Sequential Design of Computer Experiments for the Assessment of Fetus Exposure to Electromagnetic Fields

Marjorie Jala
Telecom ParisTech, and Orange Labs,
Céline Lévy-Leduc
AgroParisTech,
Éric Moulines
Telecom ParisTech,
Emmanuelle Conil and Joe Wiart
Orange Labs

September 26, 2014

Abstract

In this paper, we describe four sequential sampling strategies for estimating the quantile of \( Y = f(X) \), where \( X \) has a known distribution in \( \mathbb{R}^d \) and \( f \) is a deterministic unknown, expensive-to-evaluate real-valued function. These approaches all consist in modeling \( f \) as a sample of a well-chosen Gaussian process and aim at estimating the quantile by using as few evaluations of \( f \) as possible. The different methodologies are first compared through various numerical experiments. Then, in the framework of the ANR-JST FETUS project, we apply our strategies to a real example corresponding to the exposure of a Japanese pregnant-woman model and her 26-week-old fetus to a plane wave. Finally, we compare our methodologies on a simplified geometric model designed for modeling the fetus exposure to plane waves.

Keywords: Gaussian process, quantile estimation, sequential approach, design of experiments
1 Introduction

Over the past 30 years the use of wireless communication systems has constantly increased worldwide. However, the technologies used by the different wireless communication devices generate electromagnetic fields (EMF) in the radio frequency (RF) range, and the uncertainty about the risk of RF exposure raises a strong public concern over possible health problems. Protection limits have been established by the International Commission on Non-Ionizing Radiation Protection (see Ahlbom et al. (1998)), but questions related to exposure effects remain, in particular in the case of organisms in development such as fetuses. The risks of exposure can be studied through several approaches: long-term epidemiological studies, in vitro and in vivo studies, and numerical dosimetry based methods.

Here, we shall adopt the latter point of view, in order to evaluate the so-called Specific Absorption Rate (SAR) of the fetus. The SAR, expressed in watts per kilogram, is the measure commonly used in the field of RF exposure to quantify the rate at which EMF energy is absorbed by human tissues. In numerical dosimetry, the SAR is computed by virtually exposing human 3D-models (pregnant woman and fetus models in our case) to one source of EMF. Moreover, a single SAR computation is time-consuming for two reasons. On one hand, the preparation of a numerical dosimetry experiment might be very complex since user intervention is required to adequately deform the model when the fetus is moved inside the mother. On the other hand, once the experiment has been properly prepared, the computational load for the corresponding SAR evaluation is high.

It has to be noted that the value of the SAR depends on the geometry of the models (the morphologies and postures of the mother and the fetus), on the dielectric properties of each tissue of the numerical model according to the frequency of the incident wave, and on the type and position of the source emitting the EMF. Hence, we shall model these experimental factors as a random vector with a known distribution and thus we shall model the SAR by \( Y = f(X) \), where \( f \) is a deterministic unknown real-valued function modeling the SAR response for a given model and \( X \) is a vector whose coordinates are the input variables of the model (for example, the azimuth and the elevation of the incident wave, the dielectric properties of the tissues, etc.). It is assumed that this vector of input variables is a random vector of \( \mathbb{R}^d \) with a known distribution.

We are interested in determining the levels of RF exposure for fetuses, and since \( Y \) models the SAR in the fetus, we aim at estimating a quantile of the distribution of \( Y \). Since a single numerical evaluation of \( f \) for a given value of the input variables \( x \) in \( \mathbb{R}^d \) is very computationally expensive (from one to several hours depending on the 3D-model used, with a computer having the following...
configuration: RAM 16 GB, CPU 4 GHz), the use of classical sampling methods (such as the plain Monte Carlo or Latin hypercube techniques) or of quasi Monte-Carlo is not appropriate (see Forrester et al. (2008), Chapter 1). Hence our goal is to provide a strategy to estimate a quantile of the random variable $Y$ by using as few evaluations of $f$ as possible.

In this paper we shall propose and compare several sequential sampling strategies for quantile estimation using a Gaussian Process as a surrogate model for the unknown function $f$. This framework has already been used in several contexts; see for example Forrester et al. (2008), Santner et al. (2003) and the references therein. For instance, Seo et al. (2000) proposed two active learning strategies for estimating $f$ in a regression framework called ALM (Active Learning MacKay) and ALC (Active Learning Cohn) inspired by MacKay (1992) and Cohn (1996), respectively. These strategies have been deployed for sequential design of computer experiments by Gramacy and Lee (2009). In the context of sequential design strategies, the Gaussian process modeling has also been used for function maximization. Such an approach belongs to the set of Bayesian optimization approaches for which an overview is given in Brochu et al. (2010). The principles of these methods are close to those that we develop in this paper, as they are based on the use of a surrogate model to optimize a criterion which selects the next evaluation point of $f$. Following Brochu et al. (2010), we can distinguish two types of popular Bayesian optimization criteria, which are also called acquisition functions. The first one is based on the use of improvement functions, as it is done in expected improvement algorithms, see for example Vazquez and Bect (2010). The other one is based on confidence bounds criteria, see for instance the GP-UCB proposed by Srinivas et al. (2012) and the approach of de Freitas et al. (2012).

Other works also provide sequential approaches to estimate other quantities of interest of the function $f$. Ranjan et al. (2008) propose an improvement function adapted to the problem of contour estimation. In Bect et al. (2012) the Stepwise Uncertainty Reduction (SUR) criterion is proposed for estimating $\mathbb{P}(Y \geq s)$, where $s$ is a given threshold. This methodology has been extended by Arnaud et al. (2010) and Jala et al. (2012) in order to estimate the quantile of the distribution of $Y$. In the sequel, this approach will be referred to as the SUR algorithm. Besides, Oakley (2004) developed a non sequential method for quantile estimation using a Gaussian Process as a surrogate model for the unknown function $f$. Here we propose two sequential strategies based on a confidence bound criterion adapted to quantile estimation. Our strategies, called GPQE and GPQE+ (for Gaussian Process for Quantile Estimation), aim at learning the function $f$ in the areas of the search space which are likely to be relevant for quantile estimation. The GPQE strategy can seem close to
the methodology proposed in Oakley (2004); in that paper, the author describes a method to make inferences about the $\alpha$-quantile along with an approach to select suitable points for this purpose. This approach consists in first choosing design points all over the search space, and then identifying regions of the space that produce large values of $Y$ to choose a second set of points which are added to the design; the criterion used to identify those regions is close to the one we use in GPQE. Then, $f$ is evaluated at those design points. Our strategy is different from this approach since it is a sequential method which aims at adding only one point at a time to the set of observations. Furthermore, our method provides confidence bounds on the quantile estimation at each iteration, which is a helpful tool to determine if we need to perform more evaluations of $f$ or if the process can be stopped before the allocated budget of evaluations is reached. On the other hand, the GPQE+ strategy is an improvement of the GPQE methodology, inspired by the approach of de Freitas et al. (2012) for maximization of functions.

The paper is organized as follows. In Section 2 we describe four different sequential approaches using the Gaussian process framework for estimating a quantile of the distribution of $Y$: the two novel approaches proposed in this paper (GPQE and GPQE+), the SUR algorithm adapted to quantile estimation and an adaptation of the method proposed by Ranjan et al. (2008) for quantile estimation (called RBM in the sequel). In Section 3 we compare these four different methodologies using several synthetic real-valued functions in dimension $d = 1, 2$ and $3$. Finally, in Section 4 we apply our methodologies to real data coming from our numerical dosimetry application when $d = 2$ and 4.

