



A Gaussian-process-based framework for high-dimensional uncertainty quantification analysis in thermoacoustic instability predictions

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.

© 2020 The Combustion Institute. Published by Elsevier Inc. All rights reserved.

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).

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% ~ 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 $s = \sigma + i\omega$ (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}^ω and \hat{f}^σ 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 \mathbf{H} :

$$\omega \approx \hat{f}^\omega(G, \phi; \mathbf{H}), \quad \sigma \approx \hat{f}^\sigma(G, \phi; \mathbf{H}) \quad (1)$$

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 \mathbf{H} :

$$\begin{aligned} \hat{f}^\omega(|\mathcal{F}(\sigma + i\omega)|, \angle \mathcal{F}(\sigma + i\omega); \mathbf{H}) - \omega &= 0 \\ \hat{f}^\sigma(|\mathcal{F}(\sigma + i\omega)|, \angle \mathcal{F}(\sigma + i\omega); \mathbf{H}) - \sigma &= 0 \end{aligned} \quad (2)$$

We define the equations in Eq. (2) as the *surrogate equations*. Since \hat{f}^ω and \hat{f}^σ 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

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}^ω and \hat{f}^σ 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 $f(\mathbf{x})$ of the high-fidelity model at \mathbf{x} (a vector whose entries are the individual input parameters) as the realization of a Gaussian process:

$$f(\mathbf{x}) = \beta + Z(\mathbf{x}), \quad (3)$$

where β is a constant value and $Z(\mathbf{x})$ corresponds to the departure from β at \mathbf{x} , which is modeled as a Gaussian stochastic function with zero mean, variance σ^2 , and covariance defined as:

$$\text{Cov}[Z(\mathbf{x}^i, \mathbf{x}^j)] = \sigma^2 R(\mathbf{x}^i, \mathbf{x}^j), \quad (4)$$

where $R(\mathbf{x}^i, \mathbf{x}^j)$ is the correlation function between any two locations \mathbf{x}^i and \mathbf{x}^j in the input parameter space. Here we use a Gaussian correlation function to describe $R(\mathbf{x}^i, \mathbf{x}^j)$:

$$R(\mathbf{x}^i, \mathbf{x}^j) = \exp\left[-\sum_{k=1}^M \theta_k (\mathbf{x}_k^i - \mathbf{x}_k^j)^2\right], \quad (5)$$

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 \mathbf{x} , and $\boldsymbol{\theta} = [\theta_1, \dots, \theta_M]$ controls the level of correlation in the corresponding dimension.

Based on a set of training samples $\mathbf{X}_D = [\mathbf{x}^1, \dots, \mathbf{x}^N]^T$ and their corresponding responses $\mathbf{Y}_D = [f(\mathbf{x}^1), \dots, f(\mathbf{x}^N)]^T$, 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:

$$\hat{\beta} = (\mathbf{1}^T \mathbf{R}_D^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{R}_D^{-1} \mathbf{Y}_D \quad (6)$$

$$\hat{\sigma}^2 = \frac{1}{N} (\mathbf{Y}_D - \mathbf{1} \hat{\beta})^T \mathbf{R}_D^{-1} (\mathbf{Y}_D - \mathbf{1} \hat{\beta}) \quad (7)$$

where \mathbf{R}_D is the N-by-N correlation matrix between training samples \mathbf{X}_D , and $\mathbf{1}$ is a vector of ones of dimension M . For estimating $\boldsymbol{\theta}$, the following auxiliary optimization problem has to be solved:

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \left[-\frac{N}{2} \ln(\hat{\sigma}^2) - \frac{1}{2} \ln(|\mathbf{R}_D|) \right] \quad (8)$$

