



### **Science Arts & Métiers (SAM)**

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>  
Handle ID: <http://hdl.handle.net/10985/8640>

#### **To cite this version :**

Xavier MERLE, Paola CINNELLA - Bayesian quantification of thermodynamic uncertainties in dense gas flows - Reliability Engineering and System Safety p.in press - 2014

Any correspondence concerning this service should be sent to the repository

Administrator : [scienceouverte@ensam.eu](mailto:scienceouverte@ensam.eu)



# Bayesian quantification of thermodynamic uncertainties in dense gas flows

X. Merle<sup>a,\*</sup>, P. Cinnella<sup>a</sup>

<sup>a</sup>*DynFluid Laboratory, Arts et Métiers ParisTech  
151 boulevard de l'Hôpital, 75013 Paris, FRANCE*

---

## Abstract

A Bayesian inference methodology is developed for calibrating complex equations of state used in numerical fluid flow solvers. Precisely, the input parameters of three equations of state commonly used for modeling the thermodynamic behavior of so-called dense gas flows, – i.e. flows of gases characterized by high molecular weights and complex molecules, working in thermodynamic conditions close to the liquid-vapor saturation curve –, are calibrated by means of Bayesian inference from reference aerodynamic data for a dense gas flow over a wing section. Flow thermodynamic conditions are such that the gas thermodynamic behavior strongly deviates from that of a perfect gas. In the aim of assessing the proposed methodology, synthetic calibration data –specifically, wall pressure data– are generated by running the numerical solver with a more complex and accurate thermodynamic model. The statistical model used to build the likelihood function includes a model-form inadequacy term, accounting for the gap between the model output associated to the best-fit parameters, and the true phenomenon. Results show that, for all of the relatively simple models under investigation, calibrations lead to informative posterior probability density distributions of the input parameters and improve the predictive distribution significantly. Nevertheless, calibrated parameters strongly differ from their expected physical values. The relationship between this behavior and model-form inadequacy is discussed.

*Keywords:* Bayesian calibration, Uncertainty quantification, Dense-gas flow

---

---

\*Corresponding author

*Email addresses:* [xavier.merle@ensam.eu](mailto:xavier.merle@ensam.eu) (X. Merle), [paola.cinnella@ensam.eu](mailto:paola.cinnella@ensam.eu) (P. Cinnella)

## 1. Introduction

Flows of complex organic fluids close to saturation conditions are encountered in several engineering applications like energy conversion cycles ([1, 2, 3]), high-Reynolds wind tunnels ([4]), chemical transport and processing ([5]), and refrigeration ([6]). In thermodynamic regions close to the liquid-vapor coexistence curve, the thermodynamic behavior of fluids with high molecular complexities differs significantly from that of a perfect gas and can no longer be represented by the polytropic perfect gas law. The largest deviations are encountered in the so-called dense gas flows ([7, 8, 9]) -for which a non-ideal dependence of the speed of sound on the fluid density is observed when the flow is submitted to isentropic perturbations- and most of all for a class of fluids known as the Bethe-Zel'dovich-Thompson (BZT) fluids ([10, 11, 12]). Examples of gases exhibiting this special behavior are given by heavy hydro- and fluorocarbons and siloxanes, for which accurate and comprehensive thermodynamic data are scarce. As a consequence, high-accurate equations of state (EOS), i.e. thermodynamic laws designed to describe the fluid thermal and caloric behavior in the region of interest, are in general not available. Now, numerical simulations of dense gas flows, i.e. flows of molecularly complex gases at pressures and densities of the general order of magnitude of those of the liquid/vapor critical point, can be extremely sensitive to the model used to describe the fluid thermodynamic behavior [13]. This sensitivity is particularly large for BZT fluids, which are theoretically predicted to exhibit non-classical gas dynamic behaviors, like expansion discontinuities and splitting shocks, in a tiny thermodynamic region close to the liquid/vapor coexistence curve [14].

On the other hand, reliable simulations of compressible flows with complex thermodynamic behavior require the quantification of thermodynamic modeling errors, especially for those applications that look for improvements of the order of a few percents of the system performance, e.g. energy conversion cycles (see e.g. [3]).

For a given EOS, uncertainties of two kinds exist. The first source of uncertainty is represented by the mathematical form of the EOS to be used for a given fluid; on the other hand, the material-dependent coefficients associated to the EOS have to be calibrated

from the available experimental data, which in turn are imperfectly known. Modeling uncertainties can be reduced to levels on the same order of the experimental uncertainty for molecularly simple well known fluids like water, hydrogen, or carbon dioxide (see [15]), i.e. even as low as 0.1 %. For molecularly complex fluids reserved essentially to an industrial use, like the dense gases of interest here, high accurate experimental data are more hardly available, so that a significant uncertainty on the closure coefficients may exist. Specifically, critical-point properties and acentric factors are particularly important data for the calibration of EOS, and are commonly used as input parameters in cubic EOS as, e.g., the Soave–Redlich–Kwong [16] and Peng–Robinson–Stryjek–Vera EOS [17], and virial EOS like Martin–Hou [18] and Benedict–Webb–Rubin [19]. However, in the common practice these input parameters are most often used without including statistical uncertainty information. Precisely, they are input to the model to obtain deterministic solutions corresponding to their nominal values, despite that small errors, e.g. in the critical properties, may sometime affect the quality of final results in a quite dramatic fashion [20]. In addition to uncertainties associated to thermodynamic model parameters, previous work [13] shows that for some particularly complex gases the uncertainty associated to the mathematical form of the EOS, or model-form uncertainty, can be even overwhelming with respect to the parametric uncertainty.

In this work, we adopt a Bayesian approach to quantify modeling uncertainty associated to thermodynamic models used for dense gas flow simulations, i.e. we take the point of view that the uncertainty associated to both model parameters and mathematical form can be represented in terms of probability. A similar approach was used in the work of Robinson et al. [21], who investigated the impact of parametric and modeling uncertainties associated to the Mie–Grunheisen thermodynamic model by using, as a first step, Bayesian inference techniques to obtain a posteriori estimates of the closure coefficients associated to the EOS and, as a second step, by propagating posterior distributions through a numerical code for shock hydrodynamics. Precisely, they first calibrated the EOS coefficients using thermodynamic experimental data available for Aluminium 6061, and then propagated the posteriors through the fluid dynamic model. Here, we investigate the possibility of calibrating thermodynamic models by using directly aerodynamic

data, i.e. data describing the flow behavior, instead of strictly thermodynamic information. For some engineering applications using complex fluids, this information is expected to be more readily available, e.g. through pressure taps inserted in the flow, than thermodynamic data. Moreover, this approach allows calibrating the model directly on observed data for the output quantities of interest, like velocity, pressure, and temperature fields. The statistical model adopted is similar to that used in [22] and [23] to calibrate closure parameters of turbulence models for a boundary layer flow. The procedure includes a statistical "model-inadequacy" term accounting for the gap between the model response obtained with the best-fit parameters, and the (unobserved) true phenomenon. The feasibility of the proposed approach is investigated for a simple case study, i.e. the transonic flow of a silicon oil, namely cyclopentasiloxane (D5), past an airfoil, as in [13]. Experimental data being not available for this case, and since the scope of the present work is assessing the proposed calibration approach and not solving a realistic flow configuration, "synthetic" calibration data are generated by numerically computing the pressure distribution around the airfoil using a reference EOS based on Helmholtz free energy [24], which is considered as the best available thermodynamic model for the selected fluid. Numerical solutions of the flow equations are generated by means of a finite volume code based on a third-order accurate numerical scheme [25]. Using these synthetic data, we calibrate two well-known cubic EOS, namely, the Peng-Robinson-Stryjek-Vera [17] (PRSV) and the Soave-Redlich-Kwong [16] (RKS), and the five-term virial Martin-Hou model [18] (MAH), coupled to the numerical flow solver. This results in posterior probability distributions for the model parameters. Moreover, calibration allows quantifying the model-form uncertainty. This information is propagated back through the flow solver for providing posterior predictive distributions of the quantity of interest.

The paper is organised as follows: in Section 2 we introduce the physical problem, the governing equations used to model the fluid flow of interest, and the thermodynamic models used to describe the fluid thermodynamic behavior, as well as the numerical methods used to solve them. In Section 3 we present the reference data used for calibration. Section 4 introduces an analytical model used to replace the costly numerical flow solver to speed up prior and posterior uncertainty quantification analyses. A preliminary sensi-

tivity analysis is carried out in Section 5 to identify the most influential parameters and interactions between them. Section 6 sets the Bayesian calibration framework, including a discussion of the surrogate model used to represent the CFD code response to changes in the input parameters. Finally, section 7 illustrates and discusses the numerical results, with focus on the role played by the model-inadequacy term to capture the calibration data.

## 2. Problem statement

In this study, we develop a methodology for calibrating thermodynamic models for the numerical simulation of dense gas flows, with focus on dense gas flow through energy conversion machines, and namely turbines of Organic Rankine Cycles (ORC). These are characterized by the presence of bladed disks. Blade sections can be roughly seen as airfoils. Hereafter, we investigate the feasibility of our calibration methodology for a simplified configuration, roughly representative of a blade section, i.e. an isolated airfoil. Such kind of configuration was often used in the past to investigate qualitatively dense gas effects in turbomachinery [26, 27]. Moreover, for this simplified problem sensitivity studies of the impact of thermodynamic uncertainties on flow simulation results were carried out in [13].

Precisely, the selected case study is represented by the steady transonic flow of a dense gas over a wing section, namely, a NACA0012 airfoil. The working fluid is a siloxane (silicon oil), known with the commercial name of D5 (chemical formula  $((CH_3)_2SiO)_5$ ).

The free-stream conditions of the flow, i.e. flow velocity and thermodynamic conditions at the far field, are treated in the following as deterministic and correspond to a value of the Mach number  $M_\infty = 0.95$ , an angle of attack equal to  $0^\circ$ , a reduced pressure  $p_{r,\infty} = 0.92$  and  $\rho_{r,\infty} = 0.5263$ .

Hereafter, we briefly describe the main features of dense gas flows, then introduce the equations used to model the flow, as well as the thermodynamic submodels used to describe the gas thermodynamic behavior, and finally provide some information about the flow solver used to generate numerical solutions.

### 2.1. Dense gas flows

Dense gas flows are flows of gases at thermodynamic conditions close to the liquid/vapor critical point, where the perfect gas law is invalid. In the dense regime, some heavy polyatomic fluids, referred to as the Bethe–Zel’dovich–Thompson (BZT) fluids, can exhibit non classical nonlinearities, such as expansion shock waves, mixed shock/fan waves, and splitting shocks (see, e.g. [27] and references cited therein). These peculiar behaviors occur when the fundamental derivative of gas dynamics [12]

$$\Gamma := 1 + \frac{\rho}{a} \left( \frac{\partial a}{\partial \rho} \right)_s \quad (1)$$

exhibits a region of negative values, between the upper saturation curve and the  $\Gamma = 0$  contour. In the preceding equation,  $\rho$  is the fluid density,  $a$  the sound speed, and  $s$  the entropy. Such a region is often referred to as the inversion zone, and the  $\Gamma = 0$  contour is called the transition line.  $\Gamma$  can be interpreted as a measure of the rate of change of the sound speed with density in isentropic perturbations. If  $\Gamma < 1$ , the flow will exhibit an uncommon sound speed variation in isentropic perturbations:  $a$  grows in isentropic expansions and decreases in isentropic compressions, the opposite of what happens in “common” fluids. For perfect gases,  $\Gamma$  is equal to  $(\gamma + 1)/2$ , where the specific heats ratio  $\gamma$  is always greater than 1 for thermodynamic stability reasons; therefore,  $\Gamma > 1$  as well. The simultaneous presence of regions of negative values of the fundamental derivative within a flow field leads to a drastically different behavior compared to that of a classical gas. This is why the size and location of the inversion zone has a deep impact on the resulting flow field. The inversion zone has been found to be extremely sensitive to the equation of state in use and the associated input parameters (see e.g. [13]).

### 2.2. Governing equations

Turbomachinery dense gas flows are governed by the equations for equilibrium for non-reacting flows, i.e. the Navier-Stokes equations. Since however dense gas effects mostly affect inviscid flow behavior and in the aim of achieving a simplified model problem to assess our calibration methodology, in the present study we neglect viscous effects and restrict to the Euler equations. The use of an inviscid flow model is also justified by the fact that the data used for calibration are values of the pressure coefficient at

different locations around the blade wall. For the present case study, the boundary layer is supposed to be thin and attached (which corresponds to the hypothesis of high Reynolds number flow, thin airfoil, no angle of attack). As a consequence, according to boundary layer theory the pressure distribution at the wall is substantially the same as in the inviscid flow.

The governing equations are written in integral form for a control volume  $\Omega$  with boundary  $\partial\Omega$ :

$$\frac{d}{dt} \int_{\Omega} \mathbf{w} d\Omega + \int_{\partial\Omega} \mathbf{f} \cdot \mathbf{n} dS = 0 \quad (2)$$

In equation (2),  $\mathbf{w}$  is the conservative variable vector, where

$$\mathbf{w} = (\rho, \rho\mathbf{v}, \rho E)^T$$

$\mathbf{n}$  is the outer normal to  $\partial\Omega$ , and  $\mathbf{f}$ , is the flux density:

$$\mathbf{f} = \left( \rho\mathbf{v}, \rho\bar{\mathbf{I}}, \rho\mathbf{v}\mathbf{v}, \rho\mathbf{v}H \right)^T$$

where  $\rho$  is the density,  $\mathbf{v}$  is the velocity vector,  $E$  the specific total energy,  $H = E + p/\rho$  the specific total enthalpy,  $p$  is the pressure and  $\bar{\mathbf{I}}$  is the unit tensor. The preceding equations are completed by a thermal equation of state:

$$p = p(\rho(\mathbf{w}), T(\mathbf{w})) \quad (3)$$

with  $T$  the total pressure, and by a caloric equation of state for the specific internal energy  $e$ , which must satisfy the compatibility relation:

$$e = e(\rho(\mathbf{w}), T(\mathbf{w})) = e_0 + \int_{T_0}^T c_{v,\infty}(T') dT' - \int_{\rho_0}^{\rho} \left[ T \left( \frac{\partial p}{\partial T} \right)_{\rho} - p \right] \frac{d\rho'}{\rho'^2} \quad (4)$$

In Equation (4),  $c_{v,\infty}$  is the ideal gas specific heat at constant volume, quantities with a prime superscript are auxiliary integration variables, and subscript 0 indicates a reference state. The caloric equation of state is completely determined once a variation law for  $c_{v,\infty}$  has been specified.

