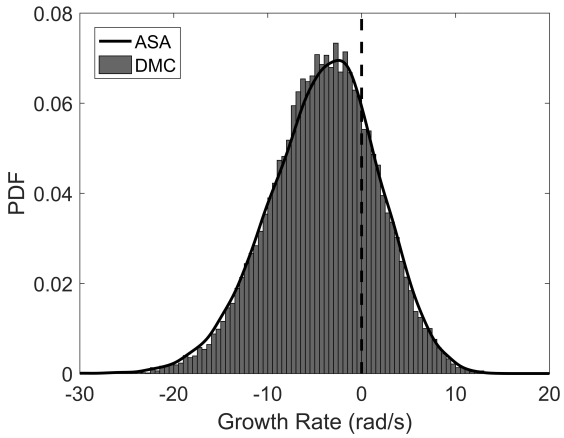


Figure 5.5: Comparison of the calculated probability density function of the modal growth rate values. ASA and DMC stand for Active Subspace approach and direct Monte Carlo, respectively.

Subsequently, surrogate models can be built to fit the trend depicted in Fig. 5.4. Since there exist only one active variable and the correlation between the active variable and the modal growth rate value is very strong, many surrogate modeling techniques are applicable to yield sufficiently accurate results. In *paper-ASME18*, a simple quadratic regression model is fitted to describe the trend shown in the Fig. 5.4.

Once the surrogate model is trained, the following Monte Carlo procedure can be directly applied to the constructed surrogate model to achieve accelerated UQ analysis. The probability density of the modal growth rate value predicted by the surrogate model matches perfectly with the results obtained via applying Monte Carlo directly to the network model (as shown in Fig. 5.5), thus indicating the accuracy of the Active Subspace approach. In terms of the computational cost, Active Subspace approach consumes a total of 400 samples (i.e., 400 times of network model calculations) to uncover the one-dimensional structure. Subsequent uncertainty propagation involves no extra cost. On the other hand, direct Monte Carlo approach based on network model takes 20000 network model calculations to obtain fully converged statistical indices. Therefore, a 50-fold increase in computational efficiency is achieved by the Active Subspace approach in the present case.

5.2 Analytical approach

The Active Subspace strategy developed in the previous section has two major limitations that are undesirable in practical applications. First of all, the identified active variable (Fig. 5.3b) may only be valid for the specific FIR model under investigation. As a consequence, each time when a new uncertain FIR model enters into analysis, another ~ 400 of thermoacoustic solver calculations would be required to re-identify the associated active variable(s). Secondly, when the FIR model under investigation has a larger number of coefficients, the computational cost required by Active Subspace to identify the low-dimensional structure would also increase sig-

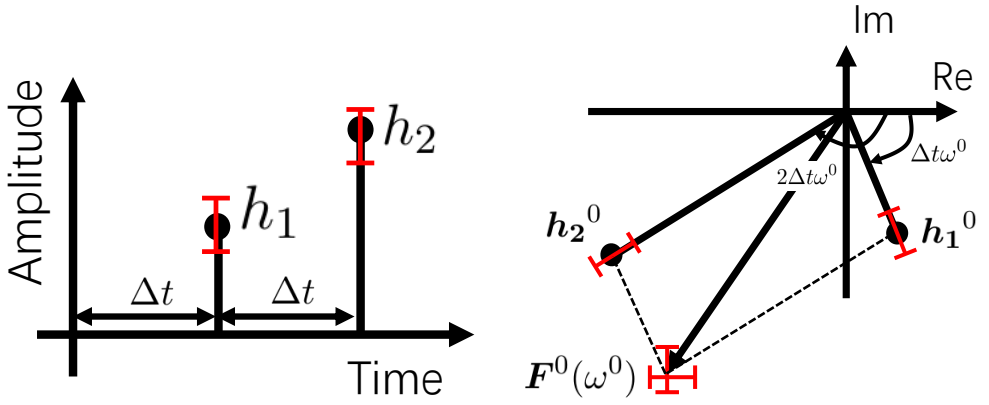


Figure 5.6: The sketch of (a) a two-coefficient FIR model and (b) its corresponding phasor plot of FRF.

nificantly. Nevertheless, the observation that a one-dimensional subspace exists that can compactly summarize the relationship between the high-dimensional FIR model coefficients and the modal growth rate is very encouraging and may serve as the inspiration for developing an analytical dimensionality reduction UQ strategy. Under this background, *paper-ISC18* approaches the problem from an analytical point of view and aims to uncover the physical interpretation of the active variable. The goal of the current work is to derive an analytical dimensionality reduction scheme to realize accelerated uncertainty propagation from the FIR model coefficients to the modal growth rate predictions.

5.2.1 Theory

The foundation of the current analytical strategy lies on the phasor plot of the flame frequency response function (FRF). An illustration of how to translate a simple two-coefficient FIR model into the corresponding phasor plot of FRF is given in Fig. 5.6. Here, $F(\omega^0)$ is written as follows:

$$F(\omega^0) = h_1 e^{-\boldsymbol{i}\Delta t\omega^0} + h_2 e^{-\boldsymbol{i}(2\Delta t)\omega^0}, \tag{5.2}$$

where ω^0 is the nominal modal frequency calculated by feeding the nominal h_1 and h_2 values into the acoustic solver.

When the FIR model coefficients are uncertain, any variations in the coefficient will change the length of the respective phasor, which subsequently will change the length and the direction of the $F(\omega^0)$ phasor (as shown in Fig. 5.6b). For samples of FIR model coefficients that produce the same growth rate values, the head locations of their corresponding $F(\omega^0)$ phasor forms the iso-growth lines in the phasor plot of FRF.

To derive the distribution pattern of the iso-growth lines in the phasor plot of FRF, *paper-ISC18* starts from the characteristic equation of the thermoacoustic feedback loop (illustrated in Fig.

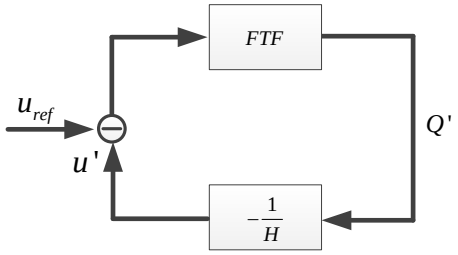


Figure 5.7: The thermoacoustic closed-loop network.

5.7):

$$\mathscr{F}(\sigma + \boldsymbol{j}\omega) = H(\sigma + \boldsymbol{j}\omega) \quad \omega, \sigma \in \mathbb{R}, \tag{5.3}$$

where H denotes the acoustic transfer function and $\mathscr{F}$ denotes the flame transfer function, which is represented by a FIR model with arbitrary number of coefficients.

By performing a Taylor expansion analysis on Eq. (5.3), it can be proven that to first order, the iso-growth lines in the phasor plot of FRF are parallel straight lines, with the line direction $\boldsymbol{l} = (G_r^0, G_i^0)$ given as:

$$G_r^0 = \partial H_r / \partial \omega|^0 + \sum_{k=0}^{L-1} h_k^0 (k+1) \Delta t \sin[(k+1)\Delta t \omega^0] \tag{5.4}$$

$$G_i^0 = \partial H_i / \partial \omega|^0 + \sum_{k=0}^{L-1} h_k^0 (k+1) \Delta t \cos[(k+1)\Delta t \omega^0], \tag{5.5}$$

where H_r and H_i denote the real and imaginary part of the acoustic transfer function, respectively, and L denotes the number of FIR model coefficients. Here, it can be seen that the components of this direction are determined by (1) the employed FIR model (h_k^0), (2) combustor geometry or boundary conditions (H_r and H_i), as well as (3) the specific thermoacoustic mode under investigation (ω^0).

The derived distribution pattern of the iso-growth lines in the phasor plot of FRF is the key to accelerate the targeted UQ analysis. As illustrated in Fig. 5.9, the variation of each coefficient can modify the modal growth rate value individually. However, what ultimately governs the modal growth rate change is the sum of the projection of the phasor $(h_k - h_k^0) e^{-\boldsymbol{i}(k+1)\Delta t \omega^0}$ on the normal direction of the iso-growth lines. This projection sum, denoted as Y in Fig. 5.9, can be formally written as:

$$Y = \sum_{k=0}^{L-1} (h_k - h_k^0) e^{-\boldsymbol{i}(k+1)\Delta t \omega^0} \cdot \frac{\boldsymbol{n}}{|\boldsymbol{n}|}, \tag{5.6}$$

where $\boldsymbol{n} = (G_i^0, -G_r^0)$ represents the normal direction of the iso-growth lines. Later on, it is possible to construct a surrogate model to map from this single variable Y to the corresponding

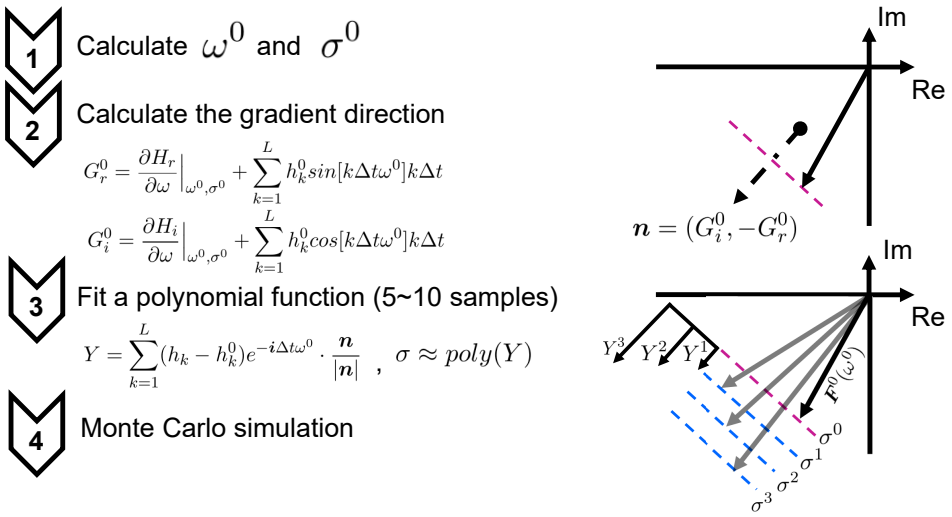


Figure 5.8: A step-by-step workflow of the proposed analytical UQ approach.

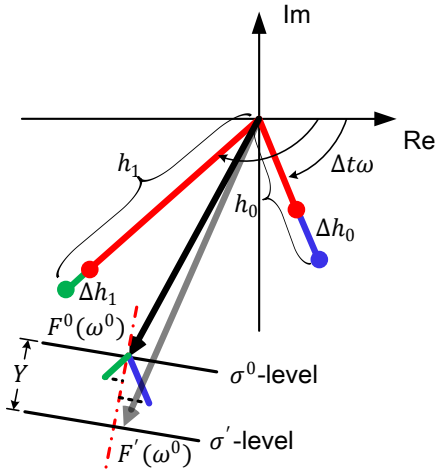


Figure 5.9: Visual demonstration of the causal relationship between the variations of FIR model coefficients and the variations of modal growth rate.

modal growth rate value, and subsequent Monte Carlo procedures can be directly applied to this univariate surrogate model to significantly improve the efficiency of the UQ analysis. A step-by-step illustration of the workflow of the proposed analytical UQ approach is given in Fig. 5.8.

The proposed analytical approach is essentially a dimensionality reduction UQ strategy: Y appears as a linear combination of FIR coefficients h_k, thus encapsulating compactly the causal relationship between the variation of h_k's and the variation of the modal growth rate value. As a result, the surrogate model can be built upon a single variable Y, instead of L FIR model coefficients. Therefore, the dimensionality of the current UQ problem shrinks from L to just one.

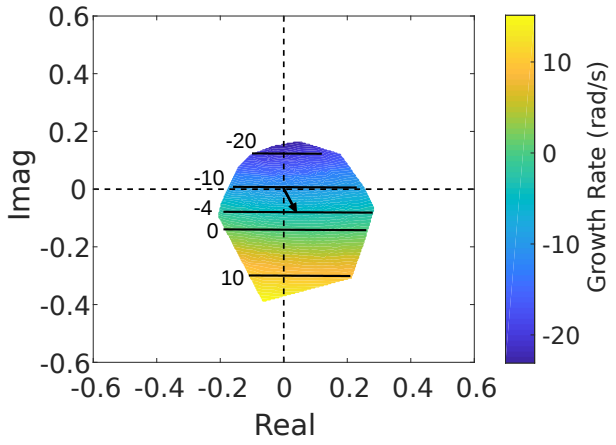


Figure 5.10: Iso-growth lines in the phasor plot of FRF. Black lines are the iso-growth lines calculated by the proposed analytical approach. The colored contours of the growth rate value are predicted by the network calculations. The black arrow line indicates the $F(\omega^0)$ phasor evaluated by using the nominal values of h_k's.

In addition, the derived analytical scheme has opened the "black-box" and shed lights on understanding the dimensionality reduction results yielded by Active Subspace approach. As a matter of fact, one-to-one correspondences can be established between the results obtained by Active Subspace approach and the current analytical strategy: (1) the previously identified active direction can now be conveniently visualized in the phasor plot of FRF and corresponds to the normal direction of the iso-growth lines $\boldsymbol{n} = (G_i^0, -G_r^0)$; (2) the previously identified active variable corresponds to Y in Eq. (5.6); (3) the linear combination coefficients that constitutes the active variable correspond to the projections of the phasor $(h_k - h_k^0)e^{-\boldsymbol{i}(k+1)\Delta t\omega^0}$ on the normal direction of the iso-growth lines. Since all the essential elements produced by the Active Subspace approach can now be retrieved analytically, the overhead cost of using samples to identify the new active variable for a new FIR model can be avoided, thus significantly expanding the applicability of the current dimensionality reduction UQ scheme.

5.2.2 Case study

The case study investigated in Section 5.1.2 is revisited here. The same acoustic network model and flame impulse response model are employed. Instead of using the Active Subspace approach, the newly derived analytical scheme is employed to reduce the problem dimensionality and accelerate the targeted UQ analysis.

As a first step, the distribution pattern of the iso-growth lines in the phasor plot of FRF is assessed. In Fig. 5.10 iso-growth lines predicted by the acoustic network model and the developed analytical scheme are compared. It can be seen that the parallel straight lines indicated by the analytical results are indeed good approximations to the distribution pattern of the iso-growth lines in the phasor plot of FRF.

The proposed analytical approach is employed to quantify the variations of the modal growth rate prediction induced by the uncertain FIR model coefficients. Figure 5.11 displays the com-

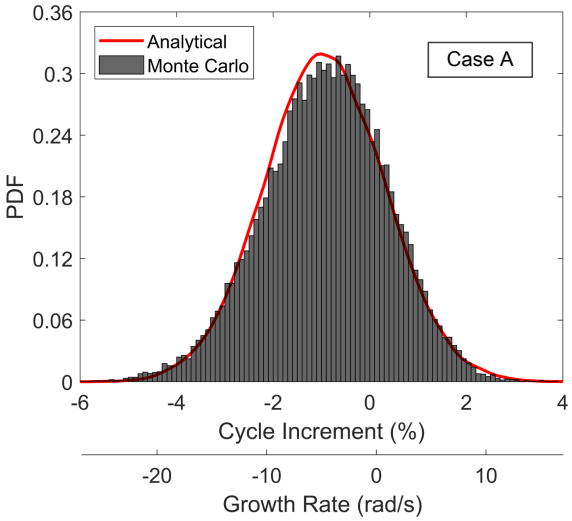


Figure 5.11: PDF of the growth rate predicted by the proposed analytical scheme (red curves) and the reference Monte Carlo simulation (bars).

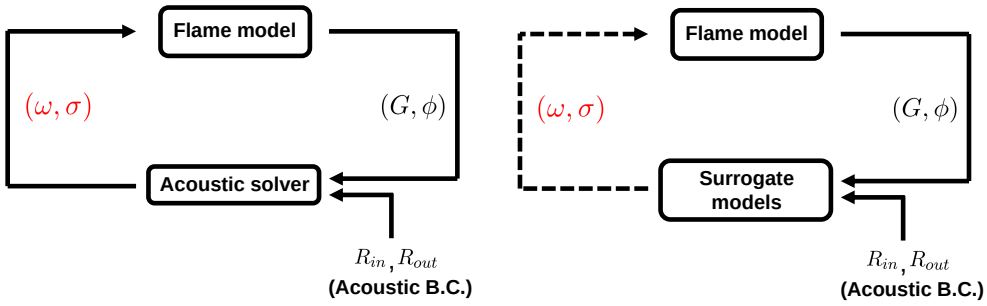
parison of the PDF of the modal growth rate yielded by the proposed analytical scheme and the reference Monte Carlo simulation, which is based on acoustic network evaluations, where excellent match between those two approaches is achieved. In terms of the computational cost, the proposed analytical strategy requires only 5 times of acoustic network calculations to establish a surrogate model to map from Y to the modal growth rate value, therefore yielding another 80-fold increase in uncertainty propagation speed compared to the Active Subspace approach.

5.3 Surrogate-based approach

The strategies presented in the previous two sections are only targeted at propagating uncertainties from FIR model to the eigenmode predictions. In practical thermoacoustic instability analysis, high-dimensional flame models other than FIR models are also widely used to describe the flame dynamic characteristics. As a result, a more general dimensionality reduction scheme needs to be developed to meet the analysis requirements in realistic settings. In this section, a novel surrogate-based dimensionality reduction scheme developed in *paper-ISC20* is introduced. This scheme leverages on the physical intuition of the coupling mechanism between the flame model and the acoustic solver in thermoacoustic instability predictions, therefore possessing the capability of handling general frequency-dependent flame models.

5.3.1 Theory

To predict thermoacoustic instability numerically, a popular approach is to combine a flame model with acoustic solvers and solving the resulting eigenvalue problems in the frequency domain. Under this framework, the flame model serves as the transfer function that takes in the



(a) Thermoacoustic coupling (b) Surrogate coupling

Figure 5.12: (a) In thermoacoustic instability prediction, the flame model takes in the modal eigenvalue and outputs the flame gain and phase, while acoustic solver takes in the flame gain and phase and outputs the modal eigenvalue, thus closing the loop. (b) A surrogate model can be trained to replace the role of the acoustic solver, thus significantly improving the efficiency in calculating the thermoacoustic modes.

eigenvalue $s = \sigma + \boldsymbol{j}\omega$ (σ denotes the modal growth rate and ω denotes the modal frequency) and outputs the flame gain $G = |F(s)|$ and the flame phase $\phi = \angle F(s)$. Meanwhile, the acoustic solver serves as the inverse transfer function that takes in the flame gain G and phase ϕ, as well as other acoustic parameters $\boldsymbol{H}$, and outputs the eigenvalue (ω and σ) of the mode under investigation. By coupling those two transfer functions (as shown in Fig. 5.12a), the eigenvalue of the investigated thermoacoustic mode can be found via iterative numerical schemes.

Since Monte-Carlo-based uncertainty propagation analysis involves repetitively solving the thermoacoustic governing equations with different realizations of uncertain flame and acoustic parameters, the key to accelerate this uncertainty propagation process is to shorten the runtime for each Monte Carlo loop. Naive surrogate modeling approach would achieve this goal by directly building surrogate models to map from flame model parameters $\boldsymbol{F}$ and acoustic system parameters $\boldsymbol{H}$ to the modal frequency and growth rate, thus replacing the expensive thermoacoustic equation-solving with an evaluation of the cheap surrogate models. However, as $\boldsymbol{F}$ is a high-dimensional vector, the "curse of dimensionality" would render this naive surrogate modeling strategy infeasible in practice.

Another way to accelerate the calculation of thermoacoustic eigenmodes lays in the coupling mechanism between the flame model and the acoustic solver. In the coupling loop illustrated in Fig. 5.12a, the path associated with the acoustic solver constitutes the most computation intensive part, since acoustic solvers may involve implementations of expensive numerical schemes (e.g., finite volume method, finite element method, etc.) to calculate the thermoacoustic eigenmodes of 2D/3D complicated geometries [208]. As a result, if two cheap-to-run surrogate models $\hat{f}^{\omega}$ and $\hat{f}^{\sigma}$ can be trained to replace the role of the acoustic solver (as shown in Fig. 5.12b), then the efficiency of evaluating $\omega \approx \hat{f}^{\omega}(G, \phi; \boldsymbol{H})$ and $\sigma \approx \hat{f}^{\sigma}(G, \phi; \boldsymbol{H})$ can be greatly improved, thus leading to accelerated calculations of the thermoacoustic mode and offering opportunities to achieve affordable UQ analysis.

Formally, this surrogate-based governing equations can be cast as follows:

$$\begin{aligned}\hat{f}^{\omega}(|\mathscr{F}(\sigma+\boldsymbol{i}\omega)|, \angle\mathscr{F}(\sigma+\boldsymbol{i}\omega); \boldsymbol{H}) - \omega = 0\\ \hat{f}^{\sigma}(|\mathscr{F}(\sigma+\boldsymbol{i}\omega)|, \angle\mathscr{F}(\sigma+\boldsymbol{i}\omega); \boldsymbol{H}) - \sigma = 0,\end{aligned} \tag{5.7}$$

where the usual iterative numerical schemes can be used to find ω and σ. Eq. (5.7) is defined as the *surrogate equations*. Since now only two flame parameters G and ϕ enter into the surrogate model instead of the primal high-dimensional flame parameters $\boldsymbol{F}$, a dimensionality reduction is achieved by the currently developed scheme. In addition, the current scheme has the benefit that the employed surrogate models only need to be trained once and can be subsequently applied to a wide range of flame models, provided that the considered $G-\phi$ values are within the $G-\phi$ ranges that the surrogate models are built upon. This is desired in realistic applications since it (1) allows using flame models with gradually increased complexity as the analysis process evolves, and (2) offers opportunities to address both the linear and nonlinear thermoacoustic UQ problems under one framework.

5.3.2 Case study

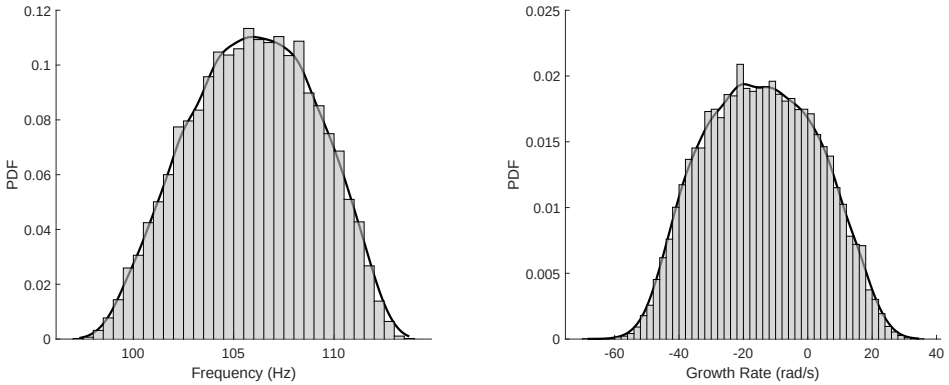
In *paper-ISC20*, the developed surrogate-based UQ strategy is adopted to address a linear case study with uncertain FIR model and acoustic system parameters, as well as a nonlinear case study with uncertain flame describing function (FDF) dataset and acoustic system parameters. Gaussian Process is employed as the surrogate model. An active learning scheme introduced in Section 4.2.2 is used to improve the training efficiency and accuracy. For both case studies, an in-house Helmholtz solver is adopted as the acoustic solver to model an EM2C laboratory-scale test rig (configuration C11, see Section 2.1).

For the linear case study, the goal is to propagate uncertainties from a 65-coefficient FIR model, as well as the magnitude of reflection coefficients $|R_{in}|$ (combustor inlet), $|R_{out}|$ (combustor outlet), and the damping coefficient α, to the modal frequency and growth rate predictions. In total, this is a 68-dimensional UQ problem.

PDF of the modal frequency and growth rate values are shown in Fig. 5.13, where excellent matches can be observed between the results yielded by the surrogate-based UQ scheme and naive Monte Carlo procedure. In terms of the computational speed, an approximately 20-fold increase is achieved for this case study. In addition, it is worth emphasizing that a more significant acceleration of the UQ analysis may be realized when a more complex configuration is considered, as in this case the runtime of a single call of Helmholtz solver would be significantly longer, while the runtime for each calculation of the surrogate equation (Eq. (5.7)) would basically stay the same.

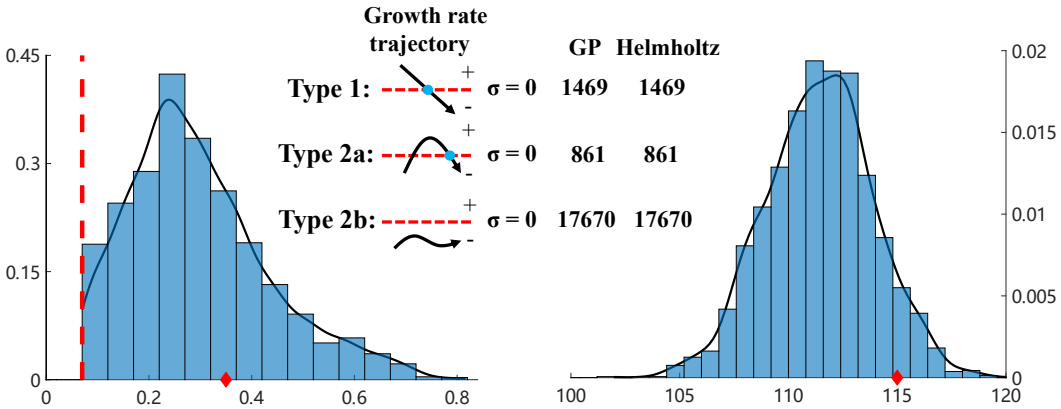
For the nonlinear case study, the goal is to propagate uncertainties from a FDF dataset with 520 uncertain data points, as well as the magnitude of reflection coefficients $|R_{in}|$ (combustor inlet), $|R_{out}|$ (combustor outlet), and the damping coefficient α, to the limit cycle frequency and amplitude. In total, this is a 523-dimensional UQ problem.

PDF of the limit cycle frequency and amplitude are shown in Fig. 5.14, where excellent matches



(a) Frequency (b) Growth Rate

Figure 5.13: PDF comparison between the surrogate-based UQ scheme (black curves) and Monte Carlo applied directly to the Helmholtz solver (bars).



(a) Limit cycle amplitude (b) Limit cycle frequency

Figure 5.14: PDF comparison between the GP-based UQ scheme (black curves) and Monte Carlo applied directly to Helmholtz solver (bars). Three types of growth rate trajectory are observed among 20000 samples (Type 1: linearly unstable; Type 2a: linearly stable, nonlinearly unstable; Type 2b: linearly stable, nonlinearly stable). Although only samples with a Type 1 or 2a trajectory are shown in the histogram, the histogram is normalized with respect to all 20000 samples. The current setup exhibits a probability of 11.6% of being unstable and 88.4% of being stable. Experimental results are shown as the diamond in the figure. Red dashed lines represent the minimum amplitude level of $(|\hat{u}|/\bar{u}_b)_m = 0.07$ in the FDF dataset.

can be observed between the results yielded by the surrogate-based UQ scheme and naive Monte Carlo procedure. In terms of the computational speed, an approximately 15-fold increase is achieved for the current case study. Both case studies demonstrate the accuracy and efficiency of the proposed surrogate-based UQ strategy.

6 Summary and Discussion of Papers

This thesis aims at developing an uncertainty management framework for thermoacoustic instability risk analysis, where the thermoacoustic eigenmodes are calculated via coupling acoustic solvers with a flame response model. Quantification and mitigation are the two consecutive management tasks with different management objectives: Quantification focuses on quantifying the modal instability risk induced by uncertain acoustic system parameters and flame model parameters, while mitigation focuses on mitigating the instability risk by implementing the principles of robust design, given the uncertain parameters. Successfully mitigating the instability risk relies on accurately quantifying the risk in the first place. Therefore, quantification and mitigation steps together constitute a coherent uncertainty management framework. This chapter intends to highlight the logical connections between the major publications discussed in the current thesis and illustrates how each individual publication fits into the overall uncertainty management framework.

6.1 Quantification

Flame model uncertainty usually plays a major role in generating the thermoacoustic modal instability risk. Although uncertainty propagation from simple $n-\tau$ models to eigenmode calculations have been well studied, for more sophisticated and realistic flame impulse response (FIR) model, efficient uncertainty propagation scheme is still lacking due to the large number of uncertain FIR model coefficients. *Paper-ASME18* for the first time assessed the impact of uncertainties in FIR model on the modal instability risk via a novel dimensionality reduction approach called Active Subspace. An one-dimensional active direction was discovered inside of the high dimensional input space. Along this direction the modal growth rate values vary the most. By projecting FIR coefficient vectors onto this active direction, the so-called active variable was obtained. Subsequently, a surrogate model to map the active variable to the modal growth rate value was built, which facilitated 50 times faster Monte Carlo simulations observed in the case studies. As a side outcome of *Paper-ASME18*, an online monitoring pipeline based on Active Subspace approach was built, which allows practitioners to effectively determine the adequate CFD simulation time to reach a satisfactory confidence of the FIR model identification.

An especially interesting observation from *Paper-ASME18* is that a one-dimensional manifold encapsulates the relationship between the variation of modal growth rate values and the variation of FIR coefficients. However, due to the black-box nature of the Active Subspace approach, obtaining a physical interpretation is a non-trivial task. In this context, *Paper-ISC18* is initiated to further investigate the physical insights conveyed by that active direction. Towards that end, linearization analysis was performed on the thermoacoustic governing equation and the phasor

plot of flame frequency response (FFR) was proposed to visualize the results. One of the main contributions of *Paper-ISC18* is that it proves that to first-order, the growth rate contours are parallel straight lines in the phasor plot of FFR. As a result, the active direction is in fact the orthogonal direction of the growth rate contours, and active variable can be perceived as the sum of the projection of individual FIR coefficient vectors onto the active direction. Later on, the same uncertainty propagation procedure developed in *Paper-ASME18* was followed to calculate the modal instability risk. The essential difference between *Paper-ASME18* and *Paper-ISC18* is that the active variable was identified via Active Subspace approach in *Paper-ASME18*, while in *Paper-ISC18* the active variable can be directly derived by using analytical formulas. As a result, dramatic efficiency improvement was further obtained without involving any complex mathematical treatments.

Despite the success of performing UQ analysis on the FIR model in *Paper-ASME18* and *Paper-ISC18*, realistic thermoacoustic UQ analysis may involve uncertainties from both acoustic system parameters and other high-dimensional flame models (e.g., flame describing function employed in nonlinear thermoacoustic instability analysis) in addition to FIR model, thus calling for a more general and powerful UQ framework than the one developed in *Paper-ASME18* and *Paper-ISC18*. To pursue this goal, *Paper-ISC20* spearheaded a general Gaussian-Process (GP) based surrogate strategy for high-dimensional UQ analysis in thermoacoustic instability prediction. The novelty of this strategy lies in the fact that the Gaussian Process models were built directly upon flame gain and phase, as well as other uncertain acoustic parameters. As a result, surrogate equations could be constructed by lining up the GP models and the high-dimensional flame models, which can be efficiently solved via simple iterative schemes to obtain the modal eigenvalues. In addition, the developed strategy possesses a highly desirable feature, i.e., it can quantify uncertainties in modal frequency and linear growth rate for linear thermoacoustic analysis, or quantify uncertainties in limit cycle frequency and amplitude for nonlinear thermoacoustic analysis, all under one surrogate modeling framework. Exceptional performances were observed for a 68-dimensional linear case study and a 523-dimensional nonlinear case study, thus demonstrating the capability of the proposed strategy in addressing thermoacoustic UQ problems in general settings.

6.2 Mitigation

After having developed competences in quantifying the modal instability risk, the next step would be to mitigate the risk given the input uncertainties. *Paper-ASME19* took an approach of implementing the principles of robust design to ensure that the thermoacoustic design is free of instability risk. More specifically, various tasks of robust design were systematically explored, where each task was formulated as an optimization problem and a GP-based UQ procedure was embedded into the optimization routine to efficiently quantify the modal instability risk at each design iteration. Finally, to reveal the interconnection between different robust design scenarios, the concept of a "risk diagram" was proposed that allows a convenient visualization of the distribution of the modal instability risk over the entire parameter space.

In practice, achieving robust thermoacoustic design relies heavily on the accuracy of the risk calculation at each optimization iteration. *Paper-ASME19* assumed that the employed GP models

were accurate with negligible model uncertainty. However, that assumption is not always valid in reality: limited computational resources means only a certain number of training samples can be afforded for GP model training. As a consequence, this "imperfect" training process may introduce epistemic uncertainty to the trained GP model, which subsequently induce the variation of modal risk calculation. *Paper-ASME20* dropped that assumption made in *Paper-ASME19* and specifically investigate strategies to propagate the GP model uncertainty to the instability risk calculation, as well as to reduce the GP model uncertainty given a limited training sample budget. Those strategies can be seamlessly integrated into the robust design workflow proposed in *Paper-ASME19*, thus contributing to mitigating modal instability risk in a reliable manner.

The main outcome of the thesis work was published in 5 papers. In the following, main findings of each paper as well as their contributions to the overall thesis are summarized. The contribution of the present author to each paper is explicitly described. The respective publications are fully reproduced in the appendix.

S. Guo, C. F. Silva, A. Ghani, W. Polifke. Quantification and propagation of uncertainties in identification of flame impulse response for thermoacoustic stability analysis. ASME Journal of Engineering for Gas Turbines and Power, 141(2): 021032, 2019. doi: 10.1115/1.4041652. (Peer-reviewed)

Label: *Paper-ASME18*

Outcome:

- A dimensionality reduction approach called Active Subspace is employed, which identifies an one-dimensional subspace that can accurately describe the relationship between the variation of high dimensional flame impulse response (FIR) model coefficients and the variation of modal growth rate values.
- Efficient uncertainty propagation scheme is proposed by exploiting the identified one-dimensional subspace.
- For the first time, the impact of uncertain FIR model on thermoacoustic stability prediction is assessed.
- An online monitoring procedure based on Active Subspace approach is proposed to assess the uncertainty of the resulting modal growth rate calculation, as longer CFD time series become available for FIR model identification. This tool is beneficial for practitioners to determine the required time series length to reach a satisfactory FIR model identification.

Relevance for the thesis:

This work marks the first step towards developing competence in propagating uncertainties from high-dimensional realistic flame models to the thermoacoustic eigenmodes calculation. Therefore, this work contributes to the quantification branch of the overall uncertainty management framework investigated in the current thesis. Flame impulse response model is investigated specifically, since its sophistication and strong representative ability have gained its popularity in thermoacoustic community, while its uncertainty induced by system identification process has been largely ignored.

Contribution: The research objective was formulated by W. Polifke. The idea of using Active Subspace approach to achieve the objective was a result of a discussion between W. Polifke, C. F. Silva and M. Bauerheim in 2014 Center for Turbulence Research Summer Program (Stanford University). The present author carries out the study, namely, implemented the Active Subspace approach to design the uncertainty propagation scheme. In addition, the present author proposed an online indicator to ensure satisfactory FIR model identification. The present author performed data post-processing and analyzing, composed and wrote the manuscript, took care of the rebuttal and implemented the changes requested by the reviewers. C. F. Silva, A. Ghani, and W. Polifke provided significant suggestions for study improvements and manuscript revision.

Reference: GuoSilva18b

Comment: A first version of this publication was published in the proceedings of the *ASME Turbo Expo 2018: Turbomachinery Technical Confernce & Exposition*.

S. Guo, C. F. Silva, M. Bauerheim, A. Ghani, W. Polifke. Evaluating the impact of uncertainty in flame impulse response model on thermoacoustic instability prediction: A dimensionality reduction approach. Proceedings of the Combustion Institute, 37(4): 5299-5306, 2019. doi: 10.1016/j.proci.2018.07.020. (Peer-reviewed)

Label: *Paper-ISC18*

Outcome:

- It can be proved that to first-order, the growth rate contours are parallel straight lines on the phasor plot of flame frequency response.
- An analytical scheme is developed accordingly to compactly describe the relationship how the variations of FIR model coefficients would lead to the variations of modal growth rate, thus effectively shrinking the problem dimensionality.
- The developed analytical scheme is further integrated into a UQ framework, which targets at propagating uncertainties from the flame impulse response model to the modal growth rate calculation. The accuracy of the above UQ framework is benchmarked in a case study, where a 5000-fold increase in computational efficiency compared to reference Monte Carlo simulation is observed.
- The developed analytical scheme directly offers sensitivity measurement for each FIR model coefficient. This information is beneficial in terms of obtaining more physical intuitions regarding the key mechanisms that controls the thermoacoustic instability.

Relevance for the thesis: This work is a natural extension of the study conducted in *paper-ASME18*. The physical intuition of the active direction and active variable identified by Active Subspace approach is uncovered in the current work, which leads to an analytical UQ scheme that is much more efficient than the one based on Active Subspace proposed in *paper-ASME18*. Same as *paper-ASME18*, this work also contributes to the quantification branch of the overall uncertainty management framework. Both *paper-ASME18* and the current study have generated extensive knowledge on the topic of quantifying modal instability risk induced by the uncertain FIR model.

Contribution: The research objective and scope was inspired by discussions between W. Polifke and M. Bauerheim. Numerous discussions between the present author, C. F. Silva, and A. Ghani generated the idea of using the phasor plot of flame frequency response (FFR) to visualize the results. The present author proved the key observation that to first-order the growth rate contours are parallel straight lines on the FFR phasor plot. In addition, the present author also formulated the relevant UQ strategy, as well as benchmarked on the case study. C. F. Silva and A. Ghani assisted in the interpretation of the results. Finally, the present author wrote the manuscript, took care of the rebuttal and implemented the changes requested by the reviewers. W. Polifke and M. Bauerheim contributed critical revisions to both results interpretation and manuscript preparation.

Reference: GuoSilva18a

S. Guo, C. F. Silva, W. Polifke. A Gaussian-Process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability prediction. Proceedings of the Combustion Institute, 2020, doi: 10.1016/j.proci.2020.06.229. (Peer-reviewed)

Label: *Paper-ISC20*

Outcome:

- A Gaussian-Process-based framework is developed to tackle UQ analyses in thermoacoustic instability prediction in practical settings. This includes using advanced acoustic solver (Helmholtz solver), considering uncertainties from both high-dimensional flame models and acoustic system parameters, as well as handling uncertain parameters with flexible statistical descriptions and large variational ranges.
- An active learning scheme is presented and deployed to accelerate Gaussian Process model training.
- The developed framework is capable of addressing UQ problems accurately and efficiently for both linear and nonlinear thermoacoustic instability predictions.
- The conducted nonlinear case study indicates that the previously observed mismatch between a numerical prediction of the limit cycle and the corresponding experimental results may be attributed to the uncertainties existed in flame describing function data and acoustic system properties, in the light of the UQ analysis delivered by the developed framework.

Relevance for the thesis: This work constitutes the final piece of contribution to the quantification branch of the overall uncertainty management framework investigated in the current thesis. The UQ methodology developed in the current work is much more comprehensive than the previous strategies investigated in *paper-ASME18* and *paper-ISC18*. Its capability meets the need for solving a wide range of thermoacoustic UQ problems in general settings, thus serving as a strong finish along the line of work in quantifying the thermoacoustic modal instability risk.

Contribution: The present author determined the research objective and the scope, as well as developing the UQ methodology and conducted case studies to benchmark the methodology. C. F. Silva provided valuable suggestions regarding selecting the case studies and formulating the overall storyline of the paper. W. Polifke contributed critical revisions and significant suggestions for polishing the study. The present author wrote the manuscript, handled the rebuttal and implemented the changes requested by the reviewers.

Reference: GuoSilva20b

S. Guo, C. F. Silva, W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. ASME Journal of Engineering for Gas Turbines and Power, 142(3): 031026, 2020. doi: 10.1115/1.4044197. (Peer-reviewed)

Label: *Paper-ASME19*

Outcome:

- Various robust design scenarios in thermoacoustic analysis are explored and summarized, with detailed mathematical formulations and efficient solution strategies proposed. These scenarios include fundamental risk analysis, ideal/realistic control designs and inverse tolerance design.
- A novel concept called "risk diagram" is proposed, which allows practitioners to conveniently visualize the modal risk distribution over the entire parameter space, locate associated feasible parameter regions of the robust design tasks, and gain intuitions of the interconnections between different goals of robust design tasks.
- Demonstrate that Gaussian process surrogate modeling approach is capable of handling a high number of uncertain parameters with large variational ranges and mixed distribution types, thus delivering highly efficient risk calculation required by each of the robust design task.

Relevance for the thesis: This work is geared towards the mitigation branch of the overall uncertainty management framework investigated in the current thesis. As a first step, it is crucial to firstly identify the scenarios where risk mitigation is desired, secondly develop the associated mathematical formulations, and finally propose corresponding solution strategies. This current work has accomplished all three points and laid solid foundation for achieving robust thermoacoustic designs.

Contribution: The research idea was conceived in 2018 ASME Turbo Expo conference, when the present author had a fruitful discussion with other colleagues from gas turbine sections. Later on, the present author refined the research objective and scope, conducted detailed research, and wrote the manuscript. C. F. Silva and W. Polifke provided valuable suggestions on literature review and polishing the manuscript. The present author wrote the rebuttal and implemented the changes suggested by the reviewers.

Reference: GuoSilva20

Comment: A first version of this publication was published in the proceedings of the *ASME Turbo Expo 2019: Turbomachinery Technical Confernce & Exposition*.

S. Guo, C. F. Silva, W. Polifke. Reliable calculation of thermoacoustic instability risk using an imperfect surrogate model. ASME Journal of Engineering for Gas Turbines and Power, 2020. doi: 10.1115/1.4049314 (Peer-reviewed)

Label: *Paper-ASME20*

Outcome:

- The compound impact of aleatory parameter uncertainties and epistemic surrogate model uncertainty on the modal instability risk calculation is investigated.
- Efficient strategy is developed to quantify the variation of modal instability risk induced by the uncertain surrogate model.
- An active learning scheme is proposed to intelligently allocate training samples to the vicinity of stability margin, thus significantly reduce the surrogate model uncertainty in calculation the modal instability risk.

Relevance for the thesis: This work makes further contribution to the mitigation perspective of the uncertainty management framework investigated in the current thesis, i.e., it provides means to obtain reliable risk calculations without assuming the employed surrogate model is "perfect". This assumption is widely adopted in the community, which is unfortunately not always valid in realistic settings. The methodology to quantify and reduce surrogate model uncertainties can be seamlessly integrated into the robust design workflow investigated in *paper-ASME19*. The combined contribution from *paper-ASME19* and the current study offers a comprehensive framework to perform robust thermoacoustic design.

Contribution: The present author defined the research objective, conducted research to develop schemes for quantifying and reducing the surrogate model uncertainty. C. F. Silva and W. Polifke reviewed the manuscript written by the present author, and provided valuable insights which helped improve the quality of the paper. The present author later handled the rebuttal and implemented the changes suggested by the reviewers.

Reference: GuoSilva20a

Comment: A first version of this publication was published in the proceedings of the *ASME Turbo Expo 2020: Turbomachinery Technical Confernce & Exposition*.

7 Conclusion and Outlook

7.1 Conclusions

The research objective of the present thesis is to develop a comprehensive uncertainty management framework to quantify and subsequently mitigate thermoacoustic modal instability risk. A network model and a Helmholtz solver were coupled with flame response models to calculate thermoacoustic eigenmodes. Flame model parameters and acoustic system parameters were considered to be uncertain. Those uncertain parameters would propagate downstream via acoustic solvers, thus leading to generate modal instability risk.

Regarding the quantification aspect of the investigated uncertainty management framework, UQ strategies were firstly developed to quantify the modal instability risk induced by an uncertain flame impulse response (FIR) model. These strategies relies on identifying and exploiting a low-dimensional manifold that exists inside of the original high-dimensional FIR coefficient space. Both data-driven approach (i.e., Active Subspace) and analytical approach (i.e., linearizing thermoacoustic governing equations) were pursued to identify this low-dimensional manifold.

Later on, more comprehensive UQ schemes were developed to address linear/nonlinear thermoacoustic UQ problems in realistic settings, featuring large variational ranges of the uncertain parameters, uncertain parameters from both system acoustic and high-dimensional flame models, as well as advanced acoustic solvers (e.g., Helmholtz solver).

Regarding the mitigation aspect of the investigated uncertainty management framework, the current thesis focused on implementing the principles of robust design in a reliable way. To start with, various scenarios of instability risk mitigation were categorized and corresponding solution strategies were proposed and benchmarked. The concept of a "risk diagram" was proposed to visualize risk distribution over the entire parameter space and identify feasible solution region for each robust design task. Subsequently, issues of calculating modal instability risk using imperfect surrogate model were addressed in detail. More specifically, methods were developed to firstly quantify the variation of risk calculation by exploiting estimates of prediction uncertainty offered by the Gaussian Process model, and secondly, to improve the robustness of risk calculation by taking advantage of an active learning scheme for efficient Gaussian Process model training.

In summary, the present thesis constituted a solid step towards efficiently managing uncertainties in thermoacoustic instability analysis. Competences were developed for uncertainty quantification and instability risk mitigation in realistic settings. Moreover, technical details of the advanced dimensionality reduction techniques and surrogate modeling techniques were thoroughly discussed and their applications in realistic case studies were demonstrated. This should be beneficial for practitioners to employ those methods to address other UQ related problems,

e.g., sensitivity analysis, inverse uncertainty propagations, etc.

7.2 Outlook

Despite the achievements made in the current thesis in terms of managing uncertainties associated with thermoacoustic instability analysis, there are several other issues that deserve further investigations.

7.2.1 Impact of uncertain operating conditions on the FIR identification

It is well-known that flame dynamics are highly sensitive to the variations in combustor operating conditions. As a result, uncertainties in operating conditions will alter the flame impulse response (FIR) model, which further induces variations and compromises robustness in modal instability risk predictions. In practice, propagate uncertainties from operating conditions to eigenmode predictions via FIR constitutes a challenging task: First of all, identifying FIR under different combustor operating conditions requires high-fidelity Large Eddy simulations (LES). The associated high computational cost prohibits a direct application of Monte Carlo simulation; Secondly, the FIR identification procedure itself introduces epistemic uncertainties due to the limited time series, combustion noise, etc. A careful treatment of the mixed uncertainty types, i.e., aleatory uncertainty in operating conditions and epistemic uncertainty in FIR identifications, is necessary to develop the full pipeline of uncertainty propagation from operating conditions to thermoacoustic eigenmode predictions. This proposed work is ongoing at the time of writing this thesis and the preliminary results have been submitted to ASME Turbo Expo 2021. It is expected that the techniques developed here can be transferred to tackle other combustion systems (e.g., spray combustion) and handle more uncertain operating condition parameters.

7.2.2 Accurate, robust and efficient identification of flame frequency response

Our experience shows that flame model usually constitutes a major source of uncertainty in the framework of combining flame models and acoustic solvers to calculate thermoacoustic eigenmodes. As a result, efforts spent on reducing the uncertainty associated with identifying flame models from LES will pay off in terms of achieving more accurate and robust eigenmode calculation. Two state-of-the-art flame frequency response (FFR) identification techniques, i.e., harmonic excitation method and broadband excitation method, are either accurate, but computationally expensive, or efficient, but with considerable uncertainty if signal noise is significant. As a result, it is natural to formulate the following research question, namely, could we obtain a better FFR identification by combining the strengths while avoid the weaknesses of those two methods? Multi-fidelity modeling framework may shed lights on this issue. More specifically, a short-time broadband analysis results can be regarded as low-fidelity, which offers a qualitatively correct trend of the FFR; Meanwhile, harmonic analysis results obtained on several carefully selected frequencies can be regarded as high-fidelity, which offers quantitatively

accurate point-estimations of the FFR. By aggregating those two sources of information, it is possible to obtain a globally more accurate, robust and efficient FFR identification. Preliminary results of the proposed work have been submitted to Journal of Sound and Vibration and the paper is under review at the time of writing this thesis.

7.2.3 Reliable calculation of thermoacoustic system instability risk using imperfect surrogate models

In *paper-ASME20*, we have demonstrated how to quantify the variation of instability risk induced by the epistemic surrogate model uncertainty, as well as how to reduce this undesirable variation by using an active learning scheme. A major limitation of the developed methodology is that it only considers one single thermoacoustic mode, which hinders its practical deployment when multiple potentially unstable modes exist. Further more, in terms of surrogate model training, as adjoint method becomes popular in the community, we now can easily obtain information regarding the sensitivity of eigenmode with respect to the input parameters. This sensitivity information should be leveraged to improve the accuracy and efficiency of constructing the surrogate model. Therefore, the current methodology in *paper-ASME20* needs to be extended so that we can assimilate gradient information in surrogate model building and robustly calculate thermoacoustic system instability risk by taking into account the uncertainties of the employed surrogate model.

7.2.4 Efficient dimensionality reduction scheme for thermoacoustic uncertainty quantification analysis for annular combustor

In *paper-ISC20*, we have demonstrated that by building surrogate models on flame gain G and phase ϕ, we are able to effectively break the "curse of dimensionality" and accurately deliver uncertainty quantification results for high-dimensional flame models. Although this idea was only benchmarked on a longitudinal burner, it has the potential to be employed in annular combustor settings. Achieving efficient surrogate modeling is especially important in annular combustor settings as they are inherently high-dimensional (due to the large number of burners) and generating a high number of training samples is not practical (acoustic solvers generally take more time to evaluate an annular configuration than a longitudinal configuration). When applying our novel surrogate modeling scheme to annular combustors, the surrogate models could still be built based on a single burner. However, the challenge has shifted to effectively model the acoustic interactions between individual burners. If those interactions could be conveniently captured, we are then able to configure a system of surrogate equations, which would allow us to efficiently compute the interested eigenmode of the annular combustor, given arbitrary flame models for individual burners. This could pave the way for further research in uncertainty quantification and risk mitigation for annular combustors.

Bibliography

[1] S. M. Correa. A Review of NOX formation under gas turbine combustion conditions. *Combust. Sci. and Tech.*, 87:329–362, 1992.

[2] S. Candel. Combustion Dynamics and Control: Progress and Challenges. *Proceedings of the Combustion Institute*, 29(1):1–28, 2002.

[3] Y. B. Zel'dovich. The Oxidation of Nitrogen in Combustion and Explosions. *Acta Physicochimica U.R.S.S.*, 21(4):577–628, 1946.