2 Description of the methods

In this section, we shall describe four sequential sampling strategies for estimating the $\alpha$-quantile $q_\alpha$ of the distribution of $Y = f(x)$, for a given $\alpha$ in $(0, 1)$,

$$q_\alpha = \inf \{q : \mathbb{P}(f(X) \leq q) > \alpha\} .$$

(1)

where $f$ is a deterministic unknown real-valued function and $X$ is a random vector having a known distribution on a compact subset $A \subseteq \mathbb{R}^d$. We shall adopt a Bayesian point of view which consists in considering $f$ as a sample of a zero-mean Gaussian process (GP) having a covariance function $k$ that we shall denote in the sequel as GP$(0, k(\cdot, \cdot))$. The advantage of this approach is that, conditionally on a set of observations $y_t = (y_1, \ldots, y_T)^\top$, where $y_i = f(x_i), 1 \leq i \leq T$, the posterior distribution
is still a GP whose mean $\mu_t(u)$ and covariance $k_t(u, v)$ are given by

\begin{align*}
\mu_t(u) &= k_t(u)K_t^{-1}y_t, \\
k_t(u, v) &= k(u, v) - k_t(u)K_t^{-1}k_t(v),
\end{align*}

where $k_t(u) = [k(x_1, u) \ldots k(x_t, u)]'$. Here $'$ denotes the matrix transposition, $u$ and $v$ are in $\mathcal{A}$ and $K_t = [k(x_i, x_j)]_{1 \leq i, j \leq t}$, where the $x_i$'s are in $\mathcal{A}$, see Rasmussen and Williams (2006) and Srinivas et al. (2012) for further details.

In our case, $f$ models a physical quantity that is supposed to be smooth, so for our applications we shall consider three covariance functions that are commonly used in this case. The first one is the squared exponential (SE) covariance function

\begin{equation}
k_{SE}(u, v) = \exp \left( -\frac{1}{2}(u - v)'M^{-1}(u - v) \right), u, v \in \mathcal{A} \subset \mathbb{R}^d,
\end{equation}

\begin{equation}
M = \text{diag} \left( \ell_1^2, \ldots, \ell_d^2 \right), \ell_1, \ell_2, \ldots, \ell_d > 0.
\end{equation}

Here the $\ell_1, \ell_2, \ldots, \ell_d$ hyperparameters are the characteristic length scales (see Section 5.1 of Rasmussen and Williams (2006)); the definition (4) allows us to model anisotropic response surfaces. As explained in Rasmussen and Williams (2006), since this covariance function is infinitely differentiable, the GP with this covariance function has mean square derivatives of all orders. As argued by Stein (1999) such strong smoothness assumptions may be unrealistic for modeling many physical processes, so we shall also consider two other covariance functions belonging to the Matérn class of covariance functions defined by

\begin{equation}
k_{\text{Matérn}}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \sqrt{2\nu r} \right)^\nu K_\nu \left( \sqrt{2\nu r} \right), \nu > 0,
\end{equation}

where $K_\nu$ is a modified Bessel function with Bessel order $\nu$, see Abramowitz and Stegun (1965, Section 9.6), and $r$ is defined by

\begin{equation}
r = \sqrt{(u - v)'M^{-1}(u - v)}, u, v \in \mathcal{A},
\end{equation}

with $M$ defined in (5). In this situation, as explained in Rasmussen and Williams (2006), the GP is $k$-times mean-square differentiable if and only if $\nu > k$. Here, as recommended by Rasmussen and Williams (2006), we shall focus on the cases where $\nu = 3/2$ or $5/2$, for which $k_{\text{Matérn}}$ has a computationally advantageous expression. Indeed, for $\nu = p + \frac{1}{2}$, where $p$ is in $\mathbb{N}$,

\begin{equation}
k_{\text{Matérn}}(r) = \exp \left( -\sqrt{2\nu r} \right) \frac{\Gamma(p + 1)}{\Gamma(2p + 1)} \sum_{i=0}^{p} \frac{(p + i)!}{i!(p - i)!} \left( \sqrt{8\nu r} \right)^{p-i},
\end{equation}

5
with \( r \) defined in (7); see Abramovitz and Stegun (1965, Equation 10.2.15) for further details. We shall compare the quantile estimators with \( \tilde{q}_{\alpha,m} \) defined by

\[
\tilde{q}_{\alpha,m} = \inf \left\{ q : \frac{1}{m} \sum_{i=1}^{m} 1\{f(x_i) \leq q\} > \alpha \right\},
\]

where \( x_1, \ldots, x_m \) are \( m \) fixed points in \( \mathcal{A} \). They can be chosen either as \( m \) independent realizations of the random vector \( X \) or by using a quasi-Monte Carlo approach, see Lemieux (2009) and Joseph et al. (2014). In the sequel, we shall denote by \( A \) the following set

\[
A = \{ x_1, \ldots, x_m \} \subset \mathcal{A}.
\]

In the following, for notational simplicity, we shall denote \( \tilde{q}_{\alpha,m} \) by \( \tilde{q}_{\alpha} \).

Given this definition of \( \tilde{q}_{\alpha} \), a first approach would be to consider that an accurate estimation of \( f \) leads to an accurate estimation of the quantile. Hence, an appropriate sequential strategy could rely on a “pure exploration criterion” (see Sacks et al. (1989), Forrester et al. (2008, Section 3) and references therein) which minimizes the global uncertainty on the estimation of the function \( f \). With this strategy the new point \( x_{t+1} \) to add to the set of \( t \) observations, at which \( f \) is evaluated, is selected in a fine grid \( A \) of \( \mathcal{A} \) as follows:

\[
x_{t+1} \in \arg \max_{x \in A} \sigma_t(x),
\]

where \( \sigma_t(x)^2 = k_t(x,x) \) defined in (3) and \( \arg \max_{x \in A} \sigma_t(x) \) is the set of \( x \in A \) where \( \sigma_t(x) \) reaches its maximum. However this strategy is not dedicated to quantile estimation so it is likely to demand a lot of evaluations of the function \( f \). Consequently, we shall propose methodologies more adapted to our quantile estimation issue.

### 2.1 The GPQE strategy

Since we are interested in the estimation of a quantile of the distribution of \( Y = f(X) \), our goal is to sample \( f \) in the areas that are most likely to be relevant for quantile estimation. Thus, instead of trying to reduce the global uncertainty on \( f \) by using the criterion (11) on all the points of \( A \), our strategy reduces, step by step, the uncertainty on \( f \) only in relevant regions of the input space for quantile estimation. Inspired by the GP-UCB algorithm proposed by Srinivas et al. (2012) for estimating the maximum of \( f \), we shall use the same definition for the Upper Confidence Bound (UCB) and its Lower Confidence Bound (LCB) counterpart for a set of \( t \) function evaluations that
we denote by $\mu_t^{U}$ and $\mu_t^{L}$, respectively, for defining a confidence region for the function $f$. Denote by $\{x_1, \ldots, x_t\}$ the values at which $f$ has been evaluated in the first $t$ rounds; let

$$
\mu_t^{U}(x) = \mu_t(x) + \sqrt{\beta_t} \sigma_t(x) \quad \text{and} \quad \mu_t^{L}(x) = \mu_t(x) - \sqrt{\beta_t} \sigma_t(x),
$$

(12)

with

$$
\beta_t = 2 \ln \left( \frac{n t^2}{6} \right) + 2 \ln \left( \frac{m}{\delta} \right),
$$

(13)

where $m$ is the cardinality of $A$ defined in (10) and $\delta$ is linked to the confidence level of the confidence interval of the quantile $\tilde{q}_\alpha$, defined in (9), as it is further explained in Proposition 1. This choice for $\beta_t$ comes actually from the proof of Proposition 1.

Given (12) and (13), we can estimate the $\alpha$-quantile of $\mu_t^{U}$ and $\mu_t^{L}$ by $\hat{q}_{\alpha,t}^{U}$ and $\hat{q}_{\alpha,t}^{L}$ by

$$
\hat{q}_{\alpha,t}^{U} = \inf \left\{ q : \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{[\mu_t^{U}(x_i) \leq q]} > \alpha \right\} \quad \text{and} \quad \hat{q}_{\alpha,t}^{L} = \inf \left\{ q : \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{[\mu_t^{L}(x_i) \leq q]} > \alpha \right\},
$$

(14)

where $x_1, \ldots, x_m$ are defined in (9). With probability greater than $(1 - \delta)$, where $\delta$ is in $(0,1)$, for any integer $t$ and any pairwise distinct $\{x_1, \ldots, x_t\} \subset A$, $\tilde{q}_\alpha$ belongs to the interval $[\hat{q}_{\alpha,t}^{L}, \hat{q}_{\alpha,t}^{U}]$. More precisely, we have the following result which is proved in the supplementary material [Jala et al. (2014)]

**Proposition 1.** Let $f$ be a sample from $GP(0, k(\cdot, \cdot))$ and $\tilde{q}_\alpha$ be defined in (9), then for all $\delta$ in $(0,1)$, $\tilde{q}_\alpha \in [\hat{q}_{\alpha,t}^{L}, \hat{q}_{\alpha,t}^{U}]$ for all $t \geq 1$, with probability greater than $(1 - \delta)$, where $\hat{q}_{\alpha,t}^{L}$ and $\hat{q}_{\alpha,t}^{U}$ are defined in (14).