### 2.3. Equations of state

In this work, we consider several alternative thermodynamic models for the thermal and caloric equations of state, which are described in the following: the Redlich-Kwong



cubic equation of state as modified by Soave [16] (denoted hereafter as the RKS equation), the Peng-Robinson-Stryjek-Vera (PRSV) cubic equation of state [17], and the multiparameter Martin-Hou (MAH) equation [18], based on a five-term virial expansion.

In addition to the preceding EOS, we consider, for calibration and validation purposes as discussed later, a more complex and accurate thermodynamic model, namely, the multiparameter technical equation of state introduced by Span and Wagner (SW) [15] with coefficients adjusted to D5 in [24].

This is a complex technical equation of state, which is considered as the most accurate and complete thermodynamic model presently available for this fluid, provided that a sufficiently large set of high-quality experimental data is available for a reliable fitting of the model coefficients, i.e. for reproducing experimental data with very high accuracy. In practice, the equation coefficients are obtained by means of a regression procedure: the experimental data used for calibration are weighted in such a way that high-quality data contribute with higher weights. Experimental data with uncertainties larger than a required threshold are discarded, or, when no or insufficient data are available for a certain region, they are weighted according to their level of uncertainty. So, the resulting equation of state is expected to be intrinsically robust to input data uncertainties.

### 2.3.1. RKS model

The Redlich-Kwong equation of state with the Soave modification [16] is commonly considered as one of the best available two-parameter equations of state. In reduced form, i.e. with thermodynamic quantities normalized with respect to their values at the liquid/vapor thermodynamic critical point, the equation writes:

$$p_r = \frac{T_r/Z_c}{v_r - b_r} - \frac{a_r(T_r)/T_r^{0.5}}{v_r(v_r + b_r)} \quad (5)$$

where the subscript  $r$  denotes a reduced quantity,  $(\bullet)_r = (\bullet)/(\bullet)_c$ , the subscript  $c$  denotes critical-point quantities,  $v$  the specific volume, and:

$$a_r(T_r) = (0.42747/Z_c)\alpha(T_r), \quad b_r = 0.08664/Z_c$$

with  $Z_c = (p_c v_c)/(RT_c)$  the critical compressibility factor and  $R$  the gas constant. The function  $\alpha(T_r)$  is an adimensional relationship depending on the reduced temperature

and the substance acentric factor  $\omega$ :

$$\alpha(T_r) = [1 + m(1 - T_r^{0.5})]^2, \quad (6)$$

with

$$m = 0.480 + 1.57\omega - 0.176\omega^2 \quad (7)$$

The critical compressibility factor is univocally determined by imposing that  $p_r$  equals 1 at the critical point, i.e. for  $v_r = 1$ ,  $T_r = 1$ , which leads to the solution of a cubic equation for  $Z_c$  with only one relevant root ( $Z_c = 0.3467$ ). Thus, the only free parameter left in the thermal equation is the acentric factor  $\omega$ .

The RKS equation is supplemented with a model for the ideal gas contribution to the specific heat at constant volume, represented here by a power law of the form:

$$c_{v,\infty}(T) = c_{v,\infty}(T_c) \left( \frac{T}{T_c} \right)^n \quad (8)$$

where the  $c$  subscript denotes critical-point values, and the exponent  $n$  and the ideal-gas-limit isocoric specific heat at the critical temperature  $c_{v,\infty}(T_c)$  are material-dependent constants.

In summary, the RKS model in non-dimensional form (5), (8) considered in the present work depends on three uncertain parameters, namely the acentric factor  $\omega$ , the exponent  $n$  and the reduced ideal-gas constant-volume specific heat at the critical temperature  $c_{v,\infty}(T_c)$ .

### 2.3.2. PRSV model

The cubic Peng and Robinson equation of state with modifications suggested by Stryjek and Vera [17] can be written, in reduced form, as:

$$p_r = \frac{T_r/Z_c}{v_r - b_r} - \frac{a_r(T_r)}{v_r^2 + 2b_r v_r - b_r^2} \quad (9)$$

with  $a_r = (0.457235/Z_c^2) \alpha(T_r)$ ,  $b_r = 0.077796/Z_c$  and the corrective factor  $\alpha(T_r)$  given again by relation (6). To improve the results, Stryjek and Vera recorrealted  $m$  as a function of acentric factor using the same polynomial form but up to third-order, obtaining

$$m = 0.378893 + 1.4897153\omega - 0.17131848\omega^2 + 0.0196554\omega^3 \quad (10)$$

Like in the RKS model  $Z_c$  is univocally determined through the solution of a cubic equation, derived by enforcing the passage of the critical isotherm through the critical point ( $Z_c = 0.3112$  for the PRSV equation). Thus, the non-dimensional PRSV model (9, 11) depends just only on the uncertain parameters  $\omega$ ,  $n$ , and  $c_{v,\infty}(T_c)$ .

### 2.3.3. MAH model

The comprehensive thermal equation of state of Martin and Hou [18] reads:

$$p_r = \frac{T_r}{Z_c(v_r - b_r)} + \sum_{i=2}^5 \frac{f_{r,i}(T_r)}{(v_r - b_r)^i} \quad (11)$$

with  $b_r = \frac{1-\beta/15}{Z_c}$ ,  $\beta = 20.533 - 31.883Z_c$ , and

$$f_{r,i}(T_r) = A_{r,i} + B_{r,i}T_r + C_{r,i} \exp(-kT_r)$$

with the Boltzmann constant  $k = 5.475$ . The gas-dependent coefficients  $A_{r,i}$ ,  $B_{r,i}$ ,  $C_{r,i}$  can be expressed in terms of the critical temperature and pressure, the critical compressibility factor, the Boyle temperature (which may be expressed as a function of the critical temperature) and one point on the vapour pressure curve. The MAH equation of state is supplemented again by Eq. (8) to compute the ideal gas contribution to the specific heat at constant volume. Globally, the MAH thermodynamic model in reduced form, Eq. (11) and (8), requires the knowledge of six material-dependent parameters, namely, the critical pressure  $p_c$ , the critical temperature  $T_c$ , the critical compressibility factor  $Z_c$ , the normal boiling temperature  $T_b$ , the exponent  $n$  and the reduced ideal-gas constant-volume specific heat at the critical temperature  $c_{v,\infty}(T_c)$ . Nevertheless, previous results [13] show that  $Z_c$ ,  $T_b$  and  $n$  have a relatively small impact on the results. As a consequence, we choose to neglect them for the following of the study. Nominal values of the different input parameters corresponding to D5 are given in Tab. 1.

### 2.3.4. SW model

The last thermodynamic model considered in this study is a 12-parameter technical equation of state based on the functional form for non polar fluids proposed by Span and Wagner ([15]). This is written as an expression for the reduced Helmholtz free energy  $\Phi$  (i.e. normalized with  $RT_c$ ), sum of an ideal-gas part,  $\Phi^0$ , function of the ideal-gas

isobaric heat capacity  $c_{p,\infty}$ , and by a residual term  $\Phi^r$  that takes into account real-gas corrections:

$$\begin{aligned}
\Phi^r(\delta, \tau) = & n_1 \delta \tau^{0.25} + n_2 \delta \tau^{1.25} + n_3 \delta \tau^{1.5} + n_4 \delta^3 \tau^{0.25} \\
& + n_5 \delta^7 \tau^{0.875} + n_6 \delta \tau^{2.375} \exp(-\delta) \\
& + n_7 \delta^2 \tau^{2.0} \exp(-\delta) + n_8 \delta^5 \tau^{2.125} \exp(-\delta) \\
& + n_9 \delta \tau^{3.5} \exp(-\delta^2) + n_{10} \delta \tau^{6.5} \exp(-\delta^2) \\
& + n_{11} \delta^4 \tau^{4.75} \exp(-\delta^2) + n_{12} \delta^2 \tau^{12.5} \exp(-\delta^3)
\end{aligned} \tag{12}$$

where  $n_1, \dots, n_{12}$  are substance-specific coefficients,  $\delta = \rho/\rho_c$  is the reduced density and  $\tau = T_c/T$  is the inverse of the reduced temperature. Material-dependent coefficients for D5 have been taken from Ref. [24], to which we refer for more details. The thermal and caloric EOS are derived from the following thermodynamic relations:

$$\frac{p}{\rho RT} = 1 + \delta \left( \frac{\partial \Phi^r}{\partial \delta} \right)_\tau ; \quad \frac{e}{RT} = \tau \left[ \left( \frac{\partial \Phi^0}{\partial \tau} \right)_\delta + \left( \frac{\partial \Phi^r}{\partial \tau} \right)_\delta \right]$$

For the calculation of caloric properties, the SW EoS (Eq. (12)) is supplemented by the ideal gas contribution to the specific heat at constant pressure, which is now approximated here as a polynomial function of the temperature:

$$\frac{c_{p,\infty}(T_r)}{R} = \frac{c_{v,\infty}(T_r)}{R} + 1 = c_1 + c_2 T + c_3 T^2 + c_4 T^3 \tag{13}$$

where the polynomial coefficients  $c_i$  depend again on the substance under consideration and are given for D5 in [24]. This model is considered as more accurate than the preceding ones. For this reason, it is used in the following to generate reference solutions used to calibrate the simpler RKS, PRSV, and MAH models.

#### 2.4. Numerical setup

Numerical solutions for the dense gas flow of interest are found by means of an in-house dense gas flow solver. The governing equations are discretized using a cell-centered finite volume scheme for structured multi-block meshes of nominal third-order accuracy, which allows computing flows governed by an arbitrary equation of state. We refer to [25] and the references therein for more details about the numerical solver and its validation.

Since NACA0012 is a symmetric airfoil and the flow has no angle of attack, the flow is steady and symmetric. For this reason, only one half of the computational domain

is simulated, with symmetry conditions enforced on the axis. The results are computed using a half C-shaped grid composed by  $100 \times 32$  cells. The mean height of the first cell closest to the wall is equal to 0.001 chords for the medium grid. The outer boundary is located 10 chords away from the airfoil. Previous studies (see [13]) show that this level of mesh resolution is fine enough to keep the numerical error on the quantity of interest to within reasonably small values. A picture of the computational mesh is provided in Fig. 1.

Obtaining a converged numerical solution for this steady flow configuration with the above numerical settings takes approximately 5 to 10 minutes according to the specific run (equation of state, choice of thermodynamic parameters) on a standard desktop personal computer. This is a relatively short computational time when seeking for deterministic solutions. Unfortunately, sensitivity analyses to uncertain input data and, even more, Bayesian calibration algorithms, may require several thousand of runs of the deterministic solver, leading to unacceptably high computational times, as discussed later. This is why we adopt in the following a surrogate model, i.e. an inexpensive analytic expression approximating the response of the numerical solver to input parameter modifications. Using the surrogate model, approximate solutions are obtained in the matter of seconds, which is compatible with the uncertainty quantification and calibration algorithms considered in this paper.

### 3. Calibration data

Unfortunately, no experimental data have been made available in the scientific literature for flows of dense gases, unlike for light fluids as air, carbon dioxide or nitrogen. The only experimental data available for a dense gas flow were produced for a shock tube flow of Freon-13, in the aim to show up the appearance of expansion waves [28]. However, these experiments were questioned by several authors (see, e.g. [29]). Efforts for carrying out new shock tube experiments were done at University of Colorado Boulder [29] and at TU Delft [30], but no data have been made available up to present. A dense gas nozzle facility is presently under construction at Politecnico di Milano [31]. Thus, no experimental data exist for flows of dense gases past airfoils.

Since the present paper aims first of all at proving the feasibility of the calibration methodology, we generate synthetic "experimental" data by running the dense gas solver with the most complex and accurate thermodynamic model, namely, the SW model.

Fig. 2 shows a deterministic numerical solution obtained by using this model, and precisely the Mach number contours around the airfoil. The flow becomes locally supersonic close to the airfoil, and the supersonic region is terminated by a shock wave located at about 80% of the airfoil chord, driving the flow back to subsonic conditions. Fig. 3 illustrates the behavior of the pressure coefficient

$$C_p = \frac{p - p_\infty}{\frac{1}{2}\rho_\infty U_\infty^2}$$

along the airfoil wall with  $U_\infty$  the free-stream velocity. As the incoming flow is deviated by the airfoil nose, a suction region (i.e. a region of  $C_p < 0$ ) is generated, corresponding to supersonic conditions. Then, the flow is first abruptly compressed by the shock wave and then further compressed downstream of the shock, so that it gradually returns to the free-stream pressure.

The location and intensity of the shock wave are extremely sensitive to the EOS in use, as discussed in [13]. Due to uncertainties intrinsic to the SW model and, more precisely, to the ancillary equation (13) for the specific heat capacity, the flow solver output is also uncertain. According to previous studies [13], this uncertainty is roughly on the order of 10% for the pressure coefficient.

In the following of this paper, we calibrate the input coefficients of three relatively simple EOS by using the synthetic data generated from SW simulations of D5 flow around the NACA0012 airfoil. Precisely, we use as calibration data the numerical values of the pressure coefficient  $C_p$ , evaluated at 17 equally distributed locations along the airfoil wall (simulated pressure taps). Figure 4 displays the iso-contours of the pressure coefficient for the SW model, as well as pressure tap locations. Then, these data are perturbed by a Gaussian noise with zero mean and standard deviation equal to 10% of the nominal value, to simulate the effect of experimental errors. The quantity of interest we want to predict is again the wall pressure distribution.

As discussed in the Section 6, the statistical model used for the present calibrations takes into account the uncertainty on the reference data.

#### 4. Surrogate model

As seen in Section 2, deterministic numerical simulations of the problem under investigation take 5 to 10 minutes on a standard personal computer. However, in order to make reliable predictions of the physical system it is necessary to understand what effect the uncertainties in the inputs have on the output Quantity of Interest (QoI), which are the most influential parameters and, if needed, to update the uncertain parameter distributions according to the available observations. A classical sensitivity analysis methodology consists in propagating the uncertain parameters through the model via some form of Monte Carlo (MC) sampling [32]. Although the convergence of the MC method is independent of the number of uncertain variables, it is slow due to the fact that the error decreases with the square root of the number of samples [33]. Since many modern computer codes are computationally expensive, utilizing MC sampling either for sensitivity analysis or calibration purposes is not tractable.