[4] T. Lieuwen and V. Yang, editors. *Combustion Instabilities in Gas Turbine Engines: Operational Experience, Fundamental Mechanisms, and Modeling*, volume 210 of *Progress in Astronautics and Aeronautics*. AIAA, 2005. ISBN 1-56347-669-X.

[5] J. O'Connor, V. Acharya, and T. Lieuwen. Transverse combustion instabilities: Acoustic, fluid mechanic, and flame processes. *Progress in Energy and Combustion Science*, 49: 1–39, Aug. 2015. ISSN 03601285. doi: 10.1016/j.pecs.2015.01.001.

[6] J. J. Keller. Thermoacoustic Oscillations in Combustion Chambers of Gas Turbines. *AIAA Journal*, 33(12):2280–2287, 1995. doi: 10.2514/3.12980.

[7] T. Lieuwen and K. McManus. Combustion Dynamics in Lean-Premixed Prevaporized (LPP) Gas Turbines. *Journal of Propulsion and Power*, 19(5):721–721, 2003. ISSN 0748-4658. doi: 10.2514/2.6171.

[8] Y. Huang and V. Yang. Dynamics and Stability of Lean-Premixed Swirl-Stabilized Combustion. *Progress in Energy and Combustion Science*, 35(4):293–364, 2009. ISSN 0360-1285. doi: 10.1016/j.pecs.2009.01.002.

[9] T. Poinsot and D. Veynante. *Theoretical and Numerical Combustion*. CNRS, Paris, third edition, 2012. ISBN 978-2-7466-3990-4.

[10] W. Polifke. Low-Order Analysis Tools for Aero- and Thermo-Acoustic Instabilities. In C. Schram, editor, *Advances in Aero-Acoustics and Thermo-Acoustics*, number VKI LS 2011-01 in VKI Lecture Series. Van Karman Institute for Fluid Dynamics, Rhode-St-Genèse, BE, 2010. ISBN 978-2-87516-012-6.

[11] A. P. Dowling. The Calculation of Thermoacoustic Oscillation. *Journal of Sound and Vibration*, 180(4):557–581, 1995. doi: 10.1006/jsvi.1995.0100.

[12] T. Emmert, S. Jaensch, C. Sovardi, and W. Polifke. taX - a Flexible Tool for Low-Order Duct Acoustic Simulation in Time and Frequency Domain. In *7th Forum Acusticum*, Krakow, Sept. 2014. DEGA.

[13] W. Polifke. Six Lectures on Thermoacoustic Combustion Instability. In *21st CISM-IUTAM Int'l Summer School on Measurement, Analysis and Passive Control of Thermoacoustic Oscillations*, Udine, Italy, 2015, June.

[14] L. Benoit and F. Nicoud. Numerical assessment of thermo-acoustic instabilities in gas turbines. *Int. J. of Numerical Methods in Fluids*, 47:849–855, 2005.

[15] C. F. Silva, F. Nicoud, T. Schuller, D. Durox, and S. Candel. Combining a Helmholtz solver with the flame describing function to assess combustion instability in a premixed swirled combustor. *Combustion and Flame*, 160(9):1743–1754, Sept. 2013. ISSN 0010-2180. doi: 10.1016/j.combustflame.2013.03.020.

[16] L. Crocco. Aspects of Combustion Stability in Liquid Propellant Rocket Motors Part 2: Low Frequency Instability with Bipropellants. High Frequency Instabilitiy. *Journal of American Rocket Society*, 22(1):7–16, 1952. doi: 10.2514/8.4410.

[17] W. Polifke. Black-Box System Identification for Reduced Order Model Construction. *Annals of Nuclear Energy*, 67C:109–128, May 2014. ISSN 0306-4549. doi: 10.1016/j.anucene.2013.10.037.

[18] W. Polifke. Modeling and Analysis of Premixed Flame Dynamics by Means of Distributed Time Delays. *Prog. Energy Combust. Sci.*, 79:100845, 2020. doi: 10.1016/j.pecs.2020.100845.

[19] N. Noiray, D. Durox, T. Schuller, and S. Candel. A Unified Framework for Nonlinear Combustion Instability Analysis Based on the Flame Describing Function. *Journal of Fluid Mechanics*, 615:139–167, 2008. doi: 10.1017/S0022112008003613.

[20] M. P. Juniper and R. I. Sujith. Sensitivity and Nonlinearity of Thermoacoustic Oscillations. *Annual Review of Fluid Mechanics*, 50(1):661–689, 2018. doi: 10.1146/annurev-fluid-122316-045125.

[21] T. Poinsot. Prediction and Control of Combustion Instabilities in Real Engines. *Proceedings of the Combustion Institute*, 36:1–28, 2017. ISSN 1540-7489. doi: 10.1016/j.proci.2016.05.007.

[22] C. Hirsch, D. Wunsch, J. Szumbarski, L. Laniewski-Wollk, and J. Pons-Prats. *Uncertainty Management for Robust Industrial Design in Aeronautics*, volume 140 of *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*. Springer Berlin Heidelberg, New York, NY, July 2018. ISBN 978-3-319-77766-5.

[23] G. Cowan. *Statistical Data Analysis*. Clarendon Press, Oxford : New York, 1 edition edition, June 1998. ISBN 978-0-19-850155-8.

[24] G. James, D. Witten, T. Hastie, and R. Tibshirani. Linear Regression. In G. James, D. Witten, T. Hastie, and R. Tibshirani, editors, *An Introduction to Statistical Learning:*

With Applications in R, Springer Texts in Statistics, pages 59–126. Springer, New York, NY, 2013. ISBN 978-1-4614-7138-7. doi: 10.1007/978-1-4614-7138-7_3.

[25] A. Ndiaye, M. Bauerheim, and F. Nicoud. Uncertainty Quantification of Thermoacoustic Instabilities on A Swirled Stabilized Combustor. In *Proceedings of the ASME Turbo Expo 2015: Turbine Technical Conference and Exposition*, GT2015-44133, Montreal, Quebec, Canada, June 2015. ASME. doi: 10.1115/GT2015-44133.

[26] M. B. Giles and N. A. Pierce. An introduction to the adjoint approach to design. *Flow, turbulence and combustion*, 65(3-4):393–415, 2000.

[27] L. Magri. Adjoint methods as design tools in thermoacoustics. *Applied Mechanics Reviews*, Feb. 2019. ISSN 0003-6900. doi: 10.1115/1.4042821.

[28] L. Magri and M. P. Juniper. Sensitivity analysis of a time-delayed thermo-acoustic system via an adjoint-based approach. *Journal of Fluid Mechanics*, 719:183–202, Mar. 2013. ISSN 1469-7645. doi: 10.1017/jfm.2012.639.

[29] L. Magri and M. P. Juniper. Global modes, receptivity, and sensitivity analysis of diffusion flames coupled with duct acoustics. *Journal of Fluid Mechanics*, 752:237–265, Aug. 2014. doi: 10.1017/jfm.2014.328.

[30] J. G. Aguilar, L. Magri, and M. P. Juniper. Adjoint-based sensitivity analysis of low-order thermoacoustic networks using a wave-based approach. *Journal of Computational Physics*, 341:163–181, July 2017. ISSN 0021-9991. doi: 10.1016/j.jcp.2017.04.013.

[31] L. Magri, M. P. Juniper, and J. P. Moeck. Sensitivity of the Rayleigh criterion in thermoacoustics. *Journal of Fluid Mechanics*, 882:R1, Jan. 2020. ISSN 0022-1120, 1469-7645. doi: 10.1017/jfm.2019.860.

[32] L. Magri, M. Bauerheim, and M. P. Juniper. Stability analysis of thermo-acoustic nonlinear eigenproblems in annular combustors. Part I. Sensitivity. *Journal of Computational Physics*, 325:395–410, Nov. 2016. ISSN 0021-9991. doi: 10.1016/j.jcp.2016.07.032.

[33] A. Orchini and M. P. Juniper. Linear Stability and Adjoint Sensitivity Analysis of Thermoacoustic Networks with Premixed Flames. *Combustion and Flame*, 165:97–108, Mar. 2016. ISSN 00102180. doi: 10.1016/j.combustflame.2015.10.011.

[34] C. F. Silva, L. Prieto, M. Ancharek, P. Marigliani, G. A. Mensah, and G. A. Mensah. Adjoint-based calculation of parametric thermoacoustic maps of an industrial combustion chamber. In *ASME Turbo Expo 2020: Turbomachinery Technical Conference and Exposition*, Virtual, Online, 2020.

[35] F. Caeiro, C. Sovardi, K. Förner, and W. Polifke. Shape Optimization of a Helmholtz Resonator using an Adjoint Method. *Int. J. Spray Combust. Dyn.*, 9(4):394–408, 2017. doi: 10.1177/1756827717703576.

[36] G. A. Mensah and J. P. Moeck. Acoustic Damper Placement and Tuning for Annular Combustors: An Adjoint-Based Optimization Study. *Journal of Engineering for Gas Turbines and Power*, 139(6):061501, June 2017. ISSN 0742-4795, 1528-8919. doi: 10.1115/1.4035201.

[37] L. Magri and M. P. Juniper. A Theoretical Approach for Passive Control of Thermoacoustic Oscillations: Application to Ducted Flames. *Journal of Engineering for Gas Turbines and Power*, 135(9):091604, Sept. 2013. ISSN 0742-4795, 1528-8919. doi: 10.1115/1.4024957.

[38] R. McClarren. *Uncertainty Quantification and Predictive Computational Science: A Foundation for Physical Scientists and Engineers*. Springer International Publishing, 2018. ISBN 978-3-319-99524-3. doi: 10.1007/978-3-319-99525-0.

[39] L. Magri, M. Bauerheim, F. Nicoud, and M. P. Juniper. Stability analysis of thermoacoustic nonlinear eigenproblems in annular combustors. Part II. Uncertainty quantification. *Computational Physics*, 325:411–421, Nov. 2016. ISSN 0021-9991. doi: 10.1016/j.jcp.2016.08.043.

[40] C. F. Silva, L. Magri, T. Runte, and W. Polifke. Uncertainty quantification of growth rates of thermoacoustic instability by an adjoint Helmholtz solver. *J. Eng. Gas Turbines and Power*, 139(1):011901, 2017. doi: 10.1115/1.4034203.

[41] P. Palies, D. Durox, T. Schuller, and S. Candel. The Combined Dynamics of Swirler and Turbulent Premixed Swirling Flames. *Combustion and Flame*, 157(9):1698–1717, Sept. 2010. ISSN 0010-2180. doi: 10.1016/j.combustflame.2010.02.011.

[42] P. Palies, D. Durox, T. Schuller, and S. Candel. Nonlinear combustion instability analysis based on the flame describing function applied to turbulent premixed swirling flames. *Combustion and Flame*, 158(10):1980–1991, Oct. 2011. ISSN 0010-2180. doi: 10.1016/j.combustflame.2011.02.012.

[43] G. A. Mensah, L. Magri, and J. P. Moeck. Methods for the Calculation of Thermoacoustic Stability Margins and Monte Carlo-Free Uncertainty Quantification. In *Proceedings of ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition*, Charlotte, NC, USA, June 2017. ASME. doi: 10.1115/GT2017-64829.

[44] O. L. Maitre and O. M. Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Scientific Computation. Springer Netherlands, 2010. ISBN 978-90-481-3519-6. doi: 10.1007/978-90-481-3520-2.

[45] D. Xiu. *Numerical Methods for Stochastic Computations - A Spectral Method Approach*. Princeton University Press, Princeton, N.J, illustrated edition edition, July 2010. ISBN 978-0-691-14212-8.

[46] G. Blatman and B. Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6):2345–2367, Mar. 2011. ISSN 0021-9991. doi: 10.1016/j.jcp.2010.12.021.

[47] P. S. Palar, T. Tsuchiya, and G. T. Parks. Multi-fidelity non-intrusive polynomial chaos based on regression. *Computer Methods in Applied Mechanics and Engineering*, 305: 579–606, June 2016. ISSN 0045-7825. doi: 10.1016/j.cma.2016.03.022.

[48] V. Yaghoubi, S. Marelli, B. Sudret, and T. Abrahamsson. Sparse polynomial chaos expansions of frequency response functions using stochastic frequency transformation.

Probabilistic Engineering Mechanics, 48:39–58, Apr. 2017. ISSN 0266-8920. doi: 10.1016/j.probengmech.2017.04.003.

[49] S. Marelli and B. Sudret. An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis. *Structural Safety*, 75: 67–74, Nov. 2018. ISSN 0167-4730. doi: 10.1016/j.strusafe.2018.06.003.

[50] E. Torre, S. Marelli, P. Embrechts, and B. Sudret. Data-driven polynomial chaos expansion for machine learning regression. *Journal of Computational Physics*, 388:601–623, July 2019. ISSN 00219991. doi: 10.1016/j.jcp.2019.03.039.

[51] G. Andrianov, S. Burriel, S. Cambier, A. Dutfoy, I. Dutka-Malen, E. de Rocquigny, B. Sudret, P. Benjamin, R. Lebrun, F. Mangeant, and M. Pendola. Open TURNS, an open source initiative to Treat Uncertainties, Risks'N Statistics in a structured industrial approach. In *Proceedings of the European Safety and Reliability Conference 2007, ESREL 2007 - Risk, Reliability and Societal Safety*, volume 2, Jan. 2007.

[52] B. M. Adams, W. Bohnhoff, K. Dalbey, J. Eddy, M. Eldred, D. Gay, K. Haskell, P. D. Hough, and L. Swiler. DAKOTA, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 5.0 user's manual. *Sandia National Laboratories, Tech. Rep. SAND2010-2183*, 2009.

[53] J. Feinberg and H. P. Langtangen. Chaospy: An open source tool for designing methods of uncertainty quantification. *Journal of Computational Science*, 11(Supplement C):46–57, Nov. 2015. ISSN 1877-7503. doi: 10.1016/j.jocs.2015.08.008.

[54] S. Marelli and B. Sudret. UQLab user manual - Polynomial chaos expansions. Technical Report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Zurich, Switzerland, 2019.

[55] D. Xiu and G. E. M. Karniadakis. The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations. *Siam J. Sci. Comput*, pages 619–644, 2002.

[56] H. N. Najm. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Annual Review of Fluid Mechanics*, 41:35–52, 2009.

[57] O. Le Maître and O. M. Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Springer Science & Business Media, 2010.

[58] Schöbi R., Sudret B., and Marelli S. Rare Event Estimation Using Polynomial-Chaos Kriging. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 3(2):D4016002, June 2017. doi: 10.1061/AJRUA6.0000870.

[59] A. Avdonin, S. Jaensch, C. F. Silva, M. Češnovar, and W. Polifke. Uncertainty quantification and sensitivity analysis of thermoacoustic stability with non-intrusive polynomial chaos expansion. *Combustion and Flame*, 189:300–310, Mar. 2018. doi: 10.1016/j.combustflame.2017.11.001.

[60] V. N. Kornilov, R. Rook, J. H. M. ten Thije Boonkkamp, and L. P. H. de Goey. Experimental and Numerical Investigation of the Acoustic Response of Multi-Slit Bunsen Burners. *Combustion and Flame*, 156(10):1957–1970, Oct. 2009. ISSN 0010-2180. doi: 10.1016/j.combustflame.2009.07.017.

[61] A. Avdonin and W. Polifke. Quantification of the Impact of Uncertainties in Operating Conditions on the Flame Transfer Function with Non-Intrusive Polynomial Chaos Expansion. *J. Eng. Gas Turbines and Power*, 141(1):011020, 2019. doi: 10.1115/1.4040745.

[62] I. Sobol'. Sensitivity Estimates for Nonlinear Mathematical Models. *Mathematical Modeling & Computational Experiment*, 1:407–414, 1993.

[63] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, Chichester, England ; Hoboken, NJ, 1. edition edition, 2008. ISBN 978-0-470-05997-5.

[64] P. G. Constantine. *Active Subspaces: Emerging Ideas in Dimension Reduction for Parameter Studies*, volume 2. SIAM, Philadelphia, US, 2015.

[65] M. Bauerheim, A. Ndiaye, P. Constantine, S. Moreau, and F. Nicoud. Symmetry breaking of azimuthal thermoacoustic modes: The UQ perspective. *Journal of Fluid Mechanics*, 789:534–566, Feb. 2016. ISSN 0022-1120, 1469-7645. doi: 10.1017/jfm.2015.730.

[66] C. F. Silva, P. Pettersson, G. Iaccarino, and M. Ihme. Uncertainty quantification of combustion noise by generalized polynomial chaos and state-space models. *Combustion and Flame*, 217:113–130, July 2020. ISSN 00102180. doi: 10.1016/j.combustflame.2020.03.010.

[67] T. Emmert, M. Meindl, S. Jaensch, and W. Polifke. Linear State Space Interconnect Modeling of Acoustic Systems. *Acta Acustica united with Acustica*, 102(5):824–833, 2016. doi: 10.3813/AAA.918997.

[68] C. F. Silva, M. Merk, T. Komarek, and W. Polifke. The Contribution of Intrinsic Thermoacoustic Feedback to Combustion Noise and Resonances of a Confined Turbulent Premixed Flame. In *International Symposium: Thermoacoustic Instabilities in Gas Turbines and Rocket Engines*, Garching, Germany, May 2016.

[69] A. Albayrak, T. Steinbacher, T. Komarek, and W. Polifke. Convective Scaling of Intrinsic Thermo-Acoustic Eigenfrequencies of a Premixed Swirl Combustor. *Journal of Engineering for Gas Turbines and Power*, 140(4):041510, Nov. 2017. ISSN 0742-4795. doi: 10.1115/1.4038083.

[70] R. Lebrun and A. Dutfoy. Do Rosenblatt and Nataf isoprobabilistic transformations really differ? *Probabilistic Engineering Mechanics*, 24(4):577–584, Oct. 2009. ISSN 0266-8920. doi: 10.1016/j.probengmech.2009.04.006.

[71] L. R. Tay Wo Chong Hilares. *Numerical Simulation of the Dynamics of Turbulent Swirling Flames*. Ph.D. Thesis, TU München, Munich, Germany, 2012.

[72] C. Sovardi, S. Jaensch, and W. Polifke. Concurrent Identification of Aero-acoustic Scattering and Noise Sources at a Flow Duct Singularity in low Mach Number Flow. *J. Sound Vibration*, 377:90–105, Sept. 2016. doi: 10.1016/j.jsv.2016.05.025.

[73] A. I. J. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1):50–79, Jan. 2009. ISSN 0376-0421. doi: 10.1016/j.paerosci.2008.11.001.

[74] S. Nannapaneni, Z. Hu, and S. Mahadevan. Uncertainty quantification in reliability estimation with limit state surrogates. *Structural and Multidisciplinary Optimization*, 54(6):1509–1526, Dec. 2016. ISSN 1615-1488. doi: 10.1007/s00158-016-1487-1.

[75] G. J. Park, T. H. Lee, K. H. Lee, and K. H. Hwang. Robust Design: An Overview. *AIAA Journal*, 44(1):181–191, Jan. 2006. doi: 10.2514/1.13639.

[76] N. V. Sahinidis. Optimization under uncertainty: State-of-the-art and opportunities. *Computers & Chemical Engineering*, 28(6):971–983, June 2004. ISSN 0098-1354. doi: 10.1016/j.compchemeng.2003.09.017.

[77] S. Bade, M. Wagner, C. Hirsch, T. Sattelmayer, and B. Schuermans. Design for Thermo-Acoustic Stability: Modeling of Burner and Flame Dynamics. *J. Eng. Gas Turbines Power*, 135(11):111502–111502–7, Sept. 2013. ISSN 0742-4795. doi: 10.1115/1.4025001.

[78] S. Bade, M. Wagner, C. Hirsch, T. Sattelmayer, and B. Schuermans. Design for Thermo-Acoustic Stability: Procedure and Database. *J. Eng. Gas Turbines and Power*, 135(12):121507–121507–8, Sept. 2013. ISSN 0742-4795. doi: 10.1115/1.4025131.

[79] J. G. Aguilar and M. P. Juniper. Adjoint Methods for Elimination of Thermoacoustic Oscillations in a Model Annular Combustor via Small Geometry Modifications. In *ASME Turbo Expo 2018: Turbomachinery Technical Conference and Exposition*, GT2018-75692, Lillestrom, Norway, 2018. ASME. doi: 10.1115/GT2018-75692.

[80] D. Yang, F. M. Sogaro, A. S. Morgans, and P. J. Schmid. Optimising the acoustic damping of multiple Helmholtz resonators attached to a thin annular duct. *Journal of Sound and Vibration*, Dec. 2018. ISSN 0022-460X. doi: 10.1016/j.jsv.2018.12.023.

[81] U. M. Diwekar. Optimization Under Uncertainty. In U. M. Diwekar, editor, *Introduction to Applied Optimization*, Applied Optimization, pages 145–208. Springer US, Boston, MA, 2003. ISBN 978-1-4757-3745-5.

[82] S. Guo, C. F. Silva, M. Bauerheim, A. Ghani, and W. Polifke. Evaluating the impact of uncertainty in flame impulse response model on thermoacoustic instability prediction: A dimensionality reduction approach. *Proceedings of the Combustion Institute*, 37:5299–5306, 2019. doi: 10.1016/j.proci.2018.07.020.

[83] S. Guo, C. F. Silva, A. Ghani, and W. Polifke. Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis. *J. Eng. Gas Turbines and Power*, 141(2):021032–10, Feb. 2019. ISSN 0742-4795. doi: 10.1115/1.4041652.

[84] S. Guo, C. F. Silva, and W. Polifke. A Gaussian-Process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability prediction. *Proceedings of the Combustion Institute*, 38, Jan. 2021.

[85] S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. *Journal of Engineering for Gas Turbines and Power*, 142(3), Mar. 2020. ISSN 0742-4795. doi: 10.1115/1.4044197.

[86] S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using an Imperfect Surrgate Model. *Journal of Engineering for Gas Turbines and Power*, Dec. 2020. ISSN 0742-4795. doi: 10.1115/1.4049314.

[87] R. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, Philadelphia, illustrated edition edition, Mar. 2014. ISBN 978-1-61197-321-1.

[88] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis, Third Edition*. CRC Press, Nov. 2013. ISBN 978-1-4398-4095-5.

[89] H. Yu, T. Jaravel, M. Ihme, M. P. Juniper, and L. Magri. Data Assimilation and Optimal Calibration in Nonlinear Models of Flame Dynamics. *Journal of Engineering for Gas Turbines and Power*, 141(12), Dec. 2019. ISSN 0742-4795. doi: 10.1115/1.4044378.

[90] H. Yu, M. P. Juniper, and L. Magri. Combined state and parameter estimation in level-set methods. *Journal of Computational Physics*, 399:108950, Dec. 2019. ISSN 00219991. doi: 10.1016/j.jcp.2019.108950.

[91] K. Law, A. Stuart, and K. Zygalakis. *Data Assimilation: A Mathematical Introduction*. Springer, Cham Heidelberg New York Dordrecht London, 1st ed. 2015 edition edition, Sept. 2015. ISBN 978-3-319-20324-9.

[92] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, July 2008. ISSN 0951-8320. doi: 10.1016/j.ress.2007.04.002.

[93] W. L. Oberkampf and C. J. Roy. *Verification and Validation in Scientific Computing*. Cambridge University Press, New York, illustrated edition edition, Oct. 2010. ISBN 978-0-521-11360-1.

[94] C. J. Roy and W. L. Oberkampf. A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. *Computer Methods in Applied Mechanics and Engineering*, 200(25):2131–2144, June 2011. ISSN 0045-7825. doi: 10.1016/j.cma.2011.03.016.

[95] D. J. Murray-Smith. *Testing and Validation of Computer Simulation Models: Principles, Methods and Applications*. Simulation Foundations, Methods and Applications. Springer International Publishing, 2015. ISBN 978-3-319-15098-7. doi: 10.1007/978-3-319-15099-4.

[96] C. F. Silva, T. Emmert, S. Jaensch, and W. Polifke. Numerical Study on Intrinsic Thermoacoustic Instability of a Laminar Premixed Flame. *Combustion and Flame*, 162(9): 3370–3378, 2015. ISSN 0010-2180. doi: 10.1016/j.combustflame.2015.06.003.

[97] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA, illustrated edition edition, Aug. 2012. ISBN 978-0-262-01802-9.

[98] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer, 1st ed. 2013, corr. 7th printing 2017 edition edition, June 2013.

[99] S. Abe. *Support Vector Machines for Pattern Classification*. Advances in Computer Vision and Pattern Recognition. Springer-Verlag, London, 2005. ISBN 978-1-84628-219-5. doi: 10.1007/1-84628-219-5.

[100] I. Steinwart and A. Christmann. *Support Vector Machines*. Information Science and Statistics. Springer-Verlag, New York, 2008. ISBN 978-0-387-77241-7. doi: 10.1007/978-0-387-77242-4.

[101] A. Basudhar and S. Missoum. Reliability Assessment using Probabilistic Support Vector Machines (PSVMs). In *51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. American Institute of Aeronautics and Astronautics, 2010. doi: 10.2514/6.2010-2764.

[102] C. Sheppard. *Tree-Based Machine Learning Algorithms: Decision Trees, Random Forests, and Boosting*. CreateSpace Independent Publishing Platform, Sept. 2017. ISBN 978-1-975860-97-4.

[103] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, USA, Oxford : New York, illustrated edition edition, Jan. 1996. ISBN 978-0-19-853864-6.

[104] B. D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, 1996. ISBN 978-0-521-71770-0. doi: 10.1017/CBO9780511812651.

[105] C. C. Aggarwal. *Neural Networks and Deep Learning: A Textbook*. Springer, Cham, 1st ed. 2018 edition edition, Sept. 2018. ISBN 978-3-319-94462-3.

[106] S. Jaensch and W. Polifke. Uncertainty Encountered When Modelling Self-Excited Thermoacoustic Oscillations with Artificial Neural Networks. *Int. J. Spray Combust. Dyn.*, 9 (4):367–379, 2017. doi: 10.1177/1756827716687583.

[107] N. Tathawadekar, A. K. Doan, C. F. Silva, and N. Thuerey. Modelling of the nonlinear flame response of a Bunsen-type flame via multi-layer perceptron. *38th Symposium on Combustion*, January 24th to 29th, 2021. doi: 10.1016/j.proci.2020.07.115.

[108] M. McCartney, M. Haeringer, and W. Polifke. Comparison of Machine Learning Algorithms in the Interpolation and Extrapolation of Flame Describing Functions. *J. Eng. Gas Turbines and Power*, 142:14, 2020. doi: 10.1115/1.4045516.

[109] A. Roy, R. Manna, and S. Chakraborty. Support vector regression based metamodeling for structural reliability analysis. *Probabilistic Engineering Mechanics*, 55:78–89, Jan. 2019. ISSN 0266-8920. doi: 10.1016/j.probengmech.2018.11.001.

[110] R. K. Tripathy and I. Bilionis. Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification. *Journal of Computational Physics*, 375: 565–588, Dec. 2018. ISSN 0021-9991. doi: 10.1016/j.jcp.2018.08.036.

[111] R. Schobi, B. Sudret, and J. Wiart. Polynomial-chaos-based kriging. *International Journal for Uncertainty Quantification*, 5(2), 2015. ISSN 2152-5080, 2152-5099. doi: 10.1615/Int.J.UncertaintyQuantification.2015012467.

[112] P. Kersaudy, B. Sudret, N. Varsier, O. Picon, and J. Wiart. A new surrogate modeling technique combining Kriging and polynomial chaos expansions – Application to uncertainty analysis in computational dosimetry. *Journal of Computational Physics*, 286: 103–117, Apr. 2015. ISSN 0021-9991. doi: 10.1016/j.jcp.2015.01.034.

[113] M. D. Spiridonakos and E. N. Chatzi. Metamodeling of dynamic nonlinear structural systems through polynomial chaos NARX models. *Computers & Structures*, 157:99–113, Sept. 2015. ISSN 0045-7949. doi: 10.1016/j.compstruc.2015.05.002.

[114] C. V. Mai, M. D. Spiridonakos, E. N. Chatzi, and B. Sudret. Surrogate modeling for stochastic dynamical systems by combining nonlinear autoregressive with exogenous input models and polynomial chaos expansions. *International Journal for Uncertainty Quantification*, 6(4), 2016. ISSN 2152-5080, 2152-5099. doi: 10.1615/Int.J.UncertaintyQuantification.2016016603.

[115] R. Schöbi. *Surrogate Models for Uncertainty Quantification in the Context of Imprecise Probability Modelling*. Ph.D. Thesis, ETH Zürich, Zürich, Switzerland, 2017.

[116] L. Brevault, S. Lacaze, M. Balesdent, and S. Missoum. Reliability Analysis in the Presence of Aleatory and Epistemic Uncertainties, Application to the Prediction of a Launch Vehicle Fallout Zone. *Journal of Mechanical Design*, 138(11):111401–111401–11, Sept. 2016. ISSN 1050-0472. doi: 10.1115/1.4034106.

[117] L. Tay-Wo-Chong, S. Bomberg, A. Ulhaq, T. Komarek, and W. Polifke. Comparative Validation Study on Identification of Premixed Flame Transfer Function. *J. of Eng. Gas Turbines Power*, 134(2):021502–1–8, 2012. doi: 10.1115/1.4004183.

[118] O. M. Knio and O. P. Le Maître. Uncertainty propagation in CFD using polynomial chaos decomposition. *Fluid Dynamics Research*, 38(9):616–640, Sept. 2006. ISSN 0169-5983. doi: 10.1016/j.fluiddyn.2005.12.003.

[119] H. Bijl, D. Lucor, S. Mishra, and C. Schwab, editors. *Uncertainty Quantification in Computational Fluid Dynamics*. Lecture Notes in Computational Science and Engineering. Springer International Publishing, 2013. ISBN 978-3-319-00884-4. doi: 10.1007/978-3-319-00885-1.

[120] L. W. Cook, A. A. Mishra, J. P. Jarrett, K. E. Willcox, and G. Iaccarino. Optimization under turbulence model uncertainty for aerospace design. *Physics of Fluids*, 31(10): 105111, Oct. 2019. ISSN 1070-6631. doi: 10.1063/1.5118785.

[121] A. A. Mishra and G. Iaccarino. Theoretical analysis of tensor perturbations for uncertainty quantification of Reynolds averaged and subgrid scale closures. *Physics of Fluids*, 31(7):075101, July 2019. ISSN 1070-6631. doi: 10.1063/1.5099176.

[122] 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. ISSN 1070-6631. doi: 10.1063/1.5086341.

[123] A. A. Mishra, J. Mukhopadhaya, G. Iaccarino, and J. Alonso. Uncertainty Estimation Module for Turbulence Model Predictions in SU2. *AIAA Journal*, 57(3):1066–1077, 2019. ISSN 0001-1452. doi: 10.2514/1.J057187.

[124] K. Duraisamy, G. Iaccarino, and H. Xiao. Turbulence Modeling in the Age of Data. *Annual Review of Fluid Mechanics*, 51(1):357–377, 2019. doi: 10.1146/annurev-fluid-010518-040547.

[125] M. E. Mueller, G. Iaccarino, and H. Pitsch. Chemical kinetic uncertainty quantification for Large Eddy Simulation of turbulent nonpremixed combustion. *Proceedings of the Combustion Institute*, 34(1):1299–1306, Jan. 2013. ISSN 1540-7489. doi: 10.1016/j.proci.2012.07.054.

[126] M. E. Mueller and V. Raman. Effects of turbulent combustion modeling errors on soot evolution in a turbulent nonpremixed jet flame. *Combustion and Flame*, 161(7):1842–1848, July 2014. ISSN 0010-2180. doi: 10.1016/j.combustflame.2013.12.020.

[127] M. Khalil, G. Lacaze, J. C. Oefelein, and H. N. Najm. Uncertainty quantification in LES of a turbulent bluff-body stabilized flame. *Proceedings of the Combustion Institute*, 35 (2):1147–1156, Jan. 2015. ISSN 1540-7489. doi: 10.1016/j.proci.2014.05.030.

[128] D. Sondak, T. Oliver, C. Simmons, and R. Moser. An Inadequacy Formulation for an Uncertain Flamelet Model. In *19th AIAA Non-Deterministic Approaches Conference*. American Institute of Aeronautics and Astronautics, 2017. doi: 10.2514/6.2017-1091.

[129] B. Enderle, B. Rauch, F. Grimm, G. Eckel, and M. Aigner. Non-intrusive uncertainty quantification in the simulation of turbulent spray combustion using Polynomial Chaos Expansion: A case study. *Combustion and Flame*, 213:26–38, Mar. 2020. ISSN 0010-2180. doi: 10.1016/j.combustflame.2019.11.021.

[130] F. Duchaine, F. Boudy, D. Durox, and T. Poinsot. Sensitivity analysis of transfer functions of laminar flames. *Combustion and Flame*, 158(12):2384–2394, Dec. 2011. ISSN 0010-2180. doi: 10.1016/j.combustflame.2011.05.013.

[131] C. Silva, K. J. Yong, and L. Magri. Thermoacoustic Modes of Quasi-One-Dimensional Combustors in the Region of Marginal Stability. *Journal of Engineering for Gas Turbines and Power*, 141(2):021022, Feb. 2019. doi: 10.1115/1.4041118.

[132] K. J. Yong, M. Meindl, W. Polifke, and C. F. Silva. Thermoacoustic spectrum of a swirled premixed combustor with partially reflecting boundaries. *J. Eng. Gas Turbines Power*, 142((1)):011005, 2020. doi: 10.1115/1.4045275.

[133] T. Emmert, S. Bomberg, S. Jaensch, and W. Polifke. Acoustic and Intrinsic Thermoacoustic Modes of a Premixed Combustor. *Proceedings of the Combustion Institute*, 36: 3835–3842, 2017. doi: 10.1016/j.proci.2016.08.002.

[134] T. Komarek and W. Polifke. Impact of Swirl Fluctuations on the Flame Response of a Perfectly Premixed Swirl Burner. *Journal of Engineering for Gas Turbines and Power*, 132(6):061503, June 2010. doi: 10.1115/1.4000127.

[135] F. Nicoud, L. Benoit, C. Sensiau, and T. Poinsot. Acoustic Modes in Combustors with Complex Impedances and Multidimensional Active Flames. *AIAA Journal*, 45(2):426–441, 2007. doi: 10.2514/1.24933.

[136] R. W. Hamming. *Introduction to Applied Numerical Analysis*. Dover Publications Inc., Mineola, N.Y, reprint edition edition, Feb. 2012. ISBN 978-0-486-48590-4.

[137] R. S. Blumenthal, P. Subramanian, R. Sujith, and W. Polifke. Novel Perspectives on the Dynamics of Premixed Flames. *Combustion and Flame*, 160(7):1215–1224, Mar. 2013. doi: 10.1016/j.combustflame.2013.02.005.

[138] H. Bodén and W. Polifke. Uncertainty quantification applied to aeroacoustic predictions. In C. Schram, editor, *Progress in Simulation, Control and Reduction of Ventilation Noise*, number VKI LS 2016-02 in VKI Lecture Series 2015. Van Karman Institute, Rhode-St-Genèse, BE, 2016. ISBN ISBN-13 978-2-87516-098-0.

[139] L. Tay-Wo-Chong, T. Komarek, R. Kaess, S. Föller, and W. Polifke. Identification of Flame Transfer Functions from LES of a Premixed Swirl Burner. In *Proceedings of ASME Turbo Expo 2010*, GT2010-22769, pages 623–635, Glasgow, UK, June 2010. ASME. ISBN 978-0-7918-4397-0. doi: 10.1115/GT2010-22769.

[140] K. J. Keesman. Time-invariant System Identification. In *System Identification*, Advanced Textbooks in Control and Signal Processing, pages 59–167. Springer London, London, 2011. ISBN 978-0-85729-521-7.

[141] M. Merk, S. Jaensch, C. Silva, and W. Polifke. Simultaneous Identification of Transfer Functions and Combustion Noise of a Turbulent Flame. *J. Sound Vibration*, 422:432–452, May 2018. ISSN 0022460X. doi: 10.1016/j.jsv.2018.02.040.

[142] S. Jaensch, M. Merk, T. Emmert, and W. Polifke. Identification of Flame Transfer Functions in the Presence of Intrinsic Thermoacoustic Feedback and Noise. *Combustion Theory and Modelling*, 22(3):613–634, Mar. 2018. ISSN 1364-7830, 1741-3559. doi: 10.1080/13647830.2018.1443517.

[143] K. Oberleithner and C. O. Paschereit. Modeling Flame Describing Functions Based on Hydrodynamic Linear Stability Analysis. In *ASME Turbo Expo 2016: Turbomachinery Technical Conference and Exposition*, page V04BT04A009, Seoul, South Korea, 2016. doi: 10.1115/GT2016-57316.

[144] A. Albayrak and W. Polifke. Propagation Velocity of Inertial Waves in Cylindrical Swirling Flow. In *23nd Int. Congress on Sound and Vibration (ICSV23)*, Athens, Greece, 2016. IIAV.

[145] D. J. Wagg, K. Worden, R. J. Barthorpe, and P. Gardner. Digital Twins: State-of-the-Art and Future Directions for Modeling and Simulation in Engineering Dynamics Applications. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, 6(3), Sept. 2020. ISSN 2332-9017. doi: 10.1115/1.4046739.

[146] P. D. Arendt, D. W. Apley, and W. Chen. Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifiability. *Journal of Mechanical Design*, 134(10), Oct. 2012. ISSN 1050-0472. doi: 10.1115/1.4007390.

[147] N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, Jan. 2002. ISBN 978-0-89871-521-7. doi: 10.1137/1.9780898718027.

[148] P. Deuflhard and A. Hohmann. *Numerical Analysis in Modern Scientific Computing: An Introduction*. Texts in Applied Mathematics. Springer-Verlag, New York, second edition, 2003. ISBN 978-0-387-95410-3. doi: 10.1007/978-0-387-21584-6.

[149] B. M. Ayyub and G. J. Klir. *Uncertainty Modeling and Analysis in Engineering and the Sciences*. Taylor & Francis Ltd, Boca Raton, FL, May 2006. ISBN 978-1-58488-644-0.

[150] M. Balesdent, J. Morio, and L. Brevault. Rare Event Probability Estimation in the Presence of Epistemic Uncertainty on Input Probability Distribution Parameters. *Methodology and Computing in Applied Probability*, 18(1):197–216, Mar. 2016. ISSN 1573-7713. doi: 10.1007/s11009-014-9411-x.

[151] J. M. Aughenbaugh and C. J. J. Paredis. The Value of Using Imprecise Probabilities in Engineering Design. *Journal of Mechanical Design*, 128(4):969–979, July 2006. ISSN 1050-0472. doi: 10.1115/1.2204976.

[152] S. S. Rao and K. K. Annamdas. A Comparative Study of Evidence Theories in the Modeling, Analysis, and Design of Engineering Systems. *Journal of Mechanical Design*, 135(6), June 2013. ISSN 1050-0472. doi: 10.1115/1.4024229.

[153] D. Sivia and J. Skilling. *Data Analysis: A Bayesian Tutorial*. Oxford University Press, Oxford, 2rev ed edition edition, July 2006. ISBN 978-0-19-856832-2.

[154] R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 2006. ISBN 978-0-387-28659-4. doi: 10.1007/0-387-28678-0.

[155] D. Sivia and J. Skilling. *Data Analysis: A Bayesian Tutorial*. Oxford University Press, Oxford, 2rev ed edition edition, July 2006. ISBN 978-0-19-856832-2.

[156] M. Grigoriu. *Stochastic Systems: Uncertainty Quantification and Propagation*. Springer Series in Reliability Engineering. Springer-Verlag, London, 2012. ISBN 978-1-4471-2326-2. doi: 10.1007/978-1-4471-2327-9.

[157] S. H. Lee and W. Chen. A comparative study of uncertainty propagation methods for black-box-type problems. *Structural and Multidisciplinary Optimization*, 37(3):239, May 2008. ISSN 1615-1488. doi: 10.1007/s00158-008-0234-7.

[158] M. Arnst and J.-P. Ponthot. An overview of nonintrusive characterization, propagation, and sensitivity analysis of uncertainties in computational mechanics. *International Journal for Uncertainty Quantification*, 4(5), 2014. ISSN 2152-5080, 2152-5099. doi: 10.1615/Int.J.UncertaintyQuantification.2014006990.

[159] A. Barbu and S.-C. Zhu. *Monte Carlo Methods*. Springer, 1st ed. 2020 edition edition, Feb. 2020. ISBN 9789811329708.

[160] T. J. Santner, B. J. Williams, and W. I. Notz. Space-Filling Designs for Computer Experiments. In T. J. Santner, B. J. Williams, and W. I. Notz, editors, *The Design and Analysis of Computer Experiments*, Springer Series in Statistics, pages 121–161. Springer, New York, NY, 2003. ISBN 978-1-4757-3799-8. doi: 10.1007/978-1-4757-3799-8_5.

[161] L. Pronzato and W. G. Müller. Design of computer experiments: Space filling and beyond. *Statistics and Computing*, 22(3):681–701, May 2012. ISSN 1573-1375. doi: 10.1007/s11222-011-9242-3.

[162] H. Dai and W. Wang. Application of low-discrepancy sampling method in structural reliability analysis. *Structural Safety*, 31(1):55–64, Jan. 2009. ISSN 0167-4730. doi: 10.1016/j.strusafe.2008.03.001.

[163] C. Lataniotis, E. Torre, S. Marelli, and B. Sudret. UQLab user manual - The Input module. Technical Report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Zurich, Switzerland, 2019.

[164] D. Chen and S. Xiong. Flexible Nested Latin Hypercube Designs for Computer Experiments. *Journal of Quality Technology*, 49(4):337–353, Oct. 2017. ISSN 0022-4065. doi: 10.1080/00224065.2017.11918001.

[165] T. Bedford and R. Cooke. *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press, Cambridge, 2001. ISBN 978-0-521-77320-1. doi: 10.1017/CBO9780511813597.

[166] T. Aven. *Foundations of Risk Analysis*. Wiley, Feb. 2012.

[167] H. G. Beyer and B. Sendhoff. Robust Optimization - A Comprehensive Survey. *Computer Methods in Applied Mechanics and Engineering*, 196(33-34):3190–3218, July 2007. doi: 10.1016/j.cma.2007.03.003.

[168] C. Audet and J. Dennis. Analysis of Generalized Pattern Searches. *SIAM Journal on Optimization*, 13(3):889–903, Jan. 2002. ISSN 1052-6234. doi: 10.1137/S1052623400378742.

[169] A. Custódio, J. Madeira, A. Vaz, and L. Vicente. Direct Multisearch for Multiobjective Optimization. *SIAM Journal on Optimization*, 21(3):1109–1140, July 2011. ISSN 1052-6234. doi: 10.1137/10079731X.

[170] T. Goel, R. Vaidyanathan, R. T. Haftka, W. Shyy, N. V. Queipo, and K. Tucker. Response surface approximation of Pareto optimal front in multi-objective optimization. *Computer Methods in Applied Mechanics and Engineering*, 196(4):879–893, Jan. 2007. ISSN 0045-7825. doi: 10.1016/j.cma.2006.07.010.

[171] R. Srinivasan. Methods of Importance Sampling. In R. Srinivasan, editor, *Importance Sampling: Applications in Communications and Detection*, pages 9–46. Springer, Berlin, Heidelberg, 2002. ISBN 978-3-662-05052-1. doi: 10.1007/978-3-662-05052-1_2.

[172] S.-K. Au and J. L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277, Oct. 2001. ISSN 0266-8920. doi: 10.1016/S0266-8920(01)00019-4.

[173] A. Forrester, A. Sobester, and A. Keane. *Engineering Design via Surrogate Modelling: A Practical Guide*. Wiley, Chichester, West Sussex, England ; Hoboken, NJ, 1. edition edition, July 2008. ISBN 978-0-470-06068-1.

[174] S. H. Kim and F. Boukouvala. Machine learning-based surrogate modeling for data-driven optimization: A comparison of subset selection for regression techniques. *Optimization Letters*, 14(4):989–1010, June 2020. ISSN 1862-4480. doi: 10.1007/s11590-019-01428-7.

[175] C. E. Rasmussen. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, Mass, Nov. 2005. ISBN 978-0-262-18253-9.

[176] J. B. Nagel. *Bayesian Techniques for Inverse Uncertainty Quantification*. Ph.D. Thesis, ETH Zürich, Zürich, Switzerland, 2017.

[177] J. A. Lee and M. Verleysen. *Nonlinear Dimensionality Reduction*. Information Science and Statistics. Springer-Verlag, New York, 2007. ISBN 978-0-387-39350-6. doi: 10.1007/978-0-387-39351-3.

[178] C. O. S. Sorzano, J. Vargas, and A. P. Montano. A survey of dimensionality reduction techniques. *arXiv:1403.2877 [cs, q-bio, stat]*, Mar. 2014.

[179] A. Bhosekar and M. Ierapetritou. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers & Chemical Engineering*, 108:250–267, Jan. 2018. ISSN 0098-1354. doi: 10.1016/j.compchemeng.2017.09.017.

[180] J. Zhang, M. Xiao, and L. Gao. An active learning reliability method combining Kriging constructed with exploration and exploitation of failure region and subset simulation. *Reliability Engineering & System Safety*, 188:90–102, Aug. 2019. ISSN 0951-8320. doi: 10.1016/j.ress.2019.03.002.

[181] H. Liu, Y.-S. Ong, and J. Cai. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. *Structural and Multidisciplinary Optimization*, 57(1):393–416, Jan. 2018. ISSN 1615-1488. doi: 10.1007/s00158-017-1739-8.

[182] E. Schneider, S. Staudacher, B. Schuermans, H. Ye, and T. Meeuwissen. Real-Time Modelling of the Thermoacoustic Dynamics of a Gas Turbine Using a Gaussian Process. In *Proceedings*, volume Volume 2 - Combustion and Fuels, pages 323–332. American Society of Mechanical Engineers, 2007. doi: 10.1115/GT2007-27468.

[183] P. Chattopadhyay, S. Mondal, C. Bhattacharya, A. Mukhopadhyay, and A. Ray. Dynamic Data-Driven Design of Lean Premixed Combustors for Thermoacoustically Stable Operations. *Mechanical Design*, 139(11):111419–111419–10, Oct. 2017. ISSN 1050-0472. doi: 10.1115/1.4037307.

[184] P. Chattopadhyay, S. Mondal, A. Ray, and A. Mukhopadhyay. Dynamic Data-Driven Combustor Design for Mitigation of Thermoacoustic Instabilities. *Journal of Dynamic Systems, Measurement, and Control*, 141(1):014501–014501–7, Sept. 2018. ISSN 0022-0434. doi: 10.1115/1.4040210.

[185] H. Kwon, S. Choi, J.-H. Kwon, and D. Lee. Surrogate-Based Robust Optimization and Design to Unsteady Low-Noise Open Rotors. *Journal of Aircraft*, 53(5):1448–1467, Mar. 2016. ISSN 0021-8669. doi: 10.2514/1.C033109.

[186] K. M. Ryan, J. Kristensen, L. Ling, S. Ghosh, I. Asher, and L. Wang. A Gaussian Process Modeling Approach for Fast Robust Design With Uncertain Inputs. In *Proceedings of ASME Turbo Expo 2018*, GT2018-77007, Oslo, Norway, June 2018. ASME. doi: 10.1115/GT2018-77007.

[187] J. Laurenceau and P. Sagaut. Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging. *AIAA Journal*, 46(2):498–507, Feb. 2008. ISSN 0001-1452. doi: 10.2514/1.32308.

[188] R. Dwight and Z.-H. Han. Efficient Uncertainty Quantification Using Gradient-Enhanced Kriging. In *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Structures, Structural Dynamics, and Materials and Co-Located Conferences. American Institute of Aeronautics and Astronautics, May 2009. doi: 10.2514/6.2009-2276.

[189] Z.-H. Han, S. Görtz, and R. Zimmermann. Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. *Aerospace Science and Technology*, 25(1):177–189, Mar. 2013. ISSN 1270-9638. doi: 10.1016/j.ast.2012.01.006.

[190] Z.-H. Han and S. Görtz. Hierarchical Kriging Model for Variable-Fidelity Surrogate Modeling. *AIAA Journal*, 50(9):1885–1896, Sept. 2012. ISSN 0001-1452. doi: 10.2514/1.J051354.

[191] I. Abdallah, C. Lataniotis, and B. Sudret. Parametric hierarchical kriging for multi-fidelity aero-servo-elastic simulators — Application to extreme loads on wind turbines. *Probabilistic Engineering Mechanics*, 55:67–77, Jan. 2019. ISSN 0266-8920. doi: 10.1016/j.probengmech.2018.10.001.

[192] H. Liu, J. Cai, and Y.-S. Ong. Remarks on multi-output Gaussian process regression. *Knowledge-Based Systems*, 144:102–121, Mar. 2018. ISSN 0950-7051. doi: 10.1016/j.knosys.2017.12.034.

[193] C. Lataniotis, D. Wicaksono, S. Marelli, and B. Sudret. UQLab user manual - Kriging (Gaussian process modeling). Technical Report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Zürich, Switzerland, 2019.

[194] Z. Ugray, L. Lasdon, J. Plummer, F. Glover, J. Kelly, and R. Martí. Scatter Search and Local NLP Solvers: A Multistart Framework for Global Optimization. *INFORMS Journal on Computing*, 19(3):328–340, July 2007. ISSN 1091-9856. doi: 10.1287/ijoc.1060.0175.

[195] R. Bostanabad, T. Kearney, S. Tao, D. W. Apley, and W. Chen. Leveraging the nugget parameter for efficient Gaussian process modeling. *International Journal for Numerical Methods in Engineering*, 114(5):501–516, 2018. ISSN 1097-0207. doi: 10.1002/nme.5751.

[196] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, Jan. 1989. ISBN 978-0-201-15767-3.

[197] R. H. Byrd, J. C. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89(1):149–185, Nov. 2000. ISSN 1436-4646. doi: 10.1007/PL00011391.

[198] J. L. Loeppky, J. Sacks, and W. J. Welch. Choosing the Sample Size of a Computer Experiment: A Practical Guide. *Technometrics*, 51(4):366–376, Nov. 2009. ISSN 0040-1706. doi: 10.1198/TECH.2009.08040.

