SEQUENTIAL DESIGN OF COMPUTER EXPERIMENTS FOR PARAMETER ESTIMATION WITH APPLICATION TO NUMERICAL DOSIMETRY

Marjorie Jala[1,2], Céline Lévy-Leduc[1], Éric Moulines[1], Emmanuelle Conil[2] and Joe Wiart[2]

1: CNRS/ LTCI/ Télécom ParisTech, 2: Orange Labs and Whist Lab
contact: jala@telecom-paristech.fr

ABSTRACT

In this paper, we propose a sequential sampling approach for estimating a parameter of interest of the distribution 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. We shall adopt a Bayesian point of view which consists in modeling f as a sample of a well-chosen Gaussian process. Our global approach aims at estimating the parameter of interest with as few evaluations of f as possible. We compare our methodology with standard approaches through numerical experiments and eventually test it on real data corresponding to the exposure of a Japanese pregnant-woman model and her 26-week-old fetus to a plane wave.

Index Terms— Sequential Design, Computer experiment, Gaussian Process, Bayesian approach.

1. INTRODUCTION

Over the past 30 years, wireless communication systems have been increasingly used. The number of mobile phones, WIFI boxes, antennas, etc., is growing together with a strong public concern over possible health problems related to the exposure to radio frequency (RF) electromagnetic fields (EMF). In order to tackle the issues raised by the human being exposure, many studies are currently carried out: long-term epidemiological studies, in vitro and in vivo studies, and numerical dosimetry based methods. Here, we shall consider the latter point of view to deal with the complex issue of fetuses exposure to EMF throughout their development. We perform numerical dosimetry simulations, virtually exposing pregnant woman and fetus 3D-models to one source of EMF; one simulation yields one value of the (whole body) Specific Absorption Rate (SAR) in the fetus, which is an evaluation of the rate at which energy is absorbed by the body of the fetus. As the SAR depends on several parameters such as the morphology and the posture of the mother and fetus, the position and the type of the wireless device, we shall model the SAR by $Y = f(\mathbf{X})$, where f is an unknown real-valued function and $\mathbf{X}$ is a random vector of $\mathbb{R}^d$ having a known distribution. Our goal is to propose a method for estimating a parameter of interest $\theta(f)$ of the unknown distribution of Y, which can be for

instance its mean, variance, the probability $\mathbb{P}(Y \geq t)$ where t is a given threshold, or a quantile. Since an evaluation of f for a given x in $\mathbb{R}^d$ is very expensive in terms of computational load, we shall propose a sequential sampling strategy where each computer trial is selected in order to estimate $\theta(f)$ as accurately as possible by performing as few evaluations of f as possible. For this, we shall adopt a Bayesian point of view, using a Gaussian Process as a surrogate model for the unknown function f; this surrogate model enables us to optimize a criterion which selects the next evaluation point of f. This framework has already been used in [1] and [2] to provide a sequential approach for contour and $\mathbb{P}(Y \geq t)$ estimation, respectively. Besides, [3], [4], [5], and references therein have also used Bayesian methods for finding the maximum of f.

This paper is organized as follows: in Section 2, we describe our approach for estimating a parameter of interest of the distribution of Y; in Section 3, we test the methodology for the estimation of a quantile of three analytic two-dimensional functions and compare it with two other methods; eventually, in Section 4 we apply our methodology to the assessment of the exposure of a 26-week-old fetus to a plane wave through numerical dosimetry simulations.

2. DESCRIPTION OF THE METHOD

In this section, we shall explain our sequential sampling strategy for estimating a parameter of interest of the distribution of $Y = f(X)$, where f is an unknown function and X is for simplicity assumed to be uniformly distributed on $[0,1]^d$. The method that we propose consists in adding one by one the points of a fine grid $\mathcal{X}$ of $[0,1]^d$ to the observation set by using a Bayesian approach. It consists in considering f as a sample of a zero-mean Gaussian process having a covariance function k that we shall denote in the sequel $\text{GP}(0, k(x,x'))$. The advantage of this approach is that the posterior distribution is still a GP such that, from a set of observations $\mathbf{y}_T = (Y_1, \ldots, Y_T)'$, the posterior mean $\mu_T(x)$

and covariance $k_T(x, x')$ are given by

$$\mu_T(x) = \mathbf{k}_T(x)' \mathbf{K}_T^{-1} \mathbf{y}_T , \qquad (1)$$

$$k_T(x, x') = k(x, x') - \mathbf{k}_T(x)' \mathbf{K}_T^{-1} \mathbf{k}_T(x') , \qquad (2)$$