To overcome this difficulty, a standard approach consists in replacing the expensive numerical model by an inexpensive analytical function approximating, as accurately as possible, the system response to the input parameters, i.e. a surrogate model. An efficient surrogate model has to provide an accurate approximation of the true solution (to within a given tolerance) by using information from a reduced set of samples (i.e. a number of samples much lower than those required by the Monte Carlo method). A comprehensive overview of different possible surrogate models for uncertainty quantification and calibration is beyond the scope of this paper. For more details, it is possible to refer to, e.g. [34, 35, 36, 37].

Since the primary objective of this work is not to construct an efficient and accurate surrogate model for the problem of interest, but rather the calibration methodology, we restrict ourselves to simple polynomial surrogates. Further research on surrogate modeling for calibration and uncertainty propagation is underway [38].

Specifically, a surrogate model approximating the response of the flow solver to the EOS parameters is constructed by using piecewise multidimensional Lagrange interpolations. This is a simple strategy, but it is found to be efficient and accurate enough for the present problem, characterized by relatively low-dimensional parameter spaces (no more than three uncertain parameters).

Two kinds of grid are employed: Gauss-Lobatto grids are used for a preliminary exploration of the parameter space, since they are expected to give more accurate results when using a small number of nodes; then, the parameter space is remeshed by means of Cartesian grids, which are easier to handle and enable local refinement if needed. Moreover, Cartesian meshes allow to significantly reduce the consuming time.

Let us represent the solution to the physical problem described in section 2, by a generic mathematical model of the form:

$$y = M(\underline{x}; \theta) \quad (14)$$

where  $M$  is a functional relation representing the physical problem (*i.e* the governing equations and the closure EOS),  $\underline{x}$  is a vector of explanatory variables, *i.e* known model inputs, including the space coordinates,  $\theta$  is a vector of input parameters associated to the model description, and  $y$  is a vector of quantities of interest output from the model. For the specific case study under investigation:

- $M$  results from the combination of the governing equations, that remain always the same, with the specific EOS in use, so that can be considered to directly represent the EOS;
- the uncertain parameter vector  $\theta$  only contains the uncertain thermodynamic input parameters associated to the chosen EOS;
- because the calibration data are values of the pressure coefficient along the airfoil wall, as discussed in Section 3, the model output  $y$  corresponds to the computed values of the pressure coefficient at different locations along the airfoil (other choices would be possible); as such it depends on the chord-wise coordinate  $x$  (explanatory variable) and on the uncertain thermodynamic inputs  $\theta$ .

Consequently, the numerical problem can be represented by:

$$y = M(x; \theta) \quad (15)$$



According to the preceding assumptions, the surrogate model  $y_S$  for problem (15) can be written

$$y_S[y = M(x; \theta); n_1, n_2, \dots, n_D] = \sum_{i_1=0}^{n_1} \sum_{i_2=0}^{n_2} \dots \sum_{i_D=0}^{n_D} \alpha_{i_1 i_2 \dots i_D}(x) \prod_{j=i_1}^{i_D} l_j(\theta_j) \quad (16)$$

where the  $l_j(\theta_j)$  are the one dimensional Lagrange characteristic polynomials in each direction, the  $\alpha_{i_1 i_2 \dots i_D}$  are interpolation coefficients dependent on the explanatory variables, and  $D$  is the dimensionality of the parameter space. With this surrogate model, we can exactly evaluate multivariate polynomial functions of the form:

$$P(\theta_1, \theta_2, \dots, \theta_D) = (a_{n_1,1}\theta_1^{n_1} + a_{n_1-1,1}\theta_1^{n_1-1} + \dots + a_{1,1}\theta_1 + a_{0,1}) \quad (17)$$

$$\times (a_{n_2,2}\theta_2^{n_2} + a_{n_2-1,2}\theta_2^{n_2-1} + \dots + a_{1,2}\theta_2 + a_{0,2}) \quad (18)$$

$$\times \dots \quad (19)$$

$$\times (a_{n_D,D}\theta_D^{n_D} + a_{n_D-1,D}\theta_D^{n_D-1} + \dots + a_{1,D}\theta_D + a_{0,D}) \quad (20)$$

where  $a_{i,j} \in \mathbb{R}$  is the  $i^{th}$ -order term coefficient of the  $j^{th}$  variable.

When Gauss-Lobatto grids are used, a global surrogate model is constructed over the whole parameter space with about 10 points per direction. With Cartesian grids, a piecewise polynomial interpolation strategy is adopted to avoid Runge oscillations close to the borders. Precisely, we construct local 5th-order polynomials by building an hypercube made of 6 point per direction, enclosing the point to be interpolated. The interpolation points are chosen in such a way as to obtain centered interpolation formulas. The local polynomial order is decreased close to the boundaries in order to keep centered approximations.

The accuracy of the surrogate models used in this study is checked by taking 10 random points in the parameter space and by comparing the surrogate model output to the exact model output. For that, we compute the relative error between the exact and the surrogate model for each abscissa and then we check that the spatial mean of the relative error is bellow 2% for each random point:

$$\text{Relative error} = \left| \frac{y(x; \theta) - y_S(x; \theta)}{y(x; \theta)} \right| \quad (21)$$

On Figure 5, we plot the relative error for two different surrogate models of the Peng-Robinson EOS. On each subfigure, we plot the pressure coefficient for the exact and the surrogate model, along with the relative error for 81 abscissas  $x_i$  along the airfoil wall. Subfigures 5a and 5b are respectively the best and the worst cases, in terms of resulting relative error, among the ten random points, for a three dimensional parameter space and a Gauss-Lobatto grid. Subfigure 5c and 5d represent the best and the worst cases for a two dimensional parameter space and a Cartesian mesh. The  $L_1$  norm of the relative error over the airfoil is 0.000719 , 0.011649, 0.000106, and 0.001052 for solutions displayed in Figs 5a, 5b, 5c, and 5d, respectively.

The  $L_\infty$  norm (maximum relative error along  $x$ ) is 0.015940, 0.889119, 0.000581, and 0.057111 for solutions displayed in Figs 5a, 5b, 5c, and 5d, respectively.

The largest errors are generally found in the region close the shock wave (when this is present in the flow) or at the leading and trailing edges ( $x = 0$  and  $x = 1$ ) when the flow is smooth.

Since the present surrogate model ensures average relative errors lower than 1%, we consider that it provides an accurate enough approximation for the present purposes. The residual errors introduced by the surrogate model are taken into account in the statistical model, as discussed in Section 6.

## 5. Sensitivity analysis

In the following we investigate a priori the impact of thermodynamic uncertainties on predictions of the transonic flow of D5 over an airfoil. Precisely, as a first step we carry out a sensitivity analysis to identify the most influential coefficients with respect to the quantity of interest used for the subsequent calibration step. Hereafter we first describe the sensitivity analysis methodology adopted for the present calculations, afterwards, we present results obtained for our specific case study.

### 5.1. Sensitivity analysis by surrogate Monte Carlo sampling

To evaluate the influence of EOS parameters on the quantity of interest  $y$  in (15), we perform a sensitivity analysis according to Sobol's framework [39]. The contribution to

the variance of  $y$  of a random parameter  $\theta_i$  belonging to the random vector  $\theta$ , *i.e.* the first order Sobol index, is defined as:

$$S_i = \frac{\text{Var}[\text{E}[y|\theta_i]]}{\text{Var}[y]} \quad (22)$$

where  $\text{E}[X]$  and  $\text{Var}[X]$  denotes respectively the expected value and the variance of the random variable  $X$ . In the above, the  $\theta_i$  are independent parameters. Thanks to the Sobol decomposition of  $y$ :

$$y = M(\theta_1, \dots, \theta_D) = M_0 + \sum_{i=1}^D M_i(\theta_i) + \sum_{1 \leq i < j \leq D} M_{ij}(\theta_i, \theta_j) + \dots + M_{1\dots D}(\theta_1, \dots, \theta_D) \quad (23)$$

where we omitted the dependency on the explanatory variable  $x$ , to simplify notations, and where:

$$\begin{aligned} M_0 &= \text{E}[y] \\ M_i(\theta_i) &= \text{E}[y|\theta_i] - \text{E}[y] \\ M_{ij}(\theta_i, \theta_j) &= \text{E}[y|\theta_i, \theta_j] - \text{E}[y|\theta_i] - \text{E}[y|\theta_j] + \text{E}[y] \\ M_{ijk}(\theta_i, \theta_j, \theta_k) &= \text{E}[y|\theta_i, \theta_j, \theta_k] - \text{E}[y|\theta_i, \theta_j] - \text{E}[y|\theta_i, \theta_k] - \text{E}[y|\theta_j, \theta_k] \dots \end{aligned}$$

Sobol showed that the variance of the model can be decomposed in:

$$\text{Var}[y] = \sum_{i=1}^D V_i + \sum_{1 \leq i < j \leq D} V_{ij} + \dots + V_{1\dots D} \quad (24)$$

where

$$\begin{aligned} V_i &= \text{Var}[\text{E}[y|\theta_i]] \\ V_{ij} &= \text{V}[\text{E}[y|\theta_i, \theta_j]] - V_i - V_j \\ V_{ijk} &= \text{V}[\text{E}[y|\theta_i, \theta_j, \theta_k]] - V_{ij} - V_{ik} - V_{jk} - V_i - V_j - V_k \\ &\vdots \\ V_{1\dots D} &= \text{Var}[y] - \sum_{i=1}^D V_i - \sum_{1 \leq i < j \leq D} V_{ij} - \dots - \sum_{1 \leq i_1 < \dots < i_{D-1} \leq D} V_{i_1 \dots i_{D-1}} \end{aligned}$$

By extension, we can also consider higher-order Sobol indices, *i.e.* the contribution to the variance of  $y$  of  $k$  parameters  $\theta_{i_1}, \theta_{i_2}, \dots, \theta_{i_k}$  varying at a time:

$$S_{i_1 i_2 \dots i_k} = \frac{V_{i_1 i_2 \dots i_k}}{\text{Var}[y]} \quad (25)$$

The quantity in (25) is called the  $k^{\text{th}}$ -order Sobol index. By construction, the sum of all Sobol indices, whatever their order, is equal to one.

Finally, the total Sobol index  $S_{T,i}$ [40] is the sum of all of Sobol indices involving the parameter  $\theta_i$ , and represents the global influence of  $\theta_i$  on  $y$ .

In this study, we compute approximations of the Sobol indices based on Monte-Carlo sampling. This method ([39],[41]) uses two  $N$ -samples of input parameters  $\theta$  to compute the Sobol indices by comparing the variance of  $y$  computed from mixed values of  $\theta$ , picked from both samples.

Consider a  $N$ -sample  $\bar{\theta}_{(N)} = (t_{k1}, \dots, t_{kD})_{k=1 \dots N}$  of realizations of the input parameters  $(\theta_1, \dots, \theta_D)$ . The expected value of  $y$ ,  $E[y]=M_0$ , and the variance  $\text{Var}[y]=V$  are estimated by:

$$\hat{M}_0 = \frac{1}{N} \sum_{k=1}^N M(t_{k1}, \dots, t_{kD}) \quad \text{and} \quad \hat{V} = \frac{1}{N} \sum_{k=1}^N M^2(t_{k1}, \dots, t_{kD}) - \hat{M}_0^2 \quad (26)$$

Then, the first order indices are computed by:

$$V_i = \text{Var}[E[y|\theta_i]] = \underbrace{E[E[y|\theta_i]^2]}_{U_i} - E[E[y|\theta_i]]^2 = U_i - E[y]^2 \quad (27)$$

where

$$\hat{U}_i = \frac{1}{N} \sum_{k=1}^N M(t_{k1}^{(1)}, \dots, t_{k(i-1)}^{(1)}, t_{ki}^{(1)}, t_{k(i+1)}^{(1)}, \dots, t_{kD}^{(1)}) M(t_{k1}^{(2)}, \dots, t_{k(i-1)}^{(2)}, t_{ki}^{(1)}, t_{k(i+1)}^{(2)}, \dots, t_{kD}^{(2)}) \quad (28)$$

where the exponents  $(1)$  and  $(2)$  denotes respectively the first and the second  $N$ -sample.

Finally

$$\hat{S}_i = \frac{\hat{V}_i}{\hat{V}} = \frac{\hat{U}_i - \hat{M}_0^2}{\hat{V}} \quad (29)$$

For the second order indices, we compute:

$$V_{ij} = \text{Var}[E[y|\theta_i, \theta_j]] - V_i - V_j = U_{ij} - E[Y]^2 - V_i - V_j \quad (30)$$

and  $U_{ij}$  is computed by

$$\begin{aligned} \hat{U}_{ij} = & \frac{1}{N} \sum_{k=1}^N M(t_{k1}^{(1)}, \dots, t_{k(i-1)}^{(1)}, t_{ki}^{(1)}, t_{k(i+1)}^{(1)}, \dots, t_{k(j-1)}^{(1)}, t_{kj}^{(1)}, t_{k(j+1)}^{(1)}, \dots, t_{kD}^{(1)}) \\ & \times M(t_{k1}^{(2)}, \dots, t_{k(i-1)}^{(2)}, t_{ki}^{(1)}, t_{k(i+1)}^{(2)}, \dots, t_{k(j-1)}^{(2)}, t_{kj}^{(1)}, t_{k(j+1)}^{(2)}, \dots, t_{kD}^{(2)}) \end{aligned} \quad (31)$$

and then

$$\hat{S}_{ij} = \frac{\hat{U}_{ij} - \hat{M}_0^2 - \hat{V}_i - \hat{V}_j}{\hat{V}} \quad (32)$$

and so on for the higher order indices.

To make sure of obtaining an accurate enough evaluation of the indices, even for coefficients with a very weak influence on  $y$ , we draw 1,000,000 samples from a joint prior distribution of the parameters. This is constructed as the product of large, non-informative uniform marginal priors for each one of the physical parameters involved in the EOS, given in Tab. 2 for RKS, PRSV, and MAH, respectively. The underlying hypothesis is that all of the parameters are independent, which represents our prior belief about the parameters. As discussed in the following, the Bayesian calibration process may show up *a posteriori* dependencies between the parameters.