Let us now denote by $U_{\alpha,t}$ and $L_{\alpha,t}$ the following sets

$$
U_{\alpha,t} = \{ x \in A : \mu_t^{U}(x) \geq \hat{q}_{\alpha,t}^{U} \} \quad \text{and} \quad L_{\alpha,t} = \{ x \in A : \mu_t^{L}(x) \leq \hat{q}_{\alpha,t}^{L} \}, \quad t \geq 1.
$$

(15)

The quantile estimator $\tilde{q}_{\alpha,t}$ computed from $t$ evaluations of $f$ is defined by

$$
\tilde{q}_{\alpha,t} = \inf \left\{ q : \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{[\mu_t(x_i) \leq q]} > \alpha \right\}.
$$

(16)

With probability greater than $(1 - \delta)$, where $\delta$ is in $(0,1)$, for any integer $t$ and any pairwise distinct $\{x_1, \ldots, x_t\} \subset A$, $\tilde{q}_\alpha$ belongs to the interval $[\tilde{q}_{\alpha,t} - \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x); \tilde{q}_{\alpha,t} + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x)]$. More precisely, we have the following result which is proved in the supplementary material [Jala et al. (2014)].
Proposition 2. Let \( \tilde{q}_{\alpha,t} \) be defined in (16) and \( \hat{q}_\alpha \) be defined in (9) then, under the assumptions of Proposition 1, with probability greater than \( \frac{1}{D_4} \frac{1}{A_1} \frac{1}{D_5} \), for all \( t \geq 1 \),

\[
|\hat{q}_{\alpha,t} - \tilde{q}_{\alpha,t}| \leq \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x) ,
\]

where \( \beta_t \) is defined in (12) and \( U_{\alpha,t} \) is defined in (15).

Based on this result, our strategy consists in adding a new point to the observation set by using (11) where the maximization of \( \sigma_t \) is only performed in \( U_{\alpha,t} \). Indeed, the principle of our strategy is to maximize the so-called pure exploration criterion but restricted to a region where the quantile should lie with high probability.

Moreover, with probability greater than \( (1 - \delta) \), where \( \delta \) is in \((0,1)\), for any integer \( t \) and any pairwise distinct \( \{x_1, \ldots, x_t\} \subset A \), and for any \( x \) in \( U_{\alpha,t} \cap L_{\alpha,t} \), \( f(x) \) belongs to the interval \( [\tilde{q}_\alpha - 4\sqrt{\beta_t} \sigma_t(x); \tilde{q}_\alpha + 4\sqrt{\beta_t} \sigma_t(x)] \). More precisely, we have the following result which is proved in the supplementary material Jala et al. (2014).

Proposition 3. Under the assumptions of Proposition 1 with probability greater than \( (1 - \delta) \), we have for all \( t \geq 1 \) and for all \( x \in U_{\alpha,t} \cap L_{\alpha,t} \) that

\[
\tilde{q}_\alpha - 4\sqrt{\beta_t} \sigma_t(x) \leq f(x) \leq \tilde{q}_\alpha + 4\sqrt{\beta_t} \sigma_t(x) ,
\]

where \( U_{\alpha,t} \) and \( L_{\alpha,t} \) are defined in (15).

This result implies that the values of \( f \) at points located in the set \( U_{\alpha,t} \cap L_{\alpha,t} \) are close to the value of the quantile.

Given these definitions, let us now summarize in Algorithm 1 our strategy that we shall call GPQE (for Gaussian Process for Quantile Estimation). We start by evaluating \( f \) at a small number \( t_1 \) of points randomly chosen among the points of the fine grid \( A \) of \( A \). Then, after \( t \) evaluations of \( f \), the new point \( x_{t+1} \) to add to the set of \( t \) observations is such that:

\[
x_{t+1} \in \arg \max_{x \in U_{\alpha,t}} \sigma_t(x) ,
\]

where \( \sigma_t(x)^2 = k_t(x, x) \) defined in (3). The process is iterated until the pre-determined allocated maximum number \( T \) of evaluation of \( f \) is reached.

2.2 GPQE+

Let us now describe our GPQE+ strategy. As in GPQE, we use an underlying grid \( A \) of cardinality \( m \) defined as in (10) for computing \( \hat{q}_{\alpha,t}, \hat{q}^U_{\alpha,t} \) and \( \hat{q}^L_{\alpha,t} \) defined in (16) and (14). The main difference...
Algorithm 1 GPQE

**Input:** $x_1, \ldots, x_t$ a small number of points of $A$ where $f$ has been evaluated

**For** $t = t_1, t_1 + 1, \ldots, T$

- Evaluate the posterior distribution of the GP using (2) and (3), and the variance $\sigma_t(x)^2$
- Evaluate $\hat{q}_{\alpha,t}$ using (16), $\hat{\mu}_t^U$ and $\hat{q}_{\alpha,t}^U$, $\hat{q}_{\alpha,t}^L$, $\hat{\mu}_t^L$ using (12), (14), and eventually $U_{\alpha,t}$ using (15)
- Choose $x_{t+1}$ using (19)
- Evaluate $f$ at this point: $y_{t+1} = f(x_{t+1})$
- Add this new observation to the set; $t = t + 1$

Evaluate the posterior distribution of the GP and $\hat{q}_{\alpha,t}$ using (16).

with GPQE is that the possible evaluation points of $f$ may be chosen outside of $A$. Indeed, as it is shown in Section 3, evaluating $f$ outside the grid $A$ may improve the estimation of $f$ and hence the quantile estimation.

Before choosing $x_{t+1}$ a new search space is defined as follows: denoting by $x_1, \ldots, x_t$ the values at which $f$ has been evaluated in the first $t$ rounds, let $\mathcal{S}_{\alpha,t} \subseteq A$ be the compact subset such that

$$\mathcal{S}_{\alpha,t} = \prod_{i=1}^d [x_{\text{min},t}^{(i)}, x_{\text{max},t}^{(i)}].$$

(20)

Here $x_{\text{min},t}^{(i)}$ denotes the smallest $i$th component of the points in $\bar{U}_{\alpha,t}$ and $x_{\text{max},t}^{(i)}$ the largest, with $\bar{U}_{\alpha,t}$ defined by

$$\bar{U}_{\alpha,t} = \{x \in \mathcal{S}_{\alpha,t-1}^\text{A} : \mu_t^U(x) \geq \hat{q}_{\alpha,t}^L \},$$

where $\mathcal{S}_{\alpha,t-1}$ is a search space at iteration $t - 1$. This search space is iteratively defined as follows:

$$\mathcal{S}_{\alpha,t} = \{x_{t,1}, \ldots, x_{t,m_t}\} \cup \bar{U}_{\alpha,t},$$

where $\{x_{t,1}, \ldots, x_{t,m_t}\}$ are $m_t$ points randomly chosen in $\mathcal{S}_{\alpha,t}$. By convention $\mathcal{S}_{\alpha,0} = A$. The new design point is then chosen as:

$$x_{t+1} \in \arg \max_{x \in U_{\alpha,t}} \sigma_t(x),$$
where $\sigma_t(x)^2 = k_t(x, x)$ is defined in (3). This process is iterated until the allowed number $T$ of function evaluations is reached. Since the size of the grid varies at each iteration of the process, $\beta_t$ has to be adapted and is now defined by:

$$\beta_t = 2 \ln \left( \frac{\pi^2 t^2}{6} \right) + 2 \ln \left( \frac{|S_{\alpha,t-1}|}{\delta} \right).$$