[199] Y. Ling, K. Ryan, I. Asher, J. Kristensen, S. Ghosh, and L. Wang. Efficient robust design optimization using Gaussian process and intelligent sampling. In *2018 Multidisciplinary Analysis and Optimization Conference, AIAA AVIATION Forum*, AIAA 2018-4175, Atlanta, Georgia, US, 2018. AIAA. doi: 10.2514/6.2018-4175.

[200] H. Liu, J. Cai, and Y.-S. Ong. An adaptive sampling approach for Kriging metamodeling by maximizing expected prediction error. *Computers & Chemical Engineering*, 106(2):171–182, Nov. 2017. doi: 10.1016/j.compchemeng.2017.05.025.

[201] B. Echard, N. Gayton, and M. Lemaire. AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. *Structural Safety*, 33(2):145–154, Mar. 2011. ISSN 0167-4730. doi: 10.1016/j.strusafe.2011.01.002.

[202] D. Chauhan and R. Mathews. Review on Dimensionality Reduction Techniques. In A. P. Pandian, R. Palanisamy, and K. Ntalianis, editors, *Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2019)*, Lecture Notes on Data Engineering and Communications Technologies, pages 356–362, Cham, 2020. Springer International Publishing. ISBN 978-3-030-43192-1. doi: 10.1007/978-3-030-43192-1_41.

[203] P. G. Constantine, M. Emory, J. Larsson, and G. Iaccarino. Exploiting active subspaces to quantify uncertainty in the numerical simulation of the HyShot II scramjet. *Journal of Computational Physics*, 302:1–20, Dec. 2015. ISSN 0021-9991. doi: 10.1016/j.jcp.2015.09.001.

[204] A. Glaws, P. G. Constantine, J. N. Shadid, and T. M. Wildey. Dimension reduction in magnetohydrodynamics power generation models: Dimensional analysis and active subspaces. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 10(5):312–325, 2017. ISSN 1932-1872. doi: 10.1002/sam.11355.

[205] P. G. Constantine and A. Doostan. Time-dependent global sensitivity analysis with active subspaces for a lithium ion battery model. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 10(5):243–262, 2017. ISSN 1932-1872. doi: 10.1002/sam.11347.

[206] Z. J. Grey and P. G. Constantine. Active Subspaces of Airfoil Shape Parameterizations. *AIAA Journal*, 56(5):2003–2017, 2018. ISSN 0001-1452. doi: 10.2514/1.J056054.

[207] P. Constantine, E. Dow, and Q. Wang. Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces. *SIAM Journal on Scientific Computing*, 36(4):A1500–A1524, Jan. 2014. ISSN 1064-8275. doi: 10.1137/130916138.

[208] M. Meindl, A. Albayrak, and W. Polifke. A state-space formulation of a discontinuous Galerkin method for thermoacoustic stability analysis. *J. Sound Vibration*, 481:115431, 2020. doi: 10.1016/j.jsv.2020.115431.

List of Figures

List of Tables

Appendices

A Reproduction of Papers

Publications that included in this dissertation are reproduced in the following appendix. Those results of the reproduced publications were presented at the following conferences:

1. ASME Turbo Expo 2018: Turbomachinery Technical Conference & Exposition, 2018, Oslo, Norway [83].
2. 37th International Symposium on Combustion, 2018, Dublin, Ireland [82]
3. ASME Turbo Expo 2019: Turbomachinery Technical Conference & Exposition, 2019, Phoenix, USA [85].
4. ASME Turbo Expo 2020: Turbomachinery Technical Conference & Exposition, 2020, Virtual, Online [86].
5. 38th International Symposium on Combustion, 2020, Adelaide, Australia [84].

Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis

Shuai Guo[1]
Professur für Thermofluiddynamik,
Technische Universität München,
Boltzmannstr. 15,
Garching D-85748, Germany
e-mail: guo@tfd.mw.tum.de

Camilo F. Silva
Professur für Thermofluiddynamik,
Technische Universität München,
Boltzmannstr. 15,
Garching D-85748, Germany
e-mail: silva@tfd.mw.tum.de

Abdulla Ghani
Professur für Thermofluiddynamik,
Technische Universität München,
Boltzmannstr. 15,
Garching D-85748, Germany
e-mail: ghani@tfd.mw.tum.de

Wolfgang Polifke
Professor Professur für Thermofluiddynamik,
Technische Universität München,
Boltzmannstr. 15,
Garching D-85748, Germany
e-mail: polifke@tfd.mw.tum.de

The thermoacoustic behavior of a combustion system can be determined numerically via acoustic tools such as Helmholtz solvers or network models coupled with a model for the flame dynamic response. Within such a framework, the flame response to flow perturbations can be described by a finite impulse response (FIR) model, which can be derived from large eddy simulation (LES) time series via system identification. However, the estimated FIR model will inevitably contain uncertainties due to, e.g., the statistical nature of the identification process, low signal-to-noise ratio, or finite length of time series. Thus, a necessary step toward reliable thermoacoustic stability analysis is to quantify the impact of uncertainties in FIR model on the growth rate of thermoacoustic modes. There are two practical considerations involved in this topic. First, how to efficiently propagate uncertainties from the FIR model to the modal growth rate of the system, considering it is a high dimensional uncertainty quantification (UQ) problem? Second, since longer computational fluid dynamics (CFD) simulation time generally leads to less uncertain FIR model identification, how to determine the length of the CFD simulation required to obtain satisfactory confidence? To address the two issues, a dimensional reduction UQ methodology called "Active subspace approach (ASA)" is employed in the present study. For the first question, ASA is applied to exploit a low-dimensional approximation of the original system, which allows accelerated UQ analysis. Good agreement with Monte Carlo analysis demonstrates the accuracy of the method. For the second question, a procedure based on ASA is proposed, which can serve as an indicator for terminating CFD simulation. The effectiveness of the procedure is verified in the paper. [DOI: 10.1115/1.4041652]

1 Introduction

Predicting thermoacoustic instability remains a major concern for industry to design more reliable and efficient combustion systems. Common practices to achieve this goal include using acoustic tools such as Helmholtz solvers [1] or network models [2] to numerically solve an eigenvalue problem from which both growth rates and frequencies of the thermoacoustic modes of interest can be obtained. However, due to the nonlinear nature of the eigenvalue equation, the predicted thermoacoustic behaviors can be highly sensitive to variations [3] in geometry, boundary conditions, operating conditions, model parameters (calibrated from noisy experimental or computational data), etc. In some occasions [4], a system estimated as stable using nominal input parameters may become unstable when inputs deviate slightly from their nominal value. Consequently, uncertainty quantification (UQ) analysis, which focuses on quantifying uncertainties on output given uncertain inputs, is essential to fully account for the impact of the variable inputs from a statistical point of view and constitutes a critical step toward a more reliable thermoacoustic stability prediction.

In the process of thermoacoustic instability estimation, one of the main uncertainty sources lies in the flame model [4] which correlates the unsteady heat release rate of the flame to upstream velocity perturbations and is employed as a source term in acoustic solvers to close the thermoacoustic system [5]. Ndiaye et al. [4], Bauerheim et al. [6], Magri et al. [7], and Silva et al. [8] individually investigated the impact of flame model uncertainties on the growth rate of thermoacoustic modes of various combustors. The n–τ flame model was employed, and gain n and time delay τ were considered as uncertain input parameters in these studies. The results highlight the importance of considering flame model uncertainties in order to obtain reliable predictions of combustion instabilities.

Compared with the n–τ model used in the previous UQ studies, the finite impulse response (FIR) model represents a more sophisticated and realistic flame model. It describes flame dynamics in the time domain, facilitates direct physical interpretation of the relevant flow-flame interaction mechanisms [9], and through z-transform [10], the flame transfer function can be easily obtained and readily integrated into acoustic tools to determine thermoacoustic behaviors of the combustion system. In addition, an efficient identification procedure for impulse response models has already been established [10] which combines computational fluid dynamics (CFD) simulation and advanced system identification, and further paves the way for its convenient implementation.

However, due to the influence of the statistical nature of the system identification process, large eddy simulation (LES), low signal-to-noise ratios, finite length of CFD time series [11], etc., the estimated coefficients of the FIR model will inevitably contain uncertainties, which may be characterized by the confidence interval of each coefficient [11]. Those coefficient uncertainties will "propagate" through the thermoacoustic model and affect the output, e.g., the growth rate of the thermoacoustic modes. This aspect of the problem has not yet received enough attention, even though the FIR model has already been employed on several occasions.

[1]Corresponding author.

Manuscript received August 28, 2018; final manuscript received September 14, 2018; published online October 23, 2018. Editor: Jerzy T. Sawicki.

S. Guo, C. F. Silva, A. Ghani, and W. Polifke. Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis. In Journal of Engineering for Gas Turbines and Power, 2018. Reprinted with permission from ASME.

To do a UQ analysis with respect to this problem, two practical aspects should be addressed: first, since FIR model usually contains 15–30 coefficients, making it eligible to be classified as a high-dimensional UQ problem, a natural question is the following: is it possible to avoid high computational cost methods like Monte Carlo simulation [12], but still efficiently propagate input uncertainties to the output? Second, since generally longer simulation time yields less uncertain identification of the FIR coefficients, is it possible to have a guideline regarding how long CFD simulations should be conducted in order to achieve target confidence?

Active subspace approach (ASA) [13], an innovative approach for large-scale UQ analysis, may be able to provide solutions to the abovementioned problems. This approach identifies directions in input parameter space that give strongest variabilities of the output and compactly summarizes them as *active variables*, which form a low-dimensional representation of the original high dimensional system. Subsequently, a surrogate model (SM) can be constructed by only exploiting these active variables, allowing a much faster UQ analysis compared with direct Monte Carlo (DMC) simulations. ASA has solid mathematical foundations [14] and its capability has already been demonstrated in different research problems [15–17].

The ASA has been introduced in the field of thermoacoustic instability by Bauerheim et al. [6], who quantified uncertainties of growth rate in an annular combustor, where 38 uncertain parameters in n–τ flame model (gain n and time delay τ for each flame were considered as uncertain parameters) were reduced to only three active variables. Subsequently, low-order models were built upon these three active variables and the predicted probability of an acoustic mode to be unstable (risk factor) agreed well with the reference Monte Carlo results. Magri et al. [7] extended the work [6] by combining the ASA with an adjoint method to efficiently calculate output gradients. Different surrogate models based on active variables were compared and their respective accuracy was discussed.

In the present work, the uncertainty of modal growth rate calculation given uncertainties in the flame impulse response model is quantified. The objectives are to provide answers for the abovementioned questions, i.e., (1) How to efficiently propagate uncertainties in FIR model to determine the variation of thermoacoustic modal growth rate? (2) How to determine the length of CFD simulations required to achieve less uncertain identification of the flame impulse response model? For the first question, ASA is applied to exploit low-dimensional approximation of the original system and a 50 times faster highly accurate UQ analysis is achieved compared with the reference Monte Carlo method. For the second question, a procedure based on ASA is proposed, which can be used as an indicator for determining adequate CFD simulation time.

The novelties of the current paper are the following: (1) to our best knowledge, this is the first time that the impact of uncertainties in FIR model on thermoacoustic stability prediction is assessed; (2) we not only demonstrate the efficiency of ASA for solving the current UQ problem, we also construct an ASA-based-procedure for CFD time series length determination, which should be highly relevant for practitioners due to the high computational cost of the CFD simulation; and (3) the presented nonintrusive ASA methodology can also be applied to more complex acoustic models (thus handling more complex combustors), where classic UQ methods would be prohibitively expensive.

The paper is organized as follows. Section 2 gives an overview of the impulse response identification procedure and quantifies input uncertainties in terms of confidence intervals. Section 3 describes the investigated combustor and corresponding acoustic network model. Section 4 presents the technical details of implementing ASA. Section 5 demonstrates the effectiveness of active subspace method by comparing the results with Monte Carlo simulations. Section 6 proposes and validates the procedure used for estimating the optimal CFD simulation time. The paper closes with the main conclusions.

2 Impulse Response Model Identification

This section starts with an overview of the flame model identification procedure, followed by applying the procedure to actual LES time series data to derive the impulse response model as well as its uncertainties, the consequences of which will be investigated in the subsequent section. More comprehensive treatments of system identification as well as its uncertainties with applications in aero/thermoacoustic are given in Refs. [10], [11], and [18].

2.1 Identification Procedure. The dynamic properties of a linear time invariant single input single output system are completely characterized by its impulse response [2]. The model structure takes the following form:

$$y[k] = G(q)u[k] + e[k] \tag{1}$$

where $u[k]$, $y[k]$, and $e[k]$ represent discrete input, output, and noise term, respectively. $G[q]$ has the form

$$G(q) = h_0 + h_1 q^{-1} + h_2 q^{-2} + \cdots + h_{(L-1)} q^{-(L-1)} \tag{2}$$

where the backward-shift operator $q^{-n}x[k] = x[k-n]$ is used to refer to past inputs and outputs; L is the order of the polynomial, i.e., the "model order"; the coefficients h_k describe the response of the system submitted to unit impulse excitation, and they are the target coefficients need to be identified. Here, $G(q)$ denotes the FIR model, due to the fact that only a finite number of impulse response coefficients are considered. In the present study, procedures described in Keesman [19] are employed to identify the coefficients h_k. A brief overview is given as follows:

The FIR coefficients of the polynomial $h_0 \ldots h_{L-1}$ can be found by solving the minimization problem given by

$$\arg\min_{\theta} \sum \left(\mathbf{y} - \widehat{\mathbf{y}}(\boldsymbol{\theta})\right)^2 \tag{3}$$

where $\widehat{\boldsymbol{y}}(\boldsymbol{\theta})$ denotes the predicted output from the model and $\boldsymbol{\theta}$ is a place holder for all coefficients. The minimization can be achieved using the constrained least-squares approach, which is also known as a regularization algorithm

$$\widehat{\boldsymbol{\theta}} = \left(\phi^{\mathrm{T}}\boldsymbol{\phi} + k\boldsymbol{I}\right)^{-1}\phi^{\mathrm{T}}\mathbf{y} \tag{4}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_n]$ is the vector containing the output at each time-step, $\boldsymbol{\phi}$ is the regressor matrix which has the form

$$\boldsymbol{\phi} = \begin{pmatrix} u[L] & u[L-1] & \ldots & u[1] \\ u[L+1] & u[L] & \ldots & u[2] \\ u[n] & u[n-1] & \ldots & u[n-(L-1)] \end{pmatrix} \tag{5}$$

the term $k\boldsymbol{I}$ represents regularization with k being the regularization parameter (set to 1 in the present study) and $\boldsymbol{I}$ being the identity matrix. Regularization can help to reduce ill-conditioning in least-squares problems and reduce covariance within estimated parameters.

Uncertainties of the estimated coefficients, which are the source of uncertainty in the focus of the present study, are represented by the covariance matrix, with its diagonal terms being the coefficient variances and its off-diagonal terms being the covariance among pairs of parameters. An estimate for the covariance matrix can be found by multiplying the Gramian matrix of ϕ with the noise variance

$$\widehat{\mathrm{cov}}(\boldsymbol{\theta}) = \left(\phi^{T}\boldsymbol{\phi} + k\boldsymbol{I}\right)^{-1}\sigma_{\varepsilon}^2 \tag{6}$$

In practice, the noise variance is estimated from the residuals $\varepsilon = \mathbf{y} - \boldsymbol{\phi}\widehat{\boldsymbol{\theta}}$ as

S. Guo, C. F. Silva, A. Ghani, and W. Polifke. Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis. In Journal of Engineering for Gas Turbines and Power, 2018. Reprinted with permission from ASME.

$$\sigma_{\varepsilon}^2 \approx \frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i)^2 \tag{7}$$

It is worth mentioning that the UQ analysis presented in this paper is independent of the specific identification procedure. As long as proper nominal values and covariance matrix of the FIR model parameters are provided, the UQ workflow proposed in the paper should work just fine. Details will be demonstrated in Sec. 5.

2.2 Identification Results. For the identification procedure, CFD time series of velocity perturbations u' (recorded at the burner mouth) were considered as input, i.e., $u = u'$, and global heat release rate fluctuations $\dot{Q}'$ were considered as output, i.e., $y = \dot{Q}'$. The data were obtained through LES by Tay-Wo-Chong et al. [20] and are presented in Fig. 1.

Figure 2 displays the impulse response coefficients h_k's identified by applying the procedure introduced in Sec. 2.1. $L = 16$ coefficients (i.e., impulse response coefficients h_k's) are chosen for the current model. A good practice of how to properly choose the number of coefficients L (or model order) is given in Ref. [11]. Both the nominal value and the 95% confidence interval are shown for each coefficient.

To summarize, through the identification process, nominal values as well as the covariance matrix of the FIR model coefficients are obtained. In the framework of UQ, FIR model coefficients will be treated as random variables following multivariate normal distribution.

3 Thermoacoustic Framework

In this section, details of the combustor configuration are presented first, followed by outlining the corresponding acoustic network model representation. The FIR model derived from Sec. 2.2 will be plugged into the acoustic network model to calculate the growth rate of thermoacoustic modes.

3.1 Combustor Configuration. The BRS combustor configuration investigated in the present study represents a turbulent premixed swirl burner test rig, which was experimentally investigated by Komarek and Polifke [21] and numerically investigated by Tay-Wo-Chong et al. [20]. The configuration consists of a plenum, a duct with an axial swirl generator, and a combustion chamber. It operates with an equivalence ratio of 0.77 of perfectly premixed methane-air mixture and a thermal power of 30 kW.

3.2 Thermoacoustic Modeling. In the present study, a low order acoustic network model is employed to predict thermoacoustic modes of the combustion system. The analysis is carried out in frequency domain by the assumption of linear and time-harmonic acoustics. Figure 3 displays the network model representing the burner test rig. The rig is discretized into interconnected acoustic elements, where each element is characterized by its acoustic transfer matrix, which relates the characteristic wave amplitude f and g [22], between the upstream and downstream ports of the elements. Mathematical formulations for different types of element used in this study, e.g., boundary conditions, duct, area jump, and flame, are given in Silva et al. [2].

The assembly of the individual transfer matrices yields a homogeneous system of equations for the unknown characteristic wave amplitudes. The corresponding characteristic equation is obtained by letting the determinant of the system matrix be equal to zero, and its solutions give the frequencies and the growth rates of the thermoacoustic modes.

In the present analysis, losses are not considered in modeling the area change, and zero Mach number is assumed for simplicity. The axial swirler is assumed to be acoustically transparent. The transfer matrix for the flame element is derived from linearized Rankine-Hugoniot relations. The FIR model is introduced to relate velocity fluctuations u' at the burner mouth to global heat release fluctuations $\dot{Q}'$ of the flame, thus closing the acoustic network model. Table 1 summaries geometrical and thermodynamic parameters used in the acoustic network model. Two pairs of combustion chamber length values and reflection coefficient values for the combustor exit are selected (marked as A and B in subscript), which are considered as two different cases to be submitted to subsequent UQ analysis.

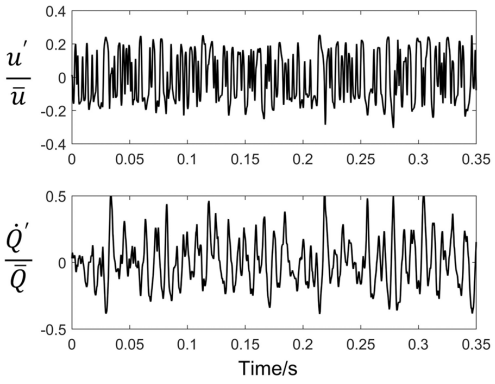


Fig. 1 Normalized velocity and global heat release rate fluctuations, the total length of the data is 350 ms

4 Active Subspace Approach

This section describes the implementation of the active subspace approach in the framework of the current study. More details on its mathematical foundation as well as its implementations in other research fields are given by Constantine [14]. Figure 4 summarizes the workflow. Details in each step are given in the following:

4.1 Preparation: Inputs Transformation. The original correlated uncertain coefficients (represented in vector form, same applies after) $\boldsymbol{h} = (h_0, \ldots, h_{L-1})$ of the FIR model need to be transformed before being submitted to active subspace approach. Orthogonal transformation [23] and normalization are adopted sequentially to transform $\boldsymbol{h}$ to $\tilde{\boldsymbol{h}}$. The latter follows independent, standard, multivariate normal distribution. The inverse of the above process can convert $\tilde{\boldsymbol{h}}$ back to $\boldsymbol{h}$, which constitutes natural inputs for the acoustic network model.

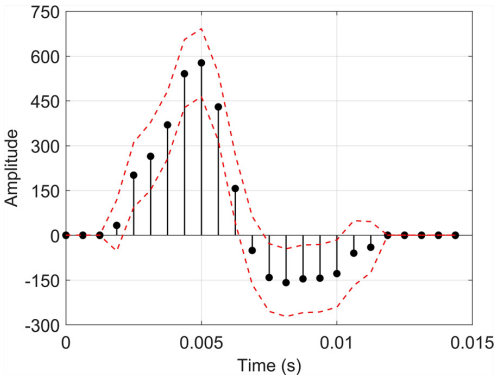


Fig. 2 Impulse response. Each discrete stem represents one coefficient h_k, upper and lower dot lines constitute the 95% confidence interval.

S. Guo, C. F. Silva, A. Ghani, and W. Polifke. Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis. In Journal of Engineering for Gas Turbines and Power, 2018. Reprinted with permission from ASME.

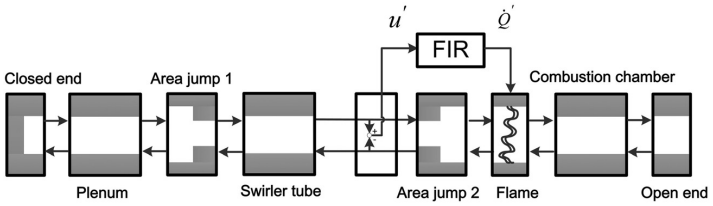


Fig. 3 Sketch of acoustic network model, flow from left to right

Table 1 Parameters of each acoustic element in the acoustic network model (Fig. 3)

Acoustic element	Parameters
Closed end	Reflection coefficient = 1
Plenum	Length = 0.17 m Sound speed = 343 m/s
Area jump 1	Area ratio = 29.76
Swirler tube	Length = 0.18 m Sound speed = 343 m/s
Area jump 2	Area ratio = 0.13
Flame	Relative temperature jump = 5.59 Ratio of specific impedances = 2.57
FIR	Impulse response model
Combustion chamber	$\text{Length}_A = 0.51$ m; $\text{Length}_B = 0.6$ m Sound speed = 880 m/s
Combustor exit	$\text{Reflection coefficient}_A = -0.9883$ $\text{Reflection coefficient}_B = -0.6351$

4.2 Data Bank Generation. We use Latin hypercube sampling to draw N samples $\widetilde{\boldsymbol{h}}^i, i = 1, 2, \ldots, N$, from an independent, standard, L-dimensional normal distribution. For each sample, we perform the inverse transformation to recover the natural inputs of the acoustic network model, then compute the corresponding growth rate. The purpose of this step is to generate a data bank for the subsequent gradient calculation.

4.3 Active Subspace Identification

(1) We use Latin hypercube sampling to draw M samples $\widetilde{(\boldsymbol{h}^i)'}, i = 1, 2, \ldots, M$, from independent, standard L-dimensional normal distribution. The guidance for determining a proper M value can be found in Ref. [14].

(2) For each sample $i = 1 \ldots M$, we retrieve the P nearest samples (according to Euclidean norm) and their corresponding growth rates in the data bank, and we build a local linear regression model

$$\text{Growth rate} \approx c + \sum_{j=0}^{L-1} b_j \widetilde{h}_j \tag{8}$$

where coefficients $\boldsymbol{b_i} = (b_0, \ldots, b_{L-1})_i^{\mathrm{T}}$ are taken as the gradients of growth rate against each of L coefficients at sample point i. This algorithm of calculating gradients is Algorithm 1.2 in Ref. [14].

(3) We construct the covariance matrix of the gradient vector and perform eigenvalue decomposition. Here, we use the superscript "ASA" to emphasize that the eigenvalues are obtained during the implementation of ASA, thus distinguishing them from the similar concepts of eigenvalues and eigenvectors involved in acoustic network calculation

$$\begin{gathered} \boldsymbol{C} = \frac{1}{M}\sum_{i}^{M} \boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{T}} = \boldsymbol{W}\Lambda^{\mathrm{ASA}}\boldsymbol{W}^{\mathrm{T}} \\ \Lambda^{\mathrm{ASA}} = \mathrm{diag}\left(\lambda_1{}^{\mathrm{ASA}}, \ldots, \lambda_L{}^{\mathrm{ASA}}\right) \\ \lambda_1{}^{\mathrm{ASA}} \geq \ldots \lambda_L{}^{\mathrm{ASA}} \end{gathered} \tag{9}$$

(4) We plot the eigenvalues $\lambda_1, \ldots \lambda_L$ on a log scale and look for "gaps." Here, gap refers to an order of magnitude decrease among adjacent eigenvalues. A prominent gap (as indicated in Fig. 8) in the eigenvalues indicates a separation between active and inactive subspaces, and we choose the dimension n of the active subspace to be the number of eigenvalues before the most prominent gap happens. Thus, we can make a corresponding partition of the eigenvector matrix, where $\boldsymbol{W}_1$ only contains n eigenvectors corresponding to the first n eigenvalues

$$\boldsymbol{W} = [\boldsymbol{W}_1 \quad \boldsymbol{W}_2] \tag{10}$$

The active variable vector $\boldsymbol{y}$ is defined as

$$\boldsymbol{y} = \boldsymbol{W}_1{}^{\mathrm{T}}\widetilde{\boldsymbol{h}} \tag{11}$$

Successively, entries in $\boldsymbol{y}$ represent the first active variable, the second active variable and so on. A total number of n active variables are obtained. Each active variable is expressed as a linear combination of the normalized input parameters, with the corresponding eigenvector (column vector of $\boldsymbol{W}_1$) being the linear combination coefficients. Statistically, each active variable follows independent standard normal distribution. When n is smaller than the number of input parameters, a model dimensionality reduction is achieved in the sense that subsequent surrogate model building will only need to be based upon these n active variables.

4.4 Surrogate Model Building. For each sample $i = 1 \ldots M$ in the data bank, there is a corresponding modal growth rate ω_i and a corresponding active variable vector $\boldsymbol{y}^i$. We select K representative samples from the data bank to pair the dataset $(\boldsymbol{y}^1, \omega_1) \ldots (\boldsymbol{y}^K, \omega_K)$, and we fit a regression model to express the growth rate ω as a function of $\boldsymbol{y}$.

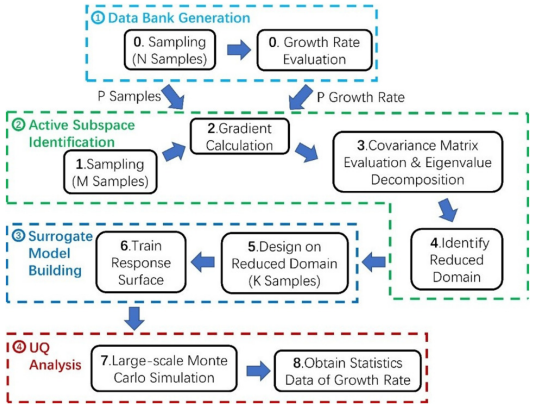


Fig. 4 Workflow of UQ based on active subspace approach

S. Guo, C. F. Silva, A. Ghani, and W. Polifke. Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis. In Journal of Engineering for Gas Turbines and Power, 2018. Reprinted with permission from ASME.

4.5 Uncertainty Quantification. We perform standard Monte Carlo simulation on surrogate model to achieve accelerated UQ analysis. We generate S samples $\widetilde{\boldsymbol{h}_{\mathrm{MC}}}^i, i = 1, 2, \ldots, S$, from independent, standard, L-dimensional normal distribution. For each sample, $\widetilde{\boldsymbol{h}_{\mathrm{MC}}}^k$ among $\widetilde{\boldsymbol{h}_{\mathrm{MC}}}^1 \ldots \widetilde{\boldsymbol{h}_{\mathrm{MC}}}^S$ calculates corresponding $\boldsymbol{y}_{\mathrm{MC}}{}^k$ as

$$\boldsymbol{y}_{\mathrm{MC}}{}^k = \boldsymbol{W}_1{}^{\mathrm{T}} \widetilde{\boldsymbol{h}_{\mathrm{MC}}}^k \tag{12}$$

then the corresponding growth rate $\omega_{\mathrm{MC}}{}^k$ can be directly computed through the surrogate regression model. Based on the obtained growth rate dataset $\omega_{\mathrm{MC}}{}^1 \ldots \omega_{\mathrm{MC}}{}^S$, a probability density function (PDF) regarding the growth rate can be constructed and relevant statistical indices can be extracted.

Figure 5 compares the ASA against DMC. To summarize, first, ASA invokes only a small number of acoustic solver calculations to identify a low-order structure within the original thermoacoustic system. Second, it replaces this expensive thermoacoustic system with a cheap algebraic surrogate model. Third, standard Monte Carlo simulations are applied only on the surrogate model with negligible cost through which a much faster UQ analysis is achieved. The most computational intensive step of ASA lays in Sec. 4.2, which requires a number of full-accuracy acoustic network calculations. For the terminology, even though Monte Carlo method is also adopted in the framework of ASA, we refer to the whole chain of analysis as ASA, while referring to the benchmark method as DMC.

5 Uncertainty Quantification Analysis

This section aims to answer the first question proposed in the introduction, i.e., to demonstrate that the ASA can indeed effectively achieve reduction of parameter dimensionality and significantly reduce the computational cost of UQ analysis, while maintaining high accuracy. Two cases with different values of combustion chamber length and reflection coefficient at combustor exit (cases A and B, as indicated in Table 1) are considered.

5.1 Thermoacoustic Mode Specification. Figure 6 shows the thermoacoustic modes up to 500 Hz for both cases, when the nominal coefficient values of the FIR model are used for acoustic network calculation. This analysis is a deterministic analysis, meaning no uncertainties are considered in the input parameters. Highly damped modes are also ignored. According to the convention adopted in the current investigation, modes with growth rates larger than zero are considered unstable.

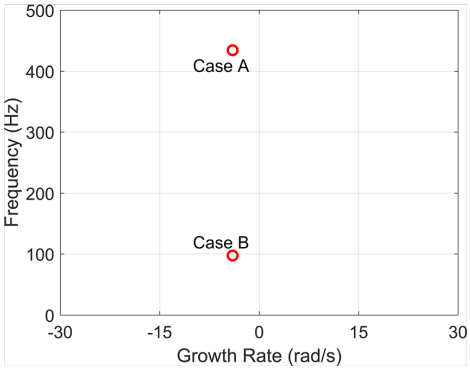


Fig. 6 Eigenmodes from deterministic analysis. For case A, the dominant mode is quarter wave mode [22], with a frequency of 434.2 Hz and a growth rate of −4 rad/s. For case B, the dominant mode is intrinsic mode [22], with a frequency of 97.5 Hz and a growth rate of −4 rad/s.

For both cases, the dominant modes are relatively close to the stability limit and may be calculated as unstable when uncertainties in FIR model coefficients are taken into account. Therefore, subsequent UQ analysis will only focus on these modes.

5.2 Direct Monte Carlo Results. Monte Carlo simulation is the classic method to propagate uncertainties from inputs to outputs. To implement the method in the present study, samples of FIR coefficients are drawn from the joint probability distribution characterized by the expectation value and the covariance matrix, which are obtained from the system identification. Each of these samples, seen as a FIR model perturbed around the model with nominal coefficient values, is subsequently fed into the acoustic network, and the corresponding perturbed modal growth rate is recorded. Based on the results of 20,000 times of acoustic network calculations, various statistical indices of the modal growth rate can be extracted.

Results of DMC are presented in Fig. 7 where contours of the joint probability density function of the modal frequency and growth rate are shown, which are constructed by 20,000 times of

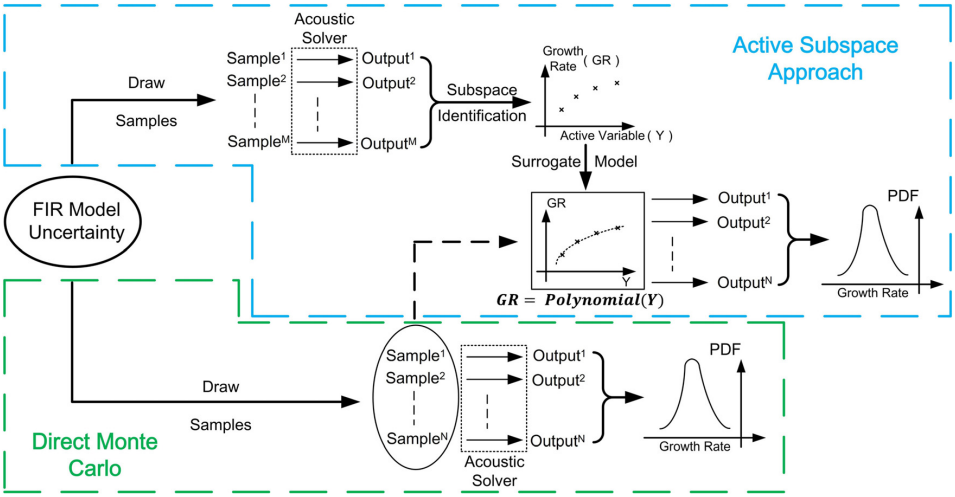


Fig. 5 Active subspace approach and DMC simulation

S. Guo, C. F. Silva, A. Ghani, and W. Polifke. Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis. In Journal of Engineering for Gas Turbines and Power, 2018. Reprinted with permission from ASME.

acoustic solver calculations. First, uncertainties in FIR model coefficients have significant impact on the predicted thermoacoustic stability in both cases, leading to 95% confidence intervals of (-4.03 ± 11.03) rad/s for case A and (-4.04 ± 7.15) rad/s for case B, respectively, in terms of the marginal distribution of the modal growth rate. Second, eigenmodes have a higher possibility to appear close to the nominal value, while lower probability is observed as eigenmodes move away from the center. Third, even though in both cases the modes have a growth rate of -4 rad/s according to the deterministic analysis, their standard deviations are different (5.63 rad/s for case A and 3.65 rad/s for case B) for the given amount of uncertainty. Those facts highlight the insufficiency of deterministic analysis.

5.3 Active Subspace Results. To accelerate the above UQ analysis, ASA is implemented for both cases with $N = 400$, $M = 110$, $P = 50$, $K = 5$, and $S = 20{,}000$, which, according to the convergence study, are considered to be sufficient.

Following the procedure outlined in Sec. 4, for each case, the covariance matrix of the gradient vector is constructed and eigenvalue decomposition is performed. Figure 8 shows the obtained eigenvalues in descending order. For both cases, since there exists a prominent gap between the first and the second eigenvalue, it is suggested that only the first eigenvector (corresponding to the largest eigenvalue in Figs. 8(*a*) and 8(*b*)), shown in Fig. 9, needs to be retained to form a single active variable, individually. Therefore, the dimensionality of the UQ problem shrinks from 16 (16 FIR model coefficients) to 1 (one active variable).

Two remarks worth mentioning here: (1) It is not a coincidence that both cases admit a one-dimensional structure. Our ongoing research has already mathematically proved that to first-order approximation the causal relationship between variations of FIR model coefficients and variations of model growth rate should collapse on a 1D subspace (i.e., a single active variable). This argument does not rely on the ASA implementation that the present manuscript relies on and is valid under the condition that the uncertainty level of the FIR model coefficients (represented as the ranges of coefficient confidence interval) is moderate. This condition is certainly fulfilled in the present case; for the FIR model investigated in our current paper, its uncertainty level can already be considered large (the maximum ratio of coefficient standard deviation to coefficient mean can be as large as 130%). These results of the physical interpretation of the active variable as well as the pertinent analytical derivations can be found in Ref. [12]. (2) The components of the first eigenvector in case B (Fig. 9(*b*)) are different from the counterparts in case A (Fig. 9(*a*)). This is expected for the following reason: according to Constantine [17], components of the eigenvector, which are also the coefficients of the linear combination that forms the active variable, reflect the sensitivity of the output (mode growth rate) against each input (individual FIR model coefficient h_k). Since two cases deal with two different combustor boundary conditions and thermoacoustic modes, the sensitivity of the modal growth rate against FIR model coefficients should be different, which explains the difference between Figs. 9(*a*) and 9(*b*).

For each sample generated in Sec. 4.2, on the one hand, we have already calculated its corresponding modal growth rate value; on the other hand, through Eq. (11), we can calculate its corresponding active variable value. By checking the relation between these two values for each sample, which is plotted in Fig. 10, it is possible to verify if a one-dimensional approximation exists. It can be seen clearly from Fig. 10 that a strong univariate trend is present and it confirms that ASA method does recover an accurate one-dimension approximation of the original system. In the present study, we construct a quadratic regression model to map from the active variable value to the modal growth rate value for each case and later large-scale Monte Carlo simulations are performed on these regression models.

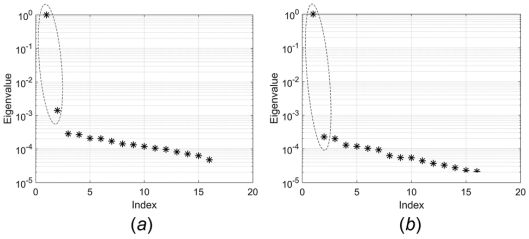


Fig. 8 Eigenvalues in λ^{ASA} in descending order. The prominent gap between the first and second eigenvalues indicates that a one-dimensional subspace exists: (*a*) case A and (*b*) case B.

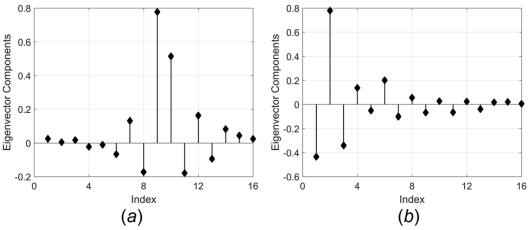


Fig. 9 Components of the first eigenvector in W_1, which will be used as the linear combination coefficients to form the single active variable: (*a*) case A and (*b*) case B

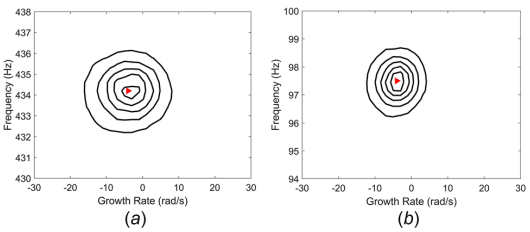


Fig. 7 Contour plot of the joint PDF of the modal growth rate and frequency. The contours (from outside to inside) correspond to 10%, 30%, 50%, 70%, and 90% of the maximum probability. The triangle is the deterministic solution (same as Fig. 6). Statistics regarding the marginal distribution of the modal growth rate are presented in Table 2: (*a*) case A and (*b*) case B.

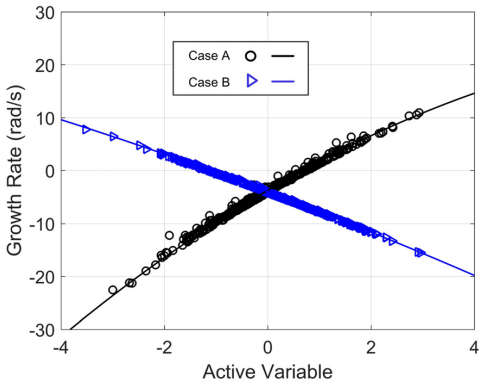


Fig. 10 Sufficient summary plot of the modal growth rate against active variable for each sample. We fit a quadratic function to link active variable and modal growth rate for each case.

S. Guo, C. F. Silva, A. Ghani, and W. Polifke. Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis. In Journal of Engineering for Gas Turbines and Power, 2018. Reprinted with permission from ASME.

Figure 11 compares the results between ASA and DMC. The PDF generated by the regression model matches well with the distribution given by the DMC, demonstrating again that one-dimensional structure identified by ASA can mimic the behaviors of the original high dimensional system with high accuracy.

5.4 Approach Assessment. In terms of computational cost, DMC requires 20,000 acoustic solver calculations for each case to obtain fully converged statistical indices. This relatively high computational cost is caused by the slow convergence of random sampling adopted in DMC. In contrast, the ASA replaces the original high dimensional system (16 parameters) with an algebraic one-dimensional model (one active variable), where subsequently UQ analysis will be applied on with negligible computational cost. As a result, only a total number of $N = 400$ acoustic solver calculations (for data bank generation) are needed for a converged and accurate UQ analysis, which significantly accelerates the analysis process.

Finally, statistical indices of modal growth rate distributions predicted by ASA are compared with DMC in Table 2. Here, RF stands for risk factor, which is defined as the probability that a mode is unstable [4] and its expression is shown in Eq. (13). To summarize, it can be concluded that ASA can indeed significantly reduce the computational cost of UQ analysis while maintaining high accuracy, which confirms the results of Bauerheim et al. [6]

$$\mathrm{RF}(\%) = 100 \int_0^\infty \mathrm{PDF}(\omega) d\omega \tag{13}$$

It is worth mentioning that a rather simple yet practical acoustic network model is adopted in the current study, where DMC is entirely feasible to obtain the PDF of the modal growth rate. For more computational intensive acoustic models, like the ones characterized by the Helmholtz equation or the Linearized Navier–Stokes equation, DMC would no longer be an option, then ASA really pays off.

6 Optimum Computational Fluid Dynamics Time Length Estimation

This section aims to answer the second question proposed in the introduction, i.e., to find a practical procedure to estimate the length of CFD time series required for FIR model identification, so that final uncertainties in modal growth rate are within a desired range. In this section, first, motivation for seeking this procedure and feasible solutions are discussed; second, a case study is presented to demonstrate the argument; and finally, steps for implementing the procedure are summarized.

6.1 Motivation and Solutions. Generally speaking, identification of FIR model from longer CFD time series yields smaller variance of model coefficients, thus leading to a less uncertain estimation of the modal growth rate. From Sec. 5, we can see that the FIR model identified from 350 ms's LES time series contains

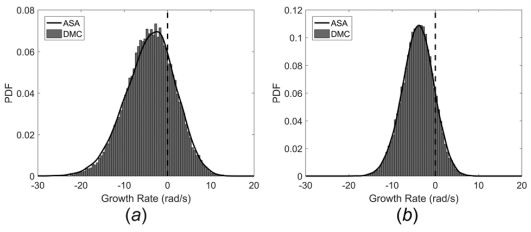


Fig. 11 Probability density function of thermoacoustic growth rate of dominant mode produced by ASA and DMC: (*a*) case A and (*b*) case B

Table 2 Statistical indices comparison: mean, standard deviation, and RF

	Case A		Case B	
	DMC	ASA	DMC	ASA
Mean	−4.03	−4.04	−4.04	−4.03
StDev	5.63	5.78	3.65	3.69
RF	24.1%	25.1%	13.3%	13.7%

relatively large uncertainty, leading to a 24% of risk factor of the modal growth rate for case A. Therefore, to obtain a more robust estimation of the modal growth rate, longer LES time series may be necessary. Then, exactly how much longer time series data are required so that a satisfactory reduction of the uncertainty of the modal growth rate estimation can be achieved? Is it possible to propose a procedure to estimate this time length?

A feasible procedure to achieve this is illustrated in Fig. 12: for an ongoing unsteady CFD run, whenever the simulation has progressed a fixed period of time, estimate FIR model coefficients as well as their uncertainties, employ ASA to evaluate uncertainty in the modal growth rate, terminate this process when the uncertainty fulfills accuracy requirement.

Compared with performing DMC every time, obviously the abovementioned procedure can significantly reduce the computational cost. However, the identification of low-dimensional structure and construct surrogate model from time to time still persist as tedious work. Then, is it possible to further improve the procedure to make it more efficient?

As a matter of fact, we argue that as long as the number of FIR model coefficients is kept constant, it is reasonable to assume that the surrogate model, which is obtained by applying ASA on the FIR model identified from t_0 length of time series (as indicated in Fig. 12), can be used without modification for subsequent UQ calculations when longer CFD time series are available. Therefore, only once ASA implementation is needed for the whole process. The rationality of this assumption lays in the fact that this surrogate model quantitatively describes the physical relations between the FIR model coefficients (varied within the individual confidence interval) and the modal growth rate. As longer time series ($t_0 +$ iterations Δt) is put into use, the newly obtained confidence

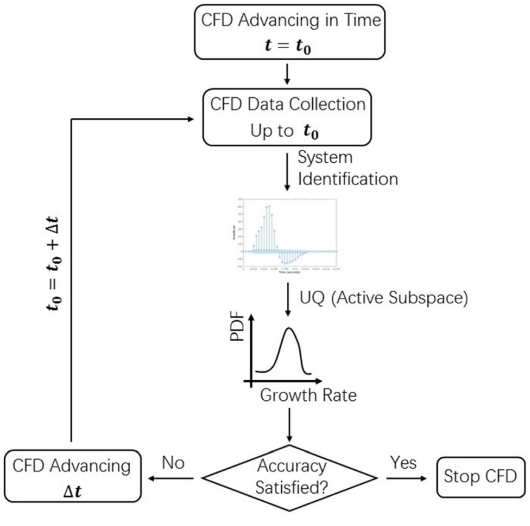


Fig. 12 A feasible workflow for estimating appropriate CFD simulation time to achieve predefined confidence requirements

S. Guo, C. F. Silva, A. Ghani, and W. Polifke. Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis. In Journal of Engineering for Gas Turbines and Power, 2018. Reprinted with permission from ASME.

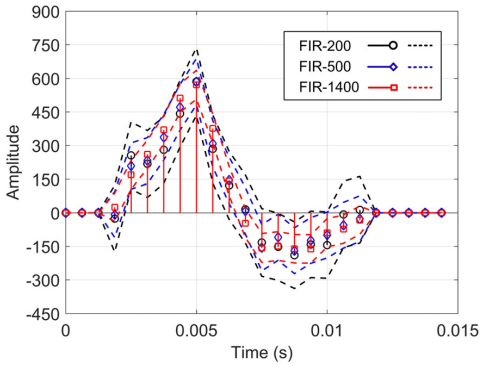


Fig. 13 Finite impulse response models identified from time series of different length. Here, confidence intervals (represented by ±3 standard deviations) of FIR model coefficients become narrower as length of time series increases.

intervals will become narrower and covered by the previous confidence intervals based on t_0 length of time series. Automatically, this same surrogate model remains valid, thus, eliminating the need for further ASA implementation.

6.2 Case Study. Case A in Sec. 5 is restudied here, but with three versions of 16-coefficient FIR model identified using different lengths of time series data (as shown in Fig. 13). Here, instead of actual LES data, synthetic time series are employed, which allows a longer time length for better illustration. To generate the synthetic time series, the FIR model in Fig. 2 is used as the reference FIR model, and a broadband velocity signal u' is applied to excite the FIR model to obtain the heat release perturbation. A Gaussian distributed white noise is added afterward to finally generate the corresponding synthetic heat release signal Q'. A similar approach to generate synthetic data for system identification is employed by Jaensch et al. [24].

The workflow for this case study is outlined in Fig. 14, where ASA will only be implemented once, on the impulse response model ("FIR-200") identified from 200 ms of synthetic time series. The goal is to prove that the derived surrogate model can also be used for UQ evaluation when longer time series are considered, thus making it possible for performing UQ analysis on the fly during an on-going CFD run. The obtained two PDFs of modal growth rate in Fig. 14 will be compared with corresponding DMC results to assess the accuracy.

The first step is to apply ASA on "FIR-200" model. Here, we assume that the FIR-200 model coefficients are independent and only the diagonal terms of the covariance matrix of FIR-200 are used for generating samples. The rationality of this treatment is the following: we are aiming to construct a surrogate model, which, in the end, forms a direct mapping from the values of h_i's in $\boldsymbol{h}$, to the value of the modal growth rate, i.e., Growth Rate $=$ surrogate$(\boldsymbol{h})$ $=$ surrogate$(h_0, h_1, \ldots, h_{15})$, and remains valid in the domain determined by the confidence interval of each FIR-200 model coefficient. To achieve this goal, we need to generate samples of $\boldsymbol{h}^i, i = 1, \ldots, N$ that can cover this domain, and the information contained in the diagonal terms of the covariance matrix is enough to guide us in generating representative samples. The benefit of this treatment is that the mathematical manipulation can be significantly simplified where no orthogonal transformation (Sec. 4.1.1) is required.

Figures 15 and 16 demonstrate the corresponding results of ASA. As expected, a one-dimensional structure is identified and only the first eigenvector needs to be retained to form a single active variable. Figure 16 plots the modal growth rate against the active variable, which further confirms that modal growth rate can be approximated as a univariate function of the active variable. We notice that the eigenvector shown in Fig. 15(*b*) is different from Fig. 9(*a*). This is because, in Fig. 15(*b*), the eigenvector is with respect to the original h_i's, while in Fig. 9(*a*) the eigenvector is with respect to the transformed $\widetilde{h}_i$'s.

The second step is to obtain the corresponding surrogate model, which is expressed as

$$Y = a_1\left(\frac{h_1 - \bar{h}_1}{\sigma_1}\right) + a_2\left(\frac{h_2 - \bar{h}_2}{\sigma_2}\right) + \cdots + a_{16}\left(\frac{h_{16} - \bar{h}_{16}}{\sigma_{16}}\right) \tag{14}$$

$$\text{Growth rate} = \beta_0 + \beta_1 Y + \beta_2 Y^2 \tag{15}$$

where Y is the active variable and a_i's are the entries in the eigenvector (shown in Fig. 15(*b*)). h_i's represent FIR model coefficients, and $\bar{h}_i$'s and σ_i's are the nominal values and standard deviations of FIR-200 model coefficients, resectively. β_i's denote the polynomial coefficients of the quadratic regression model we fit.

