Calibrating hypersonic turbulence flow models with the HIFiRE-1 experiment using data-driven machine-learned models

Kenny Chowdhary[*]   Chi Hoang[*]   Kookjin Lee[‡]   Jaideep Ray[*]

V. G. Weirs[§]   Brian Carnes[¶]

Abstract

In this paper we study the efficacy of combining machine-learning methods with projection-based model reduction techniques for creating data-driven surrogate models of computationally expensive, high-fidelity physics models. Such surrogate models are essential for many-query applications e.g., engineering design optimization and parameter estimation, where it is necessary to invoke the high-fidelity model sequentially, many times. Surrogate models are usually constructed for individual scalar quantities. However there are scenarios where a spatially varying field needs to be modeled as a function of the model's input parameters. We develop a method to do so, using projections to represent spatial variability while a machine-learned model captures the dependence of the model's response on the inputs. The method is demonstrated on modeling the heat flux and pressure on the surface of the HIFiRE-1 geometry in a Mach 7.16 turbulent flow. The surrogate model is then used to perform Bayesian estimation of freestream conditions and parameters of the SST (Shear Stress Transport) turbulence model embedded in the high-fidelity (Reynolds-Averaged Navier-Stokes) flow simulator, using shock-tunnel data. The paper provides the first-ever Bayesian calibration of a turbulence model for complex hypersonic turbulent flows. We find that the primary issues in estimating the SST model parameters are the limited information content of the heat flux and pressure measurements and the large model-form error encountered in a certain part of the flow.

[*]Extreme-scale Data Science and Analytics Department, Sandia National Laboratories, Livermore, CA 94550 (kchowdh@sandia.gov, ckhoang@sandia.gov, jairay@sandia.gov). Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

[‡]School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ 85281 (kookjin.lee@asu.edu)

[§]Computational Sciences Department, Sandia National Laboratories, Albuquerque, NM 87185-1320 (vgweirs@sandia.gov)

[¶]V&V, UQ and Credibility Process Department, Sandia National Laboratories, Albuquerque, NM 87185-0828 (bcarnes@sandia.gov)

1 Introduction

In this paper, we develop and test a method to create surrogate models that can approximate spatially varying responses (i.e., fields) generated by a high-fidelity computational model (usually a system of partial differential equations) e.g., an engineering simulator. The method will be demonstrated in the context of hypersonic turbulent flow solutions over a realistic engineering geometry with shocks, boundary layers, flow separation, and reattachment. Surrogate models are essential for many-query applications e.g., design optimization or inverse problems where the computational model has to be invoked repeatedly and sequentially, and for the prediction of quantities of interest (henceforth QoI), as a function of the computational model's inputs. In this paper, we will demonstrate the surrogate models to calibrate a turbulence model using data from a hypersonic shock-tunnel experiment, a process that will require us to simulate a turbulent flow many times. Models for predicting a single scalar variable have long existed (see Ref. [1, 2, 3] for reviews), as well as for scalar-valued fields (see Ref. [4] for a review). A review of surrogate modeling for aerodynamic applications can be found in Ref. [5].

Swischuk et al. [6] describe an alternative way of surrogate modeling fields, where they use some basic knowledge of the behavior of the fields in question to significantly simplify the architecture (and therefore the training) of the surrogate model. They realized that the spatial correlation in the fields persist and do not vary erratically as the model inputs change, and thus the modeling could admit a "separation-of-variables" approach. They modeled the spatial variation of the field using a basis set obtained by the proper orthogonal decomposition (POD) of a training dataset of fields, and captured the dependence of the weights/coefficients of the bases on the model inputs via machine-learning (ML) techniques. They found that simple ML methods such as polynomial regressions were equal to, or better than, complex methods (such as neural networks), which simplified the training of the models, and reduced requirements on the size of the training dataset (TD). In this foundational study, the problems considered were "idealized" - the fields were smooth, their dependence on model parameters relatively benign (though nonlinear) and the number of model input parameters less than half-a-dozen.

In this paper, we investigate whether a difficult and realistic engineering problem can be addressed using the "separation-of-variables" approach to surrogate modeling, and whether the usefulness of simple ML techniques still holds. We seek to construct surrogate models for the heat-flux and pressure fields on the HIFiRE-1 geometry when placed in a $M_\infty = 7.16$ hypersonic flow in the LENS-I shock-tunnel (see Ref. [7, 8] for a description of the experiment and modeling effort). In our study, the high-fidelity engineering simulator is a Reynolds-Averaged Navier–Stokes (RANS) model [9] with Menter's SST turbulence model [10] embedded in it. Each simulation takes about 384 CPU-hours[1] to converge to steady state. The QoIs (heat-flux and pressure) are obtained on the surface grid of the HIFiRE-1 geometry with 2170 grid points (1085 × 2 surface mesh with details in Sec. 3). The RANS model has 12 uncertain parameters - freestream density, temperature and velocity, as well as 9 SST turbulence model parameters - which form the input vector of the surrogate model. The hypersonic flow contains discontinuities (shock waves), regions of intense gradients (turbu-

---

[1]Each simulation is run using 128 2.3 GHz Intel Xeon Gold processors (4 nodes, each with 32 cores) for approximately 3 hours using Sandia's high performance computing resources.

lent boundary layers) and a flow separation zone on the HIFiRE-1 geometry. It is expected to pose a realistic challenge for surrogate modeling.

The usefulness of the surrogate model will be demonstrated by calibrating the 12 uncertain parameters to shock-tunnel data (heat-flux and pressure measurements on the HIFiRE-1 surface) The calibration will be Bayesian i.e., we will develop a 12-dimensional joint probability density function (JPDF) over the uncertain parameters to capture the calibrated values, as well as the uncertainty in them, due to a finite number of noisy measurements and the shotcomings of the RANS model (i.e., model-form error).

This paper has two main contributions. Firstly, we provide the first comprehensive and comparative study of data-driven ROM-based surrogate model construction for a complex realistic engineering application in hypersonic flows. In our case, the bulk of the complexity is expected to arise from the ML models that represent the influence of the freestream values and the nonlinearities engendered by the SST turbulence model. Surrogate models that leverage knowledge of the physical phenomena to simplify their architecture, training and TD requirements have their obvious attractions, especially when computationally expensive models have to be run thousands of time to populate a high-dimensional parameter space.

Our second contribution is the illustration of the potential of surrogate models in solving inverse problems, in our case, the SST model calibrated to measurements from a 2D hypersonic flow experiment. RANS models are approximate and often need to be calibrated to experimental measurements from flows similar to their final use-case (e.g., in hypersonic flows) to be predictive. To date, turbulence models calibrated to shock-tunnel data have been limited to low-speed flows (incompressible [11] and transonic [12]), and it is unclear whether the approximations inherent in RANS will even allow the estimation of SST parameters with any degree of accuracy, given separated hypersonic flows over the HIFiRE-1 geometry. While there have been attempts to calibrate turbulence models in hypersonic flows [13], they are limited to 1D (flat-plate boundary layer) problems where many turbulent processes are absent. In contrast, we provide a methodology, heavily reliant on surrogate modeling, that can be used to calibrate models with data obtained from experiments that closely resemble actual flight conditions. We also provide the model that results from it, complete with error/uncertainty estimates. Such a turbulence model, customized to hypersonic flows, does not exist in aerospace engineering literature. Note that we have not tested the calibrated model for its generalizability; rather, we have demonstrated how the model could be calibrated. A generalizable model would likely require calibration to a number of experimental datasets that, pooled together, would contain most of the physics observed in flight. That is outside the scope of the paper.

The paper is organized as follows. In Sec. 2, we review existing literature on surrogate modeling of fields and the state-of-the-art in turbulence model calibration. In Sec. 3 we describe the LENS-I experimental data, the SPARC high-fidelity flow simulator, and the setup for model calibration. In Sec. 4 we describe the construction of the surrogate model using dimension reduction and different types of machine learning regressors, whose performance is then evaluated in Sec. 5. Sec. 6 contains the formulation and results of the Bayesian calibration problem. Summary and conclusions are in Sec. 7.

2   Literature review

2.1 Surrogate models for fields

Surrogate modeling of high-fidelity models is a mature topic and contemporary reviews of the field can be found in Refs. [4, 1, 14]. We limit ourselves to surrogate models that output spatial or spatiotemporal fields (see a review in Ref. [4]). The process involves generating a large number of instances of the field by executing the high-fidelity model repeatedly for different inputs (a vector of independent scalar variables) and archiving the input-output pairs as TD. The TD is then used to compute an orthogonal basis set, usually via POD [15], though Krylov subspace bases [16, 17] and Fourier bases [18, 19] too have been explored. The output field is then represented using a weighted linear combination of the bases, with the (short) vector of weights serving as a low-dimensional representation of the field. Weights, individually or as a vector, are then modeled as a function of the uncertain inputs (of the high-fidelity model) via conventional data-driven methods. The oldest example of such an approach seems to be Ref. [20], where Rayleigh-Bénard convection was modeled in this fashion, using cubic spline interpolators to model the bases' coefficients (or weights). Gaussian Process models also have been extensively used [21, 22, 23, 24], and there have been investigations into using self-organizing maps coupled with local response surfaces [25]. Neural networks can serve as universal approximators and consequently Ref. [26, 27] explore their use as a mapping between inputs and the coefficients. In transient problems with a spatial component, POD is often used to reduce spatial dimensionality while the time-evolving coefficients are modeled as a dynamical system, using neural ordinary differential equations (NODE) and recurrent neural networks [28, 29, 30]. Dimensionality reduction of a spatial field X can also be accomplished in a nonlinear manner using a neural net encoder-decoder framework and used to predict a different, dependent spatial field Y via image-to-image regression [31]. In case of a spatiotemporal field, the time-evolution of the field in the dimensionality-reduced encoded (or latent) space has been modeled using a parameterized NODE [32]. It is also possible to generate the TD in an "intelligent" manner, sampling the input space where information on the input-output relationship is desired [33]. There are studies which explore the benefits of various surrogate modeling techniques [34, 35, 36] for a given problem, including in fluid mechanics [37].

Surrogate models have also been used in compressible aerothermodynamics (hypersonic flows), often to approximate temperatures, pressures and heating on wings and control surfaces [38, 39]. They have been used within the context of aerothermoelasticity studies [40, 41, 42, 43, 44] or to design thermal protective systems [45]. In these surrogate models, the *spatial* variation of the fields are captured using POD bases, though a few studies have used kriging [40, 39, 44]. In case of POD, the coefficient of the POD bases are modeled as a function of the environment (e.g., Mach number, altitude etc.) using kriging, radial basis functions or Chebyshev polynomials. The training dataset is generated using computational fluid dynamics (CFD) simulators to span over a parameter space ranging from two to eigth dimensions, consisting of Mach number (or speed), altitude and a host of parameters describing the attitude of the vehicle and its control surfaces e.g., angle of attack, roll angle etc. Comparisons between various surrogate modeling techniques for hypersonic flow fields can be found in Refs. [38, 40, 43, 44]. In these studies the dataset for training the surrogate model generally contained $O(10) - O(10^2)$ examples. In contrast, Ref. [45] models the *temporal* variation of temperature under the thermal protection system as a hypersonic

4

vehicle executes an uncertain trajectory using Karhunen-Loève bases. Apart from the three parameters that governed the uncertain trajectory, the study also considered 18 uncertain parameters describing the material properties of the thermal protective system. The training dataset had about 400 examples.

2.2   Calibration of turbulence models

Data-driven turbulence modeling has mostly targeted closures in RANS equations, though some work has been done for Large Eddy Simulations too [46, 47]. Refs. [48, 49, 50] contain broad reviews of data-driven models used to simulate turbulent flows. Such models fall into three categories. The first category consists of studies which seek to replace contemporary RANS closures with new forms learned from TD. This often takes the form of neural networks [51, 52, 53]. Alternatively, studies have used gene expression programming to assemble new expressions e.g., a linear eddy viscosity model augmented with additional terms [54, 55, 56, 57, 58]. The second category consists of inferring a spatially variable "correction" that modulates/multiplies certain closure terms in the RANS equations; the spatially variable term is then related to the local flow state with a data-driven model such as a neural net or a random forest [59, 60, 61, 47, 62, 63]. This field estimation has traditionally been performed using optimization, but Kalman filters have also been used [64, 65, 66].

The third category consists of conventional turbulence closures that have been calibrated to flows similar to the scenarios where they are expected to be used. Two-equation turbulence models, like the SST (Shear Stress Transport) model used in this study, contain a number of approximations [9], which makes it impossible to compute turbulence closure parameters that are universally generalizable to all types of flows. Conventionally, these parameters have been calibrated to simple turbulent boundary layer and shear flows [67], but are routinely tuned to particular classes of flows [68, 69, 70]. Due to the simplicity of the (conventional) closures' forms, it is usually not possible to estimate the closures' parameters with a great deal of certainty, and consequently, Bayesian inference is used to construct a JPDF of the parameters.

The first attempt at Bayesian calibration of closure constants used data from simple flows e.g., flat-plates and wall-bounded flows [71, 72]. The 1D flow models used in these studies did not require surrogate modeling when Bayesian inference was performed using Markov chain Monte Carlo (MCMC) techniques. In Ref. [11] the authors estimate five closure parameters of the $k - \epsilon$ turbulence model for urban canyon flows using a Gaussian Process surrogate, MCMC and 10 measurements of turbulent kinetic energy from a shock-tunnel model. A rather different approach was adopted for the 3D jet-in-crossflow problem with measurements of velocity and vorticity [12, 73, 74], where polynomial surrogates were employed to estimate three closure parameters of the $k - \epsilon$ turbulence model. Bayesian inference was also used to estimate parameters of the $k - \omega - \gamma$ turbulence model for hypersonic transitional flows, using Stanton number measurements in a turbulent flow over a flat plate and polynomial chaos expansion surrogates [75]. The same authors, in Ref. [13], redid the estimation using measurements of the skin friction, using Gaussian Process surrogates. In both cases, Direct Numerical Simulations provided the calibration data. The SST model has also been tuned for hypersonic flows, using the same HIFiRE-1 measurements used in this study [8], but the manual adjustment of a single parameter in the SST

model is not quite a formal calibration. Closure parameters of the $k - \omega$ turbulence model have also been estimated using Ensemble Transform Kalman Filters and measurements from a backward-facing step experiment [76]. The same method has been used to estimate spatially-varying turbulent viscosity fields using measurements from transonic flows over airfoils and wings [77]. In Ref. [78], the authors used Approximate Bayesian Computations (ABC) to estimate the parameters of a non-equilibrium RANS turbulence model, and quantified the dependence of those parameters on the statistical summaries used for calibration. The measurements used in the calibration were obtained from the Bachalo & Johnson experiment (transonic flow over an axisymmetric bump). Finally, in Ref. [79], the authors discovered, while calibrating RANS for Rayleigh-Taylor instabilities, a dependence of a parameter on the Atwood number and thus helped uncover and remove a model-form uncertainty (i.e., the use of a Atwood-number-indpendent parameter).

3 The HIFiRE-1 configuration and experiment

3.1 Flow configuration and RANS model
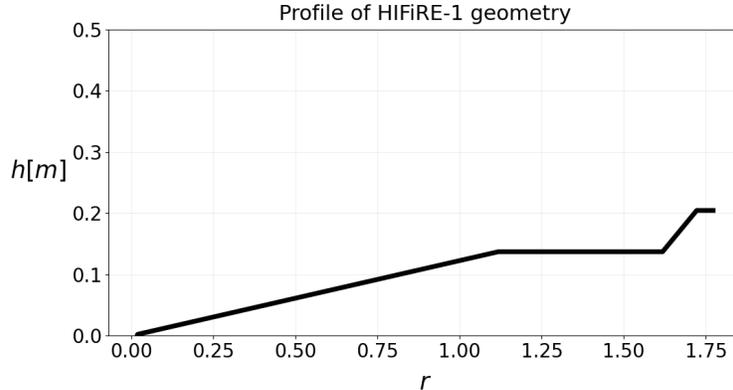
Figure 1: Profile of the cone-shaped HIFiRE-1 geometry. The heat flux and pressure fields are measured as a function of the distance from the tip of the nose along the axis of rotation, r.