where $\mathbf{k}_T(x) = [k(x_1, x) \ldots k(x_T, x)]'$, x is in $\mathcal{X}$, $'$ denotes the matrix transposition and $\mathbf{K}_T = [k(x_i, x_j)]_{1 \leq i,j \leq T}$, see [6] and [3] for further details. As usual, we shall take μ_T as an estimator of f based on the observations $\{(X_1, f(X_1)), \ldots, (X_T, f(X_T))\} = \mathcal{F}_T$. For selecting a new evaluation point of f, 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. So, the $(T+1)$th point to add to the set $\mathcal{F}_T$ is:

$$x_{T+1} = \arg\min_{x \in \mathcal{X}} \mathrm{V}_T(x) , \qquad (3)$$

where

$$\mathrm{V}_T(x) = \int \mathrm{Var}(\theta | \mathcal{F}_T^{(x,y)}) \varphi_{\mu_T(x), \sigma_T^2(x)}(y) \mathrm{d}y , x \in \mathcal{X} ,$$

with $\mathcal{F}_T^{(x,y)} = \{(X_1, f(X_1)), \ldots, (X_T, f(X_T)), (x, y)\}$, $\sigma_T^2(x) = k_T(x, x)$ 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 $\sigma_T^2(x)$. This criterion is closed to one of the improvement-based acquisition functions for Bayesian Optimization described in [5] which has been studied in [7]. Observe that this strategy corresponds to the so-called SUR criterion proposed by Bect and al. in [2] in the case of the estimation of $\mathbb{P}(Y \geq t)$ where t is a given threshold.

The criterion (3) has a closed-form expression only in particular cases; for instance, when $\theta(f) = \mathbb{E}(Y)$. Otherwise, the computation of (3) is more involved; this is pointed out for instance in [2] for the case of the estimation of $\mathbb{P}(Y \geq t)$. In those situations where no closed-form formula for (3) is available, we shall explain in Algorithm 1 how Monte-Carlo simulations can be used to approximate (3). Note that a similar algorithm has been proposed in [8] with another application.

3. NUMERICAL EXPERIMENTS

In this section, we shall apply the methodology introduced in Section 2 to the estimation of a α-quantile q_α of the distribution of $Y = f(X)$, for three instances of two-dimensional functions f; q_α is defined by:

$$\mathbb{P}(f(\mathbf{X}) \leq q_\alpha) = \alpha , \alpha \in (0, 1) .$$

The first function we shall use is a Gaussian Process path, sampled from a zero-mean GP with a squared exponential co-

---

Algorithm 1 Procedure for the estimation of $\theta(f)$

Input: $\mathcal{X}$, a fine grid of $[0, 1]^d$, T_0 a small number of points of $\mathcal{X}$ where f is evaluated $\{(X_1, f(X_1)), \ldots, (X_{T_0}, f(X_{T_0}))\} = \mathcal{F}_{T_0}$, $\mathcal{X}_0 = (X_1, \ldots, X_{T_0})$.

Processing

1. Evaluate the posterior distribution of the GP denoted ξ_{T_0} using (1) and (2)

2. Evaluate the estimation of $\theta(f)$ with the surrogate model μ_{T_0}

3. Evaluate the criterion

   (a) Simulate N paths $\xi_{T_0}^{(1)}, \ldots, \xi_{T_0}^{(N)}$ of the GP ξ_{T_0}

   (b) For each point $x \in \mathcal{X} / \mathcal{X}_0$ evaluate the posterior distribution of the GP $\xi_{T_0^{(x, \xi_{T_0}^i(x))}}$, $i \in \{1, \ldots, N\}$, conditionally to the addition of $(x, \xi_{T_0}^i(x))$ to $\mathcal{F}_{T_0}$

   (c) Simulate N' paths $\xi_{T_0^{(x, \xi_{T_0}^i(x))}}^{(1)}, \ldots, \xi_{T_0^{(x, \xi_{T_0}^i(x))}}^{(N')}$ of the GP $\xi_{T_0^{(x, \xi_{T_0}^i(x))}}$, $i \in \{1, \ldots, N\}$