Given the huge amount of calculations required for the sensitivity analysis and the cost of a dense gas solver run (see Section 2), the Monte Carlo method is not directly coupled to the computed code, but instead to the piecewise polynomial surrogate model described in Section 4. This comes to apply formulas (26 to 32) to the analytical function  $y_S$  instead of the computer code output  $y = M(\theta)$ .

## 5.2. Sensitivity analysis results

The quantity of interest  $y_S$  being the surrogate model estimate of the pressure coefficient at 17 locations along the airfoil wall, we compute Sobol indices relative to each chordwise location. Figure 6 shows distributions of the first-order Sobol indices along the airfoil wall for the three EOS under investigation. Table 3 summarizes the results of the analysis, and lists the overall most influential coefficients for each EOS. The whole set of results is given in Appendix 1. Note that the indices are sensibly constant all along the airfoil wall, with a negligibly small contribution of high-order interactions, except at

the airfoil trailing edge. In this region, strong nonlinear effects are developed due to the presence of a shock wave, and high-order interactions become important. These correspond to high values of the second-order Sobol indices between the two most influential coefficients for each EOS, i.e.  $\omega$  and  $c_{v,\infty}$  or  $T_c$  and  $p_c$  according to the EOS. We show in the following that these interactions have a strong impact on the calibration results.

From the preceding analysis it appears that, for each EOS, the dimensionality of the parameter space can be reduced by excluding from the subsequent calibration step parameters exhibiting a weak influence on  $C_p$ , typically those for which the corresponding first-order Sobol index is lower than 0.1 (i.e. that contribute to the overall variance by less than 10%) all along the profile, which are expected not to be informed significantly by the calibration process. For RKS and PRSV, which belong to the same class of EOS (cubic EOS), the most influential parameters are found to be the acentric factor  $\omega$  and the reduced ideal-gas isochoric heat capacity  $c_{v,\infty}/R$ . For MAH, only the critical-point temperature  $T_c$  and, to a lesser extent, the pressure  $p_c$  have a significant influence on  $y_S$ . Parameters that are not used for the calibrations are kept at their usual nominal values. The following calibration step is performed by considering just only the two most influential physical parameters for each EOS as random variables, and by treating the rest as deterministic. This allows reducing the overall cost of the calibration step and to improve the accuracy of the surrogate model, thanks to the reduction of the parameter space.

## 6. Bayesian calibration methodology

This section describes the statistical procedure used to calibrate the closure parameters associated to the EOS described in Section 2.3.

Consider first the generic physical problem described in equation (15). In the Bayesian framework, knowledge (or lack of knowledge) about the value of a parameter is represented by a random variable, characterized by means of its probability density function (*pdf*). Consequently,  $M$  is a stochastic model, with stochastic parameters  $\theta$ , and  $y$  is also a random vector.

The main goal of calibration is to achieve new knowledge about  $\theta$  by constructing an improved representation of its *pdf*.

Assume that new amount of information is available, represented by a random vector of observed data  $\bar{d}$ . Bayes rule states that:

$$p(\theta|d = \bar{d}, y) \propto p(d = \bar{d}|y, \theta)p(\theta). \quad (33)$$

The right-hand side of equation (33) is made of two parts:

- $p(\theta)$  is the prior *pdf* and represents the initial belief about  $\theta$ ;
- $p(d = \bar{d}|y, \theta)$  is the likelihood, and corresponds to the probability to observe  $\bar{d}$ , a realization of the random variable  $d$ , if  $y$  and  $\theta$  are known exactly.

Finally, left-hand side of equation (33) is the posterior *pdf* and represents the updated knowledge of  $\theta$  given information about  $d$  and the model  $y$ . Note that  $d$  can be any available data and not necessarily of the same nature as the quantity of interest  $y$ .

For our concerns, the uncertain parameter vector  $\theta$  is the set of closure parameters associated to a given EOS,  $x$  represents problem inputs that are modeled as deterministic, e.g. the boundary conditions and the airfoil geometry,  $y$  is a vector of outputs from the dense gas solver  $M$ , and  $d$  is a random vector of experimental data.

The EOS coefficients are calibrated by using the synthetic data generated from SW simulations of D5 flow around the NACA0012 airfoil, as discussed in Section 3, and specifically values of the pressure coefficient  $C_p$ , at 17 locations along the airfoil wall.

From equation (33), it appears that the distribution of the prior and the likelihood have a strong influence on the posterior and have to be chosen carefully.

A common practice is to chose for the priors non-informative uniform distributions:

$$p(\theta) \sim \mathcal{U}([a, b]) \quad (34)$$

where the interval  $[a, b]$  is chosen large enough to ensure a good exploration of the parameter space, while avoiding to include unphysical values.

The likelihood function is a statistical model describing the probability of observing the data for a given choice of the model parameters. It may include information both about the observational error on the data and the model error, including the effect of surrogate modeling if needed. In the following, we use a likelihood function resulting from a multiplicative/additive error model, similar to that used in [22] for calibrating the closure coefficients of turbulence model from measured velocity profiles in a turbulent boundary layer. We refer to [42], [22] and the references cited therein for more details about possible choices for the construction of likelihood functions.

Precisely, the model includes an additive error modeling observational noise and a multiplicative term representative of the model inadequacy.

First, the data  $d$  at a given location  $x_i$  along the airfoil wall are modeled as:

$$d(x_i) = \hat{d}(x_i) + e_i \quad (35)$$

with

- $e_i$ , the experimental noise at location  $x_i$ ;
- $\hat{d}(x_i)$ , the true (unobserved) pressure coefficient value at  $x_i$ .

The observational errors,  $e_i$ ,  $i = 1, \dots, N$  ( $N = 17$  being the number of data) are taken here as independent and normally distributed with zero mean and a standard deviation equal to 10% of the observed value.

On the other hand, the true process  $\hat{d}$  is in general not exactly captured by the model (15) –even with the best possible choice of the parameters  $\theta$ – because of inadequacies intrinsic to the simplifying assumption used to mathematically modeling the physical phenomenon under study.

In this work, we assume the true process to be equal to the model output  $y(x_i, \theta)$ , multiplied by an error coefficient  $\eta_i$ :

$$\hat{d}(x_i) = \eta_i y(x_i, \theta) \quad (36)$$

which takes into account the discrepancy between the simulation and the actual system. Note that the use of corrective multiplicative coefficients to account for model predictive



deficiencies (safety factors) is common engineering practice (see, e.g. [43]). The vector of model-inadequacy terms  $\eta = (\eta_1, \dots, \eta_N)$  is assumed to be well represented by a correlated Gaussian model of the form:  $\eta \sim \mathcal{N}(1, K_M)$  with

$$(K_M)_{ij} = \sigma^2 \exp \left[ - \left( \frac{x_i - x_j}{10^\alpha X} \right)^2 \right], \quad 1 \leq i, j \leq N \quad (37)$$

(see [42]), where  $x_i$  and  $x_j$  are two distinct observation abscissas separated by the length scale  $10^\alpha X$ ,  $X$  being a reference length scale, here taken equal to the airfoil chord. The coefficients  $\sigma$  and  $\alpha$  are supplementary parameters intrinsic to the statistical model. They are referred-to as hyper-parameters and are calibrated along with the parameters of the physical model. Note that  $\sigma$  represents the magnitude of the model inadequacy and thus can be taken as an indicator of the accuracy of a given model. On the other hand, the model-inadequacy term is expected to reduce the risk of over-fitting the model parameters, by introducing additional degrees of freedom (see, e.g. [44] for a discussion about the importance of model-inadequacy terms in Bayesian calibration).

With the preceding assumptions, the product  $(\eta_i y_i)_{i=1, \dots, N}$  happens to follow a normal distribution with mean  $y$  and a covariance matrix defined by:

$$(K'_M)_{ij} = y_i y_j (K_M)_{ij}, \quad 1 \leq i, j \leq N \quad (38)$$

Finally, because  $(e_i)_{i=1, \dots, N}$  is also a random vector, the likelihood can be written under the form:

$$p(d|y, \theta) = \frac{1}{\sqrt{(2\pi)^N |K|}} \exp \left[ -\frac{1}{2} (d - y)^T K^{-1} (d - y) \right] \quad (39)$$

where  $K = K_e + K'_M$ , with  $K_e$  a diagonal matrix associated to the observational error vector.

From a numerical point of view, the inference is done by drawing samples from the prior *pdf* and the likelihood. For this purpose, we use a Markov-Chain Monte-Carlo sampler, based on the Metropolis-Hastings algorithm, which is well suited to represent non-classical *pdfs*. Precisely we use the implementation made available through the *pymc*<sup>1</sup> python library.

---

<sup>1</sup><http://pymc-devs.github.io/pymc/>

Typically, a large number of samples is required to converge the posterior distributions. The results presented hereafter were obtained by running a preliminary sampling of 100000 draws followed by a subsequent sampling of 200000 after burn-in. To check that the convergence is reached, we use the method proposed by Geweke[45] available in the *pymc* package. It consists of comparing the mean and the variance of segments from the beginning and the end of the Markov chain:

$$z = \frac{E[\theta_{beginning}] - E[\theta_{end}]}{\sqrt{\text{Var}[\theta_{beginning}] + \text{Var}[\theta_{end}]}} \quad (40)$$

This  $z$ -score is theoretically distributed as standard normal variates. If the chain has converged, the majority of point should fall within 2 standard deviations of zero.

For the present application, each sample corresponds to a separate run of the dense gas solver, which takes approximately 10 min of CPU time to obtain a converged numerical solution. Clearly, running sequentially several hundreds of thousands of code runs would require an impractically long turn-around time. To overcome this difficulty, we use again a surrogate model, which replaces the computationally expensive flow solver with an inexpensive approximate function. Note that the number of code runs needed to build the surrogate model (of the order of a few thousands at most) is much smaller than that required by MCMC; furthermore, their computation can be parallelized, unlike the MCMC algorithm. The surrogate model used for the calibration step is again the one described in Section 4. As a consequence, the calibration problem is modified in the following way:

$$p(\theta|d = \bar{d}, y_S) \propto p(d = \bar{d}|y_S, \theta)p(\theta). \quad (41)$$

with  $y_S$  the surrogate model. Obviously, the surrogate model introduces an additional modeling error, which is expected to be small compared to that introduced by the true model (less than 1% in average, as discussed in section 4). In any case, this additional modeling error is expected essentially to affect the posterior distributions of the hyper-parameters, and namely the modeling error magnitude  $\sigma$ .

In summary, the steps required by the calibration procedure are:

1. a surrogate model is build over the prior ranges of the model parameters; the

deterministic computations required for sampling the parameter space can be parallelized; the CPU time required for evaluating the surrogate model at a given point of the parameter space depends on the number of parameters and the number of points per direction but is approximately equal to 0.01 seconds, i.e. it is negligible compared to the CPU time required for sampling.

2. the core of the calibration procedure is the MCMC algorithm, run on the surrogate model, that provides samples of the posterior distributions for the model and hyper parameters. This part cannot be parallelized.
3. a standard Monte-Carlo algorithm, coupled to the surrogate model, is used to propagate the posteriors into the entire statistical model. The size of the sample is chosen equal to 1000000. This part can also be parallelized.

## 7. Results and discussion

Preliminary calibration runs are carried out by using piecewise polynomial surrogate models of degree four, using  $50 \times 50$  Cartesian grids of the parameter space for each EOS. The surrogate models are built over the prior physical parameter range used for the calibrations, given in Tab. 2. For the hyper-parameters associated to the multiplicative model-inadequacy term 36, we choose the following uniform prior distributions for all of the three models:

$$\sigma \sim \mathcal{U}(0.0, 0.6), \quad \alpha \sim \mathcal{U}(-4.0, 1.0). \quad (42)$$

Markov chain Monte Carlo samplings of the posteriors are used to draw scatter plots of the calibrated physical parameters.

Figs 7 to 9) display two kinds of scatter plots. Subfigures 7a, 8a and 9a represent scatter plots between parameters, whereas subfigures 7b, 8b and 9b show isocontours of the probability density of the correlations: light-gray regions correspond to low density levels and dark-gray ones to high density levels.

These scatter plots highlight a very strong correlation between  $\omega$  and  $c_{v,\infty}/R$  for the two cubic EOS, and  $T_c$  and  $p_c$  for MAH. The correlation can be easily detected from scatter plots in the case of the PRSV model; for the other two models it can be more easily deduced by inspection of the density plots.

Note that the correlated parameters were found also exhibit very strong second-order interactions in the shock-wave region, according to the sensitivity study of Section 5.2. To a first approximation, the correlations can be seen as linear.

Inspection of the marginal posterior distributions shows that the parameters are not to be well informed from the data. For instance, Fig. 15 shows a typical marginal posterior for  $p_c$ . This is almost as spread out as the non-informative prior and does not exhibit any clearly identified mode after calibration, despite the fact that the calibration algorithm clearly detects a region of high probability density for the joint posterior (see Fig. 13 to 15). Convergence difficulties are better highlighted by plots of the Geweke  $z$ -score, depicted on the left part of Figures 10 to 12. It can be seen that the  $z$ -scores associated to the correlated parameters are far from approaching the converged value of 0, or even to stay in the  $[-2; 2]$  interval.

Note that strong correlations between parameters were not expected a priori, based on a purely thermodynamic viewpoint. They are due to the propagation of thermodynamic uncertainties through the set of highly nonlinear governing equations for compressible gas dynamics. The present case study is characterized by a region of strong nonlinearity and high sensitivity, namely, the shock region. In this region, changing pairs of parameters along certain paths in the parameter spaces enables better capturing the calibration data, and specifically the shock location. Actually, the most significant discrepancies between simple models and the reference data being located in the highly nonlinear shock region, parameter correlations happen to play a fundamental role, even if they are quite negligible elsewhere. Parameters being correlated in the shock region, the calibration problem is left with a supplementary degree of freedom that prevents the random sampling method to converge to sharp posterior marginal distributions.

In order to remove the supplementary degree of freedom, we need to provide additional information. Two possible strategies are:

- to calibrate one parameter and to calculate the other as a function of the calibrated one, by means of a statistical regression of the scatter plots,
- to fix one parameter and calibrate the other one.

To avoid the introduction of additional empiricism via regression relations, we chose the second approach. However we checked that leaving both parameters free and relating them via an empirical correlation law gives similar results.