The flow configuration: The flow configuration being simulated is the HIFiRE ground test conducted in CUBRC's (Calspan-University at Buffalo Research Center) LENS-I shock-tunnel facility, as described in Ref. [7]. The HIFiRE-1 geometry is cylindrical, and 1721.7 mm in overall length and 409.2 mm in diameter. It consists of a conical forebody, with half-angle of 7 degrees and of length 1118 mm. The cone has a blunted nose of diameter 2.5 mm and is followed by a cylindrical midbody of 400 mm. The aftbody, which is a flare of angle 33 degrees, follows the midbody and is 203.7 mm long (see schematic of the profile in Fig. 1). The test-section of the shock-tunnel is capable of accommodating test models 3 feet in diameter and 12 feet long. The geometry is aligned with the flow, leading to a nominally axisymmetric flowfield. The HIFiRE-1 surface was instrumented with pressure

and heat flux sensors. There were 42 piezoelectric pressure sensors (with a measurement error of $\pm 3\%$) and 76 thin-film heat-flux sensors (with a measurement error of $\pm 5\%$) on the surface. The flow is tripped (artificially rendered turbulent)[2] at a location 0.505 metres from the nose-tip, causing a dramatic increase in aerodynamic heating. The experiment has been modeled using the RANS equations with the Menter SST turbulence closure [10] (our "full-order" model) previously and is described in Ref. [8]. The flow configuration used in this paper is the "Condition B" of Ref. [8], corresponding to nominal HIFiRE-1 flight conditions at an altitude of 21.1 km. The inflow velocity v is 2170 m/s, with a freestream temperature u of 226.46 K and density ρ of 0.066958 kg/m^3. The Mach number is 7.16 and the unit Reynolds number $Re \approx 10.2 \times 10^6$/m. The total enthalpy of the flow is 2.38 MJ/kg. The uncertainty in the LENS-I freestream measurements are 0.5% for the Mach number, 3% for the temperature and 1% for the pressure [7]. Some simple algebra on the ideal gas model reveals that the freestream velocity and density have a measurement uncertainty of 2% each. The HIFiRE-1 body was kept at a temperature of 296.7 K. The slender cone causes oblique shock-waves to form near the nose. The turbulent flow separates in front of the flared aftbody, which also causes complex shock structures, including shock-boundary layer interactions. This is captured in both the heat-flux and pressure measurements and the full-order model calculations.

The full-order (RANS) model: The full-order flow model solves the Reynolds-Averaged Navier Stokes (RANS) equations [9], with Menter's SST turbulence closure [10]. It is implemented within Sandia's SPARC (Sandia Parallel Aerodynamics and Reentry Code) flow simulator. SPARC implements a second-order-accurate finite-volume spatial discretization of the compressible-flow RANS equations, which consists of the continuum conservations laws for mass, momentum and energy (a more detailed description of the simulator is in the appendix of Ref. [80], and some grid-convergence studies are in Ref. [81]). These are formulated for reacting gases in thermochemical non-equilibrium, though for the low total enthalpy of our flow (2.38 MJ/kg) an ideal gas approximation is used. The equations are solved using a finite-volume method for the conserved variables. SPARC can accomodate structured and unstructured meshes, though, given the simple geometry, we only use structured ones in this paper. For the simulations in this paper, we use a Steger–Warming scheme for the inviscid fluxes, extended to second-order using a MUSCL reconstruction. A mimod limiter is used within the reconstruction. Diffusion and viscous terms in the conservation laws are discretized using a central difference scheme. SPARC solves the unsteady form of the governing equations, using a second-order backward difference scheme for the time-integrator. Since all the simulations in this paper are steady, we run the time-integrator in its first-order form to accelerate the convergence to a steady-state solution. The HIFiRE-1 simulations were computed on a $1024 \times 512 \times 2$ grid-cell mesh, clustered near the HIFiRE-1 surface to resolve the boundary layer and around the region with the shock attached to the nose-tip. Two planes were used in the depth-wise direction to fit into the 3D nature of the SPARC software; the two planes were mirrored. Studies to assess the adequacy of the grid are in Appendix A. The fluid dynamical quantities on the solid boundary were interpolated onto a 1085×2 surface mesh that defined the HIFiRE-1 geometry via quadratic

---

[2]In the RANS equations, the source terms in the k and ω equations are set to zero for $r \leq 0.505$ m and the turbulent kinetic energy k is set to zero at the inflow boundary, as the freestream flow is laminar.

3.2 Model calibration and setup

The full-order solution represents the heating and pressure fields on the HIFiRE-1 geometry which we denote by

$$\mathbf{y}(\cdot; \mathbf{x}) \in \mathbb{R}^{N_s} \tag{1}$$

where N_s is the dimensionality of spatial discretization of the solution field, and $\mathbf{x} \in \mathbb{R}^d$ is the d-dimensional tuning parameter or feature space. Fig. 1 shows the profile of the HIFiRE-1 geometry. The full geometry is the shape generated by the rotation of the profile around the r axis. The heat flux and pressure fields are measured on the surface along the length of the test geometry, and, due to the symmetry of the mesh, are given as a one-dimensional function of the axis of rotation, r. Plots of the pressure and heat flux fields, computed with the nominal turbulence model and inflow conditions, can be found in Ref. [8], along with numerical Schlierens of the shock structures. We can write the solution vector in discretized form as

$$\mathbf{y}(\cdot; \mathbf{x}) \doteq [y(r_1; \mathbf{x}), y(r_2; \mathbf{x}), \dots, y(r_{N_s}; \mathbf{x})]^T, \tag{2}$$

where r_i's are the discretized mesh points along the profile in Fig. 1.

The feature space consists of 12 tunable input parameters, including three free-stream parameters, i.e., temperature, density and velocity, and 9 closure constants defined by the standard Menter two-equation (SST) model [10]. These quantities are varied by scaling (multiplying) them by a uniform random variable resulting in values $\pm 15\%$ from the nominal. Table 1 shows the three freestream quantities and their scaling parameters and Table 2 does the same for nine SST parameter constants. The $\pm 15\%$ variation for the freestream quantities' scalings are designed to bracket the uncertainty in the measured values (see above), so that the surrogate model is comfortably applicable over our prior belief regarding the freestream conditions. The uncertainties in the SST parameters are obtained from Ref. [82]. Since these uncertainties are only known as bounds, we proceed with uniform distributions (under a maximum entropy assumption) for their prior densities.

Note that we are primarily interested in the nine calibrated SST model parameters. However, the freestream density and velocity in the shock-tunnel, which are only known to within $\pm 2\%$ error, also affect the measured heat flux and pressure; the heat flux $q \sim \rho v^3$ and pressure $p \sim \rho v^2$ [80]. This strong dependence implies that the uncertainty in the freestream quantities have the potential to affect the calibrated values of the SST parameters. Thus we will perform a joint estimation of the freestream variables *and* the SST parameters, and compare it with a calibration when the freestream parameters held constant at their nominal values. This comparison will reveal the degree to which the uncertainty in the shock-tunnel inlet conditions affect the calibrated model.

The goal of calibration is to find the inputs for the flow model (set of $\mathbf{x}$'s) that minimize some measure of discrepancy between the model prediction, $\mathbf{y}$, and some observed, experimental data, $\mathbf{y}_{\text{obs}}$. Let $\mathsf{d}: \mathbb{R}^{n_{\text{obs}}} \times \mathbb{R}^{n_{\text{obs}}} \mapsto \mathbb{R}^+$ be a discrepancy function between two vectors of size n_{obs}, which we use to measure the distance between the model prediction

Scaling/multiplier	Value	Freestream quantity	Nominal value
ρ_s	$(0.85, 1.15)$	ρ (density)	0.066958
v_s	$(0.85, 1.15)$	v (velocity)	2170
u_s	$(0.85, 1.15)$	u (temperature)	226.46

Table 1: Three parameters (left in white) are the multipliers used to scale the freestream specification. The bounds of the uniform distribution for them are in the second column. The multipliers are varied by $\pm15\%$ around 1. The third column contains the freestream quantities and the fourth column their nominal values.

Par.	Value	Par.	Value	Par.	Value	Par.	Value
σ_{k_1}	$(0.7, 1.0)$	σ_{w_2}	$(0.7, 1.0)$	a_1	$(0.31, 0.40)$	β_1	$\beta^*/\beta_{1,r}$
σ_{k_2}	$(0.8, 1.2)$	β^*	$(0.0784, 0.1024)$	$\beta_{1,r}$	$(1.19, 1.31)$	β_2	$\beta^*\beta_{2,r}$
σ_{w_1}	$(0.3, 0.7)$	κ	$(0.38, 0.42)$	$\beta_{2,r}$	$(1.05, 1.45)$		

Table 2: Table showing nine parameters from the SST turbulence model and their respective parameter ranges. "Par." is an abbreviation for SST parameters. Nominal values are exact center of the specified ranges.

$\mathbf{y}(\cdot, \mathbf{x})$ and some observation $\mathbf{y}_{\text{obs}}$.[3] Then, the deterministic calibration problem can be written as

$$\arg\min_{\mathbf{x}} \; \mathsf{d}(\mathbf{y}(\cdot, \mathbf{x}), \mathbf{y}_{\text{obs}}, ; \theta), \tag{3}$$

were θ represents the parameters of the discrepancy function. A typical discrepancy is the squared error metric given by

$$\mathsf{d}_{\text{SE}}(\mathbf{y}(\cdot, \mathbf{x}), \mathbf{y}_{\text{obs}}, ; \theta) \doteq \frac{\|\mathbf{y}(\cdot, \mathbf{x}) - \mathbf{y}_{\text{obs}}\|_2^2}{\theta^2} \tag{4}$$

where the numerator is the canonical squared error norm and θ^2 the variance. In a Bayesian formulation, we can write

$$\mathbf{y}_{\text{obs}} = \mathbf{y}(\cdot, \mathbf{x}) + \epsilon, \tag{5}$$

where ϵ is the discrepancy between model predictions and measurements and is modeled as $\epsilon \sim \mathcal{N}(0, \theta^2)$, $\mathcal{N}(\cdot, \cdot)$ being a normal distribution. In such a case, problem Eq. (3) can be re-interpreted as the negative log-likelihood function. Moreover, if we place a prior distribution on θ and $\mathbf{x}$, we have fully defined a posterior distribution for the feature space parameters:

$$\log p(\mathbf{x}, \theta) \propto -\mathsf{d}(\mathbf{y}(\cdot, \mathbf{x}), \mathbf{y}_{\text{obs}}, ; \theta) + \log \pi(\theta, \mathbf{x}), \tag{6}$$

where $\pi(\theta, \mathbf{x})$ is the prior distribution on θ and $\mathbf{x}$, e.g., an inverse gamma density if θ represents the variance in a sum of squares discrepancy error and a uniform prior over some prescribed bounds, respectively. Fig. 2 illustrates the discrepancy between the full model prediction, $y(r; \mathbf{x}_i)$ evaluated at a single set of sample parameters and the experimental data for heat flux and pressure, respectively. In the heat flux profile, there is an abrupt

---

[3]We note that n_{obs} does not need to be the same as, or even a subset of N_s. In fact, in this work, the observed points fall in between the discretized mesh points.

jump at $r \approx 0.505$ where the simulation is "tripped" into turbulent flow, whereas the experimental data shows a gradual change consistent with a transition zone. Since pressure does not vary in the laminar versus turbulent zone, no changes are seen in the pressure profile (Fig. 2, bottom). Further, at $r \approx 1.25$, both pressure and heat flux decline, with the RANS model closely following experimental data; expansion fans are not difficult to model. Finally, $r \gtrsim 1.4$, the flow separates, with the heat flux and pressure climbing steeply, but significantly overpredicted by the RANS model. This is due to the model-form error in RANS and its consequent inability to model separation zone correctly. The goal
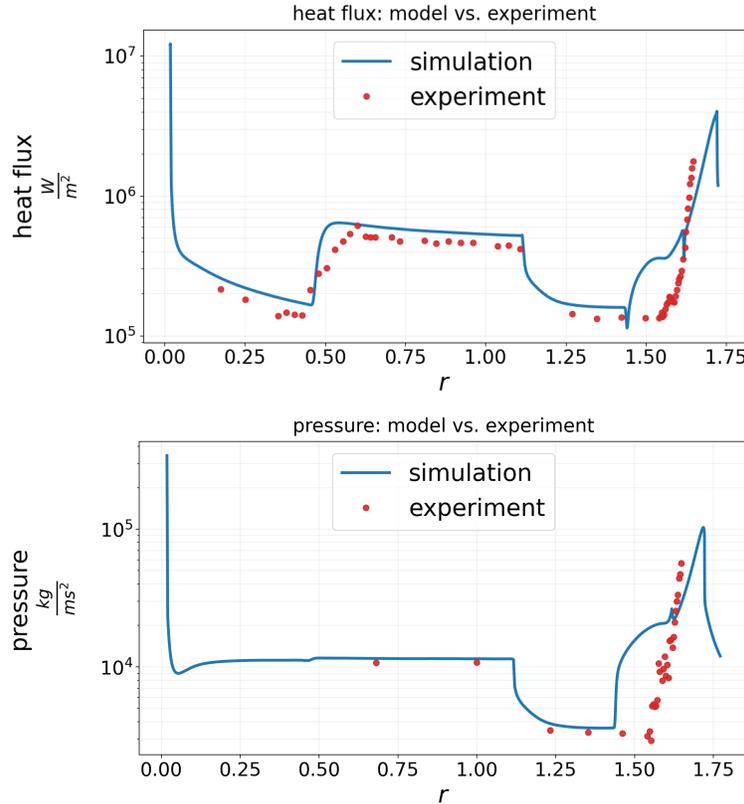
Figure 2: Model output versus the experimental HIFiRE-1 data for heat flux and pressure. The sharp changes in values correspond to changes in the geometry profile as seen in Fig. 1. The RANS model prediction, using the nominal values of the model, is plotted with a solid line. The experimental measurements are plotted with symbols.

for calibration is to find the values of $\mathbf{x}$ that result in the *best* match between the model (blue lines in Fig. 2) and the experimental data (red dots in Fig. 2). The search for the optimal $\mathbf{x}$ will require the evaluation of the RANS model repeatedly, which, at 384 CPU-hours a run[4], would make the search in 12-dimensional space intractable. Consequently, it is necessary to replace the RANS model of the HIFiRE-1 geometry with a fast-running proxy i.e., surrogate model. **In summary, the calibration procedure involves two**

---

[4]Total computational time for each simulation is roughly 3 hours on 128 cores using an Intel Xeon Gold 6140 CPU at 2.30GHz.

10

stages: (1) construct a ROM-based surrogate for the heat flux and pressure fields, and then (2) perform Bayesian inference to infer a joint density on the tuning parameters informed by the discrepancy between the surrogate and the HIFiRE-1 measurements.

4 Surrogate construction using proper orthogonal decomposition

A simple approach for constructing surrogates for spatially varying fields, sometimes called the multi-target regression problem in the machine learning literature, is to construct a (sub) surrogate for each element in the output field (see Ref. [73] for an example). The complete surrogate is then the union of all the individual surrogate models. This is cumbersome if the dimensionality of the output $\mathbf{y}(\cdot; \mathbf{x})$ is very large, as is the case for most complex problems. In addition, this method does not preserve the correlations that exist between different $y(r_i; \mathbf{x})$ in an efficient way. In contrast, we propose using proper orthogonal decomposition (POD) to transform the solution space to a low-dimensional subspace or latent space and *then* fit the handful of latent space dimensions with separate single-target surrogate models. If the transform is invertible[5], we can simply invert back to the full order solution space for direct comparison.