   (d) For each of those N' paths estimate $\theta(\xi_{T_0^{(x, \xi_{T_0}^i(x))}}^{(j)})$, $j \in \{1, \ldots, N'\}$, and the empirical variance of $\theta(f)$ on these N' estimations denoted by $s_T^{(i)}(x)$

   (e) Evaluate the mean of the N empirical variances of $\theta(f)$: $V_T(x) = \frac{1}{N} \sum_{i=1}^N s_T^{(i)}$ is the value of the criterion for the points x of the grid $x \in \mathcal{X} / \mathcal{X}_0$

   (f) Repeat steps (c), (d), (e) for each point x of the grid $x \in \mathcal{X} / \mathcal{X}_0$

4. Select the point x for which $V_T(x)$ is minimal and perform an evaluation of f at this point

5. Add this new observation to the set of observations

6. Repeat all the steps until the maximal allowed number of evaluation of f is reached

Output Estimator of $\theta(f)$

---

variance function:

$$k(x, x') = \exp\left(-\frac{(x - x')^2}{2\ell^2}\right) , x, x' \in [0, 1]^2 , \ell = 0.1 . \tag{4}$$

The second function is the function used in [9], that we shall refer to as Gramacy function:

$$f(x, y) = (8x - 2) \exp(-(8x - 2)^2 - (8y - 2)^2) ,$$
$$(x, y) \in [0, 1]^2 .$$

The third function is the so-called Ishigami function (with a fixed parameter to be two-dimensional), which is commonly used in sensitivity analysis (see for instance [10]):

$$f(x, y) = \sin(\pi(2x - 1)) + 7\sin(\pi(2y - 1))^2$$
$$+ 0.1\pi^4 \times \sin(\pi(2x - 1)), \ (x, y) \in [0, 1]^2 .$$

Those three functions are displayed in Figure 1.

(a)

(b)

(c)
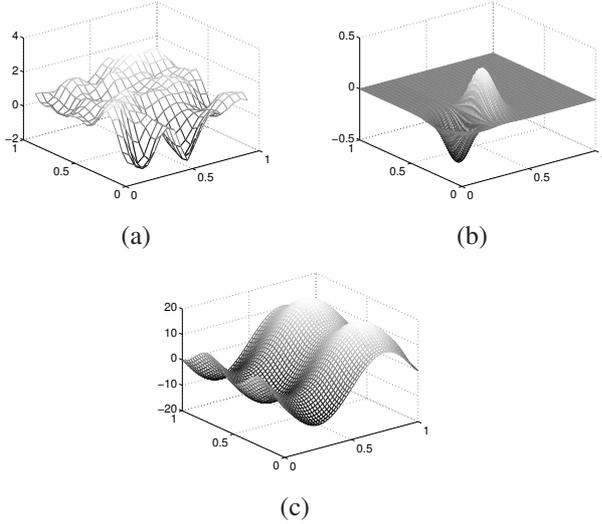
Fig. 1. 3D plots of the functions; (a) Gaussian Process path, (b) Gramacy function and (c) Ishigami function.

In the following, our approach that we shall call GPquantile is to be compared for each function with two other methods: one that we shall refer to as GPexplor, in which the point of $\mathcal{X}$ to be added to a set of T observations is x_{T+1} defined by

$$x_{T+1} = \arg\max_{x \in \mathcal{X}} \sigma_T(x) ,$$

where $\sigma_T(x)^2 = k_T(x, x)$ defined in (2). The idea here is to reduce step by step the global incertitude on the estimation of the function f. This criterion is introduced in [3] as a pure exploration strategy, hence the name we use here. With this approach, the quantile q_α is estimated by

$$\hat{q}_{\alpha, \text{GPexplor}}(T) = (\mu_T(\mathcal{X}))_{(\lceil n\alpha \rceil)} ,$$

where $(\mu_T(\mathcal{X}))_{(i)}$ denotes the ith smallest value that μ_T takes on the grid $\mathcal{X}$. The other method, that we shall denote Random, is a more naive approach which consists in randomly adding a new point at each iteration and evaluating the quantile at step T by

$$\hat{q}_{\alpha, \text{Random}}(T) = Y_{(\lceil T\alpha \rceil)} .$$

Here we shall focus on the estimation of the 95% quantile of Y, where X is uniformly distributed on $[0, 1]^2$. For each

function, we consider as a reference value the estimation of the quantile provided by all the $n = 500$ points of $\mathcal{X}$; this value will be denoted $\hat{q}_{0.95}$ in the sequel. For the GP path, $\hat{q}_{0.95} = 1.7678$, for the Gramacy function $\hat{q}_{0.95} = 0.0870$, and for the Ishigami function $\hat{q}_{0.95} = 16.0579$. We start with a set $\mathcal{X}_0$ of $T_0 = 20$ ($T_0 = 10$ for the Ishigami function) points selected with a maximin LHS obtained with the `lhsdesign` function of Matlab. This set is included in the larger set $\mathcal{X}$ also obtained with a LHS.

