

A Bayesian Approach for Quantile Optimization Problems with High-Dimensional Uncertainty Sources

Christian Sabater, Olivier Le Maitre, Pietro Marco Congedo, Stefan Görtz

► **To cite this version:**

Christian Sabater, Olivier Le Maitre, Pietro Marco Congedo, Stefan Görtz. A Bayesian Approach for Quantile Optimization Problems with High-Dimensional Uncertainty Sources. *Computer Methods in Applied Mechanics and Engineering*, Elsevier, 2021, 376, pp.113632. hal-03086453

HAL Id: hal-03086453

<https://hal.archives-ouvertes.fr/hal-03086453>

Submitted on 22 Dec 2020

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

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

A Bayesian Approach for Quantile Optimization Problems with High-Dimensional Uncertainty Sources

Christian Sabater^a, Olivier Le Maître^b, Pietro Marco Congedo^c, Stefan Görtz^a

^aGerman Aerospace Center, (DLR), Institute of Aerodynamics and Flow Technology, Braunschweig, Germany

^bCNRS, Centre de Mathématiques Appliquées, Ecole Polytechnique, Palaiseau, France

^cInria, Centre de Mathématiques Appliquées, Ecole Polytechnique, Palaiseau, France

Abstract

Robust optimization strategies typically aim at minimizing some statistics of the uncertain objective function and can be expensive to solve when the statistic is costly to estimate at each design point. Surrogate models of the uncertain objective function can be used to reduce this computational cost. However, such surrogate approaches classically require a low-dimensional parametrization of the uncertainties, limiting their applicability. This work concentrates on the minimization of the quantile and the direct construction of a quantile regression model over the design space, from a limited number of training samples. A Bayesian quantile regression procedure is employed to construct the full posterior distribution of the quantile model. Sampling this distribution, we can assess the estimation error and adjust the complexity of the regression model to the available data. The Bayesian regression is embedded in a Bayesian optimization procedure, which generates sequentially new samples to improve the determination of the minimum of the quantile. Specifically, the sample infill strategy uses optimal points of a sample set of the quantile estimator. The optimization method is tested on simple analytical functions to demonstrate its convergence to the global optimum. The robust design of an airfoil's shock control bump under high-dimensional geometrical and operational uncertainties serves to demonstrate the capability of the method to handle problems with industrial relevance. Finally, we provide recommendations for future developments and improvements of the method.

Keywords: Robust design, Optimization under uncertainty, High dimensional problems, Bayesian quantile regression, Surrogate Model, Aerodynamics, Computational Fluid Dynamics

1. Introduction

The use of robust optimization techniques is increasing in popularity to come up with configurations less sensitive to aleatory uncertainties. Indeed, most engineering systems are non-deterministic because of the presence of manufacturing tolerances, random fluctuations in the environment, and uncertainty in operational conditions [1].

Robustness can be defined as the potential for success under varying circumstances or scenarios [2]. A robust design leads to future flexibility and cost savings [3]. Another definition is related to consistency and trust. Under perturbations, a robust product is expected to perform consistently and is to be trusted. The first step to reach a robust design is to understand the potential sources of variability. Once they are understood, these can be integrated into the design process from an early stage. A complete overview on robust optimization techniques can be found in [1, 4].

The functional or Quantity of Interest (QoI) to optimize is then not deterministic, but becomes a random variable. As a result, the formulation of the optimization problem must account for the stochastic nature of the QoI. Classically, one considers some statistics of the QoI. For example, the mean value or a linear

Email address: christian.sabatercampomanes@dlr.de (Christian Sabater)

combination of the mean and standard deviation of the QoI defines objective functions for optimization under uncertainty [5]. Another possibility is to determine the design such that the probability distribution function (pdf) of the QoI matches a prescribed target pdf [6]. Alternatively, the minimization of a prescribed quantile of the QoI is a flexible approach often chosen in engineering problems [7, 8, 9], because it ensures a best minimal performance with controlled probability (design requirements).

Figure 1 illustrates different robust optimization strategies. The figure shows the QoI (objective function) as a function of the design variable x . The continuous line corresponds to the deterministic QoI in the absence of uncertainty. In the presence of environmental, operational, or manufacturing uncertainties, the QoI exhibits a variability illustrated by the shaded areas corresponding to the 50 and 90% confidence intervals. The deterministic optimum design x_1^* yields an uncertain QoI with high variability and poor expected value, as depicted by the extend of the confidence intervals and the density. The optimal design denoted x_2^* has the lowest QoI expectation but is affected by high variability. In contrast, the design x_3^* corresponding to the minimization of the QoI standard deviation exhibits a low variability but poor performance. Finally, the optimization of the 95% quantile in x_4^* ensures minimal performance for 95% of the events.

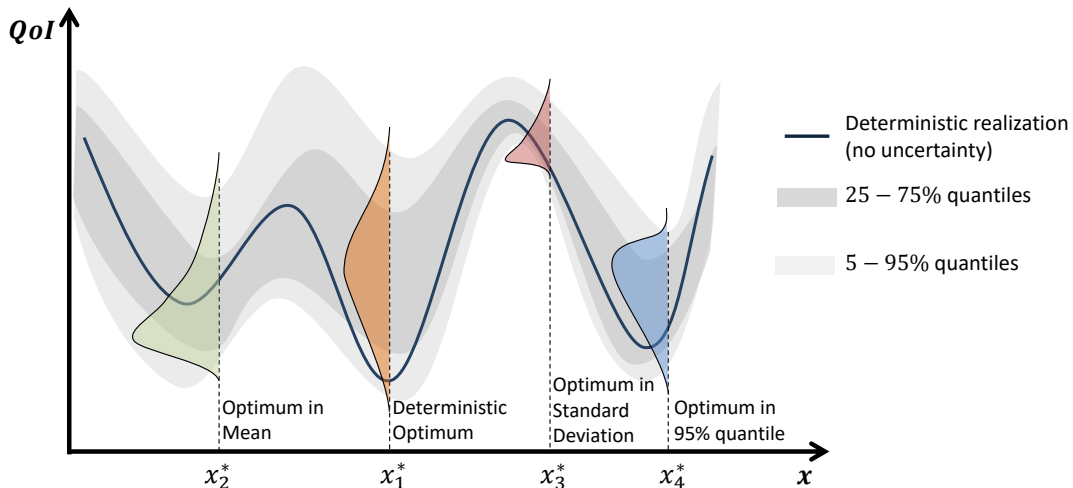


Figure 1: Comparison between deterministic and robust optimal designs.

This work focus on quantile of robust optimization problems formulated as follows. Let x be the design variables and $Y(x, \theta) \in \mathbb{R}$ the random QoI, which depends on $x \in \Omega$ and the random event $\theta \in \Theta$. The objective is to minimize the τ -quantile of Y , for some $\tau \in (0, 1)$:

$$x^* = \arg \min_{x \in \Omega} q_\tau (Y(x, \cdot)). \quad (1)$$

In particular, we are interested in solving efficiently problem (1) when $Y(\cdot, \theta)$ involves high dimensional stochastic parametrizations or is not even explicitly given.

In order to make robust design feasible for practical applications involving expensive simulations, the development of efficient methods is essential [10, 1]. Traditional approaches for robust design use surrogate models to approximate the random QoI $Y(\cdot, \theta)$. These surrogate models usually consist of Polynomial Chaos (PC) Expansions [11, 12, 13] or Gaussian Processes (GP) [5, 8, 14]. Their construction requires the evaluation of the deterministic full order model at suitable training points. These surrogates can be extensively resampled, at a low computational cost, to estimate the desired quantile through the inverse empirical cumulative distribution function. If the randomness is independent of the design variable, it is even possible to construct a global surrogate in the stochastic and design spaces ($\Omega \times \Theta$) [15].

A key feature of PC and GP surrogates is that they need a parametrization for the uncertainty sources, classically using a finite set of independent random variables. When dealing with problems requiring a high-dimensional parametrization, such as for uncertainty induced by a random field, the number of training

samples to construct accurate high-dimensional surrogates increases very quickly. It can become prohibitive unless particular structures can be exploited. The collaborative EU research project UMRIDA [16] (Uncertainty Management for Robust Design in Aeronautics, 2013-2016), tackled robust design under a large number of independent uncertainties. In this project, problems with more than 10 to 20 independent uncertain parameters were considered high dimensional and treated by dimensionality reduction techniques such as Karhunen-Loeve expansions [17, 18], and sparse Polynomial Chaos [19, 20] expansions. The construction of surrogates in high dimensions can also benefit from the gradients, estimated, for instance, by adjoint methods [21]. Dimensionality reduction techniques such as Support Vector Clustering [22] and Principal Component Analysis in combination with surrogate models [23] have been proposed for uncertainty quantification under high dimensional uncertainty sources. All these methods based on dimensionality reduction of the stochastic space are unable to deal with situations where such a reduction is not feasible or effective. This limitation makes these approaches unfeasible when dealing with complex and truly high-dimensional uncertainties. Also, the surrogate models are not suitable for non-parametric uncertainties, *i.e.*, non-amenable to a parametrization.

Another possibility when dealing with uncertainty quantification is the use of variable fidelity methods in combination with Multi Level Monte Carlo approaches [24, 25]. However, several levels of fidelity for a given black box problem are not always available, and the number of samples required to accurately compute statistics for large scale industrial problems is still very large. Also, the use of importance sampling [26] can help to reduce the computational burden associated with stochastic simulations. These sampling methods still need to be integrated with traditional optimization approaches to perform robust design making the optimization expensive.

The objective of this paper is then to develop an efficient approach for quantile optimization, that is insensitive to the structure and dimensionality of the uncertainty and even able to handle non-parametric uncertainties. To this end, we propose to construct a surrogate directly for the quantile as a function of the design variables, by-passing the estimation of a surrogate for the random objective function. To avoid any explicit parametrization of the uncertainties, we rely only on samples of the random objective function. This feature is handy when dealing with high-dimensional or non-reducible uncertainty sources. Our approach exploits the fact that sampling-based methods, such as Monte Carlo, are typically insensitive to the complexity of the random source and are therefore tractable in high-dimensional problems. Further, in our context of robust optimization, smooth dependencies with the design variables of the quantile can be exploited in the surrogate construction, sampling the stochastic and design spaces jointly.

In the following, we first propose a regression approach to estimate the quantile surrogate globally over the whole design space. We propose an original approach to estimate the quantile prediction error. We exploit this error estimation to derive a sequential construction approach in order to improve the accuracy of the quantile surrogate. The approach selects the design points where to collect new samples of the objective function. This strategy is adapted to the robust optimization problem, focusing the computational effort on the estimation and determination of the optimum of the quantile in the design space.

The organization of the paper is as follows. In Section 2, we introduce the quantile regression, which consists of a semiparametric expression of the quantile as a function of the design parameters derived from a samples set of the random function $Y(x, \theta)$. Section 3 constitutes the core of the methodological contributions of the work. In Section 3.2 a Bayesian formulation of the quantile regression problem is introduced, and we derive sampling strategies for the quantile estimator in order to estimate its error. A Bayesian sequential strategy is proposed in Section 3.3 for the optimization of the quantile. The optimization mostly uses the sampling of the quantile estimator to generate new candidates of the optimum point and infill the design domain. The Bayesian optimization methods are applied on one and two-dimensional test functions in Section 4 and to a realistic aerodynamic shape optimization problem in Section 5. Finally, Section 6 reports major conclusions of the work and recommendations for future works.

2. Quantile Regression

This section concerns the quantile regression method, which builds a deterministic model of the τ -quantile $q_\tau(x)$ of $Y(x)$, where $x \in \Omega$ is the design variable. Section 2.2 defines the quantile of a random variable

Y (non-indexed on x) and introduces its empirical estimation using the loss function (6). The estimation is further extended to Y conditioned on x , before introducing a discretization in x in Section 2.3. Finally, Section 2.4 discusses some practical aspects of the construction, emphasizing one the selection of the size of the set of regression points. The material presented in the section will be extended in Section 3 to derive a Bayesian formulation of the quantile regression and exploit this formulation to quantile optimization.

2.1. Motivation

Quantile regression, introduced in the 70's by Koenker and Basset [27], provides a richer view of applications compared to conditional mean models such as least squares regression [28]. While most regression methods focus on the mean of the response variable, it is possible to look at other statistics of the distribution.

According to Mosteller and Tukey [29], traditional regression estimates averages of the distributions at a set of points. However, it is also possible to obtain different regression curves corresponding to the percentage point of the distribution. The least-squares regression gives an incomplete picture of the distribution, where the quantile regression provides a general technique to estimate families of conditional quantiles.

The use of the quantile as an estimate of uncertainty is very useful in engineering and design optimization. As a result, the quantile regression is an appealing approach to predict any quantile as a function of the design parameters. The approach is also independent of the sources of uncertainty and does not require any parametrization.

2.2. Quantile estimation

We denote $F_Y : \mathbb{R} \mapsto [0, 1]$ the Cumulative Distribution Function (CDF) of a real-valued random variable $Y(\theta)$:

$$F_Y(y) = P(Y \leq y). \quad (2)$$

The CDF is right continuous, such that, for any $\tau \in (0, 1)$, the τ -quantile of Y can be defined as

$$q_\tau = F_Y^{-1}(\tau) = \inf \{y | F_Y(y) \geq \tau\}. \quad (3)$$

In particular $q_{0.5}$ is called the median of the distribution. A quantile can be estimated from the Empirical Cumulative Distribution Function (ECDF) [30] from a sample set $\mathcal{Y} = \{y_i, i = 1, \dots, N_{\mathcal{Y}}\}$ of $N_{\mathcal{Y}}$ random sample points of Y . The ECDF of the sample set is given by

$$\hat{F}_{Y, N_{\mathcal{Y}}}(y) = \frac{\# \text{ of samples } \leq y}{N_{\mathcal{Y}}} = \frac{1}{N_{\mathcal{Y}}} \sum_{i=1}^{N_{\mathcal{Y}}} I_y(y_i), \quad I_y(y') \doteq \begin{cases} 1, & y' \leq y \\ 0, & \text{otherwise} \end{cases}. \quad (4)$$

The ECDF converges to F_Y as $N_{\mathcal{Y}} \rightarrow \infty$ and can be used in (3) to obtain the empirical estimate of q_τ .

An alternative approach consists in estimating the quantile as the minimizer of a loss function, specifically, the optimal point estimator of the asymmetric linear loss [28] based on the check function ρ_τ :

$$\rho_\tau(u) = (\tau - I_0(u)) u = \frac{|u| + (2\tau - 1)u}{2}. \quad (5)$$

The loss function of \mathcal{Y} is

$$L_\tau(q|\mathcal{Y}) = \sum_{i=1}^{N_{\mathcal{Y}}} \rho_\tau(y_i - q) = \sum_{i=1}^{N_{\mathcal{Y}}} (\tau - I_0(y_i - q)) (y_i - q), \quad (6)$$

and the optimal estimator is

$$q_\tau^* = \arg \min_{q \in \mathbb{R}} L_\tau(q|\mathcal{Y}). \quad (7)$$

The loss function in (6) is the sum of asymmetrically weighted absolute residuals [31]. Its minimum coincides with the classical empirical quantile estimator associated with the ECDF. Roughly, the loss function

penalizes a value of q such that the fraction of samples in \mathcal{Y} greater (resp. smaller) than q exceeds τ (resp. $1 - \tau$). Absolute residuals are used to improve the quality and reliability of quantile estimator by preventing high sensitivity to outliers and dominance of extreme values $|q - y_i|$ in the loss.

The estimation above can be extended to the case where Y is conditioned on $x \in \Omega$, for instance the design variables of the optimization problem. In this case, the conditional CDF is $F_Y(y|x)$ and the τ -quantile also becomes a function of x ,

$$q_\tau(x) = F_Y^{-1}(\tau|x). \quad (8)$$

The estimation of $q_\tau(x)$ is classically restricted to a parametric family involving some parameters ω . Further, the sample set \mathcal{Y} consists now in $N_{\mathcal{Y}}$ couples (x_i, y_i) where y_i is drawn from the CDF conditioned on $x = x_i$, that is $y_i \sim F_Y(y|x_i)$. The minimization of the loss function translates to the determination of the τ -dependent optimal parameters ω_τ^* , such that

$$\omega_\tau^* = \arg \min_{\omega} \sum_{i=1}^n \rho_\tau(y_i - q(x_i|\omega)), \quad (9)$$

and the estimator is finally $q_\tau^*(x) = q(x|\omega_\tau^*)$. In the following section, we introduce a particular form to express $q(x|\omega)$.

2.3. Quantile discretization

Traditionally, multiple linear regression ($q(x|\omega = (\alpha, \beta)) = \alpha + \beta \cdot x$) have been used to estimate conditional quantiles [32]. However, more complex models are often necessary to capture real-world complexities and deal with non-linear dependencies. Among others, polynomial and Radial Basis Functions (RBF) are popular approximation methods. To remain as general as possible, we shall assume the following generic form for the quantile model,

$$q(x|\omega) = \sum_{j=1}^{N_{\mathcal{X}}} \omega_j \phi_j(x), \quad (10)$$

where $\phi_{j=1, \dots, N_{\mathcal{X}}}(x)$ are given regressors and the $\omega_{j=1, \dots, N_{\mathcal{X}}}$ are the parameters of the quantile regression. In the case of RBF models, one uses a set $\mathcal{X} \doteq \{\tilde{x}_j, j = 1, \dots, N_{\mathcal{X}}\}$ of control points in Ω , and defines

$$\phi_j(x) \doteq \phi(|x - x_j|), \quad j = 1, \dots, N_{\mathcal{X}},$$

where $|x - x'|$ denotes the Euclidean norm and $\phi : \mathbb{R}_+ \mapsto \mathbb{R}$ is a suitable kernel function. If the value of $q_\tau(x)$ was known or could be estimated at selected points, the vector of parameters $\omega = (\omega_1 \cdots \omega_{N_{\mathcal{X}}})$ could be determined by solving a standard regression problem, such as a least-squares minimization problem.