Finally, the GP model prediction $\hat{f}(\mathbf{x})$ at an arbitrary location \mathbf{x} is:

$$\hat{f}(\mathbf{x}) = \hat{\beta} + \mathbf{r}(\mathbf{x})^T \mathbf{R}_D^{-1} (\mathbf{Y}_D - \mathbf{1} \hat{\beta}), \quad (9)$$

where $\mathbf{r}(\mathbf{x})$ in Eq. (9) is the correlation vector between \mathbf{x} and all the training samples, *i.e.*, $\mathbf{r}(\mathbf{x}) = [R(\mathbf{x}, \mathbf{x}^1), \dots, R(\mathbf{x}, \mathbf{x}^N)]$. In addition, the prediction variance at location \mathbf{x} is given as:

$$\text{var}(\mathbf{x}) = \hat{\sigma}^2 [1 - \mathbf{r}^T(\mathbf{x}) \mathbf{R}_D^{-1} \mathbf{r}(\mathbf{x})], \quad (10)$$

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]:

$$EPE(\mathbf{x}) = \underbrace{(f(\mathbf{x}) - \hat{f}(\mathbf{x}))^2}_{\text{bias}^2} + \underbrace{\text{var}(\mathbf{x})}_{\text{variance}} \quad (11)$$

To calculate the EPE value at an arbitrary location \mathbf{x} , we need to estimate the bias term in Eq. (11) because of the unknown true response $f(\mathbf{x})$. 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 \mathbf{x}^i , $i = 1, \dots, N$:

$$e_{CV}^2(\mathbf{x}^i) = (f(\mathbf{x}^i) - \hat{f}^{-i}(\mathbf{x}^i))^2, \quad i = 1, 2, \dots, N, \quad (12)$$

where \mathbf{x}^i is the i -th training sample and \hat{f}^{-i} represents the GP model trained using all the training samples and their responses except $(\mathbf{x}^i, f(\mathbf{x}^i))$. For an arbitrary location \mathbf{x} , we then simply find the closest (in terms of Euclidean distance) training sample to location \mathbf{x} and assign its associated cross-validation error to $e_{CV}^2(\mathbf{x})$:

$$e_{CV}^2(\mathbf{x}) = e_{CV}^2(\mathbf{x}^i), \quad \min_i |\mathbf{x} - \mathbf{x}^i| \quad i = 1, 2, \dots, N \quad (13)$$

Finally, we replace the bias term in Eq. (11) with the cross-validation error:

$$EPE(\mathbf{x}) = e_{CV}^2(\mathbf{x}) + \text{var}(\mathbf{x}) \quad (14)$$

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 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 $l_1 = 224\text{ mm}$ and combustion chamber length $l_3 = 200\text{ mm}$. The operation condition considered is “A”, with total flame power $\bar{Q} = 1.94\text{ kW}$ and mean flow velocity $\bar{u}_{ref} = 2.67\text{ m/s}$ [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:

$$\nabla(\bar{c}^2 \nabla \hat{p}) - s^2 \hat{p} + \alpha s \hat{p} = s(\gamma - 1) \hat{q}(\mathbf{x}) \quad (15)$$

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:

$$\nabla \hat{p} + \beta \hat{p} = 0, \quad \text{where} \quad \beta = s \frac{1}{\bar{c}} \frac{(1 - R)}{(1 + R)} \quad (16)$$

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:

$$\frac{\hat{q}}{\bar{q}} = \frac{\hat{u}_{ref}}{\bar{u}} G \exp(i\phi) \quad (17)$$

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{q} is modeled as:

$$\frac{\hat{q}}{\bar{q}} = \frac{\hat{u}_{ref}}{\bar{u}} \sum_{k=0}^{L-1} h_k e^{-i(k+1)\Delta t(\omega - i\sigma)} \quad (18)$$

with FIR coefficients h_k 's as the input parameters. For (weakly) nonlinear analysis, an FDF model is employed, such that

$$\frac{\hat{q}}{\bar{q}} = \frac{\hat{u}_{ref}}{\bar{u}} G(\omega, |\hat{u}|) \exp(i\phi(\omega, |\hat{u}|)) \quad (19)$$

Table 1
Input parameters.

Parameters	Training range	Uncertainty range	
		Linear case	Nonlinear case
G	0.5 ~ 3	FIR	FDF
ϕ	0 ~ π		
R_{in}	0.7 ~ 1	0.7 ~ 1	0.9 ~ 1
$ R_{out} $	0.6 ~ 1	0.6 ~ 1	0.7 ~ 1
α	100 ~ 160	100 ~ 160	110 ~ 160

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. $\bar{q} = \bar{Q}/V_f$, where \bar{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.