Furthermore, to evaluate the criterion, we simulate $N = N' = 10$ GP sample paths. The number N of Monte Carlo may seem small, but this value yields satisfactory results as we shall see.

In the contour plots of Figure 2, the contour lines of $\hat{q}_{0.95}$ and $\hat{q}_{0.95} \pm 10\%$ are drawn, and the different points as well as their order of apparition are displayed. We can see that the points added to the set are spread out in all the space for the Gramacy function; on the contrary, most of the points have been selected near from the quantile line for the Ishigami function.

In Figure 3, the results are displayed through the relative errors $|\hat{q}_{0.95} - \hat{q}_{\alpha, \cdot}(t)|/\hat{q}_{0.95}$ at each iteration t, $t = 1, \ldots, 50$ ($t = 1, \ldots, 30$ for the Ishigami function) using the three previous approaches.. For the GP path, in this example the relative error is inferior to 5% after 18 iterations ; in other words, we achieve a good precision on a value obtained with 500 observations using only 38 evaluations of the function. Moreover, the results with the Ishigami function are even better: with a total of only 16 observations, the procedure yields an estimator of $q_{0.95}$ which has less than 1% error from the $\hat{q}_{0.95}$ estimated with 1000 evaluations of the real function. The results are less impressive for the Gramacy function, as we only achieve a 15% relative error after 50 iterations (which corresponds to 70 observations). However, our results are better than the two other methods to which we compared GPquantile. Those results are not really surprising, because the surrogate model of a classic GP seems to be too simple to estimate the function displayed in Figure 1 (b); in [9], the authors use a Bayesian treed Gaussian process as surrogate model, which means that they select the best covariance function at each iteration of the procedure.

3.1. Monte-carlo experiments

The comparison between the three methods is performed through 50 Monte-Carlo replications (on the choice of the sets $\mathcal{X}_0$) and the corresponding means of the relative errors $|\hat{q}_{0.95} - \hat{q}_{\alpha, \cdot}(T)|/\hat{q}_{0.95}$ at each iteration t, $t = 1, \ldots, 50$ ($t = 1, \ldots, 30$ for the Ishigami function) are displayed in Figure 4 for each of the three approaches. We can see from this figure that our approach outperforms the Random and the GPexplor method.

(a)   (b)

(c)
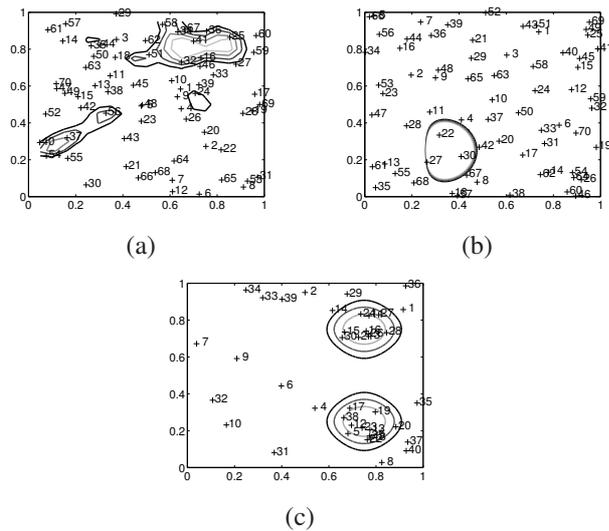
Fig. 2. Contours close to the quantile and chosen points in order of inclusion in the set for (a) Gaussian Process Path, (b) Gramacy function and (c) Ishigami function.