In the present situation, the quantile values are unknown and must be estimated from the sample set \mathcal{Y} . The vector of optimal parameters ω_τ^* is the minimizer of (9),

$$\omega_\tau^* = \arg \min_{\omega \in \mathbb{R}^{N_{\mathcal{X}}}} \rho_\tau(y - [\Phi]\omega), \quad (11)$$

where $y = (y_1 \dots y_{N_{\mathcal{Y}}})$, $[\Phi] \in \mathbb{R}^{N_{\mathcal{Y}} \times N_{\mathcal{X}}}$ is the matrix with entries $[\Phi]_{ij} = \phi_j(x_i)$, and the application of the checkfunction on a vector amounts to the sum of the checkfunctions over all the components of the vector ($\rho_\tau(y) = \sum_i \rho_\tau(y_i)$). The estimator of the quantile can be recast in

$$q_\tau^*(x) = \sum_{j=1}^{N_{\mathcal{X}}} (\omega_\tau^*)_j \phi_j(x) = \Phi(x)^T \omega_\tau^*, \quad (12)$$

where $\Phi(x) = (\phi_1(x) \cdots \phi_{N_{\mathcal{X}}}(x))^T$ is the vector of regressors evaluated at $x \in \Omega$.

In the following, we rely on a RBF approximation corresponding to

$$\phi_j(x) = \phi(|x - x_j|), \quad \phi(r) = r^2 \log r.$$

This particular kernel was proposed by Duchon [33] in the context of Thin-Plate Splines (TPS) models. The complete TPS model combines a multi-linear regression with the RBF approximation. Here, we only retain the RBF part to keep the model simple.

A detailed analysis of the RBF approximation error can be found in [34]. In particular, the approximation error depends on the density of sampling points x_j . We stress that the following developments are not dependent on the approximation method in x of the quantile, provided that it can be recast as in (12). In particular, instead of RBF, one may prefer to use Gaussian Processes models or regression in a polynomial or other finite-dimensional functional space.

2.4. Practical aspects

The estimation of the conditional quantile involves two sets, one (\mathcal{Y}) for the sample points of conditional random variable $Y(\theta|x)$, and one (\mathcal{X}) containing the control points of the quantile regression. When these two sets are fixed, the optimal parameters ω_τ^* of the quantile are computed solving the minimization problem (11). In this work, we rely on a Differential Evolution (DE) [35] algorithm to compute the optimal parameters. Differential Evolution is a global optimization strategy based on genetic algorithms. A population with size $n_{\text{pop}} = 10N_{\mathcal{X}}$ was used successfully in the examples presented in this paper. The convergence of the algorithm typically requires a few hundreds of iterations (generations). The resulting computational cost remains acceptable because of the fixed location of the control and sample points, so the matrix $[\Phi]$ in (11) can be evaluated once during the optimization procedure.

More critical is the construction of the sample sets \mathcal{Y} and \mathcal{X} . Concerning \mathcal{Y} , it is clear that the sample points x_i should cover well the domain Ω for a good approximation of the quantile at all $x \in \Omega$. This observation suggests a sampling strategy that maximizes the covering of Ω , such as low discrepancy sequences with optimal infilling properties. In this work, we take advantage of the simple structures of the domains Ω considered (hyper-rectangles) and rely on Sobol sequences [36] to generate the sample points x_i . Then, at each sampled point x_i , a single sample y_i is drawn from $Y(x_i, \theta)$. Note that multiple random replications of $Y(x_i, \theta)$ could be considered for the same x_i , thus constituting a stratified sample set \mathcal{Y} ; however, this approach would not produce better estimates of the quantiles and would not be optimal.

Regarding the set of control points, \mathcal{X} , one could consider using the set, or a subset, of point x_i in \mathcal{Y} . However, the regression method described above uses control points \tilde{x}_j that are not necessarily coinciding with sample points x_i , and we shall rely on distinct sets of points, for the data and the quantile discretization. The subsequent extension of the quantile regression method to the robust optimization problem motivates the use of distinct sets of points. The estimation of the quantile and the optimization are two distinct tasks which, although related, leads to different sequential infilling strategies for \mathcal{X} and \mathcal{Y} . These strategies will be detailed later, and we discuss here the initialization of \mathcal{X} , which follows another independent Sobol sequence of size $N_{\mathcal{X}}$.

The size $N_{\mathcal{X}}$ of \mathcal{X} must be specified and is crucial to obtain an accurate estimation of the quantile. A large number $N_{\mathcal{X}}$ of control points offers a better representation of the quantile but can induce over-fitting when $N_{\mathcal{X}}$ is too large relative to the sample set size of data ($N_{\mathcal{Y}}$). In this situation, the estimator becomes sensitive to the data in \mathcal{Y} . On the other hand, a lower number of control points will not be able to capture the details of the quantile adequately and will produce too smooth estimators with lesser accuracy. This is the well-known bias-variance trade-off. To tackle the issue of selecting the number of control points, we propose to follow a classical K -fold Cross-Validation (CV) method. In the CV method, \mathcal{Y} is first partitioned into $K > 1$ subsets $\mathcal{Y}_1, \dots, \mathcal{Y}_K$ with comparable size $\approx N_{\mathcal{Y}}/K$. The quantile is then estimated on the reduced sample set $\mathcal{Y} \setminus \mathcal{Y}_r$, yielding the loss $L_r = L_\tau(q_\tau^*|\mathcal{Y}_r)$ associated to the left-out subset \mathcal{Y}_r (see (9)). The consolidated CV-loss $L(\mathcal{X}) \doteq \sum_{r=1}^K L_r$ characterizes the loss associated to \mathcal{X} . This procedure can be repeated for a sequence of nested sets $\mathcal{X}^l \subset \mathcal{X}^{l+1} \subset \dots$, obtained by continuing the constitutive Sobol sequence, until the cross-validation loss $L(\mathcal{X}^l)$ starts to increase. The size $N_{\mathcal{X}}^l = |\mathcal{X}^l|$ giving the lowest CV-loss is finally considered as the best trade-off. In practice, we use $K = 10$ cross-validation folds and enrichment of just one new control point from \mathcal{X}^l to \mathcal{X}^{l+1} .

3. Bayesian Optimization Strategy

3.1. Motivation

The quantile regression method described in the previous section provides an estimator $q_\tau^*(x)$ of the τ -quantile of $Y(x, \theta)$ from a sample set of observation \mathcal{Y} . This estimator can substitute the true quantile to solve the robust optimization problem, that is solving

$$\hat{x}^* = \arg \min_{x \in \Omega} q_\tau^*(x), \quad q_\tau^*(x) = q(x|\boldsymbol{\omega}^*), \quad (13)$$

where $\boldsymbol{\omega}^*$ is the minimizer of the loss function associated to \mathcal{Y} . To obtain a correct approximation of the exact optimum, through this procedure, one needs a sufficiently accurate estimation of the quantile. Further, as the location of the optimum is unknown, the quantile must be accurately estimated for all $x \in \Omega$, a task that would demand a substantial sample set \mathcal{Y} , in particular for $\tau(1 - \tau) \rightarrow 0$.

This observation motivates the development of a Bayesian optimization procedure, where the estimation of the quantile is sequentially improved only in areas of interest in Ω that are susceptible to contain the optimum. Such an approach requires two main ingredients: a measure of the local error on the quantile estimation and infilling strategies to propose new estimations of Y in order to improve the estimation of the optimum x^* . The rest of the section concerns these two aspects, while the optimization procedure is discussed later.

3.2. Bayesian quantile regression

We start by focusing on the characterization of the error of the estimator $q_\tau^*(x)$. We remark that this estimator is random, as it uses a random sample set \mathcal{Y} . An immediate idea to assess the quality of $q_\tau^*(x)$ would be to measure its variance with respect to the sample set. Estimating $\mathbb{V}(q_\tau^*)$ by replicating \mathcal{Y} is not considered as a viable option, because of its computational cost. Characterizing the variability of the estimator by Bootstrapping and sub-sampling may be feasible approaches, not requiring new evaluations of $Y(x, \theta)$. However, the complete construction of the sample estimator would demand solving multiple optimization problems, for the optimal parameters $\boldsymbol{\omega}_\tau^*$ of each subsample or bootstrap replica of \mathcal{Y} proposed. Even if these optimizations can be carried out in parallel, these procedures would consume significant computational resources if the minimization of the loss function is difficult, and many samples are needed to characterize the variance of the estimator.

These considerations lead us to propose a more direct approach that does not require the minimization of the loss function. The proposed approach considers that the model parameters $\boldsymbol{\omega}$ defining the estimator are random and equipped with a joint probability density function conditioned on the sample set \mathcal{Y} . In order to derive an expression for this joint distribution, we rely on the Bayes theorem:

$$p_\tau(\boldsymbol{\omega}|\mathcal{Y}) = \frac{\mathcal{L}_\tau(\mathcal{Y}|\boldsymbol{\omega}) p(\boldsymbol{\omega})}{p_\tau(\mathcal{Y})}. \quad (14)$$

The Bayes theorem states that the posterior (after observation of the sample set \mathcal{Y}) distribution $p_\tau(\boldsymbol{\omega}|\mathcal{Y})$ of the parameters of the quantile estimator, is equal to the product of the likelihood \mathcal{L}_τ of the sample set and the prior of the parameters $p(\boldsymbol{\omega})$, divided by the evidence or absolute probability of the sample set. The critical point here is that knowing the posterior distribution of $\boldsymbol{\omega}$, we have a complete probabilistic description of the τ -quantile estimator.

3.2.1. Asymmetric Laplace Distribution

The likelihood $\mathcal{L}_\tau(\mathcal{Y}|\boldsymbol{\omega})$ measures the probability of the sample set \mathcal{Y} assuming that Y has a τ -quantile following the model with parameters $\boldsymbol{\omega}$. The main difficulty in deriving an appropriate likelihood is that we don't know the conditional distribution of $Y(x, \theta)$, such that the conditional distribution of $Y(x, \theta) - q(x|\boldsymbol{\omega})$ must be postulated [37]. Following Yu and Moyeed [38], we consider a likelihood based on the (centered) Asymmetric Laplace Distribution (ALD), whose density function is

$$f_{\text{AL}}(u; \lambda, \kappa) = \frac{\lambda}{\kappa + 1/\kappa} \begin{cases} \exp[(\lambda/\kappa) u], & u < 0, \\ \exp[-(\lambda\kappa) u], & u \geq 0. \end{cases} \quad (15)$$

Despite another approaches are available, the ALD is commonly chosen as a simpler alternative without requiring the need of complex choices of prior distributions and hyperparameters [39]. Since the introduction in [38], the choice of the ALD has proven to be effective for Bayesian quantile regression [40]. Based on empirical findings [38], the ALD is robust to the true underlying likelihood. In [41], an asymptotic justification of the ALD for quantile regression is explored, showing the posterior consistency under general conditions. Following the robustness of the ALD and its relatively simplicity, we choose it as a basis for the Bayesian formulation.

The two parameters of the distribution are the scale and asymmetry parameters, $\lambda > 0$ and $\kappa > 0$ respectively. The ALD consists of two exponential distributions in $|u|$, with rates λ/κ and $\lambda\kappa$, for negative and positive argument u respectively. Setting the rates to $1 - \tau$ and τ , the density function of the ALD can be expressed in terms of the check function ρ_τ defined in (5). After elementary calculations it comes

$$f_\tau(u) = \tau(1 - \tau) \exp[-\rho_\tau(u)]. \quad (16)$$

To derive the likelihood of \mathcal{Y} we exploit the fact that the y_i are sampled independently. Therefore, we obtain

$$\mathcal{L}_\tau(\mathcal{Y}|\boldsymbol{\omega}) = \prod_{i=1}^{N_{\mathcal{Y}}} f_\tau(y_i - q(x_i|\boldsymbol{\omega})) = \prod_{i=1}^{N_{\mathcal{Y}}} \tau(1 - \tau) \exp[-\rho_\tau(y_i - q(x_i|\boldsymbol{\omega}))]. \quad (17)$$

It is seen that the maximizer of (17) corresponds to the minimizer $\boldsymbol{\omega}_\tau^*$ of the loss function 11. In other words, $\boldsymbol{\omega}_\tau^*$ (resp. $q_\tau^*(x)$) is the Maximum Likelihood Estimator (MLE) of the quantile parameters (resp. of the quantile) for the likelihood in (17). Further, the MLE is invariant to consistent scaling of the two exponential rates preserving their ratio $\tau/(1 - \tau)$. This is due to the property of the scaling property of the check-function: $\rho_\tau(u/\alpha) = \rho_\tau(u)/\alpha$ for any $\alpha > 0$. In practice, changing the scaling factor α will affect the spread of the AL density function, as seen from the following expression:

$$f_\tau(u; \alpha) = \frac{\tau(1 - \tau)}{\alpha} \exp\left[-\frac{\rho_\tau(u)}{\alpha}\right]. \quad (18)$$

In the present work, we restrict ourselves to the case $\alpha = 1$, which was found suitable in all the tests presented hereafter. However, our approach could benefit in adapting α as to better account for the conditional distribution $Y(x, \theta)$. For instance, α could be treated as a hyper-parameter, and could be inferred. All these aspects fall beyond the scope of the present paper and will be presented elsewhere.

3.2.2. Posterior sampling

The likelihood of \mathcal{Y} being set, it remains to discuss the prior of $\boldsymbol{\omega}$ and the evidence $p_\tau(\mathcal{Y})$. Concerning the prior of the parameters, a non informative prior (improper uniform distribution) is selected in the absence of any a priori information on the parameters. However, we shall keep the prior in the subsequent development to remain general. Further, we remark that if some a priori knowledge on $q_\tau(x)$ is available at the control points $\tilde{x}_j \in \mathcal{X}$, one can easily translate it into an a priori on $\boldsymbol{\omega}$, thanks to the linear form of (10).

At this point, we have obtained an expression for the posterior distribution of the regression parameters $\boldsymbol{\omega}$, up to a multiplicative factor (the inverse of the evidence). The role of $p_\tau(\mathcal{Y})$ is to ensure that the posterior integrates to one,

$$\int_{\mathbb{R}^{N_{\mathcal{X}}}} p_\tau(\boldsymbol{\omega}|\mathcal{Y}) d\boldsymbol{\omega} = 1.$$

In practice, the evidence must be numerically estimated, if needed, as there is no closed-form expression for the likelihood. Similarly, the absence of explicit expression prevents the direct computation of the moments and statistics of posterior of $\boldsymbol{\omega}$; instead it is necessary to proceed by sampling $p_\tau(\boldsymbol{\omega}|\mathcal{Y})$ to estimate statistics.

In the following, we use a Markov Chain Monte Carlo method to sample $p_\tau(\boldsymbol{\omega}|\mathcal{Y})$. The MCMC method are appropriate to sample complex, high dimensional, or non-explicit probability distributions for which a direct sampling is difficult or not feasible [42]. The key idea of MCMC is to generate a random sequence of sample points $\boldsymbol{\omega}_k$, with a simple probabilistic transition rule from $\boldsymbol{\omega}_k$ to $\boldsymbol{\omega}_{k+1}$ ensuring that the asymptotic invariant measure of the chain corresponds to the target distribution, here $p_\tau(\boldsymbol{\omega}|\mathcal{Y})$.

The Metropolis-Hastings (MH) algorithm [43, 44] is one of the most popular MCMC sampler. It is based on a acceptance/rejection rule for the next step of the chain given the current point of the chain, ω_k . Specifically, the proposed point ω' is drawn at random from the conditional distribution $p_t(\omega'|\omega_k)$. The proposed transition is accepted, *i.e.* $\omega_{k+1} = \omega'$, with a probability π given by

$$\pi = \min \left(1, \frac{p_\tau(\omega'|\mathcal{Y})p_t(\omega_k|\omega')}{p_\tau(\omega_k|\mathcal{Y})p_t(\omega'|\omega_k)} \right). \quad (19)$$

Otherwise, the transition is rejected and $\omega_{k+1} = \omega_k$. In the case of a reversible proposals ($p_t(\omega_k|\omega') = p_t(\omega'|\omega_k)$), the rule always accepts a transition to points with higher posterior value, and accepts lower posterior points with a decreasing probability equal to the ratio (< 1) of actual and proposed posterior values. While needing the ratio of posterior values defining the rule, it is not needed to know the evidence $p_\tau(\mathcal{Y})$ when applying the MH algorithm on the posterior in (14), the product of likelihoods and priors suffices.

The choice of the proposal distribution $p_t(\omega'|\omega)$ is critical to efficiently sample the posterior distribution [42]. The objective is to generate a chain with rapid mixing properties, such that, irrespective of the starting point, ω_k quickly converges to the stationary distribution. If the proposal distribution induces increments $\omega' - \omega$ that are too small, the proposed points are accepted with high rate but many steps are required to explore the whole distribution. If on the contrary the increments tend to be too large, the MH algorithm exhibits a high rejection rate with long range correlation between the successive steps. Acceptance rates in the range of $[0.2, 0.35]$ are considered to offer a close to optimal trade-off for the mixing of the chain. In the present work, we consider (reversible) Gaussian proposals,

$$p_t(\omega'|\omega) = N(\omega, \Sigma_{p_t}^2), \quad (20)$$

with the covariance matrix $\Sigma_{p_t}^2$ proportional to the posterior covariance of the sampled quantity:

$$\Sigma_{p_t}^2 = \frac{2.38^2}{N_{\mathcal{X}}} [\mathbb{E}(\omega\omega^T) - \mathbb{E}(\omega)\mathbb{E}(\omega^T)]. \quad (21)$$

In this approach, the posterior covariance of ω must be estimated since the posterior distribution is not explicitly known. This is achieved in the burn in stage of the MH algorithm, where the proposal covariance is regularly updated as the chain advances and the estimate of the posterior covariance progresses. Note that the scaling of the covariance in (21) is optimal in the case of Gaussian posteriors [45]. This rule has provided satisfying results for all the cases presented hereafter.