The third step is to use the obtained surrogate model to perform UQ analysis against "FIR-500" model and "FIR-1400" model. For both models, full covariance matrices are considered. At this point, it is worth mentioning that since the surrogate model forms

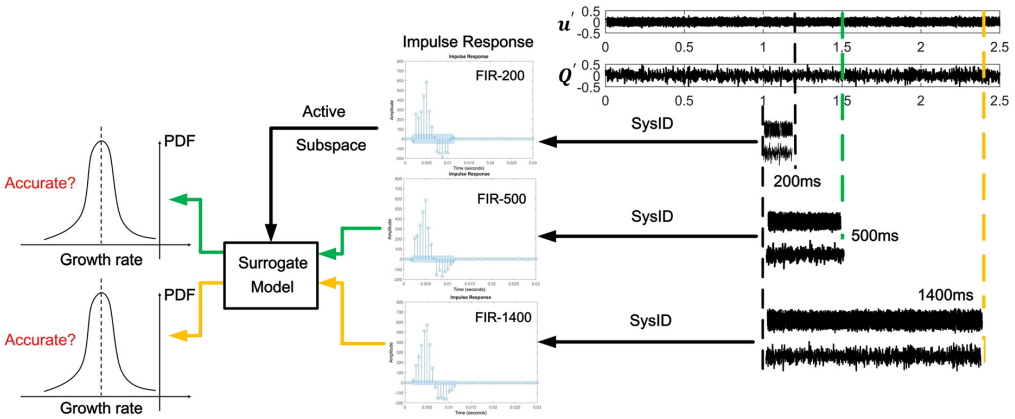


Fig. 14 200 ms, 500 ms, and 1400 ms of synthetic series are used to identify "FIR-200," "FIR-500," and "FIR-1400" model, respectively. Active subspace approach will only be implemented once on "FIR-200" to derive the SM.

S. Guo, C. F. Silva, A. Ghani, and W. Polifke. Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis. In Journal of Engineering for Gas Turbines and Power, 2018. Reprinted with permission from ASME.

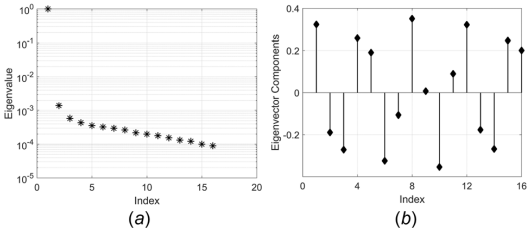


Fig. 15 (*a*) Eigenvalues in descending order and (*b*) components of the first eigenvector

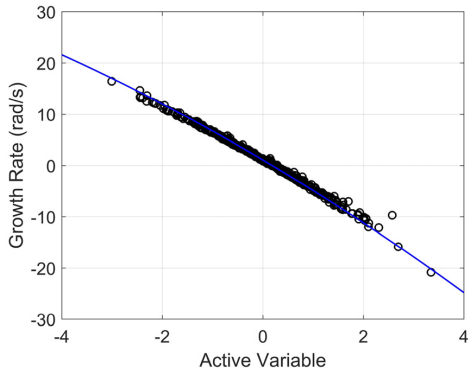


Fig. 16 Sufficient summary plot of the modal growth rate against active variable. A quadratic regression model is fitted to describe the relation between active variable and modal growth rate.

a direct link from $\boldsymbol{h}$ to modal growth rate, this surrogate model would not concern about how exactly the samples $\boldsymbol{h}^i$ are generated. As long as $\boldsymbol{h}$ with suitable values of h_i's (suitable means values of h_i's are within the confidence intervals of FIR-200 model coefficients) is provided, the calculation of modal growth rate is straightforward. Therefore, in practice, we generate samples using the full covariance matrices of FIR-500 and FIR-1400 models, and we fed those samples directly into the obtained surrogate model (Eqs. (14) and (15)) to calculate the corresponding PDF of modal growth rate. We emphasize that it is not necessary to implement ASA anymore.

Results comparison are summarized in Fig. 17 and Table 3. It can be seen that the generalized surrogate model built upon the uncertainty information of FIR-200 model is also able to achieve successful UQ analysis when new uncertainty information is provided, thus confirming our previous arguments.

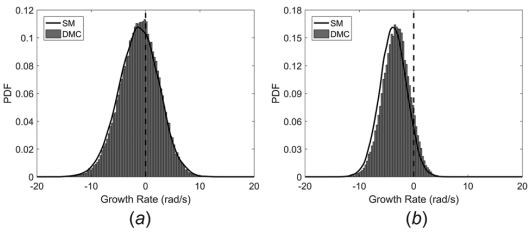


Fig. 17 Comparison of PDF results produced by SM and DME, considering the uncertainty information of (*a*) "FIR-500" model and (*b*) "FIR-1400" model

Table 3 Statistical indices comparison: mean, standard deviation, and RF

	FIR-500		FIR-1400	
	DMC	SM	DMC	SM
Mean	−1.08	−1.25	−3.23	−3.76
Std	3.57	3.68	2.48	2.45
RF	38.8%	37.3%	9.5%	6.2%

Procedure: Estimating Optimal CFD Time Length for FIR Identification

Unsteady CFD finished to t_0
Set $t_{current} = t_0$
Set *Flag* equals to **0**
While (*Flag* equals **0**)
  Data = CFD (0: $t_{current}$) # Prepare available CFD time series
  [*FIR.nominal* , *FIR.uncertainty*] = System Identification (*Data*)
  if ($t_{current}$ equals t_0)
    # Construct generalized surrogate model, only perform once
    Surrogate Model = Active Subspace (*FIR.nominal* , *FIR.uncertainty*)
    # Employ generalized surrogate model to perform fast UQ analysis
    Growth Rate PDF = *Surrogate Model*(*FIR.nominal* , *FIR.uncertainty*)
  else
    Growth Rate PDF = *Surrogate Model*(*FIR.nominal* , *FIR.uncertainty*)
  end
  if (*Growth Rate PDF* satisfies **Pre-defined Goal**)
    Flag = **1**
  else
    CFD Advancing Δt
    $t_{current} = t_{current} + \Delta t$
  end
end

Fig. 18 The proposed procedure for estimating CFD time length for FIR identification

6.3 Procedure Implementation. Figure 18 provides the pseudocode for procedure implementation. A good starting point for t_0, which is the minimum CFD time series required before executing the procedure, would be at least ten times the length of the impulse response [25].

7 Conclusions

Based on active subspace approach, UQ analyses regarding the impact of FIR model uncertainties on thermoacoustic instabilities are performed in the present paper. Answers for two practical questions are provided: First, "How to efficiently propagate uncertainties from a FIR model for the flame dynamics to the modal growth rate of the system?," the current research indicates that active subspace approach can achieve highly efficient and accurate UQ analysis by constructing and exploiting a one-dimensional surrogate model. Compared with direct Monte Carlo, 50 times faster UQ analysis was recorded with the cases considered in the paper. For the second question, i.e., "Which length of CFD time series is required for a desired confidence of system identification?," the current research further explored the potential of active subspace approach and proposed a ready-to-implement procedure in pseudocode form. A case study was performed which confirmed the effectiveness of the procedure. It can also be seen as a demonstration of the procedure workflow.

Compared with the previous achievements in the field of UQ analysis of thermoacoustic instabilities, the novelties of the current work reflect in: (1) to our best knowledge, this is the first time that the impact of uncertainties in flame FIR model, which

S. Guo, C. F. Silva, A. Ghani, and W. Polifke. Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis. In Journal of Engineering for Gas Turbines and Power, 2018. Reprinted with permission from ASME.

represents a sophisticated and realistic time-domain model of flame dynamics, on thermoacoustic stability prediction has been assessed; (2) we have gone one step further by proposing an ASA-based procedure to effectively determine the length of CFD simulation that is required to achieve FIR model identification with a desired level of uncertainty. This should be highly relevant for practitioners due to the high computational cost of the CFD simulation; (3) the presented nonintrusive ASA methodology can also be applied to more complex acoustic models (thus handling more complex combustors), where classic UQ methods would be prohibitively expensive.

Further study will include more sources of uncertainty, e.g., boundary conditions and model parameters, and investigate their combined effects in thermoacoustic instability prediction [26].

Acknowledgment

S. Guo is grateful for the financial support from doctoral scholarship of Chinese Scholarship Council. W. Polifke and C. Silva are grateful to the 2014 Center for Turbulence Research Summer Program (Stanford University), where discussions with Michael Bauerheim and Franck Nicoud instigated the ideas developed in this study.

Funding Data

- China Scholarship Council (No. 201606830045).

Nomenclature

ASA = active subspace approach
DMC = direct Monte Carlo
FIR = finite impulse response
$\boldsymbol{h}$ = FIR model coefficients in vector form
$\tilde{\boldsymbol{h}}$ = transformed FIR model coefficients in vector form
h_i = FIR model coefficient
K = number of samples in ASA for surrogate model building
L = FIR model order
M = number of samples in ASA for active subspace identification
N = number of samples in ASA for data bank generation
P = number of samples in ASA for gradient calculation
PDF = probability density function
$\dot{Q}'$ = heat release rate fluctuation
RF = risk factor
S = number of samples in ASA for Monte Carlo simulation
u = input for system identification
UQ = uncertainty quantification
u' = velocity fluctuation
$\boldsymbol{W}_1$ = matrix with column vectors being the eigenvectors that form the active variables
y = output for system identification
$\boldsymbol{y}$ = active variable vector

References

[1] Nicoud, F., Benoit, L., Sensiau, C., and Poinsot, T., 2007, "Acoustic Modes in Combustors With Complex Impedances and Multidimensional Active Flames," AIAA J., **45**(2), pp. 426–441.
[2] Silva, C. F., Emmert, T., Jaensch, S., and Polifke, W., 2015, "Numerical Study on Intrinsic Thermoacoustic Instability of a Laminar Premixed Flame," Combust. Flame, **162**(9), pp. 3370–3378.
[3] Magri, L., and Juniper, M. P., 2013, "Sensitivity Analysis of a Time-Delayed Thermo-Acoustic System Via an Adjoint-Based Approach," J. Fluid Mech., **719**, pp. 183–202.
[4] Ndiaye, A., Bauerheim, M., and Nicoud, F., 2015, "Uncertainty Quantification of Thermoacoustic Instabilities on a Swirled Stabilized Combustor," ASME Paper No. GT2015-44133.
[5] Tay-Wo-Chong, L., Bomberg, S., Ulhaq, A., Komarek, T., and Polifke, W., 2012, "Comparative Validation Study on Identification of Premixed Flame Transfer Function," ASME J. Eng. Gas Turbines Power, **134**(2), p. 021502.
[6] Bauerheim, M., Ndiaye, A., Constantine, P., Moreau, S., and Nicoud, F., 2016, "Symmetry Breaking of Azimuthal Thermoacoustic Modes: The UQ Perspective," J. Fluid Mech., **789**, pp. 534–566.
[7] Magri, L., Bauerheim, M., Nicoud, F., and Juniper, M. P., 2016, "Stability Analysis of Thermo-Acoustic Nonlinear Eigenproblems in Annular Combustors—Part II: Uncertainty Quantification," J. Comput. Phys., **325**, pp. 411–421.
[8] Silva, C., Magri, L., Runte, T., and Polifke, W., 2017, "Uncertainty Quantification of Growth Rates of Thermoacoustic Instability by an Adjoint Helmholtz Solver," ASME J. Eng. Gas Turbines Power, **139**(1), p. 011901.
[9] Blumenthal, R. S., Subramanian, P., Sujith, R., and Polifke, W., 2013, "Novel Perspectives on the Dynamics of Premixed Flames," Combust. Flame, **160**(7), pp. 1215–1224.
[10] Polifke, W., 2014, "Black-Box System Identification for Reduced Order Model Construction," Ann. Nucl. Energy, **67**, pp. 109–128.
[11] Sovardi, C., Jaensch, S., and Polifke, W., 2016, "Concurrent Identification of Aero-Acoustic Scattering and Noise Sources at a Flow Duct Singularity in Low Mach Number Flow," J. Sound Vib., **377**, pp. 90–105.
[12] Guo, S., Silva, C. F., Bauerheim, M., Ghani, A., and Polifke, W., 2018, "Evaluating the Impact of Uncertainty in Flame Impulse Response Model on Thermoacoustic Instability Prediction: A Dimensionality Reduction Approach," Proc. Combust. Inst. (in press).
[13] Constantine, P., Dow, E., and Wang, Q., 2014, "Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces," SIAM J. Sci. Comput., **36**(4), pp. A1500–A1524.
[14] Constantine, P. G., 2015, *Active Subspaces: Emerging Ideas in Dimension Reduction for Parameter Studies*, Vol. 2, SIAM, Philadelphia, PA.
[15] Lukaczyk, T. W., Constantine, P. G., Palacios, F., and Alonso, J. J., 2014, "Active Subspaces for Shape Optimization," AIAA Paper No. 2014-1171.
[16] Jefferson, J. L., Gilbert, J. M., Constantine, P. G., and Maxwell, R. M., 2015, "Active Subspaces for Sensitivity Analysis and Dimension Reduction of an Integrated Hydrologic Model," Comput. Geosci., **83**, pp. 127–138.
[17] Constantine, P. G., and Diaz, P., 2017, "Global Sensitivity Metrics From Active Subspaces," Reliab. Eng. Syst. Saf., **162**, pp. 1–13.
[18] Bodén, H., and Polifke, W., 2015, "Uncertainty Quantification Applied to Aeroacoustic Predictions," *Progress in Simulation, Control and Reduction of Ventilation Noise* (VKI Lecture Series 2015), C. Schram, ed., von Karman Institute for Fluid Dynamics, Sint-Genesius-Rode, Belgium.
[19] Keesman, K. J., 2011, "Time-Invariant System Identification," *System Identification* (Advanced Textbooks in Control and Signal Processing), Springer, London, pp. 59–167.
[20] Tay-Wo-Chong, L., Komarek, T., Kaess, R., Föller, S., and Polifke, W., 2010, "Identification of Flame Transfer Functions From LES of a Premixed Swirl Burner," ASME Paper No. GT2010-22769.
[21] Komarek, T., and Polifke, W., 2010, "Impact of Swirl Fluctuations on the Flame Response of a Perfectly Premixed Swirl Burner," ASME J. Eng. Gas Turbines Power, **132**(6), p. 061503.
[22] Emmert, T., Bomberg, S., Jaensch, S., and Polifke, W., 2017, "Acoustic and Intrinsic Thermoacoustic Modes of a Premixed Combustor," Proc. Combust. Inst., **36**(3), pp. 3835–3842.
[23] Cowan, G., 1998, *Statistical Data Analysis*, 1st ed., Clarendon Press, Gloucestershire, UK.
[24] Jaensch, S., Merk, M., Emmert, T., and Polifke, W., 2018, "Identification of Flame Transfer Functions in the Presence of Intrinsic Thermoacoustic Feedback and Noise," Combust. Theory Modell., **22**(3), pp. 613–634.
[25] Tangirala, A. K., 2014, *Principles of System Identification: Theory and Practice*, CRC Press, Boca Raton, FL.
[26] Avdonin, A., and Polifke, W., 2018, "Quantification of the Impact of Uncertainties in Operating Conditions on the Flame Transfer Function With Non-Intrusive Polynomial Chaos Expansion," ASME Paper No. GT2018-75476.

S. Guo, C. F. Silva, A. Ghani, and W. Polifke. Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis. In Journal of Engineering for Gas Turbines and Power, 2018. Reprinted with permission from ASME.

Available online at www.sciencedirect.com

ScienceDirect

Proceedings of the Combustion Institute 37 (2019) 5299–5306

Proceedings of the Combustion Institute

www.elsevier.com/locate/proci

Evaluating the impact of uncertainty in flame impulse response model on thermoacoustic instability prediction: A dimensionality reduction approach

Shuai Guo [a,*], Camilo F. Silva [a], Michael Bauerheim [b], Abdulla Ghani [a], Wolfgang Polifke [a]

[a] *Fakultät für Maschinenwesen, Technische Universität München, Boltzmannstr 15, Garching D-85748, Germany*
[b] *Institut Supérieur de l'Aéronautique et de l'Espace, 10 avenue Edouard Belin-BP 54032, Toulouse CEDEX 4 31055, France*

Received 30 November 2017; accepted 7 July 2018
Available online 27 July 2018

Abstract

The flame response to upstream velocity perturbations is properly described by a Finite Impulse Response (FIR) model. When combining an FIR model with acoustic tools to predict thermoacoustic modal growth rates, uncertainties contained in the FIR model coefficients would propagate through the acoustic model, inducing deviations of the modal growth rate from its nominal value. Therefore, an associated uncertainty quantification (UQ) analysis, which focuses on quantifying the impact of FIR model uncertainties on the modal growth rate prediction, is a necessity to obtain a more reliable thermoacoustic instability prediction. To address this UQ problem, our present work proposes an analytical strategy featuring (1) compactly summarizing the causal relationship between variations of FIR model coefficients and variations of modal growth rates; (2) Effectively shrinking the dimension of the UQ problem; (3) Requiring only negligible computational cost; (4) Involving no complex mathematical treatments. Our case studies yielded 5000 times faster yet highly accurate UQ analyses compared with reference Monte Carlo simulations, even though a significant level of FIR model uncertainty is present. The analytical approach brings additional benefits including (1) visualization of the process from the variations of FIR model coefficients to the variations of modal growth rate; (2) Easily-obtainable sensitivity measurement for each FIR model coefficient, which can help identify key mechanisms controlling the thermoacoustic instability; (3) New possibility for robust combustor design, i.e., to minimize the impact of FIR model uncertainty on the thermoacoustic instability prediction.

Keywords: Uncertainty quantification; Flame impulse response model; Thermoacoustic instability; Dimensionality reduction

1. Introduction

Combining acoustic tools (e.g., Helmholtz solvers [1] or network models [2]) with a flame

* Corresponding author.
E-mail address: guo@tfd.mw.tum.de (S. Guo).

https://doi.org/10.1016/j.proci.2018.07.020
1540-7489

response model is a popular approach to predict thermoacoustic instability. In this framework, the flame response model, which may be derived from experiment or numerical simulation, constitutes a source of uncertainty, which may have significant impact on the reliability of modal growth rate calculation, as is evidenced by the work of Nair et al. [3], Ndiaye et al. [4], Bauerheim et al. [5], Magri et al. [6] as well as Silva et al. [7]. In these studies, 2-coefficient $n-\tau$ models [8] were investigated, and the gain n and the time delay τ of the flame response were considered as uncertain input parameters.

Compared with a frequency-independent $n-\tau$ model, the finite impulse response (FIR) model represents a more sophisticated and realistic flame model, which describes flame dynamics in the time domain and facilitates direct physical interpretation of flow-flame interaction mechanisms [9]. In fact, the $n-\tau$ model can be viewed as a special case of an FIR model with only one non-zero coefficient. FIR models can be deduced from experimentally measured flame frequency response data [10] or determined through a combined CFD/System Identification procedure [11]. To further pave the way for using this advanced flame response model, it is essential to quantify the impact of uncertainties in the FIR model and obtain the associated error in predicting thermoacoustic instability. Uncertainty quantification (UQ) analysis, which focuses on propagating uncertainties from inputs to outputs, is required for that purpose.

Monte Carlo simulation [12] is a classic method for conducting UQ analysis. However, due to its slow convergence, a large number of samples (generally in the order of thousands) have to be drawn from the distribution of inputs and an equally large number of model evaluations are required to construct a converged probability density function (PDF) of the output. Therefore, previous studies investigating the impact of uncertain flame model have employed various sophisticated surrogate techniques [4–7,12], so that a smaller number of input samples and corresponding model evaluations is sufficient for UQ analysis.

In this work we propose a dimensionality reduction strategy based on analytical analysis to address our current UQ problem, i.e., evaluate the impact of FIR model uncertainties on the thermoacoustic modal growth rate calculation, so that we can (1) perform UQ analysis analytically to improve the efficiency, while avoiding sophisticated mathematical treatments as much as possible; (2) obtain further physical insights regarding the causal relationship between FIR model coefficient variations and modal growth rate variations. This paper starts with a statement of our current UQ problem, followed by introducing the way we visualize the results. Then we derive the dimensionality reduction strategy and demonstrate its effectiveness through case studies. We close the paper by pointing out further applications of the proposed UQ strategy.

2. UQ problem setting

In the present study, we investigate the uncertainty regarding the growth rate of a marginally stable thermoacoustic mode in a combustor computed with an acoustic network solver. An FIR model is introduced to describe the flame response, which links velocity fluctuations u'_u upstream of the flame to the global heat release rate fluctuations $\dot{Q}'$ in the following manner:

$$\frac{\dot{Q}'_n}{\bar{\dot{Q}}} = \frac{1}{\bar{u}_u}\sum_{k=0}^{L-1} h_k u'_{u,n-k} \tag{1}$$

where h_k's are the FIR model coefficients, which are considered uncertain. L is the model order (i.e. number of coefficients) and n is the time step. The overbar and the prime indicate average and fluctuating values, respectively.

The uncertainty of the FIR model coefficients stems from their estimation process, either through experimental measurements, which inevitably contains noise, or through a combined approach of CFD/System-identification [11], where (1) CFD simulation can be uncertain, e.g., boundary conditions, combustion model parameters, etc. and (2) identification process will be uncertain, e.g., due to its stochastic nature, low signal-to-noise ratio, finite length of CFD time series [13], etc. As a result, the estimation results for h_k's containing not only nominal values, but also an associated covariance matrix which describes the uncertainty information of h_k's. In the present work, we assume that the FIR model uncertainty is known.

Conventional thermoacoustic instability prediction only employs the nominal values h_k^0's (nominal quantities are denoted by superscript "0", the same hereinafter) and calculates the corresponding eigenvalues of the thermoacoustic system $(\boldsymbol{i}\omega^0+\sigma^0)\in\mathbb{C}$, *i.e.,* modal frequency $\omega^0\in\mathbb{R}$ and growth rate $\sigma^0\in\mathbb{R}$. This analysis is also referred to as deterministic analysis. However, conducting only deterministic analysis can be dangerous for marginally stable thermoacoustic modes [4–7], since they can become unstable when h_k's deviate from their nominal values.

In contrast to the deterministic analysis, UQ analysis takes into account variations of FIR model coefficients and propagates their uncertainties to the output, in particular, the modal growth rate σ. Therefore, the output is not just a single value σ^0, but a PDF, which describes the output in a statistical manner.

By means of the Monte Carlo method, a large number of samples of FIR model coefficients have to be drawn from the distribution of coefficients, and for each sample the corresponding modal

With permission from S. Guo, C. F. Silva, M. Bauerheim, A. Ghani and W. Polifke. Evaluating the impact of uncertainty in flame impulse response model on thermoacoustic instability prediction: A dimensionality reduction approach. Proceedings of the Combustion Institute, Elsevier, 2018.

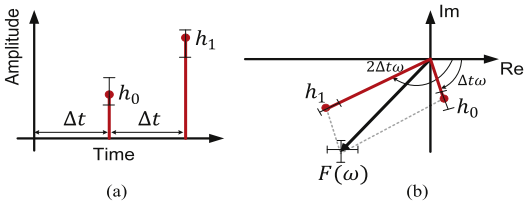


Fig. 1. The sketch of (a) a simplified FIR model and (b) the phasor plot of FRF.

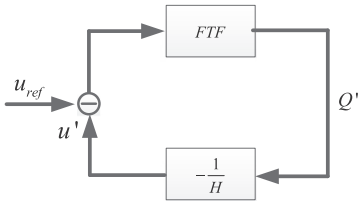


Fig. 2. The thermoacoustic closed-loop network.

growth rate has to be calculated. Clearly, direct Monte Carlo requires a large number of model evaluations, thus making it a possibly very expensive method.

3. Results visualization - phasor plot of FRF

Before addressing our current UQ problem, firstly, we need to introduce the phasor plot of the flame frequency response function (FRF), which lays the foundation for illustrating the results in the reminder of this paper.

The FRF describes the flame dynamics in frequency domain and is obtained from the general flame transfer function (FTF) in complex domain by setting the growth rate σ in FTF to zero. FRF can be obtained from the FIR model as in Eq. (2) [11], where the h_k's are the corresponding FIR model coefficients, $\omega \in \mathbb{R}$ denotes the frequency and Δt represents the sampling interval between successive FIR model coefficients.

$$\mathrm{F}(\omega) = \sum_{k=0}^{L-1} h_k e^{-\boldsymbol{i}(k+1)\Delta t\omega}, \quad \omega \in \mathbb{R} \tag{2}$$

For illustration, we consider a simplified FIR model in Fig. 1a, with only two coefficients (h_0 and h_1) as well as their uncertainty bounds. In this case, when F is evaluated at a fixed nominal frequency ω^0, Eq. (2) is written as:

$$\mathrm{F}(\omega^0) = h_0 e^{-\boldsymbol{i}\Delta t\omega^0} + h_1 e^{-\boldsymbol{i}(2\Delta t)\omega^0} \tag{3}$$

Therefore we can draw $\mathrm{F}(\omega^0)$ in Eq. (3) in Fig. 1b, which we refer to as the phasor plot of FRF. Since the values of h_k's are uncertain, any variations of the coefficient h_k will cause change of length of the respective phasor. Consequently, the length and direction of the $\mathrm{F}(\omega^0)$ phasor will change too.

We use the term *sample* to refer to one particular combination of h_k's, i.e., sample $\boldsymbol{h} = \{h_0, h_1, \ldots, h_{L-1}\}$. We are particularly interested in (1) sets of samples, that produce the same modal growth rate, and (2) the head locations of their corresponding F phasor evaluated at the fixed nominal frequency ω^0. In other words, we focus on the distribution pattern of the iso-growth lines in the phasor plot of FRF.

4. Analytical UQ procedure derivation

In this section, we demonstrate the analytical derivation and prove that to first-order, the isolines of growth rate of thermoacoustic instability are parallel straight lines in the phasor plot of FRF. Then, we discuss why this finding is crucial for addressing the current UQ problem setting in an analytical way. Finally, we summarize the analytical procedure.

4.1. Iso-growthlines in the phasor plot of FRF

The thermoacoustic system is a closed-loop network. As shown in Fig. 2, we choose the velocity at a reference location just upstream of the flame as the input and global heat release fluctuation as the output. The block *FTF* describes the link between velocity fluctuation and global heat release fluctuation, while the block $-1/H$ describes the acoustic response of the combustor to the global heat release fluctuation. Consequently, we can derive the characteristic equation of the system [14]:

$$FTF(\omega - \boldsymbol{i}\sigma) = H(\omega - \boldsymbol{i}\sigma) \quad \omega, \sigma \in \mathbb{R} \tag{4}$$

Since we are using an FIR model, Eq. (4) can be written as:

$$\sum_{k=0}^{L-1} h_k e^{-\boldsymbol{i}(k+1)\Delta t(\omega - \boldsymbol{i}\sigma)} = H(\omega - \boldsymbol{i}\sigma) \tag{5}$$

We argue that terms $e^{-(k+1)\Delta t\sigma}, k = 0...(L-1)$ can be approximated as 1 since: (1) $L\Delta t$, the flame response time to the velocity perturbation, is normally at the order of milliseconds; (2) we consider marginally stable thermoacoustic modes, which means the growth rates σ is close to zero. Therefore Eq. (5) may be approximated as:

$$\sum_{k=0}^{L-1} h_k e^{-\boldsymbol{i}(k+1)\Delta t\omega} = H(\omega - \boldsymbol{i}\sigma) \tag{6}$$

For any perturbed sample $\boldsymbol{h^*} = \boldsymbol{h^0} + \Delta\boldsymbol{h}$, we have the corresponding modal frequency $\omega^* = \omega^0 + \Delta\omega$ and growth rate $\sigma^* = \sigma^0 + \Delta\sigma$. We substitute $\boldsymbol{h^*}$, ω^* and σ^* into Eq. (6), perform first-order Taylor expansion around the nominal on both sides of the equation, and then combine like

With permission from S. Guo, C. F. Silva, M. Bauerheim, A. Ghani and W. Polifke. Evaluating the impact of uncertainty in flame impulse response model on thermoacoustic instability prediction: A dimensionality reduction approach. Proceedings of the Combustion Institute, Elsevier, 2018.

terms:

$$\underbrace{\left(\boldsymbol{i}\sum_{k=0}^{L-1}\Delta t(k+1)h_k^0 e^{-\boldsymbol{i}(k+1)\Delta t\omega^0}+\left.\frac{\partial H}{\partial\omega}\right|^0\right)\Delta\omega}_{G^0=G_r^0+\boldsymbol{i}G_i^0}+\left.\frac{\partial H}{\partial\sigma}\right|^0\Delta\sigma=\sum_{k=0}^{L-1}\Delta h_k e^{-\boldsymbol{i}(k+1)\Delta t\omega^0} \tag{7}$$

where "$|^0$" indicates evaluation using corresponding nominal values.

Now we consider two specific perturbed samples $\boldsymbol{h}'$ and $\boldsymbol{h}''$, with their modal frequencies and growth rates being (ω', σ') and (ω'', σ''). We want to determine under which condition the modal growth rate of these two samples are the same. To achieve that, we expand their characteristic equations as in Eq. (7) and subtract the two expanded characteristic equations from each other:

$$\delta\omega G^0+\delta\sigma\left.\frac{\partial H}{\partial\sigma}\right|^0=\sum_{k=0}^{L-1}\delta h_k e^{-\boldsymbol{i}(k+1)\Delta t\omega^0} \tag{8}$$

where $\delta h_k = h_k' - h_k''$, $\delta\omega = \omega' - \omega''$ and $\delta\sigma = \sigma' - \sigma''$. The r.h.s term in Eq. (8) can be written as:

$$\sum_{k=0}^{L-1}\delta h_k e^{-\boldsymbol{i}(k+1)\Delta t\omega^0}=\delta F_r+\boldsymbol{i}\delta F_i \tag{9}$$

where δF_r and δF_i represent the $\mathrm{F}(\omega^0)$ phasor displacement along the real and imaginary axes of the phasor plot of the FRF, when the coefficients of FIR change from $\boldsymbol{h}'$ to $\boldsymbol{h}''$. Also, $H(\omega-\boldsymbol{i}\sigma)$ can be explicitly written as $H(\omega-\boldsymbol{i}\sigma)=H_r(\omega,\sigma)+\boldsymbol{i}H_i(\omega,\sigma)$. Therefore, we can re-express Eq. (8) in its real part and imaginary part as a system of linear equations, with $\delta\omega$ and $\delta\sigma$ being the unknowns:

$$\begin{pmatrix} G_r^0 & \frac{\partial H_r}{\partial\sigma}\big|^0 \\ G_i^0 & \frac{\partial H_i}{\partial\sigma}\big|^0 \end{pmatrix}\begin{pmatrix}\delta\omega\\ \delta\sigma\end{pmatrix}=\begin{pmatrix}\delta F_r\\ \delta F_i\end{pmatrix} \tag{10}$$

where:

$$G_r^0=\left.\frac{\partial H_r}{\partial\omega}\right|^0+\sum_{k=0}^{L-1}h_k^0(k+1)\Delta t\sin[(k+1)\Delta t\omega^0] \tag{11}$$

$$G_i^0=\left.\frac{\partial H_i}{\partial\omega}\right|^0+\sum_{k=0}^{L-1}h_k^0(k+1)\Delta t\cos[(k+1)\Delta t\omega^0] \tag{12}$$

To obtain the iso-growth lines, the solution for $\delta\sigma$ in Eq. (10) should be zero, which requires:

$$G_i^0\delta F_r - G_r^0\delta F_i = 0 \tag{13}$$

Eq. (13) represents a line equation in the phasor plot of FRF for given G_r^0 and G_i^0, with the line direction $\boldsymbol{l}=(G_r^0, G_i^0)$ and corresponding normal direction $\boldsymbol{n}=(G_i^0, -G_r^0)$, as sketched in Fig. 3. For any two perturbed samples of $\boldsymbol{h}$, as long as they fulfill Eq. (13), they will yield to first order the same modal growth rate. Therefore, we can conclude that the iso-growth lines in the phasor plot of FRF are approximately a set of parallel straight lines approximately. Here we emphasis that (1) G_r^0 and G_i^0 are determined by the FIR model, plus the transfer function of the acoustic model (H) of the thermoacoustic system, as well as the specific thermoacoustic mode (ω^0 and σ^0). Therefore, when (a) different FIR model is employed or (b) geometry or boundary conditions of the combustor are altered or (c) different thermoacoustic modes are investigated, these directions are expected to be different; (2) it can be shown that the mathematical derivation above is also valid when a simple $n-\tau$ flame model is adopted, where both n and τ exhibit uncertainties. The same conclusion is reached eventually, *i.e.,* iso-growth lines in the phasor plot of the FRF are to first order a set of parallel straight lines to first order.

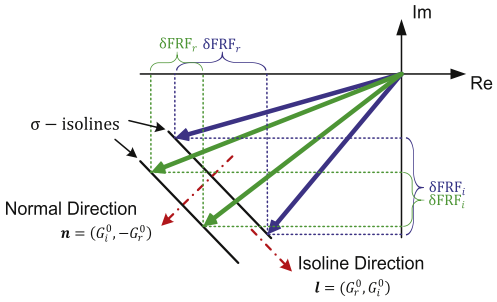


Fig. 3. Two pairs (blue and green) of perturbed samples are plotted in the phasor plot of FRF. Within each pair, two samples yield same modal growth rate and the head locations of their corresponding F phasors are along a straight line; Across pairs, these straight lines are in parallel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Analytical results discussion

Now we are ready to exploit the analytical results for conducting UQ analysis. As shown in Fig. 4 (same simplified FIR model as in Fig. 1): (1) When perturbing individual FIR model coefficient (Δh_0 and Δh_1), the head location of the $\mathrm{F}(\omega^0)$ phasor in general moves to a different level of modal growth rate. This helps to visualize the causal relationship between the FIR model coefficient variations and the modal growth rate variation; (2) Each coefficient's variation can modify the modal growth rate individually. Nevertheless, what ultimately determines the modal growth rate change is the sum of the projection of each phasor $(h_k - h_k^0)e^{-\boldsymbol{i}(k+1)\Delta t\omega^0}$ on the normal direction of the contours of modal growth rate. We use

With permission from S. Guo, C. F. Silva, M. Bauerheim, A. Ghani and W. Polifke. Evaluating the impact of uncertainty in flame impulse response model on thermoacoustic instability prediction: A dimensionality reduction approach. Proceedings of the Combustion Institute, Elsevier, 2018.

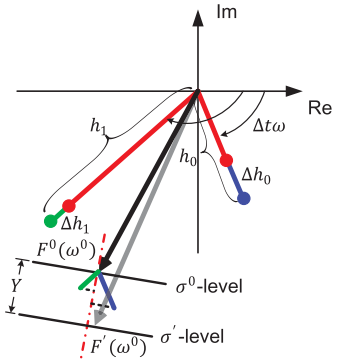


Fig. 4. Visual demonstration of the causal relationship between the variations of FIR model coefficients and the variations of modal growth rate.

Y to denote this projection sum, which can formally be written in Eq. (14). Here $\boldsymbol{n}$ is the normal direction of the iso-growth lines (as indicated in Fig. 3).

$$Y = \sum_{k=0}^{L-1} (h_k - h_k^0) e^{-\boldsymbol{i}(k+1)\Delta t\omega^0} \cdot \frac{\boldsymbol{n}}{|\boldsymbol{n}|} \tag{14}$$

Therefore, we can relate the FIR model coefficient variations to the modal growth rate variation in the following manner:

$$\sigma = f(Y) \tag{15}$$

where the surrogate function $f()$ represents a one-to-one mapping from a specific value of Y, which is the distance (can be negative) along the normal direction of the iso-growth lines starting from the head location of the nominal F phasor evaluated at nominal frequency ω^0, to a specific value of the modal growth rate variation.

At this point we can see that the modal growth rate σ of any sample $\boldsymbol{h}$ is expressed as an univariate function of Y, which depends on nominal coefficients $\boldsymbol{h^0}$ and the nominal eigen-frequency $\boldsymbol{\omega^0}$, but is essentially a linear combination of h_k's (Eq. (14)). For the current UQ problem, the L input variables (the number of h_k's is L) can be condensed into a single variable Y, thus, a potentially very significant dimensionality reduction is achieved. Interestingly, a one-dimensional structure similar to Y was also identified through Active Subspace approach [12], which is a mathematically much more demanding method.

4.3. An analytical UQ strategy

Here we outline the steps for the analytical UQ strategy, assuming that the nominal values and distributions for FIR model coefficients are known, and acoustic network model has already been set up.

Step 1: Perform deterministic thermoacoustic stability analysis to obtain the nominal frequency ω^0 and modal growth rate σ^0 of the interested thermoacoustic mode.

Step 2: Calculate the normal direction of the iso-growth lines $\boldsymbol{n}$.

Step 3: Construct function $f()$ in Eq. (15) in the following manner:

Step 3.1: Draw several samples (5–10) of FIR model coefficients from their distribution, obtain their Y values. Here we suggest the following way to determine those FIR coefficient samples: (a) draw a large number of FIR coefficient samples (e.g., 10^4 samples, store as the data bank), calculate their corresponding Y values; (b) Estimate Y's cumulative distribution function, employ space-filling sampling method (e.g., Latin Hypercube, Halton or Sobol sequences) to draw several (5–10) "ideal" samples for Y which satisfy the experimental design criteria; (c) Find those FIR coefficient samples in the data bank whose Y values are closest to the "ideal" Y sample;

Step 3.2: Calculate the modal growth rate of each sample (in *Step 3.1*) using thermoacoustic system model;

Step 3.3: Plot the modal growth rate values against corresponding Y values and fit a function f (i.e., polynomial, Kriging, Radial basis function, etc.);

Step 4: Perform standard Monte Carlo simulation, i.e., draw a large amount of samples of FIR model coefficients, for each sample, first calculate its Y value through Eq. (14), then calculate the corresponding modal growth rate through Eq. (15). Based on the obtained data bank of the modal growth rate values, a PDF can be constructed and relevant statistical indices can be extracted.

To summarize, the analytical approach manages to uncover a one-dimensional approximation (summarized as Y) of the original system, and take very few potentially expensive thermoacoustic system calculations (1 in *Step 1* and 5–10 in *Step 3.2*) to build a surrogate model only upon Y. Subsequent Monte Carlo simulations can be applied directly on this one-dimensional surrogate model, thus, achieving a significantly accelerated UQ analysis.

5. Application of analytical UQ strategy to combustion test rig

This section aims to demonstrate the accuracy and effectiveness of the UQ strategy proposed in Section 4.3. We investigate the uncertainty regarding the modal growth rate of the BRS burner [15] using an acoustic network solver combined with a 16-coefficient FIR model, assuming that the FIR model coefficients are the only source of uncertainty. The only reason for adopting a simple yet practical acoustic network model here is that brutal force Monte Carlo is entirely feasible so that we can verify our approach.

With permission from S. Guo, C. F. Silva, M. Bauerheim, A. Ghani and W. Polifke. Evaluating the impact of uncertainty in flame impulse response model on thermoacoustic instability prediction: A dimensionality reduction approach. Proceedings of the Combustion Institute, Elsevier, 2018.

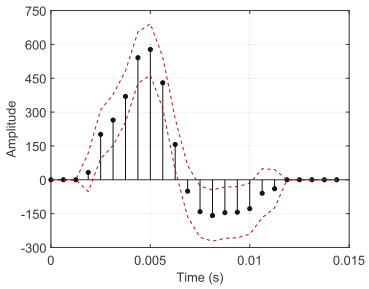


Fig. 5. Finite impulse response model. Each discrete stem represents one coefficient h_k, upper and lower red dot line constitute the 95% confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

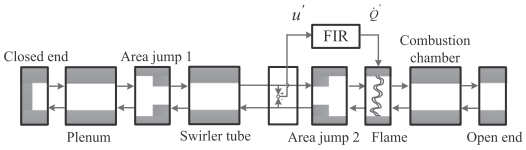


Fig. 6. Sketch of the acoustic network model. Flow from left to right.

5.1. Thermoacoustic model

Figure 5 displays the 16-coefficient FIR model adopted in the current study, which includes both the nominal value and 95% confidence interval for each coefficient. The FIR model is identified using the procedure described in Keesman [16], based on the LES data of [17], where velocity perturbations at the burner mouth u' and global heat release rate fluctuations $\dot{Q}'$ are recorded and considered as input and output, for the identification procedure, respectively. In the current UQ analysis, each FIR model coefficients are treated as a random variable, and they follow multivariate normal distribution. The covariance matrix of the FIR model coefficients is also obtained from the identification process.

The acoustic network model of the BRS burner is sketched in Fig. 6. The geometrical and thermodynamic parameters are given in [12]: two pairs of chamber length values ($l_A = 0.51$ m and $l_B = 0.6$ m) and reflection coefficient values for the combustor exit ($R_A = -0.98$ and $R_B = -0.63$) are selected and marked as case A and case B, respectively. The transfer matrix for each acoustic element and their assembly are given in [2].

Here we focus on evaluating the impact of uncertain FIR model on the dominant mode (with largest modal growth rate) of each case, which are shown in Fig. 7. These modes are calculated with standard deterministic analysis. When FIR model uncertainty is considered, these dominant modes may become unstable, since they are relatively close to the stability limit, thus highlighting the necessity of UQ analysis.

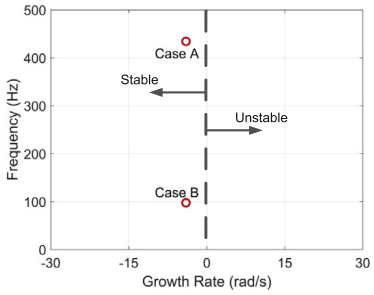
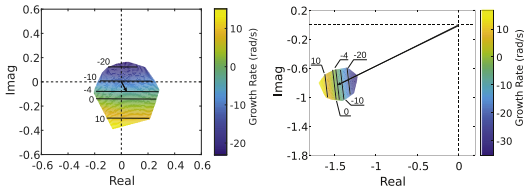


Fig. 7. Thermoacoustic modes calculated by deterministic analysis. For case A, the nominal modal frequency is 434.2 Hz and growth rate is −4 rad/s. For case B, the corresponding values are 97.5 Hz and −4 rad/s.



(a) Case A (b) Case B

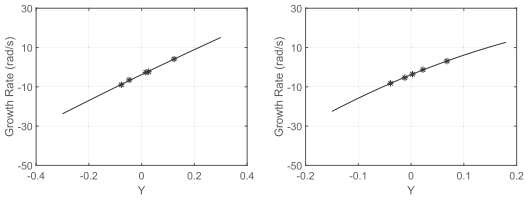
Fig. 8. Modal growth rate contours in generalized phasor plot of FRF. Black lines are the contour levels predicted by the analytical approach. For each case, Line with black arrow is the F(ω^0) phasor evaluated by using nominal values of h_k's, which corresponds to a growth rate value of −4 rad/s.

5.2. UQ analysis

First we assess if the parallel straight lines (Eq. (13)) are good approximations of the iso-growth lines in the phasor plot of FRF. Here we treat the acoustic network calculation as the full accuracy/reference solution, and its prediction of the iso-growth lines for both cases are plotted in Fig. 8. We drew 5000 samples of $\boldsymbol{h}$ for each case and for each sample we calculated the corresponding modal growth rate by acoustic network model as well as the coordinates of F(ω^0) in the phasor plot, based on which we can construct the iso-growth lines. We emphasize that such large number of samples is only necessary for this verification study, but not for the application of our proposed UQ approach. We can see from Fig. 8 that parallel straight lines, as indicated by analytical results, indeed capture the essence of the distribution of the iso-growth lines.

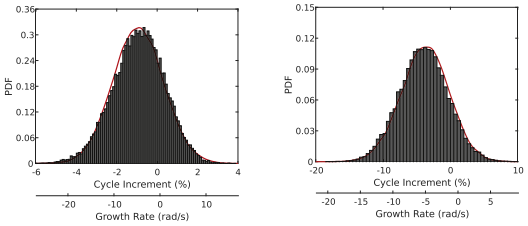
Following Section 4.3, we employ the analytical strategy to conduct UQ analysis. One critical step is ***Step 3***, i.e., constructing a surrogate function $f()$ to map Y value to modal growth rate value. For that purpose, we fit a quadratic function based on

With permission from S. Guo, C. F. Silva, M. Bauerheim, A. Ghani and W. Polifke. Evaluating the impact of uncertainty in flame impulse response model on thermoacoustic instability prediction: A dimensionality reduction approach. Proceedings of the Combustion Institute, Elsevier, 2018.



(a) Case A (b) Case B

Fig. 9. For each case, a quadratic function is fitted to link Y to modal growth rate.



(a) Case A (b) Case B

Fig. 10. PDF comparison between analytical results (red curves) and benchmark Monte Carlo results (bars). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

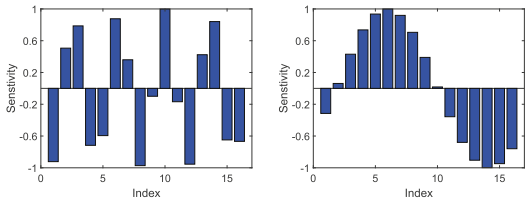
5 samples of FIR model coefficients for each case, which is shown in Fig. 9.

Figure 10 compares the PDF of the modal growth rate predicted by the analytical approach and benchmark Monte Carlo, which is based on 30,000 full-accuracy acoustic network evaluations. Corresponding cycle increments [18] of the thermoacoustic modes are also given to show the relative amplitude change per cycle.

Here it can be seen that our proposed analytical UQ strategy, though it is based on a first-order approximation, is very efficient and robust since: (1) our approach requires very few expensive thermoacoustic system evaluations; (2) in the cases studies, the uncertainty level of the adopted FIR model can already be considered as large (the maximum ratio of coefficient standard deviation to coefficient mean can be as large as 130%), yet our approach still managed to reproduce the benchmark Monte Carlo results; (3) for practical usage, the uncertainty level of the FIR model should not be larger than what we presented here, thus posing less challenge for the proposed UQ approach.

6. Conclusion

In the framework of evaluating the impact of FIR model uncertainties on the thermoacoustic instability prediction, our current work managed to uncover a one-dimensional approximation of the original thermoacoustic system, which compactly



(a) Case A (b) Case B

Fig. 11. Sensitivity measurements for each FIR model coefficients.

summarized the causal relationship between the variations of FIR model coefficients and the variations of thermoacoustic modal growth rate, and proposed a dimensionality strategy to address this UQ problem analytically. Only a handful of thermoacoustic system evaluations are required to accurately build a PDF of the growth rate of the investigated thermoacoustic mode, thus achieving dramatic efficiency improvement, while avoiding any complex mathematical treatments normally employed in other sophisticated surrogate modelling techniques. It should be noted that the convenience of the analytical procedure exposed in this work for affordable UQ studies becomes evident when considering more complex acoustic models, like the ones characterized by the Helmholtz Equation or the Linearized Navier–Stokes Equation, where a single computation is considerably more expensive than the one of a simple acoustic network model. In such models, the acoustic transfer function $H(\omega - i\sigma)$ (see Fig. 2) is directly linked to local values of the system matrix (the acoustic operator) and, therefore, readily obtainable. In addition, the analytical results can be further exploited in the following two areas:

- Sensitivity analysis: The analytical approach provides the sensitivity measurement of modal growth rate against each FIR model coefficient S_k, which can be seen as the ability of each FIR model coefficient to modify the modal growth rate, and can be formally written as:

$$S_k = e^{-j(k+1)\Delta t\omega^0}\frac{\boldsymbol{n}}{|\boldsymbol{n}|}, \quad k = 0, ...L-1 \qquad (16)$$

  For the purpose of illustration, the sensitivity measurement of each FIR model coefficient for both cases (case A and case B in Section 5.1) are plotted in Fig. 11. As described in [9], FIR model coefficients contain insightful information regarding the flow-flame interaction. Now equipped with easily-obtainable sensitivity measurements, it may help to identify the key mechanisms controlling the modal growth rate and enhance our understanding of the relevant physical phenomenon.
- Robust design: In some occasions, we are interested in optimizing combustor

With permission from S. Guo, C. F. Silva, M. Bauerheim, A. Ghani and W. Polifke. Evaluating the impact of uncertainty in flame impulse response model on thermoacoustic instability prediction: A dimensionality reduction approach. Proceedings of the Combustion Institute, Elsevier, 2018.

geometries and boundary conditions to minimize the impact of uncertain FIR model on the modal growth rate prediction. The underlying difficulty is the determination of the modal growth rate variation. Now with the capability of performing UQ analysis analytically, this index can be easily obtained with only a few thermoacoustic system calculations within each optimization algorithm loop, which could lead to a significantly accelerated optimization process.

Acknowledgments

S. Guo is grateful for the financial support from doctoral scholarship of Chinese Scholarship Council (No. 201606830045). W. Polife and C. Silva are grateful to the 2014 Center for Turbulence Research Summer Program (Standford University), where discussions with Franck Nicoud instigated the ideas developed in this study.

References

[1] F. Nicoud, L. Benoit, C. Sensiau, T. Poinsot, *AIAA J.* 45 (2) (2007) 426–441, doi:10.2514/1.24933.
[2] C.F. Silva, T. Emmert, S. Jaensch, W. Polifke, *Combust. Flame* 162 (9) (2015) 3370–3378, doi:10.1016/j.combustflame.2015.06.003.
[3] V. Nair, S. Sarkar, R.I. Sujith, *Probab. Eng. Mech.* 34 (2013) 177–188, doi:10.1016/j.probengmech.2013.09.005.
[4] A. Ndiaye, M. Bauerheim, F. Nicoud, in: Proceedings of the Turbine Technical Conference and Exposition, ASME Turbo Expo, in: GT2015-44133, ASME, Montreal, Quebec, Canada, 2015.
[5] M. Bauerheim, A. Ndiaye, P. Constantine, S. Moreau, F. Nicoud, *J. Fluid Mech.* 789 (2016) 534–566, doi:10.1017/jfm.2015.730.
[6] L. Magri, M. Bauerheim, F. Nicoud, M.P. Juniper, *J. Comput. Phys.* 325 (2016) 411–421, doi:10.1016/j.jcp.2016.08.043.
[7] C. Silva, L. Magri, T. Runte, W. Polifke, *J. Eng. Gas Turb. Power* 139 (1) (2017) 011901, doi:10.1115/1.4034203.
[8] L. Crocco, *J. Am. Rocket Soc.* 21 (6) (1951) 163–178.
[9] R.S. Blumenthal, P. Subramanian, R. Sujith, W. Polifke, *Combust. Flame* 160 (7) (2013) 1215–1224, doi:10.1016/j.combustflame.2013.02.005.
[10] P. Subramanian, R.S. Blumenthal, R. Sujith, W. Polifke, *Combust. Theory Model.* 19 (2) (2015) 223–237, doi:10.1080/13647830.2014.1001438.
[11] W. Polifke, *Ann. Nuclear Energy* 67C (2014) 109–128, doi:10.1016/j.anucene.2013.10.037.
[12] S. Guo, C.F. Silva, A. Ghani, W. Polifke, in: Proceedings of the Turbomachinery Technical Conference & Exposition, ASME Turbo Expo, in: GT2018-75644, ASME, Lillestrom, Norway, 2018.
[13] C. Sovardi, S. Jaensch, W. Polifke, *J. Sound Vibr.* 377 (2016) 90–105, doi:10.1016/j.jsv.2016.05.025.
[14] B. Schuermans, H. Luebcke, D. Bajusz, P. Flohr, in: Proceedings of the ASME Turbo Expo, in: GT2005-68393, ASME, Reno, Nevada, USA, 2005, doi:10.1115/GT2005-68393.
[15] T. Komarek, W. Polifke, *J. Eng. Gas Turb. Power* 132 (6) (2010) 061503, doi:10.1115/1.4000127.
[16] K.J. Keesman, in: System Identification, in: Advanced Textbooks in Control and Signal Processing, Springer London, London, 2011, pp. 59–167.
[17] L. Tay-Wo-Chong, T. Komarek, R. Kaess, S. Föller, W. Polifke, in: Proceedings of the ASME Turbo Expo, in: GT2010-22769, ASME, Glasgow, UK, 2010, pp. 623–635, doi:10.1115/GT2010-22769.
[18] L. Tay-Wo-Chong, S. Bomberg, A. Ulhaq, T. Komarek, W. Polifke, *J. Eng. Gas Turb. Power* 134 (2) (2012) 021502–1–8, doi:10.1115/1.4004183.

