



HAL
open science

Forward and backward uncertainty quantification with active subspaces: application to hypersonic flows around a cylinder

Andrea Cortesi, Paul Constantine, Thierry E. Magin, Pietro Marco Congedo

► **To cite this version:**

Andrea Cortesi, Paul Constantine, Thierry E. Magin, Pietro Marco Congedo. Forward and backward uncertainty quantification with active subspaces: application to hypersonic flows around a cylinder. [Research Report] RR-9097, INRIA Bordeaux, équipe CARDAMOM. 2017. hal-01592591

HAL Id: hal-01592591

<https://inria.hal.science/hal-01592591>

Submitted on 25 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Forward and backward uncertainty quantification with active subspaces: application to hypersonic flows around a cylinder

Andrea F. Cortesi , Paul G. Constantine, Thierry E. Magin, Pietro M. Congedo

**RESEARCH
REPORT**

N° 9097

August 1 2017

Project-Teams Cardamom



**Forward and backward uncertainty
quantification with active subspaces:
application to hypersonic flows around a
cylinder**

Andrea F. Cortesi ^{*}, Paul G. Constantine[†], Thierry E. Magin[‡],
Pietro M. Congedo[§]

Project-Teams Cardamom

Research Report n° 9097 — August 1 2017 — 37 pages

^{*} INRIA Bordeaux Sud-Ouest, 200 Rue de la Vieille Tour, 33405 Talence, France

[†] Department of Computer Science, University of Colorado, Boulder, CO 80309, USA

[‡] von Karman Institute for Fluid Dynamics, Chaussée de Waterloo 72, 1640 Rhode-Saint-Genèse, Belgium

[§] INRIA Bordeaux Sud-Ouest, 200 Rue de la Vieille Tour, 33405 Talence, France

**RESEARCH CENTRE
BORDEAUX – SUD-OUEST**

351, Cours de la Libération
Bâtiment A 29
33405 Talence Cedex

Abstract: We perform a Bayesian calibration of the freestream velocity and density starting from measurements of the pressure and heat flux at the stagnation point of a hypersonic high-enthalpy flow around a cylinder. The objective is to explore the possibility of using stagnation heat flux measurements, together with pressure measurements, to rebuild freestream conditions since such measurements are available for recent space missions but not exploited for freestream characterization. First, we formulate an algorithm of mesh adaptation, enabling accurate numerical solutions in an automatic way for a given set of inputs. Secondly, active subspaces are used to find a low-dimensional dependence structures in the input-to-output map of the forward numerical solver. Then, surrogate models on the active variables are used to accelerate the forward uncertainty propagation by Monte Carlo sampling and the Markov Chain Monte Carlo sampling of the posterior distribution for Bayesian inversion. A preliminary sensitivity analysis with sparse Polynomial Dimensional Decomposition is performed on the chemical model of the air mixture, to determine the most influential uncertain chemical parameters in the forward problem. Then, the forward and backward methodologies are applied to the simulation of a hypersonic flow around a cylinder, in conditions for which experimental data are available, revealing new insights towards the potential exploitation of heat flux data for freestream rebuilding.

Key-words: Hypersonic Flows, Inverse problems, Bayesian inference, Uncertainty Quantification, Active Subspaces, Surrogate modeling

Quantification des incertitudes directe et inverse avec "active subspace: application aux écoulements hypersoniques autour d'un cylindre"

Résumé : Nous effectuons une calibration bayésienne de la vitesse et de la densité au lointain à partir des mesures de la pression et du flux de chaleur au point de stagnation dans le cadre d'un écoulement hypersonique à haute enthalpie autour d'un cylindre. L'objectif est d'explorer la possibilité d'utiliser des mesures de flux de chaleur de stagnation, ainsi que des mesures de pression, pour reconstruire les conditions de freestream, car ces mesures sont disponibles pour les missions spatiales récentes mais non exploitées pour la caractérisation au lointain. Tout d'abord, nous formulons un algorithme d'adaptation de maillage, permettant des solutions numériques précises de manière automatique pour un ensemble de données d'entrée. Deuxièmement, les "active subspaces" sont utilisés pour trouver une structure à faible dimension. Ensuite, les modèles substitués sur les variables actives sont utilisés pour accélérer la propagation de l'incertitude. Ensuite, les méthodes sont appliquées à la simulation d'un écoulement hypersonique autour d'un cylindre, dans des conditions pour lesquelles des données expérimentales sont disponibles, révélant de nouvelles idées sur l'exploitation potentielle des données de flux de chaleur pour la reconstruction du freestream.

Mots-clés : Écoulements hypersoniques, problème inverse, inférence bayésienne, quantification des incertitudes, Active Subspace, fonction substituée.

Contents

1	Introduction	4
2	Formulation of the forward problem and CFD tool	8
2.1	Mesh convergence assessment in a UQ framework	9
2.2	Comparison with experimental data in the nominal conditions . .	12
3	Uncertainty Characterization	13
3.1	Comparison between air5 and air11 models	14
3.2	Sensitivity analysis on air5 chemical mechanism	15
4	Forward and Backward method: Dimension reduction and Bayesian inference	16
4.1	Bayesian setting for freestream calibration	17
4.2	Finding active subspaces for dimension reduction	18
4.3	Exploiting active subspaces in MCMC for Bayesian inversion . .	20
5	Results	22
5.1	Active Subspace for stagnation pressure and heat flux	22
5.1.1	Comparison with classical global Sensitivity Analysis . . .	24
5.2	Forward propagation of uncertainties	26
5.3	Freestream calibration by means of Bayesian inference	27
6	Conclusions	31

1 Introduction

The simulation of several engineering systems requires the numerical resolution of complex computational models, which often involves a large number of physical parameters. In some practical cases, it can be difficult, or even impossible, to obtain enough experimental data to adequately calibrate the parameters of the physical model and furthermore, in general, experimental data, when available, are intrinsically affected by measurement errors. The need for robust predictive simulations in the analysis and design process requires to quantify how these uncertainties affect the values of some quantities of interest, *i.e.* the development of efficient Uncertainty Quantification (UQ) methods.

An example of physical system where UQ is highly relevant is high-enthalpy hypersonic fluid dynamics [1, 2, 3, 4, 5], a complex problem that involves multidisciplinary aspects, from aero-thermodynamics to gas-surface interaction, and in which some experimental data are difficult to obtain, especially in-flight measurements of entering space vehicles. For example, during the reentry trajectory of a space vehicle, it is extremely challenging to obtain direct measurements of freestream quantities (velocity and thermodynamic state of the atmosphere), due to the presence of a strong bow shock in front of the vehicle nose. However, these quantities are of relevant importance in both post-flight analysis, to get accurate values of the actual flight trajectory and atmospheric conditions, and during the flight to be able to make adequate real-time corrections in case of manned flights. Different techniques for freestream characterization in post-flight analysis have been discussed in literature and applied to real flight

data [6, 7, 8]. They usually rely on Inertial Measurement Unit (IMU) and wall pressure data for the characterization of the freestream.

The motivation of this paper is to assess the interest in using heat flux measurements, which are available in recent space missions, for rebuilding freestream conditions, requiring the set-up of multiple numerical tools for tackling several issues which are analyzed and cured here. In a recent work, Tryoen et al. [9] proposed a novel approach which was exploiting wall pressure and heat flux data, and applied it to the entry trajectory of the European Experimental Reentry Test-Bed (EXPERT) vehicle. Nevertheless, the authors encountered some difficulties in the training of an adequate surrogate model for the stagnation heat flux, possibly due to a poor mesh convergence for some input conditions explored during the UQ study. This point is specifically tackled in this paper. In fact, the thermal load at the object surface can be very difficult to compute [10] and needs monitoring, especially near the shock/mesh alignment. Moreover, in a UQ study, multiple computations are required for different combinations of the input parameters, making the monitoring of the convergence of each simulation very complicated. The first contribution of this paper is the setting of a robust algorithm permitting to refine meshes in an automatic way for each sample in the stochastic space during a UQ study.

The main objective of this paper is to assess the resolvability of the problem of freestream calibration starting from stagnation pressure and heat flux data, in the same Bayesian framework as the one introduced by [9]. The posterior distribution, sampled by Markov Chain Monte Carlo accelerated by a surrogate model, trained the lower-dimensional input space constituted by the active variables. The configuration under study is the hypersonic flow around a cylinder in the HEG I configuration, described in [11, 10]. It is an experimental configuration that can be representative of some real-life hypersonic entry flows. The reason for this choice is that this is a simpler and known configuration for which experimental data is available [11] and that has already been used in literature as validation case for hypersonic CFD [10]. Therefore, this enables us to set and test an appropriate numerical simulation framework to obtain sufficiently accurate and automated evaluations of the Quantities of Interest (QoIs) in the sampling points, which are then exploited to train the metamodel. Figure 1 illustrates a sketch of the main steps and components involved in the Bayesian freestream characterization and forward uncertainty propagation problems treated in this work and their relationship.

Several sources of uncertainties have to be considered in the simulation of hypersonic entry flows: chemistry model parameters and freestream conditions have to be considered as uncertain input parameters in the forward problem, *i.e.* the CFD simulation of the HEG cylinder hypersonic flow. Moreover, in reacting flows, several constitutive relations are required to describe the chemical mechanism characterizing the gas mixture, each of which contains a large number of parameters [12]. These parameters are often the result of experimental measurements and have an associated variability. Note that direct measurements are available for pressure and heat flux data on the nose of the space vehicle, but not for the freestream quantities.

In the context of calibration of parameters that can not be directly observed, as for instance the freestream data, Bayesian inference [13, 14] is a powerful mathematical tool. This setting has received much interest in literature for application in several fields, and has been recently introduced and tested

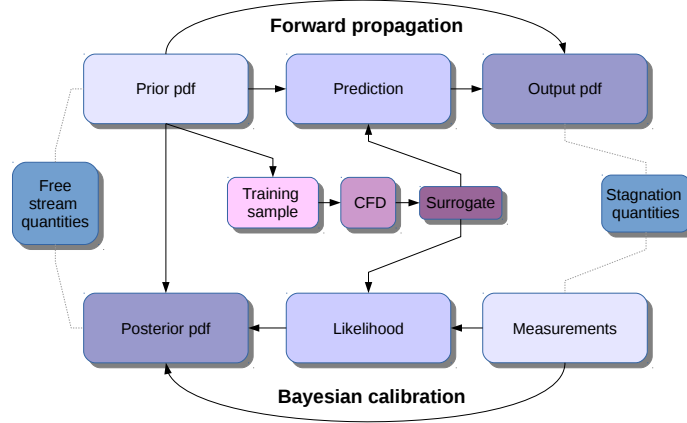


Figure 1: Diagram explaining inputs, outputs and different actors in the forward propagation and Bayesian calibration problems.

also in application related to flight condition inference in hypersonic scramjet combustor flow [15] and, as already discussed, for atmospheric entry flows [9]. Bayesian calibration offers a rigorous mathematical approach for inferring parameters values from noisy observations and accounting for uncertainties in the forward model describing the system. The output of the inverse problem in this stochastic framework is no more a single deterministic value for each inferred parameter but a *posterior* probability density function that encompasses all the available informations about the prior uncertainties and measurement errors. The technique is versatile in the sense that it can employ very complex and accurate models of the physical system and it provides a comprehensive probabilistic information about the rebuilt quantities. However, there are several well known drawbacks.

A first practical problem that arises in Bayesian calibration is that several thousands of evaluations of the quantities of interest can be needed, such as for example when Markov Chain Monte Carlo (MCMC) are used to sample the posterior density in Bayesian inverse problems. In such applications, unrealistic computational effort is required to simulate the phenomena in each configuration with complex accurate numerical tools. This leads to the employment of surrogate models [16, 17] trained on a limited number of selected solutions, to represent the functions of interest. The training process of an accurate surrogate is crucial to get meaningful UQ results. But it can be difficult, especially when the function of interest is complex.