3.2.3. Quantile error

The MH algorithm can be used to produce a large sample set of $\omega \sim p(\omega|\mathcal{Y})$ at a modest computational cost. Indeed the MH procedure involves only the evaluation of the loss function and no optimization. Typically, a long chain is computed and sample points of ω are extracted by sub-sampling the chain. Each sampled value ω_k corresponds to an estimate $\tilde{q}_\tau^k(x)$ of the quantile, namely

$$\tilde{q}_\tau^k(x) = \Phi(x)^T \omega_k, \quad \Phi_j(x) = \phi(|\tilde{x}_j - x|). \quad (22)$$

The linearity of the model (22) can be exploited to effectively compute the moments of the quantile estimate at any locations $x \in \Omega$. In particular, due to the asymmetry of the likelihood, the posterior average of the quantile, $\bar{q}_\tau(x) = \Phi(x)^T \mathbb{E}(\omega)$, does not coincide with the MLE $q_\tau^*(x)$. With sufficiently large sample sets, it is even possible to estimate the empirical distribution of the estimate and build confidence intervals on the predicted quantile value. We observe that these characterizations relate to the sample set \mathcal{Y} only and, therefore, disregard any possible source of error incurring to the quantile discretization and the selected control points in \mathcal{X} . Future works will have to focus and improve on this aspect.

The information in the Bayesian posterior distribution of the quantile estimate can be exploited to deploy infilling strategies to improve the quantile estimation over Ω . For instance, one can enrich \mathcal{Y} with new samples

of $Y(x, \theta)$ at points x where the posterior variance, or inter-quantile range, of $q_\tau^*(x)$ is the largest. Our tests (not shown) exploiting such estimates of the quantile error is effective when complemented with another infilling strategy for the set of control points \mathcal{X} . However, we focus in the present work on the optimization problem and therefore do not develop here strategies aiming at minimizing the global quantile error over the whole domain Ω .

3.3. Optimization strategy

As discussed before, solving (13) will not yield the exact minimizer of q_τ , in general, because of the error $q_\tau(x) - q_\tau^*(x)$. The objective is then to improve the estimation of the minimizer by improving the quantile estimator through a parsimonious enrichment of \mathcal{Y} . Classical Bayesian Optimization methods are based on enrichment that combines two objectives: the exploration of Ω and the exploitation of the statistical model to improve the current estimate of the minimizer. For the minimization of a deterministic objective function $f(x)$, the expected improvement (EI) infilling criteria was introduced to account for the imperfect knowledge of f [46]. The EI essentially measures the possible improvement at any point $x \in \Omega$ of the objective value from the current best estimate $f(x_k^*)$, given the statistical model (usually a Gaussian process model) of $f(x)$. In sequential approaches, the point with highest EI is selected, the objective function is evaluated at this point and the statistical model is updated with this new observation of f . The procedure is then repeated until a stopping criteria is satisfied.

The procedure outlined above must be adapted to the minimization of a quantile. Indeed, we don't want to evaluate the objective at the new point, *i.e.*, estimate $q_\tau(x)$; instead we can only draw (few) samples of $Y(x, \theta)$. In other words, the objective of the minimization procedure is never evaluated and we entirely rely on our estimation to guide the infilling procedure.

However, new sample points should preferentially be drawn in areas of interest where $q_\tau^*(x)$ is closed to the minimum of q_τ , in order to reduce the error in the estimate \hat{x}^* defined in (13).

3.3.1. Infill with samples of minimum

In the following, we denote $\mathcal{Y}^{(l)}$ the sample set $F(x, \theta)$ at the l -th iteration of the optimization procedure. We propose to directly rely on the posterior distribution of the quantile estimator based on $\mathcal{Y}^{(l)}$ to propose new points and enrich the set $\mathcal{Y}^{(l)}$. The MH algorithm detailed above is used to generate a sample set $\mathcal{Q}^{(l)}$ of $P_{\text{new}} \geq 1$ independent random realizations $\tilde{q}_\tau^k(x)$ of the estimator in (22). In practice, the sample set is obtained by sub-sampling the MCMC chain every n_s steps, where n_s is set large enough to ensure that successive sample points ω_k are uncorrelated. For each element $\tilde{q}_\tau^k(x) \in \mathcal{Q}$ we denote \hat{x}_k^* the associated global minimizer, such that

$$\hat{x}_k^* = \arg \min_{x \in \Omega} \tilde{q}_\tau^k(x). \quad (23)$$

Finally, we enrich $\mathcal{Y}^{(l)}$ by drawing randomly one sample point y_k of $Y(\hat{x}_k^*, \theta)$ and setting

$$\mathcal{Y}^{(l+1)} = \mathcal{Y}^{(l)} \cup \{(\hat{x}_k^*, y_k), k = 1, \dots, P_{\text{new}}\}. \quad (24)$$

Solving 23 requires the resolution of an optimization problem in Ω . This problem may be challenging because of the dimensionality of the design space Ω and the existence of several local minima in the current quantile model $\tilde{q}_\tau^k(x)$. However, the evaluations of the model is cheap such that robust techniques can be employed. In the computations presented we relied on the differential evolution algorithm [35] to solve (23).

It is clear that this procedure will produce an enrichment of \mathcal{Y} at "random replicas" of the quantile estimator $q_\tau^*(x)$. As shown in the examples below, sampling the estimator of $q_\tau(x)$ allows to explore multiple and disconnected potential optimum areas, as long as the quality of the estimator is not high enough to rule-out candidate areas. In fact, the approach can be seen as sampling with P_{new} points the random variable X_l^* being the (random) minimizer of the random estimator of $q_\tau(x)$. In particular, if the estimation is poor in some areas of Ω , the sampled quantile estimator $\tilde{q}_\tau(x)$ can yield a optimum \hat{x}_k^* in this area. In other words, regions with small but non zero probability of having the minimum will be sampled for large enough P_{new} or at a later iteration (larger l). As remarked, we observe that the parameter α of the ALD distribution in Section 3.2.1 can be used to tighten or relax the belief on the quality of the Bayesian estimator

of the quantile: a high value of α would induce higher deviations of the sampled parameter vector ω_k , thus promoting more exploration, while a lower value would favor exploitation of $q_\tau^*(x)$. As discussed before, the selection of α will be the focus of future works and we fix here $\alpha = 1$. Finally, we could think in other likelihoods, but the ALD is unbounded and therefore is certain that the optimum will be included. This is enough to guarantee that the optimization will converge. Also, Laplace distributions constitute robust loss functions compared to other unbounded distributions such as Gaussian, owing to flat tails. By enabling larger departures from the "best" model, the ALD provides a conservatism in the uncertainty bounds of the quantile model.

3.3.2. Refinement of the quantile model

Whence the sample set $\mathcal{Y}^{(l)}$ has been enriched, the estimation of the quantile can be updated and the procedure can be repeated as long as some convergence criteria are not satisfied. However, while $N_{\mathcal{Y}^{(l)}} = |\mathcal{Y}^{(l)}|$ is increasing, it may become necessary to jointly refine the set \mathcal{X} of control points. Without such a refinement, the quantile discretization error would be dominant, for large \mathcal{Y} , and will not necessarily reduce in the area of the minimizer, therefore preventing the convergence of the optimization procedure.

Indexing \mathcal{X} with the iteration index, it is obvious that the enrichment of $\mathcal{X}^{(l)}$ is not independent of the enrichment of $\mathcal{Y}^{(l)}$. Specifically, adding a new control point is potentially beneficial to the approximation only if it is placed in areas where the new sample points have been added. In areas that have not been re-sampled, on the contrary, the introduction of a control point can not affect the estimation but only increase the complexity of the quantile model. A naive approach would consist in adding a control point at each of the re-sampled optimum point \hat{x}_k^* . However, this would quickly result in set $\mathcal{X}^{(l)}$ with too many elements and a deteriorated estimation as a consequence. A more advanced approach is necessary.

We propose to proceed in two steps for the enrichment of $\mathcal{X}^{(l)}$. In the first step we propose a set of $P_X < P_{\text{new}}$ candidates, and in the second step we accept or reject the proposed enrichment. The first stage consists in the selection of the candidate points, and one can use a random subset of points in $\mathcal{Y}^{(l+1)} \setminus \mathcal{Y}^{(l)}$. The selection can be achieved picking randomly a fixed number $P_X < P_{\text{new}}$ of points (for instance $P_X = 1$) or ascribing a probability $r < P_X/P_{\text{new}}$ of selection to each of the points in $\mathcal{Y}^{(l+1)} \setminus \mathcal{Y}^{(l)}$. To avoid having to set a priori P_X , we used in this work an alternative approach; the points in $\mathcal{Y}^{(l+1)} \setminus \mathcal{Y}^{(l)}$ are clustered in $P_X < P_{\text{new}}$ groups using a standard K-Means Clustering algorithm. The Euclidean distance over Ω was considered to measure distance between points, and the number P_X of clusters was determined using the method of Silhouettes [47]. This procedure produces the set $\mathcal{Q} = \{\bar{x}_k, k = 1, \dots, P_X\}$ of cluster centers.

In a second stage, the insertion of \bar{x}_k in $\mathcal{X}^{(l+1)}$ is considered. Starting from $\mathcal{X}^{(l+1)} = \mathcal{X}^{(l)}$, the centroids in \mathcal{Q} are considered one after the other. Applying the Bayesian quantile estimation on $\mathcal{Y}^{(l+1)}$ and $\mathcal{X}^{(l+1)} \cup \{\bar{x}_k\}$, we accept the new point, and set $\mathcal{X}^{(l+1)} = \mathcal{X}^{(l+1)} \cup \{\bar{x}_k\}$, if it decreases the expected loss of the quantile model. Specifically, we estimate $\mathbb{E}(\mathcal{L}_\tau(\mathcal{Y}|\omega))$ by drawing samples of ω according to its density conditioned on $\mathcal{Y}^{(l+1)}$, that is $p_\tau(w|\mathcal{Y})$. In other words, we use the expected likelihood of the enriched model as a measure of goodness-of-fit. This measure was found more effective than carrying a full cross-validation procedure.

3.3.3. Optimization procedure

A schematic overview of the complete optimization procedure is shown in Figure 2. It starts, on the left, with the initial sample sets of F , $\mathcal{Y}^{(l=1)}$, and control points, $\mathcal{X}^{(1)}$, constructed using the methodology discussed in Section 2.4. The quantile model is then constructed on $\mathcal{Y}^{(l)}$ and $\mathcal{X}^{(l)}$, following the Bayesian approach and P_{new} minimums of independent realizations of $q_{\tau k}^*(x)$ are computed. The random function $Y(x, \theta)$ is sampled at these minimums to generate enriched $\mathcal{Y}^{(l+1)}$. Finally, new control points for the quantile approximation are proposed and accepted or rejected to form $\mathcal{X}^{(l+1)}$. The next optimization iteration can start, unless a stopping criterion is attained. Different stopping criteria can be thought-off. Iterations can be stopped when the computational budget is exhausted. In targeted applications, the computational budget will refer to the number of evaluations of the random function, $Y(x, \theta)$, whose cost is assumed to largely exceed all other computational costs (in particular the MH sampling steps). Stopping criteria based on the convergence of the optimum are more tricky to define as the objective function is never exactly evaluated and only estimated. A natural way to assess the convergence is to report the dispersion of the P_{new} optima generated at the l -th step of the optimization procedure. If all these optima agree well, it means that the

dispersion of the Bayesian estimate is low and their average should be representative of the true optimum x^* . Alternatively, one can look to the minimizer \tilde{x}^* of MLE $q_\tau^*(x)$ and assess the convergence of the optimization through the posterior variance of the quantile estimator at that location.

After the initial sampling, the number of required control points (model complexity) is selected by cross validation. Then, quantile regression is performed to globally obtain the quantile as a function of the design parameters. This is equivalent to the MLE estimation of the Bayesian model. The posterior distribution of the model is then obtained by MCMC. From the complete posterior distribution, P_{new} realizations are selected randomly, and the optimum location for each of these realizations is found. At these P_{new} optimum locations samples are drawn (randomly in the stochastic space) and evaluated in the full order model. After the additional sampling, it is studied whether the model complexity should be increased or not. Then the model is built again with the additional samples (and if required additional control points), and the process is repeated until convergence.

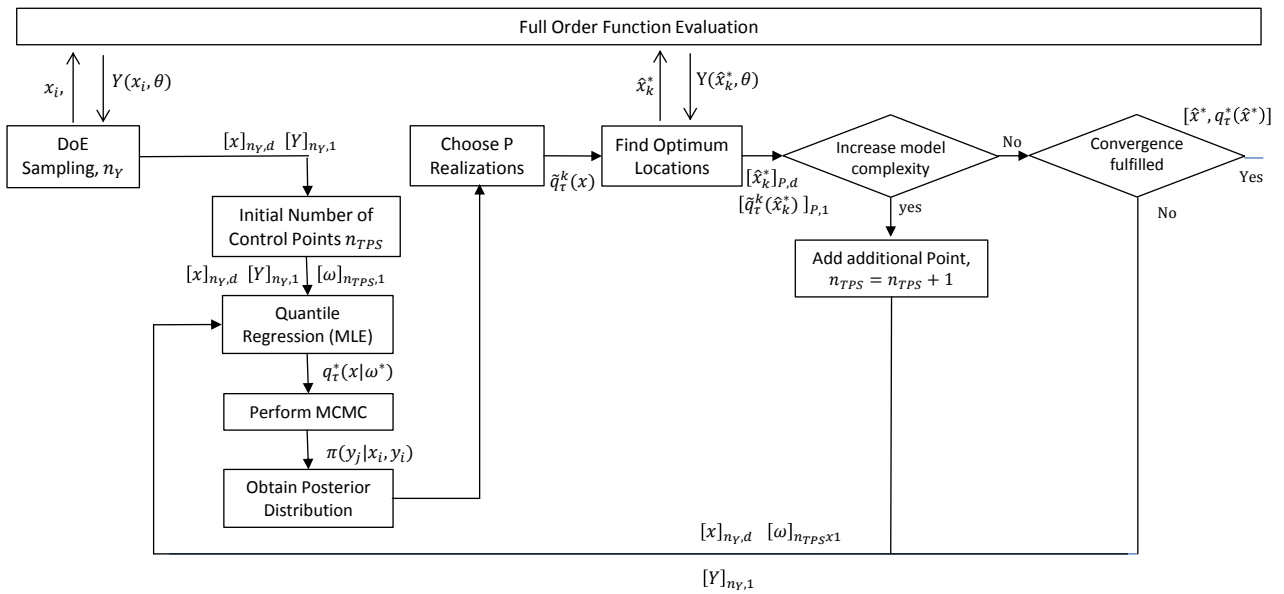


Figure 2: Overview of the optimization framework for Bayesian quantile regression.

4. Numerical Application to Test Functions

In this section, we apply the proposed framework to the optimization of multi-modal quantiles. The behaviour of the quantile estimation and of the complete optimization procedure is investigated on low dimensional settings, $\Omega \subset \mathbb{R}^d$ with $d = 1$ and 2 in Sections 4.1 and 4.2 respectively. In Section 5, the optimization framework will be applied to a complex optimization problem with engineering relevance.

4.1. 1D test function

For $x \in \Omega^{d=1} = [a, b]$, with $a = 2$ and $b = 8$, we consider the following random performance function,

$$Y^{d=1}(x, \theta) = 3 - 4 \exp(-4(x - 4)^2) - 5.2 \exp(-4(x - 6)^2) + \frac{x - a}{b - a} \xi_1(\theta) + \frac{b - x}{b - a} \xi_2(\theta), \quad (25)$$

where ξ_1 and ξ_2 are two independent random variables with uniform and Gaussian distributions, respectively,

$$\xi_1 \sim \mathcal{U}(0, 10), \quad \xi_2 \sim \mathcal{N}(1.01, 0.71). \quad (26)$$

We focus on the minimization of the 80% quantile, *i.e.*, $\tau = 0.8$. Although the stochastic dimensionality of the problem is low (having only two independent random sources), permitting the use of effective surrogate-based methods, it will be treated as high-dimensional assuming that $Y^{d=1}(\cdot, \theta)$ is infinitely dimensional and can only be sampled. In other words, we do not attempt to exploit the specific structure of the stochastic function. However, we exploit the low computational cost of the function evaluation to construct a reference estimate of the quantile using 10^6 Monte Carlo samples of (ξ_1, ξ_2) . Figure 3 shows this reference quantile, as well as 10^4 random samples of $Y^{d=1}(x, \theta)$ uniformly drawn in $\Omega^{d=1}$. The bi-modality in x of the quantile can be appreciated, with two comparable minimums located around $x = 4$ and $x = 2$. The objective is then to estimate the quantile accurately in these locations close and, eventually, determine the global optimum ($x \approx 4$) at minimal computational cost (from a minimum number of samples of Y).