The procedure for dimension reduction of the output and subsequent regression fitting is as follows. Let $\mathbf{Y} \doteq [\mathbf{y}_1, \ldots, \mathbf{y}_m]^T \in \mathbb{R}^{m \times N_s}$ be a snapshot matrix of sample solutions or ensemble runs, where each row of $\mathbf{Y}$ represents a solution field for a particular parameter set. These m samples are generated by sampling the twelve-dimensional feature space using Latin hypercube sampling over the prescribed bounds. We then perform principal component analysis (PCA) on this (centered) snapshot matrix to obtain a set of orthogonal transformations denoted by $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n] \in \mathbb{R}^{N_s \times n}$, where $\boldsymbol{\phi}_i \in \mathbb{R}^{N_s}$'s represent n new coordinate axes representing the directions of maximum variances. The associated coordinates or projection coefficients for each basis term is given by $\mathbf{c}_i \doteq \mathbf{Y}_c \boldsymbol{\phi}_i \in \mathbb{R}^m$, where $\mathbf{Y}_c$ is the centered snapshop matrix (see Algorithm 1). The empirical variance is then given by $\lambda_i \doteq \sigma_i^2 / (m-1)$, where σ_i's are the singular values associated with the SVD of $\mathbf{Y}$, or equivalently, the eigenvalues associated with the normal matrix. With $\mathbf{X} \doteq [\mathbf{x}_1, \ldots, \mathbf{x}_m]^T \in \mathbb{R}^{m \times d}$ as the data matrix for the feature space, our subsequent task is to then create a surrogate model for each of the reduced space training data pairs $\{\mathbf{X}, \mathbf{c}_i\}$ for $i = 1, \ldots, n$. While this is still a multi-target regression problem, the number of targets is $n \ll N_s$. If we denote $\hat{y}_j(\mathbf{x}) : \mathbb{R}^d \mapsto \mathbb{R}$ each of the $j = 1, \ldots, n$ surrogate models corresponding to each component, then our full surrogate model is given by the Karhunen-Loeve expansion [83]

$$\mathbf{y}(\cdot; \mathbf{x}) \approx \tilde{\mathbf{y}}(\cdot; \mathbf{x}) = \mu_0 + \sum_{j=1}^{n} \sqrt{\lambda_j} \hat{y}_j(\mathbf{x}) \boldsymbol{\phi}_j. \tag{7}$$

The complete PCA/ POD algorithm with details about automating the choice of n is shown in Algorithm 1.

---

[5]It may be the case that the inversion is not lossless, e.g., principal component analysis.

Algorithm 1 Principal Component Analysis

Input: Snapshots $\mathbf{Y} \in \mathbb{R}^{m \times N_s}$ and percentage variance threshold $\nu \in [0, 1]$.[a]

Output: Basis matrix $\mathbf{\Phi} \in \mathbb{R}^{N_s \times k^*}$, and projections $\mathbf{C} \in \mathbb{R}^{m \times k^*}$.

1: Center the snapshot data matrix, i.e. $\mathbf{Y}_c = \mathbf{Y} - \boldsymbol{\mu}$, where $\boldsymbol{\mu} = \sum_{j=1}^{m} \mathbf{y}_j$, i.e., mean w.r.t. the rows.

2: Compute thin singular value decomposition (SVD): $\mathbf{Y}_c = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{m \times K}$, $\mathbf{\Sigma} \in \mathbb{R}^{K \times K}$, $\mathbf{V} \in \mathbb{R}^{N_s \times K}$, where $K = \min(N_s, m)$.

3: Find $k^* = \arg\min\{k \in \mathbb{N}^+ : \sum_i^k \sigma_i^2 / \sum_i^K \sigma_i^2 \geq \nu\}$.

4: Set $\mathbf{\Phi} = [\mathbf{v}_1, \ldots, \mathbf{v}_{k^*}]$, where $\mathbf{v}_i$'s are the columns of $\mathbf{V}$ and compute $\mathbf{C} = \mathbf{Y}\mathbf{\Phi}$.[b]

---

[a]Each of the m snapshots corresponds to a model output evaluated at $\mathbf{x}_m$, i.e. a sample from the feature or tuning space.

[b]One can also scale the projections using $\mathbf{C} = \mathbf{Y}\mathbf{\Phi}\mathbf{\Sigma}_*^{-1}$, where $\mathbf{\Sigma}_*$ is the $k^* \times k^*$ submatrix of $\mathbf{\Sigma}$.

Once the PCA approach is performed on the solution field, the remaining effort is constructing machine learning surrogates for $\hat{y}(\mathbf{x})_j$'s. To determine the best regression, we experiment with an array of different types of machine learning regressor models including Gaussian process regression, multi-layer perceptron (fully connected neural network) models, random forests, kernel ridge regression, support vector machines, and, last but not least, polynomial chaos (or multi-variate polynomial) expansions using Legendre polynomials. Each of these regressors are hyper-parameter-tuned over a specified parameter grid, e.g., polynomial order and regularization type for polynomial fitting, using five-fold cross validation in order to perform model comparison (see Appendix B) . We briefly summarize some of the key model features of these model regressors in Sec. 5. See Ref. [84, 6] for a more thorough discussion of orthogonal polynomial interpolants for multivariate model fitting and dense neural network construction, and the Scikit-Learn documentation [85] for a brief discussion of the other five estimators and their respective implementation.

5   Surrogate models for HIFiRE-1 simulations

The training dataset (TD) for the calibration and surrogate model construction is generated using Latin hypercube sampling (LHS) of the feature space. Ranges for the LHS study are defined in Table 1 and Table 2.[6] The full-order i.e., RANS model is then evaluated at $m = 2500$ sample points and the heat flux and pressure fields are recorded to produce input/output data pairs $\{(\mathbf{X}, \mathbf{y}_i)\}$ for $i = 1, \ldots, n$. Furthermore, we use 5-fold cross validation to tune and evaluate the accuracy of each possible regression technique (see Appendix B for more details about the hyper-parameter tuning). Fig. 3 shows summary statistics of realizations or snapshots of the heat flux and pressure fields. The realizations are plotted as a function of the geometry profile shown in Fig. 1. This is data we use to train our surrogate model in order to capture the effect of perturbations in the twelve-dimension feature space on the heat flux and pressure fields.

---

[6]Note that the ranges for the training data are larger than the allowable ranges for the Bayesian calibration in order to avoid problems with extrapolation.
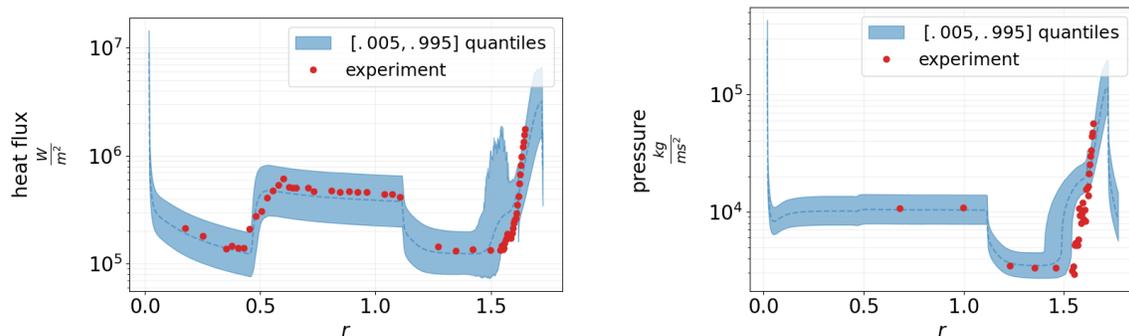
Figure 3: Ranges for of heat flux and pressure fields from the LHS study, simulated at $m = 2,500$ LHS points, plotted against the geometric profile in Fig. 1. The experimental measurements for heat flux and pressure, respectively, are also shown in red. The bottom range represents the $q = .005$ quantile while the upper range represents $q = .995$, so the total range encompasses 99% of the total data. The median is shown in the dashed blue line.

Once the TD is generated, we begin with the dimensionality reduction of the spatially varying targets using Algorithm 1. Fig. 4 shows the cumulative explained variance ratio, which can be used to determine the dimension of the latent space. From this plot we can deduce that a latent space dimension of $k^* = 4$ captures more than 99% of the total variance of the original signal. [7]
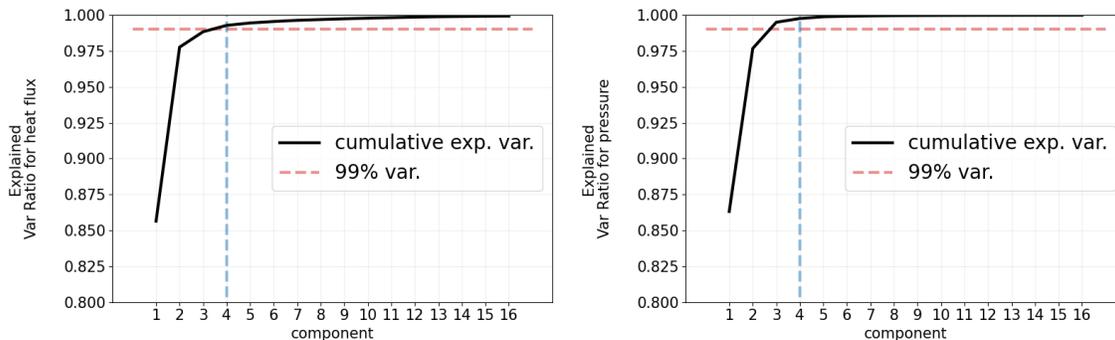
Figure 4: Cumulative explained variance ratio (i.e., Scree plots) as a function of the PCA component for both the heat flux (left) and pressure (right) fields. Four components (vertical blue line) are enough to capture more than 99% of the total variance (red dashed horizontal line) for both heat flux and pressure fields (a 1000x dimension reduction for only 4 components).

The first four components for the heat flux and pressure fields are shown in Fig. 5. The components (plotted as a function/vector of the geometric profile parameter r) represent the directions of maximum variance, in decreasing order. One can also interpret these

---

[7]We experimented with using six or eight components, but the resulting surrogate only improved test errors by less than a tenth of a percent.

components or modes as the axes of a new coordinate transformation in N_s-dimensional space. While only four components are needed to capture nearly the entirety of the variance,
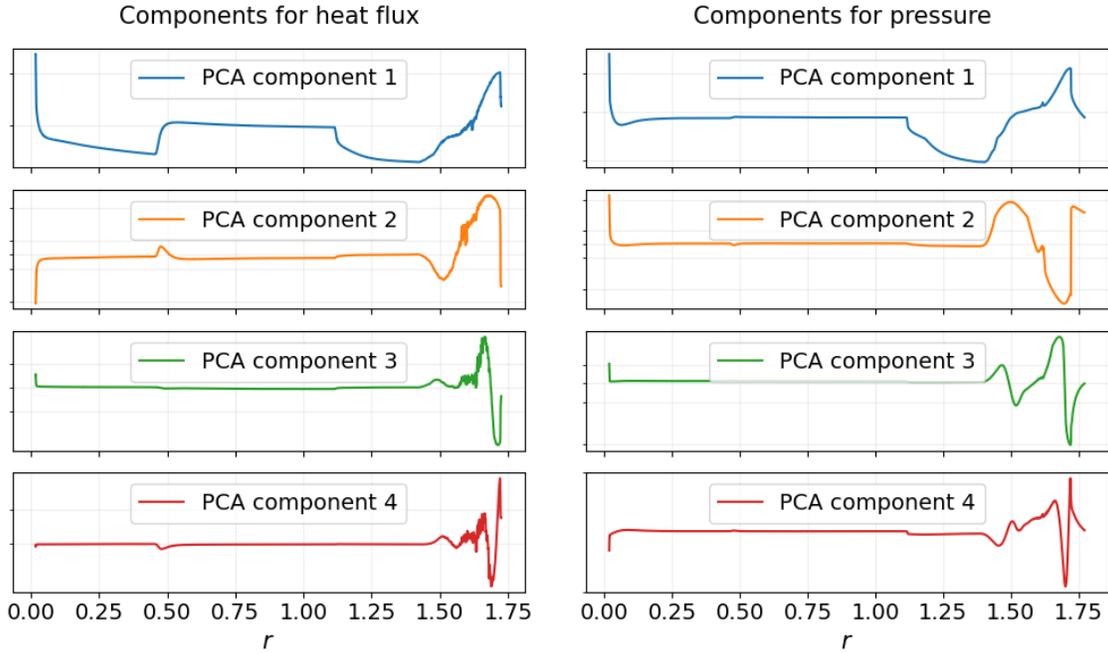
Figure 5: First four principal components for heat flux (left) and pressure (right) fields. Much of the variance is concentrated around the $r \gtrsim 1.5$ area, where the simulation struggles to capture the experimental data.

Fig. 6 shows the difference in the original versus the reconstructed signal using $k^* = 4$ components. Indeed the statistics of reconstructed signal are indeed distinguishable from the original data. This shows that the dimension reduction, while extremely accurate in capturing the total variance, is not lossless when transformed back to the original space.
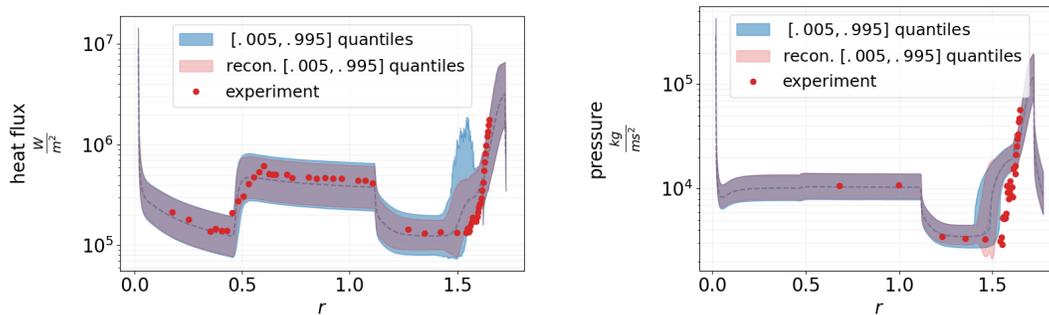
Figure 6: 99% quantile comparisons of heat flux and pressure fields from LHS study (blue) versus PCA reconstruction samples (red), compared to experimental data (red dots).

Each of the component projections are fit with six different regression models and a 5-

14

fold cross-validation score is computed. Furthermore, for added robustness, each regressor is tuned over a set of prescribed hyper-parameters, e.g., a range of maximum depth and number of estimator are tested to obtain the best overall random forests regressor, a range of polynomial orders, least squares and sparse solvers are tested to obtain the best overall polynomial regressor, and etc. (See Appendix B for more details about the hyper-parameter tuning and the final model architecture). We use a negative root mean square score function for the individual regressors. Fig. 7 shows the root mean square errors (RMSE) for $\hat{y}_1$ in Eq. (7), i.e., the surrogate for the projection coefficients of the first component of heat flux and pressure, respectively. The y-axis separates the different regressors and the x-axis shows the the RMSE (the smaller the better). Recall that the first component has the largest contribution to the total variance, and it decreases from there onwards. Thus, if we focus on the first component, we see that the two best regressors, i.e., the lowest RMSE, for fitting the component projections are the polynomial chaos expansions (PCE) with Legendre polynomials and the multi-layer perceptron (MLP) models. We left out errors for the k-nearest neighbor approach since the errors where significantly worse.
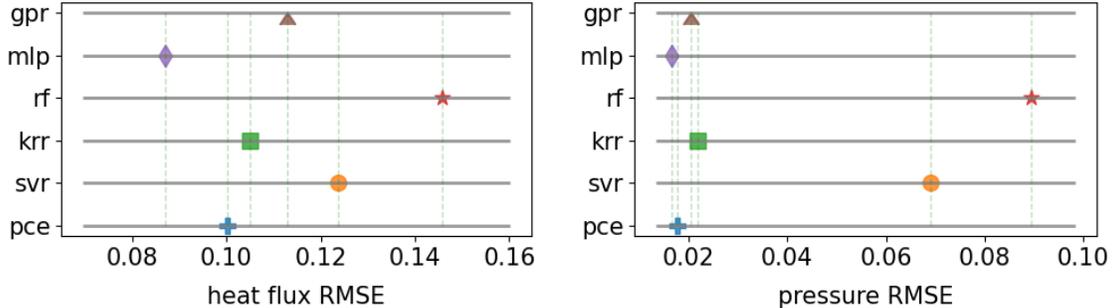
Figure 7: Root mean squared errors for regression surrogates for the first projection component of heat flux (left) and pressure (right). The six different models used for comparison are Gaussian processes (gpr), multi-layer perceptron (mlp) or dense neural networks, kernel ridge regressor (krr), random forests (rf), support vector machine (svr), and polynomial chaos expansions (pce).