A second issue related to both forward and backward UQ problems is the so-called *curse of dimensionality* [18], where the cost of a complete exploration of input parameter space scales exponentially with the dimension of the space itself. Different techniques are investigated in literature to tackle this problem by finding and exploiting lower-dimensional structures in the map between input parameters and quantities of interest. The focus of this paper is on *active subspaces* [19, 20], which could permit to enhance the efficiency of the surrogate model, without degrading the accuracy of UQ propagation results. However,

the existence of an active subspace for a function of interest depends on the regularity of the function and its behavior with respect to the inputs. For ill-posed nonlinear inverse problems, low-dimensional subspaces of the parameter space that are informed by the data can be exploited to fight the curse of dimensionality. If such subspaces exist in the given inverse problem, then the MCMC methods can focus the parameter space exploration on the coordinates of the data-informed low-dimensional subspaces. Several recent works seek to identify and exploit such subspaces for Bayesian inverse problems including the *likelihood-informed subspace*-based approaches [21, 22, 23] and *active subspace*-based approaches [24]. Broadly, the difference between these two approaches lies in the construction of the subspaces: likelihood-informed subspaces are eigenspaces of the Gauss-Newton approximation of the log-likelihood's Hessian (*i.e.*, second derivatives) averaged with respect to the posterior, while active subspaces are eigenspaces of the outer product of the log-likelihood's gradient (*i.e.*, first derivatives) with itself averaged with respect to the prior. In general, these two subspaces are distinct, and the two approaches have their own distinct computational advantages and disadvantages. Our approach in this paper takes a different angle on the active subspace-based approach. Instead of using the gradient of the log-likelihood to define the active subspace, we find an independent subspace for each of the parameter-to-observable maps, that we combine for a data-informed subspace.

Finally, the proposed forward and backward methodologies are applied to the simulation of a hypersonic flow around a cylinder, proposing for the first time in literature a comprehensive numerical platform permitting a feasibility study over the potential use of heat flux data in the rebuilding of freestream conditions.

The paper is structured as follows. In Section 2, a description of the physical configuration of interest is provided, together with the numerical method used to solve the system of equations. In Section 2.1, an algorithm permitting an automatic mesh convergence for each sample in the stochastic space is illustrated. Then, in 2.2, CFD simulations in nominal conditions are validated against experimental data. Section 3 is devoted to the uncertainty characterization of the experiment of interest. In particular, we present in Section 3.1 a justification of the choice of five species chemical model for the free stream conditions considered in this work, while a sensitivity analysis to select which uncertainties on reaction rate coefficients included in the final UQ propagation and inference problems is presented in Section 3.2. Section 4 gives a description of the mathematical framework used to perform the inverse problem: first of all the Bayesian setting is briefly recalled in 4.1, then active subspaces are reviewed in 4.2 and finally in 4.3 a detailed description is given on how to exploit the active subspace in Bayesian inference. Section 5 analyzes the results obtained for the HEG cylinder application: Sec. 5.1 shows results of the use of active subspaces on the quantities of interest, in 5.1.1 global sensitivity indices are compared to the active subspace, Sec. 5.2 presents results of the forward uncertainty propagation and results of the Bayesian inference are shown in 5.3. Finally, Section 6 draws some conclusions and perspectives.

2 Formulation of the forward problem and CFD tool

The problem of interest in this paper deals with the hypersonic high-temperature reacting aerodynamic flow around a cylinder. In this case, the reference experiment is the experimental configuration used at the HEG facility (High Enthalpy shock tunnel Göttingen), which has been described and tested in [11] and then used in [10], to assess and compare the predictions of different CFD codes. An image of the experimental setup is reported in Figure 2.

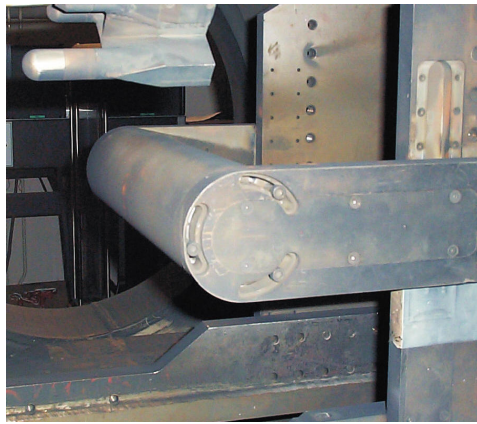


Figure 2: Image of the cylindrical experimental model in the HEG facility (taken from [11]).

The forward problem consists in computing the quantities of interest, namely the pressure p_{st} and heat flux q_{st} at the stagnation point, starting from the freestream conditions of the HEG I configuration [10] described in Table 1. The set of equations used to describe the phenomena is a combined physico-

Quantity	Value
u_∞ [m/s]	5956
ρ_∞ [kg/m ³]	1.547e-3
p_∞ [Pa]	476
T_∞ [K]	901
M_∞	8.98
$Y[N_2]_\infty$	0.75430704
$Y[O_2]_\infty$	0.00713123
$Y[NO]_\infty$	0.01026010
$Y[N]_\infty$	6.5e-7
$Y[O]_\infty$	0.22830098
T_w [K]	300

Table 1: Nominal freestream conditions and wall temperature for the simulation configuration.

chemical model, developed by Barbante in [25], able to simulate hypersonic high-temperature reacting flows. Two-dimensional Navier-Stokes equations, supplied with adequate boundary conditions, are combined with the chemical mechanism introduced by Park et al. [12] applied to a mixture of five species air (N, O, NO, N₂ and O₂). Five species are sufficient to represent the main chemical phenomena involved in the flow for the considered configuration, as motivated in Section 3.1. Therefore this simpler method is preferred over the Park model with 11 species. Furthermore, the catalyticity of the vehicle surface is taken into account, and it is modeled as a catalytic wall at imposed wall temperature T_w , with recombination coefficient $\gamma = 1$. The numerical method used to solve the governing equation is described in Section 2.1, together with the mesh adaptation technique used to obtain meaningful results.

To simulate the hypersonic flow around the cylinder, the COSMIC code developed by Barbante [25] is used. This solver was designed to approximate hypersonic flow models where chemical non-equilibrium effects need to be accounted for. It consists in a hybrid upwind splitting scheme, the hybridization of the van Leer scheme [26] and the Osher scheme [27] and includes a carbuncle fix. A planar symmetry condition is imposed on the symmetry axis, while the wall of the body is modeled by a catalytic wall at imposed wall temperature.

2.1 Mesh convergence assessment in a UQ framework

To perform the uncertainty propagation and sensitivity analysis of the HEG hypersonic flow around the cylinder, it is necessary to evaluate the stagnation quantities of interest for every sampling point in the stochastic space of uncertain parameters. The numerical simulations associated to those values are then required to be sufficiently accurate and reliable, otherwise fluctuations of the QoI from numerical errors could be misinterpreted as variations due to input uncertainties. A structured mesh of 99×26 cells has been chosen as nominal grid. The number of nodes should be a good compromise between solution accuracy and computational efficiency, since the simulation needs to be repeated several times for the UQ study. Note that UQ propagation through a CFD code is often performed by using a fixed mesh, since a converged mesh in nominal conditions can be reasonably considered as converged also if some variations of the operating conditions are taken into account. This approach is usually sufficient in the absence of shock waves. As it will be shown in the following, this can yield highly inaccurate results in hypersonic flows.

In fact, in the numerical simulation of hypersonic flows by means of a second-order accuracy cell-centered finite-volume discretization method, it is known that a proper alignment between the shock and the discretization grid is essential to get a meaningful solution. However, in an uncertainty quantification framework, the values of free stream conditions are subject to variations, which leads to different shock stand-off distances and different shock shapes for each sample of input parameters, and so to a mesh/shock misalignment, causing false heat flux trends, as clearly noticeable in figure 6b. From the same figure, it is clear that it is not possible to just use the nominal mesh, even if well refined, for all the simulations at different values of input parameters. But it is essential to adopt a simple mesh adaptation technique to realign the shock to the mesh for each different freestream configuration.

Here we describe briefly the adaptation technique adopted in this work, even

if it is a very simple tool that uses basic algorithms well-known to practitioners in the hypersonic flows field. However, it is important to stress out the fact that here the process needs to be automated and can not be directly monitored by the user, because several simulations need to be run. A first solution is computed

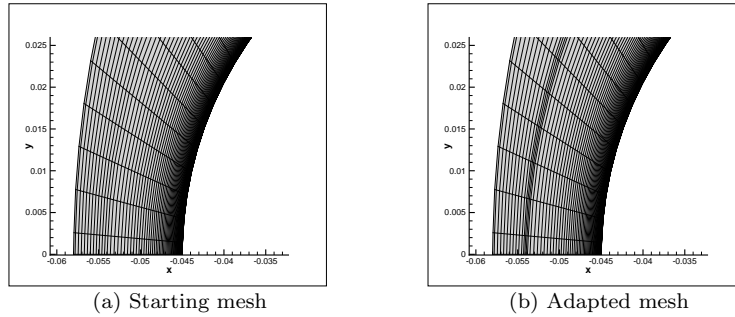


Figure 3: Comparison between a zoom of the starting nominal mesh close to the symmetry axis (left) and the same area of the adapted mesh for one of the training points (right).

on the nominal mesh, and the shock position is found by looking at the Mach number in the flow field. Then, the adaptation is performed by solving 1D linear elasticity equations along the radial directions. The alignment to the shock is enforced by imposing a rigid constraint on a row of nodes corresponding to the shock position, and also the wall surface and outer bound shape are constrained in the same way. No forcing is considered. If a small refinement near the shock is needed, two rows of nodes are added by splitting existing cells close to the shock during the first iteration, then the regularization is automatically done when performing successive iterations of the adaptation. In our case, the addition of nodes is performed just at the first adaptation step. The different steps of the algorithm are summarized in Algorithm 1.

Algorithm 1 Mesh adaptation algorithm to automatically ensure a good alignment between the shock and the mesh.

Starting from a nominal mesh and a starting solution computed at nominal condition,

1. Compute a CFD solution with the value of input parameters for a specific sample in the stochastic space
2. Find the shock position by looking at the jump in Mach number in the flow field
3. Adapt the mesh solving the solution of 1D linear elasticity equations along the radial directions, imposing a rigid constraint on a row of nodes corresponding to the shock position, and also at the wall surface and outer bound shape

Repeat this steps until the maximum number of iterations is reached.

Figure 3 shows the comparison of a zoom between the starting nominal

mesh close to the symmetry axis and the same area of the mesh adapted for input conditions corresponding to a different sample in the stochastic space. The solution temperature field obtained at the end of the adaptation process is shown in figure 4a together with the temperature profile on the stagnation line in figure 4b. An example the comparison of heat flux solution on the cylinder

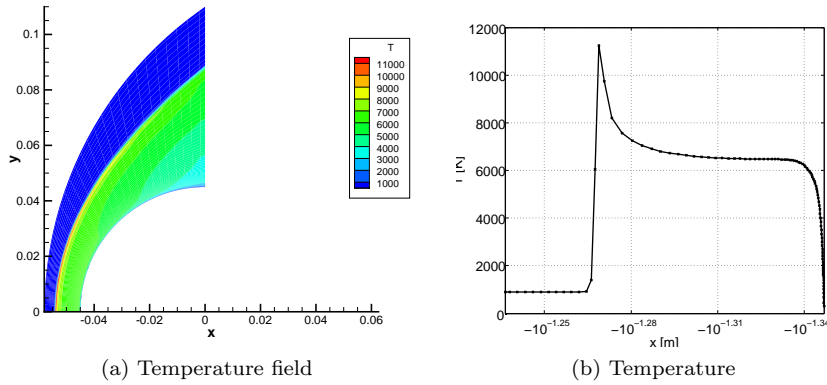


Figure 4: Solution temperature field (a) around the cylinder computed on the adapted mesh and temperature profile (b) on the stagnation line

surface before and after mesh adaptation is shown in figure 5. A non negligible difference in heat flux values, especially close the stagnation point, is noticeable.

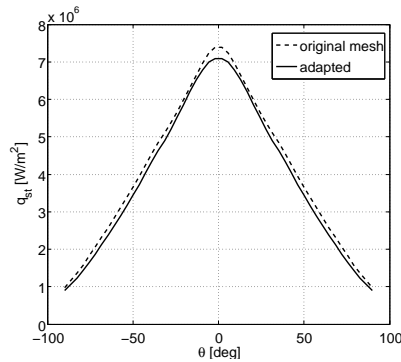


Figure 5: Solution heat flux around the cylinder surface computed on the original mesh compared to the one obtained at the end of the adaptation profile, for a speed of 5929 m/s.

To assess the algorithm efficiency, several meshes are built by considering different freestream conditions, *i.e.* the nominal conditions except for the variation of the freestream velocity, which is assumed to vary randomly in the interval of $\pm 20\%$ of the nominal velocity value. Note that three steps of the Algorithm 1 are performed. Then, the evolution of the heat flux at the stagnation point

with respect to the freestream velocity is observed for several meshes. In particular, we compare the results obtained with adaptation and different fixed meshes over the different conditions. Note that the comparison is done also with computations performed on 119×99 mesh, as reported in [10].