### 2.3 SUR strategy

In order to compare our approaches GPQE and GPQE+ to another sequential strategy using a Gaussian process prior, we describe hereafter an adaptation of the Stepwise Uncertainty Reduction criterion proposed in Bect et al. (2012) in the case of the estimation of $P(Y \geq s)$ where $s$ is a given threshold, to quantile estimation. The SUR criterion is close to the improvement-based acquisition functions for Bayesian Optimization described in Section 2.3.1 of Brochu et al. (2010). To select a new point where the function $f$ has to be evaluated, the idea is to use a myopic criterion which minimizes the deviation from the true parameter of interest $\theta(f)$ that we wish to estimate; so the new trial point should be the one which, on average, lowers most the variance of $\theta(f)$ conditionally to the $t$ previous observations. Thus, the $(t + 1)$th point to add to the set of observations is:

$$x_{t+1} = \arg\min_{x \in A} V_t(x),$$

where $A$ is a fine grid of $A$ and

$$V_t(x) = \int \text{Var}(\theta | F_t^{(x,y)}) \varphi_{\mu_t(x), \sigma_t^2(x)}(y) dy,$$

with $F_t^{(x,y)} = \{(x_1, f(x_1)), \ldots, (x_t, f(x_t)), (x, y)\}$, $\sigma_t^2(x) = k_t(x, x)$ and $\varphi_{\mu_t(x), \sigma_t^2(x)}$ is the p.d.f. of a Gaussian random variable with mean $\mu_t(x)$ and variance $\sigma_t^2(x)$. The steps of the SUR strategy are almost the same as those of Algorithm 1 except that the process of shrinking the search space and maximizing $\sigma_t^2$ on this search space is replaced by the evaluation of the criterion (21) and its minimization. The criterion (21) has a closed-form expression only in particular cases; for instance, when $\theta(f) = \mathbb{P}(Y \geq s)$. In our case where $\theta(f)$ is $q_\alpha$, the computation of (21) is more involved, and we need to perform Monte Carlo simulations to produce an approximation of (21). More precisely, after having evaluated $f$ at $t$ points, $N$ sample paths of the Gaussian process $GP(\mu_t, k_t)$ have to be simulated. For each point $x \in A$ at which $f$ has not been evaluated yet, the $N$ posterior distributions of the GP conditionally to the addition of $x$ and its value at each of those paths have to be computed. Then $N$ paths of each of those conditional GP have to be simulated, in order to estimate the empirical variance of the quantile $q_\alpha$; the mean of those $N$ empirical variances is the
value of the criterion (21) for the point \( x \in A \) at which \( f \) has not been evaluated yet. Further details about this process can be found in Jala et al. (2012).

### 2.4 RBM strategy

Another possible strategy to which we can compare our methodologies GPQE and GPQE+ is an adaptation of the method proposed by Ranjan, Bingham and Michailidis (hence the name RBM) in Ranjan et al. (2008) for contour estimation. A contour is defined by:

\[
S(a) = \{ x \in A : f(x) = a \},
\]

where \( a \in \mathbb{R} \) is a given value and \( A \) is defined in (10). Their strategy consists in estimating sequentially \( S(a) \) by selecting at each step a new design point maximizing the expected improvement criterion \((EI)_{a, \beta, t}(x)\) defined by:

\[
(EI)_{a, \beta, t}(x) = \left[ (\beta^2 - 1)\sigma_t^2(x) - (\mu_t(x) - a)^2 \right] \times \left[ \Phi \left( \frac{a - \mu_t(x)}{\sigma_t(x)} + \beta \right) - \Phi \left( \frac{a - \mu_t(x)}{\sigma_t(x)} - \beta \right) \right] \\
+ 2(\mu_t(x) - a)\sigma_t(x) \times \left[ \phi \left( \frac{a - \mu_t(x)}{\sigma_t(x)} + \beta \right) - \phi \left( \frac{a - \mu_t(x)}{\sigma_t(x)} - \beta \right) \right] \\
+ \sigma_t^2(x) \left[ \frac{a - \mu_t(x)}{\sigma_t(x)} + \beta \right] \phi \left( \frac{a - \mu_t(x)}{\sigma_t(x)} - \beta \right) - \\
\left( \frac{a - \mu_t(x)}{\sigma_t(x)} - \beta \right) \phi \left( \frac{a - \mu_t(x)}{\sigma_t(x)} - \beta \right), \tag{22}
\]

where \( \phi \) and \( \Phi \) are the p.d.f and the c.d.f of a standard Gaussian random variable, respectively. The expression of the expected improvement given in (22) corresponds to an erratum for Ranjan et al. (2008) which can found on Ranjan’s website.

We propose to adapt this strategy to quantile estimation as follows. After \( t \) evaluations of the function \( f \), the new design point \( x_{t+1} \) is such that:

\[
x_{t+1} \in \arg \max_{x \in A} (EI)_{\hat{q}_{\alpha, t}, \beta, t}(x), \tag{23}
\]

where \((EI)_{\hat{q}_{\alpha, t}, \beta, t}(x)\) is defined in (22) and \( \hat{q}_{\alpha, t} \) is defined in (16).

### 2.5 Estimating the characteristic length scale

In the previous parts, we assumed that the characteristic length scales \( \ell = (\ell_i)_{1 \leq i \leq d} \) were known. However, this is obviously not the case in real-data applications. We propose to use the maximum-likelihood strategy described in Rasmussen and Williams (2006) to estimate \( \ell \). This adds a step to

\[\text{http://www.acadiau.ca/~pranjan/}\]
the algorithms previously described, as the $\ell_i$’s have to be estimated before evaluating the posterior distribution of the GP using (2) and (3). Hence, for the observation set $\{(x_1, y_1), \ldots, (x_t, y_t)\}$ with $y_i = f(x_i), 1 \leq i \leq t$, the posterior log-likelihood given by:

$$-\frac{1}{2}y_t^T K_t^{-1} y_t - \frac{1}{2} \log |K_t| - \frac{t}{2} \log 2\pi,$$

with $y_t = (y_1, \ldots, y_t)’$ and $K_t = [k(x_i, x_j)]_{1 \leq i, j \leq t}$, has to be maximized with respect to $\ell$. In our following applications, we use the `fminsearch` function of Matlab to minimize the opposite of the posterior log-likelihood.

3 Numerical experiments

In this section, we shall apply the strategies introduced in Section 2 to the estimation of the 95% quantile of the distribution of $Y = f(X)$, where $X$ has a known distribution in $\mathcal{A} = [0, 1]^d$ for $d = 1, 2, 3$, and $f$ is either a sample path of a zero-mean GP with different covariance functions or a classical test function. We shall illustrate the different steps of our strategies GPQE and GPQE+ and investigate the properties of each methodology in those cases. In the sequel, we shall denote by $\tilde{Q}_{0.95}$ the estimator of the 95% quantile given by:

$$\tilde{Q}_{0.95} = (f(\mathcal{X}_M))_{(\lfloor M \times 0.95 \rfloor)};$$

where $f(\mathcal{X}_M) = \{f(x_1), \ldots, f(x_M)\}, x_1, \ldots, x_M$ being $M$ realizations of the random vector $X$. In (24), $(f(\mathcal{X}_M))_{(i)}$ denotes the $i$th smallest value of the set $f(\mathcal{X}_M)$ and $[z]$ denotes the nearest integer greater than or equal to the real number $z$. Following (16), the quantile estimator computed from $\mu_t$ is defined by:

$$\hat{q}_{0.95,t} = (\mu_t(\mathcal{X}_m))_{(\lfloor m \times 0.95 \rfloor)};$$

where $\mu_t(\mathcal{X}_m) = \{\mu_t(x_1), \ldots, \mu_t(x_m)\}$ and $x_1, \ldots, x_m$ are defined in (10). The values of $M$ and $m$ will be chosen in the following sections such that $M$ is much larger than $m$. The quantity $\tilde{Q}_{0.95}$ is helpful to evaluate the performance of our methodologies by computing the relative errors defined for a set of $t$ observations by $|\tilde{Q}_{0.95} - \hat{q}_{0.95,t}|/\tilde{Q}_{0.95}$. With such a choice, the performance of our strategies is not compared with the value of $\hat{q}_{0.95}$ defined in (9) which is the quantile computed on the grid $\{x_1, \ldots, x_m\}$ but with $\tilde{Q}_{0.95}$, which is a more difficult task.
### 3.1 One and two dimensional examples