With permission from S. Guo, C. F. Silva, M. Bauerheim, A. Ghani and W. Polifke. Evaluating the impact of uncertainty in flame impulse response model on thermoacoustic instability prediction: A dimensionality reduction approach. Proceedings of the Combustion Institute, Elsevier, 2018.

Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach

Shuai Guo[1]
Professur für Thermofluiddynamik,
Technische Universität München,
Boltzmannstr. 15,
Garching D-85748, Germany
e-mail: guo@tfd.mw.tum.de

Camilo F. Silva
Professur für Thermofluiddynamik,
Technische Universität München,
Boltzmannstr. 15,
Garching D-85748, Germany
e-mail: silva@tfd.mw.tum.de

Wolfgang Polifke
Professur für Thermofluiddynamik,
Technische Universität München,
Boltzmannstr. 15,
Garching D-85748, Germany
e-mail: polifke@tum.de

In the preliminary phase of analyzing the thermoacoustic characteristics of a gas turbine combustor, implementing robust design principles is essential to minimize detrimental variations of its thermoacoustic performance under various sources of uncertainties. In this study, we systematically explore different aspects of robust design in thermoacoustic instability analysis, including risk analysis, control design, and inverse tolerance design. We simultaneously take into account multiple thermoacoustic modes and uncertainty sources from both the flame and acoustic boundary parameters. In addition, we introduce the concept of a "risk diagram" based on specific statistical descriptions of the underlying uncertain parameters, which allows practitioners to conveniently visualize the distribution of the modal instability risk over the entire parameter space. Throughout this study, a machine learning method called "Gaussian process" (GP) modeling approach is employed to efficiently tackle the challenge posed by the large parameter variational ranges, various statistical descriptions of the parameters, as well as the multifaceted nature of robust design analysis. For each of the investigated robust design tasks, we propose an efficient solution strategy and benchmark the accuracy of the results delivered by GP models. We demonstrate that GP models can be flexibly adjusted to various tasks while only requiring one-time training. Their adaptability and efficiency make this modeling approach very appealing for industrial practices. [DOI: 10.1115/1.4044197]

1 Introduction

The occurrence of thermoacoustic instability [1] during the operation of a gas turbine combustor constitutes one of the major concerns for aircraft engine manufacturers due to its catastrophic effect on the structural integrity of the combustor system as well as the life-span of the gas turbine. In practice, there are two facts that must be taken into account when designing the thermoacoustic characteristics of the combustor: First of all, the thermoacoustic behavior of the combustor is highly sensitive to small parameter variations in operating conditions as well as acoustic boundary conditions [2]; Second, uncertainties in those parameters are always present during actual combustor operation, i.e., those parameters will display stochastic features instead of staying at some fixed and predetermined nominal values. In consequence, the performance of the combustor may deteriorate once the key parameters controlling the thermoacoustic interaction of the system deviate from their respective nominal conditions, and in extreme cases, a combustor that is stable at nominal operational condition may become unstable under the influence of the stochastic parameter fluctuations. Therefore, implementing the framework of probabilistic system modeling and the principle of robust design are essential to minimize the impact of operational uncertainties on the thermoacoustic performance of the combustor, thus eliminating the costly design iterations, mitigating the operational risk, and promoting an overall more reliable and competitive gas turbine product.

Robust design is generally recognized as the process of carefully selecting design parameter values such that the system behavior is insensitive to various sources of uncertainties [3], which include operation condition uncertainty, actuator imprecision, modeling error, etc. Within the scope of stochastic optimization, uncertain parameters are modeled as random variables with predefined joint probability distributions, and the robust design problem is usually cast as a two-level nested optimization problem [4]: the outer loop forms a conventional optimization problem where an optimization algorithm is employed to find the design parameters that optimize the objective function while subjected to constraint functions. For each iteration, the inner loop is called where an uncertainty quantification (UQ) analysis is performed to calculate the uncertainty measures (e.g., mean, standard deviation, quantiles, etc.) of the system outputs given various sources of uncertainty, and those uncertainty measures are fed back to the outer loop to construct the corresponding objective and constraint functions. Overall, robust design can be viewed as the integration of optimization and UQ analysis.

In the context of thermoacoustic instability analysis, pioneer works were dedicated either to stabilize thermoacoustic modes via configuration optimization or to assess the variability of modal growth rate via UQ analysis. Regarding optimization, for example, Bade et al. [5] proposed a design for thermo-acoustic stability procedure, which optimizes geometric design parameters such that the annular combustor is stable, using network models and flame dynamic models derived from test rig data [6]. In addition, the recent work of Aguilar and Juniper [7] has proposed an adjoint-augmented optimization routine, which is capable of stabilizing all thermoacoustic modes of an annular combustor with minimum geometric modification. Regarding UQ analysis, different techniques have been proposed (i.e., analytical method [8], regression [9], adjoints [10–12], active subspace [8,13], polynomial chaos expansion [14,15]), to address the impact of various uncertainty sources (i.e., flame model, acoustic boundary condition, operation condition) on the variation of modal growth rate.

Despite the remarkable progress made in terms of optimization strategy as well as accelerating the UQ analysis, an integration of these two aspects into the robust design framework is, unfortunately, not straightforward and worth further investigation. Generally, an efficient robust design strategy poses more strict requirements on the perspective of UQ technique: First of all, the robust design problem encountered in thermoacoustic analysis

[1]Corresponding author.

Manuscript received July 1, 2019; final manuscript received July 2, 2019; published online February 14, 2020. Editor: Jerzy T. Sawicki.

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

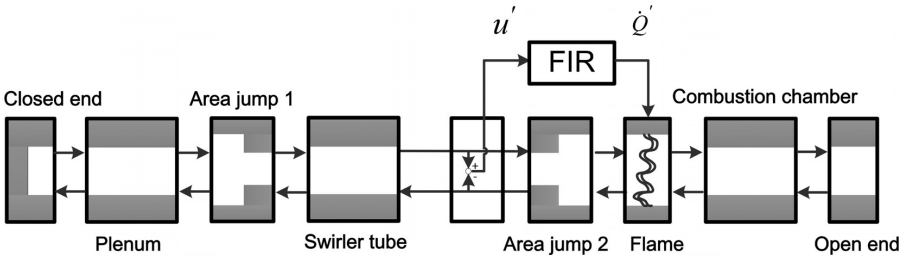


Fig. 1 Sketch of acoustic network model, flow from left to right

belongs to the "reliability-based" type [16], i.e., the constraint or objective is defined on the percentile of the modal growth rate distribution, rather than simply defined on its mean and standard deviation. As a consequence, UQ methods, which only provide mean and variance estimation, are no longer sufficient. Instead, a full probability density function (PDF) calculation of the modal growth rate is mandatory within each optimization iteration; second, the UQ technique has to be able to accommodate large parameter variation ranges so that optimization routines can fully explore different parameter combinations, thus having a better chance to locate the global optimum. And finally, the UQ technique has to be flexible enough to handle different types of PDFs associated with uncertain parameters, due to the fact that our knowledge regarding the uncertain parameters may be renewed as the design process evolves. In a word, robust design poses a serious challenge on the employed UQ strategy. To our best knowledge, no attempt has been made to integrate UQ analysis and optimization into the framework of robust design in thermoacoustic instability analysis.

The main objective of our current work is to explore this uncharted area via a machine learning method called Gaussian process (GP) modeling [17], which trains a computationally efficient surrogate model (GP model) to approximate the thermoacoustic solver. The novelties of our current work can be summarized as follows: first, for the first time we systematically investigate potential scenarios in pursuing the robust combustor thermoacoustic design. Those aspects of robust thermoacoustic design possess great significance in mitigating the operational risk of a gas turbine, yet have not received enough attention in the thermoacoustic research community; second, we identify different robust design tasks and properly categorize them as *risk analysis*, *design control,* and *inverse tolerance design*. In addition, we introduce the concept of a *risk diagram* as a high-level summary for all the investigated robust design tasks, which allows practitioners to conveniently visualize the distribution of the thermoacoustic instability risk over the entire parameter space, thus assisting optimum search performed in various robust design tasks. Third, for each identified robust design task, we not only provide rigorous mathematical formulations but also demonstrate how to exploit Gaussian process models to construct efficient solution strategy. Finally, we prove the capability of the GP modeling approach in significantly reducing the overall robust analysis time as well as flexibly coping with the multifaceted nature of the robust design process, both of which are highly favored by the industrial practices. We emphasize that, due to its "black-box" and "surrogate" nature, GP approach can make use of more sophisticated thermoacoustic and flame models than the ones used in the present work to investigate more complex combustors. Additionally, other aspects of robust design problems can be freely proposed and efficiently answered, all under one modeling framework, as will be demonstrated in our current paper.

This paper is organized as follows: Sec. 2 outlines the investigated combustor configuration, thermoacoustic modeling strategy, and the flame model. Section 3 specifies the uncertain parameters, thermoacoustic modes under consideration, and describes the individual task we propose to address. Section 4 briefly overviews the fundamental theory of GP modeling methodology and presents detailed GP model training processes. A comparison between GP approach and other UQ methods is discussed at the end. Section 5 exploits the newly trained GP models in addressing each of the proposed robust design problem. The paper closes with the main conclusions.

2 Thermoacoustic Framework

This section starts with an introduction of the investigated test rig and the employed thermoacoustic modeling approach, followed by a description of the adopted flame dynamic response model, which will be embedded into the thermoacoustic modeling approach to calculate the eigen-frequencies and growth rates of the thermoacoustic modes.

2.1 Combustor Configuration. The BRS combustor configuration is investigated in this study, which represents a turbulent premixed swirl burner test rig [18,19]. The configuration consists of a plenum, a duct with an axial swirl generator, and a combustion chamber. In this study, an equivalence ratio of 0.77 of perfectly premixed methane–air mixture and a thermal power of 30 kW are considered as the operation conditions.

2.2 Thermoacoustic Modeling. In this study, a low-order acoustic network model is employed to calculate the eigenfrequencies and growth rates of the thermoacoustic modes. Figure 1 displays the network model representation of the burner test rig. Geometrical and thermodynamic parameters used in the acoustic network model are the same as Ref. [8], except the length of the combustor chamber is 0.75 m, and the reflection coefficient at the combustor exit is $|R_{\text{out}}|e^{i\pi}$ in the current study. Other modeling details can be found in Ref. [8] as well. It is worth mentioning that we adopt a rather simple yet practical acoustic network model in the current study only for the purpose of facilitating the benchmark of the robust design results. However, due to the nonintrusive nature of the GP modeling technique, which we discuss in Sec. 4, our proposed strategy can also be extended to more computational intensive acoustic models, e.g., Helmholtz equation, Linearized Navier–Stokes equation, etc., where the benefits of the GP approach could be further amplified, leveraging on the high accuracy of thermoacoustic modeling offered by the sophisticated solvers.

2.3 Flame Model. A flame dynamic response model similar to the distributed time lag model proposed by Komarek and Polifke [18] and validated by Oberleithner and Paschereit [20] is adopted in this study. This model determines the coefficients of the flame impulse response (FIR) by taking into account the flame response to both swirl number fluctuations and axial velocity fluctuations.

This model assumes that the shape of the flame impulse response can be approximated by adding three Gaussian

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

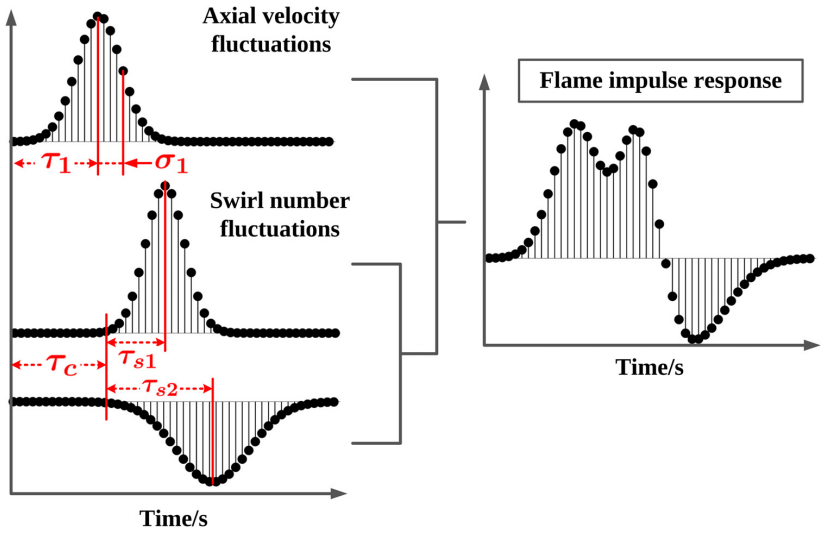


Fig. 2 Sketch of the employed FIR model for the flame dynamics. Corresponding frequency response is shown in Fig. 3(*b*).

distributions: one positive (+) Gaussian to account for the flame response to axial velocity fluctuations, and one (+) combined with another (–) Gaussian to describe the response to swirl number fluctuations.

An illustration of the flame model is given in Fig. 2, where its five model parameters, i.e., τ_1, σ_1, τ_c, τ_{s1}, τ_{s2}, are also shown. This parametrization is slightly different from Ref. [18] due to the following reasons: First, we want to explicitly express the time lag τ_c, which, according to the recent work of Albayrak and Polifke [21], represents the time needed for an inertial wave to propagate from the swirler to the flame base. Komarek and Polifke [18] have shown that the position of swirler can be changed without changing the mean flame shape and the flame response to axial velocity (τ_1, σ_1) or swirl perturbations (τ_{s1}, τ_{s2}). Therefore, τ_c can be perceived as a control parameter to stabilize the combustor; second, this parametrization allows the introduction of flame parameter uncertainty while preserving the continuous shape of the flame impulse response. The remaining parameters are: τ_1 and σ_1 represent the mean and standard deviation of the flame heat release response under an impulse axial velocity perturbation, respectively; τ_{s1} and τ_{s2} represent the characteristic time lags for the flame response to the swirl number fluctuations. The standard deviations of the associated positive and negative Gaussian functions are considered to be one third of τ_{s1} and τ_{s2}, respectively. Therefore, we can express the value of the FIR coefficient h_k as [18]

$$h_k = \frac{\Delta t}{\sigma_1\sqrt{2\pi}} e^{-\frac{(k\Delta t-\tau_1)^2}{2\sigma_1^2}} + \frac{\Delta t}{\sigma_2\sqrt{2\pi}} e^{-\frac{(k\Delta t-\tau_2)^2}{2\sigma_2^2}} - \frac{\Delta t}{\sigma_3\sqrt{2\pi}} e^{-\frac{(k\Delta t-\tau_3)^2}{2\sigma_3^2}} \quad k \in [1, 2, \ldots N] \tag{1}$$

where N is the number of FIR coefficients, and

$$\tau_2 = \tau_c + \tau_{s1} \quad \tau_3 = \tau_c + \tau_{s2} \qquad \sigma_2 = \tau_{s1}/3 \quad \sigma_3 = \tau_{s2}/3 \tag{2}$$

3 Robust Design Tasks

In this section, the investigated parameters as well as their variational ranges are introduced. Subsequently, two thermoacoustic modes are specified, which the following robust design analysis will focus on. Afterward, detailed descriptions of our proposed design tasks are provided.

3.1 Uncertain Parameters. Our proposed robust design tasks involve six parameters, including five flame model parameters τ_1, σ_1, τ_c, τ_{s1}, τ_{s2}, and one acoustic boundary condition, i.e., the magnitude of the reflection coefficient at the combustor outlet $|R_{\text{out}}|$. Their nominal values (obtained from "Table 2, 30 kW optimal fit" in Ref. [22]) as well as their investigated variational ranges are summarized in Table 1. For τ_c and $|R_{\text{out}}|$, relatively large variational ranges are investigated in order to fully explore the parameter space. In a practical scenario, τ_c is determined by the distance between the swirler and the flame base [22], while the modification of $|R_{\text{out}}|$ can represent the installation of acoustic damping devices or a change of turbine working condition [23].

In our current study, three distribution types for τ_1, σ_1, τ_{s1}, τ_{s2}, are considered: independent uniform, independent Gaussian, and correlated Gaussian. As we stated in Sec. 1, a common situation in the design process is that our knowledge regarding the uncertain parameters may be renewed as the design process evolves. Here, independent uniform indicates the least informative knowledge, i.e., uncertain parameters have equal chance to be anywhere within the variational ranges. This usually happens at the beginning stage when no prior knowledge exists; Independent Gaussian contains more information since we know uncertain parameters have a tendency to appear more frequently around their corresponding mean; Correlated Gaussian, on the other hand, gives the most complete statistical description of the uncertain parameters. This usually happens at a later design stage when a sufficient

Table 1 Uncertainty information of the investigated parameters

Parameters		Nominal	Range
Flame (units: ms)	τ_1	$\tau_1^0 = 2.85$	$0.9\tau_1^0 - 1.1\tau_1^0$
	σ_1	$\sigma_1^0 = 0.7$	$0.9\sigma_1^0 - 1.1\sigma_1^0$
	τ_c	$\tau_c^0 = 3$	2–4.8
	τ_{s1}	$\tau_{s1}^0 = 1.8$	$0.9\tau_{s1}^0 - 1.1\tau_{s1}^0$
	τ_{s2}	$\tau_{s2}^0 = 1.8$	$0.9\tau_{s2}^0 - 1.1\tau_{s2}^0$
Acoustic BC	$\lvert R_{\text{out}}\rvert$	$\lvert R_{\text{out}}\rvert^0 = 0.9$	0.6–1

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

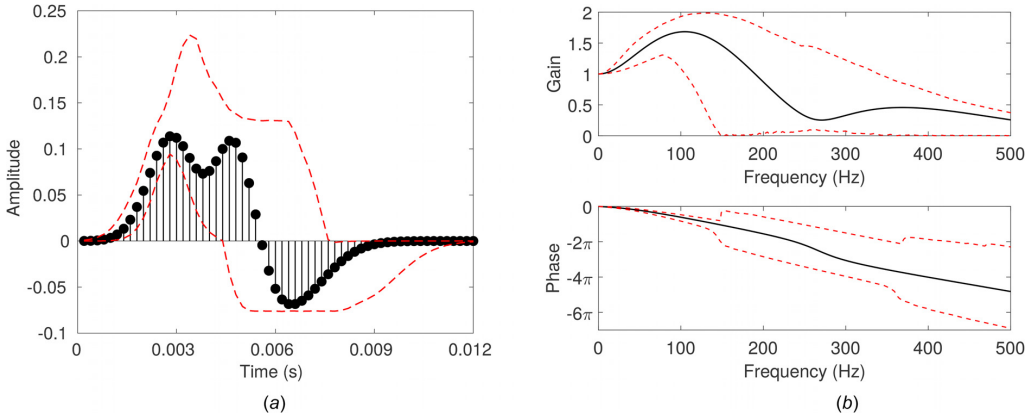


Fig. 3 Uncertainty reflected in FIR and FTF. The black stems represent the nominal values, while the red dash lines represent the envelopes recording the extreme values at each (*a*) coefficient h_k (FIR) and (*b*) frequency (FTF).

amount of experiments and simulations have been conducted so that the flame dynamics can be properly described. In Sec. 5, we start by considering τ_1, σ_1, τ_{s1}, τ_{s2} following independent uniform distribution, and then we assess the impact of different distribution types on the analysis results, all using the same GP models.

The uncertainties embedded in the flame model parameters can be easily propagated to the uncertainties in FIR coefficient h_k's and the corresponding flame transfer function (FTF) by performing Monte Carlo on Eq. (1) and subsequently on Eq. (3). The uncertainty propagation results are shown in Fig. 3

$$\mathrm{FTF}(\omega) = \sum_{k=1}^{N} h_k e^{-ik\Delta t\omega}, \quad \omega \in \mathbb{R} \tag{3}$$

3.2 Thermoacoustic Modes Specification. Figure 4 shows the thermoacoustic modes up to 400 Hz, calculated by deterministic analysis, i.e., when the nominal values of the flame model parameters and $|R_{\mathrm{out}}|$ are used in the acoustic network. Highly damped modes are ignored; thus they are not shown in the figure. Two dominant thermoacoustic modes are presented: the higher frequency mode ($\omega = 287.5$ Hz, $\alpha = -27.7$ rad/s) is identified as the quarter wave mode of the combustor [19]. We denote this mode as *cavity mode*. The other mode ($\omega = 139.3$ Hz, $\alpha = -24.8$ rad/s) is identified as the intrinsic thermoacoustic mode (ITA) [24–26]. We denote this mode as *ITA mode*. Although both modes are not immediately adjacent to the stability limit, considering the large variational ranges of the FIR and FTF (displayed in Fig. 3), there still exists risk for both modes to be unstable. Also notice that it is not uncommon that in thermoacoustic control, the effort to stabilize one mode may unexpectedly promote another mode to become unstable [1]. Therefore, in our current study, both modes will be taken into account concurrently, and the goal of robust design analysis is to mitigate the risk of instability for both modes.

3.3 Design Task Descriptions. To comprehensively investigate the problem of robust thermoacoustic design, we propose the following scenarios, which we want to address individually:

Q1-Risk analysis: what are the *risk factors* [8] of the thermoacoustic modes of the system when uncertainties are presented in the flame parameter τ_1, σ_1, τ_{s1}, and τ_{s2}?
Q2-Ideal control design: using τ_c as a control factor, what is the required minimum modification of τ_c to eliminate the risk of instability of both cavity and ITA mode simultaneously?
Q3-Realistic control design: in reality, we cannot perfectly control τ_c. Meanwhile, $|R_{\mathrm{out}}|$ is also uncertain. How would these two additional uncertain parameters affect the decision made from Q2?
Q4-Tolerance design: given certain threshold for risk factor, what are the maximum allowable variational ranges for τ_c and $|R_{\mathrm{out}}|$?
Q5-Risk diagram: is it possible to construct a diagram with risk factors showing at arbitrary combination of τ_c and $|R_{\mathrm{out}}|$, thus encompassing answers for all the previous questions?
Q6-Sensitivity analysis: if the statistical descriptions of the uncertain flame parameters are modified, then how would this affect the risk diagram derived from Q5?

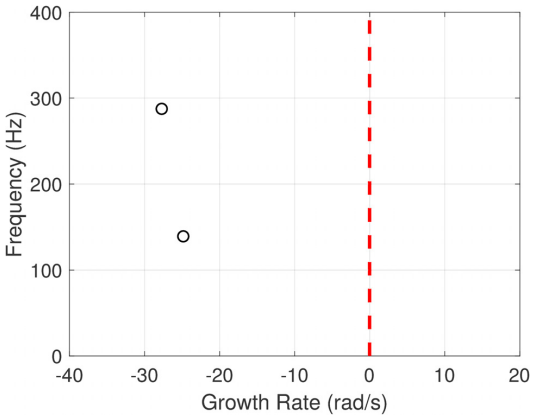


Fig. 4 Two dominant eigenmodes from deterministic analysis. In this study, modes with growth rates smaller than zero are considered stable.

Q1 sets the stage for robust design analysis. Here, risk factor $P_f(\%)$ describes the probability that a thermoacoustic mode is unstable, which can be expressed as the following [8]:

$$P_f(\%) = 100 \int_0^{\infty} \mathrm{PDF}(\alpha) d\alpha \tag{4}$$

where α stands for the modal growth rate. The risk analysis performed in Q1 serves as the foundation for the subsequent robust design analysis.

Q2 tries to increase the robustness of the system design by introducing a control factor: τ_c. In practice, τ_c can be tuned by

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

modifying the distance between the swirler and the combustor dump plane [27]. By adjusting the constructive/destructive interferences between the flame response to axial velocity fluctuations and the response to swirl number fluctuations, we are aiming at making the growth rates of both modes vary only in the stable region, thus mitigating the risk.

Q3 goes one step further toward reality compared to Q2: τ_c can be easily influenced by the perturbations of the convective velocity in the swirler tube, and the reflection coefficient $|R_{\text{out}}|$, which largely depends on the turbine operating conditions, is inherently uncertain. By additional acknowledgment of the uncertainties in flame and acoustic boundary parameters, the robustness of the system can be further enhanced.

Q4 approaches the robust design via the perspective of an *inverse problem*: searching the maximum allowable variational ranges for the uncertain parameters while still satisfying the constraints. This is beneficial toward understanding, e.g., the trade-off between combustor and turbine operational uncertainties, thus achieving a better coordination between combustor and turbine design groups.

Q5 aims higher by proposing the concept of a risk diagram, which provides the distribution of risk factor over the entire the parameter space. This task extends Q1 to enable the practitioner to conveniently locate the desired parameter regions for achieving efficient optimum search performed in Q2, Q3, and Q4.

Finally, Q6 mimics the situation that our knowledge of the uncertain parameters, which is reflected in their assigned PDF, will be updated as the design process evolves. Therefore, it is necessary to reconduct the risk analysis and assess the sensitivity of results to different distribution types of parameters.

As progressing from Q1 to Q6, the dramatically increased problem complexity and associated computational cost prohibit a direct application of acoustic solvers, including the relatively cheap acoustic network model. However, surrogate modeling technique turns this around and opens new possibilities for addressing robust design problems efficiently, which we discuss in Sec. 4.

4 Gaussian Process Modeling

To efficiently address the robust design tasks outlined in Sec. 3.3, a machine learning method called GP is employed and corresponding solution strategies are constructed. In the context of thermoacoustic analysis, Schneider et al. [28] employed GP to construct a real-time modeling approach to simulate pressure pulsation amplitudes of an annular gas turbine combustor. The accuracy of their approach was further verified on the measured data for a wide range of operating conditions. In a recent work, Chattopadhyay et al. [29,30] adopted a GP model to predict combustor instability at untried operational conditions and successfully benchmarked the results with the experimental data. In the following, we first briefly review the fundamentals of GP modeling. Afterward, we train a total of four surrogate models for predicting modal frequencies and growth rates of the ITA and cavity modes, respectively, which will be repetitively used for addressing different robust design tasks in the subsequent sections. We close this section with a brief comparison of the GP approach with other UQ methods. The GP model training is performed via UQLab [31].

4.1 Fundamentals Overview. Gaussian process modeling is a supervised learning technique in machine learning domain [32]. Based on carefully selected training samples and their corresponding responses, GP is capable of training a computationally efficient surrogate model to approximate the original high-fidelity solver. As a result, the UQ analysis nested in the robust design procedure can be performed directly on the trained GP model; thus, repetitive high-fidelity solver calls can be avoided and the overall efficiency of the robust design analysis can be significantly improved. In the following, we summarize the key features of GP modeling. Jones et al. [33] provides a complete derivation of GP method and Forrester and Keane [34] review the topic of employing GP in the context of surrogate modeling. For the application of GP modeling in robust design in other domains, we refer readers to Refs. [35] and [36].

GP modeling treats the output $f(\boldsymbol{x})$ of the high fidelity model at $\boldsymbol{x}$ (a vector with the entries being the individual input parameter) as the realization of a Gaussian process

$$f(\boldsymbol{x}) = \beta + Z(\boldsymbol{x}) \tag{5}$$

where β is a constant value and $Z(\boldsymbol{x})$ corresponds to the departure from β at $\boldsymbol{x}$, which is modeled as a Gaussian stochastic function with zero mean, variance σ^2, and covariance defined as

$$\text{Cov}[Z(\boldsymbol{x}^i, \boldsymbol{x}^j)] = \sigma^2 R(\boldsymbol{x}^i, \boldsymbol{x}^j) \tag{6}$$

where $R(\boldsymbol{x}^i, \boldsymbol{x}^j)$ is the correlation function between any two locations $\boldsymbol{x}^i$ and $\boldsymbol{x}^j$ in the input parameter space. In this study, a Gaussian correlation function is adopted to describe $R(\boldsymbol{x}^i, \boldsymbol{x}^j)$, which can be written as

$$\boldsymbol{R}(\boldsymbol{x}^i, \boldsymbol{x}^j) = \exp\left[-\sum_{k=1}^{M} \theta_k (x_k^i - x_k^j)^2\right] \tag{7}$$

where M denotes the dimension of the input (i.e., number of input parameters), subscript k represents the kth component of the input vector $\boldsymbol{x}$, and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_M]$ controls the level of correlation in the corresponding dimension: a high value of θ_i signifies a high rate of decay of correlation in the ith dimension, and vice versa.

Based on a set of training samples $\boldsymbol{X}_D = [\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N]^{\text{T}}$ and their corresponding responses $\boldsymbol{Y}_D = [f(\boldsymbol{x}^1), \ldots, f(\boldsymbol{x}^N)]^{\text{T}}$, we can train the GP model by finding values for β, σ^2, and $\boldsymbol{\theta}$ such that the likelihood of achieving the observations (training samples and their corresponding responses) is maximized. For any given GP model parameter set $(\beta, \sigma^2, \boldsymbol{\theta})$, the likelihood function L can be directly obtained from the probability density function of a multivariate normal distribution

$$L(\beta, \sigma^2, \boldsymbol{\theta} | \boldsymbol{X}_D, \boldsymbol{Y}_D) = \frac{1}{(2\pi\sigma^2)^{N/2} |\boldsymbol{R}|^{1/2}} \times \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y}_D - 1\beta)^{\text{T}} \boldsymbol{R}^{-1} (\boldsymbol{Y}_D - 1\beta)\right] \tag{8}$$

where $\mathbf{1}$ is a vector of ones of dimension M. In practice, the logarithm of the above likelihood function is being maximized, which can be written as

$$\ln\left(L(\beta, \sigma^2, \boldsymbol{\theta} | \boldsymbol{X}_D, \boldsymbol{Y}_D)\right) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\ln(|\boldsymbol{R}_D|) - \frac{1}{2\sigma^2}(\boldsymbol{Y}_D - 1\beta)^{\text{T}} \boldsymbol{R}_D^{-1} (\boldsymbol{Y}_D - 1\beta) \tag{9}$$

By setting the derivatives of Eq. (9) with respect to β and σ^2 to zeros, we can obtain their maximum likelihood estimation

$$\widehat{\beta} = (1^{\text{T}} \boldsymbol{R}_D^{-1} 1)^{-1} 1^{\text{T}} \boldsymbol{R}_D^{-1} \boldsymbol{Y}_D \tag{10}$$

$$\widehat{\sigma^2} = \frac{1}{N}\left(\boldsymbol{Y}_D - 1\widehat{\beta}\right)^{\text{T}} \boldsymbol{R}_D^{-1} \left(\boldsymbol{Y}_D - 1\widehat{\beta}\right) \tag{11}$$

For estimating $\boldsymbol{\theta}$, the following auxiliary optimization problem has to be solved:

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\text{argmax}}\left[-\frac{N}{2}\ln\left(\widehat{\sigma^2}\right) - \frac{1}{2}\ln(|\boldsymbol{R}_D|)\right] \tag{12}$$

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

Finally, the GP model prediction $\widehat{f}(\boldsymbol{x})$ at an arbitrary location $\boldsymbol{x}$ is

$$\widehat{f}(\boldsymbol{x}) = E[G(\boldsymbol{x})|\boldsymbol{X}_D, \boldsymbol{Y}_D] = \widehat{\beta} + \boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{R}_D^{-1}(\boldsymbol{Y}_D - 1\beta) \tag{13}$$

which is interpreted as the conditional expectation of the Gaussian process $G(\boldsymbol{x})$, conditional on the known responses of the training samples [37]. $\boldsymbol{r}(\boldsymbol{x})$ in Eq. (13) is the correlation vector between $\boldsymbol{x}$ and all the training samples, i.e., $\boldsymbol{r}(\boldsymbol{x}) = [\boldsymbol{R}(\boldsymbol{x}, \boldsymbol{x}^1), \ldots, \boldsymbol{R}(\boldsymbol{x}, \boldsymbol{x}^N)]$.

At training sample locations $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^N$, the following condition is fulfilled:

$$\widehat{f}(\boldsymbol{x}^i) = f(\boldsymbol{x}^i) \tag{14}$$

indicating that GP surrogate modeling is an interpolation method that produces the exact responses at corresponding training sample locations.

4.2 Gaussian Process Model Training. In this study, we strive to train four separate, generally applicable GP models for approximating the frequency and growth rate of ITA and cavity modes, respectively, covering the whole parameter range specified in Table 1. The robust design analysis conducted in Sec. 5 will be configured completely upon the models trained in this section, i.e., no additional training will be needed.

Figure 5 summarizes the flowchart for training the GP models. The training (as well as the employment of GP models in Sec. 5) is performed on an Intel Core i5-6200 U CPU 2.30 GHz laptop PC. Details for each step are given in the following:

Step 1: We start the training process by uniformly drawing 18 samples of $(\tau_1, \sigma_1, \tau_c, \tau_{s1}, \tau_{s2}, |R_{\text{out}}|)$ from the parameter space specified in Table 1, via a Halton sequence [38], which is a low-discrepancy sampling method that possesses an excellent space-filling property. Loeppky et al. [39] suggested that approximately a sample size of $10M$ (M is the number of input parameters) is required when using GP to approximate the target response. Therefore, we use a third of $10M$ (18 in this case) as our initial sample size. Subsequently, we can construct the design matrix $\boldsymbol{X_D}$ as Eq. (15), with each row representing one sample

$$\boldsymbol{X_D} = \begin{bmatrix} \tau_1^{(1)} & \sigma_1^{(1)} & \tau_c^{(1)} & \tau_{s1}^{(1)} & \tau_{s2}^{(1)} & |R_{\text{out}}|^{(1)} \\ \tau_1^{(2)} & \sigma_1^{(2)} & \tau_c^{(2)} & \tau_{s1}^{(2)} & \tau_{s2}^{(2)} & |R_{\text{out}}|^{(2)} \\ & & & \cdots & & \\ \tau_1^{(18)} & \sigma_1^{(18)} & \tau_c^{(18)} & \tau_{s1}^{(18)} & \tau_{s2}^{(18)} & |R_{\text{out}}|^{(18)} \end{bmatrix} \tag{15}$$

Step 2: For each sample, we calculate the corresponding output quantities via acoustic network: ITA mode frequency ω^{ITA}, ITA mode growth rate α^{ITA}, cavity mode frequency ω^{CAV}, and cavity mode growth rate α^{CAV}. We store them in the response matrix $\boldsymbol{Y_D} = [\boldsymbol{\omega}^{\text{ITA}}, \boldsymbol{\alpha}^{\text{ITA}}, \boldsymbol{\omega}^{\text{CAV}}, \boldsymbol{\alpha}^{\text{CAV}}]$, with each column vector containing the corresponding response values of all the samples in $\boldsymbol{X_D}$.

Step 3: We train four separate GP models by using the following "sample-response" pairs: $(\boldsymbol{X_D}, \boldsymbol{\omega}^{\text{ITA}})$, $(\boldsymbol{X_D}, \boldsymbol{\alpha}^{\text{ITA}})$, $(\boldsymbol{X_D}, \boldsymbol{\omega}^{\text{CAV}})$ and $(\boldsymbol{X_D}, \boldsymbol{\alpha}^{\text{CAV}})$.

Step 4: We perform leave-one-out cross-validation [31] to judge the prediction accuracy of all four GP models. Leave-one-out cross-validation estimates the generalization error (GE) of the GP model in the following manner:

$$\text{GE} = \frac{1}{N}\sum_{i=1}^{N}\left(f_i - \widehat{f}_i^{(-i)}\right)^2 \tag{16}$$

where N denotes the total number of training samples, f_i corresponds to the known response of the training sample $\boldsymbol{x}^i$, and $\widehat{f}_i^{(-i)}$ represents the prediction at $\boldsymbol{x}^i$ using the GP model constructed upon all training samples except $(\boldsymbol{x}^i, f_i)$.

Step 5: If the prediction accuracy for any GP models is not satisfied, we draw six new samples from the Halton sequence and expand $\boldsymbol{X_D}$, run the acoustic network solver to calculate the responses of these new samples, augment $\boldsymbol{Y_D}$ by the newly obtained responses, and start another round of model training and validating. We choose to update six samples at once to take advantage of parallel computing, which would be especially important when expensive solvers are employed.

Step 6: The process continues until the model generalization errors are saturated.

Figure 6 displays the convergence history of the generalization errors of all four GP models as training samples are enlarged iteratively. It can be seen that after 15 iterations, i.e., a total of 102 $(18 + 14 \times 6)$ training samples are employed, the generalization errors for frequency and growth rate predictions have saturated, thus implying that all four GP models have converged. The total training time is 93 s.

To further test the accuracy of the newly trained GP model, another 50 samples of $(\tau_1, \sigma_1, \tau_c, \tau_{s1}, \tau_{s2}, |R_{\text{out}}|)$ were randomly draw, which are not included in the training samples, and compare the GP-predicted responses with the true responses calculated via acoustic network at these sample locations. We emphasize that this is only for the purpose of verification, but not necessary for the GP model training.

The comparisons shown in Fig. 7 confirm the results given by the error diagnoses in Fig. 6. Therefore, we can see that with only a modest number of training samples, the obtained GP models have already achieved high predicting accuracy and applicable over a large parameter space.

4.3 Comparisons With Other Uncertainty Quantification Methods. In the context of accelerating the UQ analysis in thermoacoustic instability analysis, various methods have been proposed and validated on their corresponding problem settings: a polynomial-based regression method [9] and an analytical method [8], which enjoy simplicity but so far have only dealt with flame model uncertainties; Adjoint-based methods [10–12], which enjoy mathematical elegance and are free of the "curse of dimensionality," but most effective for small parameter variational ranges and may miss the global optimum when combined with a gradient-based optimization routine; polynomial chaos expansion [14,15,40], which shows computational efficiency for a small number of uncertain parameters, but relies on the specific PDFs of the input parameters to determine the expansion and also works best on a rather small parameter variational ranges; *active*

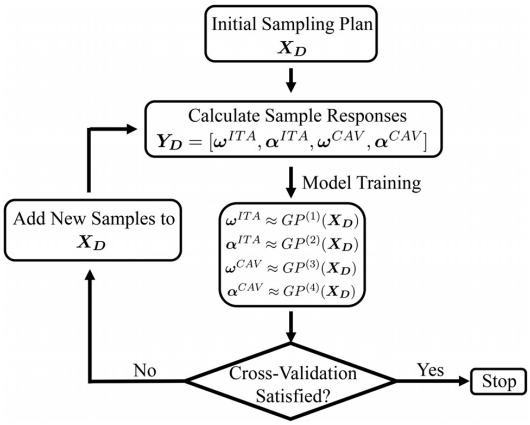


Fig. 5 Flowchart of GP model training

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

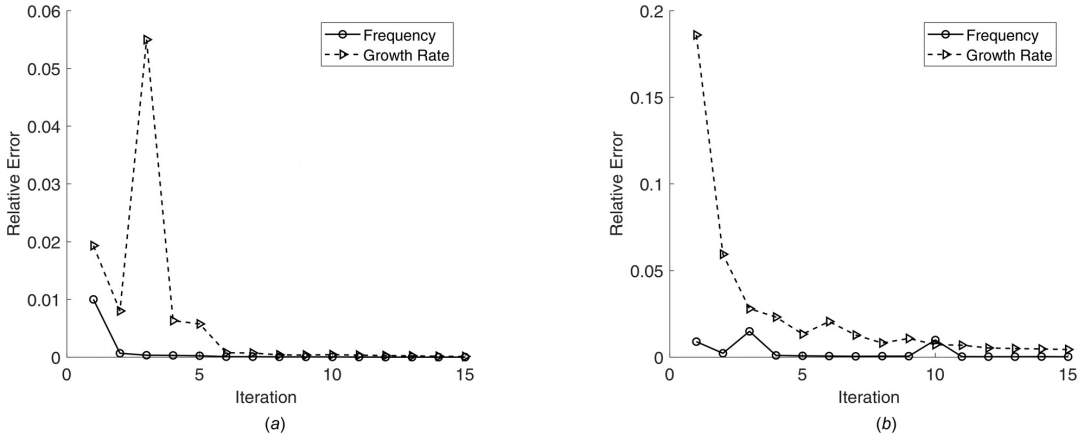


Fig. 6 Convergence history of the GP models for approximating (*a*) ITA and (*b*) cavity mode. For each GP model, the GE calculated via Eq. (16) is scaled by the variance of the responses of the training sample and shown as the relative error in the figure. For both modes, frequency converges faster than growth rate. The relative error of the ITA mode is smaller than the cavity mode.

subspace approach [8,13], which leverages on the low-dimensional subspace of the original system to accelerate the analysis, but the existence of this low-dimensional subspace is not guaranteed for general problems.

Gaussian process approach, on the other hand, fits perfectly in the context of robust design, with its capability of handling large parameter variational ranges and various probability distribution types, provided that the number of the input parameters is moderate (smaller than 20 [41]). We will demonstrate its accuracy and efficiency in Sec. 5.

5 Robust Design Analysis

In this section, we exploit the GP models trained from the last section to address the proposed robust design problems outlined in Sec. 3.3. Each subsection is devoted to one problem, starting with the problem statement, followed by the solution strategy, results, demonstration, and validation.

5.1 Risk Analysis

"Q1: what is the risk factor of the system when uncertainties are presented in the flame parameter τ_1, σ_1, τ_{s1}, and τ_{s2}?"

Here, we consider τ_c and $|R_{\text{out}}|$ to be fixed at their nominal values, while τ_1, σ_1, τ_{s1}, τ_{s2} follow independent uniform distributions with the parameter range indicated in Table 1. When flame parameters are uncertain, the growth rate values for both modes will also exhibit variation, thus having the potential to be unstable. Therefore, this is a typical *forward uncertainty quantification* [8] problem.

To calculate the risk factor, we can approximate Eq. (4) via Monte Carlo simulation [13] by repeatedly calculating the growth rates of different samples of (τ_1, σ_1, τ_{s1}, τ_{s2}) and then compute the ratio between the number of samples with growth rate larger than zero and the total number of samples. Instead of calculating the growth rate with the acoustic solver, which might be computationally very expensive, we perform Monte Carlo directly on the GP models, thus significantly accelerating the analysis due to its negligible predicting cost. We randomly draw 20,000 samples of (τ_1, σ_1, τ_{s1}, τ_{s2}) from their distributions for Monte Carlo simulation, which are large enough to ensure the statistical convergence of the obtained PDFs.

Probability density functions of growth rate and frequency of both ITA mode and cavity mode are shown in Fig. 8. The diagonal figures display the marginal distributions while the off-diagonal figure shows their joint distribution. Comparisons against the reference PDFs, which are obtained via applying the Monte Carlo directly on the acoustic network solver using the same 20,000 samples, are also made and shown in Fig. 8. R^2 coefficient (defined in Appendix C) is employed to quantify the accuracy of the GP approximation. Excellent matches between GP models and full thermoacoustic model for both modes are observed, thus demonstrating the high predictive accuracy of the trained GP models. In terms of computation time, the full Monte Carlo simulation performed via acoustic network solver takes 271 s, while GP models only require 1.3 s. We emphasize that this acceleration of UQ analysis will be even more significant when GP method is adopted to approximate other more sophisticated thermoacoustic models, and when a complete robust design task is required where multiple times of UQ analyses have to be performed.

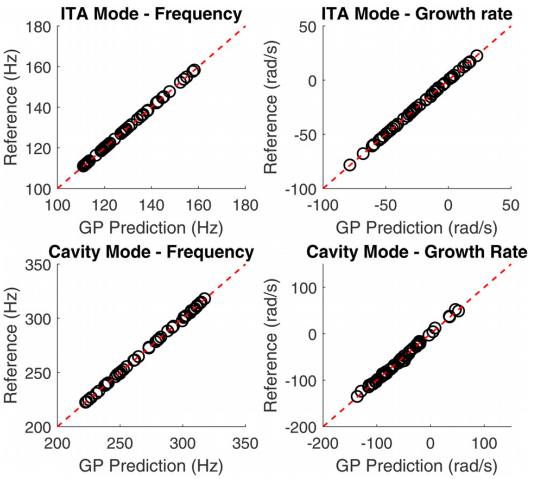


Fig. 7 Comparisons between GP model predictions and the reference values produced by acoustic network solver

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

5.2 Ideal Control Design

"Q2: using τ_c as a control factor, what is the required minimum modification of τ_c to eliminate the risk of instability of both cavity and ITA mode simultaneously?"

As in Sec. 5.1, we consider independent uniform distributions with variational ranges specified in Table 1 for τ_1, σ_1, τ_{s1}, τ_{s2} and assume that the exit reflection coefficient $|R_{\text{out}}|$ is fixed at its nominal value. We try to optimize the time lag τ_c of swirl fluctuations under the constraint that the risk factors of both modes ($P_f^{(I)}$ for the ITA mode and $P_f^{(C)}$ for the cavity mode) are small. This optimization problem is explicitly expressed as follows:

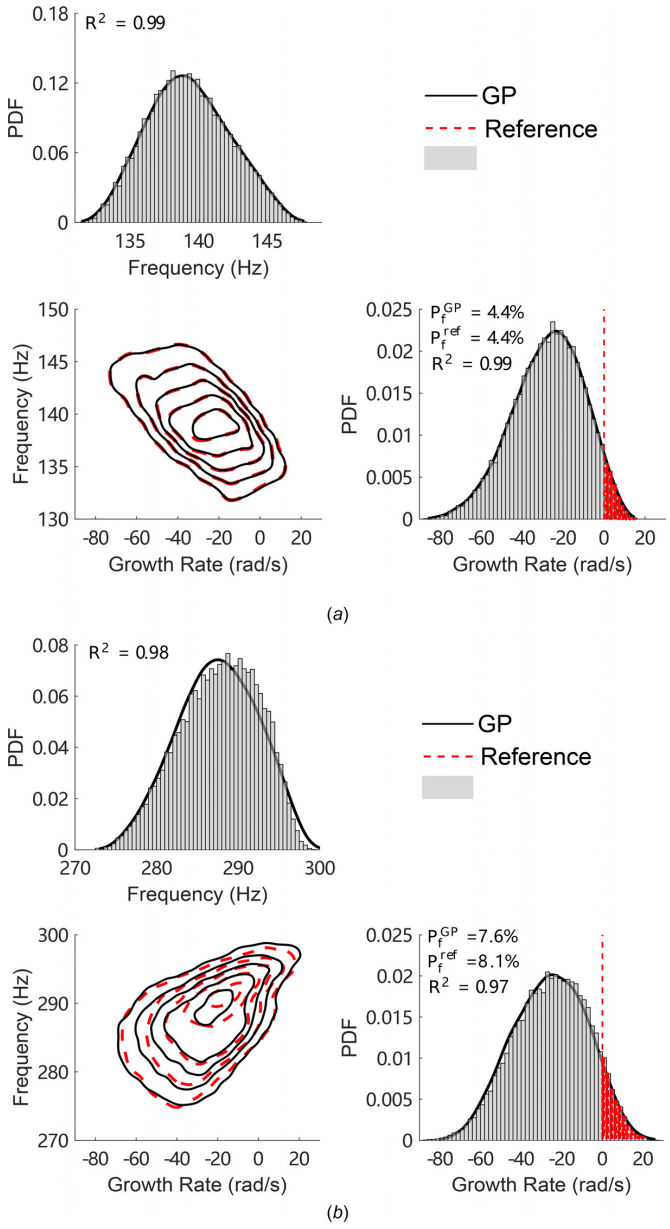


Fig. 8 PDFs of the modal frequency and growth rate given by the GP models and reference acoustic solver are compared. Diagonal figures are marginal distributions while off-diagonal figure is the joint distribution. Same Monte Carlo samples are used for both GP models and acoustic network solver: (*a*) ITA mode and (*b*) cavity mode.

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

$$\begin{aligned} \min_{\tau_c} & f(\tau_c) = (\tau_c - \tau_c^0)^2 \\ \text{subject to:} \quad & P_f^{(I)}(\tau_c) \leq 0.1\% \\ & P_f^{(C)}(\tau_c) \leq 0.1\% \\ & 0.66\tau_c^0 \leq \tau_c \leq 1.6\tau_c^0 \end{aligned} \tag{17}$$

The first line represents the objective of the optimization, i.e., that the modification of τ_c from its nominal value τ_c^0 is minimum. Notice that the goal of mitigating instability risk is formulated as constraints on $P_f^{(I)}$ and $P_f^{(C)}$ instead of as objectives. The optimal value of τ_c should lie within the range in Table 1. This condition is also expressed as a constraint. The formulation (17) of the optimization problem Q2 assures its well posedness, since the minimum possible value of P_f is zero and there may exist multiple τ_c's that lead to zero P_f. Thus, we seek the *minimum* modification of τ_c from τ_c^0 that satisfies the constraint of zero instability risk, thus ensuring that only a single optimum exists.