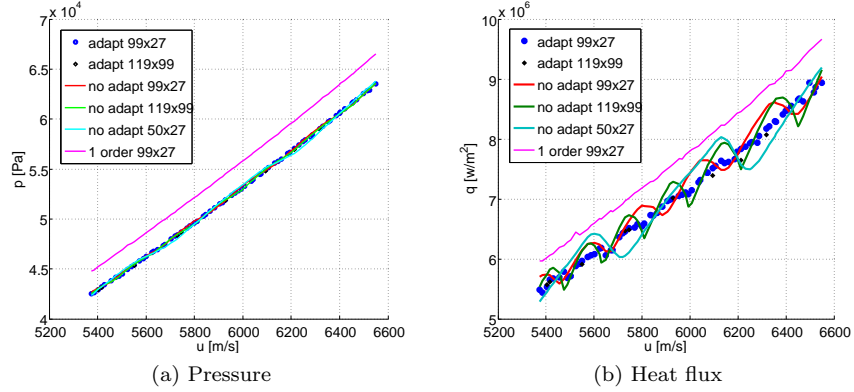


Figure 6: Comparison between numerical results obtained with grids of increasing size, to test the mesh convergence.

From the comparison (Figure 6), it can be observed the necessity to adapt the mesh over each different condition, while the use of a fixed mesh yields a large numerical error. As a consequence, it is possible to state that the Algorithm 1 allows an automatic mesh convergence assessment, and the highest influence on the output value is caused by the proper alignment between the shock and the grid, rather than the grid refinement itself. From these results, it is possible to estimate that the numerical error due to mesh alignment associated to the output value of the heat flux is contained in the $\pm 2\%$ interval, while, as known, the stagnation pressure is much less affected by this issue.

2.2 Comparison with experimental data in the nominal conditions

A comparison with the experimental data presented in [11, 10] is used to validate the convergence of the simulations at nominal conditions. The nominal mesh with three steps of the adaptation algorithm is used to compute the solution. In Figure 7, the pressure and heat flux at the wall of the cylinder are compared with experimental data. For the sake of comparison, for the heat flux some simulations from [10] are reported too, together with a computation performed with COSMIC code on a finer 119×99 mesh. It can be noticed that the agreement is very good for the pressure, and quite good for the heat flux, but worse than the pressure, as expected due to the known difficulties in the simulation of the heat flux at the wall for hypersonic flows. Other codes, such as the one of Nompelis [10] show better performance, while the refined mesh does not improve substantially the result. However, results can be considered satisfying, especially near the stagnation, which is where we will focus for the UQ study.

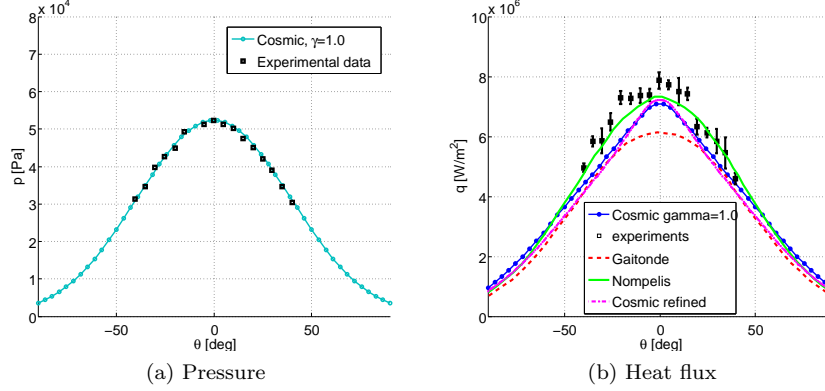


Figure 7: Comparison between numerical and experimental data for the nominal conditions.

3 Uncertainty Characterization

The characterization of the uncertainties is here performed, basing on the experimental configuration under study, which has been depicted in Figure 2. The first source of uncertainty is the chemical mechanism used to model nonequilibrium effects. Uncertainty is taken into account on preexponential factors A_r for the rate coefficients of 6 elementary chemical processes (dissociation of N_2 and NO and exchange reaction of formation of NO). These coefficients are assumed distributed with log-normal distributions [28] (see Table 2). The choice of the atmospheric model and uncertain variables is based on the analysis presented in the following Sections 3.1 and 3.2.

Variable	Index	Gas reaction	Density of $\log_{10} A_r$	σ_r
A_1	13	$NO + O \rightarrow N + O + O$	Normal	0.12
A_2	12	$NO + N \rightarrow N + O + N$	Normal	0.12
A_3	16	$N_2 + O \rightarrow NO + N$	Normal	0.10
A_4	4	$N_2 + O \rightarrow 2N + O$	Normal	0.15
A_5	5	$N_2 + N \rightarrow 2N + N$	Normal	0.15
A_6	14	$NO + N_2 \rightarrow N + O + N_2$	Normal	0.12

Table 2: Uncertainties on gas reaction rates. The index is referred to the notation used in Section 3.2.

We consider to have a second source of uncertainty related to freestream conditions. In particular, the freestream density ρ_∞ and velocity u_∞ are assumed to be unknown and to follow uninformative uniform distributions, as described in Table 3. The uncertainties on ρ_∞ and u_∞ are chosen to be of 20% of their nominal value. Note that these ranges are considered for the forward UQ problem and as priors for the inverse problem. These priors are chosen so that the standard deviations of the quantities of interest obtained by propagation of the uncertainties (see Section 5.2) are bigger than the ones associated to measurement errors. This because we do not want the prior to be informative

and to have a strong impact on the posterior result of the Bayesian inference.

Variable	Distribution	Minimum	Maximum
ρ_∞ , kg/m ³	Uniform	$1.237 \cdot 10^{-3}$	$1.856 \cdot 10^{-3}$
u_∞ m/s	Uniform	4764.8	7147.2

Table 3: Uncertainties on freestream conditions.

Hence, a total number of 8 input uncertain variables is taken into account for the sensitivity analysis, forward propagation and Bayesian inference.

Another source of uncertainty is related to the experimental measurement errors which are associated to pressure and heat flux data at the stagnation point. Experimental data are considered to be affected by Gaussian noise. The standard deviation model for the stagnation pressure noise is $\sigma_p = 0.02 p_{st}^*$, and the standard deviation model for stagnation heat flux is $\sigma_q = 0.1 q_{st}^*$; where p_{st}^* and q_{st}^* are the measured values. The differences in noise levels model the engineers' trust in the sensors for pressure and heat flux; in other words, we expect roughly 2% error in pressure measurements and 10% error in heat flux measurements. These errors will be used to construct the likelihood model for the Bayesian inference.

In the following two sections, a comparison and sensitivity analysis is performed on the chemical model for the Earth atmosphere, in order to fulfill the conditions associated to the experiment. The goal is to choose the proper model for the following CFD computations (Section 3.1) and to select, among the several chemical reactions constituting the mechanism, the ones that most influence the variability of the species mass fractions when considering uncertainties on the Arrhenius coefficients (Section 3.2).

3.1 Comparison between air5 and air11 models

A first step in the analysis of the atmospheric chemistry model is to verify which species are mostly influent in the post shock region, to be able to choose the simpler mixture able to represent the phenomena. Here, the Park's air model [12] with 5 species (air5) is compared to the one with 11 species (air11). Since, in [10], every simulation is performed with just 5 species air model, we are expecting to observe a good agreement between the two different mixtures.

For this analysis, a simpler simulation of the post-shock flow field is performed by means of the Shocking code (details about this code are provided in [29]), which computes the downstream flow field using the 1D conservation equations of mass, momentum and global energy plus the conservation of vibrational energy of the nitrogen molecules. This code is not able to account for the deceleration due to the body and the stand-off distance is imposed a priori to an approximate value equal to 9.5 mm.

The comparison between the two gas mixtures is done by looking at the differences between mass fractions of the common species and by checking if a strong presence of ionized species, typical of air11 mixture, is observed, especially at the position corresponding to the stand off distance. As it can be noticed from Figure 8, the two mixtures behave practically in the same way, and the mass fractions of ionized species, and especially electrons, are very

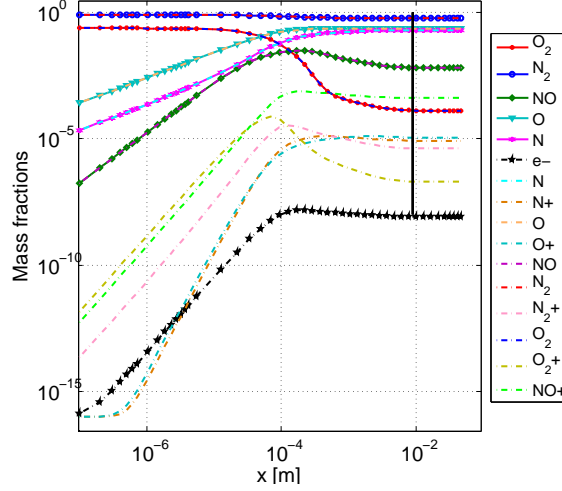


Figure 8: Comparison of mass fraction post-shock variations between air 5 and air 11 mixtures.

small with respect to the ones of the species composing the air5 mixture. Hence it is safe to continue the analysis on the simpler air5 mixture, knowing that the main chemical characteristics of the flow are correctly described.

3.2 Sensitivity analysis on air5 chemical mechanism

Here, the two Quantities of Interest chosen for the sensitivity analysis are the mass fractions of nitrogen $[N]_w$ and oxygen $[O]_w$ at a distance from the shock corresponding to the stand-off distance d_{so} between the shock and the cylinder wall. These two QoIs are chosen since they have a major influence in the recombination of atoms at the surface of a reusable thermal protection material.

Seventeen (17) uncertain variables are considered for this problem, which are the preexponential factors A_r of the Arrhenius rate coefficients $k_r = A_r T^\beta \exp\{-C_r/T\}$, because other parameters such as the activation energy are supposed to be known with a higher accuracy. They are supposed to be distributed with a lognormal distribution centered on their mean values and standard deviations given in Table 4, which corresponds to consider a normal uncertainty on their exponent. The freestream variables are fixed to their nominal values described in Table 1, and the standoff distance is taken $d_{so} = 0.0095\text{m}$.

Latin hypercube sampling [30] is used to select 20000 sampling points in the stochastic input space, and the Shocking code is run for each sample to get the values of the outputs for each input condition. Then, a sparse Polynomial Dimensional Decomposition (PDD) [31] is built on the QoIs in order to get their first-order sensitivity indices.

In Tables 5 and 6, the sorted first order sensitivity indices for the two quantities of interest are shown. This illustrates the hierarchy of the uncertain Arrhenius coefficients with respect to their contribution to the variance of the outputs. It can be noticed that the first two reactions, namely the dissociation of NO with third body being respectively O and N, contribute to about the 98%

Variable	Reaction	Deviation σ
1-3	$N_2 + M \rightarrow 2N + M$ where $M = N_2, NO, O_2$	0.11
4,5	$N_2 + M \rightarrow 2N + M$ where $M = O, N$	0.15
6-10	$O_2 + M \rightarrow 2O + M$ where $M = N_2, NO, O_2, O, N$	0.10
11-15	$NO + M \rightarrow N + O + M$ where $M = NO, N, O, N_2, O_2$	0.12
16	$N_2 + O \rightarrow NO + N$	0.10
17	$NO + O \rightarrow O_2 + N$	0.10

Table 4: Standard deviations of the pdfs for the reaction rate coefficients pre-exponential factors.

Variable	Reaction	SI 1st order	% Variance
13	$NO + O \rightarrow N + O + O$	0.5128	0.51287
12	$NO + N \rightarrow N + O + N$	0.4636	0.97657
16	$N_2 + O \rightarrow NO + N$	0.9996e-2	0.98656
4	$N_2 + O \rightarrow 2N + O$	0.4294e-2	0.99086
5	$N_2 + N \rightarrow 2N + N$	0.3559e-2	0.99442
14	$NO + N_2 \rightarrow N + O + N_2$	0.3412e-2	0.99783
9	$O_2 + O \rightarrow 2O + O$	0.6639e-3	0.99849
10	$O_2 + N \rightarrow 2O + N$	0.5532e-3	0.99905
11	$NO + NO \rightarrow N + O + NO$	0.3693e-3	0.99942
1	$N_2 + N_2 \rightarrow 2N + N_2$	0.2516e-3	0.99967
6	$O_2 + N_2 \rightarrow 2O + N_2$	0.1309e-3	0.99980
17	$NO + O \rightarrow O_2 + N$	0.1051e-3	0.99990
7	$O_2 + NO \rightarrow 2O + NO$	0.5214e-4	0.99996
3	$N_2 + O_2 \rightarrow 2N + O_2$	0.1544e-4	0.99997
2	$N_2 + NO \rightarrow 2N + NO$	0.1157e-4	0.99998
15	$NO + O_2 \rightarrow N + O + O_2$	0.6492e-5	0.99999
8	$O_2 + O_2 \rightarrow 2O + O_2$	0.6155e-5	1.0000

Table 5: Sensitivity Indices for HEG cylinder chemistry, $QoI [N]_w$.

of the variance of the QoIs in this particular configuration. For this reason, it is important to take into account at least the uncertainties associated to these two reactions in the following UQ analysis performed by means of CFD tools. For sake of completeness, we consider also the four following reactions, thus arriving to the 99.7% of the output variance contribution.