In order to illustrate the way our strategies work, we shall apply them to the function $f(x) = \exp(-1.4x) \cos(7\pi x/2), \ x \in [0, 1]$ described in Example 3.3 of [Santner et al. (2003)](#), denoted $f_1$ and displayed in Figure 1(a). We shall also apply our strategies to functions sampled from a zero-mean GP with two different covariance functions. We use a SE covariance function defined in (4) for $d = 2$, with $\ell_1 = \ell_2 = 0.1$; this function, denoted $f_2$, is displayed in Figures 1(b) and (d). More precisely, Figure 1(b) displays the 3D graph and Figure 1(d) displays the contour graph for $d = 2$. We also propose to apply our methodology using the Matérn covariance function defined in (6) with parameters $\nu = 1/2$ and $\ell_1 = 0.1$; the corresponding function, denoted $f_3$, is displayed in Figure 1(c).

![Figure 1](#)

**Figure 1:** (a): $f_1$, (b): $f_2$, (c): $f_3$, (d): contour graph for $f_2$.

**GPQE step by step** To illustrate how the GPQE strategy works, we show in Figure 2 the different steps of this methodology when it is applied to the function $f_1$ displayed in Figure 1(a) and $X$ has either a uniform distribution on $[0, 1]$ (left column of Figure 2) or a Beta(2,2) distribution (right
column of Figure 2. We start with a set of $t_1 = 2$ points randomly chosen in a grid $A$ of cardinality $100$ ($m = 100$). The points of $A$ are chosen as follows: a Latin Hypercube design of cardinality $100$ is first generated, then we apply to each of these points the inverse of the c.d.f of $X$ in order to generate points having the same distribution as $X$.

The points which have been added to the observation set are easy to discern because they are located in the regions where $\mu_U^t$ and $\mu_L^t$ coincide, $\mu_U^t$ and $\mu_L^t$ being defined in (12), respectively. Thus, in Figures 2(a) and (b), the process is at its beginning: no points have been added yet, the function $f$ has only been evaluated at the two starting points. Then, in Figures 2(c) and (d), four iterations later, four points have been added. In Figure 2(c), there are clearly no more points in the search space around 0.2. In this region, we can see that the bold dark line representing $\mu_U^t$ is under the light gray line representing the 95%-quantile of $\mu_L^t$, so, as specified by our criterion, this region is discarded from the search space. In Figure 2(e) and (f), eight points have been added to the observation set, and the search space is reduced to four points for the uniform distribution and to two points for the Beta distribution; here we can see that the upper quantile is almost equal to $\tilde{q}_{0.95}$ defined in (9), and that the lower quantile is very close too, both for the uniform and Beta distributions.

**GPQE+ step by step** Here we illustrate the GPQE+ strategy on the function $f_2$ displayed in Figure 1(b) and (d), where the lightest zones correspond to the highest values of $f_2$. Four steps of this approach are displayed in the contour plots of Figure 3. In this example, $X$ is assumed to be uniformly distributed on $[0, 1]^2$. In Figure 3(a), $f_2$ is evaluated at two points (dark gray dots) randomly chosen in a grid $A$ of cardinality 200 ($m = 200$) displayed with black dots. This grid has been generated by using Latin hypercube sampling. Here, $m_t$ is chosen as follows: $m_t = |\bar{U}_{\alpha,t}|$. In this example, the algorithm needs more exploration time than in dimension 1. Indeed, after 60 iterations of the procedure (hence after adding 60 points to the set of observations), we obtain Figure 3(b). In this figure, the dark gray dots display the points at which $f_2$ has been evaluated and the black dots correspond to the set $\bar{U}_{\alpha,t}$. After 100 iterations (Figure 3(c)) and 160 iterations (Figure 3(d)) we can see that the regions in which the next evaluation point of $f_2$ is evaluated is chosen, are localized in the zones where $f_2$ is greater than the quantile. For this two dimensional example, we evaluate the quantile $\bar{Q}_{0.95}$ with $M = 10^5$ points.

**Monte-Carlo replications** The comparison between GPQE, GPQE+, SUR and RBM for the estimation of the 95% quantile is performed through Monte Carlo replications on the choices of the
grid A and the initial set of observations for the functions $f_2$ and $f_3$ described at the beginning of Section 3.1. Here, $X$ is assumed to be uniformly distributed on $[0, 1]^2$ for $f_2$ and on $[0, 1]$ for $f_3$. The grid A has been generated thanks to a Latin Hypercube design. The results are displayed in Figure 4 through the mean of the relative errors computed at each addition of a new design point in the observation set. Note that, since no recommendation for the choice of the constant $\beta$ (see (22)) is available, we tried several values of this parameter and only displayed the best results.

For the function $f_2$ displayed in Figures 1 (b) and (d), 60 Monte Carlo replications are per-
formed on the grid $A$ and hence on the initial set of observations, which is randomly picked from $A$ and of cardinality 2. The results are displayed in Figure 4 (left). The SUR strategy with $N = 50$ performs better than GPQE, GPQE+, RBM and SUR with $N = 10$, which give very close results.

We can see from Figure 5 that these results are very general and not only valid for the $f_2$ function. Indeed, this figure displays the means of the relative errors for the 95% quantile estimation obtained from 100 Monte-Carlo replications on the grid $A$ as explained for $f_2$ but for 100 different sample paths of Gaussian processes having $k_{SE}$ as a covariance function and $\ell_1 = \ell_2 = 0.1$.

The behavior of the different strategies are quite different for the function $f_3$ displayed in Figure 4 (c). In this case, 100 Monte Carlo replications are performed on the grid $A$ and hence on the initial set of observations, which is randomly picked from $A$ and of cardinality 2. We can see from Figure 4 (right) that the RBM strategy performs poorly for all the $\beta$ that we tried: after a few iterations of the algorithm, it seems to converge to a value which is not the 95% quantile we are looking for. The performances of GPQE, GPQE+ and SUR (with $N = 10$) are equivalent until 100

Figure 3: GPQE+ step by step: (a) $t = 2$, (b) $t = 62$, (c) $t = 102$, (d) $t = 162$; contour lines of the function $f_2$, observations (‘*’), and points of the densified shrunk search space (‘.’).
iterations, which is the number of points contained in the underlying grid $A$: after 100 iterations all of them attain an error rate of 4%. However the GPQE+ strategy, which is devised for adaptively adding points to the original grid, can attain an error rate of 1%.

Figure 4: Means of the relative errors at each iteration. GPQE (black squares and plain line), GPQE+ (black crosses and plain line), RBM (light gray triangles and plain line), SUR with $N = 10$ (light gray circles and plain line) and $N = 50$ (light gray ‘- -’ line) for $f_2$ (left) and $f_3$ (right).

Figure 5: Left: Means of the relative errors at each iteration. GPQE (black squares and plain line), GPQE+ (black crosses and plain line), RBM (light gray triangles and plain line) and SUR with $N = 10$ (light gray circles and plain line). Right: Zoom of the left part of the figure.

Since the performance of the RBM methodology are closely linked to the value of $\beta$, which is very difficult to choose in practice, we shall not use this method for our applications to real data.

3.2 Three dimensional example

As we elaborated our strategies in order to apply them to dimension greater than one or two, we study another synthetic example, where $f$ is a function sampled from a zero-mean GP with a SE covariance function defined in (4) with $d = 3$ and $\ell_1 = \ell_2 = \ell_3 = 0.1$ and $X$ is uniformly
distributed on $[0,1]^3$. As we have done in Section 3.1, we compare the three strategies on this function, starting with the same two points picked among the same underlying grid $A$ of cardinality 1000 generated by using a Latin Hypercube design. Here we shall compare the different quantile estimators with $\tilde{Q}_{0.95} = 1.7109$ computed with $M = 10^5$.