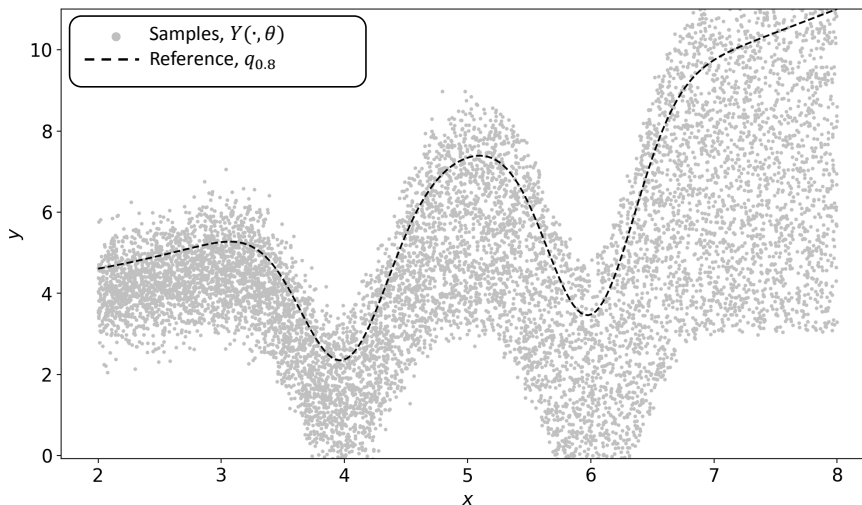


Figure 3: Reference quantile $q_\tau(x)$, for $\tau = 0.8$ and 10,000 random samples of $Y^{d=1}(x, \theta)$.

4.1.1. Control points selection

We first focus on the quantile estimator q_τ^* and the selection of the control points for the regression. Figure 4 reports the best estimators $q_\tau^*(x)$ for $N_{\mathcal{Y}} = 200$ and using $N_{\mathcal{X}} = 5$ (left), $N_{\mathcal{X}} = 13$ (center) and $N_{\mathcal{X}} = 20$ (right) control points for the regression (black squares), following Sobol sequences. The plots also report the reference quantile (dashed line) for comparison. We see that for $N_{\mathcal{X}} = 5$, the quantile estimator misses some features of the reference quantile (bias). On the other hand, the predictor for $N_{\mathcal{X}} = 20$ control points over-fits the data and becomes noisy (high variance). In contrast, the case with $N_{\mathcal{X}} = 13$ leads to an estimator with both low bias and low variance.

In Section 2.4 we discussed the selection of the number of control points supporting the quantile approximation through Cross-Validation based on the loss function in (6). Figure 5 shows typical evolutions of the CV-loss with $N_{\mathcal{X}}$. The CV loss is reported for two sample sets \mathcal{Y} with respective size $N_{\mathcal{Y}} = 50$ (left plot) and $N_{\mathcal{Y}} = 200$ (right plot), the latter corresponding to the case of Figure 4. In order to demonstrate that the CV loss is an appropriate measure of the quality of the estimator, we also report the RMSE of the estimator with the reference solution, $q_\tau^*(x) - q_\tau(x)$, computed at 500 linearly spaced locations along the design space. We observe a very high correlation between the RMSE and the CV-loss estimate, demonstrating that the latter is relevant to select the $N_{\mathcal{X}}$ when no reference for the quantile is available.

Further, the CV-losses are seen to present a minimum at $N_{\mathcal{X}} = 10$ and $N_{\mathcal{X}} = 13$ for $N_{\mathcal{Y}} = 50$ and $N_{\mathcal{Y}} = 200$ respectively. This finding is consistent with the expectation that a larger sample set allows for learning more structures in the quantile and thus necessitating larger $N_{\mathcal{X}}$. This aspect is illustrated further in Figure 6 where the RMSE of the best estimators $q_\tau^*(x)$, for fixed values of $N_{\mathcal{X}}$ and increasing sample sets \mathcal{Y} , are reported. When $N_{\mathcal{X}}$ is small, the RMSE of the estimator quickly stagnates when $N_{\mathcal{Y}}$ increases,

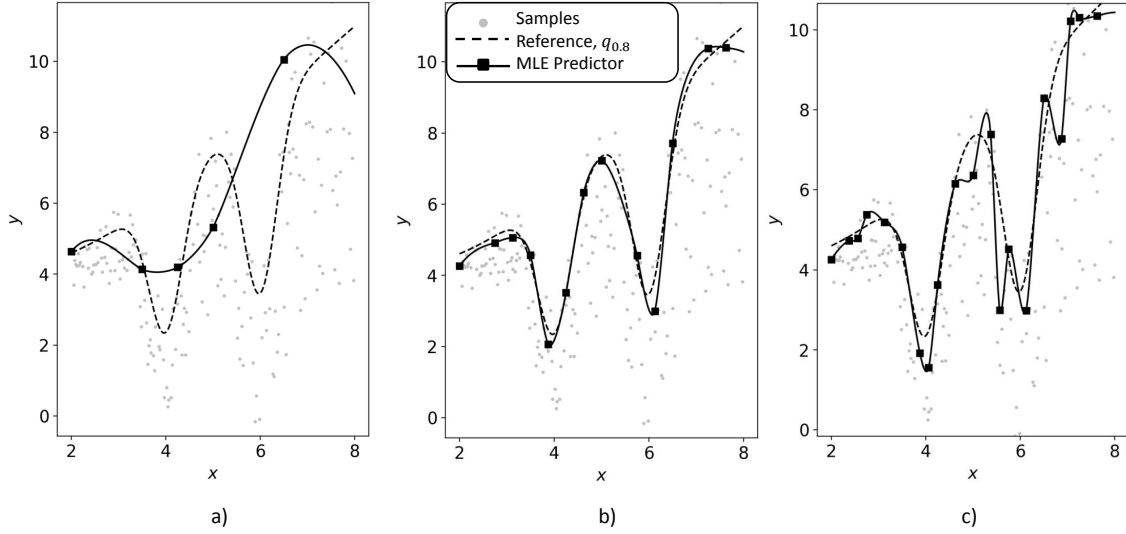


Figure 4: Quantile estimators q_τ^* for $N_{\mathcal{X}} = 5, 13$ and 20 (from left to right). The plots also depicts the $N_{\mathcal{Y}} = 200$ sample points of \mathcal{Y} and the reference quantile. The control points are reported as black squares.

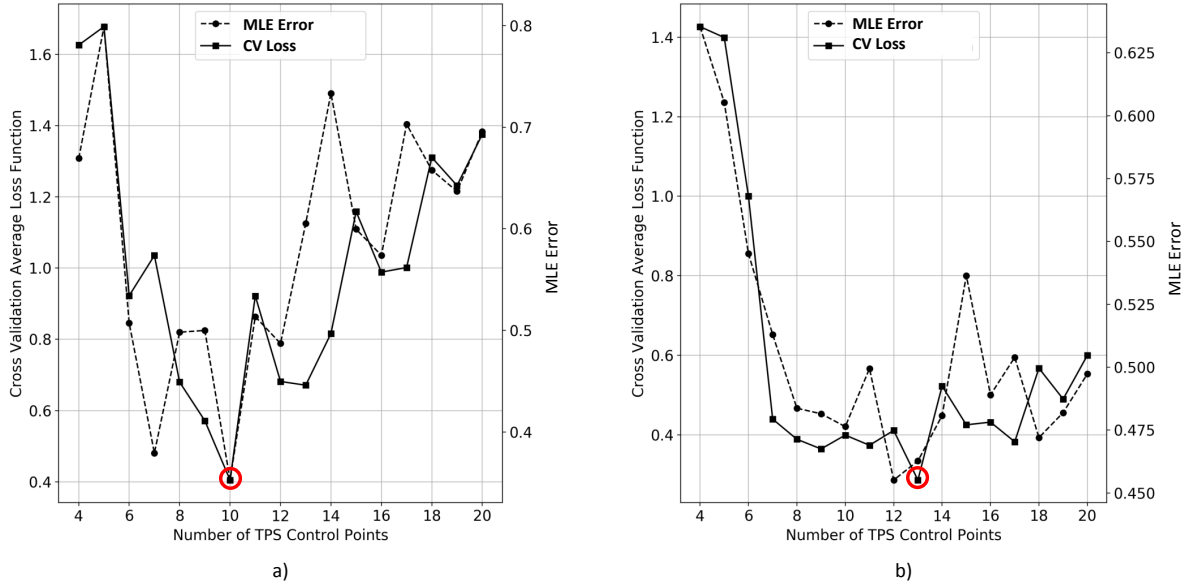


Figure 5: Cross-validation loss and RMSE (with respect to the reference) of the best estimator $q_\tau^*(x)$ as functions of $N_{\mathcal{X}}$ the number of control points: a) Case of $N_{\mathcal{Y}} = 50$, b) case of $N_{\mathcal{Y}} = 200$.

denoting a poor quantile discretization enable to fit the feature of the true quantile. The intermediate value $N_{\mathcal{X}} = 13$ is providing the best estimator among the values tested till $N_{\mathcal{Y}} \approx 1,000$, when the model with $N_{\mathcal{X}} = 20$ becomes better while no improvement is observed for $N_{\mathcal{X}} = 13$ after $N_{\mathcal{Y}}$ exceeds 4,000. These results illustrate the importance of balancing the quantile discretization error to the sample set available for the construction.

4.1.2. Optimization

We return to the robust optimization problem and the minimization of the τ -quantile.

The optimization procedure is initialized with $N_{\mathcal{Y}} = 50$ random samples uniformly drawn in $\Omega^{d=1}$. The initial estimate of the quantile uses $N_{\mathcal{X}} = 10$ control points, following the results presented in Figure 5. The

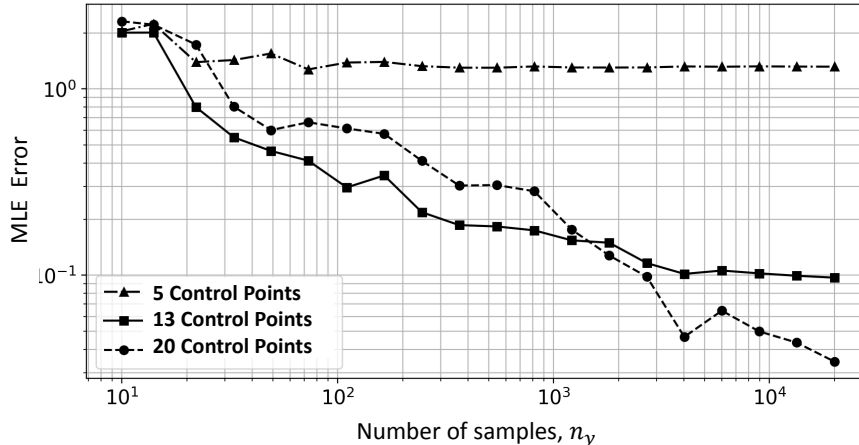


Figure 6: RMSE of $q_r^*(x)$ as a function of the sample set size N_Y and for different number of control points ($N_{\mathcal{X}}$) as indicated.

Bayesian approach is subsequently employed to draw a sample set of 30,000 realizations of the estimator from its posterior. The 95% confidence interval of the estimator is derived from the sample set and reported as the shaded area in the top plot of Figure 7. Also shown are the reference, MLE, and posterior averaged estimators of the quantile. The confidence interval indicates a high estimation variability, which is explained by the small size of the sample set \mathcal{Y} . The bottom plot of Figure 7 depicts the histogram of the posterior minimum x_k^* of the Bayesian quantile estimator. The histogram is built on a subset of 1,000 realizations extracted from the MCMC sample set. The histogram shows that the sample set of minimums concentrates in two areas around $x \in [3.5, 4.2]$ and $x \in [5.8, 6.2]$. The two domains correspond to the two most significant local minima, and their presence highlights the impossibility to decide on the global minimum for such low size \mathcal{Y} . The sample-set \mathcal{Y} is then enriched with $P = 10$ new sample points selected randomly from the set of realizations of the optimum x_k^* , so the new sample points follow the distribution depicted by the histogram in the bottom plot of Figure 7.

After the addition of these new sample points, the quantile estimation is updated, testing the insertion of new control points in the quantile discretization. The tentative new control points are proposed as the cluster's centers of the sample set of minima x_k^* . In this example, the method of silhouettes and K-means clustering split the sample set into 2 clusters, such that two new control points are proposed at this stage of the algorithm. However, none of the points is accepted, as it does not improve the marginal likelihood of the model, and \mathcal{X} is not enriched. Still, the procedure continues, restarting from the sampling of the posterior, using now the likelihood of the enriched \mathcal{Y} .

Figure 8 presents the Bayesian estimator of the quantile after four iterations of the optimization procedure (so now $N_Y = 90$). We observe that, compared to the previous case, the MLE predictor fits better the reference quantile. The additional sample points (dark grey dots) have been inserted at locations x close to the two local minima, where the estimation has improved the most. Further, the confidence interval in the quantile prediction around the local minima has drastically reduced, denoting that the information mostly concerns the regions of interest.

At this stage of the optimization, the information in \mathcal{Y} is sufficient to make beneficial the insertion of a new control point near the local minima at $x = 4$. The effect of this insertion is shown in Figure 9. Specifically, the new control point reduces the bias of the MLE estimator without introducing significant variance (without a noticeable increase of the confidence intervals).

Figure 10 shows the estimator after 20 optimization iterations, corresponding to a sample set with size $N_Y = 250$. The accuracy of the estimation near the two local minima has greatly improved, with restricted confidence intervals, as most samples fall in these two regions. This improvement leads to the determination with high confidence in the global optimum location in the neighborhood of $x = 4$.

To better appreciate the convergence of the optimization procedure, Figure 11 details the evolutions

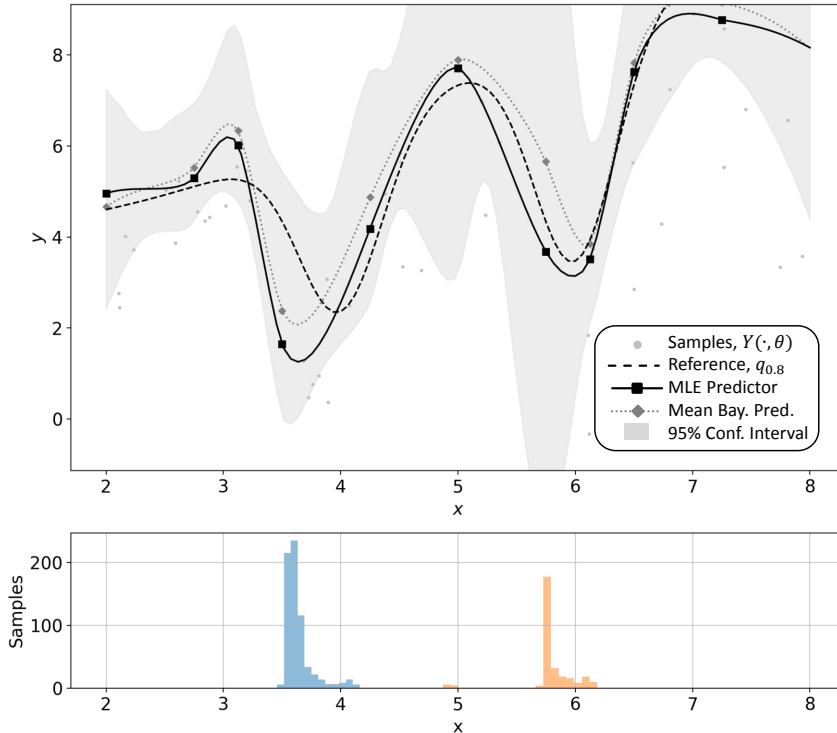


Figure 7: Bayesian estimator of the quantile based on the initial sample sets with $N_{\mathcal{Y}} = 50$ sample points.

with the optimization iterations (so, essentially a measure the computational cost of the method) of the best estimate of the quantile minimum location x^* and value $q_r^*(x^*)$. The figure compares our Bayesian approach to draw the new sample points and the insertion of new control points, with a naive approach that enriches \mathcal{Y} with the same number of samples but selecting their locations x_i randomly in the domain Ω . The second approach is not converging, for the considered number of iterations, while our approach achieves fairly precise estimates of $q_r^*(x^*)$ and location x^* . The naive approach overestimates the value at the optimum and keeps switching the best location from one local minimum to another. In contrast, our procedure converges well on the location of the reference global optimum, although the quantile MLE at this point is still affected by a noticeable error ($\approx 5\%$) after 20 iterations. The bars on the value of q_r^* corresponds to the 95% confidence interval of the Bayesian analysis. That means, they represent the confidence in distribution of the possible optimum values from different realizations of the model that originate from the Bayesian perspective. One can verify that the bars are much smaller in our approach, reflecting higher confidence in the optimal quantile value, and contain the reference value. The level of error in the minimal quantile value achieved by our method is consistent with the amount of information in the sample set with a size of $N_{\mathcal{Y}} = 250$ only.

This simple experiment illustrates the effectiveness of the method to determine fairly accurately the location of the global optimum of the quantile, without having to rely on estimates of the quantile at each new evaluation point x_i of the optimization procedure. We underline that at each iteration, $P = 10$ samples only need to be generated. If one assumes that $\mathcal{O}(100)$ samples would be necessary for the point-wise estimation of the 80% quantile at a given x , only 2 to 3 optimization points would be feasible for the same computational budget. Exploiting the structure of Y and observing that it is a mixture of a uniform and a Gaussian random variables at all x , one could rely on (linear) surrogates of $Y^{d=1}(\cdot, \theta)$ to compute exactly (or with high accuracy) from a low number of samples, in fact, three samples for each x_i . Thus, even in this ideal situation, less than 100 optimizations points could be evaluated for the same computational budget; we also recall that not all situations of interest will be reducible to such low dimensional surrogate evaluation.