4 Forward and Backward method: Dimension reduction and Bayesian inference

In this section, first the Bayesian framework used for freestream conditions calibration is described. Then, the active subspaces are presented for exploring the possibility to reduce the dimensionality of input space of uncertain variables. If a low-dimensional structure exists, it can be usefully exploited to accelerate

Variable	Reaction	SI 1st order	% Variance
13	$\text{NO} + \text{O} \rightarrow \text{N} + \text{O} + \text{O}$	0.5165	0.51652
12	$\text{NO} + \text{N} \rightarrow \text{N} + \text{O} + \text{N}$	0.4683	0.98489
4	$\text{N}_2 + \text{O} \rightarrow 2\text{N} + \text{O}$	0.3978e-2	0.98887
14	$\text{NO} + \text{N}_2 \rightarrow \text{N} + \text{O} + \text{N}_2$	0.3450e-2	0.99232
5	$\text{N}_2 + \text{N} \rightarrow 2\text{N} + \text{N}$	0.3284e-2	0.99560
16	$\text{N}_2 + \text{O} \rightarrow \text{NO} + \text{N}$	0.1997e-2	0.99760
9	$\text{O}_2 + \text{O} \rightarrow 2\text{O} + \text{O}$	0.8116e-3	0.99841
10	$\text{O}_2 + \text{N} \rightarrow 2\text{O} + \text{N}$	0.6824e-3	0.99909
11	$\text{NO} + \text{NO} \rightarrow \text{N} + \text{O} + \text{NO}$	0.3827e-3	0.99948
1	$\text{N}_2 + \text{N}_2 \rightarrow 2\text{N} + \text{N}_2$	0.2328e-3	0.99971
6	$\text{O}_2 + \text{N}_2 \rightarrow 2\text{O} + \text{N}_2$	0.1435e-3	0.99985
7	$\text{O}_2 + \text{NO} \rightarrow 2\text{O} + \text{NO}$	0.6008e-4	0.99991
17	$\text{NO} + \text{O} \rightarrow \text{O}_2 + \text{N}$	0.3895e-4	0.99995
3	$\text{N}_2 + \text{O}_2 \rightarrow 2\text{N} + \text{O}_2$	0.1665e-4	0.99997
2	$\text{N}_2 + \text{NO} \rightarrow 2\text{N} + \text{NO}$	0.1420e-4	0.99998
15	$\text{NO} + \text{O}_2 \rightarrow \text{N} + \text{O} + \text{O}_2$	0.7047e-5	0.99999
8	$\text{O}_2 + \text{O}_2 \rightarrow 2\text{O} + \text{O}_2$	0.6463e-5	1.0000

Table 6: Sensitivity Indices for HEG cylinder chemistry, QoI $[O]_w$.

both the forward uncertainty propagation and the Bayesian inverse problem by fighting the curse of dimensionality.

4.1 Bayesian setting for freestream calibration

This section is focused on the problem of rebuilding freestream density and velocity $\mathbf{m} = \{\rho_\infty, u_\infty\}$ starting from experimental (*i.e.* noisy) measurements of pressure and heat flux at stagnation point of the cylinder $\mathbf{d} = \{p_{st}^*, q_{st}^*\}$. We also want to be able to account for uncertainties in the forward models, namely the uncertainties on the chemical reaction rate coefficients $\mathbf{c} = \{(A_r)_{r=1, \dots, n_r}\}$. Thus, we solve this calibration problem in the Bayesian sense [13], that is, upon choosing a prior probability density on the parameters and a likelihood function of the measurements and the parameters, we seek to characterize a posterior conditional probability density function of the parameters given measurements. The priors represent the degree of belief about possible uncertain parameters values before observing the data. The prior densities we use on each parameter are given in Tables 2 and 3. The joint prior $\sigma(\mathbf{x})$, with $\mathbf{x} = \{\mathbf{m}, \mathbf{c}\}$, is the product of these univariate priors, which reflects our prior belief that these parameters are independent. Measurements enter the formulation through the likelihood distribution $\ell(\mathbf{d}|\mathbf{x})$, which is proportional to the conditional probability of the observations given the parameters \mathbf{m} and \mathbf{c} . Measurements of p_{st} and q_{st} can be assumed independent and affected by independent additive errors: $\mathbf{d} = \mathbf{f}(\mathbf{m}, \mathbf{c}) + \boldsymbol{\eta}$. It is also common to assume errors as Gaussian noise with mean zero and standard deviation respectively σ_p and σ_q , representing measurement errors: $\boldsymbol{\eta} \sim \mathcal{N}(0, S)$, where $S = \text{diag}(\sigma_p^2, \sigma_q^2)$. In this case the likelihood distribution can be written as:

$$\ell(p_{st}^*, q_{st}^*|\mathbf{x}) \propto \exp\left(-\frac{(p_{st}^* - p_{st}(\mathbf{x}))^2}{\sigma_p^2} - \frac{(q_{st}^* - q_{st}(\mathbf{x}))^2}{\sigma_q^2}\right), \quad (1)$$

where $p_{\text{st}}(\mathbf{x})$ and $q_{\text{st}}(\mathbf{x})$ are computational model's prediction of stagnation pressure and heat flux given input parameters \mathbf{x} . There is a proportionality constant that makes the right hand side of (1) a probability density function (*i.e.*, such that it integrates to 1). However, this constant is not needed for the Markov Chain Monte Carlo (MCMC) method that will be used to characterize the posterior.

Bayes rule can be used to define the posterior probability density $\sigma^{\text{pos}}(\mathbf{x}|\mathbf{d})$ for the uncertain quantities given measurements of the stagnation data:

$$\sigma^{\text{pos}}(\mathbf{x}|\mathbf{d}) = \frac{\ell(\mathbf{d}|\mathbf{x})\sigma(\mathbf{x})}{\int \ell(\mathbf{d}|\mathbf{x})\sigma(\mathbf{x}) d\mathbf{x}}. \quad (2)$$

In order to compute means, moments and other statistical indicators, often the only possible way is by drawing samples from the posterior. The most common class of methods for characterizing the posterior are Markov Chain Monte Carlo methods [32, 33]. They are a broad class of methods used to draw samples from a probability density by constructing a Markov chain whose stationary distribution converges to the desired posterior. In this work, the Metropolis-Hastings algorithm with Gaussian proposal density is used to draw samples from the normalized posterior distribution. Samples from the chain are correlated samples from the posterior. Given such samples, we can compute posterior statistics of the parameters, such as moments, quantiles, or estimated marginal densities. It is well known that MCMC-based methods struggle to produce accurate estimates when the parameter space dimension is large. In effect, the Markov chain must explore the high dimensional space to reveal all regions of high posterior probability. Such exploration suffers from the *curse of dimensionality* [18], where the cost of thorough parameter space exploration scales exponentially with the dimension of the space.

4.2 Finding active subspaces for dimension reduction

To review active subspaces, we closely follow Chapter 3 in Constantine [20]. We use generic notation for a given function of m -variables, $f : \mathbb{R}^m \rightarrow \mathbb{R}$, and a joint probability density function, $\sigma : \mathbb{R}^m \rightarrow \mathbb{R}$, where $\mathbf{x} \in \mathbb{R}^m$ is a vector of input parameters and $f(\mathbf{x}) \in \mathbb{R}$ is the output of interest. Assume f is continuous and differentiable in the support of σ with continuous and square-integrable (with respect to σ) derivatives. The active subspaces of the pair (f, σ) are the eigenspaces of the symmetric, positive semidefinite matrix \mathbf{C} defined as

$$\mathbf{C} = \int \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T \sigma(\mathbf{x}) d\mathbf{x} = \mathbf{W} \mathbf{\Lambda} \mathbf{W}^T, \quad (3)$$

where \mathbf{W} is the orthogonal matrix of eigenvectors and $\mathbf{\Lambda}$ is diagonal matrix of non-negative eigenvalues arranged in descending order. The eigenpairs reveal structure in the map f , since (see [20, Lemma 3.1]),

$$\lambda_i = \int (\mathbf{w}_i^T \nabla f(\mathbf{x}))^2 \sigma(\mathbf{x}) d\mathbf{x}, \quad i = 1, \dots, m. \quad (4)$$

In words, the i th eigenvalue measures the average, squared directional derivative of f along the corresponding eigenvector. Thus, if $\lambda_i > \lambda_j$, then perturbing \mathbf{x} along \mathbf{w}_i changes f more, on average, than perturbing \mathbf{x} along \mathbf{w}_j . Moreover,

if \mathbf{C} is rank deficient, then its null space contains directions in the space of \mathbf{x} along which f is constant.

When simulation codes have gradient capabilities (*e.g.*, adjoint-based derivatives or algorithmic differentiation [34]), then we can estimate \mathbf{C} from (3) with numerical integration such as simple Monte Carlo; the eigenpairs of \mathbf{C} 's estimate provide estimates of the active subspaces [35]. However, many complex simulation codes do not have subroutines for gradients. In this case, we resort to modeling the gradients using function evaluations; for example, finite difference approximations of partial derivatives are the slope of a plane that interpolates two nearby function evaluations. The approach outlined in Algorithm 2, based on a least-squares-fit linear approximation of $f(\mathbf{x})$, has been unreasonably and surprisingly effective for uncovering one-dimensional active subspaces in a range of applications from integrated hydrological modeling [36] to multiphysics scramjet modeling [1] to satellite system modeling [37].

Algorithm 2 Linear model-based approach for estimating a one-dimensional active subspace; see [20, Algorithm 1.3].

1. Choose $N > m + 1$ points $\mathbf{x}_j \in \mathbb{R}^m$ in the support of ρ that are poised to fit a linear model.
2. For each \mathbf{x}_j , run the simulation model to compute $f_j = f(\mathbf{x}_j)$.
3. Find the coefficients \hat{a}_0 and $\hat{\mathbf{a}}$ of the least-squares-fit linear model such that

$$f_j \approx \hat{a}_0 + \hat{\mathbf{a}}^T \mathbf{x}_j, \quad j = 1, \dots, N. \quad (5)$$

4. Compute the vector $\hat{\mathbf{w}}$ that is the normalized gradient of the linear model,

$$\hat{\mathbf{w}} = \hat{\mathbf{a}} / \|\hat{\mathbf{a}}\|, \quad (6)$$

where $\|\cdot\|$ is the vector 2-norm.

There are no eigenvalues computed in Algorithm 2, but the vector it computes often defines a one-dimensional active subspace. The algorithm is motivated as follows. Suppose that $f(\mathbf{x})$ is nearly linear and its gradient is nearly constant, *i.e.*, for some constants $a_0 \in \mathbb{R}$ and $\mathbf{a} \in \mathbb{R}^m$,

$$f(\mathbf{x}) \approx a_0 + \mathbf{a}^T \mathbf{x}, \quad \nabla f(\mathbf{x}) \approx \mathbf{a}. \quad (7)$$

Then \mathbf{C} from (3) becomes

$$\mathbf{C} \approx \int \mathbf{a} \mathbf{a}^T \sigma(\mathbf{x}) d\mathbf{x} = \mathbf{a} \mathbf{a}^T = \mathbf{w} \lambda \mathbf{w}^T, \quad (8)$$

where $\mathbf{w} = \mathbf{a} / \|\mathbf{a}\|$ and $\lambda = \|\mathbf{a}\|^2$. This linear model-based approach is closely related to the *ordinary least-squares* method [38] for *sufficient dimension reduction* [39] in statistical regression. Although the algorithms across the two literatures are the same, the interpretation differs substantially between methods for data sets with random noise compared to methods for deterministic computer simulations.

To verify that the computed vector $\hat{\mathbf{w}}$ from Algorithm 2 has identified a one-dimensional active subspace, it is possible to make a *summary plot*, which is a

scatter plot of $\hat{\mathbf{w}}^T \mathbf{x}_j$ versus f_j . Such plots are common in regression graphics for identifying low-dimensional structures in regression data sets [39], and they are related to statistical techniques for *projection pursuit* [40]. If the plot reveals a strong univariate trend, then we can confidently approximate

$$f(\mathbf{x}) \approx g(\hat{\mathbf{w}}^T \mathbf{x}), \quad (9)$$

where $g : \mathbb{R} \rightarrow \mathbb{R}$ is fitted (*e.g.*, a least-squares-fit polynomial) with the pairs $(\hat{\mathbf{w}}^T \mathbf{x}_j, f_j)$. The plot enables the engineer to assess the quality of the fitted surface visually without the need to interpret complicated response surface quality metrics.