We also compute the root mean square relative error (RMSRE) for the full field solution $\tilde{y}(\boldsymbol{x})$, which we obtain by projecting the latent space surrogate back to the original space, for the two best performing surrogates in the latent space, i.e., the PCEs and the MLPs. The error for the full field solution is defined by

$$\epsilon_i \doteq \frac{\|\mathbf{y}_i - \tilde{\mathbf{y}}_i\|^2}{\|\mathbf{y}_i\|^2},\tag{8}$$

where $\tilde{\mathbf{y}}$ is the ML surrogate, i.e. either PCE or MLP, evaluated at the ith training or observed data point. Here we haven chosen to use relative error explicitly in order to get a sense of the relative magnitude of the surrogate construction error, and not just as a tool for model selection. Not shown in this plot is the baseline error computed by creating a "dummy" mean predictor which had a RMSRE of .1. Thus, our surrogate model reduces the relative error by almost an order of magnitude (10x) from the mean predictor.

15
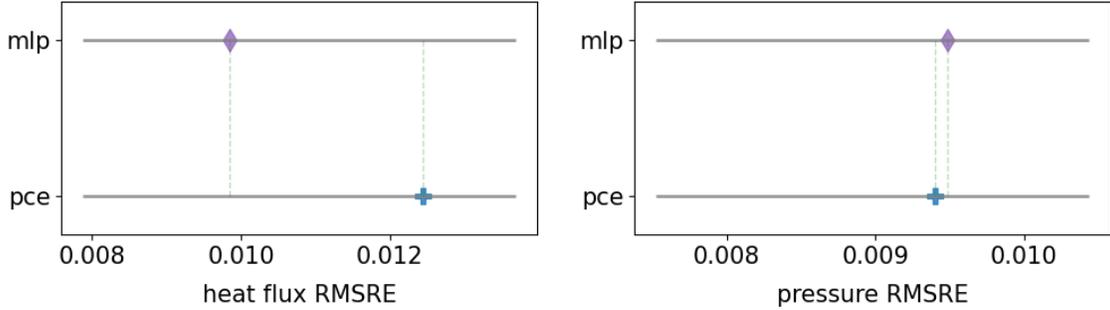
Figure 8: Root mean squared relative errors (RMSRE) for the full field solution surrogate. We calculate the 5-fold cross-validation error only for two best performing models on the projection coefficients: multi-layer perceptron (mlp) or dense neural networks and polynomial chaos expansions (pce). Heat flux is shown in the left and pressure on the right. For a baseline comparison, the mean "dummy" predictor has an RMSRE of .1, which is roughly ten times worse.

When choosing our final surrogate model, we have to balance simplicity with accuracy. With that in mind, for the purposes of the Bayesian calibration, we will use Legendre PCEs for our surrogate model construction. MLPs suffered from long training times, difficult hyper-parameter tuning spaces, and sensitivities to random seeds, which made them less reliable and less robust. PCEs offered simpler models and fitting approaches with far fewer tuning parameters and better reproducability.

Before we move on the Bayesian calibration we would like to make a quick remark about the speed up in computational time. Recall that a single run for the full simulation takes approximately three hours on a supercomputer. The surrogate, on the other hand, can run in less than a tenth of a second on a standard laptop. Thus, the surrogate runs about one hundred thousand times faster, i.e., a speed up of five orders of magnitude, with far less computational power! The caveat is that, of course, the generation of the training data and the subsequent training of the surrogate model itself are not trivial tasks. Still, without the surrogate, it would be impossible to perform any sort of gradient-based parameter optimization or Bayesian calibration, which requires thousands or even millions of model evaluations, in any reasonable amount of time.

Next, in Sec. 6, we use the newly constructed ROM-based PCE surrogate to perform Bayesian calibration of the HIFiRE-1 experiment. We will show that the Bayesian approach not only provides a better overall model, i.e., improved match to the HIFiRE experiment, but also a measure of the uncertainty for the heat flux and pressure fields via the joint parameter probability density function.

6 Bayesian calibration of the HIFiRE-1 experiment

As in Section 3.2, we use a squared error discrepancy term, which equates to standard additive Gaussian white noise model (Eq. 5; also Ref. [86, Section 2.5]), between the surrogate predictions and the experimental data. Since we are attempting to simulatenously

use both the heat flux and the pressure field during the calibration, our final discrepancy error is actually a normalized average of the error in the heat flux and the pressure field, with the same shared $\mathbf{x}$ parameter, but separate noise models θ_q and θ_p. This implies that the mismatch between model predictions and measurements at a point for the heat flux is assumed to be independent of the mismatch observed for the pressure. Denote $\mathsf{d}_q(\mathbf{x}; \theta_q)$ and $\mathsf{d}_p(\mathbf{x}; \theta_p)$ to be the heat flux and pressure log-likelihood discrepancy terms respectively, where θ_p and θ_q represent the noise parameters of the log-likelihood. Then, the full model posterior form is given by

$$\log p(\mathbf{x}; \theta_q, \theta_p) \propto -w_q \mathsf{d}_q(\mathbf{x}; \theta_q) - w_p \mathsf{d}_p(\mathbf{x}; \theta_p) + \log \pi(\theta_q) + \log \pi(\theta_p) + \log \pi(\mathbf{x}), \quad (9)$$

where $\log \pi(\theta_q)$, $\log \pi(\theta_p)$, and $\log \pi(\mathbf{x})$ are the log priors, and w_p, w_q are weights chosen to give equal weighting to pressure and heat flux calibration.[8] The explicit posterior is given by

$$\begin{aligned}
\log p(\mathbf{x}; \theta_q, \theta_p) \propto &-\frac{w_q}{2\theta_q^2} \sum_{i=1}^{n_{\mathrm{obs}}^q} \left(\tilde{y}_q(r_i^{\mathrm{Exp}}; \mathbf{x}) - y_{q,i}^{\mathrm{Exp}} \right)^2 - w_q n_{\mathrm{obs}}^q \log \theta_q \\
&-\frac{w_p}{2\theta_p^2} \sum_{i=1}^{n_{\mathrm{obs}}^p} \left(\tilde{y}_p(r_i^{\mathrm{Exp}}; \mathbf{x}) - y_{p,i}^{\mathrm{Exp}} \right)^2 - w_p n_{\mathrm{obs}}^p \log \theta_p \\
&+ \log \pi(\theta_q) + \log \pi(\theta_p) + \log \pi(\mathbf{x}), \quad (10)
\end{aligned}$$

where n_{obs}^q is the number of experimental observations for heat flux, r_i^{Exp}'s are the locations at which the observations were made (different from r_i), $y_{q,i}^{\mathrm{Exp}}$ is the experimental observations for heat flux, and $\tilde{y}_q(r_i^{\mathrm{Exp}})$ is the surrogate prediction at the observed location. The terms are analogous for pressure. It is clear from Fig. 3 that the discrepancy between the model and the observations are higher for the heat flux than the pressure (even after weighting), thus the reason for the two θ's.

We briefly discuss the choice of the prior distributions next. The feature vector $\mathbf{x}$ are divided into two sets. The first set, consisting of the (scalings for) freestream (or shock-tunnel inlet) conditions, are modeled using uniform distributions as $\rho_s \sim \mathcal{U}(0.98, 1.02)$, $v_s \sim \mathcal{U}(0.98, 1.02)$ and $u_s \sim \mathcal{U}(0.85, 1.15)$, where $\mathcal{U}(a, b)$ denotes a uniform distribution between (a, b). The bounds for ρ_s and v_s reflect the $\pm 2\%$ uncertainty in the freestream conditions for density and velocity (see Sec. 3). The measured quantities are only weakly sensitive to the freestream temperature and so the bounds on u_s are the same as those used for training the surrogate model (Table 1). The second set consists of the SST model parameters whose prior densities are cast as uniform distributions with the bounds specified in Table 2. We use a gamma prior on the inverse variance parameter, which is the conjugate prior for the Gaussian likelihood, where we denote the precision as $\tau_q = \theta_q^{-2}$ and $\tau_p = \theta_p^{-2}$ with shape and scale parameters for their respective gamma densities set to $k = 2, \theta = 2$, chosen to encapsulate the "model errors".[9] "Model errors" here refer to a composite of model-form errors, measurement errors and discretization error of the mesh. For our HIFiRE-1 case,

---

[8]Since the number of heat flux, n_{obs}^q, and pressure, n_{obs}^p, observations are different, we set $w_q = 1/n_{\mathrm{obs}}^q$ and $w_p = 1/n_{\mathrm{obs}}^p$ or, equivalently, $w_q = 1$ and $w_p = n_{\mathrm{obs}}^q/n_{\mathrm{obs}}^p$, to give equal weight to all observations.

[9]The log gamma prior for τ is $\log(\tau; k, \theta) \propto (k-1)\log(\tau) - \frac{\tau}{\theta}$

the model-form errors dominate due to the existence of a separation zone, which is poorly modeled by RANS.

6.1 MCMC results

In order to obtain samples from our posterior distribution, we use Markov Chain Monte Carlo (MCMC) [87, 88]. The idea of MCMC is to derive a Markov chain, with a prescribed transition probability, such that the chain converges to a stationary distribution equal to the posterior distribution. A modified version of the classic Metropolis-Hastings algorithm [86] which adapts the covariance kernel of the transition probabilities [89] was used, often referred to as an adaptive MCMC methods (AMCMC). 32 parallel chains were run, each with a 50,000 burn-in period and 750,000 post burn-in runs, for a total of 24 million samples. The resulting chains had an average autocorrelation of ≤ 500 and all chains had an acceptance rate of $0.21 - 0.22$. We aggressively thinned the chain by 1000 for a total of effective samples size of 24,000 samples. See Appendix C for autocorrelation diagnostics.

The thinned chains provide samples from the 12-dimensional JPDF that is the solution of the Bayesian inverse problem for the freestream scalings and SST parameters. These samples are marginalized (integrated over all dimensions except one) to compute the posterior probability density functions (PDFs) of each of the features i.e., elements of $\mathbf{x}$. These are plotted in Fig. 9 with a solid line. The prior distribution (plotted with a red dashed line) and the nominal value taken from Table 1 and Table 2 (vertical dashed line), are also shown. A posterior PDF that differs significantly from the prior density implies a calibrated parameter that has assimilated information from the measurements. It is clear that the freestream scaling (ρ_s, v_s, u_s) can be inferred quite easily - the PDFs' peaks are sharp and distinct from the nominal values. The performance of the SST parameters are mixed. Some like a_1 and σ_{k_2} have sharp PDFs whereas others such as σ_{k_1} and κ are not well informed by the measurements.

Scaling/multiplier	MAP Value	Freestream quantity	Nominal value
ρ_s	0.981	ρ (density)	0.066958
v_s	0.980	v (velocity)	2170
u_s	0.856	u (temperature)	226.46

Table 3: Same as Table 1 but with the optimal MAP parameters.

Par.	Value	Par.	Value	Par.	Value	Par.	Value
σ_{k_1}	0.717	σ_{w_2}	0.715	a_1	0.399	β_1	$\beta^*/\beta_{1,r}$
σ_{k_2}	1.183	β^*	0.079)	$\beta_{1,r}$	1.304	β_2	$\beta^*\beta_{2,r}$
σ_{w_1}	0.686	κ	0.382	$\beta_{2,r}$	1.219		

Table 4: Same as Table 2 but with MAP values

The final analysis involves pushing through a few hundred samples of the posterior back through the RANS model in order to determine if our Bayesian procedure actually results in a better calibration. These "pushed-forward-posterior" simulations lead to a distribution

18
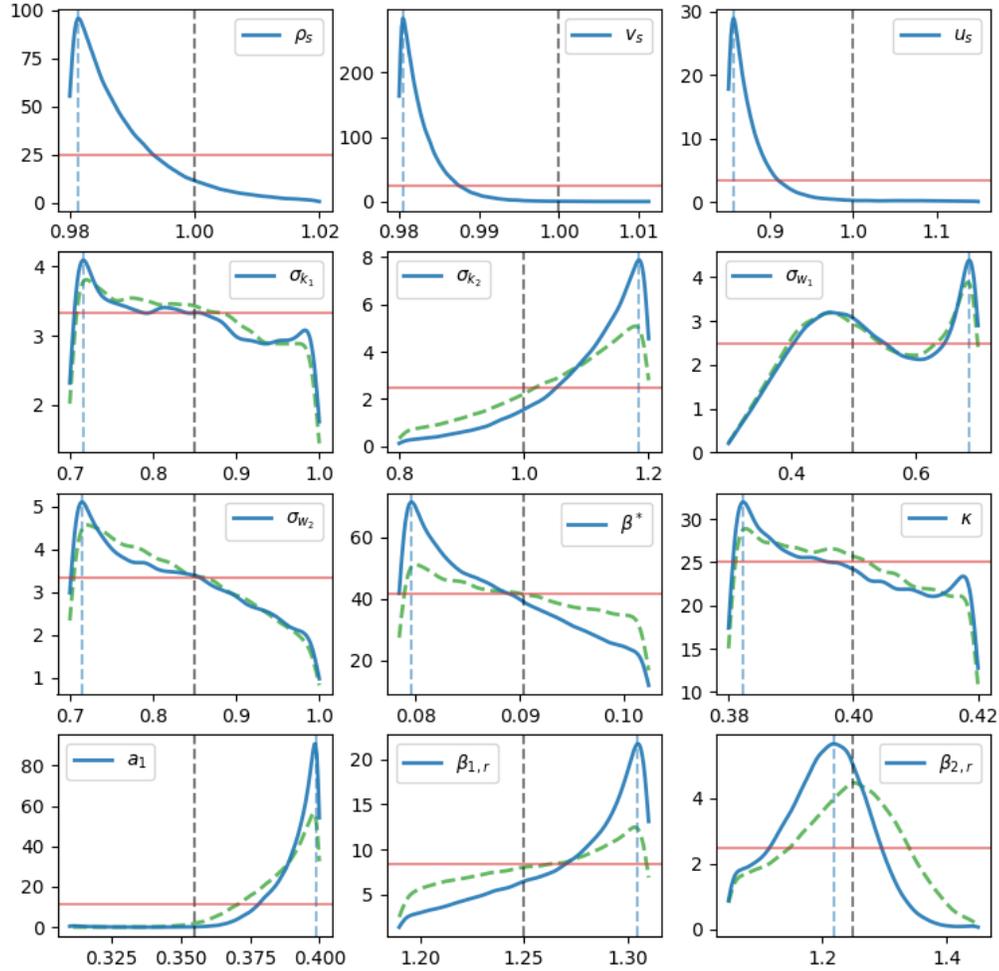
Figure 9: Marginalized posterior PDFs for the 12 features, plotted using a solid line. The uniform priors are plotted using red horizontal lines. The vertical line denotes the nominal value of the parameters. The green dashed line plots are the same except that the freestream parameters are fixed at the nominal. A one-dimensional Gaussian kde algorithm, with automatic bandwidth determination, was used to smooth out the univariate histograms.

of predicted heat fluxes and pressures, which are summarized in Fig. 10 for the heat flux predictions and Fig. 11 for the pressure. The prediction using the nominal SST model is plotted with a dashed red line, the median prediction with a solid blue line, the $[.025, .75]$ quantiles with a gray band and the $[.05, .95]$ quantiles with a blue band. The experimental

data is plotted with symbols. We see that in the region over the cone with turbulent flow (0.4 ≤ r ≤ 1.1), calibration reduces the agreement with heat flux measurements (Fig. 10) though the agreement improves in the highly instrumented separation zone at the back of the HIFiRE-1 geometry (1.5 ≤ r ≤ 1.6). A similar effect, though much more muted, occurs with pressure predictions (Fig. 11). Despite the improvement, RANS's predictive skill in the separation zone remains low. It is unlikely that any parameter (SST or freestream conditions) adjustment, within physical bounds, will drive RANS predictions closer to the experimental data. This is the definition of model-form error, caused by physics missing in the model.
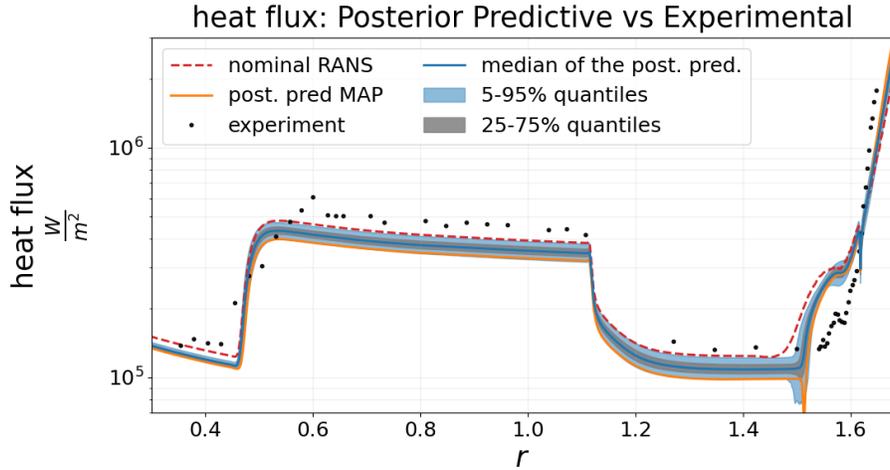
Figure 10: Comparison of heat flux predictions before and after calibration. The prediction using the nominal SST model is plotted dashed red line, the median prediction with a solid blue line, the first and third quartiles with a dashed blue line and the [.05, .95] quantiles with a dotted blue line. The experimental data is plotted with symbols. The MAP estimate is show in orange.