Subsequently, a new series of calibration runs is carried out by taking into account just only the most influential physical coefficient for each model, namely,  $\omega$  for RKS and PRSV and  $T_c$  for MAH. Using information from the preceding scatter plots, we chose to fix the correlated parameters  $c_{v,\infty}/R$  to a value of 300.0 and 200.0, for the RKS and the PRSV models, respectively, and  $p_c$  to a value of 16.0 for the MAH model. Thanks to the further reduction of the parameter space, very accurate surrogate models are generated by using 100 points uniformly spaced along the intervals spanned by the priors for  $\omega$  and  $T_c$ , for the cubic EOS and the virial EOS, respectively. These are the same given in Tab. 2. Priors for the hyperparameters are also left unchanged.

Marginal posterior distributions are depicted in Figures 16, 17, and 18, for RKS, PRSV, and MAH, respectively. Posterior expectancies and standard deviations of the physical and hyper-parameters are summarized in Tabs. 4a, 4b and 4c.

First of all, we notice that all of the parameters are now well informed by the calibration. Precisely, posterior ranges for the physical parameters  $\omega$  and  $T_c$  are strongly reduced with respect to the prior distributions and exhibit a well-defined mode. This is confirmed by the fact that the Geweke  $z$ -scores, depicted on the right part of Figures 10b to 12b, are now close to 0. In other terms, fixing a parameter removes the indetermination introduced by parameter correlations and helps the Markov chain to converge toward a well-defined solution for the parameter posterior marginals. Concerning the two cubic EOS, PRSV exhibits a much more peaky posterior than RKS, consistently with the spread of the scatter plots of Figs 7, 8.

Secondly, we observe that the hyper-parameter  $\sigma$ , which represents the magnitude of the model-inadequacy term  $\eta$  in our statistical model, takes relatively high values for all EOS. The lowest value of  $\sigma$  (about 17.5% of the model output  $y$ ) is obtained for the PRSV model, the highest (more than 47.0%) for MAH. This means that calibration of the physical parameters does not suffice to obtain a good fitting of the model output, and

the calibration procedure still relies on  $\sigma$  for capturing the data. Hyper-parameter  $\alpha$ , on the other hand, has a more secondary effect, since it essentially measures the level of smoothness of the model-inadequacy term. This is found to be correlated over a distance roughly equal to one hundredth of the airfoil chord for all of the models.

The posterior expectancies of the physical parameters differ significantly from the expected physical values: this shows that, despite the introduction of a model-inadequacy term, small modifications of the standard values of the physical parameters are not sufficient to correct the model output, and specifically the shock location. As a consequence, the parameters lose their underlying physical meaning and tend to become tuning parameters. A thorough discussion of the importance of the model discrepancy term for capturing physically relevant values of physical parameters is provided in [42]. Future research is warranted to explore calibration strategies involving more realistic descriptions of model discrepancy.

The role of the model-inadequacy term is better understood by inspection of Figs. 19 to 21, showing posteriors of the hyper-parameters obtained by running calibration runs with all of the physical parameters taken fixed at their nominal values. As a consequence, calibrations now rely entirely on the model-inadequacy term for capturing the data, all degrees of freedom associated to tunable closure parameters having been suppressed. As expected, posterior distributions of  $\sigma$  for all of the three models are shifted toward higher values with respect to the preceding case, with expectancies higher than 300% of the model output in the worst case (MAH EOS). This demonstrates the crucial importance of the model-inadequacy term for the present highly sensitive flow problem. Unlike to the preceding calibration, RKS model inadequacy is found to be lower than for PRSV (see Tabs 5a–5c), MAH still being the worst case.

Finally, posterior parameter distributions are propagated through the flow solver to obtain posterior estimates of the quantity of interest. Figures 22 to 24 compare solution expectancies and error bars (corresponding to  $\pm$  three times the standard deviation) for the calibrated models against the reference data and the deterministic solutions obtained by using the nominal values of the parameters. Generally speaking, calibration improves dramatically model predictions for all of the three models. Note that the three

calibrated models have to rely on the model-inadequacy term to be consistent with the experimental data, the posterior uncertainty about the parameters being smaller than the experimental one. The effect of the model inadequacy term is to enlarge numerical error bars with respect to those obtained just only by propagating a posteriori parameter distributions through the model. This effect is particularly significant at the rear part of the airfoil, where model uncertainties have a strong influence on shock location and the data is captured essentially through the model-inadequacy term. Error bars are greater for models characterized by higher values of the hyper-parameter  $\sigma$ . Specifically, for RKS and, even more, MAH, the model-inadequacy uncertainty is extremely large, denoting that these models are not reliable enough, even after calibration. On the contrary, the residual model-inadequacy uncertainty for PRSV is of the same order of the experimental one.

In summary, the preceding calibration results show that, out of the three alternative thermodynamic models under investigation, PRSV gives the more informative parameter posteriors along with the lowest model-inadequacy term contribution, even if this remains relatively important for all of the models.

### *7.1. Conclusions*

In this work we proposed and assessed a Bayesian calibration strategy for quantifying parametric uncertainties associated to thermodynamic models used in dense gas flow solvers. The objective is to use data for aerodynamic quantities, and specifically wall pressures, to inform thermodynamic data input to the code. Precisely, we calibrated the input parameters of three well-known real-gas equations of state by using reference data available for the transonic flow of a dense gas, a siloxane known as D5, over a wing section. Since no experimental data are available for the configuration of interest, and in the aim of assessing the feasibility of the proposed calibration strategy more than solving a real calibration problem, we used synthetic reference data generated by running a more complex and accurate equation of state than the three to be calibrated.

A sensitivity study based on Monte Carlo sampling of the parameter space was carried out prior to the calibration step, to identify the most influential parameters, as well as possible interactions between them. Inspection of the Sobol indices of second order shows

that strong higher-order interactions developed in the flow region characterized by the presence of a shock wave, dominated by nonlinear effects.

For the calibration step, we built a statistical model accounting both for uncertainties intrinsic to the calibration data, via an additive error term, as well as model-form uncertainties, via a multiplicative error term applied to the computational model output. Both terms were modeled as Gaussian processes with a given correlation structure. Nevertheless, given the complexity of the flow model, represented by a set of highly nonlinear partial differential equations supplemented by thermodynamic submodels, namely, equations of state, there is no possible analytical solution for the posterior probability density of the parameters, and numerical approximations are obtained by sampling the likelihood function by means of a Markov-chain Monte Carlo algorithm.

The large amount of flow solver runs required by both the sensitivity analysis and the algorithm is drastically reduced by introducing a surrogate model approximating the response of the numerical solver to variations of the input parameters. The surrogate model used in the present study is a simple polynomial model based on Gauss-Lobatto interpolation or piecewise polynomial interpolation on Cartesian grids. A preliminary accuracy study demonstrated that, for the present problem, such a simple strategy ensures an accurate enough representation of the response surface.

Calibration results showed that model parameters are well informed by the data available for pressure coefficient distributions along the airfoil wall. Nevertheless, parameter interactions also play a major role, due to strong flow nonlinearities associated to shock waves, and need to be taken into account in the calibration in order to obtain informative posteriors. The calibration step also highlighted a huge impact of the model-inadequacy term on calibration results. However, the use of a multiplicative model-error term alone is not enough, for the present highly sensitive and highly nonlinear flow problem, to prevent posterior parameter expectancies from deviating significantly from the physically expected ranges. For this reason, in future research we plan to investigate advanced strategies for representing model inadequacy. One of them is to use Bayesian model averaging for taking into account uncertainties associated to the existence of several alternative models, and improving predictive robustness (see e.g. [46]).

Future work includes the application of the present calibration strategy to more real-



istic flow configurations, like flows through turbine cascades, and the use of experimental data instead of synthetic data. Moreover, we are planning to investigate the use of aerodynamic data of a different nature, like wall temperature distributions, velocity profiles, and/or global force coefficients.

## Acknowledgements

The present work was supported by the French 'Agence Nationale de la Recherche'(ANR) under contract ANR-11-MONU-008-002.

## 8. References

- [1] Horen J., Talonpoika T., Larjola J., Siikonen T., "Numerical simulation of real-gas flow in a supersonic turbine nozzle ring", *J Eng Gas Turb Power* , 124(2), 395–403, 2002.
- [2] Monaco J., Cramer M. S., Watson L., "Supersonic flows of dense gases in cascade configurations", *J Fluid Mech* , 330, 31–59, 1997.
- [3] Brown B., Argrow B., "Application of Bethe–Zel'dovich–Thompson fluids in organic Rankine cycle engines", *J Propul Power* , 16(6), 1118-23, 2000.
- [4] Anderson K. W., "A numerical study on the use of sulfur hexafluoride as a test gas for wind tunnels", *AIAA J* , 29, 2179–2181, 1991.
- [5] Kirillov N., "Analysis of modern natural gas liquefaction technologies", *Chem Petrol Eng* , 40(7-8), 401–406, 2004.
- [6] Zamfirescu C., Dincer I., "Performance investigation of high-temperature heat pumps with various BZT working fluids", *Thermochim Acta* , 488, 66–77, 2009.
- [7] Cramer M. S., "On the Mach number variation in steady flows of dense hydrocarbons", *J Fluid Eng* , 113, 675–680, 1991.
- [8] Cramer M. S., Best L. M., "Steady isentropic flows of dense gases", *Phys Fluids A* , 3(1), 219–226, 1991.
- [9] Kluwick A., "Internal flows of dense gases", *Acta Mech* , 169, 123–143, 2004.
- [10] Bethe H. A., "The theory of shock waves for an arbitrary equation of state", Technical Paper 545, Office of Science, Research and Development, 1942.
- [11] Zel'dovich Y. B., "On the possibility of rarefaction shock waves", *Zh Exsp Teor Fiz* , 4, 363–364, 1946.
- [12] Thompson P. A., "A fundamental derivative in gas dynamics", *Phys Fluids* , 14(9), 1843–1849, 1971.
- [13] Cinnella P., Congedo P. M., Pediroda V., Parussini L., "Sensitivity analysis of dense gas flow simulations to thermodynamic uncertainties", *Phys. Fluids*, 23, 116101, 2011.

- [14] Cramer M. S., Kluwick A., “On the propagation of waves exhibiting both positive and negative nonlinearity”, *J Fluid Mech* , 142, 9–37, 1984.
- [15] Span R., Wagner W., “Equations of state for technical applications. I. Simultaneously optimized functional forms for non polar and polar fluids”, *Int. J. Thermophys* , 24, 1–39, 2003.
- [16] Soave G., “Equilibrium constants from a modified Redlich-Kwong equation of state”, *Chem. Eng. Sci.* , 27, 1197, 1972.
- [17] Stryjek R., Vera J. H., “An improved Peng-Robinson equation of state for pure compounds and mixtures” , *Can. J. Chem. Eng.* , 64, 323, 1986.
- [18] Martin J. J., Hou Y. C., “Development of an equation of state for gases”, *AIChE J.* , 1, 142, 1955.
- [19] Benedict M., Webb G. B., Rubin L. C., “An Empirical Equation for Thermodynamic Properties of Light Hydrocarbons and Their Mixtures: I. Methane, Ethane, Propane, and n-Butane”, *J Chem Phys* , 8 (4), 334–345, 1940.
- [20] Hajipour S., Satyro M. A., “Uncertainty analysis applied to thermodynamic models and process design –I. Pure components”, *Fluid Phase Equilib* , 307, 78–94, 2011.
- [21] Robinson A. C., Berry R. D., Carpenter J. H., Debusschere B., Drake R. R., Mattson A. E., Rider W. J., “Fundamental issues in the representation and propagation of uncertain equation of state information in shock hydrodynamics”, *Comp Fluids* , 83, 187–193, 2013.
- [22] Cheung S. H., Oliver T. H., Prudencio E. E., Prudhomme S., Moser R. D., “Bayesian uncertainty analysis with applications to turbulence modeling”, *Reliability Engineering and System Safety* , 96, 1137–1149, 2011.
- [23] Edeling W. N., Cinnella P., Dwight R., Bijl H., “Bayesian estimates of parameter variability in the  $k - \epsilon$  turbulence model”, *J Comp Phys* , 258, 73–94, 2014.
- [24] Colonna P., Nannan N. R., Guardone A., Lemmon E. W., “Multiparameter equations of state for selected siloxanes”, *Fluid Phase Equilib* , 244, 193–211, 2006.
- [25] Cinnella P., Congedo P. M., “Numerical solver for dense gas flows”, *AIAA J* , 43(11), 2458–2461, 2005.
- [26] Cramer M. S., Tarkenton G. M., “Transonic flows of Bethe–Zel’dovich–Thompson fluids”, *J Fluid Mech* , 240, 197–298, 1992.
- [27] Cinnella P., Congedo P. M., “Optimal airfoil shapes for viscous transonic flows of Bethe–Zel’dovich–Thompson fluids”, *Computers & Fluids* , 37(3), 250–264, 2008.
- [28] Borisov A. A., Kutateladze S. S., Nakaryakov V. E., “Rarefaction shock waves near the critic liquid-vapour point”, *J Fluid Mech* , 126, 59, 1983.
- [29] Fergason S. H., Argrow B. M., Emanuel G., “Theory for producing a single-phase rarefaction shock-wave in a shock tube”, *J Fluid Mech* , 445, 37, 2001.
- [30] Colonna P., Guardone A., Nannan N. R., Zamfirescu C., “Design of the dense gas flexible asymmetric shock tube”, *J Fluid Eng* , 130, 034501, 2008.
- [31] Spinelli A., Pini M., Dossena V., Gaetani P., Casella F., “Design, simulation, and construction of a Test Rig for Organic Vapors”, *J of Eng For Gas Turb And Power* , 135, 042303, 2013.
- [32] Metropolis N., Ulam S., “The Monte Carlo method”, *J Am Statist Assoc* , 44, 335–341, 1949.