4.3 Exploiting active subspaces in MCMC for Bayesian inversion

In this section, we describe a way to exploit the low-dimensional structure in the parameter-to-observable map revealed by the active subspaces to enable efficient Bayesian inversion with MCMC. Constantine et al. [24] used the active subspaces of the log-likelihood function to reduce the dimension of the Markov chain, which led to smaller autocorrelation in the chain. In contrast, here the two one-dimensional active subspaces in the parameter-to-observable maps are exploited, as opposed to the log-likelihood (or misfit) function.

As will be shown in the results (see Sec. 5.1), stagnation pressure and heat flux each depend on one active variable. The two directions, $\hat{\mathbf{w}}_p$ and $\hat{\mathbf{w}}_q$, that define these two active variables are correlated (see Figures 9c and 9d) but not identical. Define the matrix \mathbf{W} as

$$\mathbf{W} = [\hat{\mathbf{w}}_p \quad \hat{\mathbf{w}}_q], \quad (10)$$

and let $\mathbf{V} \in \mathbb{R}^{8 \times 6}$ contain a basis for the null space of \mathbf{W}^T . For convenience, we orthogonalize the basis for the range of \mathbf{W} . Let

$$\mathbf{W} = \mathbf{U}\mathbf{R}, \quad \mathbf{U} \in \mathbb{R}^{8 \times 2}, \quad \mathbf{R} \in \mathbb{R}^{2 \times 2}, \quad (11)$$

be the tall QR factorization of \mathbf{W} where \mathbf{U} has orthogonal columns and \mathbf{R} is upper triangular. Any parameter point $\mathbf{x} \in \mathbb{R}^8$ can be decomposed into a projection on the range of \mathbf{W} and projection onto its orthogonal complement,

$$\mathbf{x} = \mathbf{U}\mathbf{U}^T \mathbf{x} + \mathbf{V}\mathbf{V}^T \mathbf{x} = \mathbf{U}\mathbf{y} + \mathbf{V}\mathbf{z}, \quad (12)$$

where $\mathbf{y} = \mathbf{U}^T \mathbf{x}$ and $\mathbf{z} = \mathbf{V}^T \mathbf{x}$. The summary plots in Figures 9a and 9b provide strong evidence that perturbing \mathbf{z} in (12) has little-to-no effect on either $p_{\text{st}}(\mathbf{x})$ or $q_{\text{st}}(\mathbf{x})$. Therefore, such perturbations should not affect the likelihood (1). In other words, we can ignore the variables \mathbf{z} for the Bayesian inversion, since we cannot calibrate them anyway. To exploit this structure, we derive a prior on the active variables as follows. For the prior probability density $\sigma(\mathbf{x})$ on the parameter space,

$$\begin{aligned} \sigma(\mathbf{x}) &= \sigma(\mathbf{U}\mathbf{y} + \mathbf{V}\mathbf{z}) \\ &= \pi(\mathbf{y}, \mathbf{z}) \\ &= \pi_{\mathbf{y}}(\mathbf{y}) \pi_{\mathbf{z}|\mathbf{y}}(\mathbf{z}|\mathbf{y}), \end{aligned} \quad (13)$$

where $\pi(\mathbf{y}, \mathbf{z})$ is a joint density on the active and inactive variables, $\pi_{\mathbf{y}}$ is the marginal density on the active variables, and $\pi_{\mathbf{z}|\mathbf{y}}$ is the conditional density on the inactive variables given the active variables. The change of variables is straightforward since the determinant of the Jacobian $\begin{bmatrix} \mathbf{U} & \mathbf{V} \end{bmatrix}$ of the linear transformation is 1, by orthogonality.

Let $g_p(\hat{\mathbf{w}}_p^T \mathbf{x})$ and $g_q(\hat{\mathbf{w}}_q^T \mathbf{x})$ be the response surfaces of the form (9) with vectors $\hat{\mathbf{w}}_p$ and $\hat{\mathbf{w}}_q$ computed from Algorithm 2. From the QR decomposition in (11), any point \mathbf{y} in the active variable space can be written

$$\mathbf{y} = \mathbf{U}^T \mathbf{x} = \mathbf{R}^{-1} \mathbf{W}^T \mathbf{x}, \quad (14)$$

which implies

$$\begin{bmatrix} \mathbf{r}_1 \mathbf{y} \\ \mathbf{r}_2 \mathbf{y} \end{bmatrix} = \mathbf{R} \mathbf{y} = \mathbf{W}^T \mathbf{x} = \begin{bmatrix} \hat{\mathbf{w}}_p^T \mathbf{x} \\ \hat{\mathbf{w}}_q^T \mathbf{x} \end{bmatrix}, \quad (15)$$

where \mathbf{r}_1 and \mathbf{r}_2 are the first and second row of \mathbf{R} , respectively. In other words, given a sample of \mathbf{y} , we get the arguments to the response surfaces by a linear transformation with \mathbf{R} . Then the likelihood in the active variables $\ell_{\mathbf{y}}(p_{st}^*, q_{st}^* | \mathbf{y})$, derived from (1), is

$$\ell_{\mathbf{y}}(p_{st}^*, q_{st}^* | \mathbf{y}) \propto \exp \left(-\frac{(p_{st}^* - g_p(\mathbf{r}_1 \mathbf{y}))^2}{\sigma_p^2} + \frac{-(q_{st}^* - g_q(\mathbf{r}_2 \mathbf{y}))^2}{\sigma_q^2} \right). \quad (16)$$

Knowing the marginal prior of the active variables $\pi_{\mathbf{y}}$ (which in some simpler cases can be computed analytically from the one of the physical parameters, but often only a numerical estimation $\hat{\pi}_{\mathbf{y}}$ is possible, for example by kernel density estimation [41]) and the likelihood $\ell_{\mathbf{y}}$, we can run MCMC to draw samples of the posterior on the active variables \mathbf{y} . Algorithm 3 shows a Metropolis-Hastings MCMC method from Kaipio and Sommersalo [14, Chapter 3.6.2] where the continuous state space for the Markov chain is the space of active variables \mathbf{y} .

Algorithm 3 A Metropolis-Hastings MCMC (see [14, Chapter 3.6.2]) on the active variables using the likelihood $\ell_{\mathbf{y}}$ from (16) and the kernel density estimate of the prior $\hat{\pi}_{\mathbf{y}}$.

Given an initial point $\mathbf{y}_0 \in \mathbb{R}^2$. For $k = 0, 1, 2, \dots$,

1. Draw $\mathbf{y}' \in \mathbb{R}^2$ from a symmetric proposal distribution centered at \mathbf{y}_k .
2. Compute the acceptance ratio,

$$\alpha(\mathbf{y}', \mathbf{y}_k) = \text{minimum} \left(1, \frac{\ell_{\mathbf{y}}(p_{st}^*, q_{st}^* | \mathbf{y}') \hat{\pi}_{\mathbf{y}}(\mathbf{y}')}{\ell_{\mathbf{y}}(p_{st}^*, q_{st}^* | \mathbf{y}_k) \hat{\pi}_{\mathbf{y}}(\mathbf{y}_k)} \right). \quad (17)$$

3. Draw t uniformly from $[0, 1]$. If $\alpha(\mathbf{y}', \mathbf{y}_k) > t$, set $\mathbf{y}_{k+1} = \mathbf{y}'$. Otherwise, set $\mathbf{y}_{k+1} = \mathbf{y}_k$.
-

The analysis of the posterior density on the active variables provides insights into the workings of the MCMC for Bayesian inversion—particularly since having only two active variables in this case permits helpful visualization. However, our goal is to estimate a posterior density on the model's original parameters

\mathbf{x} . To achieve this goal, we exploit a relationship similar to (13). Let $\sigma^{\text{pos}}(\mathbf{x})$ be the desired posterior density on the input parameters \mathbf{x} . Then,

$$\begin{aligned}\sigma^{\text{pos}}(\mathbf{x}|p_{\text{st}}^*, q_{\text{st}}^*) &= \sigma^{\text{pos}}(\mathbf{U}\mathbf{y} + \mathbf{V}\mathbf{z}|p_{\text{st}}^*, q_{\text{st}}^*) \\ &= \pi^{\text{pos}}(\mathbf{y}, \mathbf{z}|p_{\text{st}}^*, q_{\text{st}}^*) \\ &= \pi_{\mathbf{y}}^{\text{pos}}(\mathbf{y}|p_{\text{st}}^*, q_{\text{st}}^*) \pi_{\mathbf{z}|\mathbf{y}}^{\text{pos}}(\mathbf{z}|\mathbf{y}, p_{\text{st}}^*, q_{\text{st}}^*) \\ &= \pi_{\mathbf{y}}^{\text{pos}}(\mathbf{y}|p_{\text{st}}^*, q_{\text{st}}^*) \pi_{\mathbf{z}|\mathbf{y}}(\mathbf{z}|\mathbf{y}).\end{aligned}\tag{18}$$

The second line is a change of variables, where, similar to (13), the determinant of the linear transformation $[\mathbf{U} \ \mathbf{V}]$ is 1. The last line uses the fact that the inactive variables are independent of the data, so the conditional posterior of \mathbf{z} given \mathbf{y} (and the measurements $p_{\text{st}}^*, q_{\text{st}}^*$) is equal to the conditional prior on \mathbf{z} given \mathbf{y} . Therefore, given \mathbf{y} drawn according to the posterior on the active variables $\pi_{\mathbf{y}}^{\text{pos}}$ (*i.e.*, from the MCMC on the active variables), we can draw \mathbf{z} according to the conditional prior $\pi_{\mathbf{z}|\mathbf{y}}$; then a sample \mathbf{x} from the posterior σ^{pos} with the construction (12). Algorithm 4 shows a Metropolis-Hastings MCMC for drawing \mathbf{z} according to the conditional prior $\pi_{\mathbf{z}|\mathbf{y}}$.

Algorithm 4 A Metropolis-Hastings MCMC [14, Chapter 3.6.2] on the inactive variables \mathbf{z} given a value for the active variables \mathbf{y} .

Given \mathbf{y} and an initial point $\mathbf{z} \in \mathbb{R}^6$. For $k = 0, 1, 2, \dots$,

1. Draw $\mathbf{z}' \in \mathbb{R}^6$ from a symmetric proposal distribution centered at \mathbf{z}_k .
2. Compute the acceptance ratio,

$$\alpha(\mathbf{z}', \mathbf{z}_k) = \text{minimum} \left(1, \frac{\sigma(\mathbf{U}\mathbf{y} + \mathbf{V}\mathbf{z}')}{\sigma(\mathbf{U}\mathbf{y} + \mathbf{V}\mathbf{z}_k)} \right).\tag{19}$$

3. Draw t uniformly from $[0, 1]$. If $\alpha(\mathbf{z}', \mathbf{z}_k) > t$, set $\mathbf{z}_{k+1} = \mathbf{z}'$. Otherwise, set $\mathbf{z}_{k+1} = \mathbf{z}_k$.
-

5 Results

In this section, the results of the application of Active Subspaces and Bayesian inference on the HEG cylinder study case are presented. A Latin Hypercube sample of 576 points is sampled from the 8-dimensional input space characterized by the probability densities in Tables 3 and 2 and used as training set. The COSMIC code is then run for each sample to get the value of the outputs. Note that each computation includes three iterations of the adaptation strategy (which leads to a cost of about ten hours per simulation) Once the evaluations are performed, it is possible to proceed with the forward and backward UQ analysis, which are presented in the following sections.

5.1 Active Subspace for stagnation pressure and heat flux

As a first step, using the bounds in Table 3 and the standard deviations in Table 2, we shift and scale the parameter space to $[-1, 1]^2 \times \mathbb{R}^6$. This normalization

is convenient for studying and exploiting the active subspaces. The thrust of active subspaces is to rotate high-dimensional surfaces until they reveal low-dimensional structure manifesting as directions along which the surface is flat, globally. Centering the parameter space at the origin ensures that all rotations occur about the origin. The prior densities on the normalized space are $1/4$ (the uniform density on the box $[-1, 1]^2$) times a multivariate standard Gaussian on \mathbb{R}^6 .