(a)   (b)

(c)
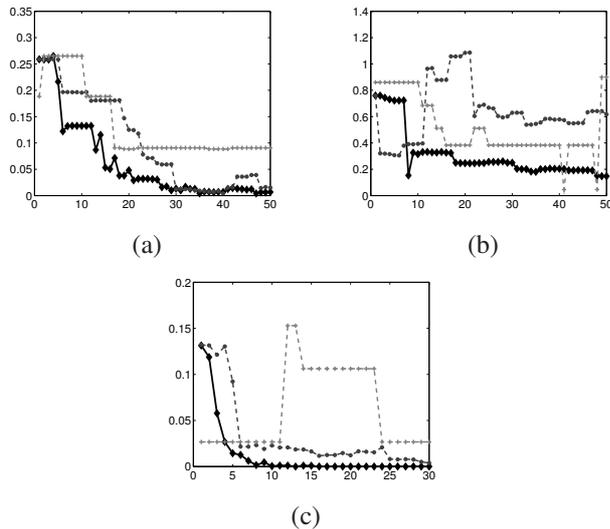
Fig. 3. Relative errors at each iteration (for (a) Gaussian Process Path, (b) Gramacy function and (c) Ishigami function). GPquantile (black diamonds and plain line), GPexplor (dark gray circles and dotted line) and Random (light gray crosses and dotted line).

4. APPLICATION TO REAL DATA

In this section, we shall apply the methodology developed in Section 2 to the estimation of the 95% quantile of the SAR of a 26-week-old fetus. Our application is all the more interesting since most of dosimetric studies are carried out with deterministic approach, which means with one human model in a given posture and one configuration of exposure (such as a frontal incident plane wave). As it has been shown

(a)   (b)
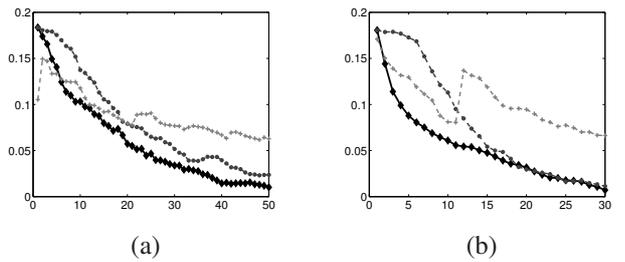
Fig. 4. Means of the relative errors at each iteration. GPquantile (black diamonds and plain line), GPexplor (dark gray circles and dotted line) and Random (light gray crosses and dotted line) for (a) GP path and (b) Ishigami function.

in [11] and [12], that morphology, posture and position of the EMF source have an high influence on the exposure, the JST-ANR Fetus project in which this work is included aims at statistically analyzing the exposure of fetuses. We shall use an anatomically realistic woman model corresponding to the average dimensions of Japanese women, in which a 26-week-old fetus model has been inserted (see [13] and references therein); indeed, whole body pregnant woman models do not exist, as medical data needed to build them is not always available. In our application, the pregnant woman model is exposed to 900 MHz vertically polarized electromagnetic plane waves with a 1 Volt per meter amplitude.

The SAR (expressed in W/kg) of the fetus will be considered as a function of two parameters: 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 [14], [11] and [15]. We performed 500 evaluations of the SAR in the fetus, from a set of 20 points joint to a set of 480 points, both obtained with the lhsdesign function of Matlab. The results are displayed on the surface plot of Figure 5 (a). With those 500 points, we evaluate the estimator of the quantile to be used as a reference: $\hat{q}_{0.95} = 4.2204 \times 10^{-4}$. Then, we run the procedure starting with the $T_0 = 20$ points of the LHS design. In Figure 5 (b) the contour plot shows that the points are picked in the area close to the quantile; the values of the 95% quantile estimators at each iteration t obtained using the GPquantile are displayed in Table 1. We observe that the value is stabilized at 4.2204×10^{-4} after 10 iterations, which is exactly $\hat{q}_{0.95}$.

5. CONCLUSION

In this paper, we have developed a sequential sampling strategy based on a Bayesian approach for estimating a parameter of interest of the distribution of an expensive to evaluate black-box function. Our method outperforms other strategies when tested on synthetic examples, and yields very good results when applied to real data.

Iteration (t)	1	2	3	4	5
$\hat{q}(t) \times 10^4$	3.4030	3.5877	3.8057	3.8057	4.0260
Iteration (t)	6	7	8	9	10
$\hat{q}(t) \times 10^4$	4.1687	4.1736	4.1991	4.2275	4.2204
Iteration (t)	15	20	30	40	50
$\hat{q}(t) \times 10^4$	4.2204	4.2204	4.2204	4.2204	4.2204

Table 1. The quantile estimators at the different iteration steps.