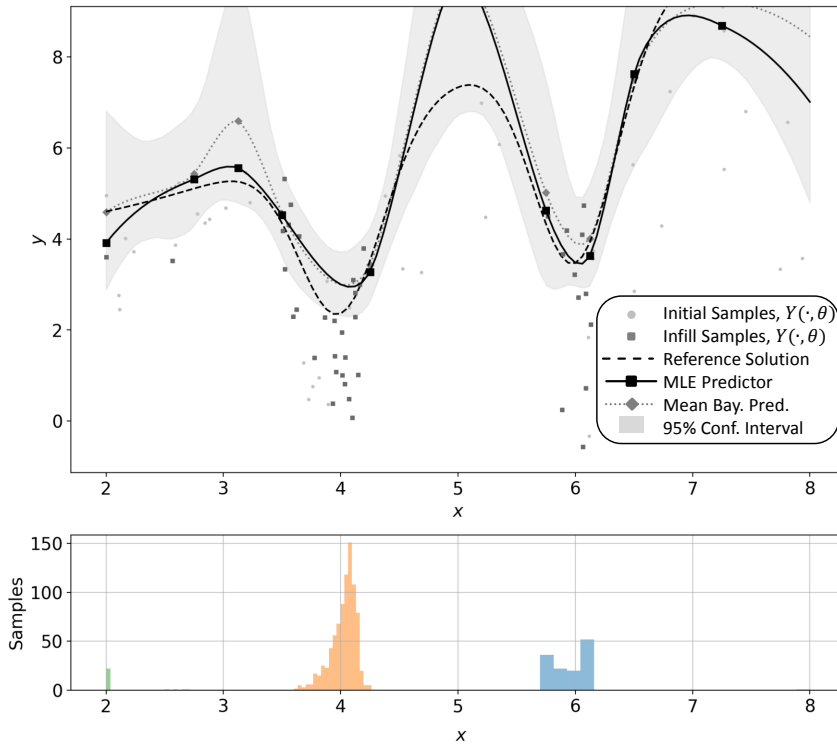


Figure 8: Bayesian quantile estimator after 4 iterations of the optimization procedure.

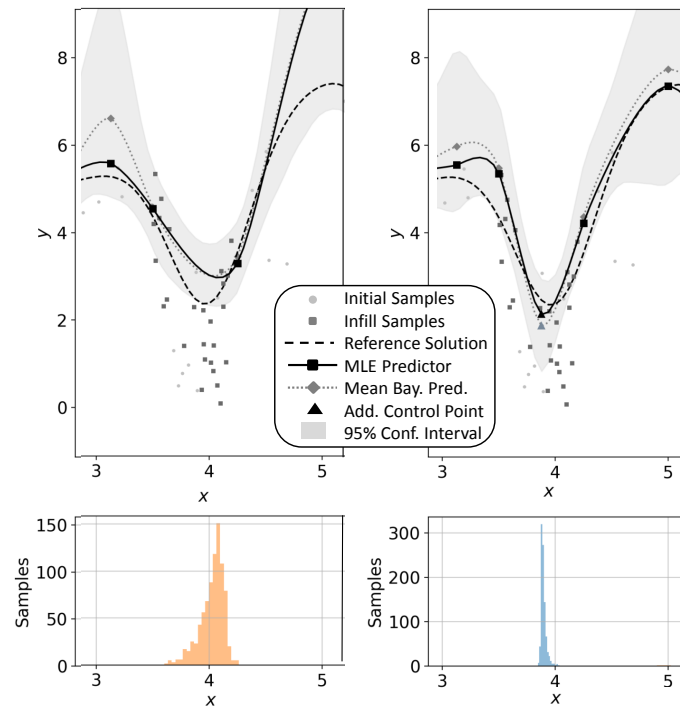


Figure 9: Effect of a new control point on the quantile estimation. Left: before insertion. Right: after insertion.

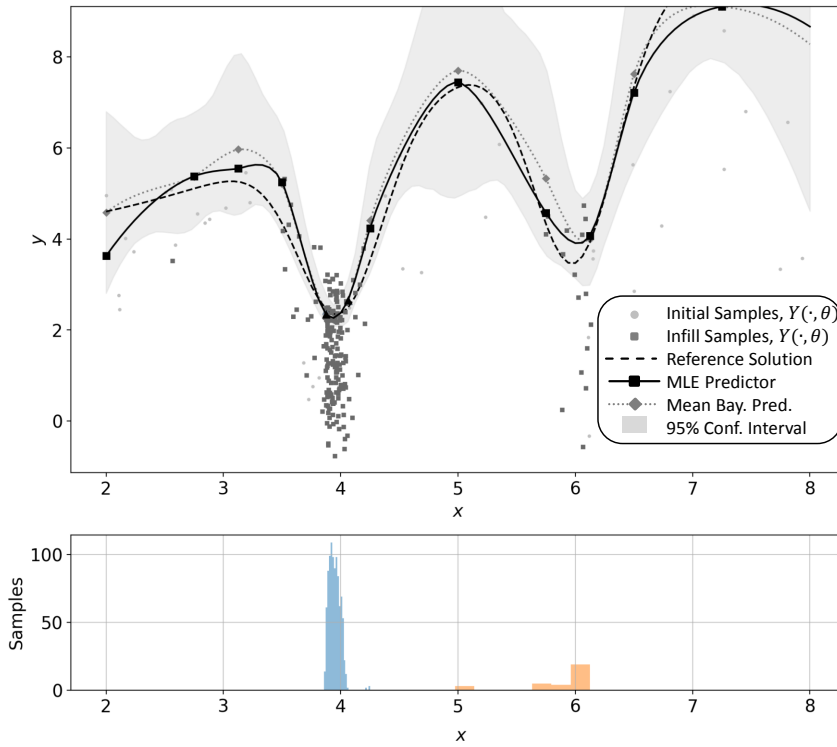


Figure 10: Quantile estimation after 20 iterations of the optimization procedure ($N_y = 250$).

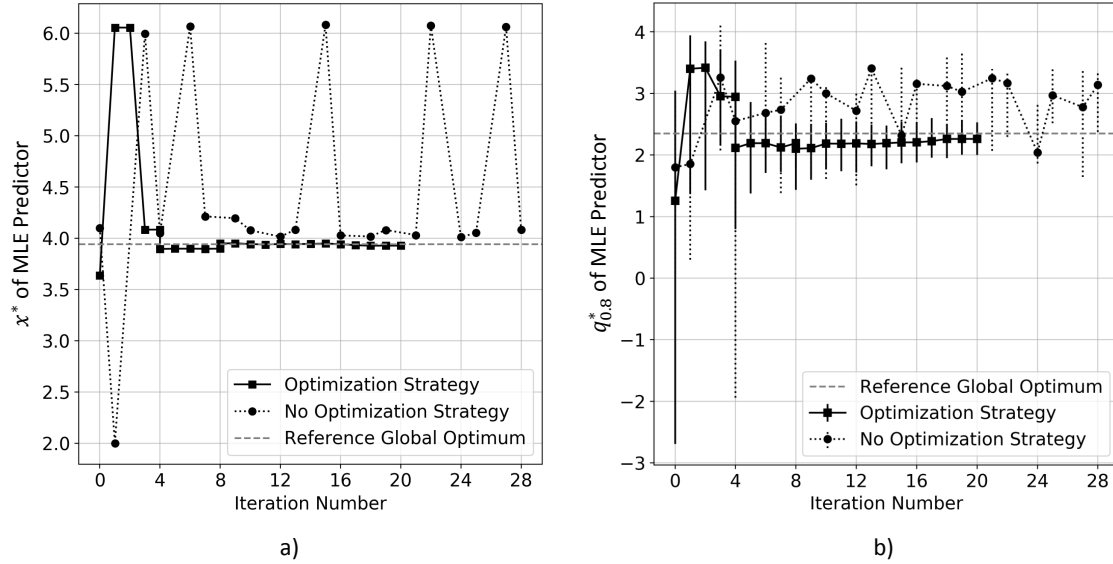


Figure 11: Convergence of the optimization process for our Bayesian and naive approaches for 1D test function. Left: estimate of the optimal quantile. Right: estimate of optimal quantile location.

4.2. 2D test function

In this section we extend the two dimensions ($d = 2$) the previous test case, setting $x = (x_1, x_2) \in \Omega^{d=2} = \Omega^{d=1} \times \Omega^{d=1}$, and letting

$$Y^{d=2}(x, \theta) = Y^{d=1}(x_1, \theta) + Y^{d=1}(x_2, \theta). \quad (27)$$

Note that in the definition of the this test case, x_i refers to the i -th component of x and not to a sample index as previously. The reference estimate of the 80%-quantile is shown in the left plot of Figure 12. The quantile $q_{0.8}(x)$ has 4 local minima in $\Omega^{d=2}$, the global one being at $x^* \approx (4, 4)$. The right plot of Figure 12 corresponds to the initial quantile MLE obtained for an initial sample set \mathcal{Y} of $N_{\mathcal{Y}} = 200$ points (small black dots). It was determined that $N_{\mathcal{X}} = 32$ control points (black circles), from a Sobol sequence, was providing the minimum CV-loss. Despite the limited sample set, the quantile regression captures the multi-modality of the $q_{\tau}(x)$.

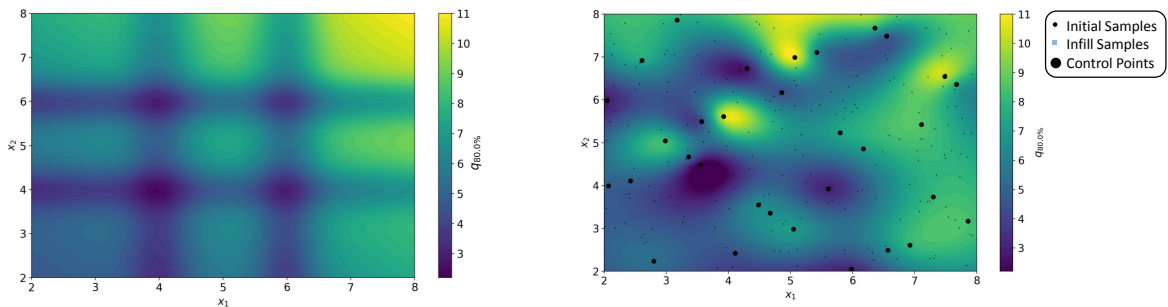


Figure 12: Reference 80%-quantile of $Y^{d=2}$ defined in (27) (left) and its initial MLE (right) based on 32 control points and 200 samples.

Proceeding with the optimization, we consider the enrichment of \mathcal{Y} with $P = 6$ new sample points at each iteration. Figure 13 shows the sample set points after 42 optimization iterations, as well as the resulting quantile MLE and the added control points (black triangles). The figure shows how the new samples are drawn in the neighborhood of the global optimum.

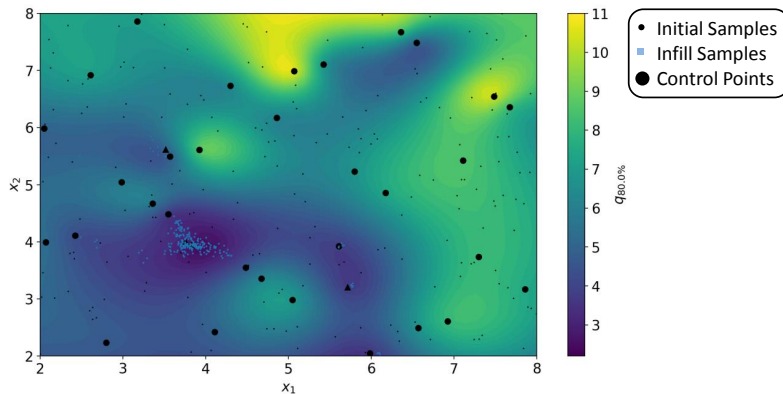


Figure 13: MLE of the 80% quantile after 42 optimization iterations.

Figure 14 reports the convergence of the estimated location of the optimal point x^* and its quantile value $q_{\tau}^*(x^*)$. The case of the naive sampling is also reported for comparison. Concerning the convergence of the optimum point x^* , it is seen that our method can provide a correct estimate in only ten iterations ($N_{\mathcal{Y}} = 260$), and a relatively well converged minimal point after roughly 28 iterations ($N_{\mathcal{Y}} = 368$), while the naive approach is still hesitating between the different local minima after twice as many iterations

($N_y = 536$). Concerning the convergence of the minimum value, about 40 iterations ($N_y = 440$) are necessary to obtain a reasonable estimate with an error of less than a few percents, while the naive approach exhibits no convergence at all. We also remark that the real optima lies within the 95% confidence interval when using our method.

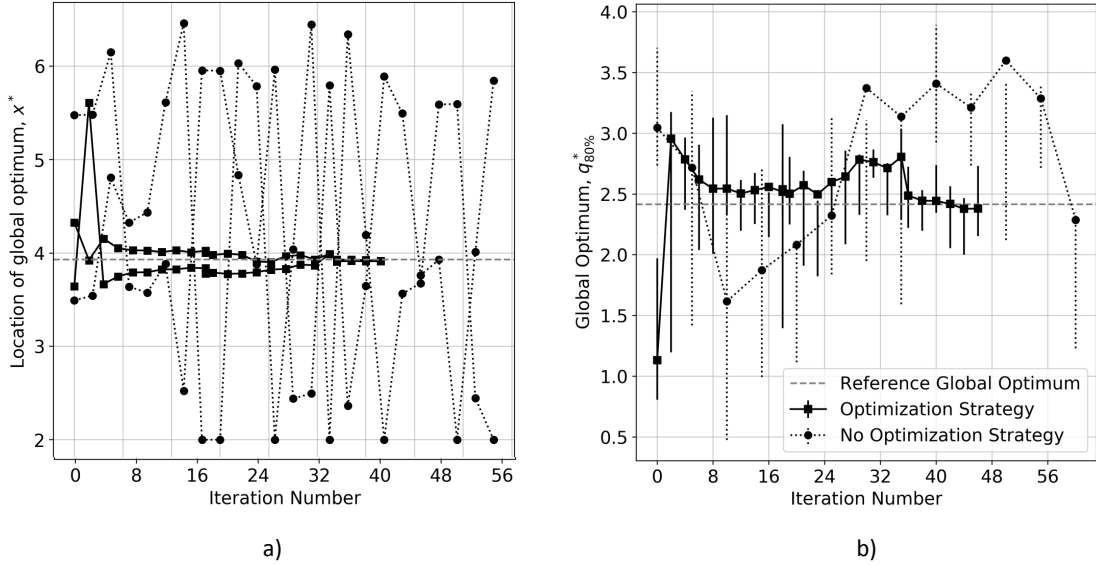


Figure 14: Convergence of the optimization process for our Bayesian and naive approaches for 2D test function. Left: estimate of optimal quantile location. Right: estimate of the optimal quantile.

5. Optimization of a Shock Control Bump under Geometrical and Operational Uncertainties

In this section, we apply the quantile optimization framework to a representative aerodynamic shape optimization problem under a total of 382 independent uncertainty sources.

5.1. Problem definition

The problem concerns the robust design of a Shock Control Bump (SCB) retrofitted to an airfoil subject to operational and geometrical uncertainties. SCBs are passive devices that mitigate normal shock waves over the airfoil in the transonic flow regime [48, 49]. The effectiveness of these devices is highly sensitive to the operational conditions (*e.g.*, free-stream Mach number, angle of attack or lift coefficient), and manufacturing tolerances (geometry) [8]. These uncertainties can move the shock wave location upstream or downstream, and the effect of the bump becomes much less effective when the shock departs from its design location. Therefore, it is crucial to account for uncertainties when designing the SCB.

The geometry of the SCB is parametrized by a modified Hicks-Henne bump function, following Maza-heri [50] and Tian [51]:

$$Z_{\text{bump}}(X) = H_{\text{bump}} \left[\sin \left(\pi \left(\frac{X - X_{\text{start}}}{L_{\text{bump}}} \right)^{M_{\text{bump}}} \right) \right]^{T_{\text{bump}}}, \quad X_{\text{start}} \leq X \leq X_{\text{start}} + L_{\text{bump}}, \quad (28)$$

where

$$M_{\text{bump}} = \frac{\log(0.5)}{\log(X_{H_{\text{bump}}})}. \quad (29)$$

The five design parameters are the bump maximum height H_{bump} , the maximum height location $X_{H_{\text{bump}}}$, the starting location X_{start} , the bump length L_{bump} and the slope factor T_{bump} . The locations and height

are normalized by the chord-length, and the extent of the design domain Ω is specified in Table 1. Figure 15 illustrates the parametrization of the bump and how the deformation Z_{bump} modifies the original airfoil geometry, $Z(X)$, to the retrofitted geometry.

Parameter	Description	Lower bound	Upper bound
H_{bump}	Bump maximum height	0	0.015
$X_{H_{bump}}$	Bump maximum height location (as fraction of L_{bump})	0.4	0.85
L_{bump}	Bump length	0.15	0.45
X_{start}	Bump starting location	0.3	0.54
T_{bump}	Bump slope factor	0.2	2

Table 1: Design range for the optimization of the SCB.

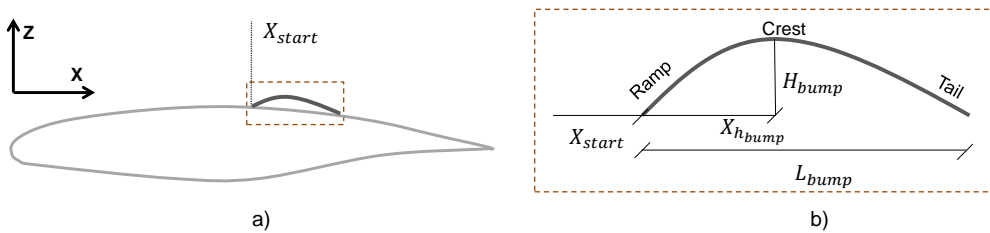


Figure 15: a) Bump relative to the airfoil; b) design parameters defining the bump shape

Consistently with the notation of the previous sections, the design vector is $x = (\in \Omega \subset \mathbb{R}^5)$, and we consider an optimization based on C_D , the airfoil's drag coefficient, at a constant speed and lift. Because of operational and geometrical uncertainties, C_D is not a deterministic quantity, and we seek for the optimal design that minimizes the τ -quantile of $C_D(x, \theta)$. We consider two different optimizations: one focusing on adverse situations (those that would lead to a high value of drag) using the 95% quantile, and the second focusing on day-to-day conditions with the minimization of the 50% quantile (median) of the drag.