- [33] Hammersley J. M., Handscom D. C., “Monte Carlo methods”, Vol. 1, Springer, 1964.
- [34] Forrester A. I., Sobester A., and Keane A. J., “Engineering design via surrogate modeling: a practical guide”, John Wiley & Sons, 2008.
- [35] Forrester A. I., KEane A. J., “Recent advances in surrogate-based optimization”, *Prog Aerosp Sci* , 45, 50, 2009.
- [36] Queipo N. V., Haftka R. T., Shyy W., Goel T., Vaidyanathan R., Kevin Tucker P., “Surrogate-based analysis and optimization”, *Prog Aerosp Sci* , 41, 1, 2005.
- [37] Marzouk Y. M., Najm H. N., “Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems”, *J Comp Phys* , 228, 1862, 2009.
- [38] Edeling W. N., Dwight R., Cinnella P., “Adaptive stochastic collocation methods for high-dimensional uncertainty quantification problems”, Workshop on uncertainty quantification in Computational Fluid Dynamics, May 26-27, 2014, Pisa, Italy.
- [39] Sobol’ I. M., “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates”, *Mathematics and Computers in Simulation* , 55, 271–280, 2001.
- [40] Homma T., Saltelli A. “Importance measures in global sensitivity analysis of non linear models”, *Reliability Engineering and System Safety* , 52, 1–17, 1996.
- [41] Jacques J. “Contributions à l’analyse de sensibilité et à l’analyse discriminante généralisée”, *PhD Thesis* , Université Joseph Fourier, 2005.
- [42] Kennedy M. C., O’Hagan A., “Bayesian calibration of computer models”, *J. R. Stat. Soc. Ser. B Stat. Methodol.* , 63, 425–464, 2001.
- [43] Ullman D. G., “The mechanical design process, 2nd ed.”, McGraw-Hill, 1992.
- [44] Brynjarsdottir J., O’Hagan A., “Learning about physical parameters: The importance of model discrepancy.”, *Inverse Probl.* , to appear, 2014.
- [45] Geweke J. “Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments”, pp. 169–193. *In Bernardo, Berger, Dawid, and Smith (1992)*. .
- [46] Edeling W., Cinnella P., Dwight R., “Predictive RANS simulations via Bayesian Model-Scenario Averaging”, *J Comput Phys* , 275, 65–91, 2014.

## 9. Appendix

### 9.1. Sobol indices

#### 9.1.1. RKS model

$x$	$(c_{v,\infty})$	$(\omega)$	$(n)$	$(c_{v,\infty},\omega)$	$(c_{v,\infty},n)$	$(\omega,n)$	$(c_{v,\infty},\omega,n)$
1.0049	0.7670	0.0755	0.0010	0.1570	-0.0010	-0.0010	0.0012
$9.3519 \cdot 10^{-1}$	0.3423	0.2414	-0.0013	0.4159	0.0015	0.0014	-0.0013
$8.5879 \cdot 10^{-1}$	0.8229	0.1781	0.0089	-0.0009	-0.0089	-0.0090	0.0089
$7.7770 \cdot 10^{-1}$	0.8550	0.1538	0.0110	-0.0089	-0.0110	-0.0111	0.0111
$6.9139 \cdot 10^{-1}$	0.8546	0.1543	0.0108	-0.0089	-0.0109	-0.0109	0.0109
$5.9969 \cdot 10^{-1}$	0.8535	0.1554	0.0105	-0.0089	-0.0105	-0.0106	0.0106
$5.0220 \cdot 10^{-1}$	0.8525	0.1566	0.0103	-0.0091	-0.0104	-0.0105	0.0105
$3.9849 \cdot 10^{-1}$	0.8505	0.1585	0.0099	-0.0092	-0.0099	-0.0101	0.0101
$2.8830 \cdot 10^{-1}$	0.8475	0.1615	0.0093	-0.0090	-0.0093	-0.0094	0.0094
$1.8739 \cdot 10^{-1}$	0.8417	0.1668	0.0083	-0.0085	-0.0083	-0.0084	0.0084
$1.1800 \cdot 10^{-1}$	0.8323	0.1745	0.0070	-0.0068	-0.0070	-0.0070	0.0070
$7.1130 \cdot 10^{-2}$	0.8148	0.1865	0.0050	-0.0014	-0.0050	-0.0050	0.0050
$3.9689 \cdot 10^{-2}$	0.7558	0.2162	0.0008	0.0277	-0.0007	-0.0007	0.0007
$1.9039 \cdot 10^{-2}$	0.8643	0.1536	0.0337	-0.0182	-0.0334	-0.0342	0.0342
$6.2130 \cdot 10^{-3}$	0.8133	0.1985	0.0143	-0.0116	-0.0145	-0.0146	0.0146
$-4.9549 \cdot 10^{-4}$	0.8101	0.2057	0.0187	-0.0153	-0.0188	-0.0191	0.0187
$-2.4699 \cdot 10^{-3}$	0.8150	0.2087	0.0248	-0.0228	-0.0247	-0.0260	0.0251

9.1.2. PRSV model

$x$	$(c_{v,\infty})$	$(\omega)$	$(n)$	$(c_{v,\infty},\omega)$	$(c_{v,\infty},n)$	$(\omega,n)$	$(c_{v,\infty},\omega,n)$
1.0049	0.5736	0.3242	-0.0571	0.1009	0.0576	0.0585	-0.0578
$9.3519 \cdot 10^{-1}$	0.2785	0.1218	-0.0005	0.5986	0.0010	0.0009	-0.0004
$8.5879 \cdot 10^{-1}$	0.5912	0.2685	0.0006	0.1399	-0.0006	-0.0006	0.0006
$7.7770 \cdot 10^{-1}$	0.1195	0.1112	-0.0091	0.7682	0.0098	0.0098	-0.0094
$6.9139 \cdot 10^{-1}$	0.6098	0.3488	-0.0332	0.0409	0.0335	0.0338	-0.0338
$5.9969 \cdot 10^{-1}$	0.6065	0.3592	-0.0313	0.0337	0.0319	0.0318	-0.0320
$5.0220 \cdot 10^{-1}$	0.6063	0.3609	-0.0299	0.0321	0.0305	0.0305	-0.0305
$3.9849 \cdot 10^{-1}$	0.6058	0.3635	-0.0282	0.0299	0.0286	0.0286	-0.0285
$2.8830 \cdot 10^{-1}$	0.6051	0.3682	-0.0250	0.0260	0.0255	0.0255	-0.0256
$1.8739 \cdot 10^{-1}$	0.6032	0.3751	-0.0206	0.0212	0.0210	0.0211	-0.0211
$1.1800 \cdot 10^{-1}$	0.5992	0.3839	-0.0157	0.0164	0.0160	0.0161	-0.0161
$7.1130 \cdot 10^{-2}$	0.5914	0.3947	-0.0097	0.0134	0.0100	0.0100	-0.0100
$3.9689 \cdot 10^{-2}$	0.5724	0.4087	-0.0005	0.0186	0.0007	0.0007	-0.0007
$1.9039 \cdot 10^{-2}$	0.4888	0.1940	-0.1752	0.3142	0.1778	0.1795	-0.1792
$6.2130 \cdot 10^{-3}$	0.5466	0.4094	-0.0215	0.0434	0.0221	0.0219	-0.0221
$-4.9549 \cdot 10^{-4}$	0.5466	0.4023	-0.0432	0.0505	0.0437	0.0442	-0.0443
$-2.4699 \cdot 10^{-3}$	0.5345	0.3990	-0.0581	0.0648	0.0597	0.0595	-0.0594

9.1.3. MAH model

$x$	$(T_c)$	$(p_c)$	$(c_{v,\infty})$	$(T_c, p_c)$	$(T_c, c_{v,\infty})$	$(p_c, c_{v,\infty})$	$(T_c, p_c, c_{v,\infty})$
1.0049	0.6961	0.0473	0.0029	0.2246	0.0074	0.0018	0.0196
$9.3519 \cdot 10^{-1}$	0.9527	0.0036	0.0003	0.0273	0.0015	0.0014	0.0129
$8.5879 \cdot 10^{-1}$	0.9079	0.0074	0.0024	0.0409	0.0099	0.0075	0.0236
$7.7770 \cdot 10^{-1}$	0.6217	0.0592	0.0021	0.2627	0.0168	0.0100	0.0270
$6.9139 \cdot 10^{-1}$	0.9210	0.0071	-0.0020	0.0350	0.0064	0.0110	0.0214
$5.9969 \cdot 10^{-1}$	0.8168	0.0112	0.0004	0.0914	0.0300	0.0150	0.0348
$5.0220 \cdot 10^{-1}$	0.7663	0.0154	0.0048	0.1463	0.0185	0.0085	0.0398
$3.9849 \cdot 10^{-1}$	0.7060	0.0149	0.0126	0.1832	0.0602	-0.0006	0.0234
$2.8830 \cdot 10^{-1}$	0.9337	0.0126	0.0030	0.0309	0.0114	-0.0023	0.0104
$1.8739 \cdot 10^{-1}$	0.9653	0.0254	0.0077	0.0024	-0.0014	-0.0023	0.0028
$1.1800 \cdot 10^{-1}$	0.9623	0.0272	0.0076	0.0025	4.6661	-0.0013	0.0014
$7.1130 \cdot 10^{-2}$	0.9516	0.0322	0.0082	0.0055	0.0021	-0.0004	0.0005
$3.9689 \cdot 10^{-2}$	0.9223	0.0430	0.0105	0.0182	0.0057	-2.9407	0.0002
$1.9039 \cdot 10^{-2}$	0.9420	0.0444	0.0258	0.0073	-0.0204	-0.0225	0.0234
$6.2130 \cdot 10^{-3}$	0.9509	0.0412	0.0030	0.0050	-0.0001	-0.0007	0.0005
$-4.9549 \cdot 10^{-4}$	0.9488	0.0415	0.0028	0.0066	-0.0003	-0.0013	0.0018
$-2.4699 \cdot 10^{-3}$	0.9493	0.0410	0.0039	0.0070	-0.0011	-0.0024	0.0021

## 10. Figures

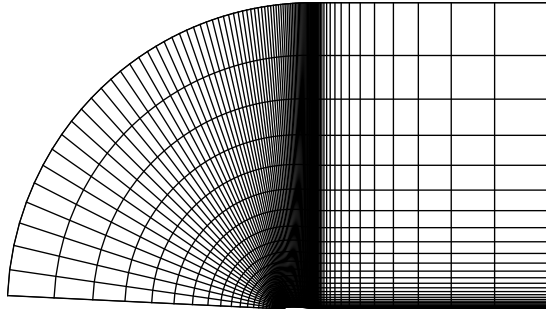


Figure 1: Computational grid.

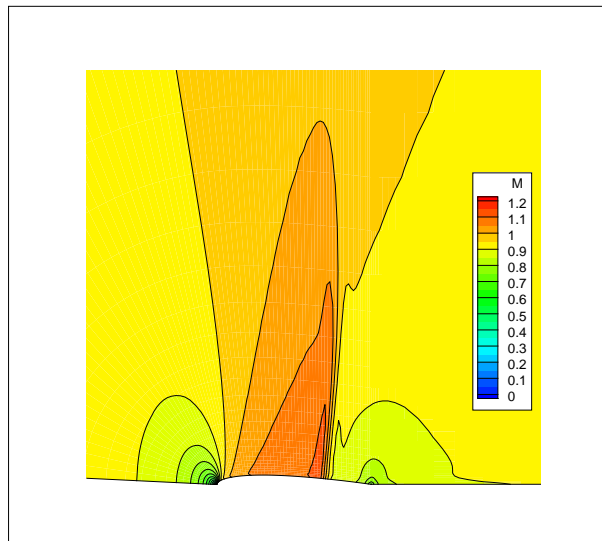


Figure 2: Iso-contours of Mach number.

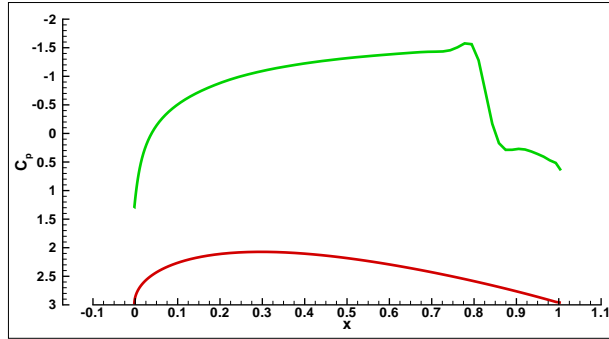


Figure 3: Pressure coefficient all along the profile.

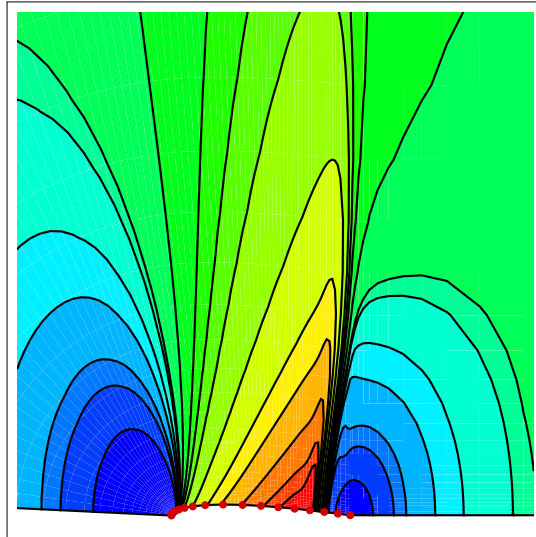


Figure 4: Iso-contours of pressure coefficient. •: numerical pressure sensors.