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

Algorithm 1 GP model training via active learning

- 1: Generate L samples X_U as a pool of candidate samples for sample enrichment, set iteration number $q = 0$
- 2: Generate m initial samples $X_D = [x^1, \dots, x^m]^T$
- 3: Call Helmholtz solver to evaluate their frequency responses $y_D^{(\omega)} = [\omega^1, \dots, \omega^m]^T$ and growth rate responses $y_D^{(\sigma)} = [\sigma^1, \dots, \sigma^m]^T$
- 4: Construct the initial GP models $\hat{f}^{(\omega)} \leftarrow (X_D, y_D^{(\omega)})$ and $\hat{f}^{(\sigma)} \leftarrow (X_D, y_D^{(\sigma)})$
- 5: Identify $x^{(\omega)} \leftarrow \max_{x \in X_U} EPE(x)$ based on $\hat{f}^{(\omega)}$, let $EPE_0^\omega = EPE(x^{(\omega)})$
- 6: Identify $x^{(\sigma)} \leftarrow \max_{x \in X_U} EPE(x)$ based on $\hat{f}^{(\sigma)}$, let $EPE_0^\sigma = EPE(x^{(\sigma)})$
- 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 $y^{(\omega)} = [\omega^{(\omega)}, \omega^{(\sigma)}]$ and growth rate responses of $y^{(\sigma)} = [\sigma^{(\omega)}, \sigma^{(\sigma)}]$ of $x^{(\omega)}$ and $x^{(\sigma)}$, respectively.
- 10: Enrich samples: $X_D = X_D \cup x^{(\omega)} \cup x^{(\sigma)}$, $y_D^{(\omega)} = y_D^{(\omega)} \cup y^{(\omega)}$, $y_D^{(\sigma)} = y_D^{(\sigma)} \cup y^{(\sigma)}$
- 11: Re-train GP models $\hat{f}^{(\omega)} \leftarrow (X_D, y_D^{(\omega)})$ and $\hat{f}^{(\sigma)} \leftarrow (X_D, y_D^{(\sigma)})$
- 12: Identify $x^{(\omega)} \leftarrow \max_{x \in X_U} EPE(x)$ based on $\hat{f}^{(\omega)}$, let $EPE^\omega = EPE(x^{(\omega)})$
- 13: Identify $x^{(\sigma)} \leftarrow \max_{x \in X_U} EPE(x)$ based on $\hat{f}^{(\sigma)}$, let $EPE^\sigma = EPE(x^{(\sigma)})$
- 14: **end while**

Fig. 1. Algorithm for GP model training.

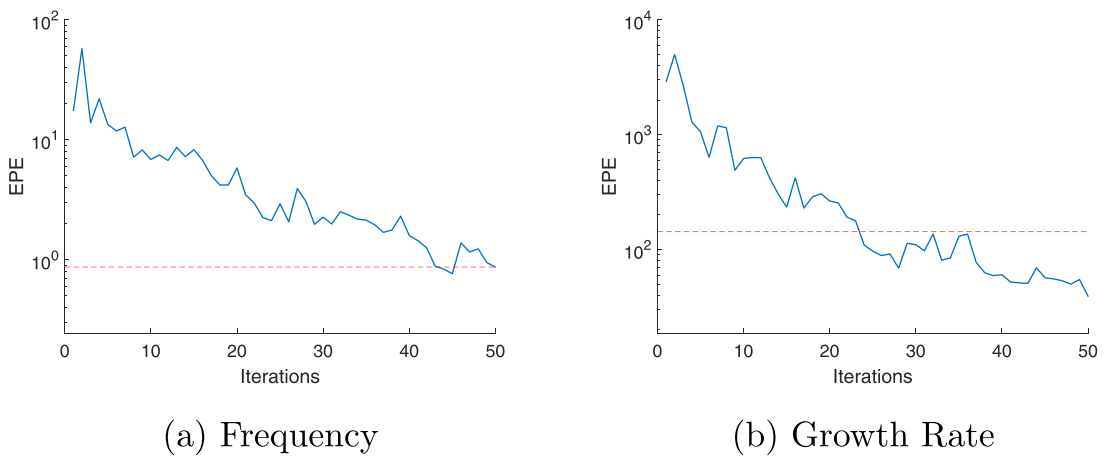


Fig. 2. EPE convergence history: After 50 iterations, maximum EPE values for both GP models drop below 5% of their initial maximum EPE values.