5.1.1. Geometrical Uncertainties

Several approaches are able to handle uncertainty quantification of partial differential equations with random geometries [52, 53, 54]. A basic procedure in Computational Fluid Dynamics (CFD) based optimization consists in considering random perturbations $Z_{Geo}^{\pm}(X)$ of the nominal geometry, where the \pm indicates the intrados and extrados of the airfoil and X is again the normalized coordinate along the chord line. In the present work, the nominal geometry is the RAE2822 airfoil [55]. For simplicity, we shall assume here that $Z_{Geo}^{\pm}(X)$ have zero mean and follow Gaussian distributions with covariance function

$$C(X, X') = \sqrt{\sigma_{Geo}^2(X)\sigma_{Geo}^2(X')} \exp\left(-\frac{|X - X'|^2}{2L^2}\right), \quad (30)$$

where the variance σ_{Geo}^2 of Z_{Geo}^{\pm} varies with X , and L is the correlation length of the stochastic processes. The covariance in (30) is symmetric, semi positive definite for $\sigma_{Geo}^2 \geq 0$, and corresponds to smooth processes [56]. In the following, we set $L = 0.1$ and $\sigma_{Geo}^2(X) = (\sigma_{max}X(X - 1))^2$ such that Z_{Geo}^{\pm} vanish at the leading and trailing edges. This behavior at the leading and trailing edges is necessary to avoid non-matching deviations at $X = 0, 1$ that could lead to convergence problems with the mesh generator (negative cells) and CFD solver. The geometrical uncertainty is the largest at mid-chord ($X = 0.5$), where the standard deviation tops to $\sigma_{max} = 0.00022$ (unit of chord length). This standard deviation corresponds to 2.3 mm in the case of the Airbus A330-300 wing with a chord root of 10.56 m [57].

In practice, the discretization of the airfoil uses computational nodes X_i^{\pm} on the upper and lower surface of the nominal geometry. This set of nodes supports the discretization of the covariance functions into covariance matrices Σ_{Geo}^{\pm} . The sampling of $Z_{Geo}^{\pm}(X_i)$ proceeds by standard techniques relying on a spectral

decomposition of the respective covariance matrices $\Sigma_{\text{Geo}}^{\pm}$. The numerical example below uses a total of 380 computational nodes on the two surfaces, leading to 380 geometrical uncertainties. Figure 16 shows, in the left plot, 20 realizations of the Gaussian Field Z_{Geo}^+ (extrados). The right plot shows 20 realizations of the uncertain airfoil geometry; the deviations have been increased ten times for clarity.

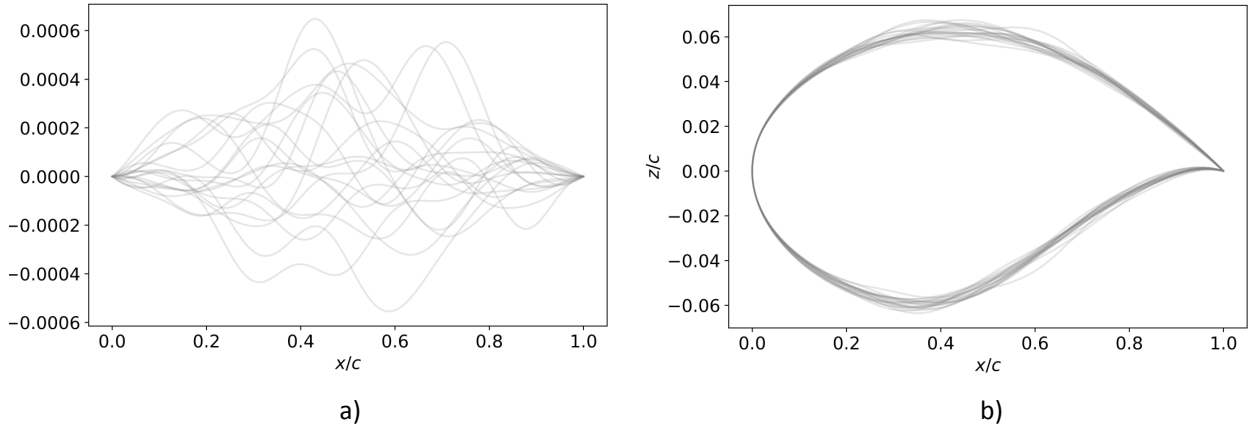


Figure 16: Geometrical uncertainties with 20 random realizations of $Z_{\text{Geo}}^+(X)$ (a) and 20 realizations of the resulting airfoil geometry (b) where Z_{Geo}^+ has been amplified ten times for visibility.

We observe that the sampling approach of realizations is direct, in contrast to alternative reduction approaches that would have exploited the correlation of Z_{Geo}^{\pm} to derive lower-dimensional parametrizations of Z_{Geo}^{\pm} , for instance by truncated Karhunen-Loeve expansions (KLE) [5, 17, 58]. Such reduction is unnecessary in our approach as the quantile regression is insensitive to the dimensionality of the uncertainty sources. Finally, although the normal distributions used to sample the stochastic process are unbounded, the standard deviations considered are small enough to always result in well-posed problems in our numerical simulations.

5.1.2. Operational Uncertainties

We consider two operational uncertainties, the lift coefficient C_L and Mach number Ma , which have the most substantial influence in the shock wave location. The Mach number relates to the cruising speed of the aircraft, while the C_L relates to its weight. For these two operational parameters, we assume symmetric Beta distributions centered on the nominal condition. The shape parameters are set to $\alpha_1 = \alpha_2 = 5$ in both cases. For these parameters, the Beta-distribution resembles truncated normal distributions. The location β_1 and scale β_2 parameters are fixed to obtain prescribed mean μ and standard deviation σ as reported in Table 2. The standard deviations of the Ma and C_L represent typical changes in flight speed and aircraft weight during a flight segment [5, 8].

Parameter	Description	Mean value	Standard Deviation
Ma	Freestream Mach number	0.734	0.0045
C_L	Lift coefficient	0.791	0.0045

Table 2: Operational Uncertainties

5.1.3. CFD Solver

The flow simulations for the evaluation of the drag coefficient rely on DLR CFD solver TAU [59]. This code solves the compressible Navier-Stokes equations, complemented with Reynolds Average Navier Stokes (RANS) Spalart-Allmaras turbulence model. In each simulation, the solver adjusts the airfoil's angle of attack to obtain the prescribed C_L . The discretization uses a central flux discretization, upper/lower symmetric Gauss-Seidel scheme, with a backward Euler time-implicit integration scheme. Figure 17 shows

the hybrid unstructured mesh of the nominal geometry; it consists of approximately 29,000 mesh nodes constituting quasi-two-dimensional, tetrahedral, and prismatic cells. The mesh deformation tool developed by DLR [60] is employed to deform the meshes around the uncertain airfoil and the SCB. The deformation uses the linear elasticity theory to propagate the displacement from the nominal geometry of the surface nodes X_i^\pm inside the computational domain. The TAU built-in partitioner is used to decompose the computational domain into four regions, for parallel computations. The density equation residual is reduced to 10^{-8} using a multigrid V-cycle. The discretization parameters were set according to previous studies on shape optimization [5], where the discretization error was estimated to be ≈ 0.1 Drag Count (DC, $C_D = 1 = 10,000$ DC). The drag coefficient, C_D , is obtained by the integration of the pressure and viscous forces along the airfoil surface. DLR’s Flow Simulator simulation environment [61] controls the process chain and manages convergence, restarts, and parallel evaluations. Overall, the computational time of a single sample amounts to roughly 2 minutes on average.

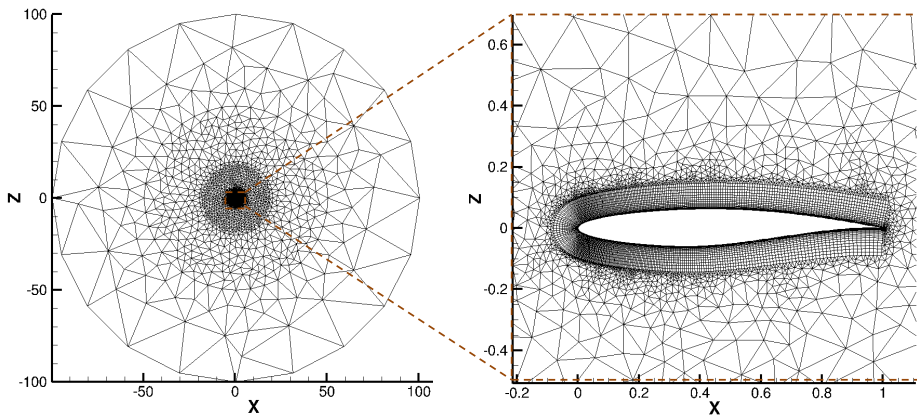


Figure 17: Typical CFD mesh for the estimation of C_D .

Before engaging in the robust optimization, we start by presenting the deterministic optimization solution, corresponding to the nominal geometry and mean values of the operational conditions. Classical Surrogate Based Optimization [46] is used for this optimization. The optimal SCB is found in 116 iterations. The bump reduces the drag of the baseline configuration by 23.9% through an isentropic deceleration of the shock wave. This effect is visible in Figure 18, which reports the pressure coefficients distributions on the baseline airfoil (nominal geometry) without (left) and with (right) the deterministic optimal SCB. As mentioned before, this optimum is not robust to operational and geometrical uncertainties.

5.2. Robust Quantile Optimization

We use $N_{\mathcal{Y}} = 500$ samples computed in parallel to initialize the optimization procedure. The initial samples are chosen according to a Sobol sequence [36] in the five-dimensional design space Ω . For each sample that represents a different design (different SCB shape), we draw a realization of the uncertain geometry (Z_{Geo}) and operational conditions (Ma, C_L), and compute the corresponding realization of the drag coefficient. With the initial samples set, quantile regressions are performed with the CV procedure to select the numbers of control points $N_{\mathcal{X}}$ in the quantile regression. The CV losses are shown Figure 19 for the two quantiles and variable $N_{\mathcal{X}}$. For $\tau = 0.5$ the CV-loss is minimal at $N_{\mathcal{X}} = 70$, while it is minimal at $N_{\mathcal{X}} = 50$ for $\tau = 95$. This difference is due to the size of the sample set, which, for the same number of control points (same model complexity), is likely to induce more over-fitting when τ approaches 1 or 0. Note also that the minimum is more pronounced in the latter case, consistently with the previous remark.

The robust optimization procedure is then engaged. At each iteration, we perform the selection of 12 new samples to enrich \mathcal{Y} , following the Bayesian approach. This calls for the sampling of high dimensional posteriors, with initially 70 ($\tau = 0.95$) and 50 ($\tau = 0.5$) dimensions respectively. In our numerical experiments, the adaptive MCMC sampler was robust enough and capable of exploring these high dimensional

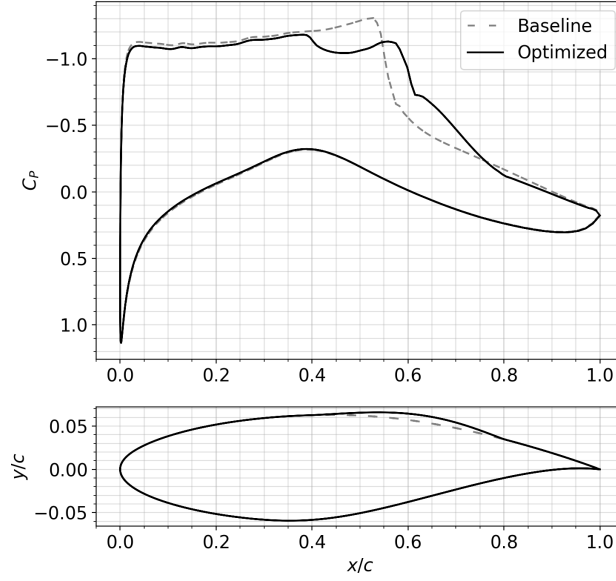


Figure 18: Pressure coefficients on the baseline airfoil without (left) and with (right) the deterministic optimal SCB.

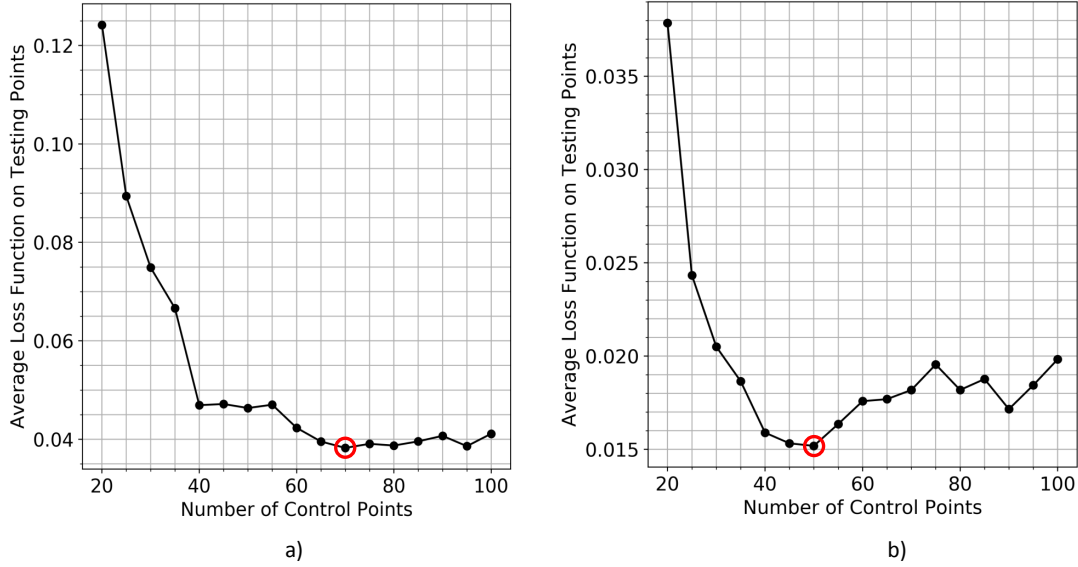


Figure 19: CV-losses as a function of the number $N_{\mathcal{X}}$ of support points. $N_{\mathcal{Y}} = 500$ and case of $\tau = 0.5$ (a) and 0.95 (b).

posterior distributions properly. After the selection of the new design points, their associated operational conditions and geometrical deviations are randomly set, and their drag evaluated in parallel. The quantile estimations are updated, and the procedure iterates until the optimization converges.

5.2.1. Convergence of the optimization

The optimization converges in 175 iterations for $\tau = 0.5$. The left plot in Figure 20 shows the evolution with the iterations of the estimated minimal drag quantile expressed in drag count units. The error bars on the quantile values represent the 95% confidence interval according to the distribution of the candidate optima solutions at each iteration. These are quickly reducing denoting the confidence in the final estimate

of the optimum. Also, two additional control points were automatically introduced at iterations 6 and 92 to increase the accuracy in the neighborhood of the global minima.

In the case of $\tau = 0.95$, shown in the right plot of Figure 20, more iterations are needed to converge, and even after ≈ 220 iterations the error bars on the estimator are not entirely reduced. In this case, the regression has used a total of five additional control points, introduced at iteration 20, 74, 123, 136, and 171. The effect on the optimum estimates of the new control points is seen, especially in the early stages of the optimization.

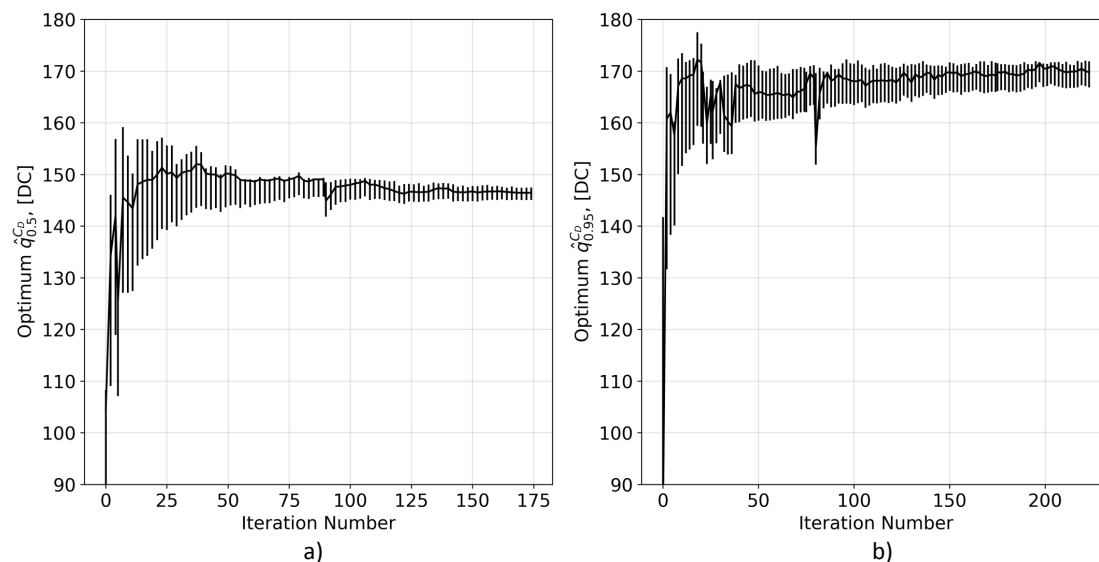


Figure 20: Convergence history of the optimum quantile for $\tau = 0.5$ (left) and $\tau = 0.95$ (right).