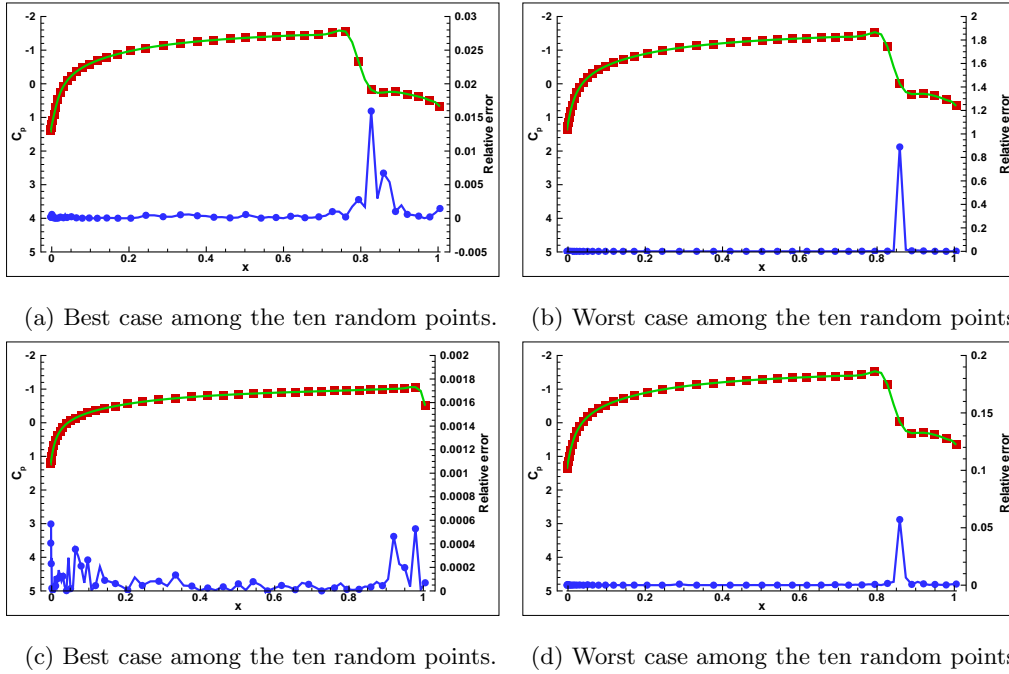
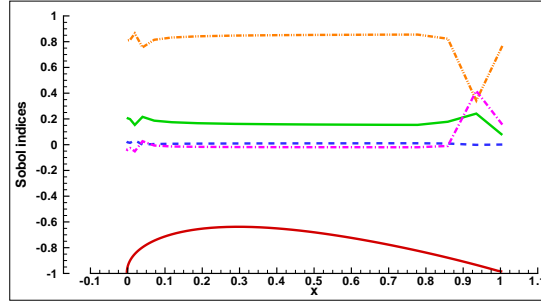
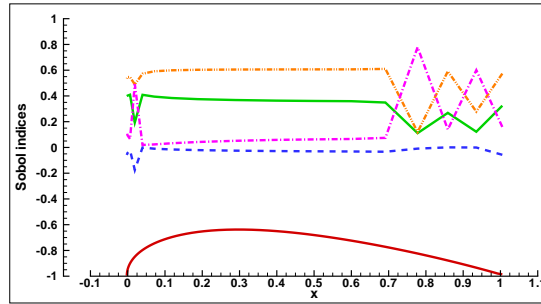


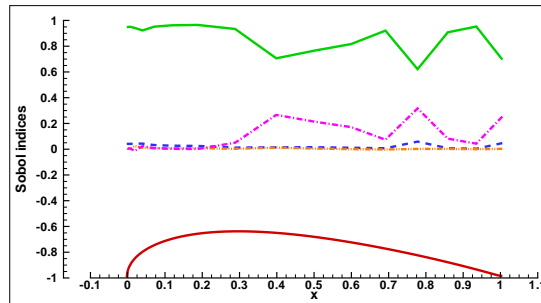
Figure 5: Relative error (—●—) between the exact solution (■) and the surrogate model (—) in the PRSV case. The top figures correspond to a three dimensional surrogate model with a Gauss-Lobatto grid whereas the bottom part correspond to a two dimensional surrogate model with a Cartesian mesh.



(a) RKS:  $\omega$  (—),  $n$  (- - -),  $c_{v,\infty}$  (- · - · -), high order Sobol indices (- · - · -).

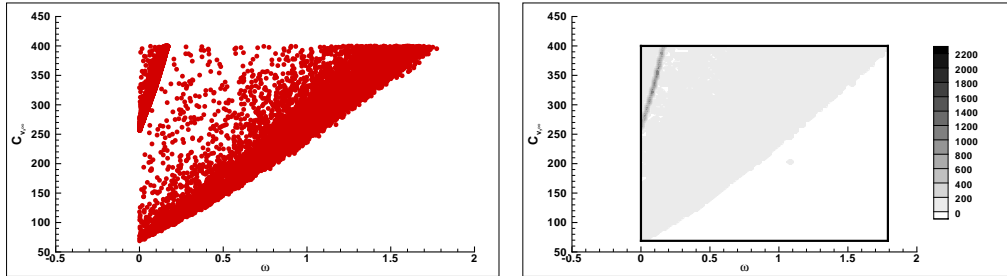


(b) PRSV:  $\omega$  (—),  $n$  (- - -),  $c_{v,\infty}$  (- · - · -), high order Sobol indices (- · - · -).



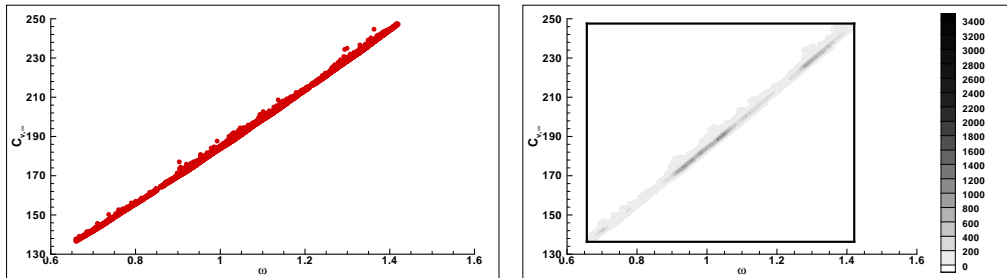
(c) MAH:  $T_c$  (—),  $p_c$  (- - -),  $c_{v,\infty}$  (- · - · -), high order Sobol indices (- · - · -).

Figure 6: Distributions of the Sobol indices along the airfoil wall.



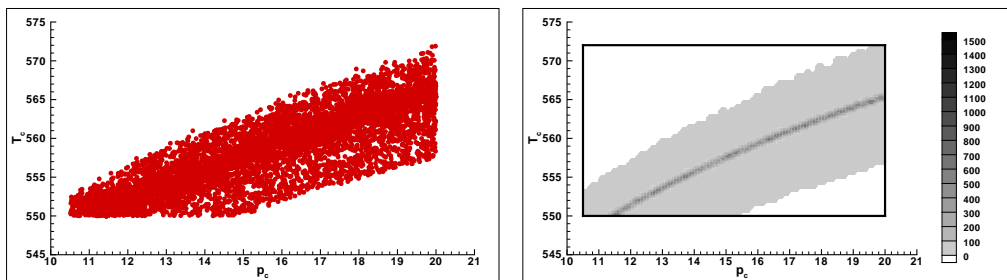
(a) Simple correlations between  $c_{v,\infty}$  and  $\omega$ . (b) Density of correlations between  $c_{v,\infty}$  and  $\omega$ .

Figure 7: Scatter plot for  $c_{v,\infty}$  and  $\omega$ , RKS model.



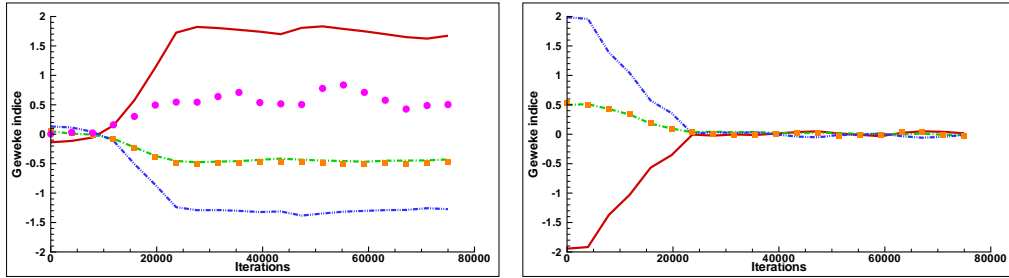
(a) Simple correlations between  $c_{v,\infty}$  and  $\omega$ . (b) Density of correlations between  $c_{v,\infty}$  and  $\omega$ .

Figure 8: Scatter plot for  $c_{v,\infty}$  and  $\omega$ , PRSV model.



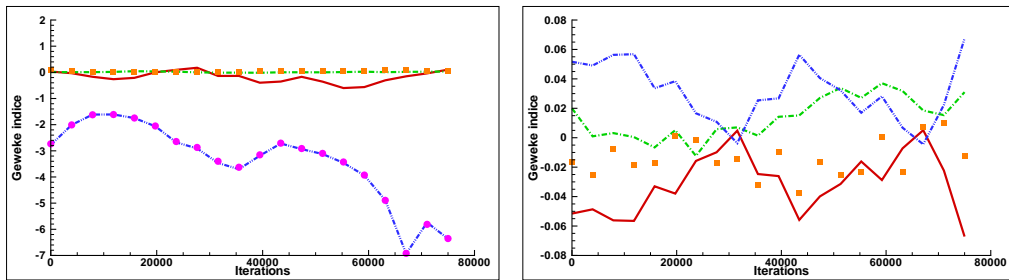
(a) Simple correlations between  $T_c$  and  $p_c$ . (b) Density of correlations between  $T_c$  and  $p_c$ .

Figure 9: Scatter plot for  $T_c$  and  $p_c$ , MAH model.



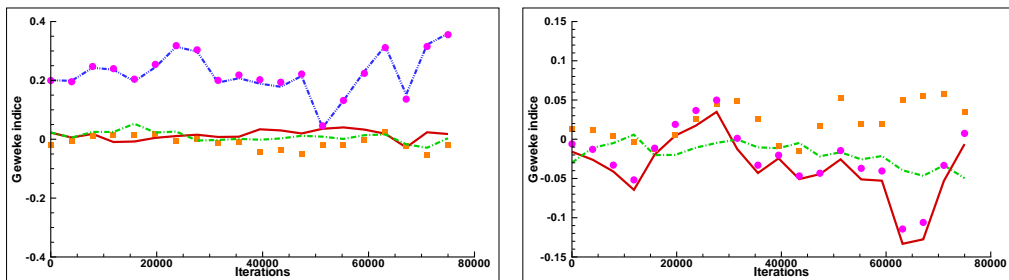
(a) Two parameters calibration:  $c_{v,\infty}$  ( $\bullet$ ),  $\omega$  ( $-\cdot-\cdot-$ ),  $y$  ( $-$ ),  $\sigma$  ( $\blacksquare$ ),  $\alpha$  ( $- \cdot -$ ). (b) Single parameter calibration:  $\omega$  ( $-\cdot-\cdot-$ ),  $y$  ( $-$ ),  $\sigma$  ( $\blacksquare$ ),  $\alpha$  ( $- \cdot -$ ).

Figure 10: Geweke  $z$ -scores for RKS model.



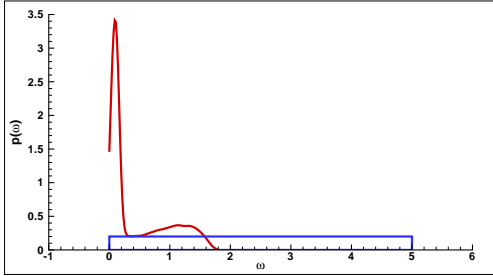
(a) Two parameters calibration:  $c_{v,\infty}$  ( $\bullet$ ),  $\omega$  ( $-\cdot-\cdot-$ ),  $y$  ( $-$ ),  $\sigma$  ( $\blacksquare$ ),  $\alpha$  ( $- \cdot -$ ). (b) Single parameter calibration:  $\omega$  ( $-\cdot-\cdot-$ ),  $y$  ( $-$ ),  $\sigma$  ( $\blacksquare$ ),  $\alpha$  ( $- \cdot -$ ).

Figure 11: Geweke  $z$ -scores for PRSV model.

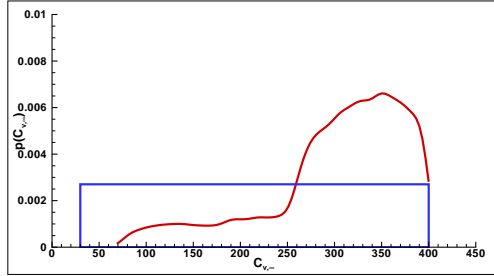


(a) Two parameters calibration:  $T_c$  ( $\bullet$ ),  $p_c$  ( $-\cdot-\cdot-$ ),  $y$  ( $-$ ),  $\sigma$  ( $\blacksquare$ ),  $\alpha$  ( $- \cdot -$ ). (b) Single parameter calibration:  $T_c$  ( $\bullet$ ),  $y$  ( $-$ ),  $\sigma$  ( $\blacksquare$ ),  $\alpha$  ( $- \cdot -$ ).

Figure 12: Geweke  $z$ -scores for MAH model.

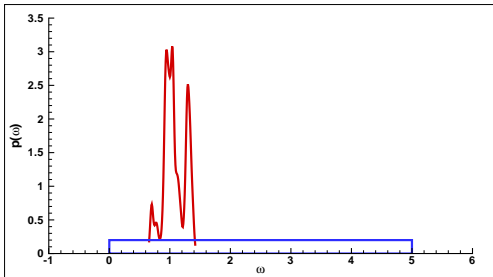


(a) Posterior distribution for  $\omega$ .

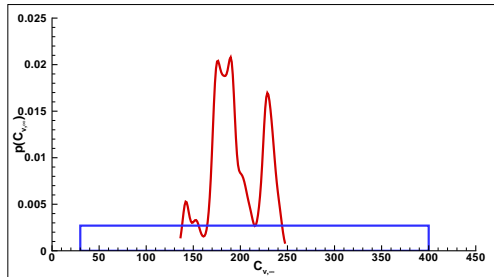


(b) Posterior distribution for  $c_{v,\infty}$ .

Figure 13: RKS model - Posterior distributions for the calibrated parameters. (—), prior (—).

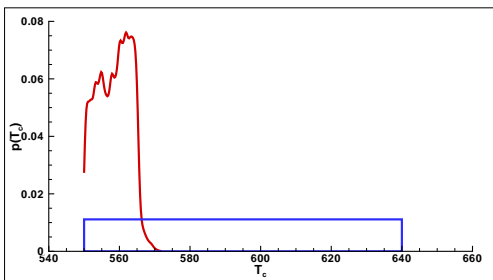


(a) Posterior distribution for  $\omega$ .

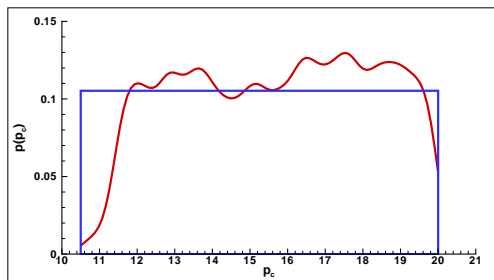


(b) Posterior distribution for  $c_{v,\infty}$ .

Figure 14: PRSV model - Posterior distributions for the calibrated parameters. (—), prior (—).