for GP training, which amounts to 90s of computation time. In the next two sections, these two 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 coef-

ficients $\mathbf{h} = [h_0, h_1, \dots, h_{64}]$ 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

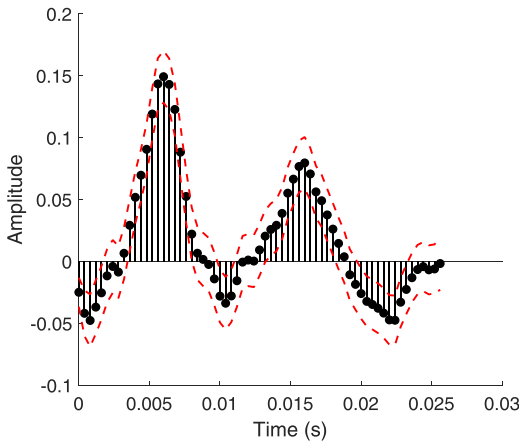


Fig. 3. Finite impulse response model. Each discrete stem represents one coefficient h_k ; upper and lower bounds constitute the 95% confidence interval.

time series of u'_{ref} and q' by forcing the system with a broadband velocity signal and a colored 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 $[\mathbf{h}, R_{in}, R_{out}, \alpha]^{(i)}, i = 1, \dots, 20,000$. We solve surrogate equations once for each sample, where $\mathbf{H} = [R_{in}, R_{out}, \alpha]$ 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

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

more significant acceleration of the UQ analysis may be achieved.

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

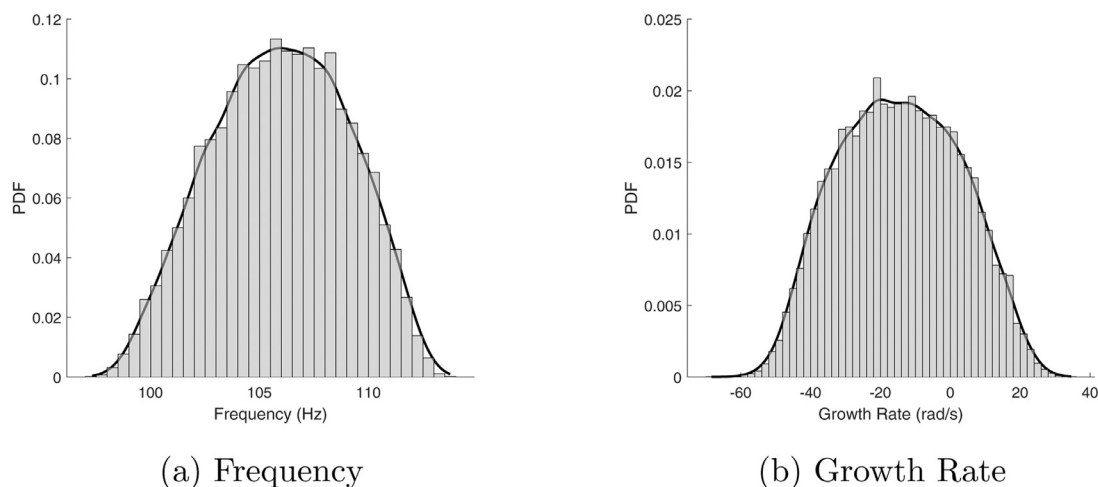


Fig. 4. PDF comparison between the GP-based UQ scheme (black curves) and Monte Carlo applied directly to Helmholtz solver (bars).

follows a normal distribution, with the mean being 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 $0 \sim 250\text{Hz}$ 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