Figure 21 reports the standard deviation of the possible candidate optima at each iteration. The standard deviations are globally decaying with the iteration index, but noisily with spikes and jumps to higher values. This noise is due to the probabilistic nature of the quantile estimation and to the introduction of new control points that, although reducing the CV loss, induce a local increase of the regression's standard deviation. However, one can appreciate the global convergence of the optimization by taking the lower envelope of the curves. As one would have expected, for the same sample set size, the standard deviation of the 50% quantile at the optimum is lower than for the 95% quantile, underlying the more challenging character of the estimation. Further, the global reduction in the standard deviation seems to agree with the expected asymptotic rate of samples-based quantile estimation in $\mathcal{O}(1/\sqrt{N_{\mathcal{Y}}})$.

Figure 22 shows the location in Ω of the sample points in \mathcal{Y} and \mathcal{X} at the end of the 95% quantile optimization. Each plot corresponds to the projection of the sample points on a 2D plane of 2 design parameters. The SCB parameters x have been mapped to the unit-hypercube by an affine transformation. Histograms of the sample points in \mathcal{Y} , for each design parameter, are also provided. The plots indicate that the majority of the samples added to \mathcal{Y} during the optimization (blue points) fall in a few clusters underlying the existence of competing minima in the quantile. Most samples not belonging to a cluster are the uniformly distributed initial sample points. The multi-modality of the histograms also reflects the presence of competing minima. The clustering of the new samples (blue), in particular in the neighborhood of the final optimal design point (red circle), demonstrates that the Bayesian quantile regression can identify regions of interest and focus the computational effort in these areas. Concerning the additional control points (black), they all fall within the clusters, where it matters to improve the regression.

5.2.2. Analysis of the robust optima

We generated a sample set of 10,000 realizations of the operational and geometrical uncertainties and computed the corresponding realizations of the drag for the different optima: the deterministic optimum,

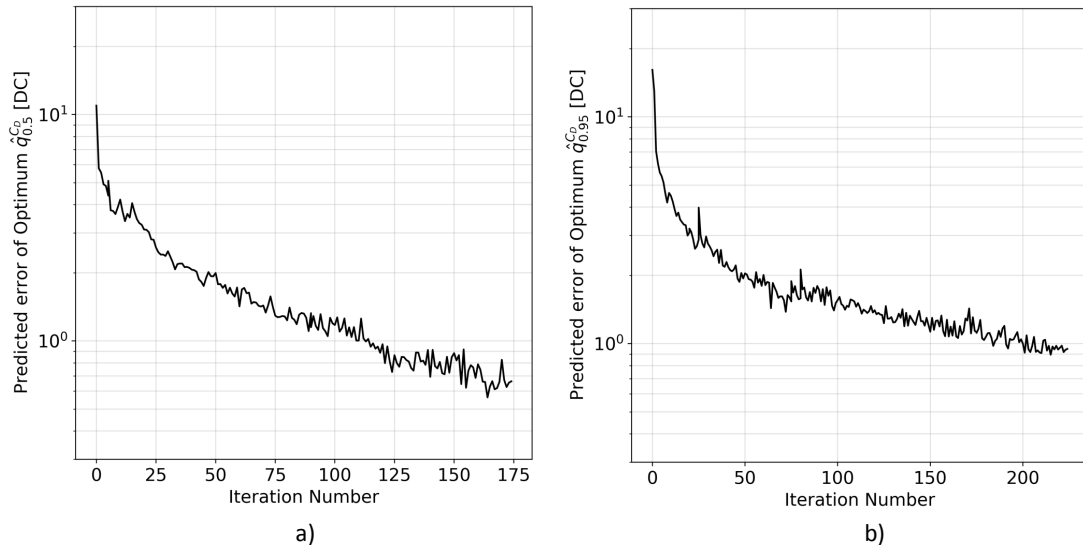


Figure 21: Convergence history of standard deviation of possible candidate optima for $\tau = 0.5$ (a) and $\tau = 0.95$ (b).

the $\tau = 0.5$, and $\tau = 0.95$ optima. These realizations allow assessing the quality of the quantiles Bayesian estimate. For $\tau = 0.5$, the differences between the MLE estimate of the quantile at the optimal design has a difference of less than 0.3% (1.35 drag counts) with the samples estimate. The difference is more significant for $\tau = 0.95$, as expected, and is equal to 1.33% (3.4 drag counts). These differences are consistent with the posterior's standard deviation. In our opinion, the accuracy of these estimates is quite reasonable, considering the total number of samples computed.

Figure 23 illustrates the effects of the different optimal SCB. For each design, the plot shows the distribution of the stochastic C_D resulting from the uncertain operational condition and geometrical variabilities. We used violin representations, with horizontal thickness proportional to the density of C_D (estimated by a kernel method from the samples set). The violins are superposed with the quartiles and the 5% and 95% quantiles. The baseline case, which corresponds to the airfoil with geometrical and operational uncertainties but no SCB, is also presented in the leftmost part of the plot. The improvement due to the deterministic optimal SCB is evident (see orange violin), with a general reduction of the drag compared to the reference. However, we observe a rather long-tailed distribution. The case of the 50% quantile optimum is performing quite similarly to the deterministic optimum, with slightly improved quartile values (in particular, the median gains more than one drag count). However, the tail, and generally the support, of the drag distribution remains quite broad. In contrast, the 95% quantile optimum presents a significantly more compact drag distribution. The inter-quartile distances are roughly halved compared to the deterministic optimum. This higher robustness comes at the expense of the median drag value that increases by less than six drag counts. Table 5.2.2 summarizes the performance of the different airfoils and provides the optimal parameters and number of simulations performed. The robust optimizations require about 20 to 30 times more CFD evaluations to converge than the deterministic optimization. In the present example, the hardest optimization problem ($\tau = 0.95$) required a total of 3,320 CFD computations. Further, the new samples (here 12) at a given iteration are computed in parallel, such that the computational overhead of the robust optimization is much less in terms of wall-clock time to get the solution.

The optimum parameters reported in Table 5.2.2 indicate that the robustness of the $\tau = 0.95$ optimum comes from an SCB with higher and longer shapes, placed slightly more downstream and with lower slope factor T_{bump} than the $\tau = 0.5$ configuration. Figure 24 reports the different optimal SCB shapes. The downstream displacement of the bump enables dealing with a broader range of shock waves, leading to an effective drag reduction for more events, in particular for the most substantial shock waves occurring at large Mach and lift. As a result, the robust optimal SCB reduces the upper tail of the probability distribution

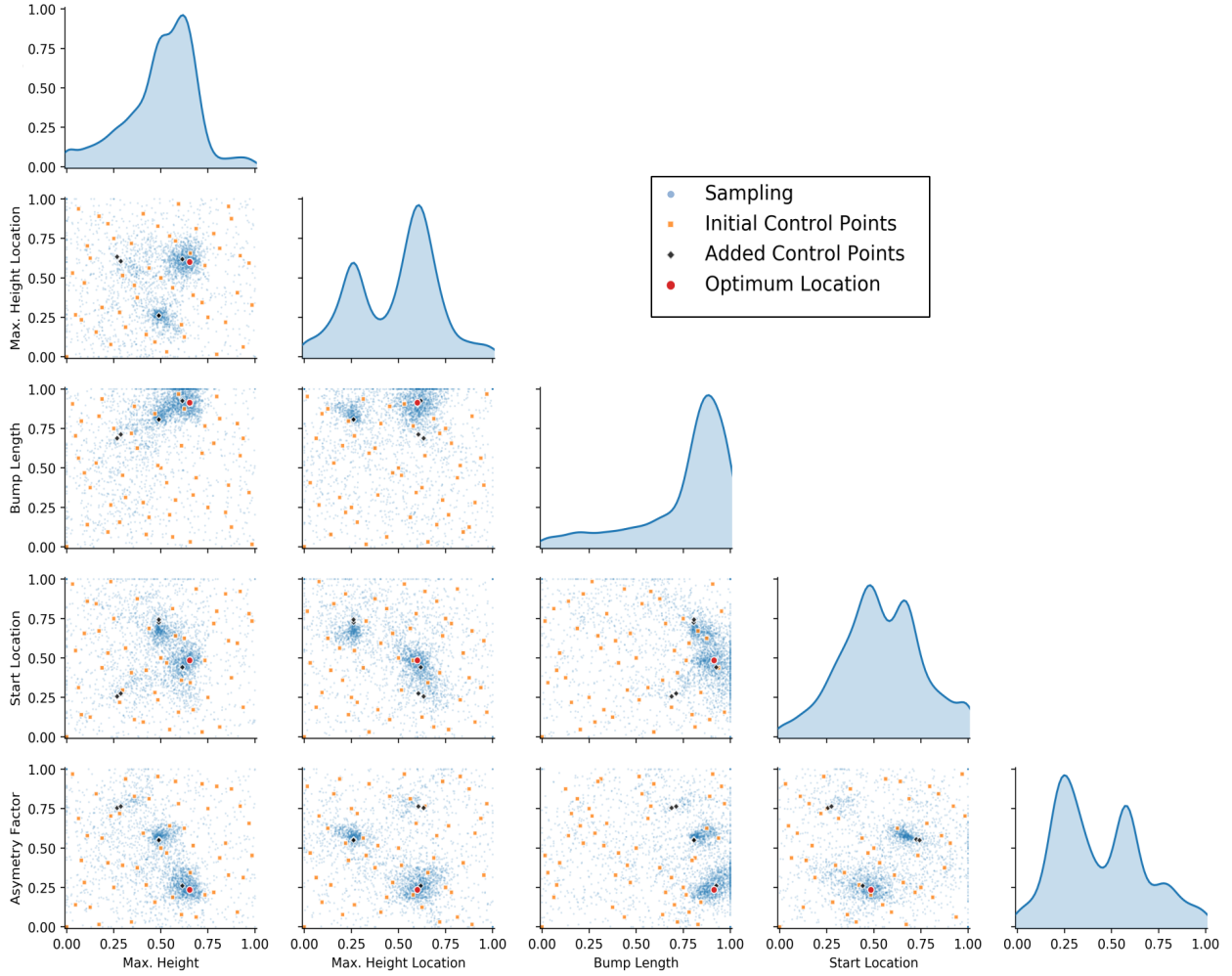


Figure 22: Pair-plot representation of the sample sets in design space at the end of the optimization: sample points in \mathcal{Y} (blue) and \mathcal{X} (initial control points in orange and new points in black). Case of $\tau = 0.95$.

Configuration	Optimal SCB parameters					C_D quantiles		CFD Samples
	H_{bump}	$X_{H_{\text{bump}}}$	L_{bump}	X_{start}	T_{bump}	$\tau = 0.5$	$\tau = 0.95$	
No SCB	—	—	—	—	—	189.7	232.05	1
Deterministic Opt.	0.00772	0.598	0.402	0.398	1.111	148.5	185.92	116
50% Optimum	0.00739	0.676	0.421	0.374	0.789	147.3	184.2	2600
95% Optimum	0.00979	0.670	0.424	0.416	0.623	154.0	172.5	3320

Table 3: Optimal parameters, $q_{0.5}^*$ and $q_{0.95}^*$ of the drag (in drag count units) for the different airfoils. Also provided in the last column are the numbers of simulations to solve the optimization problems.

but is less effective in other situations (lower lift and lower Mach in particular). This mechanism is visible in Figure 25, which shows 100 realizations of the pressure coefficients distribution over the airfoil for the two robust optima. Realizations for the $\tau = 0.95$ optima presents two successive shocks (sharp drops of the pressure coefficient on the extrados) while the $\tau = 0.5$ optimal SCB tends instead to have a single, more pronounced, shock.

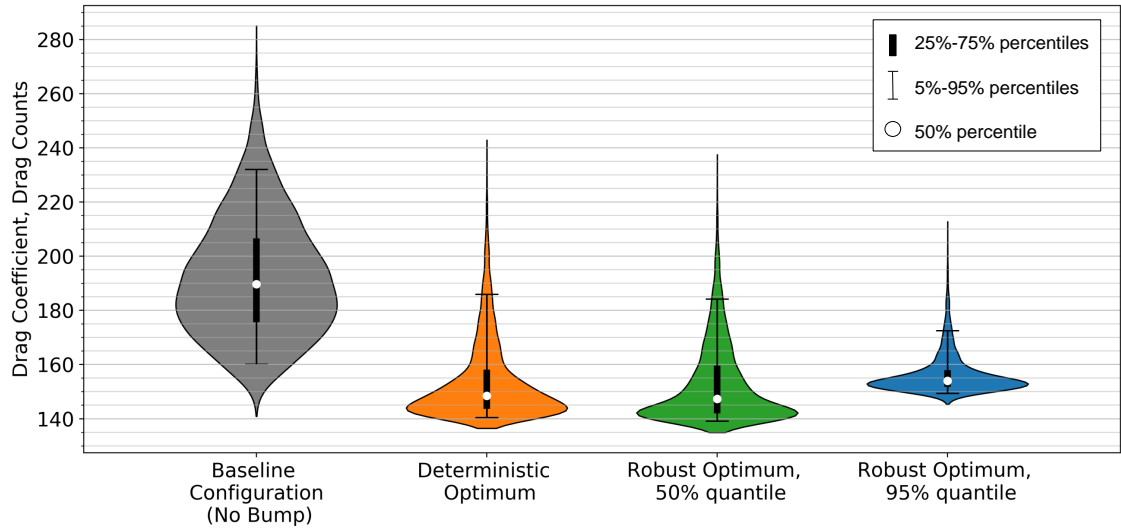


Figure 23: Drag distributions due to geometrical and operational uncertainties and for different airfoils. From left to right: airfoil without SCB, deterministic optimal SCB, $\tau = 0.5$ optimal SCB and $\tau = 0.95$ optimal SCB.

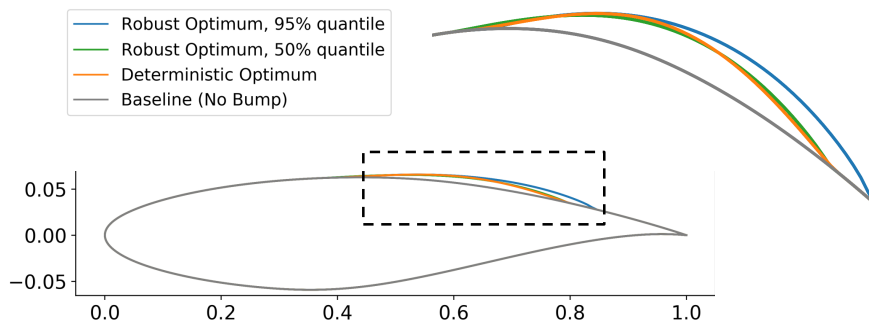


Figure 24: Comparison of the different optimum bump shapes.

6. Conclusions

In this paper, we have proposed a new approach for robust optimization based on quantile minimization. Our approach mainly targets problems that are not amenable to low dimensional parametrizations of the uncertainties, such that a surrogate-based approach is a priori not feasible. The approach proposed combines a Bayesian regression method for the estimation of the quantile over the design space, with a sequential Bayesian procedure for the optimization. We have shown that the Bayesian regression constitutes a flexible approach to characterize the quality of the quantile estimator, beyond the MLE quantile, through the sampling of the quantile's posterior. Further, we have proposed a Cross-Validation strategy to select the number of control points to build the a priori estimator of the quantile. Although the procedure is presented for RBF approximation, the procedure can apply to other types of models for the quantile discretization. Finally, the optimization procedure enriches the samples set and the quantile's model using points drawn from the current distribution of the optimum, which accounts for the posterior distribution of the quantile estimator. These active infilling strategies can come up with the global optimum and adjust the regression quality in areas of interest.

The method was first tested and validated in simple one and two-dimensional multi-modal explicit test functions and on a complex aerodynamic shape optimization problem requiring CFD simulations to compute the quantity of interest (the airfoil's drag). In the latter case, we also considered the minimization of two

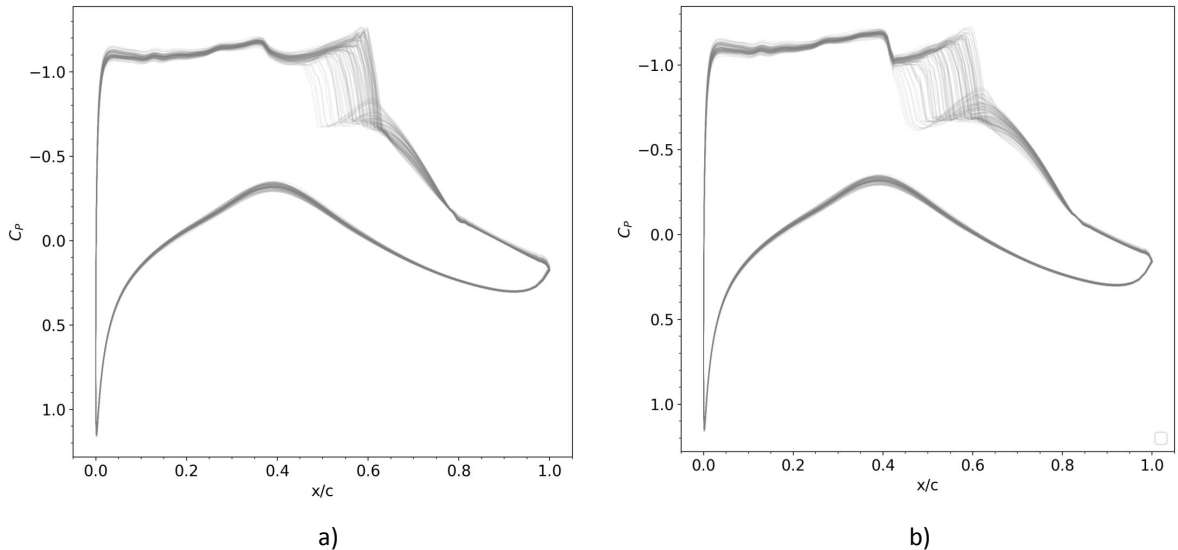


Figure 25: Realizations of the pressure coefficient along the airfoil with SCB optimized for $\tau = 0.5$ (left) and $\tau = 0.95$ (right).

quantiles ($\tau = 0.5$ and $\tau = 0.95$) to illustrate the impact of more challenging problems on the behavior of the method. These examples show that the Bayesian quantile optimization method provides reasonably accurate results for limited computational resources.