The improvement in predictive skill, post calibration, can be quantified using both the mean absolute error (MAE; [90, 91]) and continuous rank probability score (CRPS; [90, 91]). These can be computed for the predictions using samples picked from the posterior JPDF (as illustrated in Fig. 10 and Fig. 11) and compared to their counterparts computed using samples picked from the prior density [92]. Plots of the actual distributions (not summaries) of heat flux and pressure predictions are in Appendix C (Fig. 17 and Fig. 18). For each of the experimental data points, we compute the CRPS scores in Fig. 12. Overall, the CRPS averaged over all experimental observations is reduced by the Bayesian calibration procedure compared with the prior predictive, which is an indication of success (CRPS error is roughly the same for the the heat flux, but reduced by roughly 10% for pressure). Likewise, if we look at the MAE averaged over all experimental points in Fig. 13, we see the same result, to a slightly higher degree (a decrease in error of about 5% for the heat flux and 20% for pressure). We note that in order to average different MAE and CRPS score over different experimental points over different scales, we weigh each of the experimental data points by the inverse mean squared error of the magnitude of observations. The net
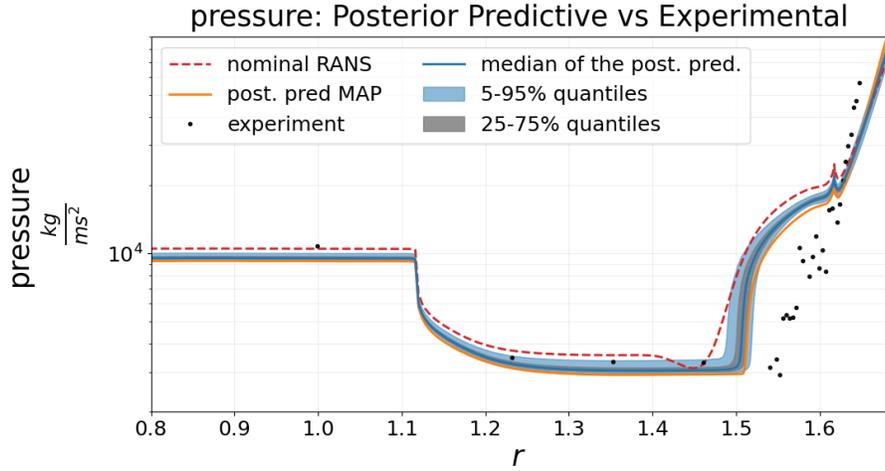
Figure 11: Comparison of pressure predictions before and after calibration. The prediction using the nominal SST model is plotted dashed red line, the median prediction with a solid blue line, the first and third quartiles with a dashed blue line and the $[.05, .95]$ quantiles with a dotted blue line. The experimental data is plotted with symbols. The MAP estimate is shown in orange.

field/ prediction	prior predictive CRPS error	posterior predictive CRPS error
heat flux	0.302	0.307
pressure	0.976	0.837

Table 5: Average CRPS errors for heat flux and pressure fields for the prior predictive and posterior predictive densities. A decrease in CRPS is preferred.

effect of this is that the errors can be interpreted as relative errors. In both cases, the calibration results seem to favor improvement of the pressure field, as opposed to the heat flux which sees almost no change between prior and posterior predictive results. This is because the model discrepancy error in the pressure field dominates the heat flux errors (see Fig. 3 which shows how the LHS runs envelope heat flux better than pressure). A summary of the average CRPS and MAE errors for the prior predictive versus the posterior predictive are shown in Table 5 and Table 6.

field/ prediction	prior predictive MAE error	posterior predictive MAE error
heat flux	0.366	0.342
pressure	1.072	0.878

Table 6: Average MAE errors for heat flux and pressure fields for the prior predictive and posterior predictive densities. A decrease in MAE is preferred.
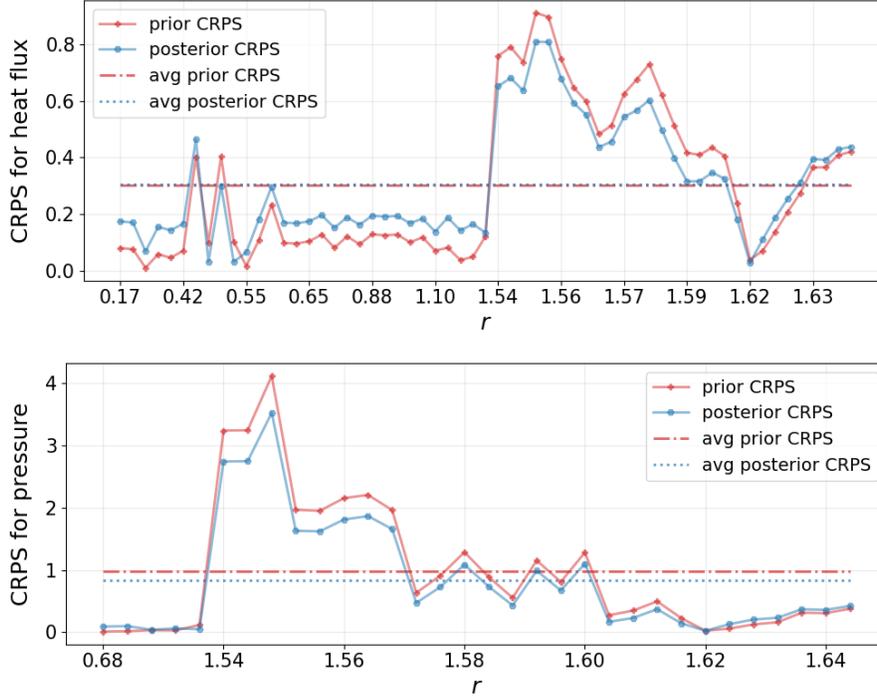
Figure 12: Continuous rank probability score (CRPS) for posterior predictive versus the prior predictive for heat flux (top) and pressure (bottom). The dotted horizontal lines represent the average CRPS scores at the different observation points, r_i^{Exp}'s. The average CRPS scores are simply the uniform average over the different observed data points.

6.2 Discussion

The marginal posterior PDFs in Fig. 9 (solid lines) show that only a handful of SST parameters can be estimated well from the heat flux and pressure measurements. Since the heat flux and pressure measurements depend strongly on freestream quantities, there is always a doubt whether the difficulty in estimating SST parameters could be due to the uncertainties in the freestream quantities. Therefore, we reran the Bayesian procedure while the freestream quantities were held at their nominal values i.e., $\rho_s = v_s = u_s = 1$. The marginalized posterior PDFs so obtained are plotted in Fig. 9 using a green dashed line. There is not a big change in the posterior PDFs, indicating that the inclusion of the freestream quantities did not negatively impact the estimation problem. This is probably due to the very narrow priors $\mathcal{U}(0.98, 1.02)$ used for ρ_s and v_s. Temperature, as expected, had no effect on the estimation of SST parameters. The prior and posterior for σ_{k_1} are about the same, indicating that there is little information about it in the observations. Note that sensors are densely clustered in the separation zone and the parameters' PDFs are strongly influenced by it.

We also notice that the peaks of the parameters' PDFs are near their prior bounds. This a classic sign of large model-form errors. The Bayesian calibration drives the parameters to extremes in an effort to match experimental data, and in the process, adjust parameters
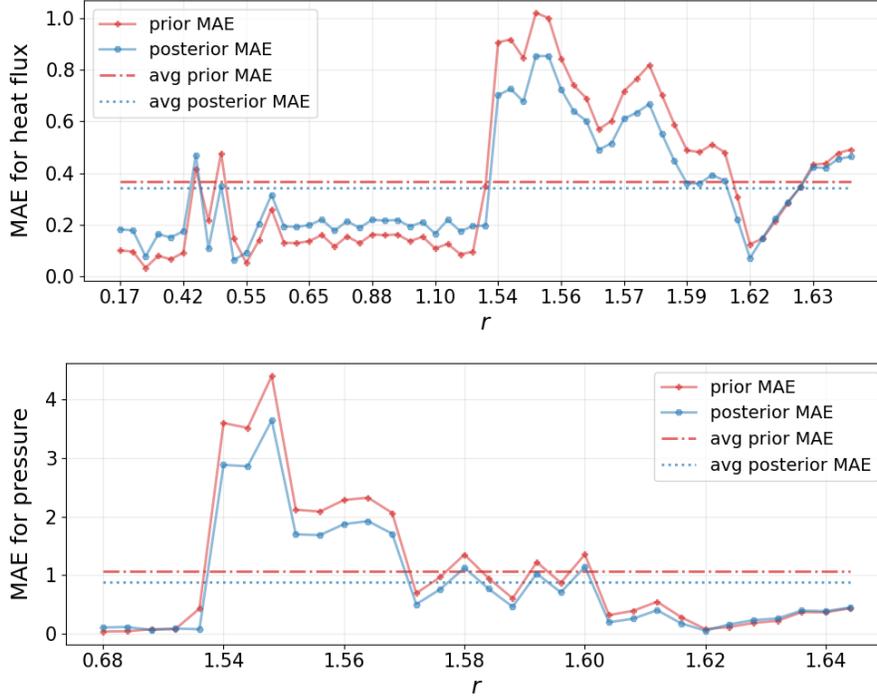
Figure 13: Mean absolute errors (MAE) for posterior predictive versus the prior predictive for heat flux (top) and pressure (bottom). The dotted horizontal lines represent the average MAE scores.

to compensate for model-form errors, the primary one being RANS's inability to model separation zones correctly. However, the parameters are expected to stay within physical bounds, causing the most promising values to be drawn from the extremities of their range (and hence the peak in the PDF). One could presumably expand the priors in an effort to match experimental data. However, for SST parameters, they would then violate the range of values observed in literature. For the freestream parameters, it would require us to assume that the measurement error in LENS-I are far larger than usually stated. That is unlikely - LENS-I is very well-characterized facility.

If we ignore the PDFs (in Fig. 9) that show peaks only at their extremities (since they are adversely affected by model-form errors), we find that σ_{w1} and $\beta_{2,r}$ can be estimated quite well; they show a well-defined peak within the support of their prior. Some confirmation of this behavior can be drawn from Ref. [93] where the authors performed a global sensitivity analysis of the SST turbulence model. They used a double ellipsoid ($M = 7.8$, unit $Re = 2.2 \times 10^7$/m) and the X-33 launch vehicle ($M = 7.4$, unit $Re = 1.64 \times 10^7$/m) as their test cases. These flow conditions are similar to our HIFiRE-1 test case ($M = 7.16$, unit $Re = 1.02 \times 10^7$/m), as are the characteristics of the flow, especially the existence of flow separation and reattachment. The studies found that σ_{w1} was by far the most influential parameter, as it appears very prominently in the equation for ω, the specific rate of turbulent dissipation. Since ω controls the rate of conversion of turbulent kinetic energy into heat, and since

23

the wall heat flux is measured and used as calibration data in our estimation process, it is not surprising that σ_{w1} can be estimated well, even in the presence of model-form errors. In addition the σ_{w1} PDF peaks close to the nominal value, providing some degree of confidence in it. $\beta_{2,r}$, the other parameter that we can estimate stably, was not found to be very influential in Ref. [93]. However, it also appears in the $\omega-$equation (and only there) and is informed by the heat flux. The heat flux undergoes a large change when the flow transitions to turbulence and again at the junction of the cone and cylindrical mid-section (which contains an expansion fan), and these variations likely help constrain the PDFs of the parameters that are strongly informed by the heat flux. The model-form errors in these regions are much smaller than those observed near the separation zone, which allows the PDFs for σ_{w1} and $\beta_{2,r}$ to emerge from the data, clearly distinguished from the prior.

The sharpness of the PDFs for the SST parameters could perhaps be improved by removing some of the hard-to-estimate SST parameters (i.e., the ones whose posterior and prior PDFs do not differ significantly). Such an exercise is left for future work, but it would, undoubtedly, require the use of surrogate models (Sec. 4) to perform global sensitivity analysis (GSA) to choose the influential subset of SST parameters. Fig. 16 in the Appendix C shows, via pair plots, that the SST parameters are not very correlated in the 12-dimensional JPDF, indicating that the removal of less influential SST parameters will not adversely affect the performance of the calibrated SST model.[10]

Fig. 10 and Fig. 11 show the effect of model-form errors in the SST model which prevent it from modeling the separation zone accurately. The net effect of calibration is to improve the prediction of pressure and heat flux in the separation zone while degrading it elsewhere. The large number of measurements in the separation zone also contributed to the outsized importance of this zone during calibration. It may be possible to obtain an arguably better SST model by removing all measurements from the separation zone i.e. $r \gtrsim 1.5$. While such a model would not be very predictive in the separation zone, it would be highly accurate over the cone and the cylindrical sections $0.4 \lesssim r \lesssim 1.5$ which accounts for a large fraction of the heating of the HIFiRE-1 geometry. The poor agreement of RANS versus experiments, even after calibration, in the separation zone ($1.4 \leq r \leq 1.7$) is expected. RANS's model-form errors do not allow it to model separation zone well, and the calibration resulted in predictions marginally better than the nominal model. However, the MCMC calibration yielded signatures (parameter PDFs peaked at the lower or upper bounds) that allow us to diagnose the data - model disagreement as model-form errors, rather than an issue that could be ameliorated with more sensors or an error model more sophisticated than the Gaussian used in this paper.

Note that in this study we have not used the enthalpy of the incoming flow and its static pressure, both of which were measured in the HIFiRE-1 experiment. This is because they only help with estimation of the freestream quantities and carry no information at all about the turbulence model.

The preceding paragraphs reveal some very useful and practical information about hypersonic turbulence and shock-tunnel experiments. The freestream uncertainties in the HIFiRE-1 experiment were inconsequential to the turbulence model calibration. The information content in the heat flux and pressure measurements on the HIFiRE-1 geometry is

---

[10]The ability to easily perform sensitivity analysis using Sobol indices is another motivation for using PCEs, from which Sobol indices can be easily extracted due to their orthogonal basis representation.

limited and informs only a few turbulence model parameters. However, it is the RANS's model-form uncertainties that are the calibration bottleneck. These findings do not exist in fluid dynamics literature for realistic hypersonic flows in engineering geometries, and were made possible only by our ability to construct surrogate models of fields encountered in hypersonic turbulent flows and use them within a Bayesian inference framework. The closest equivalent to our work is Ref. [75] which targets a Mach 6.1 hypersonic turbulent flow over a flat plate.