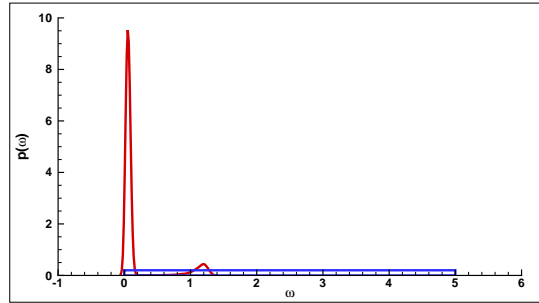


(a) Posterior distribution for  $T_c$ .

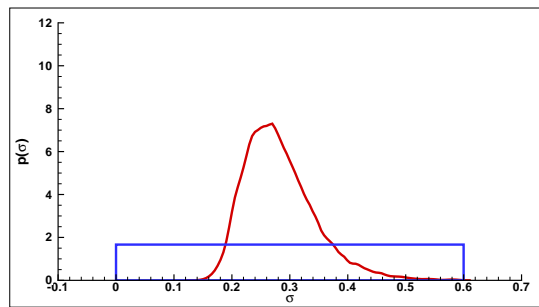


(b) Posterior distribution for  $p_c$ .

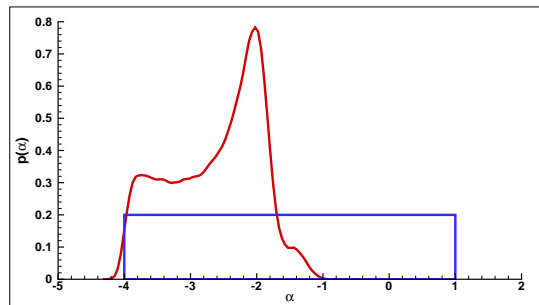
Figure 15: MAH model - Posterior distributions for the calibrated parameters. (—), prior (—).



(a) Posterior distribution for  $\omega$ .

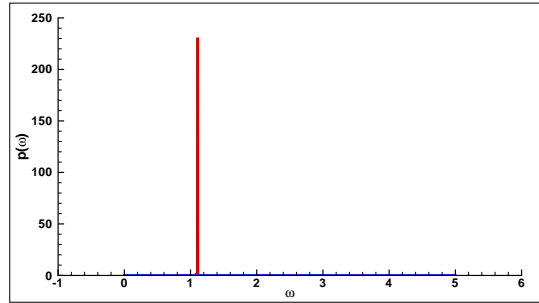


(b) Posterior distribution for  $\sigma$ .

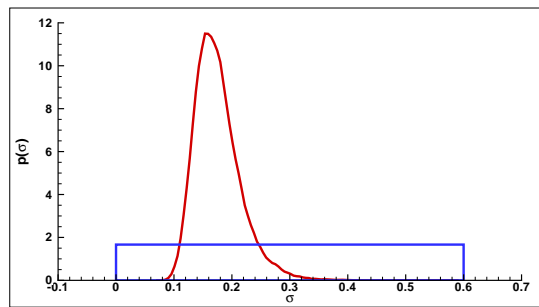


(c) Posterior distribution for  $\alpha$ .

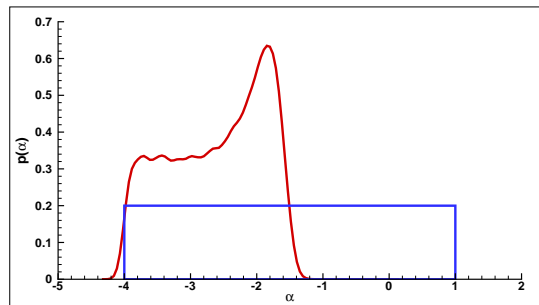
Figure 16: RKS model - Posterior distributions for the calibrated parameters. (—), prior (—).



(a) Posterior distribution for  $\omega$ .

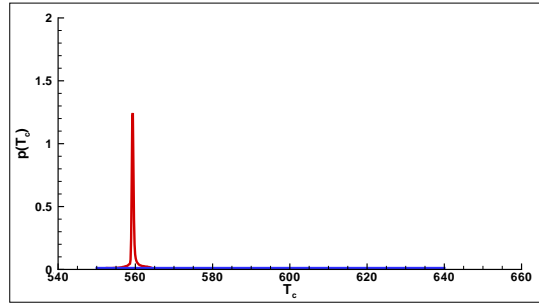


(b) Posterior distribution for  $\sigma$ .

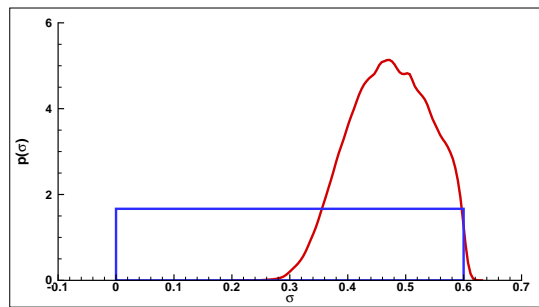


(c) Posterior distribution for  $\alpha$ .

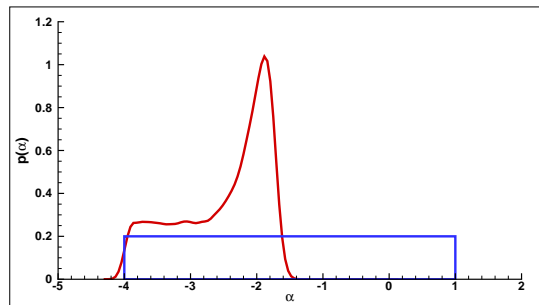
Figure 17: PRSV model - Posterior distributions for the calibrated parameters. (—), prior (—).



(a) Posterior distribution for  $T_c$ .



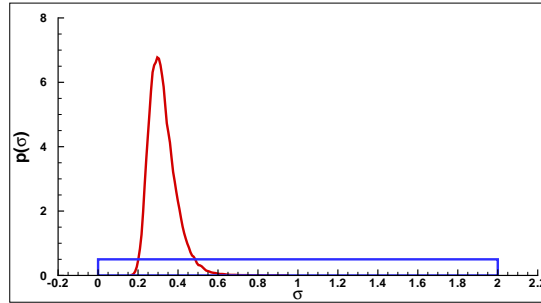
(b) Posterior distribution for  $\sigma$ .



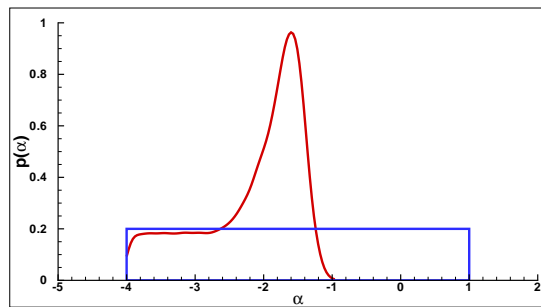
(c) Posterior distribution for  $\alpha$ .

Figure 18: MAH model - Posterior distributions for the calibrated parameters. (—), prior (—).



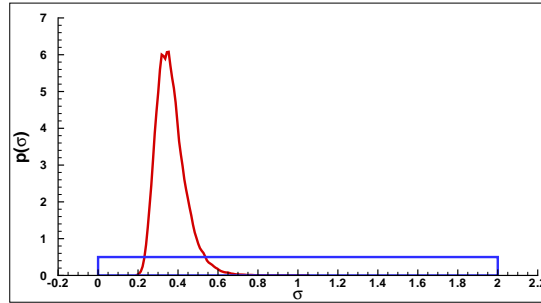


(a) Posterior distribution for  $\sigma$ .

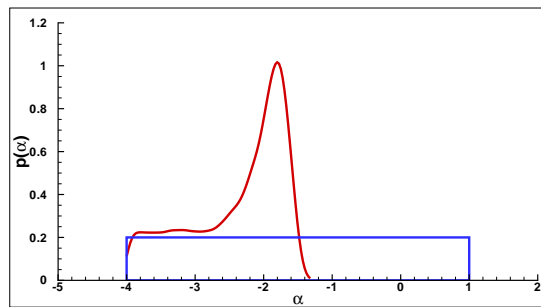


(b) Posterior distribution for  $\alpha$ .

Figure 19: RKS model - posterior distributions of  $\sigma$  and  $\alpha$  with fixed physical parameters: posterior (—), prior (—).

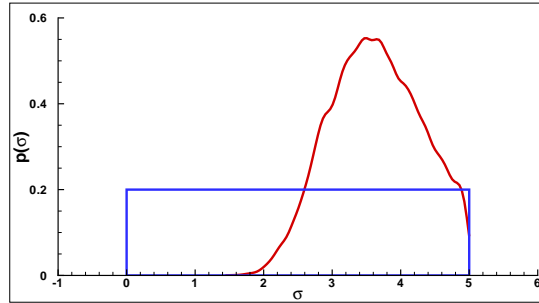


(a) Posterior distribution for  $\sigma$ .

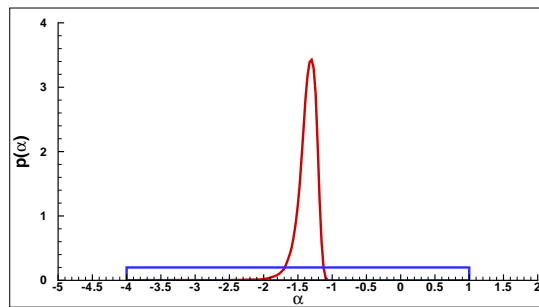


(b) Posterior distribution for  $\alpha$ .

Figure 20: PRSV model - posterior distributions of  $\sigma$  and  $\alpha$  with fixed physical parameters: posterior (—), prior (—).



(a) Posterior distribution for  $\sigma$ .



(b) Posterior distribution for  $\alpha$ .

Figure 21: MAH model - posterior distributions of  $\sigma$  and  $\alpha$  with fixed physical parameters: posterior (—), prior (—).

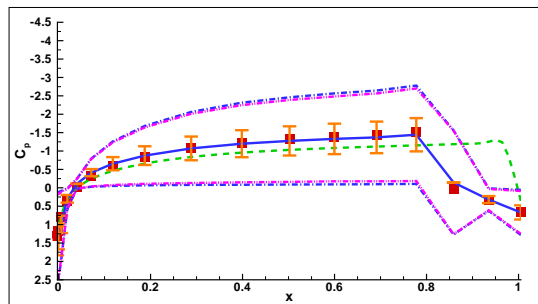


Figure 22: Redlich-Kwong -  $C_p$  vs  $x$ . Solutions: reference (■), nominal (— — —), a posteriori (—). Error bars (three times the standard deviation): experimental (—). Error ranges (three times the standard deviation): parametric + model-inadequacy (— — —), model (— · — ·).

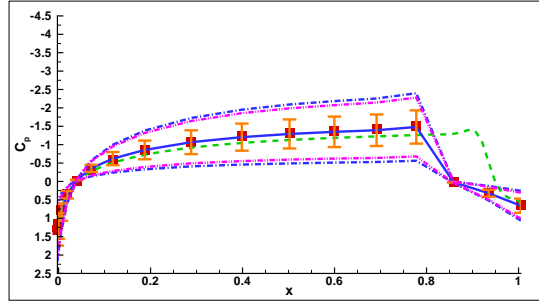


Figure 23: Peng-Robinson -  $C_p$  vs  $x$ . Solutions: reference (■), nominal (---), a posteriori (—). Error bars (three times the standard deviation): experimental (—). Error ranges (three times the standard deviation): parametric + model-inadequacy (---), model (—).

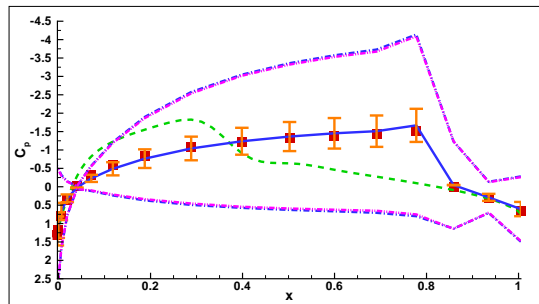


Figure 24: Martin-Hou -  $C_p$  vs  $x$ . Solutions: reference (■), nominal (---), a posteriori (—). Error bars (three times the standard deviation): experimental (—). Error ranges (three times the standard deviation): parametric + model-inadequacy (---), model (—).

## 11. Tables

$T_c$ (K)	617.4
$p_c$ (atm)	10.2
$Z_c$ —	0.237
$T_e$ (K)	484.1
$\omega$	0.6658
$c_{\nu,\infty}(T_c)/R$	76
$n$	0.5208

Table 1: Nominal values of thermodynamic properties and parameters for D5.

	RKS	PRSV	MAH
$T_c$	—	—	$\mathcal{U}(550, 640)$
$p_c$	—	—	$\mathcal{U}(10.5, 20)$
$c_{\nu,\infty}(T_c)$	$\mathcal{U}(30, 400)$	$\mathcal{U}(30, 400)$	$\mathcal{U}(75, 80)$
$\omega$	$\mathcal{U}(0, 5)$	$\mathcal{U}(0, 5)$	—
$n$	$\mathcal{U}(0, 1)$	$\mathcal{U}(0, 1)$	—

Table 2: Prior distributions of physical parameters for the three EOS under investigation.

	Strong influence on $C_p$	Weak influence on $C_p$
RKS	$\omega, c_{\nu,\infty}(T_c)$	$n$
PRSV	$\omega, c_{\nu,\infty}(T_c)$	$n$
MAH	$T_c, p_c$	$c_{\nu,\infty}(T_c)$

Table 3: Summary of the Sobol analysis for each EOS.

	Mean	Standard deviation
$\omega$	0.16940	0.33381
$\sigma$	0.28270	0.06128
$\alpha$	-2.61809	0.69334

(a) RKS model

	Mean	Standard deviation
$\omega$	1.10834	0.00244
$\sigma$	0.17540	0.03903
$\alpha$	-2.61000	0.72742

(b) PRSV model

	Mean	Standard deviation
$T_c$	559.28678	1.64511
$\sigma$	0.47193	0.06809
$\alpha$	-2.51034	0.68401

(c) MAH model

Table 4: Posterior parameter statistics.

	Mean	Standard deviation
$\sigma$	0.32378	0.06911
$\alpha$	-2.17570	0.74721

(a) RKS model.

	Mean	Standard deviation
$\sigma$	0.36567	0.07473
$\alpha$	-2.38574	0.70160

(b) PRSV model.

	Mean	Standard deviation
$\sigma$	3.63605	0.66215
$\alpha$	-1.37589	0.19883

(c) MAH model.

Table 5: Hyperparameter statistics with fixed physical parameters.