Although successful in the examples presented, the method needs more investigations to be fully understood and operated with full confidence. A first aspect to explore is the generalization of the quantile regression to other types of representations, for instance, substituting the RBF approximation with more general procedures. Second, in this work, we have restricted the selection of the control points to the first $N_{\mathcal{X}}$ elements of a fixed sequence. A more effective strategy would be to select each $N_{\mathcal{X}}$ points individually in Ω , possibly from a finite set. Although CV-loss criteria would continue to apply, more advanced algorithms would be needed to select the set of control points. Similarly, we have proposed to use sampled optimal points x^* as candidates for the enrichment of the regression space during the optimization process. Although this heuristic gives good results, one may be interested in enriching the quantile discretization in a better and possibly more effective way in the future. Also, concerning potential improvements of the Bayesian optimization procedure, future works should focus on the role of the shape parameter α of the Asymmetric Laplace Distribution on the distribution of the minima x^* , and the introduction of a strategy for learning its value from the samples set. Similarly, the direct draw of the new sample points from the current distribution of x^* could be substituted by more advanced selection criteria, favoring for instance points with higher potential in reducing the regression error in areas of interest. Indeed, the current infilling procedure only incorporates the probability of being the minimal point of the quantile, without accounting for the impact of the new sample on the estimation of the quantile. Finally, an increase in the sample size with the quantile value is expected. Currently, a complete random sampling is used. For higher quantiles such as the 99% or 99.9%, more advance sampling techniques will be necessary.

Regarding the application to aerodynamic shape optimization, the parallelization of the sampling at each iteration enables the application of this framework to more complex problems. Special interest lies in the application to the robust design of transonic wings under operational and geometrical uncertainties. In that case, the random field over the surface of the wing will lead to thousands of uncertainties.

At a more general level, we are planning to explore the extension of the proposed optimization procedure to other robust optimization problems. These problems naturally include the minimization of the mean, higher-order moments, and conditional value at risk, but also the extension to robust multi-objective optimization.

Acknowledgments

This work is funded by the European Commission’s H2020 programme, through the UTOPIAE Marie Curie Innovative Training Network, H2020-MSCA-ITN-2016, Grant Agreement number 722734.

References

- [1] H.-G. Beyer, B. Sendhoff, Robust optimization – a comprehensive survey, *Computer Methods in Applied Mechanics and Engineering* 196 (33-34) (2007) 3190–3218. doi:10.1016/j.cma.2007.03.003.
- [2] R. A. Bettis, M. A. Hitt, The new competitive landscape, *Strategic Management Journal* 16 (S1) (1995) 7–19. doi:10.1002/smj.4250160915.
URL <https://doi.org/10.1002/smj.4250160915>
- [3] K. S. Swan, M. Kotabe, B. B. Allred, Exploring robust design capabilities, their role in creating global products, and their relationship to firm performance, *Journal of Product Innovation Management* 22 (2) (2005) 144–164. doi:10.1111/j.0737-6782.2005.00111.x.
URL <https://doi.org/10.1111/j.0737-6782.2005.00111.x>
- [4] Z. Huan, G. Zhenghong, X. Fang, Z. Yidian, Review of robust aerodynamic design optimization for air vehicles, *Archives of Computational Methods in Engineering* 26 (3) (2018) 685–732. doi:10.1007/s11831-018-9259-2.
URL <https://doi.org/10.1007/s11831-018-9259-2>
- [5] D. Maruyama, D. Liu, S. Görtz, An efficient aerodynamic shape optimization framework for robust design of airfoils using surrogate models, in: *Proceedings of the VII European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2016)*, NTUA Greece, 2016. doi:10.7712/100016.2450.8838.
- [6] P. Seshadri, P. Constantine, G. Iaccarino, G. Parks, A density-matching approach for optimization under uncertainty, *Computer Methods in Applied Mechanics and Engineering* 305 (2016) 562–578. doi:10.1016/j.cma.2016.03.006.
- [7] M. Moustapha, B. Sudret, J.-M. Bourinet, B. Guillaume, Quantile-based optimization under uncertainties using adaptive kriging surrogate models, *Structural and Multidisciplinary Optimization* 54 (6) (2016) 1403–1421. doi:10.1007/s00158-016-1504-4.
- [8] C. Sabater, S. Görtz, An efficient bi-level surrogate approach for optimizing shock control bumps under uncertainty, in: *AIAA Scitech 2019 Forum*, American Institute of Aeronautics and Astronautics, 2019. doi:10.2514/6.2019-2214.
- [9] N. Razaaly, G. Persico, G. Gori, P. M. Congedo, Quantile-based robust optimization of a supersonic nozzle for organic rankine cycle turbines, *Applied Mathematical Modelling* 82 (2020) 802–824. doi:10.1016/j.apm.2020.01.048.
- [10] G. Schuëller, H. Jensen, Computational methods in optimization considering uncertainties – an overview, *Computer Methods in Applied Mechanics and Engineering* 198 (1) (2008) 2–13. doi:10.1016/j.cma.2008.05.004.
- [11] M. Dodson, G. T. Parks, Robust aerodynamic design optimization using polynomial chaos, *Journal of Aircraft* 46 (2) (2009) 635–646. doi:10.2514/1.39419.
- [12] V. Keshavarzadeh, R. G. Ghanem, D. A. Tortorelli, Shape optimization under uncertainty for rotor blades of horizontal axis wind turbines, *Computer Methods in Applied Mechanics and Engineering* 354 (2019) 271–306. doi:10.1016/j.cma.2019.05.015.
- [13] L. P. Swiler, M. S. Eldred, Efficient algorithms for mixed aleatory-epistemic uncertainty quantification with application to radiation-hardened electronics. part i, algorithms and benchmark results. (9 2009). doi:10.2172/972887.
- [14] V. Picheny, D. Ginsbourger, Y. Richet, G. Caplin, Quantile-based optimization of noisy computer experiments with tunable precision, *Technometrics* 55 (1) (2013) 2–13. arXiv:<https://doi.org/10.1080/00401706.2012.707580>, doi:10.1080/00401706.2012.707580.
URL <https://doi.org/10.1080/00401706.2012.707580>
- [15] J. Zhang, A. Taflanidis, J. Medina, Sequential approximate optimization for design under uncertainty problems utilizing kriging metamodeling in augmented input space, *Computer Methods in Applied Mechanics and Engineering* 315 (2017) 369–395. doi:10.1016/j.cma.2016.10.042.
- [16] C. Hirsch, D. Wunsch, J. Szumbariski, L. Laniewski-Wołk, J. Pons-Prats (Eds.), *Uncertainty Management for Robust Industrial Design in Aeronautics*, Springer International Publishing, 2019. doi:10.1007/978-3-319-77767-2.
- [17] D. Liu, D. Maruyama, S. Görtz, Geometrical uncertainties—accuracy of parametrization and its influence on UQ and RDO results, in: *Uncertainty Management for Robust Industrial Design in Aeronautics*, Springer International Publishing, 2018, pp. 785–798. doi:10.1007/978-3-319-77767-2_51.
- [18] D. Wunsch, R. Nigro, G. Coussement, C. Hirsch, Non-intrusive probabilistic collocation method for operational, geometrical, and manufacturing uncertainties in engineering practice, in: *Uncertainty Management for Robust Industrial Design in Aeronautics*, Springer International Publishing, 2018, pp. 143–167. doi:10.1007/978-3-319-77767-2_9.
- [19] A. Clarich, R. Russo, Innovative methodologies for robust design optimization with large number of uncertainties using ModeFRONTIER, in: *Computational Methods in Applied Sciences*, Springer International Publishing, 2018, pp. 335–351. doi:10.1007/978-3-319-89988-6_20.
- [20] P. G. Constantine, M. S. Eldred, E. T. Phipps, Sparse pseudospectral approximation method, *Computer Methods in Applied Mechanics and Engineering* 229-232 (2012) 1 – 12. doi:<https://doi.org/10.1016/j.cma.2012.03.019>.
URL <http://www.sciencedirect.com/science/article/pii/S0045782512000953>
- [21] D. Maruyama, D. Liu, S. Görtz, Surrogate model-based approaches to UQ and their range of applicability, in: *Uncertainty Management for Robust Industrial Design in Aeronautics*, Springer International Publishing, 2018, pp. 703–714. doi:10.1007/978-3-319-77767-2_43.

- [22] C. Shang, F. You, Robust optimization in high-dimensional data space with support vector clustering, *IFAC-PapersOnLine* 51 (18) (2018) 19–24. doi:10.1016/j.ifacol.2018.09.238.
- [23] C. Lataniotis, S. Marelli, B. Sudret, Extending classical surrogate modelling to high dimensions through supervised dimensionality reduction: a data-driven approach, *International Journal for Uncertainty Quantification* (01 2020). doi:10.1615/Int.J.UncertaintyQuantification.2020031935.
- [24] M. Pisaroni, F. Nobile, P. Leyland, A continuation multi level monte carlo (c-mlmc) method for uncertainty quantification in compressible inviscid aerodynamics, *Computer Methods in Applied Mechanics and Engineering* 326 (08 2017). doi:10.1016/j.cma.2017.07.030.
- [25] M. Pisaroni, F. Nobile, P. Leyland, Continuation multilevel monte carlo evolutionary algorithm for robust aerodynamic shape design, *Journal of Aircraft* 56 (2018) 1–16. doi:10.2514/1.C035054.
- [26] J. C. Medina, A. A. Taflanidis, Adaptive importance sampling for optimization under uncertainty problems, *Computer Methods in Applied Mechanics and Engineering* 279 (2014) 133–162. doi:https://doi.org/10.1016/j.cma.2014.06.025. URL <http://www.sciencedirect.com/science/article/pii/S0045782514002114>
- [27] R. Koenker, G. Bassett, Regression quantiles, *Econometrica* 46 (1) (1978) 33. doi:10.2307/1913643.
- [28] R. Koenker, *Quantile Regression*, CAMBRIDGE UNIV PR, 2010. URL https://www.ebook.de/de/product/3499890/roger_koenker_quantile_regression.html
- [29] F. Mosteller, J. W. Tukey, *Data Analysis and Regression: A Second Course in Statistics*, Pearson, 1977. URL <https://www.amazon.com/Data-Analysis-Regression-Second-Statistics/dp/020104854X?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=020104854X>
- [30] A. W. van der Vaart, *Asymptotic Statistics*, CAMBRIDGE UNIV PR, 2011. URL https://www.ebook.de/de/product/3338604/a_w_van_der_vaart_asymptotic_statistics.html
- [31] R. Koenker, K. F. Hallock, Quantile regression, *Journal of Economic Perspectives* 15 (4) (2001) 143–156. doi:10.1257/jep.15.4.143. URL <https://www.aeaweb.org/articles?id=10.1257/jep.15.4.143>
- [32] B. S. Cade, B. R. Noon, A gentle introduction to quantile regression for ecologists, *Frontiers in Ecology and the Environment* 1 (8) (2003) 412–420. doi:10.1890/1540-9295(2003)001[0412:agitqr]2.0.co;2.
- [33] J. Duchon, Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces, *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique* 10 (R3) (1976) 5–12. URL http://www.numdam.org/item/M2AN_1976__10_3_5_0
- [34] R. Schaback, Improved error bounds for scattered data interpolation by radial basis functions, *Mathematics of Computation* 225 (68) (1999) 201–216.
- [35] R. Storn, K. Price, Differential evolution. a simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization* 11 (4) (1997) 341–359. doi:10.1023/a:1008202821328.
- [36] I. Sobol, On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Computational Mathematics and Mathematical Physics* 7 (4) (1967) 86–112. doi:10.1016/0041-5553(67)90144-9.
- [37] Y. Feng, Y. Chen, X. He, Bayesian quantile regression with approximate likelihood, *Bernoulli* 21 (2) (2015) 832–850. doi:10.3150/13-bej589.
- [38] K. Yu, R. A. Moyeed, Bayesian quantile regression, *Statistics & Probability Letters* 54 (4) (2001) 437–447. doi:10.1016/S0167-7152(01)00124-9.
- [39] D. F. Benoit, D. V. den Poel, bayesQR: A bayesian approach to quantile regression, *Journal of Statistical Software* 76 (7) (2017). doi:10.18637/jss.v076.i07. URL <https://doi.org/10.18637/jss.v076.i07>
- [40] K. Sriram, P. Shi, P. Ghosh, A bayesian quantile regression model for insurance company costs data, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179 (1) (2015) 177–202. doi:10.1111/rssa.12111. URL <https://doi.org/10.1111/rssa.12111>
- [41] K. Sriram, R. Ramamoorthi, P. Ghosh, Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density, *Bayesian Analysis* 8 (2) (2013) 479–504. doi:10.1214/13-ba817. URL <https://doi.org/10.1214/13-ba817>
- [42] C. Andrieu, N. de Freitas, A. Doucet, M. I. Jordan, An introduction to MCMC for machine learning, *Machine Learning* 50 (1) (2003) 5–43. doi:10.1023/A:1020281327116. URL <https://doi.org/10.1023/A:1020281327116>
- [43] W. K. Hastings, Monte carlo sampling methods using markov chains and their applications, *Biometrika* 57 (1) (1970) 97–109. doi:10.1093/biomet/57.1.97.
- [44] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of state calculations by fast computing machines, *The Journal of Chemical Physics* 21 (6) (1953) 1087–1092. doi:10.1063/1.1699114.
- [45] G. O. Roberts, J. S. Rosenthal, Optimal scaling for various Metropolis-Hastings algorithms, *Statistical Science* 16 (4) (2001) 351–367.
- [46] D. R. Jones, M. Schonlau, W. J. Welch, Efficient global optimization of expensive black-box functions, *Journal of Global Optimization* 13 (4) (1998) 455–492. doi:10.1023/a:1008306431147.
- [47] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics* 20 (1987) 53–65. doi:10.1016/0377-0427(87)90125-7.
- [48] P. Ashill, J. Fulker, J. Shires, A novel technique for controlling shock strength of laminar-flow airfoil sections, in: *Proceedings of the 1st European Forum on Laminar Flow Technology*, Hamburg, 1992, pp. 175–183.
- [49] P. J. K. Bruce, S. P. Colliss, Review of research into shock control bumps, *Shock Waves* 25 (5) (2014) 451–471. doi:

- 10.1007/s00193-014-0533-4.
- [50] K. Mazaheri, A. Nejati, The multi-point optimization of shock control bump with constant-lift constraint enhanced with suction and blowing for a supercritical airfoil, *Flow, Turbulence and Combustion* 96 (3) (2015) 639–666. doi:10.1007/s10494-015-9671-8.
- [51] Y. Tian, P. Liu, P. Feng, Shock control bump parametric research on supercritical airfoil, *Science China Technological Sciences* 54 (11) (2011) 2935–2944. doi:10.1007/s11431-011-4582-y.
- [52] D. M. Tartakovsky, D. Xiu, Stochastic analysis of transport in tubes with rough walls, *Journal of Computational Physics* 217 (1) (2006) 248–259. doi:10.1016/j.jcp.2006.02.029.
URL <https://doi.org/10.1016/j.jcp.2006.02.029>
- [53] D. Guignard, F. Nobile, M. Picasso, A posteriori error estimation for the steady navier–stokes equations in random domains, *Computer Methods in Applied Mechanics and Engineering* 313 (2017) 483–511. doi:10.1016/j.cma.2016.10.008.
URL <https://doi.org/10.1016/j.cma.2016.10.008>
- [54] H. Harbrecht, M. Peters, M. Siebenmorgen, Analysis of the domain mapping method for elliptic diffusion problems on random domains, *Numerische Mathematik* 134 (4) (2016) 823–856. doi:10.1007/s00211-016-0791-4.
URL <https://doi.org/10.1007/s00211-016-0791-4>
- [55] P. Cook, M. McDonald, M. Firmin, Aerofoil rae 2822 - pressure distributions, and boundary layer and wake measurements, Tech. Rep. 138, AGARD (1979).
- [56] J. E. T. R. J. Adler, *Random Fields and Geometry*, Springer-Verlag GmbH, 2009.
URL https://www.ebook.de/de/product/12469889/r_j_adler_jonathan_e_taylor_random_fields_and_geometry.html
- [57] A. S. C. Services, Airbus a330 aircraft characteristics - airport and maintenance planning, Tech. rep., Airbus S.A.S. Customer Services (2005).
- [58] D. Liu, S. Görtz, Efficient quantification of aerodynamic uncertainty due to random geometry perturbations, in: *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, Springer International Publishing, 2014, pp. 65–73. doi:10.1007/978-3-319-03158-3_7.
- [59] T. Gerhold, Overview of the hybrid RANS code TAU, in: *MEGAFLOW - Numerical Flow Simulation for Aircraft Design*, Springer Berlin Heidelberg, 2015, pp. 81–92. doi:10.1007/3-540-32382-1_5.
- [60] T. Gerhold, J. Neumann, The parallel mesh deformation of the DLR TAU-code, in: *Notes on Numerical Fluid Mechanics and Multidisciplinary Design (NNFM)*, Springer Berlin Heidelberg, 2006, pp. 162–169. doi:10.1007/978-3-540-74460-3_20.
- [61] M. Meinel, G. Einarsson, The flowsimulator framework for massively parallel cfd applications, *PARA 2010* (jun 2010).