In Figure 6(a), the estimation of $\hat{q}^L_{0.95,t}$ and $\hat{q}^U_{0.95,t}$ are displayed for GPQE+. We can see that they converge to $\tilde{Q}_{0.95}$ as the number of points added to the observation set increases.

With a computer having the following configuration: RAM 12 GB, CPU 3.2 GHz, for the SUR strategy with $N = 50$ it takes about 415 seconds to add only one point. Thus, performing the 500 iterations of the SUR strategy takes more than two days (about 55 hours). Those computation times prevent us from performing Monte Carlo replications. Indeed, the computation time for 50 Monte Carlo replications would be more than 3 months. However, for the GPQE strategy, adding one point, with the same computer, takes no more than 0.3 seconds.

The relative errors associated to the SUR, GPQE and GPQE+ strategies are displayed in Figure 6(b) for one random choice of the grid $A$. From this figure, we can see that GPQE and GPQE+ perform better than SUR. We saw in Section 3.1 that increasing $N$ in the SUR strategy would improve those results, but it would also increase the associated computational load. Actually, in view of those statistical results and the prohibitive computational times for SUR, we shall focus in the sequel on GPQE and GPQE+.

![Figure 6: (a) Quantiles estimated with the GPQE+ strategy at each iteration: $\hat{q}^L_{0.95,t}$ (black bold line), $\hat{q}^U_{0.95,t}$ and $\tilde{q}^L_{0.95,t}$ (up and down, light gray line), and $\tilde{Q}_{0.95}$ (black ‘-’ line); (b) Relative errors at each iteration for GPQE (black line), GPQE+ (black bold line) and SUR (light gray line).](image)
3.3 Additional remarks

Before applying our methods on real data let us first give some hints on how to choose the number of points in A and the number of points from which we start with respect to the dimension $d$. Figure 7 displays the average relative errors of our quantile estimator obtained after 20 iterations for different sizes of the grid A, two different numbers of starting points and two values of $d$. The functions $f$ that we used here are sample of Gaussian processes in the one and two dimensional cases obtained with the SE covariance function. We can see from this figure that increasing the number of starting points does not improve the quantile estimation and that even in the 2D case we obtain an average relative error of 0.01 when $|A| = 200$.

![Figure 7: Average relative errors for different sizes of the grid A in the 1D case (left) and the 2D case (right) with 2 (‘○’) and 10 (‘*’) points in the initial set.](image)

4 Application to real data

4.1 The Japanese pregnant woman model

In this section, we shall consider the case of the estimation of the 95% quantile of the whole body Specific Absorption Rate (SAR) of a fetus. As explained in Section 1, the SAR is evaluated by exposing an anatomically realistic pregnant woman model to a source of EMF; it is important to state that very few of those 3D models are available and reliable for numerical dosimetry. Indeed, whole body pregnant woman models do not exist, as medical data needed to build them is not always available. So here we shall use an anatomically realistic woman model designed by Nagaoka et al. (2004) corresponding to the average dimensions of Japanese women, in which a 26-week-old fetus model has been inserted; this model has already been used for a study in Nagaoka et al. (2007), who provided this model in the framework of the JST-ANR Fetus project. In our application, the
Figure 8: (a) Front view of the Japanese pregnant woman model; (b) side view, with the variation of the elevation; (c) view from above, with the variation of the azimuth.

The pregnant woman model is exposed to 900 MHz vertically polarized electromagnetic plane waves with a 1 Volt per meter amplitude. Plane wave exposure is currently used to assess exposure to far field sources such as base station antennas. The SAR (expressed in W/kg) of the fetus will be considered as a function of two input variables: the azimuth and the elevation of the incident wave. The value of the SAR for a given value of the azimuth and elevation is computed through the Finite Difference in Time Domain (FDTD) method, which is commonly used in the field of dosimetry, see for instance Hirata et al. (2007), Conil et al. (2008) and Wiert et al. (2008). The pregnant woman model is displayed in Figure 8, along with the variation of the angles of the incident wave.

We performed 1000 evaluations of the SAR in the fetus, from a set of points \( A \) obtained with a Latin Hypercube Design which was used in computer experiments for the first time by McKay et al. (1979). The results are displayed on the surface plot of Figure 9 (a). With those 1000 points, we evaluate the quantile estimator to be used as a reference: \( \hat{Q}_{0.95} \). Indeed in this case we cannot evaluate \( \tilde{Q}_{0.95} \), since the available observations are those obtained from the grid.

Since we could not have access to the values of \( f \) at other points than those of the grid \( A \) we cannot use GPQE+ strategy and we will thus focus on the GPQE approach. From an initial set of observations of 5 points randomly selected in \( A \), the GPQE method with a GP prior having a SE covariance function provides the results displayed in Figure 9 (b). It shows that the method leads to pick new observations in an area where the values of \( f \) are high, thus close to the quantile.

In Figure 9 (c), we compared, through 50 Monte Carlo replications on the choices of the initial set of observations, the performance of the GPQE method when different covariance functions are used for the GP prior: squared exponential (SE) and Matérn covariance functions with \( \nu = 3/2 \) and \( \nu = 5/2 \). In Figure 9 (e) and (f) are displayed the means of the estimators of \( \ell_1 \) and \( \ell_2 \). We can observe that after 20 iterations the estimations of \( \ell_1 \) and \( \ell_2 \) are stabilized for all the covariance functions. All of them provide accurate estimations of the quantiles: SE obtained the
best performance at the beginning while after 20 iterations the Matérn covariance function ($\nu = 5/2$) is the most reliable. These facts are confirmed by the results displayed in Figure 9(d) where the means of $\hat{q}^U_{0.95,t}$ and $\hat{q}^L_{0.95,t}$ at each iteration obtained through 50 Monte Carlo replications using a SE or Matérn covariance function are displayed.

### 4.2 The simplified geometric model

In order to have an analytic multidimensional function whose behavior is close to our application, we created a simplified geometric model which represents the fetus exposure to a plane wave. Our modeling relies on two physical facts. On the one hand, at the telecommunication frequencies (from 900 to 2400 MHz), the absorption in the human body is mainly on the surface; thus, the absorption in the human body depends on the area covered by the incident wave, see Conil et al. (2011) and references therein. On the other hand, in the case of a fetus, the incident wave has to go through some of the mother’s tissues before it reaches the baby. The power lost by the wave when crossing the mother’s tissues can be described by an exponential decrease. Hence, if we model the fetus as an ellipsoid inserted in a cylinder which models the mother (see Figure 10), the power absorbed by the fetus is given by:

$$f(x) = aS(x) \exp (-bd(x)),$$  \hspace{1cm} (25)