(a)

(b)   (c)
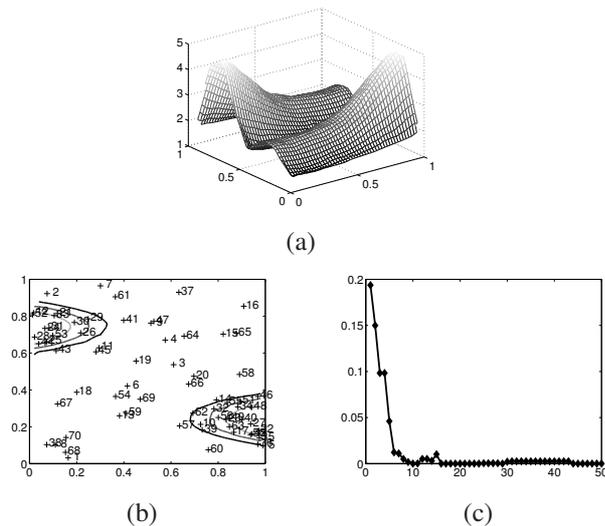
Fig. 5. (a) Contour plot, (b) selected points for the quantile estimation and (c) relative errors at each iteration for the quantile estimation.

6. REFERENCES

[1] P. Ranjan, D. Bingham, and G. Michailidis, "Sequential experiment design for contour estimation from complex computer codes," *Technometrics*, vol. 50, no. 4, pp. 527–541, 2008.

[2] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez, "Sequential design of computer experiments for the estimation of a probability of failure," *Statistics and Computing*, pp. 1–21, 2011.

[3] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Gaussian process bandits without regret: An experimental design approach," *CoRR*, vol. abs/0912.3995, 2009.

[4] E. Vazquez and J. Bect, "Convergence properties of the expected improvement algorithm with fixed mean and covariance functions," *Journal of Statistical Planning and Inference*, vol. 140, no. 11, pp. 3088 – 3095, 2010.

[5] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *CoRR*, vol. abs/1012.2599, 2010.

[6] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press, December 2005.

[7] J. Mockus, V. Tiesis, and A. Zilinskas, "The application of bayesian methods for seeking the extremum," *Towards Global Optimization*, vol. 2, pp. 117–129, 1978.

[8] A. Arnaud, M. Bect, J.and Couplet, A. Pasanisi, and E. Vazquez, "Évaluation d'un risque d'inondation fluviale par planification séquentielle d'expériences," in *42èmes Journées de Statistique*, Marseille, France, France, 2010.

[9] Gramacy, R. B. and Lee, H. K. H., "Adaptive Design and Analysis of Supercomputer Experiments," *Technometrics*, 2009.

[10] A. Saltelli, K. Chan, and E. Scott, *Sensitivity Analysis*, J. Wiley & Sons, 2000.

[11] E. Conil, A. Hadjem, F. Lacroux, M. F. Wong, and J. Wiart, "Variability analysis of sar from 20 MHz to 2.4 GHz for different adult and child models using finite-difference time-domain," *Physics in Medicine and Biology*, vol. 53, no. 6, pp. 1511, 2008.

[12] A. El Habachi, E. Conil, A. Hadjem, E. Vazquez, M. F. Wong, A. Gati, G. Fleury, and J. Wiart, "Statistical analysis of whole-body absorption depending on anatomical human characteristics at a frequency of 2.1 GHz," *Physics in Medicine and Biology*, vol. 55, no. 7, pp. 1875, 2010.

[13] T. Nagaoka, T. Togashi, K. Saito, M. Takahashi, K. Ito, and S. Watanabe, "An anatomically realistic whole-body pregnant-woman model and specific absorption rates for pregnant-woman exposure to electromagnetic plane waves from 10 MHz to 2 GHz," *Physics in Medicine and Biology*, vol. 52, no. 22, pp. 6731, 2007.

[14] A. Hirata, S. Kodera, J. Wang, and O. Fujiwara, "Dominant factors influencing whole-body average SAR due to far-field exposure in whole-body resonance frequency and GHz regions," *Bioelectromagnetics*, vol. 28, no. 6, pp. 484–487, 2007.

[15] J. Wiart, A. Hadjem, M. F. Wong, and I. Bloch, "Analysis of rf exposure in the head tissues of children and adults," *Physics in Medicine and Biology*, vol. 53, no. 13, pp. 3681, 2008.