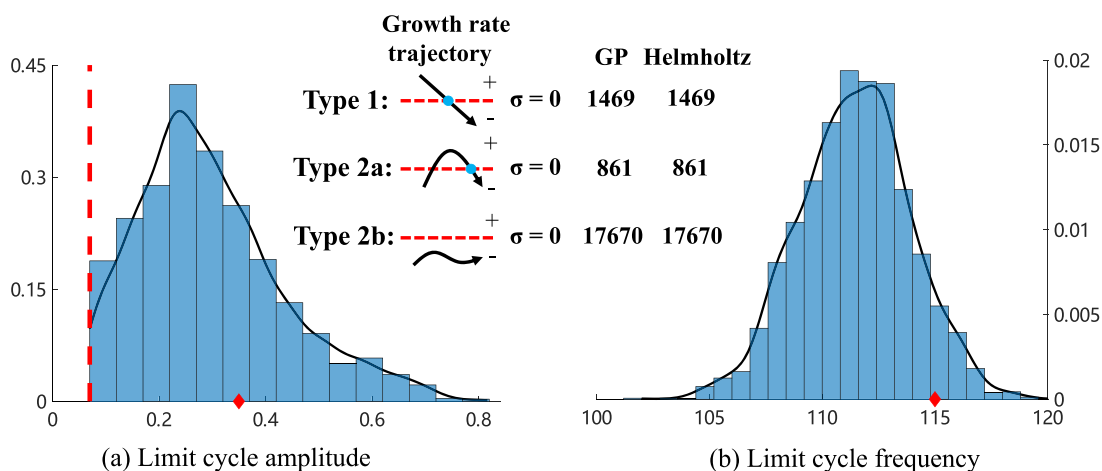


Fig. 5. PDF comparison between the GP-based UQ scheme (black curves) and Monte Carlo applied directly to Helmholtz solver (bars). Three types of growth rate trajectory are observed among 20,000 samples (Type 1: linearly unstable; Type 2a: linearly stable, nonlinearly unstable; Type 2b: linearly stable, 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.)

cases where experimental measurement uncertainties 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 emphasize that the success of GP model training depends on all the responses in 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.

Acknowledgments

S. Guo is grateful for the financial support in the form of a doctoral scholarship from the Chinese Scholarship Council (No. 201606830045).

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](https://doi.org/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](https://doi.org/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](https://doi.org/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.ccp.2016.08.043](https://doi.org/10.1016/j.ccp.2016.08.043).
- [7] S. Guo, C.F. Silva, A. Ghani, W. Polifke, *J. Eng. Gas Turbines and Power* 141 (2) (2019), doi:[10.1115/1.4041652](https://doi.org/10.1115/1.4041652). 021032–10
- [8] S. Guo, C.F. Silva, M. Bauerheim, A. Ghani, W. Polifke, *Proceedings of the Combustion Institute* 37 (2019) 5299–5306, doi:[10.1016/j.proci.2018.07.020](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1115/1.4040210).
- [13] S. Guo, C.F. Silva, W. Polifke, *J. Eng. Gas Turbines and Power* (2019), doi:[10.1115/1.4044197](https://doi.org/10.1115/1.4044197).
- [14] T.F. Coleman, Y. Li, *SIAM Journal on Optimization* 6 (2) (1996) 418–445, doi:[10.1137/0806023](https://doi.org/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](https://doi.org/10.1016/j.compchemeng.2017.05.025).
- [17] W. Polifke, *Prog. Energy Combust. Sci.* 79 (2020) 100845, doi:[10.1016/j.peccs.2020.100845](https://doi.org/10.1016/j.peccs.2020.100845).
- [18] W. Polifke, *Annals of Nuclear Energy* 67C (2014) 109–128, doi:[10.1016/j.anucene.2013.10.037](https://doi.org/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](https://doi.org/10.1080/13647830.2018.1443517).
- [20] M. McCartney, M. Haeringer, W. Polifke, in: *GT2019-91319*, ASME, Phoenix, Arizona, 2019, p. 10, doi:[10.1115/GT2019-91319](https://doi.org/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](https://doi.org/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](https://doi.org/10.1017/S0022112008003613).
- [23] B. Gustavsen, A. Semlyen, *IEEE Transactions on Power Delivery* 14 (3) (1999) 1052–1061, doi:[10.1109/61.772353](https://doi.org/10.1109/61.772353).