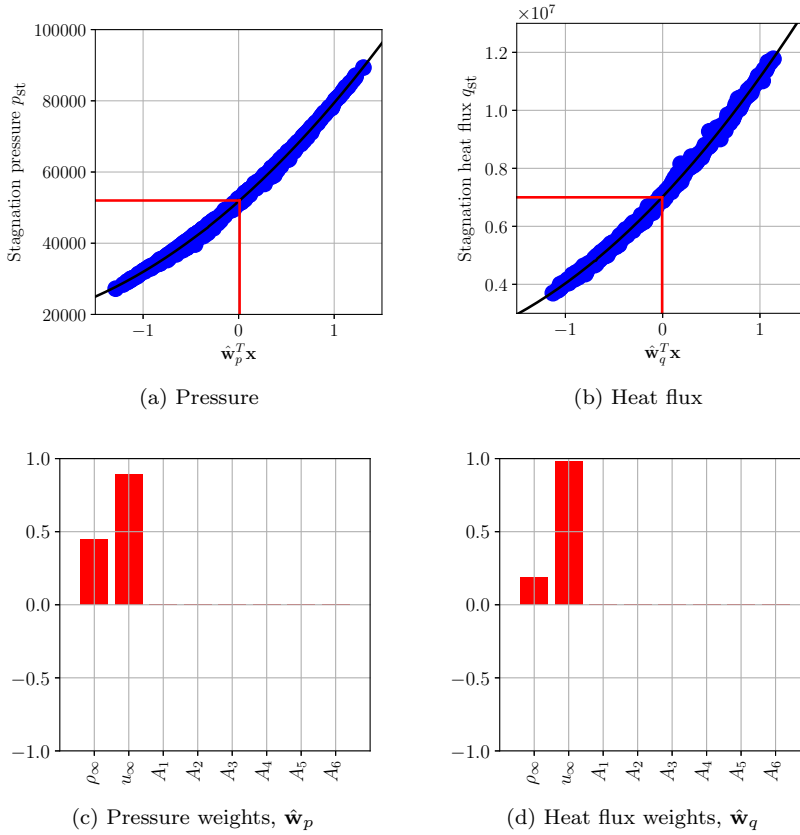


Figure 9: One-dimensional summary plots (top row) for pressure (left) and heat flux (right) using the vectors $\hat{\mathbf{w}}_p$ and $\hat{\mathbf{w}}_q$, respectively, from Algorithm 2, along with bar plots (bottom row) of the components of $\hat{\mathbf{w}}_p$ and $\hat{\mathbf{w}}_q$. The black lines show the one-dimensional response surfaces. The red horizontal lines are at the given measurements, p_{st}^* and q_{st}^* . The vertical red lines show the value of the respective active variables, $\hat{\mathbf{w}}_p^T \mathbf{x}$ and $\hat{\mathbf{w}}_q^T \mathbf{x}$, that maps to the measurements.

Figure 9 shows the summary plots and weight vector components for stagnation pressure p_{st} and heat flux q_{st} . The blue dots represent 576 independent runs of COSMIC code with inputs drawn from the probability densities in Tables 3 and 2. We use these runs within Algorithm 2 to compute the associated vectors $\hat{\mathbf{w}}_p$ and $\hat{\mathbf{w}}_q$. Figure 9a is the summary plot for pressure, and Figure 9b is the summary plot for heat flux. The strong univariate trend in the blue dots for each output suggests that orthogonal to the active direction, the sur-

face is (relatively) constant. The black lines show the one-dimensional response surfaces (g from (9)) fitted to the data pairs; each surface is a simple univariate quadratic functions with three coefficients. The data varies roughly 1 or 2% from the fitted surface, but this variation is almost indistinguishable from numerical errors in the complex flow simulation, especially for heat flux values. Figures 9c and 9d shows bar plots of the components of $\hat{\mathbf{w}}_p$ and $\hat{\mathbf{w}}_q$, respectively. The summary plots in Figures 9a and 9b reveal low-dimensional structure in the map from input parameters to output observables. The vector components in Figures 9c and 9d give insight into the important input parameters. The relative insensitivity of pressure and heat flux to the reaction rates suggests that it is not possible to reliably calibrate the reaction rates using stagnation pressure and heat flux measurements, but at the same time that uncertainties on the chemical model will not influence strongly rebuilt freestream values.

5.1.1 Comparison with classical global Sensitivity Analysis

In this section, we compare results in Figures 9c and 9d to the variance-based global sensitivity analysis performed on the HEG cylinder study case using the adaptive sparse polynomial dimensional decomposition (PDD) by Tang et al. [31]. When the summary plots (*i.e.*, Figures 9a and 9b) show such strong evidence of univariate structure, the components of the vectors $\hat{\mathbf{w}}_p$ and $\hat{\mathbf{w}}_q$ can be used as global sensitivity metrics. Since these weights are derived from the coefficients of the least-squares-fit linear model, they are similar to the *standard regression coefficients* [42, Chapter 1.2.5] with different normalization—though their interpretation is significantly different. For these two outputs, the weights suggest that neither pressure nor heat flux is sensitive to changes in the reaction rate coefficients—compared to their sensitivity with respect to free stream conditions. For more general connections between active subspaces and global sensitivity analysis, see Constantine and Diaz [43].

Here, sparse PDD [31] is used on the 576 points Experimental Design to perform a global Sensitivity Analysis of the two functions of interests. The method easily provides the first order Sobol’ Sensitivity Indices (SI), which are used to hierarchically rank the input variables according to their contribution to the variance of the output. Computations are performed considering a maximum ANOVA order of interaction $\nu = 3$ and a maximum polynomial order $m = 4$. The obtained Sobol’ indices are given in Table 7 for the pressure and 8 for the heat flux.

Variable	SI 1st order	% Variance	Total SI
u_∞	0.7970115	0.7970115	0.802208
ρ_∞	0.2027181	0.9997297	0.208394
A_5	0.8159018E-04	0.9998113	0.520080E-06
A_6	0.7354608E-04	0.9998848	0.510802E-07
A_2	0.4810281E-04	0.9999329	0.261344E-06
A_3	0.4020067E-04	0.9999731	0.170147E-05
A_4	0.2178639E-04	0.9999949	0.175886E-06
A_1	0.5055511E-05	1.000000	0.376313E-06

Table 7: Sorted first order and total Sensitivity Indices for HEG cylinder simulation, QoI p_{st} .

Variable	SI 1st order	% Variance	Total SI
u_∞	0.9644307	0.9644307	0.965089
ρ_∞	0.3547857E-01	0.9999093	0.375331E-01
A_6	0.3577629E-04	0.9999450	0.465707E-05
A_5	0.2533707E-04	0.9999704	0.474139E-05
A_1	0.1540559E-04	0.9999858	0.104035E-04
A_2	0.5799109E-05	0.9999916	0.353491E-05
A_3	0.5159200E-05	0.9999967	0.260056E-04
A_4	0.3205809E-05	0.9999999	0.626155E-05

Table 8: Sorted first order and total Sensitivity Indices for HEG cylinder simulation, QoI q_{st} .

What can be noticed from this results is that the 20% of the uncertainty on the velocity value has the greatest impact on both outputs. Especially for the heat flux, where the theory suggests a dependence with the cube of the velocity, this variable is responsible of more than the 96% of the total output variance. Uncertainties on the chemistry play a very small role in the variance of the output for both quantities, and values of their sensitivity indices are mostly related to numerical noise. This was expected for the stagnation pressure, since it is known to be mainly dependent on the free stream density and velocity. For the heat flux, this can be explained firstly by the fact that the 20% uniform uncertainty on the velocity causes big variations of the heat flux. Furthermore, in the considered configuration the Mach number is relatively small, and only the molecular oxygen dissociates. In addition, the imposed value of $\gamma = 1$ for the catalicity is forcing the chemical equilibrium at the wall. Therefore, the fact that small sensitivity indices are associated to the chemistry should not surprise. Also Total SIs are reported for both QoIs in Tables 7 and 8. The fact that their values are quite similar to the ones of the first order indices means that the interaction between variables is not very strong. From Figure 10, it is

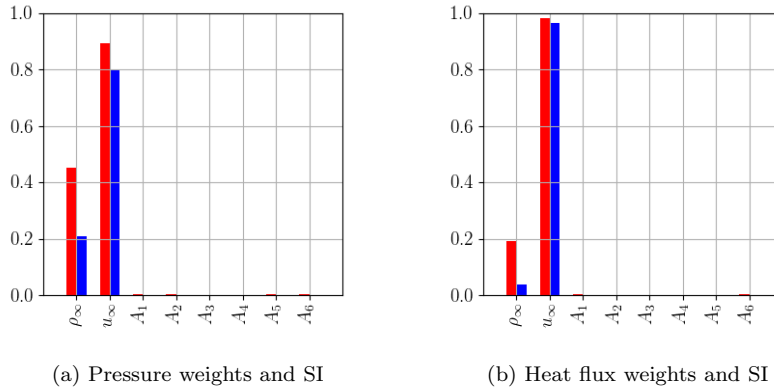


Figure 10: Bar plots of the components of $\hat{\mathbf{w}}_p$ and $\hat{\mathbf{w}}_q$ (red) along with the global sensitivity indices (blue).

also possible to notice that global sensitivity indices show a hierarchy of input

variables that is consistent with the one of vectors $\hat{\mathbf{w}}_p$ and $\hat{\mathbf{w}}_q$.

5.2 Forward propagation of uncertainties

Exploiting the very reduced cost of evaluating the surrogate model, a simple Monte Carlo (MC) sampling technique can be applied to the two 1D polynomial response surfaces in the active variables in order to perform a forward uncertainty propagation and compute the statistics of the quantities of interest. In order to get a set of sampling points of the active variables, 100000 Monte Carlo samples are drawn from the priors of the physical variables and then projected on the active subspace by means of the linear transformation $\mathbf{y} = \hat{\mathbf{w}}^T \mathbf{x}$. Their kernel density estimate is reported in Figure 11, top row. The set of sampling points in the original space is also exploited to perform a propagation on a quadratic polynomial regression surrogate model of the QoIs in the original 8-dimensional space, that is used as a comparison. Mean and variance of p_{st} and q_{st} are reported in Table 9. Differences between results of active subspace and physical variables are very small, being inferior to 1% for both means and variances. This means that the reduced order approximation of the functions of interest introduced by the active subspace (see Equation (9)) is able to conserve all the important information in the input-to-output map, as expected. Coefficients of variation are respectively about 26% for the pressure and 28% for the heat flux. It can be noticed that the forward propagation of prior probability densities produces coefficients of variation of the outputs that are bigger than the percentage measurement errors. This allows the posteriors to be mainly informed by the measurements and not by the priors, which in this work are just uninformative intervals chosen by the authors. The probability

		AS var.	Physical var.
p_{st} [Pa]	μ	53033	53050
	σ	13782	13760
q_{st} [W/m ²]	μ	7.2029 e6	7.2054 e6
	σ	2.0586 e6	2.0596 e6

Table 9: Mean value and standard deviation of the quantities of interest computed with the response surfaces built on the pressure and heat flux active variables. Values are computed with 100000 Latin Hypercubes samples and compared to results obtained with a propagation on a surrogate in the 8-dimensional physical input.

density functions of the output quantities, estimated by means of kernel density estimations on the Monte Carlo samples, are reported in Figure 11, bottom row. Comparison between results obtained on the active subspace metamodel (blue) and the ones of the physical variables metamodel (green) confirm a very good agreement. The figure shows the propagation of the PDF of the two active variables for p_{st} and q_{st} through the one-dimensional surrogate models and the obtained PDF. The fact of having reduced the input dimensionality of the problem makes the visualization of the uncertainty propagation process very easy.

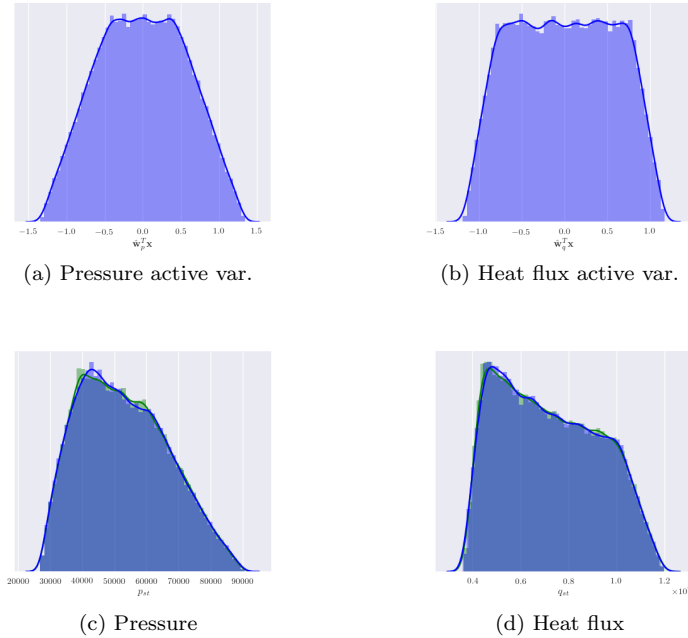


Figure 11: Uncertainty propagation process from the probability densities of the first active variables (top row) to the ones of the stagnation pressure and heat flux (bottom row) through the one-dimensional response surfaces built on the first active variable. Comparison between results obtained on the active subspace metamodel (blue) and the ones of the physical variables metamodel (green) confirm a very good agreement.