7 Summary and conclusions

In this paper we formulated a surrogate which combines projection-based model reduction techniques with machine learning regressors for the prediction of scalar-valued fields. We used principal component analysis to perform the dimension reduction and then explored a variety of different machine learning regressors to fit the projection coefficients of the learned components. We experimented with Gaussian process regression, polynomial chaos expansions, random forests, kernel ridge regression, support vector machines, and multi-layer perceptron models. In order to tune each scalar regressor over a given set of hyperparameters and perform model selection, k-fold cross validation was used. Ultimately, for the final surrogate, a multivariate polynomial representation, i.e., a polynomial chaos expansion with multivariate Legendre polynomials, was chosen to fit our reduced space projection coefficients. The ML experiments showed that polynomials provided the greatest amount of expressivity and accuracy, while being the easiest and simplest to train. They also provided the most consistent answers (e.g., without dependence on say random seeds like MLPs) which is critical for reproducibility. We demonstrated the efficacy and accuracy of these surrogates for predicting the heat flux and pressure fields on the surface of the HIFiRE geometry in a Mach 7.16 turbulent flow using $m = 2500$ simulation runs. The surrogate was then used in a first-ever Bayesian calibration of the HIFiRE experiment, using an adaptive MCMC method to construct a joint density. The posterior predictive samples from the JPDF matched the experiment data better than the prior predictive samples for both heat flux and the pressure fields in terms of both the CRPS (continuous rank probability score) and MAE (mean absolute error), thus resulting in an improved predictive model and reduced mismatch between prediction and experimental data.

The Bayesian calibration of the SST model parameters was also able to construct posterior PDFs, compare them with the prior and discern which SST model parameters could be estimated well from the heat flux and pressure measurements. We discovered that the range over which the freestream quantities were controlled in the LENS-I shock-tunnel during the HIFiRE-1 experiment was sufficiently narrow that the uncertainty did not impact the turbulence model estimation problem. The limiting factor was the model-form error in the RANS model which made it infeasible to capture the separation zone at the extreme aft of the test geometry. Further improvement in the calibrated model may be had by performing a GSA to pick the most sensitive SST model parameters and calibrating them to the same dataset. These findings are novel and were made possible by the numerical and statistical tools developed in this paper. In addition, the same tools can be used to perform the GSA.

Last, but not least, we have provided software, *tesuract* (Tensor Surrogate Automation and Computation), to build the types of surrogates used in this paper. *tesuract* is built on

25

top of the scikit-learn API [85] and utilizes scikit-learn's vast library of machine learning estimators, model selection techniques, and dimension reduction methods in a unique ML pipeline which allows the construction of surrogates for both single target scalar outputs and scalar-valued fields (i.e., multi-target correlated outputs). This allows flexibility, utility and easy-of-use for many applications related to surrogate construction. This software is freely available on github (https://github.com/kennychowdhary/tesuract).

8   Funding and Acknowledgment

A   Grid resolution studies for HIFiRE-1

A solution verification study was performed for HIFiRE-1 at the flow conditions used in this paper. Three axisymmetric structured meshes were made, with $256 \times 512, 512 \times 1024$, and 1024×2048 cells in the wall normal $\times$ streamwise directions. The coarser meshes were constructed from the finer by removing every other mesh node, preserving regular refinement. The solution verification procedure produces numerical error estimates for the heat flux and the pressure on the wall at locations where the corresponding experimental diagnostics were installed. The solution values on each mesh are interpolated to these locations, and an extrapolated value at each location is computed from the well-known order-of-accuracy equation. Second order accuracy is assumed, which holds except possibly for isolated locations. Fig. 14 (top) shows the heat flux for the solutions on the three meshes as lines and the extrapolated solution at points. We see that in the transition region (Fig. 14 (top left)) the solutions from the 3 meshes are visually indistinguishable, but the extrapolated heat fluxes (under a second order assumption) do not agree with any of the numerical solutions. In the aft region (Fig. 14 (top right)) a similar issue is seen near the separation region abutting the flare. Elsewhere, the numerical simulations on the three meshes and the extrapolated values are visually indistinguishable. Recall that the heat flux on the flare is highly sensitive to the separation bubble size and the consequent shock structure. Fig. 14 (bottom) presents corresponding pressure data. Again, the only visible differences are on the flare, for the same reasons as for the heat flux. Since the extrapolated values from the 256×512 and 512×1024 meshes agree with what is achieved on the 1024×2048 mesh, we see that SPARC is largely achieving second-order behavior.
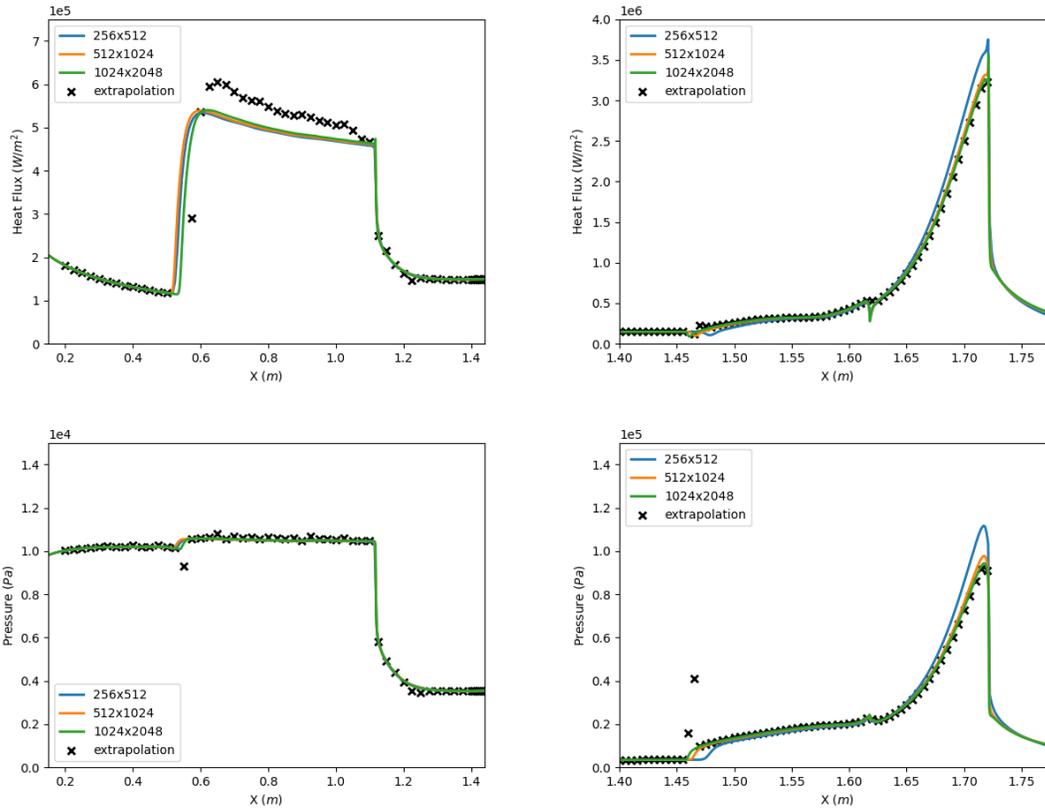
Figure 14: Top: Heat flux profiles computed on 256×512, 512×1024 and 1024×2048 meshes. The symbols are values of heat flux at the sensor locations, extrapolated from 256×512 and 512×1024, meshes to the 1024×2048 mesh, assuming second-order accuracy. The left figure shows the results upstream of the separation point while the right figure shows results downstream of separation. Bottom: The same, but done for the pressure field.

B   Final ROM machine learning models

Here we discuss the hyper-parameter tuning procedure and the final ROM-based model architectures for both heat flux and pressure fields. Recall that we have chosen four PCA components for our reduced order model. The corresponding projection coefficients associated for each of the four components is then fit with six different machine learning regressor models, and each of these regressors is hyper-parameter tuned over a set of possible parameter combinations. These parameter combinations are listed below, of which more detail can be found in the documentation of our surrogate construction software *tesuract* and sklearn's website. A 5-fold cross validation score was computed for every single combination of grid values.

```
# polynomial chaos regressor
pce_grid = {
    'order': list(range(1,11)),
```

27

```
723    'mindex_type': ['total_order','hyperbolic'],
724    'fit_type': ['LassoCV','ElasticNetCV','linear']}
725
726  # random forest regressor
727  rf_grid = {
728    'n_estimators': [200,500,1000,5000],
729    'max_features': [3,'sqrt','auto'],
730    'max_depth': [5,10,50]}
731
732  # multi-layer perceptron regressor
733  mlp_grid = {
734    'hidden_layer_sizes': [(50,),(50,)*2,(50,)*4,(50,)*6
735                           (100,),(100,)*2,(100,)*4,(100,)*6
736                           (500,),(500,)*2,(500,)*4,(500,)*6],
737    'solver': ['lbfgs','adam','sgd'],
738    'activation': ['relu'],
739    'max_iter': [2500],
740    'batch_size': ['auto'],
741    'learning_rate': ['invscaling'],
742    'alpha': [1e-4,1e-6,1e-2],
743    'tol': [1e-6,1e-4],
744    'random_state': [0,99,324]}
745
746  # kernel ridge regression regressor
747  krr_grid = {
748    'kernel': ['polynomial'],
749    'kernel_params': [{'degree':1},{'degree':2},{'degree':3},{'degree':4}],
750    'alpha': [1e-4,1e-2,1e-1,1.0]}
751
752  # gaussian process regressor
753  gpr_grid = {
754    'kernel': [1.0 * RBF(.1) + .1**2 *WhiteKernel(.1),
755               1.0 * RBF(.1) + .1**2 *WhiteKernel(.1) + 1.0*DotProduct(.1),
756               1.0 * Matern(length_scale=.1, nu=1.5) + .1**2 *WhiteKernel(.1)],
757    'alpha': [1e-10],
758    'optimizer': ['fmin_l_bfgs_b'],
759    'n_restarts_optimizer': [2],
760    'random_state': [0]}
761
762  # k-nearest neighbor regressor
763  knn_grid = {
764    'n_neighbors': (1,5,8,10),
765    'leaf_size': (20,30,40,1),
766    'p': (1,2),
767    'weights': ('uniform', 'distance'),
```

```
768    'metric': ('minkowski', 'chebyshev')}
769
770  # support vector machine regressor
771  svr_grid = {
772    'kernel': ('linear','poly', 'rbf', 'sigmoid'),
773    'degree': (2,4,8),
774    'gamma': ('scale','auto'),
775    'C': (1,5,10)}
```

To give an example, consider the the polynomial chaos regressor parameter grid above. A 5-fold cross validation score was computed for each and every combination of the `order`, the polynomial total degree, `mindx_type`, which controls the number of interaction terms, and `fit_type`, the algorithm for solving the least squares problem. There are 60 parameter combinations and, thus, a total of 300 PCE fits were computed (five for each of the 60 parameter combinations since we are using five-fold cross-validation) for each of the four components. This method was repeated for each of the regressors listed above in order to obtain the penultimate model for both heat flux and pressure fields. Thus, each model comparison involves hundreds or thousands of ML regression fits, all of which is handled neatly and efficiently within the tesuract and the sklearn framework so that the user does not need to bother with the cumbersome nesting and splitting of the test and train data.

The PCE and MLP models had the highest cross-validation scores among the regressors. For heat flux, the PCE model with the highest cross-validation score had polynomial orders of degrees $\{2, 2, 4, 4\}$ for each of the four components, and for pressure the PCE model with the highest cross-validation score had polynomial orders of $\{4, 4, 4, 4\}$. In contrast, the best MLP network for each projection coefficient had 4 hidden layers of 50 nodes each for both the pressure and heat flux fields. The network used rectified linear units for the activation functions, a tolerance of 10^{-6} for the LBFGS solver, and an inverse scaling for the learning rate, which gradually decreases as the time step progresses. The rest of the parameters were set to their default values [85].

C   Bayesian calibration results

The adaptive MCMC algorithm described in Sec. 6 yielded a chain of that was thinned to reduce the autocorrelation in the sequence of samples. The autocorrelation vs lag time is plotted in Fig. 15. [11]

The full JPDF, illustrated as a matrix of pair plots, is shown in Fig. 16. While Fig. 9 shows the SST parameters that could be estimated from the HIFiRE-1 measurements, Fig. 16 shows the correlations that exist between the various SST parameters in the posterior JPDF. We see that the correlations are mild i.e., the structures in the 2D plots are mostly horizontally or vertically aligned. This is fortunate as it implies that the SST parameters that cannot be inferred well (i.e., where prior and posterior PDFs in Fig. 9 are similar) can

---

[11]The aggressive thinning accounted for the variability in the autocorrelation amongst the different parameters and was chosen so that all twelve parameters has sufficiently converged according to the autocorrelation lag plot in Fig. 15.
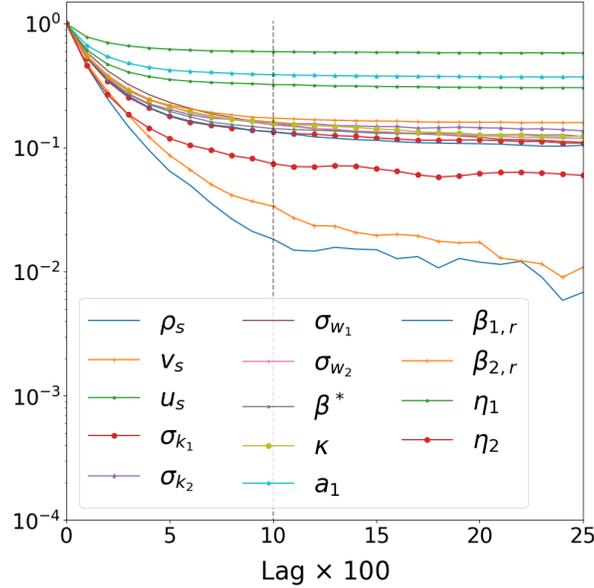
Figure 15: Autocorrelation time for each of the twelve calibration parameters, including the two noise parameters, η_1, η_2. The horizontal dotted line represents the amount by which we thin the chain, giving us an effective sample size of roughly 24k samples in our twelve-dimensional sample space (plus two noise parameters, i.e., Γ distributed random variables η_1 and η_2 representing the inverse variance of the model discrepancy errors).

be simply removed to yield a smaller estimation problem without materially (negatively) impacting the accuracy of the SST turbulence model.