where $S$ is the surface of the ellipsoid cast by the incident wave, which depends on 8 input variables: the size of the three semi-axes of the ellipsoid, three rotations around the three axes of the ellipsoid, and the azimuth and the elevation of the incident wave; $d$ the distance covered by the incident wave from its entry into the mother tissues to the point where it reaches the fetus, which depends on three other input variables: the radius of the cylinder, and the coordinates of the center of the ellipsoid in the horizontal plane; $x$ is therefore a vector of $\mathbb{R}^11$; $a, b \in \mathbb{R}$ are coefficients to be set for a given pregnant woman model, corresponding to the morphology of the model and the dielectric properties of the tissues. In the sequel, we shall consider $f$ as a function of only 4 input variables. The first ones are the azimuth and the elevation of the incident wave. The others are the rotation of the ellipsoid around its $x$ axis, and the rotation of the ellipsoid around its $y$ axis. As we assume those rotations uniformly distributed, we chose small sizes for the semi-axes of the ellipsoid (2, 1.25 and 5 centimeters); indeed, only small fetuses move in all directions: the bigger they grow, the less they move, so to be realistic, we opted for a small ellipsoid-fetus. We set $a = 5.70 \times 10^{-9}$ and $b = 0.07$, the radius of the cylinder is 13 centimeters and the center of the ellipsoid is positioned as displayed in Figure 10.
Figure 9: (a) Surface plot of the SAR in the fetus according to the elevation and azimuth of the incident wave; (b) contour plot of (a) and points selected for one run of the GPQE strategy on the data; (c), (d), (e) and (f) Monte Carlo replications, using a SE (black crosses), a Matérn with $\nu = 3/2$ (black squares) or a Matérn with $\nu = 5/2$ (gray circles) covariance function for the GP prior: (c) means of the relative errors at each iteration for quantile estimation; (d) means of $\hat{q}_{0.95,t}^U$ and $\hat{q}_{0.95,t}^L$ at each iteration (the black ‘.-’ line on (d) is $\tilde{q}_{0.95}$); means of the estimated (e) $\ell_1$ and (f) $\ell_2$ at each iteration.
As in the previous applications, we are interested in estimating the 95% quantile of the distribution of $f$. We begin by computing the value of $\tilde{Q}_{0.95}$ with a sample of $10^5$ observations; $\tilde{Q}_{0.95} = 8.6808 \times 10^{-4}$. We start with 10 points randomly chosen from a grid $A$ of cardinality 5000 points in $[\frac{\pi}{4}, \frac{3\pi}{4}] \times [0; 2\pi]^3$. We perform the GPQE and GPQE+ strategies using three different covariance functions as in Section 4.1: the SE covariance function defined in (6) and the Matérn covariance function defined in (6) with $\nu = 3/2$ and $\nu = 5/2$. The parameter $\ell$ is estimated as explained in Section 2.5. The relative errors for the different quantile estimators are displayed in Figure 11 (a) and 11 (b) for GPQE and GPQE+ respectively. From these figures, we can see that the two strategies provide accurate estimators of the quantile and that the best performance is obtained by using the SE covariance function.

We display in Figure 11 (c) and 11 (d) the estimated values of the quantiles $\hat{q}_{0.95,t}$, $\hat{q}_{0.95,t}^U$, and $\hat{q}_{0.95,t}^L$, for the GPQE+ strategy, using the SE covariance function. The values of the estimates of $\ell_1$, $\ell_2$, $\ell_3$ and $\ell_4$ for this strategy are displayed in Figures 11 (e) and (f). We can see that by using only 600 evaluations of $f$ we obtain an error of 2% on the value of $\tilde{Q}_{0.95}$ that had been computed with $10^5$ evaluations of $f$. Thus, with 166 times fewer evaluations of $f$ we obtain a very accurate estimation of the 95% quantile applying our adaptive approaches.

5 Conclusion

In this paper, we propose and compare two novel sequential strategies for estimating the quantile of $Y = f(X)$, where $X$ has a known distribution in $\mathbb{R}^d$ and $f$ is an unknown, expensive-to-evaluate real-valued function. Our methodologies are based on confidence bound criteria coming from Bayesian optimization and adapted to our quantile estimation issue. In the course of this study, we have shown that GPQE and GPQE+ have two main features that make them very attractive. Firstly, both of them are very efficient for providing accurate quantile estimations by using
Figure 11: (a) and (b): relative errors at each iteration for the GPQE (a) and GPQE+ (b) strategies, comparing the SE (black bold line), the Matérn with $\nu = 3/2$ (black line) and the Matérn with $\nu = 5/2$ (light gray line) covariance functions; (c): Estimation of $\hat{q}_{0.95,t}$ (black bold line), $\hat{q}_{0.95,t}^U$ and $\hat{q}_{0.95,t}^L$ (up and down, light gray line) at each iteration for GPQE+, and $\tilde{Q}_{0.95}$ (black ’.-’ line) and $\tilde{q}_{0.95} \pm 10\%$ (light gray ’.-’ line). (d): estimation of $\ell_1$ (triangle), $\ell_2$ (diamond), $\ell_3$ (dot) and $\ell_4$ (square) at each iteration for GPQE+.

few evaluations of the function $f$ even for large $d$. Secondly, their computational simplicity and efficiency make their implementation feasible. This implementation has been successfully tested in a real data application context corresponding to the exposure of a Japanese pregnant-woman model and her 26-week-old fetus to a plane wave in the framework of the ANR-JST FETUS project.

6 Supplementary material

The proofs of Propositions 1, 2 and 3 are given in the supplementary material Jala et al. (2014).
References


Jala, M., C. Lévy-Leduc, E. Moulines, E. Conil, and J. Wiart (2014). Supplementary material to the paper: “sequential design of computer experiments for the assessment of fetus exposure to electromagnetic fields”.


Supplementary material: Proofs of the paper “Sequential Design of Computer Experiments for the Assessment of Fetus Exposure to Electromagnetic Fields”

We begin with a Lemma from Srinivas et al. (2012):

**Lemma 1.** (Srinivas et al., 2012, Lemma 5.1) Given any finite subset \( A \) of \( \mathbb{A} \), any integer \( t \) and any pairwise distinct \( \{x_1, \ldots, x_t\} \subset A \), and \( f : A \rightarrow \mathbb{R} \) a sample from \( \text{GP}(0,k(\cdot, \cdot)) \), for all \( \delta \) in \( (0, 1) \), we have

\[
\mathbb{P} \left( \forall x \in A : \forall t \geq 1 : f(x) \in [\mu_t^L(x); \mu_t^U(x)] \right) \geq 1 - \delta ,
\]

(26)

where \( \mu_t^L(x) \) and \( \mu_t^U(x) \) are defined in (12).

The proof of Lemma 1 is given in Srinivas et al. (2012), but in order to make our paper self-contained, we give it hereafter.

**Proof of Lemma 1** Fix \( t \geq 1 \) and \( x \in A \). In the case of noise free observations, for a given set \( \{y_1, \ldots, y_t\} \) of evaluations of \( f \) at the points \( \{x_1, \ldots, x_t\} \) such that \( y_i = f(x_i) \), \( 1 \leq i \leq t \), according to the prior we have

\[
f(x) \mid (x_1, y_1), \ldots, (x_t, y_t) \sim \mathcal{N} \left( \mu_t(x), \sigma_t^2(x) \right),
\]

(27)

where \( \mu_t(x) \) is defined in (2) and \( \sigma_t^2(x) = k_t(x, x) \) defined in (3). Now, let \( Z \) be a random variable such that \( Z \sim \mathcal{N}(0, 1) \); thus, for all \( c > 0 \),

\[
\mathbb{P}(Z > c) = \frac{1}{\sqrt{2\pi}} \int_c^{+\infty} e^{-\frac{1}{2}r^2} \, dr = \frac{e^{-\frac{1}{2}c^2}}{\sqrt{2\pi}} \int_c^{+\infty} e^{-\frac{1}{2}(r-c)^2} \, dr .
\]

For \( r > c > 0 \), \( e^{-c(r-c)} \leq 1 \), hence we get that

\[
\mathbb{P}(Z > c) \leq \frac{e^{-\frac{1}{2}c^2}}{\sqrt{2\pi}} \int_c^{+\infty} e^{-\frac{1}{2}(r-c)^2} \, dr = e^{-\frac{1}{2}c^2} \mathbb{P}(Z > 0) = e^{-\frac{1}{2}c^2} .
\]

By (27), \( \frac{f(x) - \mu_t(x)}{\sigma_t(x)} \sim \mathcal{N}(0, 1) \), and for \( c = \sqrt{\beta_t} \) we have

\[
\mathbb{P} \left( |f(x) - \mu_t(x)| > \sqrt{\beta_t} \sigma_t(x) \mid (x_1, y_1), \ldots, (x_t, y_t) \right) \leq e^{-\frac{c^2}{2}} .
\]