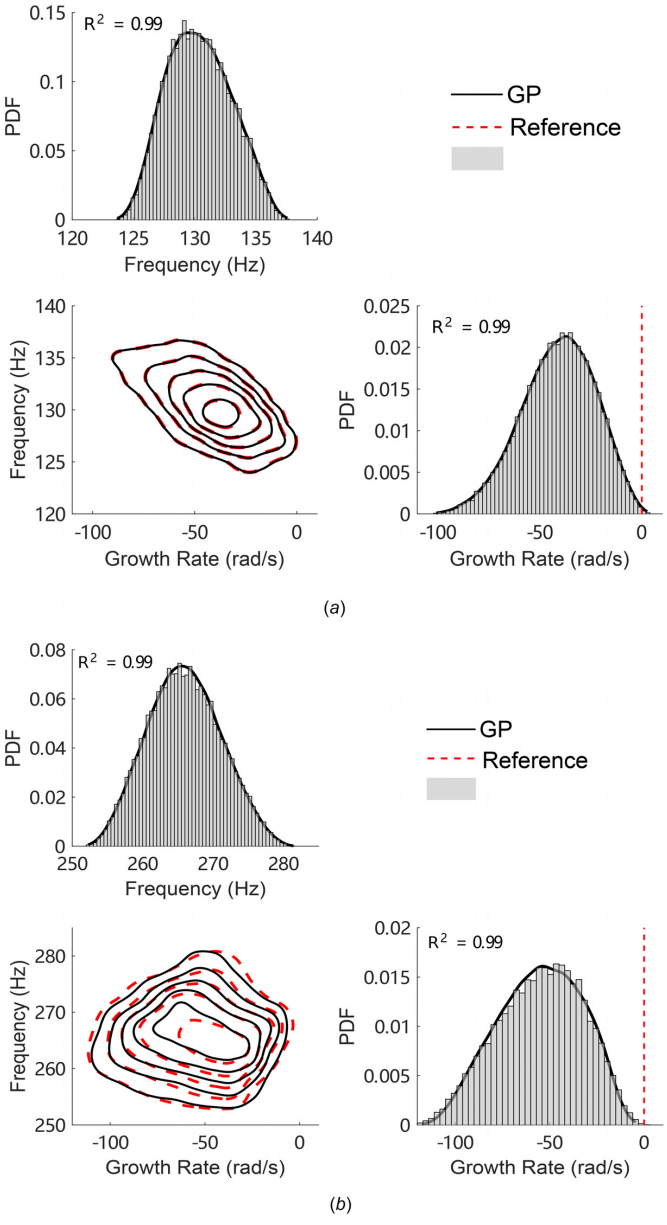


Fig. 9 PDFs of the modal frequency and growth rate predicted by the GP models and reference acoustic solver are compared. Indeed, with $\tau_c = \tau_c^{opt}$, the instability risks of both modes are basically eliminated: (*a*) ITA mode and (*b*): cavity mode.

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

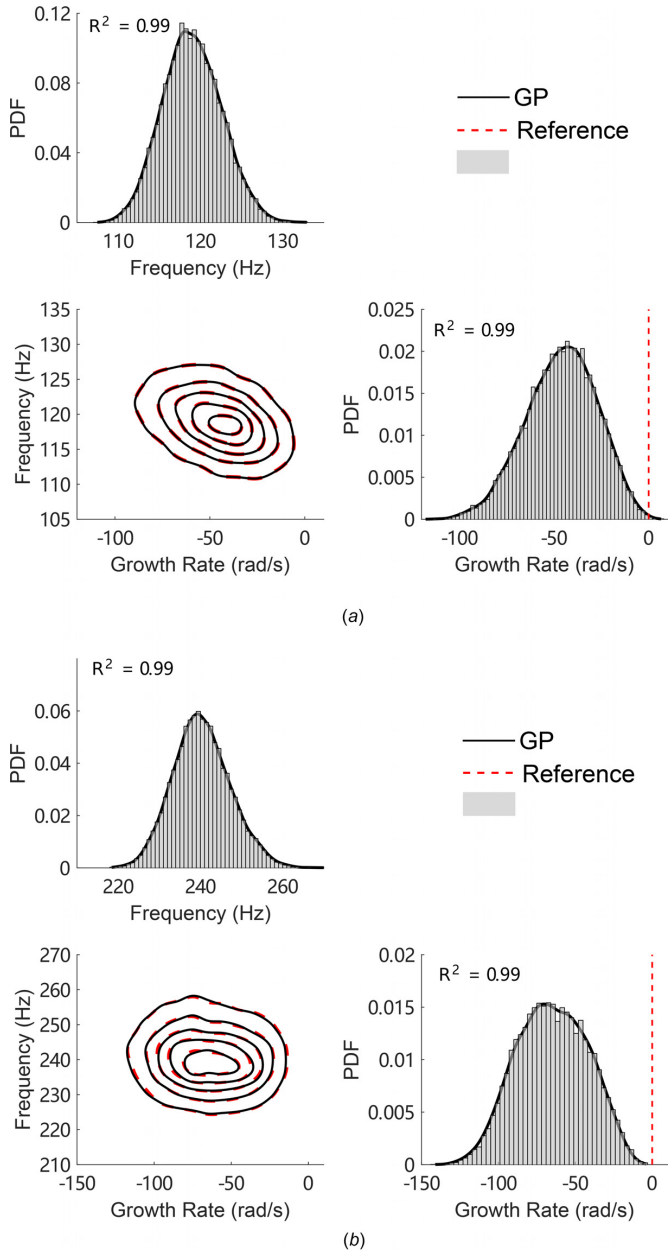


Fig. 10 PDFs of the modal frequency and growth rate predicted by the GP models and reference acoustic solver are compared. Here, $\tau_c \sim \mathcal{N}(\tau_c^{\text{opt}}, (0.05\tau_c^0)^2)$, $|R_{\text{out}}| \sim \mathcal{U}(0.7, 0.9)$ and τ_1, σ_1, τ_{s1}, τ_{s2} follow independent uniform distribution with variational ranges specified in Table 1: (*a*) ITA mode and (*b*) cavity mode.

Since $P_f^{(I)}$ and $P_f^{(C)}$ are both nonlinear functions of τ_c, Eq. (17) constitutes a single-variable, nonlinear constrained global optimization problem. Therefore, we employ a *pattern search* algorithm [42] to optimize τ_c. For each optimization iteration, $P_f^{(I)}$ and $P_f^{(C)}$ are computed by the same way as presented in Q1. Since a global optimization search usually involves more than a handful of iterations, using efficient GP models to repeatedly calculate P_f within each iteration could potentially achieve a significant reduction of computational cost. It is worth mentioning that other more advanced global optimization algorithms can also be employed here. A thorough comparison between different optimization routines is beyond the scope of this paper.

The global optimum result identified by the pattern search algorithm is $\tau_c^{\text{opt}} = 3.52\,\text{ms}$. To verify this result, we perform a Monte Carlo simulation as in Q1 by setting $\tau_c = \tau_c^{\text{opt}}$. Twenty thousand samples of $(\tau_1, \sigma_1, \tau_{s1}, \tau_{s2})$ are employed to ensure statistical

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

convergence. PDFs of frequency and growth rate for both modes are shown in Fig. 9, where the reference PDFs obtained by applying Monte Carlo directly on acoustic solver are also shown. It can be seen that, with this value of τ_c, P_f for both modes are basically zero, and the PDFs predicted by GP models match very well with the reference results.

5.3 Realistic Control Design

"Q3: in reality, we cannot perfectly control τ_c. Meanwhile, $|R_{out}|$ is also uncertain. Then how would these affect the decision made from Q2?"

As in Sec. 5.1, we consider τ_1, σ_1, τ_{s1}, and τ_{s2} to follow independent uniform distribution with variational ranges specified in Table 1. Now, we consider $|R_{\text{out}}|$ to be uncertain, following a uniform distribution in the range 0.7–0.9. For τ_c, a Gaussian distribution is assigned with mean $\bar{\tau}_c$ and standard deviation 5% of τ_c^0. Here, the standard deviation is used to mimic the situation that τ_c cannot be perfectly predicted, and statistically exhibits fluctuation around its design value $\bar{\tau}_c$. We aim to optimize $\bar{\tau}_c$ with the constraints being P_f of both modes are sufficiently small, and objective function being the modification of $\bar{\tau}_c$ from its nominal value is minimum. We can explicitly express this optimization problem as follows:

$$\begin{aligned} \min_{\bar{\tau}_c} \; & f(\bar{\tau}_c) = (\bar{\tau}_c - \tau_c^0)^2 \\ \text{subject to:} \quad & P_f^{(I)}(\tau_c) \le 0.1\% \\ & P_f^{(C)}(\tau_c) \le 0.1\% \\ & \tau_c \sim \mathcal{N}(\bar{\tau}_c, (0.05\tau_c^0)^2) \end{aligned} \tag{18}$$

The global optimum result is $\bar{\tau}_c^{\text{opt}} = 4.06\,\text{ms}$, which is obtained via the same optimization algorithm employed in Sec. 5.2. To verify this result, we perform a Monte Carlo simulation by setting $\tau_c \sim \mathcal{N}(\bar{\tau}_c^{\text{opt}}, (0.05\tau_c^0)^2)$, $|R_{\text{out}}| \sim \mathcal{U}(0.7, 0.9)$ and τ_1, σ_1, τ_{s1}, τ_{s2} follow independent uniform distribution with variational ranges specified in Table 1. Twenty thousand samples of $(\tau_1, \sigma_1, \tau_c, \tau_{s1}, \tau_{s2}, |R_{\text{out}}|)$ are used to ensure statistical convergence. PDFs of both frequency and growth rates of both modes are shown in Fig. 10, where the reference PDFs given by applying Monte Carlo directly on acoustic solver are also shown.

It can be seen that the PDFs predicted by GP models match very well with the reference results. Also, we notice that when taking into account the uncertainties in $|R_{\text{out}}|$ and imperfect control of τ_c, the optimum decision made from Q2 is not valid anymore. Instead, the design value for τ_c should increase from τ_c^{opt} to $\bar{\tau}_c^{\text{opt}}$ to eliminate the instability risk for both modes, as indicated in Fig. 10. Here, we emphasize that although we take into account more sources of uncertainties with mixed distribution type compared to Q2, our GP models manage to capture the induced variations of both eigenmodes, thus showing a high level of both flexibility and accuracy.

5.4 Tolerance Design

"Q4: given a certain threshold for risk factor, what are the maximum allowable variational ranges for τ_c and $|R_{out}|$?"

Here, we aim to investigate the compromise between the allowable uncertainty level of τ_c and $|R_{\text{out}}|$, based on the optimum result obtained from Q3. To be more specific, by considering $\tau_c \sim \mathcal{N}(\bar{\tau}_c^{\text{opt}}, (\sigma_{\tau_c})^2)$ and $|R_{\text{out}}| \sim \mathcal{U}(R_L, 0.9)$, we seek the trade-off between σ_{τ_c} (standard deviation of τ_c) and R_L (lower bound of $|R_{\text{out}}|$), both of which serve as the indications of the corresponding parameter allowable uncertainty level. For τ_1, σ_1, τ_{s1}, and τ_{s2}, we consider them to follow independent uniform distribution with variational ranges specified in Table 1. We formulate Q4 as a multiobjective optimization problem, which can be explicitly expressed as

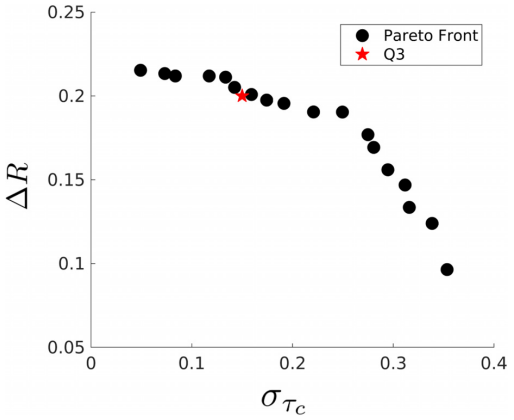


Fig. 11 Pareto front of the two objectives: maximize σ_{τ_c} and maximize the variational range for $|R_{\text{out}}|$. The *y*-axis $\Delta R = |R_{\text{out}}|^0 - R_l$. Red star marks the optimum result of Q3, from which we can see its relative location on the overall Pareto front.

$$\begin{aligned} \max_{\sigma_{\tau_c}} \; & f(\sigma_{\tau_c}) = \frac{\sigma_{\tau_c}}{\tau_c^0} \\ \min_{R_L} \; & g(R_L) = \frac{R_L}{|R_{\text{out}}|^0} \\ \text{subject to:} \quad & P_f^{(I)}(\tau_c, |R_{\text{out}}|) \le 0.1\% \\ & P_f^{(C)}(\tau_c, |R_{\text{out}}|) \le 0.1\% \\ & \tau_c \sim \mathcal{N}\left(\bar{\tau}_c, (\sigma_{\tau_c})^2\right) \\ & |R_{\text{out}}| \sim \mathcal{U}(R_L, 0.9) \end{aligned} \tag{19}$$

Notice that we are dealing with two variables, two objective functions, and $P_f^{(I)}$ and $P_f^{(C)}$ are both nonlinear functions of τ_c and $|R_{\text{out}}|$. Therefore, Eq. (19) constitutes a multivariable, multiobjective, nonlinear constrained global optimization problem. For multi-objective optimization problems, it is usually the case that not all the objectives can be optimized simultaneously. Indeed, there must be a trade-off in some way. Therefore, it would be more informative to introduce the concept of a *noninferior* solution, i.e., a solution where neither objective can be improved without degrading the other objectives. By locating a set of these noninferior solutions in the objective space, which is also called *Pareto front* [43], we can visualize the trade-offs between different objectives conveniently.

Toward that end, we employ the *multi-objective pattern search* algorithm [44] to locate the Pareto front of the two objectives. $P_f^{(I)}$ and $P_f^{(C)}$ are efficiently calculated via the same way presented in Q1.

The results are shown in Fig. 11, where the optimum result obtained from Q3 is also displayed. It can be seen that two objectives are competing with each other: toward the right of the Pareto front, a larger σ_{τ_c} would require smaller ΔR; toward the left of the Pareto front, a smaller σ_{τ_c} would allow larger variational range for $|R_{\text{out}}|$. The qualitatively described trade-offs are beneficial to understand the interaction of the allowable uncertainty level between τ_c and $|R_{\text{out}}|$.

5.5 Risk Diagram

"Q5: is it possible to construct a diagram with risk factors shown at arbitrary combinations of τ_c and $|R_{out}|$, thus encompassing answers for all the previous questions?"

Here, we aim to compute the P_f's for both modes at arbitrary location in the parameter space of τ_c and $|R_{\text{out}}|$, and illustrate them in a single risk diagram. As in Sec. 5.1, we consider τ_1, σ_1, τ_{s1}, τ_{s2} to

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

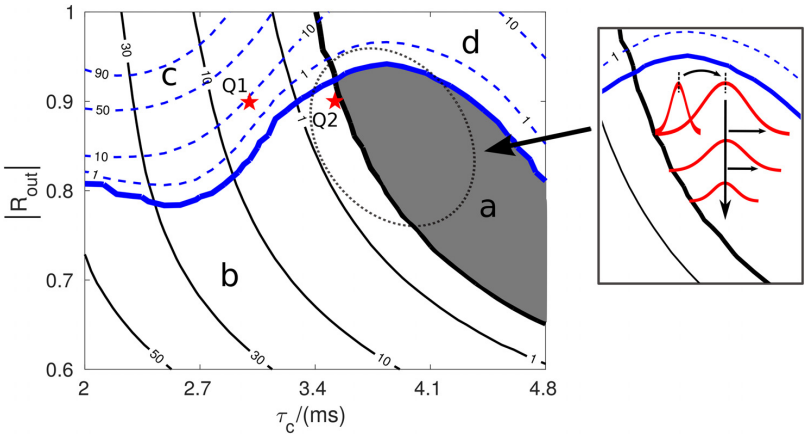


Fig. 12 Risk factor contours for both modes (blue for cavity mode and black for ITA mode). Zero risk lines (thick lines) divide the whole parameter space into four regions, with region a (shaded in grey) being the risk-free region. Results for Q1 and Q2 can be directly obtained from the diagram. For Q3 and Q4, the risk diagram offers a convenient visualization of the trade-off between τ_c's mean location, τ_c's uncertainty level and $|r_{\text{out}}|'$s uncertainty level. (For color figure please refer to online version.)

follow independent uniform distribution with variational ranges specified in Table 1.

To obtain such a risk diagram, we conduct a full factorial design [45] with a 30×30 uniformly spaced grid spanned over the parameter space of τ_c and $|R_{\text{out}}|$ (ranges indicated in Table 1), and we calculate the P_f's for both modes at each combination of $(\tau_c, |R_{\text{out}}|)$ using the same method as presented in Q1. The potential excessive computational cost induced by this repetitive task is efficiently mitigated with GP models.

The risk diagram is depicted in Fig. 12, where risk factor contours for both ITA mode (black) and cavity mode (blue) are displayed. The thick lines mark the zero risk factor contour and they divide the parameter space of $(\tau_c, |R_{\text{out}}|)$ into four regions: region a is the risk-free region, i.e., when $(\tau_c, |R_{\text{out}}|)$ values are within this region, there is no risk for both modes to be unstable even though τ_1, σ_1, τ_{s1}, τ_{s2} are uncertain. Region b is dominated by the ITA mode, i.e., in this parameter region, cavity mode is stable while ITA mode exhibits instability risk. In region c, both modes have a certain level of risk to be unstable. Finally, region d is dominated by the cavity mode, i.e., in this parameter region, ITA mode is stable while cavity mode exhibits instability risk.

For the BRS burner investigated in this current study, ITA mode and cavity mode response differently to the change of $|R_{\text{out}}|$, as observed and thoroughly discussed in Refs. [25] and [46]. This phenomenon is also reflected in the obtained risk diagram (Fig. 12): For the cavity mode, above the threshold (zero risk line), an increase of $|R_{\text{out}}|$ promotes instability risk; contrarily, for the ITA mode, increasing $|R_{\text{out}}|$ tends to reduce the instability risk. However, we caution the reader that care must be exercised when interpreting physics from Fig. 12: risk diagram is not a modal stability map, and zero risk factor lines presented here are not zero growth rate lines. Each location in the risk diagram is associated with a PDF of the modal growth rate, and the risk factor value at that location is determined by not only the mean but also the standard deviation as well as the skewness of that PDF.

This risk diagram also indicates all the results obtained from Q1 to Q4: For Q1, we can directly read the risk factor values for both modes based on the $(\tau_c, |R_{\text{out}}|)$ location on the diagram; For Q2, we can conveniently locate the τ_c^{opt} to mitigate instability risk while minimizing the modification of τ_c. For Q3 and Q4, we cannot directly obtain the answers due to the uncertainties exhibited in both τ_c and $|R_{\text{out}}|$. Nevertheless, the risk diagram shows the trade-off between $\bar{\tau}_c$, σ_{τ_c} and $\Delta|R_{\text{out}}|$: larger $\bar{\tau}_c$ accommodates larger uncertainty of τ_c and $|R_{\text{out}}|$; For fixed $\bar{\tau}_c$, larger $\Delta|R_{\text{out}}|$ requires smaller σ_{τ_c}, while smaller $\Delta|R_{\text{out}}|$ allows larger σ_{τ_c}. Such information can guide detailed optimization analysis.

5.6 Sensitivity Analysis

"Q6: if the statistical descriptions of the uncertain flame parameters are modified, how would this affect the risk diagram derived from Q5"

Up till now, we have only considered τ_1, σ_1, τ_{s1}, τ_{s2} following independent uniform distribution. In Q6, we relieve this assumption and assess the sensitivity of the zero risk lines against the statistical descriptions of the uncertain flame parameters τ_1, σ_1, τ_{s1}, τ_{s2}. Toward this end, we examine two other distributions for τ_1, σ_1, τ_{s1}, τ_{s2}, i.e., uncorrelated and correlated Gaussian. Distribution parameters for different cases are summarized in Table 2. Here, we label case A as the uniform distribution case, case B as the uncorrelated Gaussian case, and case C as the correlated Gaussian case, where the mean $\boldsymbol{M} = (\tau_1^0, \sigma_1^0, \tau_{s1}^0, \tau_{s2}^0)$ and the covariance matrix $\boldsymbol{C}$ are determined via a bootstrap analysis of surrogate time series of velocity fluctuations and flame heat release rate fluctuations. Details are given in Appendix A.

Figure 13 demonstrates the zero risk factor lines for cases A, B, and C. "Star" location is chosen for accuracy benchmark of the GP models (details in Appendix B). For Gaussian distribution cases B and C, $P_f \leq 0.1\%$ is considered as zero risk since the probability of instability is statistically sufficiently small that it basically would never happen in reality. Arrows indicate the evolution of zero risk factor lines from case A to case C. It can be seen that the distribution type of flame parameters has a significant impact on the risk diagram: compare case A with case B, when distribution type changes from uniform to Gaussian, zero risk factor lines for both modes move outward, thus expanding the

Table 2 Summary of different distributions under investigation

Parameters	Case A	Case B	Case C
τ_1	$\mathcal{U}(0.9\tau_1^0, 1.1\tau_1^0)$	$\mathcal{N}(\tau_1^0, (0.03\tau_1^0)^2)$	
σ_1	$\mathcal{U}(0.9\sigma_1^0, 1.1\sigma_1^0)$	$\mathcal{N}(\sigma_1^0, (0.03\sigma_1^0)^2)$	$\mathcal{N}(\boldsymbol{M}, \boldsymbol{C})$
τ_{s1}	$\mathcal{U}(0.9\tau_{s1}^0, 1.1\tau_{s1}^0)$	$\mathcal{N}(\tau_{s1}^0, (0.03\tau_{s1}^0)^2)$	
τ_{s2}	$\mathcal{U}(0.9\tau_{s2}^0, 1.1\tau_{s2}^0)$	$\mathcal{N}(\tau_{s2}^0, (0.03\tau_{s2}^0)^2)$	

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

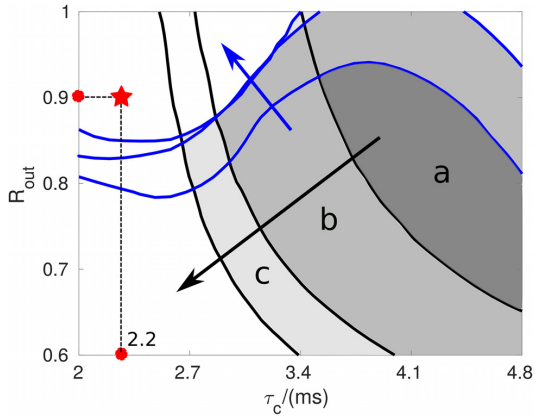


Fig. 13 Zero risk factor lines for cases a, b, and c, where blue and black lines correspond to the cavity mode and ITA mode, respectively. Risk-free zone keeps expanding from "a" (case a, uniform) to "a + b" (case b, uncorrelated Gaussian) and "a + b + c" (case c, correlated Gaussian). "star" location is chosen for accuracy benchmark of the GP models (details in Appendix B).

risk-free region to "a + b"; compare case B with case C, taking into account the correlations between the flame parameters results in further changes for cavity and ITA mode, thus gaining another risk-free region "c." This expansion of risk-free zone from case A to C is due to the fact that a Gaussian distribution requires more samples to appear near the mean compared to an uniform distribution, and a correlated Gaussian further constrains the variational possibility of the parameters compared to an uncorrelated Gaussian. It is also interesting to see that the zero risk line of the ITA mode is more sensitive to the distribution type of the uncertain flame parameters, when comparing case B with C. However, this statement may not be generalized. Risk factor is a joint effect of the mean, standard deviation, and skewness of the PDF of the modal growth rate, all of which will be simultaneously influenced by the uncertainties of the flame parameters.

6 Conclusions

In this study, we systematically explored different scenarios in thermoacoustic robust design, ranging from basic risk analysis to control design and inverse tolerance design. In addition, we provided mathematical formulation and proposed efficient solution strategy for each of these problems. We introduced the concept of a risk diagram, which displays the risk factor values of multiple thermoacoustic modes at different locations in the parameter space under investigation, and we also assessed its sensitivity to the statistical descriptions of the underlying uncertain parameters.

Throughout this work, the Gaussian process modeling technique was employed, which trains a computationally efficient yet highly accurate surrogate model to approximate the thermoacoustic solver, thus efficiently mitigating the high computational cost induced by the forward uncertainty quantification embedded in the robust design analysis. The solutions for all the robust design tasks pursued in this work are based on one-time trained GP models. Considering the number, large variational ranges and mixed distribution types of the investigated uncertain parameters, GP models managed to deliver highly accurate results for multifaceted robust design tasks. We emphasize that the flexibility and efficiency offered by the GP approach would be more prominent when combined with more sophisticated and more comprehensive thermoacoustic solvers (e.g., Helmholtz solver, Linearized Navier–Stokes solver), leveraging on the better modeling capability they offer.

Future studies will focus on the following three aspects: (1) since GP modeling is a sampling-based approach, a limited number of training samples may introduce prediction uncertainties in the GP model. Therefore, it is necessary to treat the parametric uncertainties and the surrogate modeling uncertainty concurrently in robust thermoacoustic design; (2) GP models also yield highly accurate approximations for modal frequency variations, thus opening new possibilities to include modal frequency as another constraint in robust thermoacoustic design; (3) we will employ GP method to approximate other more sophisticated thermoacoustic models, e.g., Helmholtz solver or linearized Navier–Stokes solver. Our preliminary work of extending GP to Helmholtz solver showed promising results, which will be demonstrated in the forthcoming papers. All the source code and data to reproduce the results presented in the current paper can be found online.[2]

Funding Data

- Chinese Scholarship Council (Grant No. 201606830045).

Nomenclature

FIR = flame impulse response
FTF = flame transfer function
GP = Gaussian process
h_i = FIR model coefficient
ITA = intrinsic thermoacoustic mode
N = number of FIR model coefficients
PDF = probability density function
P_f = risk factor (%)
UQ = uncertainty quantification
α = thermoacoustic modal growth rate
σ_1 = standard deviation of the distributed time lags of flame response for axial velocity perturbation
τ_c = time lag for swirl fluctuation traveling from swirler to flame base
τ_{s1} = mean of the distributed time lags of flame response for swirl fluctuation (+)
τ_{s2} = mean of the distributed time lags of flame response for swirl fluctuation (−)
τ_1 = mean of the distributed time lags of flame response for axial velocity perturbation
ω = thermoacoustic modal frequency
$|R_{out}|$ = magnitude of reflection coefficients at combustor outlet

Appendix A: Determination of Covariance Matrix

We adopt the following procedure to determine the covariance matrix $\boldsymbol{C}$ for case C in Q6:

(1) Configure the reference FIR model using $(\tau_1^0, \sigma_1^0, \tau_{s1}^0, \tau_{s2}^0, \tau_c^0)$ via Eq. (1), use the time series of velocity perturbations $(\boldsymbol{u}'_{\text{ref}})$ from Ref. [8] (Fig. 1) as the reference input signal, compute the corresponding time series of heat release rate fluctuations $(\boldsymbol{Q}'_{\text{ref}})$ via the following equation:

$$Q'_{\text{ref},n} = \sum_{k=1}^{L} h_k u'_{n-k} \tag{A1}$$

(2) Bootstrap analysis: For $i = 1, 2, \ldots, 1000$, perform
  (a) Generate a random white noise vector $\boldsymbol{Q}'_e$ from $\mathcal{N}(0, \sigma_e^2)$ so that the signal-to-noise ratio is 5 (i.e., $\sigma^2_{\boldsymbol{Q}'_{\text{ref}}}/\sigma_e^2 = 5$).
  (b) Let $\boldsymbol{Q}' = \boldsymbol{Q}'_e + \boldsymbol{Q}'_{\text{ref}}$, fix $\tau_c = \tau_c^0$, optimize $(\tau_1, \sigma_1, \tau_{s1}, \tau_{s2})$ so that the corresponding FIR model can produce $\boldsymbol{Q}'_{\text{FIR}}$ which minimize $||\boldsymbol{Q}'_{\text{FIR}} - \boldsymbol{Q}'||_2$ given $\boldsymbol{u}'_{\text{ref}}$.
  (c) Store the optimized results as $\boldsymbol{X}(i,:) = (\tau_1, \sigma_1, \tau_{s1}, \tau_{s2})^{(i)}$

(3) (Calculate the 4-by-4 covariance matrix $\boldsymbol{C} = [C_{jk}]$ as

$$C_{jk} = \frac{1}{1000-1}\sum_{i=1}^{1000}(x_{ij} - \bar{x_j})(x_{ik} - \bar{x_k}) \tag{A2}$$

[2] https://github.com/ShuaiGuo16/ASME19

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

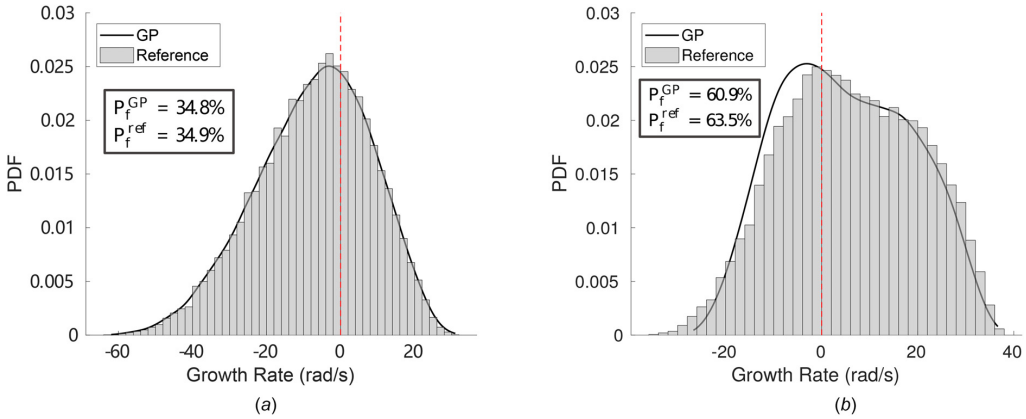


Fig. 14 PDF comparison: case A. (*a*) ITA mode and (*b*) cavity mode.

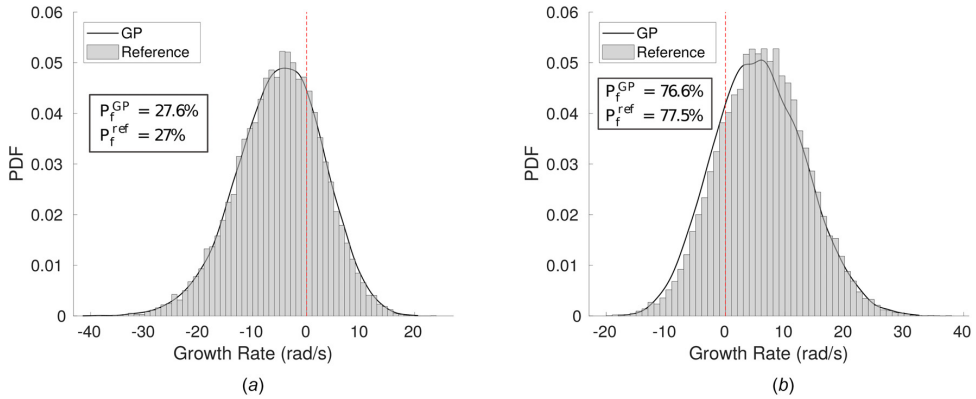


Fig. 15 PDF comparison: case B. (*a*) ITA mode and (*b*) cavity mode.

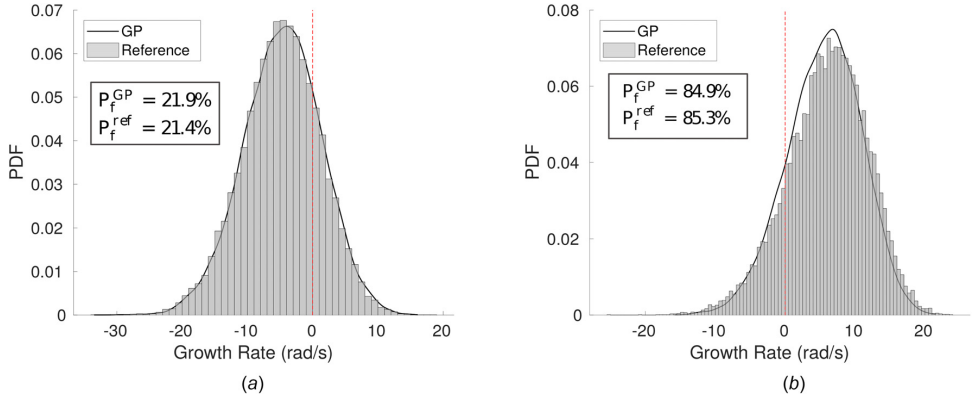


Fig. 16 PDF comparison: case C. (*a*) ITA mode and (*b*) cavity mode.

Appendix B: Accuracy Demonstrations of GP Models

We compare the PDFs of the modal growth rate predicted by the GP models and the acoustic solver. We choose $(\tau_c, |R_{\text{out}}|)$ to be the "star" location marked in Fig. 13. The results for three different cases are shown in Figs. 14–16, respectively. For each case, we employ 20,000 samples in Monte Carlo simulation to obtain the converged results.

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

Appendix C: R^2 Coefficient

We adopt R^2 coefficient in Figs. 8–10 to quantitatively assess the accuracy of GP models in approximating eigenmodes (frequency and growth rate). R^2 coefficient is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (X_i - Y_i)^2}{\sum_{i=1}^{N} (Y_i - \bar{Y}_i)^2} \tag{C1}$$

where X and Y denote the quantity computed by GP models and acoustic network solver, respectively. $\bar{Y}_i$ represents the mean value of the acoustic network solver predictions, and N represents the total number of Monte Carlo samples.

References

[1] Lieuwen, T., and Yang, V., eds., 2005, *Combustion Instabilities in Gas Turbine Engines: Operational Experience, Fundamental Mechanisms, and Modeling* (Progress in Astronautics and Aeronautics, Vol. 210), AIAA, Reston, VA.

[2] Juniper, M. P., and Sujith, R. I., 2018, "Sensitivity and Nonlinearity of Thermoacoustic Oscillations," Annu. Rev. Fluid Mech., **50**(1), pp. 661–689.

[3] Park, G. J., Lee, T. H., Lee, K. H., and Hwang, K. H., 2006, "Robust Design: An Overview," AIAA J., **44**(1), pp. 181–191.

[4] Ling, Y., Ryan, K., Asher, I., Kristensen, J., Ghosh, S., and Wang, L., 2018, "Efficient Robust Design Optimization Using Gaussian Process and Intelligent Sampling," AIAA Paper No. 2018-4175.

[5] Bade, S., Wagner, M., Hirsch, C., Sattelmayer, T., and Schuermans, B., 2013, "Design for Thermo-Acoustic Stability: Modeling of Burner and Flame Dynamics," ASME J. Eng. Gas Turbines Power, **135**(11), p. 111502.

[6] Bade, S., Wagner, M., Hirsch, C., Sattelmayer, T., and Schuermans, B., 2013, "Design for Thermo-Acoustic Stability: Procedure and Database," ASME J. Eng. Gas Turbines Power, **135**(12), p. 121507.

[7] Aguilar, J. G., and Juniper, M. P., 2018, "Adjoint Methods for Elimination of Thermoacoustic Oscillations in a Model Annular Combustor Via Small Geometry Modifications," ASME Paper No. GT2018-75692.

[8] Guo, S., Silva, C. F., Ghani, A., and Polifke, W., 2019, "Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis," ASME J. Eng. Gas Turbines Power, **141**(2), p. 021032.

[9] Ndiaye, A., Bauerheim, M., and Nicoud, F., 2015, "Uncertainty Quantification of Thermoacoustic Instabilities on a Swirled Stabilized Combustor," ASME Paper No. GT2015-44133.

[10] Magri, L., Bauerheim, M., Nicoud, F., and Juniper, M. P., 2016, "Stability Analysis of Thermo-Acoustic Nonlinear Eigenproblems in Annular Combustors—Part II: Uncertainty Quantification," J. Comput. Phys., **325**, pp. 411–421.

[11] Mensah, G. A., Magri, L., and Moeck, J. P., 2017, "Methods for the Calculation of Thermoacoustic Stability Margins and Monte Carlo-Free Uncertainty Quantification," ASME Paper No. GT2017-64829.

[12] Silva, C., Magri, L., Runte, T., and Polifke, W., 2016, "Uncertainty Quantification of Growth Rates of Thermoacoustic Instability by an Adjoint Helmholtz Solver," ASME J. Eng. Gas Turbines Power, **139**(1), p. 011901.

[13] Bauerheim, M., Ndiaye, A., Constantine, P., Moreau, S., and Nicoud, F., 2016, "Symmetry Breaking of Azimuthal Thermoacoustic Modes: The UQ Perspective," J. Fluid Mech., **789**, pp. 534–566.

[14] Avdonin, A., Jaensch, S., Silva, C. F., Češnovar, M., and Polifke, W., 2018, "Uncertainty Quantification and Sensitivity Analysis of Thermoacoustic Stability With Non-Intrusive Polynomial Chaos Expansion," Combust. Flame, **189**, pp. 300–310.

[15] Avdonin, A., and Polifke, W., 2018, "Quantification of the Impact of Uncertainties in Operating Conditions on the Flame Transfer Function With Non-Intrusive Polynomial Chaos Expansion," ASME J. Eng. Gas Turbines Power, **141**(1), p. 011020.

[16] Clarich, A., and Russo, R., 2018, "Formulations for Robust Design and Inverse Robust Design," *Uncertainty Management for Robust Industrial Design in Aeronautics* (Notes on Numerical Fluid Mechanics and Multidisciplinary Design, Vol. 140), Springer, pp. 447–462.

[17] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Kevin Tucker, P., 2005, "Surrogate-Based Analysis and Optimization," Prog. Aerosp. Sci., **41**(1), pp. 1–28.

[18] Komarek, T., and Polifke, W., 2010, "Impact of Swirl Fluctuations on the Flame Response of a Perfectly Premixed Swirl Burner," ASME J. Eng. Gas Turbines Power, **132**(6), p. 061503.

[19] Tay-Wo-Chong, L., Bomberg, S., Ulhaq, A., Komarek, T., and Polifke, W., 2012, "Comparative Validation Study on Identification of Premixed Flame Transfer Function," ASME J. Eng. Gas Turbines Power, **134**(2), p. 021502.

[20] Oberleithner, K., and Paschereit, C. O., 2016, "Modeling Flame Describing Functions Based on Hydrodynamic Linear Stability Analysis," ASME Paper No. GT2016-57316.

[21] Albayrak, A., and Polifke, W., 2016, "Propagation Velocity of Inertial Waves in Cylindrical Swirling Flow," 23nd International Congress on Sound and Vibration (ICSV23), Athens, Greece, July 10–14, pp. 1–8.

[22] Tay-Wo-Chong, L., Komarek, T., Kaess, R., Föller, S., and Polifke, W., 2010, "Identification of Flame Transfer Functions From LES of a Premixed Swirl Burner," ASME Paper No. GT2010-22769.

[23] Poinsot, T., 2017, "Prediction and Control of Combustion Instabilites in Real Engines," Proc. Combust. Inst., **36**(1), pp. 1–28.

[24] Hoeijmakers, M., Kornilov, V., Lopez Arteaga, I., de Goey, P., and Nijmeijer, H., 2014, "Intrinsic Instability of Flame-Acoustic Coupling," Combust. Flame, **161**(11), pp. 2860–2867.

[25] Emmert, T., Bomberg, S., Jaensch, S., and Polifke, W., 2017, "Acoustic and Intrinsic Thermoacoustic Modes of a Premixed Combustor," Proc. Combust. Inst., **36**(3), pp. 3835–3842.

[26] Albayrak, A., Steinbacher, T., Komarek, T., and Polifke, W., 2017, "Convective Scaling of Intrinsic Thermo-Acoustic Eigenfrequencies of a Premixed Swirl Combustor," ASME J. Eng. Gas Turbines Power, **140**(4), p. 041510.

[27] Candel, S., Durox, D., Schuller, T., Bourgouin, J. F., and Moeck, J. P., 2014, "Dynamics of Swirling Flames," Annu. Rev. Fluid Mech., **46**(1), pp. 147–173.

[28] Schneider, E., Staudacher, S., Schuermans, B., Ye, H., and Meeuwissen, T., 2007, "Real-Time Modelling of the Thermoacoustic Dynamics of a Gas Turbine Using a Gaussian Process," ASME Paper No. GT2007-27468.

[29] Chattopadhyay, P., Mondal, S., Bhattacharya, C., Mukhopadhyay, A., and Ray, A., 2017, "Dynamic Data-Driven Design of Lean Premixed Combustors for Thermoacoustically Stable Operations," ASME J. Mech. Des., **139**(11), p. 111419.

[30] Chattopadhyay, P., Mondal, S., Ray, A., and Mukhopadhyay, A., 2018, "Dynamic Data-Driven Combustor Design for Mitigation of Thermoacoustic Instabilities," ASME J. Dyn. Syst., Meas., Control, **141**(1), p. 014501.

[31] Lataniotis, C., Marelli, S., and Sudret, B., 2017, "UQLab User Manual—Kriging (Gaussian Process Modelling)," Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Zurich, Switzerland, Report No. 3.

[32] Rasmussen, C. E., and Williams, C. K. I., 2006, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning Series)*, The MIT Press, Cambridge, MA.

[33] Jones, D. R., Schonlau, M., and Welch, W. J., 1998, "Efficient Global Optimization of Expensive Black-Box Functions," J. Global Optim., **13**(4), pp. 455–492.

[34] Forrester, A. I. J., and Keane, A. J., 2009, "Recent Advances in Surrogate-Based Optimization," Prog. Aerosp. Sci., **45**(1–3), pp. 50–79.

[35] Kwon, H., Choi, S., Kwon, J.-H., and Lee, D., 2016, "Surrogate-Based Robust Optimization and Design to Unsteady Low-Noise Open Rotors," J. Aircr., **53**(5), pp. 1448–1467.

[36] Ryan, K. M., Kristensen, J., Ling, L., Ghosh, S., Asher, I., and Wang, L., 2018, "A Gaussian Process Modeling Approach for Fast Robust Design With Uncertain Inputs," ASME Paper No. GT2018-77007.

[37] Liu, H., Cai, J., and Ong, Y.-S., 2017, "An Adaptive Sampling Approach for Kriging Metamodeling by Maximizing Expected Prediction Error," Comput. Chem. Eng., **106**(2), pp. 171–182.

[38] Swiler, L., Slepoy, R., and Giunta, A., 2006, "Evaluation of Sampling Methods in Constructing Response Surface Approximations," AIAA Paper No. 2006-1827.

[39] Loeppky, J. L., Sacks, J., and Welch, W. J., 2009, "Choosing the Sample Size of a Computer Experiment: A Practical Guide," Technometrics, **51**(4), pp. 366–376.

[40] Silva, C. F., Pettersson, P., Iaccarino, G., and Ihme, M., 2018, "Generalized Chaos Expansion of State Space Models for Uncertainty Quanti Cation in Thermoacoustics," Summer Program, Center for Turbulence Research, Stanford University, Stanford CA, June 24–July 20, pp. 339–348.

[41] Ahmed, M. Y. M., and Qin, N., 2009, "Comparison of Response Surface and Kriging Surrogates in Aerodynamic Design Optimization of Hypersonic Spiked Blunt Bodies," 13th International Conference on Aerospace Sciences and Aviation Technology, Military Technical College, Cairo, Egypt, May 28–30, Paper No. ASAT-13-AE-15.

[42] Audet, C., and Dennis, J., 2002, "Analysis of Generalized Pattern Searches," SIAM J. Optim., **13**(3), pp. 889–903.

[43] Goel, T., Vaidyanathan, R., Haftka, R. T., Shyy, W., Queipo, N. V., and Tucker, K., 2007, "Response Surface Approximation of Pareto Optimal Front in Multi-Objective Optimization," Comput. Methods Appl. Mech. Eng., **196**(4–6), pp. 879–893.

[44] Custódio, A., Madeira, J., Vaz, A., and Vicente, L., 2011, "Direct Multisearch for Multiobjective Optimization," SIAM J. Optim., **21**(3), pp. 1109–1140.

[45] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., 2008, *Global Sensitivity Analysis: The Primer*, Wiley, Hoboken, NJ.

[46] Silva, C. F., Merk, M., Komarek, T., and Polifke, W., 2017, "The Contribution of Intrinsic Thermoacoustic Feedback to Combustion Noise and Resonances of a Confined Turbulent Premixed Flame," Combust. Flame, **182**, pp. 269–278.

S. Guo, C. F. Silva, and W. Polifke. Efficient Robust Design for Thermoacoustic Instability Analysis: A Gaussian Process Approach. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

Proceedings of ASME Turbo Expo 2020
Turbomachinery Technical Conference and Exposition
GT2020
September 21-25, 2020, Virtual, Online

GT2020-14434

RELIABLE CALCULATION OF THERMOACOUSTIC INSTABILITY RISK USING AN IMPERFECT SURROGATE MODEL

Shuai Guo*, Camilo F. Silva, Wolfgang Polifke
Technical University of Munich
Department of Mechanical Engineering
D-85747 Garching, Germany
Email: guo@tfd.mw.tum.de

ABSTRACT

One of the fundamental tasks in performing robust thermoacoustic design of gas turbine combustors is calculating the modal instability risk, i.e., the probability that a thermoacoustic mode is unstable, given various sources of uncertainty (e.g., operation or boundary conditions). To alleviate the high computational cost associated with conventional Monte Carlo simulation, surrogate modeling techniques are usually employed. Unfortunately, in practice it is not uncommon that only a small number of training samples can be afforded for surrogate model training. As a result, epistemic uncertainty may be introduced by such an "inaccurate" model, provoking a variation of modal instability risk calculation. In the current study, using Gaussian Process (GP) as the surrogate model, we address the following two questions: Firstly, how to quantify the variation of modal instability risk induced by the epistemic surrogate model uncertainty? Secondly, how to reduce the variation of risk calculation given a limited computational budget for the surrogate model training? For the first question, we leverage on the Bayesian characteristic of the GP model and perform correlated sampling of the GP predictions at different inputs to quantify the uncertainty of risk calculation. We show how this uncertainty shrinks when more training samples are available. For the second question, we adopt an active learning strategy to intelligently allocate training samples, such that the trained GP model is highly accurate particularly in the vicinity of the zero growth rate contour. As a result, a more accurate and robust modal instability risk calculation is obtained without increasing the computational cost of surrogate model training.

*Address all correspondence to this author.

NOMENCLATURE

FIR	Flame Impulse Response
UQ	Uncertainty Quantification
GP	Gaussian Process
PDF	Probability Density Function
h_i	FIR model coefficient
N	Number of FIR model coefficients
P_f	Risk factor (%)
α	Thermoacoustic modal growth rate
τ_1	Mean of the distributed time lags of flame response for axial velocity perturbation
σ_1	Standard deviation of the distributed time lags of flame response for axial velocity perturbation
τ_c	Time lag for swirl fluctuation traveling from swirler to flame base
τ_{s1}	Mean of the distributed time lags of flame response for swirl fluctuation (+)
τ_{s2}	Mean of the distributed time lags of flame response for swirl fluctuation (-)
$\lvert R_{out}\rvert$	Magnitude of reflection coefficients at combustor outlet

INTRODUCTION

The thermoacoustic behavior of a gas turbine combustor usually exhibits a high level of sensitivity to uncertain parameters such as operation conditions and acoustic boundary conditions [1]. As a result, conventional thermoacoustic instability predictions, which employ only the nominal values of the system parameters when calculating the modal frequencies and growth rates, may yield unreliable results, *e.g.*, a combustor that

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

is predicted as stable with nominal input parameters may become unstable when inputs deviate only slightly from their nominal values. Consequently, uncertainty quantification (UQ) analysis, which focuses on quantifying uncertainties on outputs given uncertain inputs, is essential to achieve reliable thermoacoustic instability predictions and constitutes a fundamental step towards robust thermoacoustic design.

One of the main tasks in thermoacoustic UQ analysis is calculating the modal instability risk (P_f), *i.e.*, the probability that a thermoacoustic mode is unstable, given various uncertain parameters and their corresponding probability density functions (PDF). To calculate P_f induced by those uncertainties, a straightforward method is to perform Monte Carlo simulation: firstly a large number of samples (at the order of 10^4) are drawn according to the PDFs of uncertain inputs. Then for each sample, its corresponding modal growth rate values are calculated via thermoacoustic solvers (*e.g.*, acoustic network model [2], Helmholtz solver [3], *etc*). Finally P_f can be simply estimated as the number of samples with positive modal growth rate values divided by the total sample number.

It is obvious that a direct application of Monte Carlo simulation using expensive thermoacoustic solvers would induce prohibitive computational cost, thus rendering it infeasible for realistic UQ analysis. In order to solve the efficiency problem, various surrogate modeling techniques have been explored in previous studies. In this framework, firstly a small number of training samples (around 1% of full Monte Carlo sample size) are carefully selected according to the input PDFs. Their corresponding outputs (growth rate values) are calculated via thermoacoustic solver. Then machine learning techniques are usually employed to train a cheap surrogate model, based on the training samples and their responses, to approximate the thermoacoustic solver. Subsequently, Monte Carlo simulation can be applied on the surrogate model to derive P_f with negligible cost. Typical examples of surrogate modeling techniques include polynomial regression [4], adjoint method [3, 5, 6], polynomial chaos expansion [7–9], as well as Gaussian Process approach [10].

Despite the remarkable progress made in terms of accelerating P_f calculation, the uncertainty contained in the surrogate model itself has been largely ignored in previous studies. This issue becomes particularly prominent as the parameter variation ranges become larger and the fidelity of the thermoacoustic solver improves: a larger number of training samples is then required to build a surrogate model that is accurate over a larger parameter space. Meanwhile, every calculation of a training sample induces significant computational cost. Consequently, in many practical cases, only a small training sample size can be afforded, which leads to a potentially "inaccurate" surrogate model. As a result, epistemic uncertainty, which is defined as uncertainty induced by lack of knowledge or simplifications and can be reduced by collecting more information [11], is introduced by such a model. In this situation, P_f is not a deterministic value anymore, but rather an uncertain variable itself. To summarize, in the framework of surrogate model based uncertainty quantification, P_f exists due to the presence of system parameter uncertainty; P_f varies due to the presence of surrogate model uncertainty.

Among various surrogate modeling techniques, the *Gaussian Process* (GP) approach stands out. This is due to the fact that GP not only provides predicted values at unsampled inputs, but also estimates of the prediction variances, thanks to its Bayesian nature [12]. This unique feature of GP allows us to quantify the associated epistemic uncertainty when using GP to calculate P_f.