The posterior predictive densities, computed by simulating the HIFiRE-1 experiments with **x** drawn from the posterior JPDF (as plotted in Fig. 16 and marginalized in Fig. 9) are plotted in Fig. 17 (heat flux) and Fig. 18 (pressure) in blue. The prior predictive are plotted in red. These predictive densities were used to compute the CRPS and MAE in Fig. 12 and Fig. 13. These plots are colloquially known as *joy plots*. The y-axis represents the locations of the observed data point and the x-axis represents the log values of the heat flux and/or pressure.
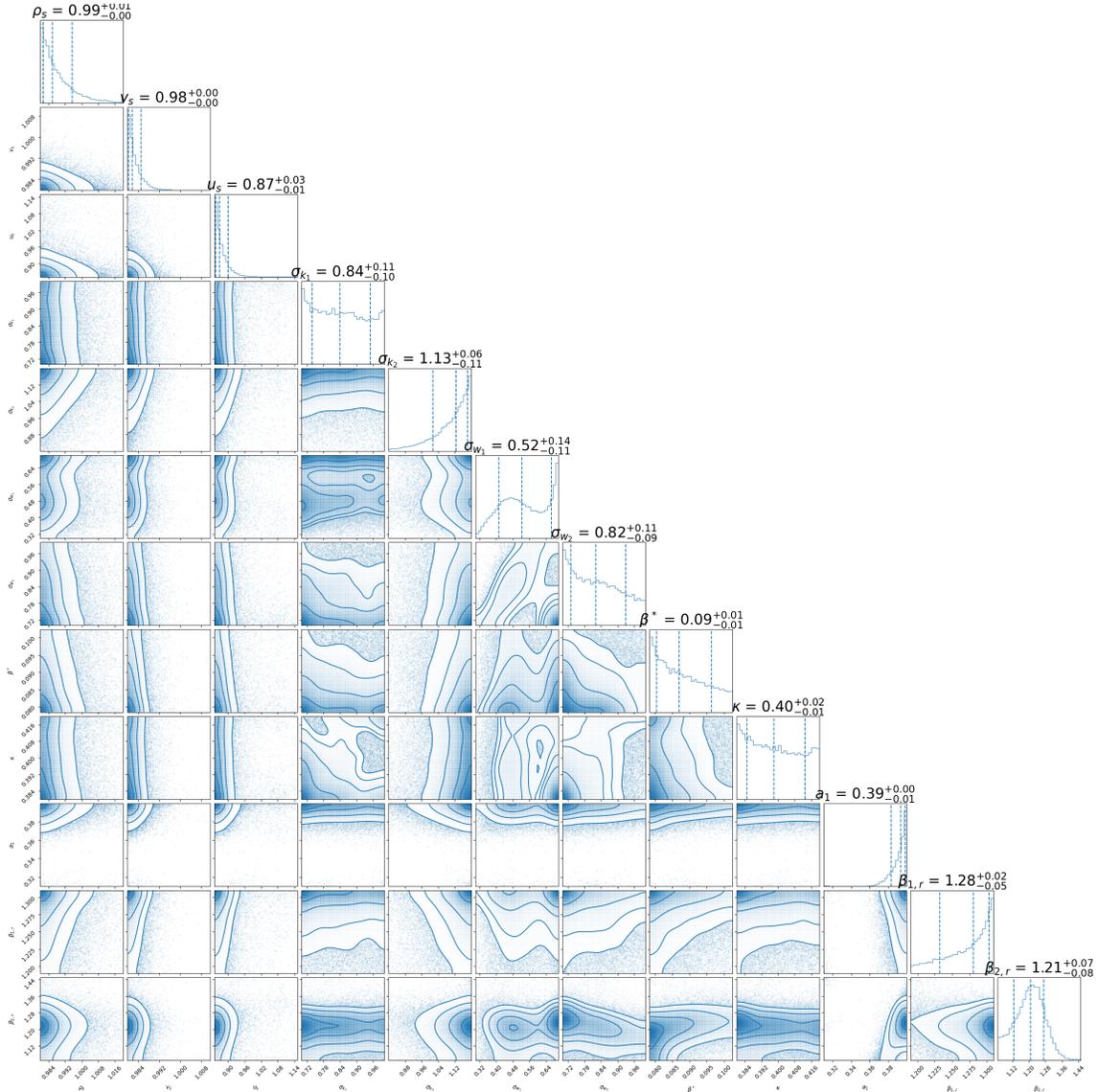
Figure 16: Corner (pair) plot for the posterior distribution over the twelve calibration or tuning parameters. The maximum *a posteriori* estimate is displayed above the univariate plots. Given that the priors are chosen to be uniform, all parameters are informed, to some extent, from the observed data.
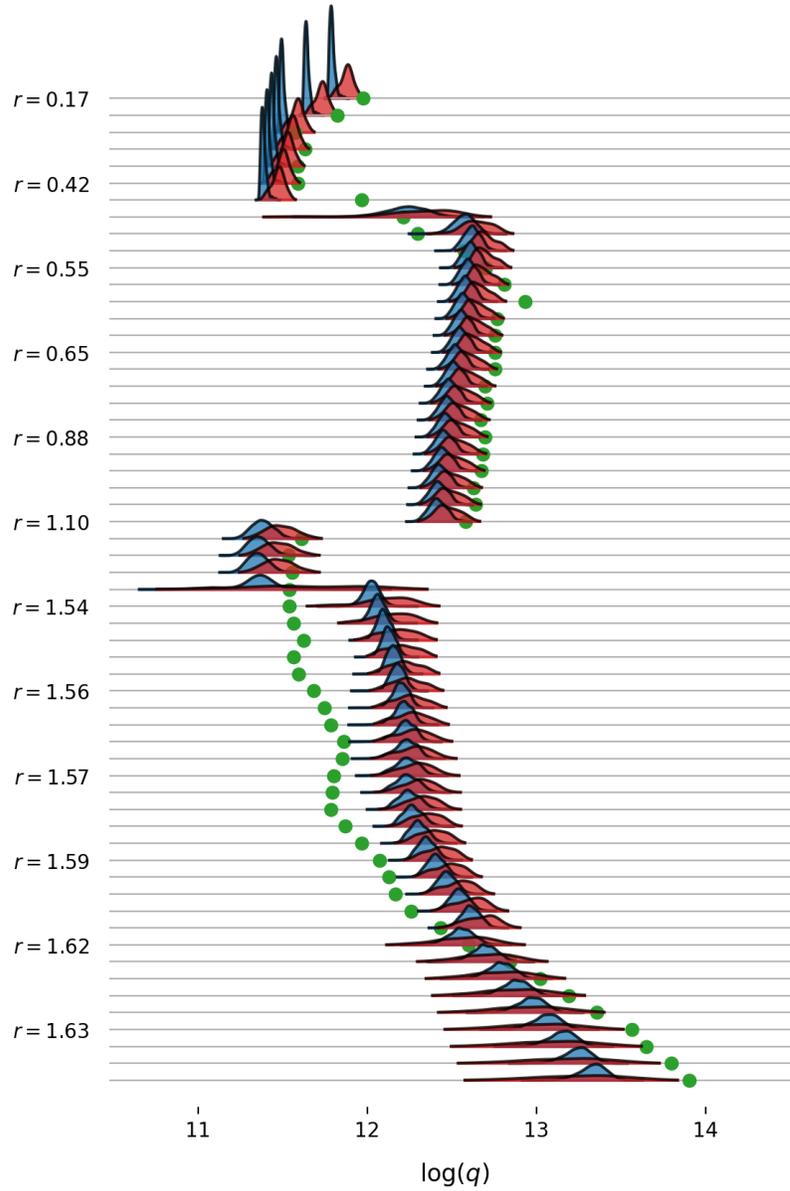
Figure 17: Prior versus the posterior predictive distributions for log-scaled heat flux, denoted by $\log(q)$. r-axis is displayed on the vertical, the blue histogram represents the posterior predictive, while the red shows the prior. The green dot shows the single observed data.
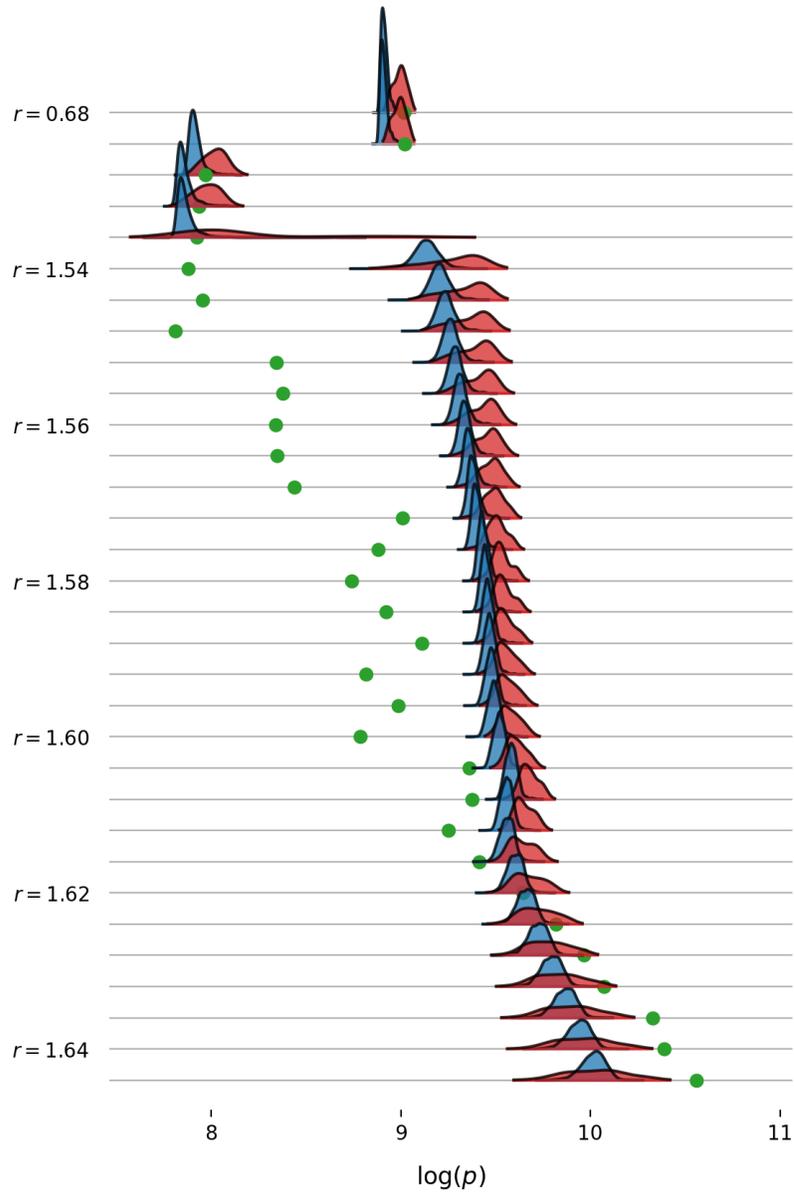
Figure 18: Prior versus the posterior predictive distributions for log-scaled pressure, denoted by $\log(p)$. r-axis is displayed on the vertical, the blue histogram represents the posterior predictive, while the red shows the prior. The green dot shows the single observed data.

References

[1] Alizadeh, R., Allen, J. K., and Mistree, F., "Managing computational complexity using surrogate models: a critical review," *Research in Engineering Design*, Vol. 31, 2020, pp. 275–298.

[2] Yondo, R., Andrés, E., and Valero, E., "A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses," *Progress in Aerospace Sciences*, Vol. 96, 2018, pp. 23–61.

[3] Sudret, B., Marelli, S., and Wiart, J., "Surrogate models for uncertainty quantification: An overview," *2017 11th European Conference on Antennas and Propagation (EUCAP)*, 2017, pp. 793–797.

[4] Asher, M. J., Croke, B. F. W., Jakeman, A. J., and Peeters, L. J. M., "A review of surrogate models and their application to groundwater modeling," *Water Resources Research*, Vol. 51, No. 8, 2015, pp. 5957–5973.

[5] Kou, J. and Zhang, W., "Data-driven modeling for unsteady aerodynamics and aeroelasticity," *Progress in Aerospace Sciences*, Vol. 125, 2021, pp. 100725.

[6] Swischuk, R., Mainini, L., Peherstorfer, B., and Willcox, K., "Projection-based model reduction: Formulations for physics-based machine learning," *Computers and Fluids*, Vol. 179, 2019, pp. 704–717.

[7] Wadhams, T. P., Mundy, E., MacLean, M. G., and Holden, M. S., "Ground test studies of the HIFiRE-1 transition experiment Part 1: Experimental results," *Journal of Spacecraft and Rockets*, Vol. 45, No. 6, 2008, pp. 1134–1148.

[8] MacLean, M., Wadhams, T., Holden, M., and Johnson, H., "Ground test studies of the HIFiRE-1 transition experiment Part 2: Computational analysis," *Journal of Spacecraft and Rockets*, Vol. 45, No. 6, 2008, pp. 1149–1164.

[9] Wilcox, D. C., *Turbulence Modeling for CFD*, D C W Industries, La Canada, CA, USA, 3rd ed., 2006.

[10] Menter, F. R., "Two-equation eddy-viscosity turbulence models for engineering applications," *AIAA Journal*, Vol. 32, No. 8, August 1994, pp. 1598–1605.

[11] Guillas, S., Glover, N., and Malki-Epshtein, L., "Bayesian calibration of the constants of the k-ϵ turbulence model for a CFD model of street canyon flow," *Computer Methods in Applied Mechanics and Engineering*, Vol. 279, 2014, pp. 536–553.

[12] Ray, J., Dechant, L., Lefantzi, S., Ling, J., and Arunajatesan, S., "Robust Bayesian calibration of a k-ϵ model for compressible jet-in-crossflow simulations," *AIAA Journal*, Vol. 56, No. 12, 2018, pp. 4893–4909.

[13] Zhang, J. and Fu, S., "An efficient approach for quantifying parameter uncertainty in the SST turbulence model," *Computers and Fluids*, Vol. 181, 2019, pp. 173–187.

[14] Viana, F. A. C. and Subramaniyan, A. K., "A Survey of Bayesian Calibration and Physics-informed Neural Networks in Scientific Modeling," *Archives of Computational Methods in Engineering*, Vol. 28, 2021, pp. 3801–3830.

[15] McPhee, J. and Yeh, W. W.-G., "Groundwater Management Using Model Reduction via Empirical Orthogonal Functions," *Journal of Water Resources Planning and Management*, Vol. 134, No. 2, 2008, pp. 161–170.

[16] Dunbar, W. S. and Woodbury, A. D., "Application of the Lanczos Algorithm to the solution of the groundwater flow equation," *Water Resources Research*, Vol. 25, No. 3, 1989, pp. 551–558.

[17] Woodbury, A. D., Dunbar, W. S., and Nour-Omid, B., "Application of the Arnoldi Algorithm to the solution of the advection-dispersion equation," *Water Resources Research*, Vol. 26, No. 10, 1990, pp. 2579–2590.

[18] Willcox, K. and Megretski, A., "Fourier Series for Accurate, Stable, Reduced-Order Models in Large-Scale Linear Applications," *SIAM Journal on Scientific Computing*, Vol. 26, No. 3, 2005, pp. 944–962.

[19] Gugercin, S. and Willcox, K., "Krylov projection framework for Fourier model reduction," *Automatica*, Vol. 44, No. 1, 2008, pp. 209–215.

[20] Ly, H. V. and Tran, H. T., "Modeling and control of physical processes using proper orthogonal decomposition," *Mathematical and computer modelling*, Vol. 33, No. 1-3, 2001, pp. 223–236.

[21] Higdon, D., Gattiker, J., Williams, B., and Rightley, M., "Computer model calibration using high-dimensional output," *Journal of the American Statistical Association*, Vol. 103, No. 482, 2008, pp. 570–583.

[22] Audouze, C., De Vuyst, F., and Nair, P., "Reduced-order modeling of parameterized PDEs using time–space-parameter principal component analysis," *International journal for numerical methods in engineering*, Vol. 80, No. 8, 2009, pp. 1025–1057.

[23] Audouze, C., De Vuyst, F., and Nair, P. B., "Nonintrusive reduced-order modeling of parametrized time-dependent partial differential equations," *Numerical Methods for Partial Differential Equations*, Vol. 29, No. 5, 2013, pp. 1587–1628.

[24] Wirtz, D., Karajan, N., and Haasdonk, B., "Surrogate modeling of multiscale models using kernel methods," *International Journal for Numerical Methods in Engineering*, Vol. 101, No. 1, 2015, pp. 1–28.

[25] Mainini, L. and Willcox, K., "Surrogate modeling approach to support real-time structural assessment and decision making," *AIAA Journal*, Vol. 53, No. 6, 2015, pp. 1612–1626.

[26] Ulu, E., Zhang, R., and Kara, L. B., "A data-driven investigation and estimation of optimal topologies under variable loading configurations," *Computer Methods in*

888 *Biomechanics and Biomedical Engineering: Imaging & Visualization*, Vol. 4, No. 2,
889 2016, pp. 61–72.

890 [27] Hesthaven, J. S. and Ubbiali, S., "Non-intrusive reduced order modeling of nonlinear
891 problems using neural networks," *Journal of Computational Physics*, Vol. 363, 2018,
892 pp. 55–78.