Applying the union bound, we get that

\[
\mathbb{P} \left( \forall x \in A : \forall t \geq 1 : |f(x) - \mu_t(x)| \leq \sqrt{\beta_t} \sigma_t(x) \mid (x_1, y_1), \ldots, (x_t, y_t) \right) \geq 1 - |A| \sum_{t \geq 1} e^{-\frac{\beta_t}{2}} ,
\]

where \( |A| \) denotes the cardinality of the set \( A \). By definition of \( \beta_t \) in (13) we get (26), which concludes the proof of Lemma 1. □
Proof of Proposition 1. By Lemma (1), with probability greater than \((1 - \delta)\), for any finite subset \(A = \{x_1, \ldots, x_m\}\), for any integer \(t\) and any pairwise distinct \(\{x_1, \ldots, x_t\} \subset A\), we have for all \(x\) in \(A\)

\[
\mu_t^L(x) \leq f(x) \leq \mu_t^U(x),
\]

where \(\beta_t\) is defined in (13). Thus, for all \(q \in \mathbb{R}\),

\[
\frac{1}{m} \sum_{i=1}^{m} \mathbbm{1}_{\{\mu_t^U(x_i) \leq q\}} \leq \frac{1}{m} \sum_{i=1}^{m} \mathbbm{1}_{\{f(x_i) \leq q\}} \leq \frac{1}{m} \sum_{i=1}^{m} \mathbbm{1}_{\{\mu_t^L(x_i) \leq q\}}.
\]

(28)

Let \(q = \hat{q}_{\alpha,t}^U\) in (28) then, by (14), we get that

\[
\frac{1}{m} \sum_{i=1}^{m} \mathbbm{1}_{\{f(x_i) \leq \hat{q}_{\alpha,t}^U\}} \geq \alpha.
\]

Thus, by (9), we obtain that \(\hat{q}_{\alpha} \leq \hat{q}_{\alpha,t}^U\). Let \(q = \hat{q}_{\alpha}\) in (28) then, by (9), we get that

\[
\frac{1}{m} \sum_{i=1}^{m} \mathbbm{1}_{\{\mu_t^L(x_i) \leq \hat{q}_{\alpha}\}} \geq \alpha.
\]

Thus, by (14), we obtain that \(\hat{q}_{\alpha,t}^L \leq \hat{q}_{\alpha}\), which concludes the proof of Proposition 1. 

Proof of Proposition 2. Let us first prove that for any integer \(t\) and any pairwise distinct \(\{x_1, \ldots, x_t\} \subset A\),

\[
\hat{q}_{\alpha,t} \leq \hat{q}_{\alpha,t}^L + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x).
\]

(29)

By (14), for all \(t \geq 1\),

\[
\frac{1}{m} \sum_{i=1}^{m} \mathbbm{1}_{\{\mu_t^L(x_i) \geq \hat{q}_{\alpha,t}^L\}} < 1 - \alpha.
\]

By definition of \(U_{\alpha,t}\) given in (15), we get that

\[
\frac{1}{m} \sum_{1 \leq i \leq m, x_i \in U_{\alpha,t}} \mathbbm{1}_{\{\mu_t^L(x_i) > \hat{q}_{\alpha,t}^L\}} = \frac{1}{m} \sum_{i=1}^{m} \mathbbm{1}_{\{\mu_t^L(x_i) \geq \hat{q}_{\alpha,t}^L\}} < 1 - \alpha.
\]

Since by definition \(\mu_t^L(x_i) = \mu_t(x_i) - \sqrt{\beta_t} \sigma_t(x_i)\), we obtain that

\[
\frac{1}{m} \sum_{1 \leq i \leq m, x_i \in U_{\alpha,t}} \mathbbm{1}_{\{\mu_t(x_i) > \hat{q}_{\alpha,t}^L + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x_i)\}} \leq \frac{1}{m} \sum_{1 \leq i \leq m, x_i \in U_{\alpha,t}} \mathbbm{1}_{\{\mu_t(x_i) > \hat{q}_{\alpha,t}^L + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x_i)\}} < 1 - \alpha.
\]

By definition of \(U_{\alpha,t}\) given in (15), we get that

\[
\frac{1}{m} \sum_{i=1}^{m} \mathbbm{1}_{\{\mu_t(x_i) > \hat{q}_{\alpha,t}^L + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x_i)\}} = \frac{1}{m} \sum_{1 \leq i \leq m, x_i \in U_{\alpha,t}} \mathbbm{1}_{\{\mu_t(x_i) > \hat{q}_{\alpha,t}^L + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x_i)\}} < 1 - \alpha.
\]
By (16), we get (29). By Proposition 1 with probability greater than \((1 - \delta)\), we get that for any integer \(t\) and any pairwise distinct \(\{x_1, \ldots, x_t\} \subset A\),
\[
\hat{q}_{\alpha,t} \leq \tilde{q}_\alpha + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x) .
\] (30)

Let us now prove that for all \(t \geq 1\),
\[
\hat{q}^U_{\alpha,t} \leq \tilde{q}_{\alpha,t} + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x) .
\] (31)

Since for all \(t \geq 1\), \(\mu^U_t(x) \leq \mu_t(x) \leq \mu^U_t(x)\), we get by using the same arguments as those used in Proposition 1 that \(\hat{q}^L_{\alpha,t} \leq \tilde{q}_{\alpha,t} \leq \hat{q}^U_{\alpha,t}\), for all \(t \geq 1\). Thus, by definition of \(U_{\alpha,t}\) given in (15), we get that
\[
\frac{1}{m} \sum_{i=1}^{m} \left[ \mu^U_t(x_i) > \hat{q}_{\alpha,t} + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x_i) \right] = \frac{1}{m} \sum_{i=1}^{m} \left[ \mu^U_t(x_i) > \hat{q}_{\alpha,t} + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x_i) \right] .
\] (32)

Since \(\mu^U_t(x) = \mu_t(x) + \sqrt{\beta_t} \sigma_t(x)\),
\[
\frac{1}{m} \sum_{i \in \{1, \ldots, m\} \atop x_i \notin U_{\alpha,t}} \left[ \mu^U_t(x_i) > \hat{q}_{\alpha,t} + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x_i) \right] 
\leq \frac{1}{m} \sum_{i \in \{1, \ldots, m\} \atop x_i \notin U_{\alpha,t}} \left[ \mu_t(x_i) + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x_i) > \hat{q}_{\alpha,t} + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x_i) \right]
\leq \frac{1}{m} \sum_{i \in \{1, \ldots, m\} \atop x_i \notin U_{\alpha,t}} \left[ \mu_t(x_i) > \hat{q}_{\alpha,t} \right] < 1 - \alpha .
\]

By (32) and (14), we get (31). By Proposition 1 with probability greater than \((1 - \delta)\), we get that for all \(t \geq 1\),
\[
\hat{q}_\alpha \leq \tilde{q}_{\alpha,t} + \sqrt{\beta_t} \sup_{x \in U_{\alpha,t}} \sigma_t(x) .
\] (33)

Using (30) and (33), we get (17) which concludes the proof of Proposition 2. \(\Box\)

**Proof of Proposition 3** By (26), with probability greater than \((1 - \delta)\), for any finite subset \(A = \{x_1, \ldots, x_m\}\), for any integer \(t\) and any pairwise distinct \(\{x_1, \ldots, x_t\} \subset A\), for all \(x\) in \(A\), we have
\[
\mu^L_t(x) \leq f(x) \leq \mu^U_t(x) ,
\]
where \(\beta_t\) is defined in (13). Thus, with probability greater than \((1 - \delta)\), by definition of \(\mu^U_t(x)\) and \(\mu^L_t(x)\) given in (12),
\[
f(x) - \sqrt{\beta_t} \sigma_t(x) \leq \mu_t(x) \leq f(x) + \sqrt{\beta_t} \sigma_t(x) .
\] (34)
Since for all $x$ in $U_{\alpha,t}$,

$$
\mu_t(x) + \sqrt{\beta_t} \sigma_t(x) \geq \hat{q}_{\alpha,t}^L,
$$

we deduce from (34) that for all $x$ in $U_{\alpha,t}$,

$$
\hat{q}_{\alpha,t}^L \leq f(x) + 2\sqrt{\beta_t} \sigma_t(x),
$$

which gives the first inequality of (18) by (29) and (33). Since for all $x$ in $L_{\alpha,t}$,

$$
\mu_t(x) \leq \hat{q}_{\alpha,t}^U + \sqrt{\beta_t} \sigma_t(x),
$$

we deduce from (34) that for all $x$ in $L_{\alpha,t}$,

$$
\hat{q}_{\alpha,t}^U \geq f(x) - 2\sqrt{\beta_t} \sigma_t(x),
$$

which gives the last inequality of (18) by (31) and (30) and concludes the proof of Proposition 3. \qed