Therefore, in the current study, we employ GP as the surrogate model. We aim to address the following two questions that are closely related to a realistic application of GP approach in thermoacoustic UQ analysis: (Q1) how to quantify the variation of modal instability risk P_f induced by the epistemic GP model uncertainty? Although surrogate modeling techniques have become popular in the thermoacoustic community to perform risk analysis, the extra uncertainties induced by imperfectly-trained surrogate models have been largely ignored until now. To our best knowledge, our work is the first to quantify such uncertainties as well as the impact of these uncertainties on the variation of thermoacoustic risk calculation. (Q2) How to reduce the variation of P_f given a limited computational budget for the GP model training? Q2 directly builds on Q1: after understanding how to quantify the uncertainty of risk calculation, we now focus on how to reduce this uncertainty. This point is of particular interest to industrial applications. For the first question, we leverage on the prediction variance naturally offered by the GP model and perform correlated sampling of the GP predictions at different inputs [13] to quantify the uncertainty of P_f calculation. For the second question, we apply active learning methods [12] from the machine learning community, where the training samples are sequentially allocated, such that the trained GP model is gradually refined particularly in the vicinity of the zero growth rate contour. As a result, we are able to obtain a more accurate and robust P_f calculation without increasing the computational cost for surrogate model training.

The current work is closely related to our previous work [10], where our focus was to employ the GP approach to efficiently address various robust thermoacoustic design tasks. In that work, P_f calculations were carried out with the assumption that GP models have negligible epistemic uncertainty. In our current work, we will drop this assumption and investigate how GP model uncertainty impacts the P_f calculation, as well as how to reduce this impact given a limited training sample budget. It is worth mentioning that the procedure of a reliable calculation of P_f established in the current work can be seamlessly integrated into the robust design analysis investigated in our previous work [10], without compromising the associated workflow developed there.

This paper is organized as follows. Section "Thermoacoustic Framework" outlines the thermoacoustic problem under in-

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

vestigation. Section "Gaussian Process Modeling" presents the technical details of the employed GP approach and gives the formula for estimating the prediction variance. Section "Uncertainty of P_f Calculation" focuses on quantifying the P_f calculation uncertainty induced by the epistemic GP model uncertainty. Procedures are introduced first, followed by a case study to show how the uncertainty of P_f calculation shrinks when more GP training samples are available. Section "Reliable P_f Calculation" introduces the active learning scheme for adaptive GP model training. The case study in the previous section are revisited and the effectiveness of the adopted active learning scheme is demonstrated. All the code and data to produce the results presented in the current paper can be found at https://github.com/tfd/GuoSilva20a_ASME.

THERMOACOUSTIC FRAMEWORK

The thermoacoustic problem investigated in the current study is taken from our previous work [10]. Here, we briefly review the key aspects to ensure the self-containedness of the present paper.

For the combustor configuration, we investigate a turbulent premixed swirl combustor test rig [14, 15]. An equivalence ratio of 0.77 of perfectly premixed methane-air mixture and a thermal power of 30kW are considered as the operation conditions.

We employ a low-order acoustic network model (shown in Fig. 1) to calculate the eigenmode of the burner test rig. Geometry, thermodynamic parameters as well as modeling details can be found in [2, 10]. The reflection coefficient at the combustor exit is modeled as $|R_{out}|e^{i\pi}$.

A flame impulse response (FIR) model describing the swirling flame dynamics is adopted in the current work. This model was proposed by Komarek *et al.* [14] and validated by Oberleithner *et al.* [16]. This model decomposes the shape of the flame impulse response into three Gaussian distributions: one positive (+) Gaussian to account for the flame response to axial velocity fluctuations, and one (+) combined with another (-) Gaussian to describe the response to swirl number fluctuations. The parametrization of the model is shown in Fig. 2. Here, τ_c represents the time needed for an inertial wave to propagate from the swirler to the flame base [17]. τ_1 and σ_1 represent the mean and standard deviation of the flame heat release response under an impulse axial velocity perturbation, respectively. τ_{s1} and τ_{s2} represent the characteristic time lags for the flame response to the swirl number fluctuations. The standard deviations of the associated positive and negative Gaussian functions are assumed to be one third of τ_{s1} and τ_{s2}, respectively. Therefore, we can express the value of the FIR coefficient h_k as [14]:

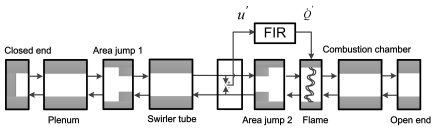


FIGURE 1: SKETCH OF ACOUSTIC NETWORK MODEL, FLOW FROM LEFT TO RIGHT.

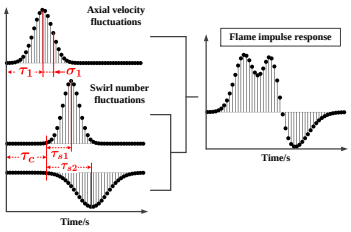


FIGURE 2: SKETCH OF THE EMPLOYED FIR MODEL FOR THE FLAME DYNAMICS.

$$h_k = \frac{\Delta t}{\sigma_1\sqrt{2\pi}} e^{-\frac{(k\Delta t-\tau_1)^2}{2\sigma_1^2}} + \frac{\Delta t}{\sigma_2\sqrt{2\pi}} e^{-\frac{(k\Delta t-\tau_2)^2}{2\sigma_2^2}} - \frac{\Delta t}{\sigma_3\sqrt{2\pi}} e^{-\frac{(k\Delta t-\tau_3)^2}{2\sigma_3^2}}, \quad k \in [1,2,...N], \tag{1}$$

where N is the number of FIR coefficients, and

$$\tau_2 = \tau_c + \tau_{s1}, \quad \tau_3 = \tau_c + \tau_{s2}, \quad \sigma_2 = \tau_{s1}/3, \quad \sigma_3 = \tau_{s2}/3, \tag{2}$$

In the current study, we consider six uncertain parameters: five flame model parameters $\tau_1, \sigma_1, \tau_c, \tau_{s1}, \tau_{s2}$, and one acoustic boundary condition, *i.e.*, the magnitude of the reflection coefficient at the combustor outlet $|R_{out}|$. Their nominal values and their respective uncertain ranges are taken from [10] and summarized in Tab. 1. In practice, it is preferable to construct a GP model that is valid over a large parameter space, so that it can be reused for different tasks in the analysis process (*e.g.*, various robust design tasks detailed in [10]). Therefore, we will build the GP model on the full parameter space as described in Tab. 1. In section "Uncertainty of P_f Calculation", we will re-investigate 'Q1: Risk Analysis" in [10] to assess the impact of epistemic surrogate model uncertainty on the calculation of P_f. Since only τ_1,

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

TABLE 1: UNCERTAINTY INFORMATION OF THE INVESTIGATED PARAMETERS.

Parameters		Nominal	Range
Flame (units: ms)	τ_1	$\tau_1^0 = 2.85$	$0.9\tau_1^0 \sim 1.1\tau_1^0$
	σ_1	$\sigma_1^0 = 0.7$	$0.9\sigma_1^0 \sim 1.1\sigma_1^0$
	τ_c	$\tau_c^0 = 3$	$2 \sim 4.8$
	τ_{s1}	$\tau_{s1}^0 = 1.8$	$0.9\tau_{s1}^0 \sim 1.1\tau_{s1}^0$
	τ_{s2}	$\tau_{s2}^0 = 3.3$	$0.9\tau_{s2}^0 \sim 1.1\tau_{s2}^0$
Acoustic BC	$\lvert R_{out}\rvert$	$\lvert R_{out}\rvert^0 = 0.9$	$0.6 \sim 1$

σ_1, τ_{s1} and τ_{s2} are considered as the uncertain parameters in that case study, we can effectively test the flexibility and robustness of the GP modeling approach.

GAUSSIAN PROCESS MODELING

In this section, we summarize the key features of GP surrogate modeling that are relevant for our current study. For an comprehensive introduction of GP approach, readers are referred to [18]. For the application of GP approach in the context of thermoacoustic instability analysis, readers are referred to [10, 19–21].

GP approach models the output $f(\boldsymbol{x})$ of the high fidelity model at $\boldsymbol{x}$ as the realization of a Gaussian process:

$$f(\boldsymbol{x}) = \beta + Z(\boldsymbol{x}), \tag{3}$$

where β is a constant value and $Z(\boldsymbol{x})$ corresponds to a Gaussian process with zero mean and covariance defined as:

$$Cov[Z(\boldsymbol{x}^i,\boldsymbol{x}^j)] = \sigma^2 R(\boldsymbol{x}^i,\boldsymbol{x}^j), \tag{4}$$

where σ^2 is the process variance and $R(\boldsymbol{x}^i,\boldsymbol{x}^j)$ is the correlation function between any two locations $\boldsymbol{x}^i$ and $\boldsymbol{x}^j$ in the input parameter space. In the present study, we adopt a Gaussian correlation function to describe $R(\boldsymbol{x}^i,\boldsymbol{x}^j)$:

$$R(\boldsymbol{x}^i,\boldsymbol{x}^j) = \exp[-\sum_{k=1}^{M} \theta_k (x_k^i - x_k^j)^2], \tag{5}$$

where M denotes the dimension of the input (*i.e.,* number of input parameters), and θ_k is a hyper-parameter that controls the correlation strength between the points within dimension k.

Based on a set of training samples $\boldsymbol{X}_D = [\boldsymbol{x}^1,...,\boldsymbol{x}^N]^T$ and their corresponding responses $\boldsymbol{Y}_D = [f(\boldsymbol{x}^1),...,f(\boldsymbol{x}^N)]^T$, we can estimate β and σ^2 according to [18]:

$$\widehat{\beta} = (\mathbf{1}^T \boldsymbol{R}_D^{-1} \mathbf{1})^{-1} \mathbf{1}^T \boldsymbol{R}_D^{-1} \boldsymbol{Y}_D, \tag{6}$$

$$\widehat{\sigma}^2 = \frac{1}{N} (\boldsymbol{Y}_D - \mathbf{1}\widehat{\beta})^T \boldsymbol{R}_D^{-1} (\boldsymbol{Y}_D - \mathbf{1}\widehat{\beta}), \tag{7}$$

where $\mathbf{1}$ is a vector of ones of dimension M. $\boldsymbol{R}_D$ is the N-by-N correlation matrix between training samples in $\boldsymbol{X}_D$. However, since Eq. (6) and (7) depends on θ_k through correlation matrix $\boldsymbol{R}_D$, it is first required to estimate θ_k using maximum likelihood estimation:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left[-\frac{N}{2} \ln(\widehat{\sigma}^2) - \frac{1}{2} \ln(|\boldsymbol{R}_D|) \right]. \tag{8}$$

Finally, the GP model predictions $\boldsymbol{Y}_p$ at unknown locations $\boldsymbol{X}_p = [\boldsymbol{x}^1,...,\boldsymbol{x}^L]^T$ follow a multivariate normal distribution, with the mean predictions $\boldsymbol{\mu}(\boldsymbol{Y}_p)$ and prediction covariance $cov(\boldsymbol{Y}_p)$ given as:

$$\boldsymbol{\mu}(\boldsymbol{Y}_p) = \mathbf{1}\widehat{\beta} + \boldsymbol{R}_{PD}^T \boldsymbol{R}_D^{-1} (\boldsymbol{Y}_D - \mathbf{1}\widehat{\beta}), \tag{9}$$

$$cov(\boldsymbol{Y}_p) = \widehat{\sigma}^2 (\boldsymbol{R}_P - \boldsymbol{R}_{PD}^T \boldsymbol{R}_D^{-1} \boldsymbol{R}_{PD}), \tag{10}$$

where $\boldsymbol{R}_{PD}$ represents the N-by-L correlation matrix between the prediction inputs $\boldsymbol{X}_P$ and the training inputs $\boldsymbol{X}_D$. $\boldsymbol{R}_P$ represents the L-by-L correlation matrix between the prediction inputs $\boldsymbol{X}_P$.

Specifically, when GP prediction y_p at only a single point $\boldsymbol{x}$ is required, the corresponding prediction mean $\mu(y_p)$ and variance $\sigma^2(y_p)$ can be directly derived from Eq. (9) and (10) as:

$$\mu(y_p) = \widehat{\beta} + \boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R}_D^{-1} (\boldsymbol{Y}_D - \mathbf{1}\widehat{\beta}), \tag{11}$$

$$\sigma^2(y_p) = \widehat{\sigma}^2 (1 - \boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R}_D^{-1} \boldsymbol{r}(\boldsymbol{x})). \tag{12}$$

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

In section "Uncertainty of P_f Calculation", Eq. (9)-(10) are used to generate realizations of GP predictions to facilitate the quantification of P_f calculation uncertainty. In section "Reliable P_f Calculation", Eq. (11)-(12) are used to derive the learning function to facilitate adaptive GP model training.

UNCERTAINTY OF P_f CALCULATION

In this section, we investigate the procedure to quantify the P_f calculation uncertainty and demonstrate the procedure through a case study. We start with the case set-up, followed by detailing the uncertainty quantification procedure. Finally, we examine how the P_f calculation uncertainty evolves once more training samples are made available for GP model training.

Case Set-up

In the current work, we focus on one of the eigenmodes investigated in [10]: the quarter wave mode of the combustor, which is labeled as a *cavity mode* [22]. Given nominal values of the flame model parameters and $|R_{out}|$, the nominal values of the modal frequencies (ω) and growth rates (α) of the cavity mode are calculated via the acoustic network model (Fig. 1), *i.e.*, $(\omega = 287.5Hz, \alpha = -27.7rad/s)$ for cavity mode.

When uncertainties are present in flame model parameters and $|R_{out}|$, the cavity mode may exhibit a certain level of risk to be unstable. Here, we re-investigate the problem outlined in "Q1: Risk Analysis" in [10]. For that problem, we need to calculate the P_f of cavity mode when τ_c and $|R_{out}|$ are fixed at their nominal values, while the rest of the parameters follow independent uniform distributions with the parameter range indicated in Tab. 1.

In GP-based Monte Carlo simulation, P_f of a thermoacoustic mode can be calculated as [13]:

$$P_f = \frac{1}{n_{MC}} \sum_{i=1}^{n_{MC}} I(y_p(\boldsymbol{x}^i)), \quad y_p(\boldsymbol{x}^i) \in \boldsymbol{Y}_p, \tag{13}$$

where n_{MC} denotes the number of Monte Carlo samples, $\boldsymbol{x}^i$ is the ith Monte Carlo sample, $y_p(\boldsymbol{x}^i)$ represents the GP prediction of the modal growth rate value given the ith Monte Carlo sample. $I(x)$ constitutes an indicator function, which equals 1 if $x > 0$ and equals 0 if $x \leqslant 0$. Note that there are other more sophisticated approaches (*e.g.* importance sampling [23], subset sampling [24], *etc.*) to compute P_f. However, we choose direct Monte Carlo simulation due to the fact that it is extremely easy to implement and versatile, therefore making it especially favorable for industrial applications. In addition, by adopting straight-forward Monte Carlo method, we can better focus on quantifying the impact of GP model uncertainty on the variation of risk calculation, which is the main purpose of our current work.

Here, due to the stochastic nature of the GP model, each $y_p(\boldsymbol{x}^i), i = 1, ..., n$ is a random variable, following a normal distribution with the mean and variance given by Eq. (11)-(12). This is the marginal distribution of the ensemble of all the $y_p(\boldsymbol{x}^i)$'s, which follows a multivariate normal distribution with the mean and covariance given by Eq. (9)-(10). In our previous work [10], we assume that the GP model uncertainty is negligible, *i.e.*, we directly substitute $y_p(\boldsymbol{x}^i), i = 1, ..., n$ with $\mu(y_p(\boldsymbol{x}^i)), i = 1, ..., n$ in Eq. (13), thus leading to a deterministic value of P_f. However, when $cov(\boldsymbol{Y}_p)$ (determined by Eq. (10)) is sufficiently large, the training sample size is insufficient and GP model predictions exhibit non-negligible uncertainties. In that case, $y_p(\boldsymbol{x}^i)$'s have to be treated as a random vector. As a result, P_f would not be a deterministic value, but rather a random variable with an associated PDF. In the following section, we adopt an efficient procedure proposed by [13] to account for the randomness of $y_p(\boldsymbol{x}^i)$'s and using Monte Carlo technique to derive the PDF of P_f for the cavity mode.

Quantification of GP Model Uncertainty

To determine the uncertainty of P_f, we employ a direct Monte Carlo approach: Firstly, we generate realizations of GP predictions $\boldsymbol{Y}_p$ as Monte Carlo samples. Then, for each sample of $\boldsymbol{Y}_p$, we calculate its corresponding P_f value via Eq. (13). Finally, based on an ensemble of P_f values, we can obtain its histogram, thus quantifying the uncertainty of P_f induced by the epistemic uncertainty of the GP model.

However, the Monte Carlo sample size n_{MC} in Eq. (13) is usually very large (at the order of 10^4). As a result, the covariance matrix of $\boldsymbol{Y}_p$ would be at the order of 10^4-by-10^4. Since drawing samples from a multivariate normal distribution usually involves performing eigenvalue decomposition on the covariance matrix [25], the resulting computational cost of drawing samples of $\boldsymbol{Y}_p$ would be prohibitively expensive.

To resolve this efficiency issue, Nannapaneni *et al.* [13] proposed a procedure to reduce the problem dimensionality, *i.e.*, the number of random variables in $\boldsymbol{Y}_p$ that has to be taken into account. The fundamental idea is illustrated in Fig. 3: since the output of the indicator function $\boldsymbol{I}(x)$ only depends on the sign of its input, not all the variations of $y_p(\boldsymbol{x}^i)$ will lead to a change of P_f. Figure 3 illustrates two potential scenarios: For the GP prediction of the first Monte Carlo sample $y_p(\boldsymbol{x}^1)$, its mean and variance values can be calculated via Eq. (11)-(12). Since more than 99.7% (three sigma) of the associated PDF lies in the unstable side, our GP model is extremely confident to classify this sample to be unstable, *i.e.,* $\boldsymbol{I}(y_p(\boldsymbol{x}^1)) = 1$, although its exact growth rate value may not be estimated precisely. In another word, there is negligible probability that our GP model makes a mistake in estimating the value of $\boldsymbol{I}(y_p(\boldsymbol{x}^1))$. On the other hand, for the GP prediction of the second Monte Carlo sample $y_p(\boldsymbol{x}^2)$, our GP model no longer exhibits a high level of confidence regarding the stabil-

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

ity nature of the sample: there are large portions of its PDF that lie in both the unstable side and stable side, respectively. Consequently, there is a high probability that our GP model makes a mistake in estimating the value of $\boldsymbol{I}(y_p(\boldsymbol{x}^2))$.

Based on the probability that our GP model makes a mistake in estimating the sign of $y_p(\boldsymbol{x}^i)$, we can partition the n_{MC} Monte Carlo samples into two groups: The first group corresponds to predictions $y_p(\boldsymbol{x}^j), j = 1,...,n_1$, for which the probability of making mistakes in estimating the signs of $\boldsymbol{I}(y_p(\boldsymbol{x}^j)), j = 1,...,n_1$ is sufficiently low (*i.e.*, below 0.3%). For this group, $y_p(\boldsymbol{x}^j)$ are not treated as random variables and the mean predictions $\boldsymbol{\mu}(y_p(\boldsymbol{x}^j)), j = 1,...,n_1$ are directly used in Eq. (13). The remaining samples $y_p(\boldsymbol{x}^j), j = 1,...,n_2$ (with $n_1 + n_2 = n_{MC}$) belong to the second group. They are treated as random variables, following a multivariate normal distribution, with mean and covariance determined by Eq. (9) -(10). Therefore, Eq. (13) can be rewritten as:

$$P_f = \frac{1}{n_{MC}}\left(\sum_{j=1}^{n_1} I(\boldsymbol{\mu}(y_p(\boldsymbol{x}^j))) + \sum_{j=1}^{n_2} I(y_p(\boldsymbol{x}^j))\right), \tag{14}$$

To achieve the partition, we adopt the following classification function proposed by Echard *et al.* [12]:

$$U(\boldsymbol{x}^i) = \frac{|\mu(y_p(\boldsymbol{x}^i))|}{\sigma(y_p(\boldsymbol{x}^i))}. \tag{15}$$

Here, $U(\boldsymbol{x}^i)$ is used to estimate the probability of making a mistake in estimating the sign of $y_p(\boldsymbol{x}^i)$ (see Appendix A). The first group consists of samples with a U-value larger than 3, and the rest of the samples fall into the second group.

The following procedure summarizes the procedure to derive the uncertainty of P_f calculation. We assume the GP model has already been trained.

(1) Generate n_{MC} Monte Carlo samples according to the distributions of uncertain flame and acoustic parameters;

(2) Use GP model to calculate the mean and variance values of each sample (Eq. (11)-(12));

(3) Calculate U-value for each sample (Eq. (15)). All the samples with a U-value larger than 3 fall into group 1, the rest of the samples fall into group 2. Define n_1 and n_2 as the total sample number in two groups, respectively. Re-index the samples in the first group as $g_1(1),...g_1(n_1)$. Re-index the samples in the second group as $g_2(1),...g_2(n_2)$.

(4) Using Eq. (9)-(10) to calculate the mean and covariance matrix of the random vector $[y_p(\boldsymbol{x}^{g_2(1)}),...,y_p(\boldsymbol{x}^{g_2(n_2)})]$. Generate L realizations of $[y_p(\boldsymbol{x}^{g_2(1)}),...,y_p(\boldsymbol{x}^{g_2(n_2)})]$. For each realization, using Eq. (14) to calculate the corresponding value of $P_f^k, k = 1,...,L$;

(5) Construct the histogram of P_f and extract desired statistical indices.

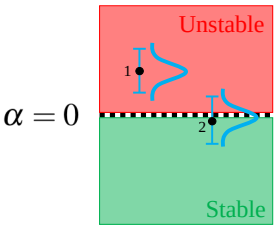


FIGURE 3: AN ILLUSTRATION OF THE FUNDAMENTAL IDEA TO PARTITION ALL THE MONTE CARLO SAMPLES INTO TWO GROUPS. TWO SAMPLES ARE DEPICTED HERE, AND THEIR ASSOCIATED GP PREDICTION PDFs ARE ALSO PLOTTED. SAMPLE 1 BELONGS TO THE FIRST GROUP SINCE ITS PREDICTION VARIATIONS WOULD NOT CHANGE P_f. WHILE FOR SAMPLE 2, IT CAN HOP BETWEEN STABLE AND UNSTABLE BY CHANCE, THUS DIRECTLY AFFECTING P_f CALCULATION. THEREFORE, WE ONLY NEED TO CONSIDER SAMPLES IN GROUP 2 AS RANDOM VARIABLES, THUS SIGNIFICANTLY IMPROVING THE UQ EFFICIENCY.

Case study

We train our GP model based on the full parameter space described in Tab. 1. To generate training samples, we adopt the Halton sampling technique [26], which is a low-discrepancy sampling method with excellent space-filling property. In our current study, we label this sampling method as *passive sampling*, to indicate that this sampling technique merely tries to distribute the training samples as evenly as possible in the parameter space. This is in direct contrast with the *adaptive sampling* technique, which we will discuss in the following section, where the sampling method exploits the already learned "landscape" of the underlying function, and adaptively allocate samples to the region where prediction accuracy is not satisfactory.

In line with the problem setting in "Q1: Risk Analysis" in [10], we consider four uncertain parameters: τ_1, σ_1, τ_{s1} and τ_{s2}, which follow independent uniform distribution with the parameter ranges indicated in Tab. 1. τ_c and $|R_{out}|$ are fixed at their nominal values. Here we use 102 training samples (same as in [10]) to train a GP model and follow the procedure outlined in section "Quantification of GP model Uncertainty" to derive the PDF of P_f. Monte Carlo sample size is 20000 for both n_{MC} and L. After partitioning the Monte Carlo samples, only 1630 out

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

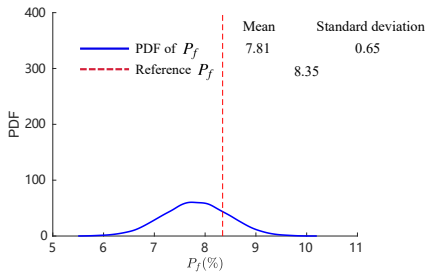


FIGURE 4: THE PDF OF P_f IS CALCULATED FOLLOWING THE PROCEDURE OUTLINED IN SECTION "Quantification of GP model Uncertainty". REFERENCE VALUE OF P_f IS CALCULATED BY APPLYING MONTE CARLO DIRECTLY ON THE ACOUSTIC NETWORK MODEL (Fig. 1). ALTHOUGH REFERENCE VALUE IS COVERED BY THE PDF, P_f PREDICTION EXHIBITS A CERTAIN LEVEL OF VARIATION, WHICH IS INDUCED BY A LACK OF SAMPLES FOR GP MODEL TRAINING.

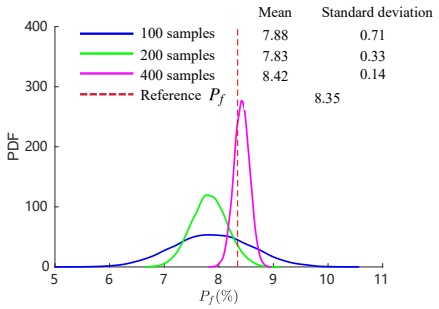


FIGURE 5: MORE TRAINING SAMPLES LEAD TO A MORE ACCURATE AND ROBUST CALCULATION OF P_f.

of 20000 samples belong to the second group, resulting in significantly lowered computational cost associated with generating realizations.

The calculated PDF of P_f for the cavity mode is shown in Fig. 4. Also shown in Fig. 4 is the reference value for P_f, which is calculated by the network model (Fig. 1) using the same n_{MC} Monte Carlo sample. It can be seen that although the reference value is covered by the PDF, P_f prediction exhibits a high level of variation, ranging from 5.5% to 10%, indicating a non-negligible epistemic GP model uncertainty induced by a lack of training samples. No GP model uncertainty was considered in our previous work [10]. The predicted P_f has a deterministic value of 7.6%, which approximately corresponds to the mean of the P_f PDF shown in Fig. 4.

It is worth mentioning that the calculation of the reference P_f value via direct Monte Carlo is entirely feasible with an acoustic network model, thus making it possible to benchmark our current approach to quantify the uncertainty of P_f calculation. For other more computational intensive acoustic models, like the ones characterized by the Helmholtz equation or the Linearized Navier-Stokes equation, direct Monte Carlo would no longer be an option, then a Gaussian Process surrogate model with quantified uncertainty becomes really valuable for a robust calculation of P_f.

It would also be interesting to see how the PDF of P_f varies when the number of training samples is gradually increased. Figure 5 demonstrates the PDFs of P_f when the number of training samples are 100, 200 and 400, respectively. It can be seen clearly that with more training samples, the mean value of P_f prediction moves closer to the reference P_f value, thus indicating an improvement of accuracy in P_f calculation; in addition, the PDF becomes more concentrated, indicating that the epistemic uncertainty of GP model reduces with more available information (training samples).

In the next section, we will investigate how to intelligently train the GP model, so that the improvement of both accuracy and robustness of P_f calculation can be maximized.

RELIABLE P_f CALCULATION

In this section, we aim to employ an adaptive sampling scheme to enrich the training samples $\boldsymbol{X}_D$ sequentially, thus enabling us to iteratively increase the accuracy and robustness of the GP model for P_f calculation, while keeping the number of runs of the acoustic solver as low as possible. We start with the motivation for adaptive sampling, followed by outlining the detailed procedures. Subsequently, we revisit the case study investigated in the previous section and demonstrate the superior performance of the adaptive sampling scheme over the passive sampling scheme. It is worth mentioning that we have also applied our proposed workflow to quantify the uncertainty of P_f based on a Helmholtz solver. The results are presented in Appendix B.

Motivations

We can see that the convergence of the PDF of P_f is rather slow in Fig. 5. This is a direct consequence of the employed passive sampling strategy, which evenly distributes the training samples in the parameter space (shown in Fig. 6, left), with the goal of improving the accuracy and robustness of the GP predictions throughout the entire domain. This practice may induce significant wasted expense since as illustrated in Fig. 3, only the predictions of the Monte Carlo samples close to the zero growth rate contour are crucial to the accurate calculation of P_f. Those

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

Monte Carlo samples are associated with low values of $U(\boldsymbol{x})$ (since their prediction means are small), indicating higher probabilities of misclassification and contributing to the variations of P_f calculation. Therefore, instead of making the GP model accurate everywhere in the parameter space, it would be more beneficial to make the GP model particularly accurate in the vicinity of the stability border. To achieve that, more training samples should be allocated near the zero growth rate contour, as shown in Fig. 6.

Adaptive sampling scheme

In practice, we can adopt adaptive sampling strategy to intelligently allocate training samples: we start with an initial sampling plan and train an initial GP model. Then, a learning function is applied to determine the location of the next sample, which will be added to the current sampling plan. Afterwards, an updated GP model is trained based on the enriched training samples. This iteration is terminated when a stopping criterion is met.

In order to put training samples in the vicinity of the zero growth rate contour, $U(\boldsymbol{x})$ in Eq. (15) is employed as the learning function, as proposed by Echard et al. [12]. Algorithm 1 summarizes the complete procedure for adaptive GP model training. For the stopping criterion, we require that $\min(U(\boldsymbol{x})) \geq 1.65$, which corresponds to a probability of misclassification of $\Phi(-1.65) = 0.05$.

Algorithm 1 GP model training via adaptive sampling [12]

1: Generate L samples $\boldsymbol{X}_U$ as a pool of candidate samples for sample enrichment, set iteration number $q = 0$
2: Generate m initial samples $\boldsymbol{X}_D = [\boldsymbol{x}^1, ..., \boldsymbol{x}^m]^T$
3: Call network model to evaluate their growth rate responses $\boldsymbol{Y}_D = [\alpha^1, ..., \alpha^m]^T$
4: Construct the initial GP model $\hat{f} \leftarrow (\boldsymbol{X}_D, \boldsymbol{Y}_D)$
5: Identify $\boldsymbol{x} \leftarrow \min_{\boldsymbol{x} \in \boldsymbol{X}_U} U(\boldsymbol{x})$ based on $\hat{f}$
6: **while** $U(\boldsymbol{x}) \leq 1.65$ **do**
7: $q = q + 1$.
8: Call network model to calculate the growth rate response σ of $\boldsymbol{x}$
9: Enrich samples: $\boldsymbol{X}_D = \boldsymbol{X}_D \cup \boldsymbol{x}$, $\boldsymbol{Y}_D = \boldsymbol{Y}_D \cup \sigma$
10: Re-train GP model $\hat{f} \leftarrow (\boldsymbol{X}_D, \boldsymbol{Y}_D)$
11: Identify $\boldsymbol{x} \leftarrow \min_{\boldsymbol{x} \in \boldsymbol{X}_U} U(\boldsymbol{x})$ based on $\hat{f}$
12: **end while**

Case study: A revisit

We employ the Algorithm 1 to retrain the GP model. The initial sample size m is set to be 60, which is 10 times the number of input parameters, as suggested in [27]. The candidate sample size $L = 20000$. The Latin-hypercube method is employed to generate samples.

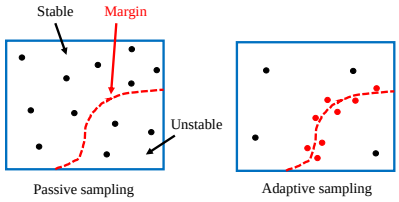


FIGURE 6: PASSIVE SAMPLING SCHEME DISTRIBUTE TRAINING SAMPLES EVENLY ACROSS THE PARAMETER SPACE, WHILE ADAPTIVE SAMPLING SCHEME ALLOCATE SAMPLES IN THE VICINITY OF THE ZERO GROWTH RATE CONTOUR. IN THIS WAY, THE CALCULATION OF P_f WOULD BE MORE ACCURATE WITH LESS UNCERTAINTY.

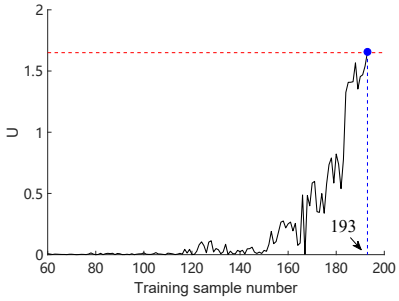


FIGURE 7: THE MINIMUM VALUE OF U IN EACH ITERATION IS RECORDED AND PLOTTED HERE. AFTER 193 TRAINING SAMPLES ARE EMPLOYED, THE MINIMUM U VALUE REACHES 1.65, INDICATING A CONVERGED GP TRAINING PROCESS.

Figure 7 displays the convergence history of the GP model training process. Here, the minimum U value among the candidate samples in each iteration is recorded and plotted. After allocating 193 training samples, *i.e.*, 133 iterations, the minimum U value reaches 1.65, therefore the training is deemed to be converged.

The case study investigated in the section "3.3 Case study" is revisited here, using the GP model trained via the adaptive sampling strategy. The calculated PDF of P_f for the cavity mode is demonstrated in Fig. 8. Also shown in the figure is the PDF of P_f calculated by the GP model, which is trained by using the passive sampling scheme with 193 training samples. We can clearly see

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

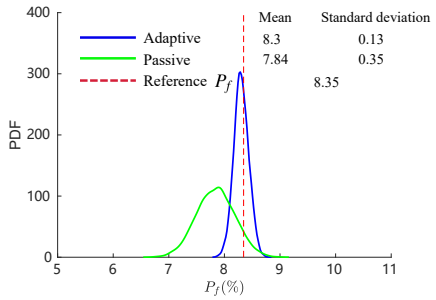


FIGURE 8: GIVEN THE SAME NUMBER OF TRAINING SAMPLES (*i.e.*, 193), ADAPTIVE SAMPLING SCHEME IS ABLE TO MAKE BETTER USE OF THE COMPUTATIONAL RESOURCES, YIELDING A MORE ACCURATE AND ROBUST P_f CALCULATION. THE RESULTS DELIVERED BY THE ADAPTIVE SAMPLING SCHEME IS COMPARABLE TO THE PASSIVE SAMPLING SCHEME WITH DOUBLE THE COMPUTATIONAL COST.

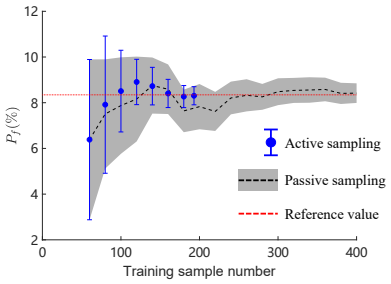


FIGURE 9: THE VARIATIONS OF THE ACCURACY AND ROBUSTNESS OF P_f CALCULATION WITH RESPECT TO THE NUMBER OF TRAINING SAMPLES. FOR BOTH METHODS, 95% CONFIDENCE INTERVAL OF THE P_f PREDICTION ARE GIVEN, WHICH SERVES AS AN INDICATOR FOR PREDICTION ROBUSTNESS.

that the adaptive training strategy significantly improved the accuracy and robustness of the P_f calculation, *i.e.*, the PDF centers closer to the reference value with a smaller level of scattering.

To provide a more convincing visual demonstration of the effectiveness of the active learning approach, Fig. 9 displays the evolutions of the P_f mean prediction as well as the associated 95% confidence interval against the number of training samples. We can observe that the adaptive sampling scheme leads to a faster convergence of the mean prediction and a more dramatic reduction of the prediction uncertainty. As a matter of fact, the calculation quality of P_f at the end of the adaptive sampling iteration (193 samples employed) is comparable to the results achieved by the passive sampling scheme with 400 samples. In addition, we emphasize that the GP model trained via the adaptive training scheme is not only limited to the current case study. Other types of studies illustrated in [10] can also be investigated with the same GP model without further computational cost.

CONCLUSION

The current work investigates the combined effects of aleatory parameter uncertainty and epistemic GP model uncertainty on the calculation of thermoacoustic instability risk (P_f). We addressed two practical questions in the study. For the first question, *i.e.*, "how to quantify the variation of modal instability risk induced by the epistemic surrogate model uncertainty?", the current study leveraged on the prediction uncertainty naturally provided by the GP model and performed correlated sampling of the GP predictions at different inputs to quantify the uncertainty of P_f calculation. We demonstrated that the P_f calculation uncertainty shrinks as more training samples are available. For the second question, *i.e.*, "how to reduce the variation of P_f calculation given a limited computational budget for the surrogate model training?", the current study adopted an adaptive sampling strategy to allocate training samples in the vicinity of the zero growth rate contour, due to the fact that the accuracy of GP predictions in that region plays a major role in achieving a reliable calculation of p_f. Our studies showed that the adaptive sampling strategy can significantly improve the accuracy and robustness of P_f calculation: With half the computational cost, the results delivered by the adaptive training strategy are comparable to the passive training strategy.

Note that in our previous work [10], we have extensively investigated various scenarios of robust design for thermoacoustic stability. However, the foundation of our strategies in [10], *i.e.,* how to reliably calculate P_f, especially in face of limited training samples, is not addressed directly, but simply assumed feasible. Our current work addresses this shortcoming. Therefore, in combination with our previous work [10], the current study is not only of academic interest, but also of industrial relevance.

Future studies should focus on the following aspects: (1) when high-dimensional flame/ acoustic models are employed, the efficiency of building a GP model may drop significantly due to the "curse of dimensionality". Therefore, it is important to investigate how the proposed workflow can cope with high-dimensional. (2) When partitioning Monte Carlo samples into two groups to improve the efficiency of GP sampling, a significant number of "group 2" samples may still exist when high-dimensional flame/acoustic models are employed. (3) It is not uncommon that in thermoacoustic control, the effort to stabilize one mode may unexpectedly promote another mode to become

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

unstable [28]. Therefore, it is necessary that our current adaptive strategy can be extended such that it can efficiently track multiple stability margins.

ACKNOWLEDGMENT

S. Guo is grateful for the financial support from doctoral scholarship of Chinese Scholarship Council (No. 201606830045).

REFERENCES

[1] Juniper, M. P., and Sujith, R. I., 2018. "Sensitivity and Nonlinearity of Thermoacoustic Oscillations". *Annual Review of Fluid Mechanics,* ***50***(1), pp. 661–689.

[2] Guo, S., Silva, C. F., Ghani, A., and Polifke, W., 2019. "Quantification and Propagation of Uncertainties in Identification of Flame Impulse Response for Thermoacoustic Stability Analysis". *J. Eng. Gas Turbines and Power,* ***141***(2), Feb., pp. 021032–10.

[3] Silva, C., Magri, L., Runte, T., and Polifke, W., 2017. "Uncertainty quantification of growth rates of thermoacoustic instability by an adjoint Helmholtz solver". *J. Eng. Gas Turbines and Power,* ***139***(1), p. 011901.

[4] Ndiaye, A., Bauerheim, M., and Nicoud, F., 2015. "Uncertainty Quantification of Thermoacoustic Instabilities on A Swirled Stabilized Combustor". In Proceedings of the ASME Turbo Expo 2015: Turbine Technical Conference and Exposition, GT2015-44133, ASME.

[5] Magri, L., Bauerheim, M., Nicoud, F., and Juniper, M. P., 2016. "Stability analysis of thermo-acoustic nonlinear eigenproblems in annular combustors. Part II. Uncertainty quantification". *Computational Physics,* ***325***, Nov., pp. 411–421.

[6] Mensah, G. A., Magri, L., and Moeck, J. P., 2017. "Methods for the Calculation of Thermoacoustic Stability Margins and Monte Carlo-Free Uncertainty Quantification". In Proceedings of ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition, ASME.

[7] Avdonin, A., Jaensch, S., Silva, C. F., Češnovar, M., and Polifke, W., 2018. "Uncertainty quantification and sensitivity analysis of thermoacoustic stability with non-intrusive polynomial chaos expansion". *Combustion and Flame,* ***189***, Mar., pp. 300–310.

[8] Avdonin, A., and Polifke, W., 2019. "Quantification of the Impact of Uncertainties in Operating Conditions on the Flame Transfer Function with Non-Intrusive Polynomial Chaos Expansion". *J. Eng. Gas Turbines and Power,* ***141***(1), p. 011020.

[9] Silva, C. F., Pettersson, P., Iaccarino, G., and Ihme, M., 2018. "Generalized chaos expansion of state space models for uncertainty quanti cation in thermoacoustics". In Proceedings of the Summer Program, Center for Turbulence Research, Stanford University.

[10] Guo, S., Silva, C. F., and Polifke, W., 2019. "Efficient Robust Design For Thermoacoustic Instability Analysis: A Gaussian Process Approach". *J. Eng. Gas Turbines and Power*.

[11] Balesdent, M., Morio, J., and Brevault, L., 2016. "Rare Event Probability Estimation in the Presence of Epistemic Uncertainty on Input Probability Distribution Parameters". *Methodology and Computing in Applied Probability,* ***18***(1), Mar., pp. 197–216.

[12] Echard, B., Gayton, N., and Lemaire, M., 2011. "AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation". *Structural Safety,* ***33***(2), Mar., pp. 145–154.

[13] Nannapaneni, S., Hu, Z., and Mahadevan, S., 2016. "Uncertainty quantification in reliability estimation with limit state surrogates". *Structural and Multidisciplinary Optimization,* ***54***(6), Dec., pp. 1509–1526.

[14] Komarek, T., and Polifke, W., 2010. "Impact of Swirl Fluctuations on the Flame Response of a Perfectly Premixed Swirl Burner". *Journal of Engineering for Gas Turbines and Power,* ***132***(6), June, p. 061503.

[15] Tay-Wo-Chong, L., Bomberg, S., Ulhaq, A., Komarek, T., and Polifke, W., 2012. "Comparative Validation Study on Identification of Premixed Flame Transfer Function". *J. of Eng. Gas Turbines Power,* ***134***(2), pp. 021502–1–8.

[16] Oberleithner, K., and Paschereit, C. O., 2016. "Modeling Flame Describing Functions Based on Hydrodynamic Linear Stability Analysis". In ASME Turbo Expo 2016: Turbomachinery Technical Conference and Exposition, p. V04BT04A009.

[17] Albayrak, A., Juniper, M. P., and Polifke, W., 2019. "Propagation speed of inertial waves in cylindrical swirling flows". *J. Fluid Mech.,* ***879***, pp. 85–120.

[18] Jones, D. R., Schonlau, M., and Welch, W. J., 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization,* ***13***(4), Dec., pp. 455–492.

[19] Schneider, E., Staudacher, S., Schuermans, B., Ye, H., and Meeuwissen, T., 2007. "Real-Time Modelling of the Thermoacoustic Dynamics of a Gas Turbine Using a Gaussian Process". pp. 323–332.

[20] Chattopadhyay, P., Mondal, S., Bhattacharya, C., Mukhopadhyay, A., and Ray, A., 2017. "Dynamic Data-Driven Design of Lean Premixed Combustors for Thermoacoustically Stable Operations". *Mechanical Design,* ***139***(11), Oct., pp. 111419–111419–10.

[21] Chattopadhyay, P., Mondal, S., Ray, A., and Mukhopadhyay, A., 2018. "Dynamic Data-Driven Combustor Design for Mitigation of Thermoacoustic Instabilities". *Journal of Dynamic Systems, Measurement, and Control,* ***141***(1),

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

Sept., pp. 014501–014501–7.
[22] Albayrak, A., Steinbacher, T., Komarek, T., and Polifke, W., 2017. "Convective Scaling of Intrinsic Thermo-Acoustic Eigenfrequencies of a Premixed Swirl Combustor". *Journal of Engineering for Gas Turbines and Power,* ***140***(4), Nov., p. 041510.
[23] Echard, B., Gayton, N., Lemaire, M., and Relun, N., 2013. "A combined Importance Sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models". *Reliability Engineering & System Safety,* ***111***, Mar., pp. 232–240.
[24] Miao, F., and Ghosn, M., 2011. "Modified subset simulation method for reliability analysis of structural systems". *Structural Safety,* ***33***(4), July, pp. 251–260.
[25] Smith, R., 2014. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, Philadelphia, Mar.
[26] Swiler, L., Slepoy, R., and Giunta, A., 2006. "Evaluation of Sampling Methods in Constructing Response Surface Approximations". In 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, AIAA 2006-1827, AIAA.
[27] Loeppky, J. L., Sacks, J., and Welch, W. J., 2009. "Choosing the Sample Size of a Computer Experiment: A Practical Guide". *Technometrics,* ***51***(4), Nov., pp. 366–376.
[28] Lieuwen, T., and Yang, V., eds., 2005. *Combustion Instabilities in Gas Turbine Engines: Operational Experience, Fundamental Mechanisms, and Modeling*, Vol. 210 of *Progress in Astronautics and Aeronautics*. AIAA.
[29] Palies, P., Durox, D., Schuller, T., and Candel, S., 2011. "Nonlinear combustion instability analysis based on the flame describing function applied to turbulent premixed swirling flames". *Combustion and Flame,* ***158***(10), Oct., pp. 1980–1991.
[30] Silva, C. F., Nicoud, F., Schuller, T., Durox, D., and Candel, S., 2013. "Combining a Helmholtz solver with the flame describing function to assess combustion instability in a premixed swirled combustor". *Combustion and Flame,* ***160***(9), Sept., pp. 1743–1754.

Appendix A: $U(\boldsymbol{x})$ and probability of misclassification

Misclassification at a Monte Carlo sample $\boldsymbol{x}$ happens when its corresponding GP prediction mean $\mu(y_p) > 0$, whereas the true response $f(\boldsymbol{x}) < 0$, or vice versa. Since the GP prediction y_p follows a normal distribution with a mean $\mu(y_p)$ (Eq. (11)) and a variance $\sigma^2(y_p)$ (Eq. (12)), we can write out the probability of misclassification at $\boldsymbol{x}$ as:

$$P_m(\boldsymbol{x}) = \Phi\left[-\frac{\mu(y_p)}{\sigma(y_p)}\right], \tag{16}$$

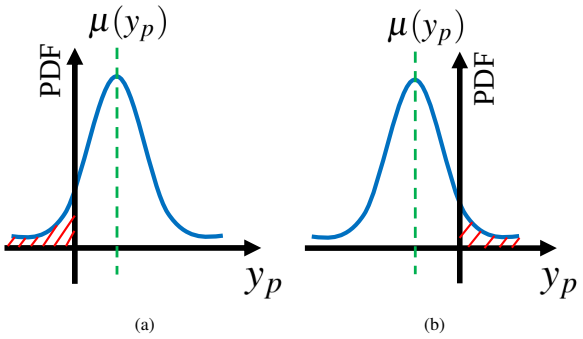


FIGURE 10: TWO SITUATIONS FOR MISCLASSIFICATION: (a) REPRESENTS THE SITUATION THAT THE TRUTH IS NEGATIVE WHILE GP PREDICTS POSITIVE; (b) REPRESENTS THE SITUATION THAT THE TRUTH IS POSITIVE WHILE GP PREDICTS NEGATIVE.

or

$$P_m(\boldsymbol{x}) = \Phi\left[\frac{\mu(y_p)}{\sigma(y_p)}\right], \tag{17}$$

where P_m stands for probability of misclassification, and Φ represents the cumulative density function of the normal distribution. Eq. (16) corresponds to the case when $\mu(y_p) > 0$ while the true response $f(\boldsymbol{x}) < 0$ (as shown in Fig. 10a); Eq. (17) corresponds to the case when $\mu(y_p) < 0$ while the true response $f(\boldsymbol{x}) > 0$ (as shown in Fig. 10b).

Combining Eq. (16)-(17), we can derive the probability of misclassification as:

$$P_m(\boldsymbol{x}) = \Phi\left[-\frac{|\mu(y_p)|}{\sigma(y_p)}\right] = \Phi[-U(\boldsymbol{x})]. \tag{18}$$

Therefore, a larger $U(\boldsymbol{x})$ indicates a smaller probability of misclassification.

Appendix B: Case study with Helmholtz solver

To further demonstrate the effectiveness of our proposed workflow to quantify and control GP model uncertainty, here we conduct another case study, where a finite-volume-based Helmholtz solver [3] is adopted.

The configuration under investigation is the EM2C turbulent swirler combustor [29]. Here we consider configuration

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

C4 [30] and the operation condition B of [29]. An in-house finite-volume-based Helmholtz solver [3] is adopted to calculate the first longitudinal thermoacoustic mode of the test rig, and its variation under input uncertainties will be the focus of the current study. The governing equation is given as:

$$\nabla(\bar{c}^2\nabla\hat{p}) - s^2\hat{p} + \alpha s\hat{p} = s(\gamma-1)\hat{q}(\boldsymbol{x}) \tag{19}$$

where $\hat{p}$ and $\hat{q}$ denote the amplitude of pressure and heat release rate distributions, respectively. $\bar{c}$ represents the mean value of the local speed of sound. A damping coefficient α is adopted to globally model the acoustic energy dissipation effect, which rests on the assumption that the acoustic system behaves like a second-order harmonic oscillator [30]. In this study, the damping rate of the combustor was evaluated to be $\alpha = 170s^{-1}$. Robin boundary conditions are employed:

$$\nabla\hat{p} + \beta\hat{p} = 0, \qquad \text{where} \qquad \beta = s\frac{1}{\bar{c}}\frac{(1-R)}{(1+R)} \tag{20}$$

where R is the reflection coefficient at a given boundary; $R_{in} = |R_{in}|$ at the combustor inlet and $R_{out} = |R_{out}|e^{i\pi}$ at the combustor outlet. A $n-\tau$ model is adopted to describe the flame heat release rate response to velocity perturbation at the reference position, which can be written as:

$$\frac{\hat{q}}{\bar{q}} = \frac{\hat{u}_{ref}}{\bar{u}} n \exp(i\omega\tau) \tag{21}$$

In this case study, we consider the flame gain n, flame time delay τ and the magnitude of reflection coefficient at the combustor outlet $|R_{out}|$ to be uncertain. Their variational ranges are displayed in table 2. An independent uniform distribution is assigned to all three parameters.

TABLE 2: Input uncertain parameters

Parameters	Range
n	$0.4 \sim 2$
τ	$3.5 \sim 6.5(ms)$
R_{out}	$0.5 \sim 1$

We start with 30 training samples (10 times of the number of uncertain parameters), and employ the active learning scheme to adaptively train the GP model. Figure 11 displays the convergence history of the GP model training process. Here, the minimum U value among the candidate samples in each iteration is recorded and plotted. After allocating 53 samples, the minimum U value is above 1.65, indicating the training process is converged.