893 [28] Maulik, R., Mohan, A., Lusch, B., Madireddy, S., Balaprakash, P., and Livescu, D.,
894 "Time-series learning of latent-space dynamics for reduced-order model closure," *Phys-*
895 *ica D: Nonlinear Phenomena*, Vol. 405, Apr 2020, pp. 132368.

896 [29] Rahman, S. M., Pawar, S., San, O., Rasheed, A., and Iliescu, T., "Nonintrusive reduced
897 order modeling framework for quasigeostrophic turbulence," *Phys. Rev. E*, Vol. 100,
898 Nov 2019, pp. 053306.

899 [30] Wang, Z., Xiao, D., Fang, F., Govindan, R., Pain, C. C., and Guo, Y., "Model iden-
900 tification of reduced order fluid dynamics systems using deep learning," *International*
901 *Journal for Numerical Methods in Fluids*, Vol. 86, 2018, pp. 255–268.

902 [31] Mo, S., Zhu, Y., Zabaras, N., Shi, X., and Wu, J., "Deep Convolutional Encoder-
903 Decoder Networks for Uncertainty Quantification of Dynamic Multiphase Flow in Het-
904 erogeneous Media," *Water Resources Research*, Vol. 55, No. 1, 2019, pp. 703–728.

905 [32] Lee, K. and Parish, E. J., "Parameterized neural ordinary differential equations: appli-
906 cations to computational physics problems," *Proceedings of the Royal Society A: Math-*
907 *ematical, Physical and Engineering Sciences*, Vol. 477, No. 2253, 2021, pp. 20210162.

908 [33] Chen, W., Hesthaven, J. S., Junqiang, B., Qiu, Y., Yang, Z., and Tihao, Y., "Greedy
909 nonintrusive reduced order model for fluid dynamics," *AIAA Journal*, Vol. 56, No. 12,
910 2018, pp. 4927–4943.

911 [34] Fang, J., Sun, G., Qiu, N., Kim, N. H., and Li, Q., "On design optimization for
912 structural crashworthiness and its state of the art," *Structural and Multidisciplnary*
913 *Optimization*, Vol. 55, 2017, pp. 1091–1119.

914 [35] Dey, S., Mukhopadhyay, T., and Adhikari, S., "Metamodel based high-fidelity stochas-
915 tic analysis of composite laminates: A concise review with critical comparative assess-
916 ment," *Composite Structures*, Vol. 171, 2017, pp. 227–250.

917 [36] Laurent, L., Riche, R. L., Soulier, B., and Boucard, P., "An Overview of Gradient-
918 Enhanced Metamodels with Applications," *Archive of Computational Methods in En-*
919 *gineering*, Vol. 26, 2019, pp. 61–106.

920 [37] Swischuk, R., Mainini, L., Peherstorfer, B., and Willcox, K., "Projection-based model
921 reduction: Formulations for physics-based machine learning," *Computers & Fluids*,
922 Vol. 179, 2019, pp. 704–717.

923 [38] Cao, C., Nie, C., Pan, S., Cai, J., and Qu, K., "A constrained reduced-order method for
924 fast prediction of steady hypersonic flows," *Aerospace Science and Technology*, Vol. 91,
925 2019, pp. 679–690.

[39] Dreyer, E. R., Grier, B. J., McNamara, J. J., and Orr, B. C., "Rapid Steady-State Hypersonic Aerothermodynamic Loads Prediction Using Reduced Fidelity Models," *Journal of Aircraft*, Vol. 58, No. 3, 2021, pp. 663–676.

[40] Chen, X., Liu, L., Long, T., and Yue, Z., "A reduced order aerothermodynamic modeling framework for hypersonic vehicles based on surrogate and POD," *Chinese Journal of Aeronautics*, Vol. 28, No. 5, 2015, pp. 1328–1342.

[41] Chen, X., Zuo, G., Shi, Y., and Liu, L., *An Efficient Integrated Aerothermoelasticity Analysis System Based on Surrogate-based Reduced Order Modeling for Hypersonic Vehicles*, 2017.

[42] Chen, Z. and Zhao, Y., "Aerothermoelastic Analysis of a Hypersonic Vehicle Based on Thermal Modal Reconstruction," *International Journal of Aerospace Engineering*, Vol. 2019, 2019, Article ID 8384639.

[43] Crowell, A. R. and McNamara, J. J., "Model Reduction of Computational Aerothermodynamics for Hypersonic Aerothermoelasticity," *AIAA Journal*, Vol. 50, No. 1, 2012, pp. 74–84.

[44] Xiaoxuan, Y., Jinglong, H., Bing, Z., and Haiwei, Y., "Model reduction of aerothermodynamic for hypersonic aerothermoelasticity based on POD and Chebyshev method," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, Vol. 233, No. 10, 2019, pp. 3734–3748.

[45] Zhang, K., Yao, J., He, Z., Xin, J., and Fan, J., "Probabilistic Transient Heat Conduction Analysis Considering Uncertainties in Thermal Loads Using Surrogate Model," *Journal of Spacecraft and Rockets*, Vol. 58, No. 4, 2021, pp. 1030–1042.

[46] Vollant, A., Balarac, G., and Corre, C., "Subgrid-scale scalar flux modelling based on optimal estimation theory and machine-learning procedures," *Journal of Turbulence*, Vol. 18, No. 9, 2017, pp. 854–878.

[47] Matai, R. and Durbin, P., "Large-eddy simulation of turbulent flow over a parametric set of bumps," *Journal of Fluid Mechanics*, Vol. 866, 2019, pp. 503–525.

[48] Duraisamy, K., Iaccarino, G., and Xiao, H., "Turbulence Modeling in the Age of Data," *Annual Review of Fluid Mechanics*, Vol. 51, No. 1, 2019, pp. 357–377.

[49] Xiao, H. and Cinnella, P., "Quantification of model uncertainty in RANS simulations: A review," *Progress in Aerospace Sciences*, Vol. 108, 2019, pp. 1–31.

[50] Zhang, X., Wu, J., Coutier-Delgosha, O., and Xiao, H., "Recent progress in augmenting turbulence models with physics-informed machine learning," *Journal of Hydrodynamics*, Vol. 31, 2019, pp. 1153–1158.

[51] Ling, J., Kurzawski, A., and Templeton, J., "Reynolds-averaged turbulence modelling using deep neural networks with embedded invariance," *Journal of Fluid Mechanics*, Vol. 807, 2016, pp. 155–166.

[52] Zhang, Y., Dwight, R. P., Schmelzer, M., Gómez, J. F., hua Han, Z., and Hickel, S., "Customized data-driven RANS closures for bi-fidelity LES–RANS optimization," *Journal of Computational Physics*, Vol. 432, 2021, pp. 110153.

[53] Zhu, L., Zhang, W., Kou, J., and Liu, Y., "Machine learning methods for turbulence modeling in subsonic flows around airfoils," *Physics of Fluids*, Vol. 31, No. 1, 2019, pp. 015105.

[54] Weatheritt, J. and Sandberg, R., "A novel evolutionary algorithm applied to algebraic modifications of the RANS stress–strain relationship," *Journal of Computational Physics*, Vol. 325, 2016, pp. 22–37.

[55] Zhao, Y., Akolekar, H. D., Weatheritt, J., Michelassi, V., and Sandberg, R. D., "RANS turbulence model development using CFD-driven machine learning," *Journal of Computational Physics*, Vol. 411, 2020, pp. 109413.

[56] Weatheritt, J. and Sandberg, R., "The development of algebraic stress models using a novel evolutionary algorithm," *International Journal of Heat and Fluid Flow*, Vol. 68, 2017, pp. 298–318.

[57] Schmelzer, M., Dwight, R. P., and Cinnella, P., "Discovery of Algebraic Reynolds-Stress Models Using Sparse Symbolic Regression," *Flow, Turbulence and Combustion*, Vol. 104, 2020, pp. 579–603.

[58] Sotgiu, C., Weigand, B., Semmler, K., and Wellinger, P., "Towards a general data-driven explicit algebraic Reynolds stress prediction framework," *International Journal of Heat and Fluid Flow*, Vol. 79, 2019, pp. 108454.

[59] Singh, A. P. and Duraisamy, K., "Using field inversion to quantify functional errors in turbulence closures," *Physics of Fluids*, Vol. 28, No. 4, 2016, pp. 045110.

[60] Parish, E. J. and Duraisamy, K., "A paradigm for data-driven predictive modeling using field inversion and machine learning," *Journal of Computational Physics*, Vol. 305, 2016, pp. 758–774.

[61] Singh, A. P., Medida, S., and Duraisamy, K., "Machine-Learning-Augmented Predictive Modeling of Turbulent Separated Flows over Airfoils," *AIAA Journal*, Vol. 55, No. 7, 2017, pp. 2215–2227.

[62] Dow, E. and Wang, Q., *Quantification of Structural Uncertainties in the $k - \omega$ Turbulence Model*, 2011.

[63] Wu, J.-L., Wang, J.-X., and Xiao, H., "A Bayesian calibration method for reducing model-form uncertainties with application in RANS simulations," *Flow, Turbulence and Combustion*, Vol. 97, No. 3, 2016, pp. 761–786.

[64] Xiao, H., Wu, J.-L., Wang, J.-X., Sun, R., and Roy, C. J., "Quantifying and reducing model-form uncertainties in Reynolds-Averaged Navier–Stokes simulations: A data-driven, physics-informed Bayesian approach," *Journal of Computational Physics*, Vol. 324, 2016, pp. 115–136.

[65] Meldi, M. and Poux, A., "A reduced order model based on Kalman filtering for sequential data assimilation of turbulent flows," *Journal of Computational Physics*, Vol. 347, 2017, pp. 207–234.

[66] Meldi, M., "Augmented prediction of turbulent flows via sequential estimators," *Flow, Turbulence and Combustion*, Vol. 101, 2018, pp. 389–412.

[67] Thies, A. T. and Tam, C. K. W., "Computation of turbulent axisymmetric and non-axisymmetric jet flows using the $k - \epsilon$ model," *AIAA Journal*, Vol. 34, No. 2, 1996, pp. 309–316.

[68] Pope, S. B., "An explanation of the turbulent round-jet/plane-jet anomaly," *AIAA Journal*, Vol. 16, No. 3, 1978, pp. 279–281.

[69] Sarkar, S. and Lakshmanan, B., "Application of a Reynolds stress turbulence model to the compressible shear layer," *AIAA Journal*, Vol. 29, No. 5, 1991, pp. 743–749.

[70] Shirzadi, M., Mirzaei, P. A., and Naghashzadegan, M., "Improvement of $k - \epsilon$ turbulence model for CFD simulation of atmospheric boundary layer around a high-rise building using stochastic optimization and Monte Carlo Sampling technique," *Journal of Wind Engineering and Industrial Aerodynamics*, Vol. 171, 2017, pp. 366–379.

[71] Cheung, S. H., Oliver, T. A., Prudencio, E. E., Prudhomme, S., and Moser, R. D., "Bayesian uncertainty analysis with applications to turbulence modeling," *Reliability Engineering and System Safety*, Vol. 96, No. 9, 2011, pp. 1137–1149, Quantification of Margins and Uncertainties.

[72] Edeling, W. N., Cinnella, P., Dwight, R. P., and Bijl, H., "Bayesian estimates of parameter variability in the k–ϵ turbulence model," *Journal of Computational Physics*, Vol. 258, 2014, pp. 73–94.

[73] Ray, J., Lefantzi, S., Arunajatesan, S., and Dechant, L., "Bayesian Parameter Estimation of a $k - \epsilon$ Model for Accurate Jet-in-Crossflow Simulations," *AIAA Journal*, Vol. 54, No. 8, 2016, pp. 2432–2448.

[74] Ray, J., Lefantzi, S., Arunajatesan, S., and Dechant, L., "Learning an eddy viscosity model using shrinkage and Bayesian calibration: A jet-in-crossflow case study," *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, Vol. 4, No. 1, 09 2017, 011001.

[75] Zhang, J. and Fu, S., "An efficient Bayesian uncertainty quantification approach with application to $k - \omega - \gamma$ transition modeling," *Computers and Fluids*, Vol. 161, 2018, pp. 211–224.

[76] Kato, H., Ishiko, K., and Yoshizawa, A., "Optimization of Parameter Values in the Turbulence Model Aided by Data Assimilation," *AIAA Journal*, Vol. 54, No. 5, 2016, pp. 1512–1523.

[77] Kato, H., Yoshizawa, A., Ueno, G., and Obayashi, S., "A data assimilation methodology for reconstructing turbulent flows around aircraft," *Journal of Computational Physics*, Vol. 283, 2015, pp. 559–581.

[78] Doronina, O. A., Murman, S. A., and Hamlington, P. E., "Parameter Estimation for Reynolds-Averaged Navier–Stokes Models Using Approximate Bayesian Computation," *AIAA Journal*, Vol. 59, No. 11, 2021, pp. 4703–4718.

[79] Nadiga, B., Jiang, C., and Livescu, D., "Leveraging Bayesian analysis to improve accuracy of approximate models," *Journal of Computational Physics*, Vol. 394, 2019, pp. 280–297.

[80] Ray, J., Kieweg, S., Dinzl, D., Carnes, B., Weirs, V. G., Freno, B., Howard, M., Smith, T., I. Nompelis, I., and Candler, G. V., "Estimation of Inflow Uncertainties in Laminar Hypersonic Double-Cone Experiments," *AIAA Journal*, Vol. 58, No. 10, 2020, pp. 4461–4474.

[81] Kieweg, S., Ray, J., Weirs, V. G., Carnes, B., Dinzl, D., Freno, B., Howard, M., Rider, W. J., and Smith, T., "Uncertainty quantification, sensitivity analysis, and validation assessment of laminar hypersonic double-cone flow simulations," *AIAA 2019 Aerosciences Conference, AIAA SciTech 2019 (AIAA-2019-2278)*, 2019.

[82] Schaefer, J., Hosder, S., West, T., Rumsey, C., Carlson, J.-R., and Kleb, W., "Uncertainty Quantification of Turbulence Model Closure Coefficients for Transonic Wall-Bounded Flows," *AIAA Journal*, Vol. 55, No. 1, 2017, pp. 195–213.

[83] Chowdhary, K. and Najm, H. N., "Bayesian estimation of Karhunen–Loève expansions; A random subspace approach," *Journal of Computational Physics*, Vol. 319, 2016, pp. 280–293.

[84] Xiu, D. and Hesthaven, J. S., "High-Order Collocation Methods for Differential Equations with Random Inputs," *SIAM Journal on Scientific Computing*, Vol. 27, No. 3, 2005, pp. 1118–1139.

[85] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, Vol. 12, 2011, pp. 2825–2830.

[86] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B., *Bayesian Data Analysis*, Chapman and Hall/CRC, 2nd ed., 2004.

[87] Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., *Markov Chain Monte Carlo in Practice*, Chapman & Hall / CRC, Boca Raton, Florida, USA, 1996.

[88] Liang, F., Liu, C., and Carroll, R. J., *Advanced Markov Chain Monte Carlo Methods*, Wiley, Chichester, West Sussex, UK, 2010.

[89] Haario, H., Laine, M., Mira, A., and Saksman, E., "DRAM: Efficient adaptive MCMC," *Statistics and Computing*, Vol. 16, No. 4, 2006, pp. 339–354.

[90] Gneiting, T., Balabdaoui, F., and Raftery, A. E., "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 69, 2007, pp. 243–268.

[91] Gneiting, T. and Raftery, A. E., "Strictly proper scoring rules, prediction, and estimation," *Journal of the American statistical Association*, Vol. 102, No. 477, 2007, pp. 359–378.

[92] Zamo, M. and Naveau, P., "Estimation of the Continuous Ranked Probability Score with limited information and applications to ensemble weather forecasts," *Mathematical Geosciences*, Vol. 50, No. 2, 2018, pp. 209–234.

[93] Zhao, Y., Yan, C., Wang, X., Liu, H., and Zhang, W., "Uncertainty and sensitivity analysis of SST turbulence model on hypersonic flow heat transfer," *International Journal of Heat and Mass Transfer*, Vol. 136, 2019, pp. 808–820.