5.3 Freestream calibration by means of Bayesian inference

In this section, we illustrate the results for the Bayesian calibration of free stream parameters. Results are presented by following the sampling procedure based on MCMC in the active subspace presented in Section 4.3. The first step that must be done to draw samples from the posterior of the active variables with Algorithm 3 is to define the prior $\pi_{\mathbf{y}}$ of the active variables starting from the one of the physical parameters $\sigma(\mathbf{x})$. If $\sigma(\mathbf{x})$ was Gaussian, then $\pi_{\mathbf{y}}$ would also have been a Gaussian density. However, prior components are not all Gaussian (see Table 3), hence $\pi_{\mathbf{y}}$ must be estimated numerically. Since it is only a bivariate density and the forward map $\mathbf{y} = \mathbf{U}^T \mathbf{x}$ is simple (with no dependence on the complex physics model), we can estimate $\pi_{\mathbf{y}}$ to sufficient accuracy with a bivariate kernel density estimate. Figure 12b shows the contours of a bivariate kernel density estimate using the Python SciKit Learn's `KernelDensity` function [44] with parameters `bandwidth=0.15` and `kernel='tophat'` constructed from 50000 samples of $\mathbf{y} = \mathbf{U}^T \mathbf{x}$, where \mathbf{x} 's are drawn according to the normalized prior. Figure 12a shows a subset of 200 samples for reference. The small scale contour fluctuations near the center of the domain are due to finite sampling. Essentially, the marginal on the two active variables $\mathbf{y} = [y_1, y_2]^T$ is

a rotated uniform density. This is consistent with the weights from Figures 9c and 9d, since the largest weights are associated with the two parameters (ρ_∞ and u_∞) that have uniform priors. Denote the kernel density estimate of $\pi_{\mathbf{y}}$ by $\hat{\pi}_{\mathbf{y}}$. We use the response surfaces shown in Figures 9a and 9b to build a

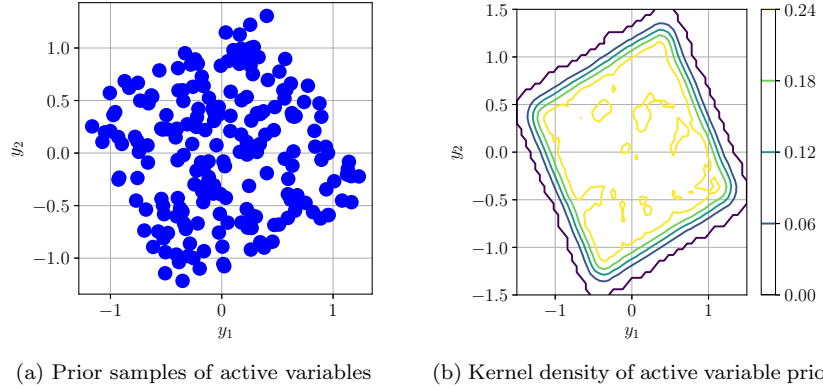


Figure 12: The left figure shows 200 samples of the active variables $\mathbf{y} = \mathbf{U}^T \mathbf{x}$ with \mathbf{x} drawn from its normalized prior. The right figure shows the contours of the bivariate kernel density estimate for the active variables constructed from 50000 samples; it is essentially a rotate bivariate uniform density. The small scale fluctuations are a result of the finite sampling.

likelihood function on the active variables. The given artificial measurements are stagnation pressure $p_{\text{st}}^* = 5.2 \cdot 10^4$ Pa and stagnation heat flux $q_{\text{st}}^* = 7.0 \cdot 10^6$ W/m².

In our numerical implementation of Algorithm 3, we use a Gaussian proposal density centered at the current state (\mathbf{y}_k , see Step 1) with covariance matrix $\text{diag}(0.23^2, 1.02^2)$, where the two variance values are issue from an adaptation of the exploration step. We run one chain of 250000 steps, and, following the suggestion of Brooks et al. [33, Chapter 6.5], we discard the first part of the chain's samples as *burn in*, and we take in the final sample one sample each 10 samples, which leaves us with 20000 samples. This is sufficient to cover the two-dimensional space of active variables. Figure 13a shows 200 samples from the MCMC. The samples' apparent alignment with the axes y_1 and y_2 is consistent with the rotation induced by \mathbf{R} from (11) to create orthogonal coordinates (*i.e.*, the columns of \mathbf{U} from (11) are orthonormal). The difference in the samples' spread along the axes is consistent with the different noise levels for each output. In effect, the relatively large noise (10%) in the heat flux measurement q_{st}^* permits a larger range of y_2 values that are consistent with the noisy measurements. The relatively low noise (2%) in the pressure measurement p_{st}^* leads to less spread in the y_1 samples. Figure 13b shows contours (purple) of a kernel density estimate of the posterior on \mathbf{y} from the MCMC samples; the estimate uses a Gaussian kernel with bandwidth 0.15 implemented in SciKit Learn's `KernelDensity` [44]. The axis alignment is consistent with the discussion of the samples in Figure 13a. The apparent slight asymmetry about the y_2 axis results from the mild nonlinearity in the response surfaces. The same asymmetry appears in the y_1 axis, though it is not as apparent since the support

of the contours is relatively narrow. The posterior contours are superimposed on top of the contours from the density estimate of the prior $\hat{\pi}_{\mathbf{y}}$. The support of the posterior is smaller than the prior, which implies that the measurements (pressure p_{st}^* and heat flux q_{st}^*) have provided additional information about the parameter uncertainty in the active variables.

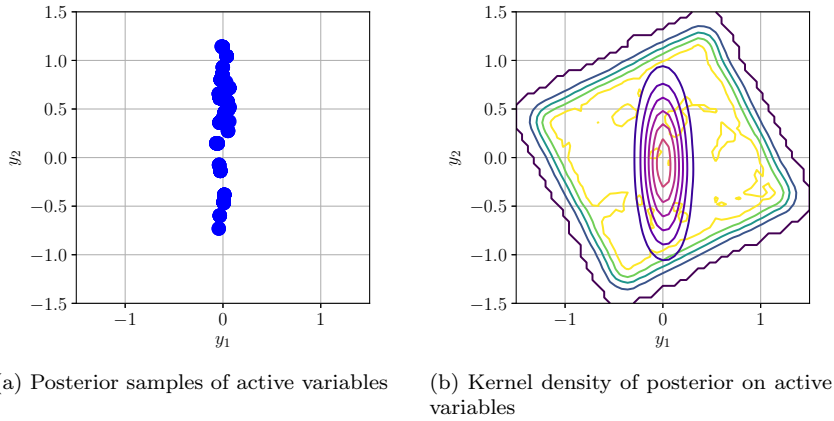


Figure 13: The left figure shows 200 samples from the response surface-enabled MCMC on the active variables. The spread along coordinate axes is consistent with the different noise assumptions in the measurements. The right figure shows contours of a bivariate kernel density estimate of the MCMC samples compared to the prior contours.

The final goal of the work is to have samples from the posterior density of the physical parameters. It is possible to restore to this set variables starting from the posteriors of the active variables by means of Eq. (18). If the joint prior on (\mathbf{y}, \mathbf{z}) was Gaussian, then the conditional $\pi_{\mathbf{z}|\mathbf{y}}$ would also have been Gaussian. Instead, in our case, the joint prior on (\mathbf{y}, \mathbf{z}) is not Gaussian (see section 3); hence the Metropolis-Hastings MCMC provides a general way to sample from the conditional pdf. In our implementation of Algorithm 4, we use a Gaussian proposal in Step 1 with covariance matrix $(0.8)^2 \mathbf{I}$, where \mathbf{I} is the 6×6 identity matrix. In our numerical experiments, we draw 100 samples of \mathbf{z} per MCMC sample of \mathbf{y} from Algorithm 3. From 20000 samples of \mathbf{y} , we get 2000000 samples of \mathbf{x} with (12). Note that since \mathbf{z} is independent of the measurements, the algorithm for drawing \mathbf{z} given \mathbf{y} does not call the simulation model or its response surface approximation. Figure 14 compares the univariate marginal posterior densities (blue lines) on the model parameters to the associated marginal prior densities (red lines); see Tables 3 and 2. Notice that the only two parameters that are informed by the stagnation point measurements, p_{st}^* and q_{st}^* , (*i.e.*, their posteriors differ substantially from the priors) are the freestream conditions ρ_∞ and u_∞ . Since the measurements cannot inform the reaction rates, the reaction rate posteriors are the same as the priors. A further comparison is proposed with the posteriors (green lines) obtained with a single-site updating Metropolis-Hastings MCMC sampling directly in the 8-dimensional physical parameters space, considering 8D quadratic regressions as metamodels for the stagnation pressure and heat flux. The chain ran for

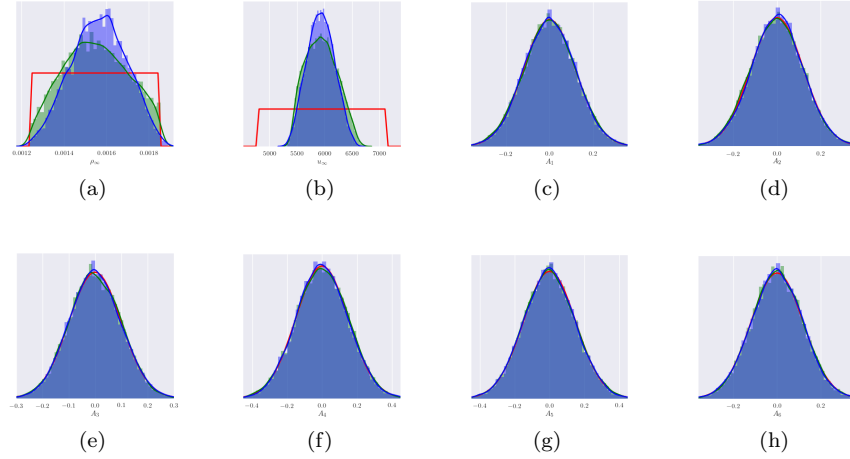


Figure 14: Posterior marginal densities (blue lines) on the eight model parameters compared to their marginal priors (red lines); see Tables 3 and 2. A comparison is shown also with the posteriors obtained by MCMC sampling directly in the physical space (green lines). The reaction rate parameters are not informed by the data, so their posterior is essentially the prior.

500000 steps, with adapted explorations steps for the Gaussian proposals in every dimension. It can be noticed that the posteriors present a good qualitative agreement. Figure 15 shows the most interesting bivariate marginal posterior,

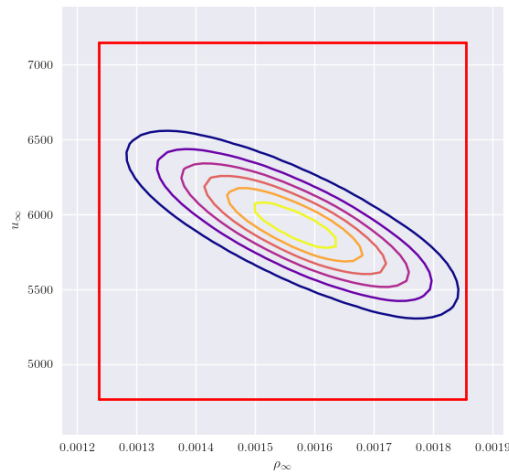


Figure 15: Bivariate marginal posterior contours for free stream parameters ρ_∞ and u_∞ compared to the joint marginal prior; see Table 3. Compare these contours to the posterior on the active variables in Figure 13b.

which is on the freestream parameters informed by the measurements. The contours are the posterior and the red box is the joint prior. The freestream parameters are highly correlated in their posterior, so the univariate marginals alone can be misleading. Comparing the contours in Figure 15 to the posterior on the active variables in Figure 13b, the shapes are essentially the same but rotated; the correlation in the freestream parameters is related to the rotation induced by the matrix \mathbf{R} from (11).

Posterior means and standard deviations of the free stream are then computed exploiting the MCMC samples of the posteriors and are reported in Table 10. A comparison is also performed with the posterior values sampled in the physical space. Values show a good agreement, with differences in the mean values smaller than 1% for both the density and the velocity.