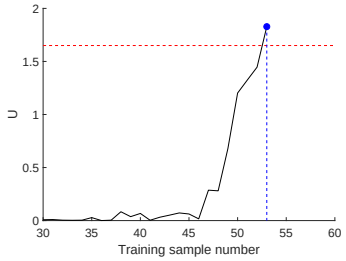


FIGURE 11: THE MINIMUM VALUE OF U IN EACH ITERATION IS RECORDED AND PLOTTED HERE. AFTER 53 TRAINING SAMPLES ARE EMPLOYED, THE MINIMUM U VALUE REACHES 1.65, INDICATING A CONVERGED GP TRAINING PROCESS.

Figure 12 shows the position of the training samples in the 3D parameter space, where blue cubics represent the initial samples (only 10 of them are shown for clear illustration), and red circles represent enriched samples. It can be seen that all the enriched samples are located in the vicinity of the stability margin, thus indicating our algorithm is behaving as we expected.

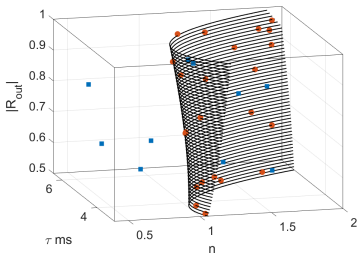


FIGURE 12: THE LOCATION OF TRAINING SAMPLES IN THE 3D PARAMETER SPACE. THE INITIAL TRAINING SAMPLES ARE SHOWN IN BLUE CUBIC (ONLY 10 OF THEM ARE SHOWN FOR CLEAR ILLUSTRATION). THE ENRICHED SAMPLES ARE SHOWN IN RED CIRCLES.

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

Compared with the passive sampling scheme, the active sampling approach significantly improved the accuracy and robustness of the P_f calculation. This point is illustrated in Fig. 13, where the predicted mean value of P_f as well as its uncertainty (95% confidence interval) are plotted against the available sample numbers. For the active sampling scheme, firstly, the predicted mean value of P_f converges much faster towards the reference value. Secondly, the prediction uncertainty of P_f reduces monotonically and with a faster speed. Finally, the provided confidence interval always covers the reference P_f value. Therefore, we are able to show that for a more complicated thermoacoustic model (*i.e.*, Helmholtz-equation-based model), active sampling scheme remains to be effective and delivers a more accurate and less uncertain calculation of P_f, compared with the passive sampling scheme with the same computational budget.

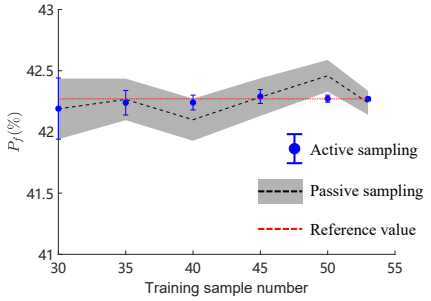


FIGURE 13: THE VARIATIONS OF THE ACCURACY AND ROBUSTNESS OF P_f CALCULATION WITH RESPECT TO THE NUMBER OF TRAINING SAMPLES. FOR BOTH METHODS, 95% CONFIDENCE INTERVAL OF THE P_f PREDICTION ARE GIVEN, WHICH SERVES AS AN INDICATOR FOR PREDICTION ROBUSTNESS.

S. Guo, C. F. Silva, and W. Polifke. Reliable Calculation of Thermoacoustic Instability Risk Using An Imperfect Surrogate Model. In Journal of Engineering for Gas Turbines and Power, 2020. Reprinted with permission from ASME.

ARTICLE IN PRESS
JID: PROCI [mNS;September 15, 2020;2:8]

Available online at www.sciencedirect.com

ScienceDirect

Proceedings of the Combustion Institute xxx (2020) xxx–xxx

Proceedings of the Combustion Institute

www.elsevier.com/locate/proci

A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions

Q1 Shuai Guo*, Camilo F. Silva, Kah Joon Yong, Wolfgang Polifke

Fakultät für Maschinenwesen, Technische Universität München, Boltzmannstr. 15, Garching D-85748, Germany

Received 7 November 2019; accepted 15 June 2020
Available online xxx

Abstract

When combining a flame model with acoustic tools to predict thermoacoustic instability, uncertainties embedded in the flame model and acoustic system parameters propagate through the thermoacoustic model, inducing variations in calculation results. Therefore, uncertainty quantification (UQ) analysis is essential for delivering a reliable prediction of thermoacoustic instability. The present paper proposes a general, surrogate-based framework to efficiently perform UQ analysis in thermoacoustic instability predictions that (1) can handle large variational ranges and flexible statistical descriptions of the uncertain parameters, (2) takes into account uncertainties from both acoustic system parameters and high-dimensional flame response models (*e.g.* the finite impulse response model (FIR), the flame describing function (FDF), *etc.*), (3) quantifies uncertainties in modal frequency and linear growth rate for linear thermoacoustic analysis, or (4) quantifies uncertainties in limit cycle frequency and amplitude for nonlinear thermoacoustic analysis. The framework is built upon Gaussian process (GP) surrogate models. An active learning strategy from the machine learning community has been adopted to significantly enhance the efficiency of GP model training, thus achieving a significant reduction in computational cost. The effectiveness of the proposed UQ framework is demonstrated by two case studies: one linear case with an uncertain FIR model and acoustic system parameters, and one nonlinear case with an uncertain FDF dataset and acoustic system parameters. Compared with reference Monte Carlo simulations, the case studies reveal UQ analyses that are, respectively, 20 and 15 times faster, but nevertheless highly accurate. The proposed GP-based framework also forms an efficient foundation on which to address other types of studies, in which repetitive thermoacoustic calculations are required, such as parametric investigations, sensitivity analyses, nonlinear bifurcation studies and robust design.

Keywords: Uncertainty quantification; Gaussian process; Thermoacoustic instability; Machine learning

1. Introduction

When combining acoustic solvers with a flame frequency response model to predict thermoacoustic instability, uncertainties embedded in the

* Corresponding author.
E-mail address: guo@tfd.mw.tum.de (S. Guo).

https://doi.org/10.1016/j.proci.2020.06.229

ARTICLE IN PRESS
JID: PROCI [mNS;September 15, 2020;2:8]

acoustic system parameters and flame models may lead to significant variations in the results. *Uncertainty quantification* (UQ) analysis is therefore essential to realize a more reliable prediction of thermoacoustic instability.

Pioneer work on accurate and affordable methods of UQ analysis in the context of thermoacoustic instability focused on the development of various surrogate techniques, in which a cheap surrogate model is first of all trained to approximate the acoustic solver and Monte Carlo simulations are then directly applied to the surrogate model to enable accelerated UQ analysis. These studies concentrated on UQ analysis in the context of linear thermoacoustic instability prediction, *i.e.* modal frequency and growth rate calculations. Some of the studies considered uncertainties from both acoustic system parameters and simplistic $n-\tau$ models for the flame dynamics [1–4], while others solely considered uncertainties in high-dimensional flame models, *i.e.* those featuring with a large number of uncertain parameters [5–8].

Despite the remarkable progress made by these studies, critical limitations still exist that could potentially hinder an effective UQ analysis in a real-world setting: (1) A general UQ framework is desirable, which is capable of considering both uncertain acoustic system parameters as well as a sophisticated flame model with a large number of uncertain coefficients; (2) the effectiveness of the above approaches is unclear when the UQ analysis is extended to the prediction of nonlinear thermoacoustic instability, where the prediction of limit cycle frequency and amplitude may be influenced by uncertain acoustic damping or an uncertain nonlinear flame response model, as indicated by the work of Palies et al. [9] and Silva et al [10]; (3) the parameter variation ranges considered were generally small (*e.g.* within 10% $\sim$ 20% of the corresponding nominal values). However, larger input parameter spaces may be required, so as to reflect realistic uncertainty levels.

The main objective of the current work is to fill this gap: we propose a general surrogate-based framework for performing UQ analysis in the prediction of thermoacoustic instability, to enable us to (1) handle large variational ranges and flexible statistical descriptions of the uncertain parameters; (2) take into account uncertainties from both acoustic system parameters and high-dimensional flame response models (such as the finite impulse response model (FIR), the flame describing function (FDF), *etc.*); (3) quantify uncertainties in modal frequency and linear growth rate in linear thermoacoustic analysis, and also quantify uncertainties in limit cycle frequency and amplitude in nonlinear thermoacoustic analysis, with a single UQ framework.

The framework is built upon *Gaussian process* (GP) surrogate models, which have previously been successfully utilized in thermoacoustic instability analysis by Schneider et al. [11], Chattopadhyay et al. [12] and Guo et al. [13]. Our current work extends the GP methodology employed in [13] in the following ways: firstly, we use GP models to approximate a 3D Helmholtz solver, as this possesses better modeling capability than an acoustic network model for addressing real-world combustors with complicated geometries, and secondly, we adopt an *active learning* strategy from machine learning to further enhance the efficiency of GP model training.

The paper is organized as follows. We start by proposing a general surrogate-based UQ framework, followed by a review of the fundamentals of GP modeling and the details of the active learning strategy. We then go on to demonstrate the effectiveness of the proposed GP-based framework through two case studies: one linear case with an uncertain FIR model and acoustic system parameters, and one nonlinear case with an uncertain FDF dataset and acoustic system parameters. All the code and data used to produce the results presented in this paper can be found at https://github.com/ShuaiGuo16/ISC20.

2. Surrogate-based UQ strategy

In general, a flame model $\mathcal{F}$ inputs a complex frequency $s=\sigma+\boldsymbol{i}\omega$ (where σ and ω are the modal growth rate and frequency, respectively) and outputs the flame gain $G=|\mathcal{F}(s)|$ and phase $\phi=\angle\mathcal{F}(s)$. Suppose it were possible to build two cheap surrogate models $\hat{f}^{\omega}$ and $\hat{f}^{\sigma}$ to accurately approximate the Helmholtz solver, such that we could quickly evaluate the corresponding thermoacoustic mode of interest, given a flame gain G, phase ϕ and acoustic parameter vector $\boldsymbol{H}$:

$$\omega\approx\hat{f}^{\omega}(G,\phi;\boldsymbol{H}),\quad \sigma\approx\hat{f}^{\sigma}(G,\phi;\boldsymbol{H}) \tag{1}$$

It would then be possible to derive the governing equations with which to calculate the thermoacoustic mode of interest under any given flame model $\mathcal{F}$ and system acoustic parameters $\boldsymbol{H}$:

$$\begin{aligned}&\hat{f}^{\omega}(|\mathcal{F}(\sigma+\boldsymbol{i}\omega)|,\angle\mathcal{F}(\sigma+\boldsymbol{i}\omega);\boldsymbol{H})-\omega=0\\&\hat{f}^{\sigma}(|\mathcal{F}(\sigma+\boldsymbol{i}\omega)|,\angle\mathcal{F}(\sigma+\boldsymbol{i}\omega);\boldsymbol{H})-\sigma=0\end{aligned} \tag{2}$$

We define the equations in Eq. (2) as the *surrogate equations*. Since $\hat{f}^{\omega}$ and $\hat{f}^{\sigma}$ can be evaluated quickly, surrogate equations can be efficiently solved by means of iterative algorithms. Experience shows that the Matlab function **fsolve** used together with the "*trust-region*" algorithm [14] is suitable for this purpose.

The newly proposed surrogate-based iterative scheme provides us with an opportunity to significantly speed up the process of UQ analysis: if we wish to perform Monte Carlo simulation with $o(10^4)$ realizations of uncertain acoustic parameters and flame model parameters, we can directly

Please cite this article as: S. Guo, C.F. Silva and K.J. Yong et al., A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.06.229

ARTICLE IN PRESS
JID: PROCI [mNS;September 15, 2020;2:8]

adopt surrogate equations to calculate the corresponding modal frequency and growth rate much more quickly than a Helmholtz solver, thus making UQ analysis affordable. In the current study, $\hat{f}^{\omega}$ and $\hat{f}^{\sigma}$ are modeled by Gaussian processes.

3. Gaussian process modeling

This section gives a brief overview of the key features of the GP modeling and presents details of the active learning strategy. The GP model training is performed with UQLab [15].

3.1. Fundamentals

GP modeling treats the output $f(\boldsymbol{x})$ of the high-fidelity model at $\boldsymbol{x}$ (a vector whose entries are the individual input parameters) as the realization of a Gaussian process:

$$f(\boldsymbol{x}) = \beta + Z(\boldsymbol{x}), \tag{3}$$

where β is a constant value and $Z(\boldsymbol{x})$ corresponds to the departure from β at $\boldsymbol{x}$, which is modeled as a Gaussian stochastic function with zero mean, variance σ^2, and covariance defined as:

$$Cov[Z(\boldsymbol{x}^i, \boldsymbol{x}^j)] = \sigma^2 R(\boldsymbol{x}^i, \boldsymbol{x}^j), \tag{4}$$

where $R(\boldsymbol{x}^i, \boldsymbol{x}^j)$ is the correlation function between any two locations $\boldsymbol{x}^i$ and $\boldsymbol{x}^j$ in the input parameter space. Here we use a Gaussian correlation function to describe $R(\boldsymbol{x}^i, \boldsymbol{x}^j)$:

$$R(\boldsymbol{x}^i, \boldsymbol{x}^j) = \exp[-\sum_{k=1}^{M} \theta_k (x_k^i - x_k^j)^2], \tag{5}$$

where M denotes the dimension of the input (*i.e.* number of input parameters), subscript k represents the k-th component of the input vector $\boldsymbol{x}$, and $\boldsymbol{\theta} = [\theta_1, ..., \theta_M]$ controls the level of correlation in the corresponding dimension.

Based on a set of training samples $\boldsymbol{X}_D = [\boldsymbol{x}^1, ..., \boldsymbol{x}^N]^T$ and their corresponding responses $\boldsymbol{Y}_D = [f(\boldsymbol{x}^1), ..., f(\boldsymbol{x}^N)]^T$, we can train the GP model by finding values for β, σ^2 and $\boldsymbol{\theta}$ such that the likelihood of achieving the observations (training samples and their corresponding responses) is maximized. The maximum likelihood estimations of β and σ^2 can be derived analytically:

$$\widehat{\beta} = (\mathbf{1}^T \boldsymbol{R}_D^{-1} \mathbf{1})^{-1} \mathbf{1}^T \boldsymbol{R}_D^{-1} \boldsymbol{Y}_D \tag{6}$$

$$\widehat{\sigma^2} = \frac{1}{N} (\boldsymbol{Y}_D - \mathbf{1}\widehat{\beta})^T \boldsymbol{R}_D^{-1} (\boldsymbol{Y}_D - \mathbf{1}\widehat{\beta}) \tag{7}$$

where $\boldsymbol{R}_D$ is the N-by-N correlation matrix between training samples $\boldsymbol{X}_D$, and $\boldsymbol{I}$ is a vector of ones of dimension M. For estimating $\boldsymbol{\theta}$, the following auxiliary optimization problem has to be solved:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left[-\frac{N}{2} \ln(\widehat{\sigma^2}) - \frac{1}{2} \ln(|\boldsymbol{R}_D|) \right] \tag{8}$$

Finally, the GP model prediction $\widehat{f}(\boldsymbol{x})$ at an arbitrary location $\boldsymbol{x}$ is:

$$\widehat{f}(\boldsymbol{x}) = \widehat{\beta} + \boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R}_D^{-1} (\boldsymbol{Y}_D - \mathbf{1}\beta), \tag{9}$$

where $\boldsymbol{r}(\boldsymbol{x})$ in Eq. (9) is the correlation vector between $\boldsymbol{x}$ and all the training samples, *i.e.*, $\boldsymbol{r}(\boldsymbol{x}) = [R(\boldsymbol{x}, \boldsymbol{x}^1), ..., R(\boldsymbol{x}, \boldsymbol{x}^N)]$. In addition, the prediction variance at location $\boldsymbol{x}$ is given as:

$$var(\boldsymbol{x}) = \widehat{\sigma^2}[1 - \boldsymbol{r}^T(\boldsymbol{x}) \boldsymbol{R}_D^{-1} \boldsymbol{r}(\boldsymbol{x})], \tag{10}$$

which indicates the uncertainty of the GP prediction.

3.2. Active learning strategy

To enhance the training efficiency, we adopt an active learning strategy to sequentially enrich the training samples. The learning strategy selects potential samples with the maximum *expected prediction error* (EPE) values, thus maximizing the reduction of the GP generalization error. The expression of the EPE value can be written in [16]:

$$EPE(\boldsymbol{x}) = \underbrace{(f(\boldsymbol{x}) - \widehat{f}(\boldsymbol{x}))^2}_{\text{bias}^2} + \underbrace{var(\boldsymbol{x})}_{\text{variance}} \tag{11}$$

To calculate the EPE value at an arbitrary location $\boldsymbol{x}$, we need to estimate the bias term in Eq. (11) because of the unknown true response $f(\boldsymbol{x})$. Following the work of Liu et al. [16], we adopt leave-one-out cross validation to estimate the bias term. Firstly, we estimate cross-validation errors at all training sample locations $\boldsymbol{x}^i$, $i = 1, ..., N$:

$$e_{CV}^2(\boldsymbol{x}^i) = (f(\boldsymbol{x}^i) - \widehat{f}^{-i}(\boldsymbol{x}^i))^2, \quad i = 1, 2, ..., N, \tag{12}$$

where $\boldsymbol{x}^i$ is the i-th training sample and $\widehat{f}^{-i}$ represents the GP model trained using all the training samples and their responses except $(\boldsymbol{x}^i, f(\boldsymbol{x}^i))$. For an arbitrary location $\boldsymbol{x}$, we then simply find the closest (in terms of Euclidean distance) training sample to location $\boldsymbol{x}$ and assign its associated cross-validation error to $e_{CV}^2(\boldsymbol{x})$:

$$e_{CV}^2(\boldsymbol{x}) = e_{CV}^2(\boldsymbol{x}^i), \quad \min_i |\boldsymbol{x} - \boldsymbol{x}^i| \quad i = 1, 2, ..., N \tag{13}$$

Finally, we replace the bias term in Eq. (11) with the cross-validation error:

$$EPE(\boldsymbol{x}) = e_{CV}^2(\boldsymbol{x}) + var(\boldsymbol{x}) \tag{14}$$

4. Case studies

In this section, we demonstrate the effectiveness of the proposed GP-based UQ framework by way of two case studies. We start with the layout of the thermoacoustic problem setting, and follow it with a description of the GP model training process. The

Please cite this article as: S. Guo, C.F. Silva and K.J. Yong et al., A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.06.229

With permission from S. Guo, C. F. Silva, K. J. Yong and W. Polifke. A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions. Proceedings of the Combustion Institute, Elsevier, 2020.

ARTICLE IN PRESS
JID: PROCI [mNS;September 15, 2020;2:8]

trained GP models will subsequently be employed in the proposed UQ framework to address both the linear and the nonlinear UQ problems. All calculations were performed on a personal computer with a 3.20GHz Intel Core i5-6500.

4.1. Thermoacoustic problem setting

The combustor in question is a turbulent swirl EM2C laboratory-scale test rig [9]. The chosen configuration was C11 (c.f. Silva et al. [10]), with plenum length $l_1 = 224\,mm$ and combustion chamber length $l_3 = 200\,mm$. The operation condition considered is "A", with total flame power $\bar{\dot{Q}} = 1.94\,kW$ and mean flow velocity $\bar{u}_{ref} = 2.67\,m/s$ [10].

An in-house finite-volume-based Helmholtz solver [1] is adopted to calculate the first longitudinal thermoacoustic mode of the test rig. The current study will focus on its variation under input uncertainties. The governing equation is given as:

$$\nabla(\bar{c}^2\nabla\hat{p}) - s^2\hat{p} + \alpha s\hat{p} = s(\gamma-1)\hat{q}(\boldsymbol{x}) \tag{15}$$

where $\hat{p}$ and $\hat{q}$ respectively denote the amplitude of pressure and heat release rate distributions. $\bar{c}$ represents the mean local speed of sound. A damping coefficient α is adopted to globally model the effect of acoustic energy dissipation, which is based on the assumption that the acoustic system behaves like a second-order harmonic oscillator [10]. Robin boundary conditions are employed:

$$\nabla\hat{p} + \beta\hat{p} = 0, \qquad \text{where} \qquad \beta = s\frac{1}{\bar{c}}\frac{(1-R)}{(1+R)} \tag{16}$$

in which R is the reflection coefficient at a given boundary; $R_{in} = |R_{in}|$ at the combustor inlet and $R_{out} = |R_{out}|e^{i\pi}$ at the combustor outlet.

GP models are trained on the basis of a simple transfer function definition of the flame response:

$$\frac{\hat{\dot{q}}}{\bar{\dot{q}}} = \frac{\hat{u}_{ref}}{\bar{u}} G\exp(\boldsymbol{i}\phi) \tag{17}$$

with flame gain G and phase ϕ as the input parameters. To enable linear thermoacoustic stability analysis, an FIR model, which models the flame dynamics by means of distributed time delays [17], is employed, whereby $\hat{\dot{q}}$ is modeled as:

$$\frac{\hat{\dot{q}}}{\bar{\dot{q}}} = \frac{\hat{u}_{ref}}{\bar{u}} \sum_{k=0}^{L-1} h_k e^{-i(k+1)\Delta t(\omega - i\sigma)} \tag{18}$$

with FIR coefficients h_k's as the input parameters. For (weakly) nonlinear analysis, an FDF model is employed, such that

$$\frac{\hat{\dot{q}}}{\bar{\dot{q}}} = \frac{\hat{u}_{ref}}{\bar{u}} G(\omega, |\hat{u}|)\exp(\boldsymbol{i}\phi(\omega, |\hat{u}|)) \tag{19}$$

with data of gain G and phase ϕ at various frequencies and velocity perturbation amplitudes as the parameters. In Eq. (17)–(19), overbar denotes the temporal mean quantity. $\bar{\dot{q}} = \bar{\dot{Q}}/V_f$, where $\bar{\dot{Q}}$ is the global heat release rate and V_f denotes the volume of the flame region. We assume that the flame is acoustically compact. Accordingly, we consider that the gain and phase are uniform within the flame region and zero elsewhere. In accordance with the conservation of momentum, we have $\hat{u}_{ref} = -\nabla\hat{p}_{ref}/(s\bar{\rho})$, where $\bar{\rho}$ is the mean flow density, thus closing Eq. (15). The nonlinear eigenvalue problem Eq. (15) is solved using an iterative scheme based on the secant algorithm.

Table 1
Input parameters.

Parameters	Training range	Uncertainty range	
		Linear case	Nonlinear case
G	$0.5\sim3$	FIR	FDF
ϕ	$0\sim\pi$		
R_{in}	$0.7\sim1$	$0.7\sim1$	$0.9\sim1$
$\lvert R_{out}\rvert$	$0.6\sim1$	$0.6\sim1$	$0.7\sim1$
α	$100\sim160$	$100\sim160$	$110\sim160$

4.2. GP model training

In the present study, we strive to train two separate, generally applicable GP models $\hat{f}^{(\omega)}(G, \phi, R_{in}, R_{out}, \alpha)$ and $\hat{f}^{(\sigma)}(G, \phi, R_{in}, R_{out}, \alpha)$ to approximate, respectively, the frequency and growth rate of the first longitudinal thermoacoustic mode of the investigated rig. The input ranges considered in the current study are listed in Table 1, which are significantly larger than the ranges investigated in the previous studies.

Our training strategy consists of 4 major steps: (1) train an initial GP model based on the initial training dataset; (2) from the pool of candidate samples, identify the sample with the largest EPE value and employ the Helmholtz solver to compute its corresponding frequency and growth rate responses; (3) enrich the training dataset with the selected sample and its corresponding frequency and growth rate responses, then re-train the GP model with the enriched training dataset; (4) repeat steps (2)-(3) until the GP model is deemed converged. The pseudo-code in Fig. 1 provides further details of the active learning process when our two GP models are trained simultaneously. Here, the initial sample size $m = 50$ and the candidate sample size $L = 10{,}000$. The training stops when the maximum EPE value of both GP models drops below 5% of their individual maximum EPE values before submitting to active learning.

Figure 2 displays the convergence histories for training the GP models. In total, we employed 150 ($50 + 2 \times 50$) input samples (*i.e.* 150 Helmholtz solver calculations) to reach the target accuracy for GP training, which amounts to 90s of computation time. In the next two sections, these two

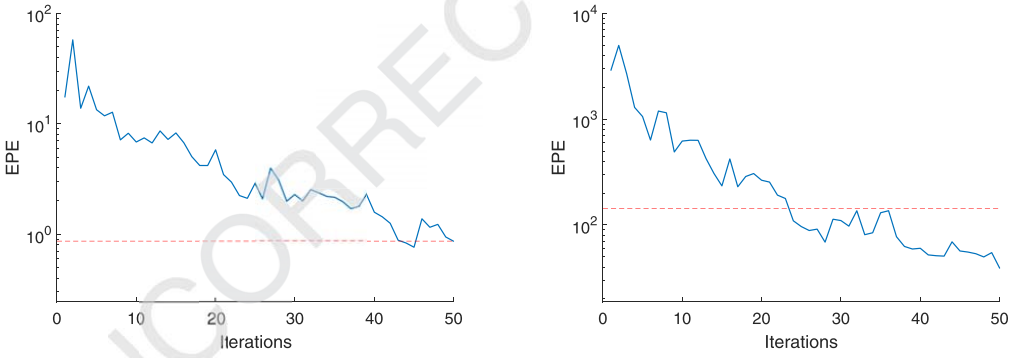
Please cite this article as: S. Guo, C.F. Silva and K.J. Yong et al., A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.06.229

ARTICLE IN PRESS
JID: PROCI [mNS;September 15, 2020;2:8]

Algorithm 1 GP model training via active learning

1: Generate L samples $\boldsymbol{X}_U$ as a pool of candidate samples for sample enrichment, set iteration number $q = 0$
2: Generate m initial samples $\boldsymbol{X}_D = [\boldsymbol{x}^1, ..., \boldsymbol{x}^m]^T$
3: Call Helmholtz solver to evaluate their frequency responses $\boldsymbol{y}_D^{(\omega)} = [\omega^1, ..., \omega^m]^T$ and growth rate responses $\boldsymbol{y}_D^{(\sigma)} = [\sigma^1, ..., \sigma^m]^T$
4: Construct the initial GP models $\hat{f}^{(\omega)} \leftarrow (\boldsymbol{X}_D, \boldsymbol{y}_D^{(\omega)})$ and $\hat{f}^{(\sigma)} \leftarrow (\boldsymbol{X}_D, \boldsymbol{y}_D^{(\sigma)})$
5: Identify $\boldsymbol{x}^{(\omega)} \leftarrow \max_{\boldsymbol{x} \in \boldsymbol{X}_U} EPE(\boldsymbol{x})$ based on $\hat{f}^{(\omega)}$, let $EPE_0^{\omega} = EPE(\boldsymbol{x}^{(\omega)})$
6: Identify $\boldsymbol{x}^{(\sigma)} \leftarrow \max_{\boldsymbol{x} \in \boldsymbol{X}_U} EPE(\boldsymbol{x})$ based on $\hat{f}^{(\sigma)}$, let $EPE_0^{\sigma} = EPE(\boldsymbol{x}^{(\sigma)})$
7: **while** $EPE^{\omega} > 0.05 \cdot EPE_0^{\omega}$ and $EPE^{\sigma} > 0.05 \cdot EPE_0^{\sigma}$ **do**
8: $q = q + 1$.
9: Call Helmholtz solver to calculate the frequency responses $\boldsymbol{y}^{(\omega)} = [\omega^{(\omega)}, \omega^{(\sigma)}]$ and growth rate responses of $\boldsymbol{y}^{(\sigma)} = [\sigma^{(\omega)}, \sigma^{(\sigma)}]$ of $\boldsymbol{x}^{(\omega)}$ and $\boldsymbol{x}^{(\sigma)}$, respectively.
10: Enrich samples: $\boldsymbol{X}_D = \boldsymbol{X}_D \cup \boldsymbol{x}^{(\omega)} \cup \boldsymbol{x}^{(\sigma)}$, $\boldsymbol{y}_D^{(\omega)} = \boldsymbol{y}_D^{(\omega)} \cup \boldsymbol{y}^{(\omega)}$, $\boldsymbol{y}_D^{(\sigma)} = \boldsymbol{y}_D^{(\sigma)} \cup \boldsymbol{y}^{(\sigma)}$
11: Re-train GP models $\hat{f}^{(\omega)} \leftarrow (\boldsymbol{X}_D, \boldsymbol{y}_D^{(\omega)})$ and $\hat{f}^{(\sigma)} \leftarrow (\boldsymbol{X}_D, \boldsymbol{y}_D^{(\sigma)})$
12: Identify $\boldsymbol{x}^{(\omega)} \leftarrow \max_{\boldsymbol{x} \in \boldsymbol{X}_U} EPE(\boldsymbol{x})$ based on $\hat{f}^{(\omega)}$, let $EPE^{\omega} = EPE(\boldsymbol{x}^{(\omega)})$
13: Identify $\boldsymbol{x}^{(\sigma)} \leftarrow \max_{\boldsymbol{x} \in \boldsymbol{X}_U} EPE(\boldsymbol{x})$ based on $\hat{f}^{(\sigma)}$, let $EPE^{\sigma} = EPE(\boldsymbol{x}^{(\sigma)})$
14: **end while**

Fig. 1. Algorithm for GP model training.



(a) Frequency (b) Growth Rate

Fig. 2. EPE convergence history: After 50 iterations, maximum EPE values for both GP models drop below 5% of their initial maximum EPE values.

newly-trained GP models will be employed to address two case studies.

4.3. Case study: linear thermoacoustic UQ analysis

The FIR model represents a sophisticated and realistic flame model, which describes flame dynamics in the time domain. In this first case study, we propagate the uncertainties in the FIR model parameters and acoustic system parameters to the modal frequency and growth rate calculations.

Here, we consider a 65-coefficient FIR model, which is displayed in Fig. 3. The FIR model coefficients $\boldsymbol{h} = [h_0, h_1, ..., h_{64}]$ are treated as random variables, following a multivariate normal distribution. The FIR model is primarily obtained by unsteady numerical simulation, and the uncertainties associated with the FIR coefficients are evaluated during the system identification process [18], where low signal-to-noise ratio may be encountered, and only short CFD time series are available. In this study, we mimic such a procedure in three steps: first, we perform an inverse Fourier transformation on the flame frequency response data given by 'Fig. 3, Flame A, $|\hat{u}/\bar{u}_b| = 0.07$' in [10], to obtain a reference flame impulse response model; we then insert this reference FIR model into a time-domain thermoacoustic network model [19], to obtain the time series of u'_{ref} and $\dot{q}'$ by forcing the system with a broadband velocity signal and a colored

Please cite this article as: S. Guo, C.F. Silva and K.J. Yong et al., A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.06.229

With permission from S. Guo, C. F. Silva, K. J. Yong and W. Polifke. A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions. Proceedings of the Combustion Institute, Elsevier, 2020.

ARTICLE IN PRESS
JID: PROCI [mNS;September 15, 2020;2:8]

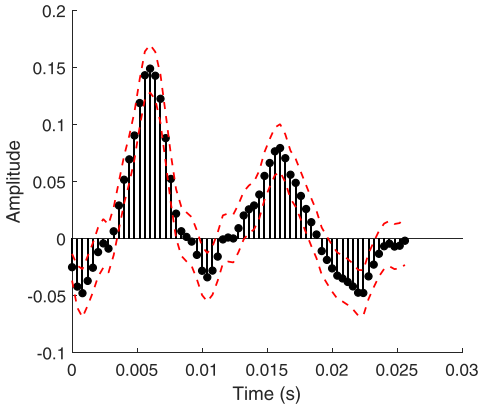


Fig. 3. Finite impulse response model. Each discrete stem represents one coefficient h_k; upper and lower bounds constitute the 95% confidence interval.

combustion noise signal [19]. Finally, we employ system identification [18] on the time series obtained to reidentify the FIR model (results shown in Fig. 3). The identification process provides us with both the mean and covariance matrix of the FIR model coefficients.

In addition to the FIR coefficient uncertainty, we also consider the reflection coefficients R_{in}, R_{out} and the damping coefficient α of the acoustic model to be uncertain. They are assumed to exhibit independent uniform distributions, with the ranges given in Table 1. This results, in total, in a 68-dimensional UQ problem. This dimensionality is significantly higher than in previous studies.

To perform the uncertainty propagation, we perform Monte Carlo simulation by adopting the proposed GP-based UQ strategy, *i.e.* we draw 20,000 samples of $[\boldsymbol{h},\ R_{in},\ R_{out},\ \alpha]^{(i)}, i = 1, ..., 20,000$. We solve surrogate equations once for each sample, where $\boldsymbol{H} = [R_{in}, R_{out}, \alpha]$ and $\mathcal{F}(\sigma + j\omega) = \sum_{k=0}^{64} h_k e^{-i(k+1)\Delta t(\omega - i\sigma)}$.

The probability density functions (PDF) of growth rate and frequency are shown in Fig. 4. Excellent matches are achieved between the results yielded by the GP-based UQ scheme and the reference results obtained using direct Helmholtz solver calculations. In terms of computational time, a GP-based iterative scheme requires 463s, while direct Helmholtz solver calculations cost 9094s, thus achieving an approximately 20 times increase in the speed of UQ analysis. It is anticipated that when a more complex configuration is considered, the computational cost of a single Helmholtz solver calculation would be significantly longer, while the cost of each iteration in a GP-based UQ scheme would basically stay the same. In this way, a much more significant acceleration of the UQ analysis may be achieved.

Table 2
Comparisons with Active Subspace approach.

Method	GP-based UQ strategy	Active Subspace [7]
Uncertainty sources	Flame & acoustic	Flame
Fidelity	$\approx$ Helmholtz solver	$\approx$ acoustic network solver
Cost/Effectiveness	150 solver evaluations/ 68 parameters	400 solver evaluations/ 16 parameters
Applicability	General flame model	Only FIR model

Previously we successfully applied an *Active Subspace* approach [7] to address a high-dimensional thermoacoustic uncertainty quantification problem. Table 2 summarizes the two strategies. Our newly proposed strategy performs better in terms of applicability, fidelity and cost-effectiveness ratio. Note, however, that [8] indicated that the active subspace identified in [7] provides new insights regarding the causal relationship between the variation of FIR coefficients and the variation of modal growth rate, which is beyond the scope of the current study.

4.4. Case study: nonlinear thermoacoustic UQ analysis

In this second case study, FDF gain and phase data at discrete frequency and velocity perturbation levels are used as a nonlinear flame model and are considered uncertain. Together with the acoustic system parameters, we will propagate those uncertainties to the predictions of limit cycle frequency and amplitude.

In practice, FDF model uncertainty is induced by both *data quantity* (measurements only available at limited frequencies and amplitude combinations) and *data quality* (resulting from the fact that measurements have errors/uncertainties). To mimic this situation, we introduce FDF model uncertainty in three steps:

Step 1: we use the FDF gain and phase data in 'Fig. 3, Flame A' [10] as our reference FDF model. To mimic uncertainty induced by data quantity, we only use the gain and phase data measured at $f_m = [0, 30, 60, 80, 140, 170, 190, 210, 230, 250](Hz)$ for all available amplitude levels $((|\hat{u}|/\bar{u}_b)_m = [0.07, 0.15, 0.3, 0.41, 0.51, 0.71])$.

Step 2: To mimic uncertainty induced by data quality, we further assume at each measurement location that gain measurement follows a normal distribution, with the mean being the original experimental value and the standard deviation being 10% of the mean; phase measurement also follows a normal distribution, with the mean being

Please cite this article as: S. Guo, C.F. Silva and K.J. Yong et al., A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.06.229

With permission from S. Guo, C. F. Silva, K. J. Yong and W. Polifke. A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions. Proceedings of the Combustion Institute, Elsevier, 2020.

ARTICLE IN PRESS
JID: PROCI [mNS;September 15, 2020;2:8]
S. Guo, C.F. Silva and K.J. Yong et al. / Proceedings of the Combustion Institute xxx (xxxx) xxx 7

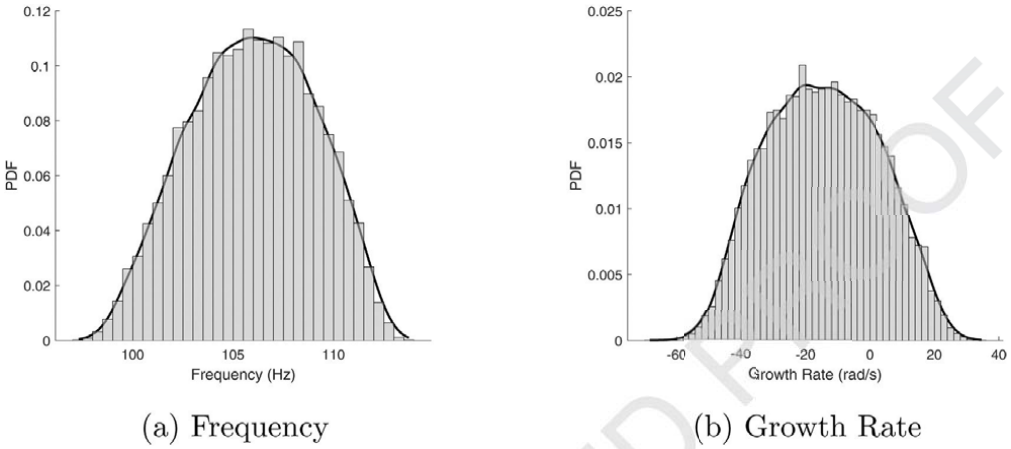


Fig. 4. PDF comparison between the GP-based UQ scheme (black curves) and Monte Carlo applied directly to Helmholtz solver (bars).

the original experimental value and the standard deviation being 5% of the mean.

Step 3: to facilitate nonlinear thermoacoustic analysis, we interpolate the uncertain measurements at combinations $(f_m, (|\hat{u}|/\bar{u}_b)_m)$ to a finer grid consisting of 26 levels of frequency and 10 levels of amplitude that are evenly distributed between $0 \sim 250Hz$ and $0.07 \sim 0.71$, respectively. Following the work of McCartney et al. [20], we use stochastic Gaussian process regression [21] to interpolate FDF gain and phase data, individually. As a result, all 260 data points for gain (and another 260 data points for phase) become uncertain, which follow a multivariate normal distribution, with the mean and covariance determined by the corresponding stochastic GP model. Hence, we are able to generate realizations of FDF gain and phase data.

In addition, we also consider R_{in}, R_{out} and α to be uncertain, which follow independent uniform distributions, with the ranges given in Table 1. In total, this is a 523-dimensional (260+260+3) UQ problem.

To perform Monte Carlo simulation, we draw 20,000 samples of the uncertain inputs, where each sample contains one realization of R_{in}, R_{out} and α, respectively, as well as one realization of a 26-by-10 gain dataset and one realization of a 26-by-10 phase dataset.

To perform nonlinear thermoacoustic analysis for each sample, we follow the framework proposed by Noiray et al. [22] and later successfully applied to Helmholtz solver [10]: for each amplitude level, we obtain the corresponding flame model by using the Matlab function 'rationalfit' to fit a rational function [23] based on 26 pairs of gain and phase data at that amplitude level. We then insert this flame model into surrogate equations to calculate the frequency and growth rate for that amplitude. Based on the trajectory of the growth rate over the amplitude, we can determine the limit cycle frequency and amplitude when instability occurs.

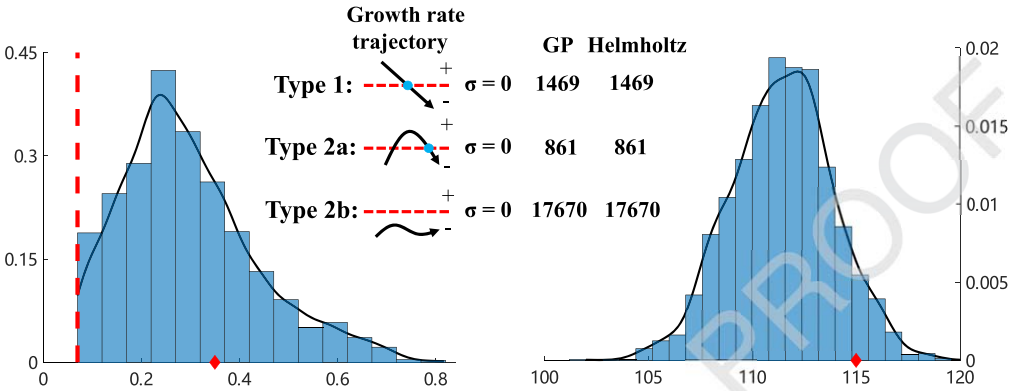
Aligning with the definitions given in [22], three types of growth rate trajectory are observed among 20,000 Monte Carlo samples; their decompositions predicted by the GP-based UQ scheme and Helmholtz solver are given in Fig. 5. Note that while only samples with a Type 1 or 2a trajectory (where limit cycle oscillation actually occurs) are shown in the histogram of limit cycle amplitude and frequency, the histogram is actually normalized with respect to all 20,000 samples. We observe perfect matches between both methods. In terms of computational time, the GP-based iterative scheme requires 2800s, while direct Helmholtz solver calculations costs 42,700s, which shows that an increase in speed of approximately 15 times acceleration is achieved, thus demonstrating the effectiveness of the newly proposed UQ scheme.

Also shown in Fig. 5 are the experimentally measured limit cycle amplitude and frequency (red diamond). A significant mismatch was previously observed between the numerical prediction and the experimental results in the configuration C11 [10]. In the light of our current UQ analysis, this mismatch may be attributed to the uncertainties in FDF data and acoustic system properties.

Before concluding this section, we would like to make two remarks. First of all, in the current case study, we unfortunately do not have any quantitative uncertainty information available regarding the experimental measurements, which would have been beneficial for shedding light on the credibility of the numerical solver (in our case, a Helmholtz solver) as well as the specification of the flame and acoustic parametric uncertainty descriptions. Nevertheless, the proposed GP-based UQ framework can be robustly and conveniently adopted in other cases where experimental measurement uncertain-

Please cite this article as: S. Guo, C.F. Silva and K.J. Yong et al., A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.06.229

ARTICLE IN PRESS
JID: PROCI [mNS;September 15, 2020;2:8]
8 S. Guo, C.F. Silva and K.J. Yong et al. / Proceedings of the Combustion Institute xxx (xxxx) xxx



(a) Limit cycle amplitude (b) Limit cycle frequency

Fig. 5. PDF comparison between the GP-based UQ scheme (black curves) and Monte Carlo applied directly to Helmholtz solver (bars). Three types of growth rate trajectory are observed among 20,000 samples (Type 1: linearly unstable; Type 2a: linearly stable, nonlinearly unstable; Type 2b: linearly stable, nonlinearly stable). Although only samples with a Type 1 or 2a trajectory [22] are shown in the histogram, the histogram is normalized with respect to all 20,000 samples. The current setup exhibits a probability of 11.6% of being unstable and 88.4% of being stable. Experimental results are shown as the diamond in the figure. Red dashed lines represent the minimum amplitude level of $(|\hat{u}|/\bar{u}_b)_m = 0.07$ in the FDF dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ties are available, thus facilitating calibration of the numerical solver and inverse uncertainty propagation to derive the input uncertainty descriptions. These ideas will be incorporated in our further studies. Secondly, our preliminary tests have shown that the variations of the limit cycle frequency and amplitude are sensitive to the choice of uncertainty generation procedure for FDF. In particular, compared with the data quality aspect (*i.e.* level of measurement error), the data quantity aspect, *i.e.* the number and the frequency locations of FDF gain/phase measurements, plays a bigger role in determining the FDF model uncertainty, thus it has a more significant impact on the variation of limit cycle frequency and amplitude. Nevertheless, since our current focus is on developing a general surrogate-based framework to efficiently perform high-dimensional UQ analysis, a rigorous sensitivity analysis is out of the scope of the present paper and will be incorporated in future studies. Fortunately, the UQ framework developed here makes such sensitivity analysis entirely feasible.

5. Conclusion

In the current study, we developed a Gaussian-process-based iterative UQ scheme to perform effectively high-dimensional uncertainty quantification analysis for both linear and nonlinear thermoacoustic instability analysis. We employed an active learning strategy to efficiently train Gaussian process models, thus achieve maximum computational cost savings. The accuracy and efficiency of the procedure were demonstrated by addressing two UQ case studies.

Equipped with this fast tool for evaluating a wide design space of thermoacoustic instability, a new array of research opportunities in the field of uncertainty management has now become possible. Examples include but are not limited to sensitivity analysis, robust design, inverse uncertainty propagation and computational model calibration.

Finally, we wish to emphasis that the success of GP model training depends on all the responses in $\boldsymbol{Y}_D$ being correctly calculated and corresponding to the particular thermoacoustic mode under investigation. Our experience has shown that during GP model training, if the acoustic solver returns the frequency and growth rate of another mode, the EPE value does not converge during the training process. Considering that accurately classifying thermoacoustic modes may not always be an easy task in practice, close attention must be paid to this when employing the current GP-based UQ framework.

Declaration of Competing Interest

None. Q3

Acknowledgments

S. Guo is grateful for the financial support in the form of a doctoral scholarship from the Chinese Scholarship Council (No. 201606830045). Q4

Please cite this article as: S. Guo, C.F. Silva and K.J. Yong et al., A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.06.229

With permission from S. Guo, C. F. Silva, K. J. Yong and W. Polifke. A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions. Proceedings of the Combustion Institute, Elsevier, 2020.

ARTICLE IN PRESS
JID: PROCI [mNS;September 15, 2020;2:8]

References

[1] C.F. Silva, L. Magri, T. Runte, W. Polifke, *J. Eng. Gas Turbines and Power* 139 (1) (2017) 011901, doi:10.1115/1.4034203.
[2] G.A. Mensah, L. Magri, J.P. Moeck, in: Proceedings of ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition, ASME, Charlotte, NC, USA, 2017, doi:10.1115/GT2017-64829.
[3] C.F. Silva, P. Pettersson, G. Iaccarino, M. Ihme, in: Proceedings of the Summer Program, Center for Turbulence Research, Stanford University, Stanford, USA, 2018.
[4] A. Avdonin, S. Jaensch, C.F. Silva, M. Češnovar, W. Polifke, *Combustion and Flame* 189 (2018) 300–310.
[5] M. Bauerheim, A. Ndiaye, P. Constantine, S. Moreau, F. Nicoud, *Journal of Fluid Mechanics* 789 (2016) 534–566, doi:10.1017/jfm.2015.730.
[6] L. Magri, M. Bauerheim, F. Nicoud, M.P. Juniper, *Computational Physics* 325 (2016) 411–421, doi:10.1016/j.jcp.2016.08.043.
[7] S. Guo, C.F. Silva, A. Ghani, W. Polifke, *J. Eng. Gas Turbines and Power* 141 (2) (2019a), doi:10.1115/1.4041652. 021032–10
[8] S. Guo, C.F. Silva, M. Bauerheim, A. Ghani, W. Polifke, *Proceedings of the Combustion Institute* 37 (2019b) 5299–5306, doi:10.1016/j.proci.2018.07.020.
[9] P. Palies, D. Durox, T. Schuller, S. Candel, *Combustion and Flame* 158 (10) (2011) 1980–1991, doi:10.1016/j.combustflame.2011.02.012.
[10] C.F. Silva, F. Nicoud, T. Schuller, D. Durox, S. Candel, *Combustion and Flame* 160 (9) (2013) 1743–1754, doi:10.1016/j.combustflame.2013.03.020.
[11] E. Schneider, S. Staudacher, B. Schuermans, H. Ye, T. Meeuwissen, *Real-Time Modelling of the Thermoacoustic Dynamics of a Gas Turbine Using a Gaussian Process* (2007) 323–332, doi:10.1115/GT2007-27468.
[12] P. Chattopadhyay, S. Mondal, A. Ray, A. Mukhopadhyay, *Journal of Dynamic Systems, Measurement, and Control* 141 (1) (2018) 014501–014501–7, doi:10.1115/1.4040210.
[13] S. Guo, C.F. Silva, W. Polifke, *J. Eng. Gas Turbines and Power* (2019), doi:10.1115/1.4044197.
[14] T.F. Coleman, Y. Li, *SIAM Journal on Optimization* 6 (2) (1996) 418–445, doi:10.1137/0806023.
[15] C. Lataniotis, S. Marelli, B. Sudret, UQLab User Manual - Kriging (Gaussian Process Modelling), *Technical Report*, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Zurich, Switzerland, 2017.
[16] H. Liu, J. Cai, Y.-S. Ong, *Computers & Chemical Engineering* 106 (2) (2017) 171–182, doi:10.1016/j.compchemeng.2017.05.025.
[17] W. Polifke, *Prog. Energy Combust. Sci.* 79 (2020) 100845, doi:10.1016/j.pecs.2020.100845.
[18] W. Polifke, *Annals of Nuclear Energy* 67C (2014) 109–128, doi:10.1016/j.anucene.2013.10.037.
[19] S. Jaensch, M. Merk, T. Emmert, W. Polifke, *Combustion Theory and Modelling* 22 (3) (2018) 613–634, doi:10.1080/13647830.2018.1443517.
[20] M. McCartney, M. Haeringer, W. Polifke, in: GT2019-91319, ASME, Pheonix, Arizona, 2019, p. 10, doi:10.1115/GT2019-91319.
[21] V. Picheny, T. Wagner, D. Ginsbourger, *Structural and Multidisciplinary Optimization* 48 (3) (2013) 607–626, doi:10.1007/s00158-013-0919-4.
[22] N. Noiray, D. Durox, T. Schuller, S. Candel, *Journal of Fluid Mechanics* 615 (2008) 139–167, doi:10.1017/S0022112008003613.
[23] B. Gustavsen, A. Semlyen, *IEEE Transactions on Power Delivery* 14 (3) (1999) 1052–1061, doi:10.1109/61.772353.

Please cite this article as: S. Guo, C.F. Silva and K.J. Yong et al., A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions, Proceedings of the Combustion Institute, https://doi.org/10.1016/j.proci.2020.06.229

With permission from S. Guo, C. F. Silva, K. J. Yong and W. Polifke. A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions. Proceedings of the Combustion Institute, Elsevier, 2020.