Looking at the potential interest in using heat flux data for reconstruction, note that the coefficient of variation for freestream density and velocity are 8.3% and 4.1%, respectively. This shows how heat flux data could provide additional insights towards the reconstruction of freestream conditions. However, the obtained posterior uncertainties are quite big if compared to other freestream reconstructions in literature (for example [8]). This because the accuracy on velocity is much higher when rebuilt using IMU data, allowing also for a better estimation of the freestream density. Of course an improvement of results would be obtained assuming smaller noise associated to heat flux measurements. Another possibility would be to include IMU results in the Bayesian setting, to get a smaller uncertainty on the velocity, and using heat flux measurements to calibrate other uncertain freestream or model parameters. These aspects should clearly be tested in the future, also in the context of a real space mission.

		AS var.	Physical var.
ρ_∞ [kg/m ³]	μ	1.5666e-03	1.5545e-03
	σ	1.2954e-04	1.4806e-04
u_∞ [m/s]	μ	5930.18	5939.10
	σ	246.50	279.13

Table 10: Posterior mean value and standard deviation of the inferred free stream quantities computed with the MCMC samples. Comparison between values obtained for the regular chain directly run on the physical variables and the chain in the active subspace.

6 Conclusions

In this paper, we have illustrated how to perform a Bayesian calibration of the free stream parameters of a hypersonic high-enthalpy flow around a cylinder, exploiting active subspaces for the reduction of the dimensionality of the input space. The configuration taken into account was the HEG I configuration, known in literature as a validation test-case for hypersonic CFD. The goal of the Bayesian inversion was to show the feasibility in using measurements of pressure and heat flux at the stagnation point for rebuilding freestream velocity and density.

First, we have proposed an algorithm allowing an automatic mesh convergence assessment, which is required in the context of hypersonic flows with uncertainty. Then, the validation of the numerical solver used to simulate the hypersonic flow has been performed against experimental data.

Concerning the uncertainty characterization for chemistry model, a comparison between five and eleven species chemistry model allowed to assess that the simpler model was accurate enough for the case under study. Global Sensitivity indices were used to chose which uncertain reaction rates to keep in the final uncertainty quantification.

We described an alternative way to exploit the low-dimensional structure in the parameter-to-observable map revealed by the active subspaces to enable efficient Bayesian inversion with MCMC. Instead of using the active subspaces of the log-likelihood (or misfit) function to reduce the dimension of the Markov chain as in [24], here the two one-dimensional active subspaces in the parameter-to-observable maps are exploited. As all the active-subspace-based approaches, it enables efficient MCMC in the active variables for Bayesian inversion, which is normally an easier task than performing a MCMC sampling directly in the higher-dimensional physical space. This enables to use more efficient and fast implementations of the MCMC algorithm that require a smaller number of sampling points to explore the lower-dimensional active subspace. This aspect has not been deeply explored in the paper, but it could be relevant to practical applications. One advantage with respect to the misfit-based method is that the active subspaces and metamodels for the functions of interest, which in this case correspond to stagnation pressure and heat flux, do not depend anymore on the measurements but just on the prior densities. This makes the approach more versatile as the surrogates can be used also for forward uncertainty propagation, and reused if changing measurements values. Furthermore, discovering and exploiting the active subspaces in the simulation model's parameter-to-observable map provides global sensitivity metrics, which give comparable rankings of the input as global variance-based sensitivity indices.

The proposed approach has one drawback that arises in cases where a non-Gaussian prior is present in the physical space. In order to have a sample of the posterior density in physical variables, one has to draw samples from the conditioned probability of the inactive variables given the active variables, which is not analytical when the joint prior on (\mathbf{y}, \mathbf{z}) is not Gaussian. Hence, one must run a second MCMC to sample from the conditional pdf, which adds numerical effort and complication. However, it must be noticed that the algorithm for drawing \mathbf{z} given \mathbf{y} does not need to call the simulation model or its response surface approximation.

To summarize the Bayesian inversion results, the noisy measurements of stagnation pressure and heat flux can inform a correlated posterior probability density on the freestream density and velocity, but they cannot inform the reaction rates. On the other hand, this also means that a prior uncertainty on chemical parameters does not affect strongly the rebuilt free stream parameters. Moreover, insight rich visualizations are possible because of the dimension reduction. Of course, the visualizations would not have been possible if the number of active variables was greater than two; *e.g.*, if either of the observables depended on more than one linear combination of the eight model parameters. However, the methodology that exploits the active subspace for efficient MCMC extends without change. The obtained posterior on the freestream density presented a

large support when compared to the one of its prior. This can be explained by the fact that a quite large measurement error has been considered for the heat flux, that represent current heat sensor capabilities. An improvement that is not tested in this work would be to reduce the heat flux measurement noise deviation until an acceptable 1-sigma interval is obtained on the posterior. Globally, the proposed framework shows the potential interest in using heat flux data for rebuilding freestream conditions. Future works will be oriented to an accurate comparison of the heat flux based reconstruction with respect to other existing techniques in the context of a real space mission.

Acknowledgments

Andrea F. Cortesi has been supported by DGA (Direction Générale de l'Armement) from French Government. This material is based on work supported by Department of Defense, Defense Advanced Research Project Agency's program Enabling Quantification of Uncertainty in Physical Systems. The second author's work is partially supported by (i) DARPA's Enabling Quantification of Uncertainty in Physical Systems (EQUiPS) program and (ii) U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under Award Number DE-SC- 0011077. Authors wish to thank Georgios Bellas-Chatzigeorgis and Paolo Barbante for their support with regard to numerical simulations with the COSMIC code.

References

- [1] 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, 2015.
- [2] Brian Lockwood and Dimitri Mavriplis. Gradient-based methods for uncertainty quantification in hypersonic flows. *Computers & Fluids*, (85):27 – 38, 2013.
- [3] François Sanson, Nadège Villedieu, Francesco Panerai, Olivier Chazot, Pietro M. Congedo, and Thierry E. Magin. Quantification of uncertainty on the catalytic property of reusable thermal protection materials from high enthalpy experiments. *Experimental Thermal and Fluid Science*, 82:414 – 423, 2017.
- [4] Thomas K. West and Serhat Hosder. Uncertainty quantification of hypersonic reentry flows with sparse sampling and stochastic expansions. *Journal of Spacecraft and Rockets*, 52:120 – 133, 2015.
- [5] James S. Strand and David B. Goldstein. Global sensitivity analysis for dsmc simulations of hypersonic shocks. *Journal of Computational Physics*, 246:184 – 206, 2013.
- [6] P. zur Nieden and H. Olivier. Determination of atmospheric densities from reentry data flight data. *Journal of Spacecraft and Rockets*, 44(2):332 – 337, 2007.
- [7] S. Dutta and R. D. Braun. Statistical entry, descent, and landing performance reconstruction of the mars science laboratory. *Journal of Spacecraft and Rockets*, 51(4):1048 – 1061, 2014.
- [8] B. Van Hove and Ö. Karatekin. Atmospheric reconstruction with stagnation pressure flight data from mars science laboratory. *Journal of Spacecraft and Rockets*, 54(3):609 – 620, 2017.
- [9] J. Tryoen, P. M. Congedo, R. Abgrall, N. Villedieu, and T.E. Magin. Bayesian-based method with metamodels for rebuilding freestream conditions in atmospheric entry flows. *AIAA Journal*, 52(10):2190 – 2197, 2014.
- [10] Doyle Knight, José Longo, Dimitris Drikakis, Datta Gaitonde, Andrea Lani, Ioannis Nompelis, Bodo Reimann, and Louis Walpot. Assessment of CFD capability for prediction of hypersonic shock interactions. *Progress in Aerospace Sciences*, 48 - 49:8 – 26, 2012.
- [11] Sebastian Karl, Jan Martinez Schramm, and Klaus Hannemann. High enthalpy cylinder flow in HEG: A basis for CFD validation. In *33rd AIAA Fluid Dynamics Conference and Exhibit, Fluid Dynamics and Co-located Conferences*. American Institute of Aeronautics and Astronautics, 2003.
- [12] C. Park, R. Jaffe, and H. Partridge. Chemical-kinetic parameters of hyperbolic earth entry. *Journal of Thermophysics and Heat Transfer*, 15(1):76 – 90, 2001.

-
- [13] Daniela Calvetti and Erkki Somersalo. *Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*. Springer, New York, 2007.
- [14] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160 of *Applied Mathematical Sciences*. Springer, 2005.
- [15] P. G. Constantine, A. Doostan, Q. Wang, and G. Iaccarino. A surrogate accelerated bayesian inverse analysis of the HyShot II supersonic combustion data. In *Proceedings of the 49th Aerospace Meeting*, 2011.
- [16] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical sciences*, 4(4):409 – 435, 1989.
- [17] T.W. Simpson, J.D. Poplinski, P. N. Koch, and J.K. Allen. Metamodels for computer-based engineering design: Survey and recommendations. *Engineering with Computers*, 17(2):129–150, 2001.
- [18] D. L. Donoho. High-dimensional data analysis: The curses and blessings of dimensionality. In *AMS Conference on Math Challenges of the 21st Century*, 2000.
- [19] P. G. 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, 2014.
- [20] P. G. Constantine. *Active Subspaces: Emerging Ideas in Dimension Reduction for Parameter Studies*. SIAM, Philadelphia, 2015.
- [21] T Cui, J Martin, Y M Marzouk, A Solonen, and A Spantini. Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems*, 30(11):114015, 2014.
- [22] Tiangang Cui, Kody J.H. Law, and Youssef M. Marzouk. Dimension-independent likelihood-informed MCMC. *Journal of Computational Physics*, 304:109–137, 2016.
- [23] Tiangang Cui, Youssef Marzouk, and Karen Willcox. Scalable posterior approximations for large-scale Bayesian inverse problems via likelihood-informed parameter and state reduction. *Journal of Computational Physics*, 315:363–387, 2016.
- [24] Paul G. Constantine, Carson Kent, and Tan Bui-Thanh. Accelerating Markov chain Monte Carlo with active subspaces. *SIAM Journal on Scientific Computing*, 38(5):A2779–A2805, 2016.
- [25] P. Barbante. *Accurate and efficient modelling of high temperature non-equilibrium air flows*. PhD thesis, Von Karman Institute, 2001.
- [26] B. Van Leer. Towards the ultimate conservative difference scheme. V. A second-order sequel to godunov’s method. *Journal of Computational Physics*, 32(1):101–136, 1979.

-
- [27] S. Osher and F. Solomon. Upwind Difference Schemes for Hyperbolic Systems of Conservation Laws. *Mathematics of Computation*, 38(158):339–374, 1982.
- [28] D. Bose, M. Wright, and T. Gokçen. Uncertainty and Sensitivity Analysis of Thermochemical Modeling for Titan Atmospheric Entry. *37th AIAA Thermophysics Conference, Portland, Oregon, 2004*.
- [29] T. E. Magin, L. Caillault, A. Bourdon, and C. O. Laux. Nonequilibrium radiative heat flux modeling for the Huygens entry probe. *Journal of Geophysical Research: Planets*, 111(E7), 2006. E07S12.
- [30] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239 – 245, 1979.
- [31] K. Tang, P. M. Congedo, and R. Abgrall. Adaptive surrogate modeling by ANOVA and sparse polynomial dimensional decomposition for global sensitivity analysis in fluid simulation. *Journal of Computational Physics*, 314:557–589, 2016.
- [32] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in practice*. Chapman & Hall/CRC Interdisciplinary Statistics, 1996.
- [33] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. Chapman & Hall / CRC, Boca Raton, 2011.
- [34] Andreas Griewank and Andrea Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, Philadelphia, 2008.
- [35] Paul Constantine and David Gleich. Computing active subspaces with Monte Carlo. *arXiv preprint arXiv:1408.0545v2*, 2015.
- [36] J. L. Jefferson, J. M. Gilbert, P. G. Constantine, and R. M. Maxwell. Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. *Computers & Geosciences*, 83:127–138, 2015.
- [37] Xingzhi Hu, Geoffrey T. Parks, Xiaoqian Chen, and Pranay Seshadri. Discovering a one-dimensional active subspace to quantify multidisciplinary uncertainty in satellite system design. *Advances in Space Research*, 57(5):1268–1279, 2016.
- [38] Ker-Chau Li and Naihua Duan. Regression analysis under link violation. *The Annals of Statistics*, 17(3):pp. 1009–1052, 1989.
- [39] R. Dennis Cook. *Regression Graphics: Ideas for Studying Regression through Graphics*. John Wiley & Sons, New York, 1998.
- [40] Peter J. Huber. Projection pursuit. *The Annals of Statistics*, 13(2):435–475, 1985.
- [41] Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

-
- [42] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, Inc., Hoboken, 2008.
- [43] Paul G Constantine and Paul Diaz. Global sensitivity metrics from active subspaces. *Reliability Engineering & System Safety*, 162:1–13, 2017.
- [44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.



**RESEARCH CENTRE
BORDEAUX – SUD-OUEST**

351, Cours de la Libération
Bâtiment A 29
33405 Talence Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399