ANNALES DE LA FACULTÉ DES SCIENCES DE TOULOUSE Mathématiques

NABIL RACHDI, JEAN-CLAUDE FORT, THIERRY KLEIN
Stochastic Inverse Problem with Noisy Simulator. Application to aeronautical model

Stochastic Inverse Problem with Noisy Simulator
– Application to aeronautical model –

Nabil Rachdi[1], Jean-Claude Fort[2], Thierry Klein[3]

Abstract. — Inverse problem is a current practice in engineering where the goal is to identify parameters from observed data through numerical models. These numerical models, also called *Simulators*, are built to represent the phenomenon making possible the inference. However, such representation can include some part of variability or commonly called *uncertainty* (see [4]), arising from some variables of the model. The phenomenon we study is the fuel mass needed to link two given countries with a commercial aircraft, where we only consider the *Cruise* phase. From a data base of fuel mass consumptions during the cruise phase, we aim at identifying the *Specific Fuel Consumption* (*SFC*) in a robust way, given the uncertainty of the *cruise speed V* and the *lift-to-drag ratio F*. In this paper, we present an estimation procedure based on Maximum-Likelihood estimation, taking into account this uncertainty.

Résumé. — Le problème inverse est une pratique assez courante en ingénierie, où le but est de déterminer les causes d'un certain phénomène à partir d'observations de ce dernier. Le phénomène mis en jeu est représenté par un modèle numérique, dont certaines composantes peuvent comporter une part de variabilité (voir [4]). Ici nous étudions la masse de fuel nécessaire pour effectuer une liaison fixée avec un avion commercial, en ne considérant que la phase de *Croisière*. Le but est, à partir de données de masses de fuel consommées en croisière, d'identifier de manière robuste la consommation spécifique *SFC* de la motorisation en tenant compte de l'incertitude sur la vitesse de croisière *V* et sur la *finesse F* de l'avion. Nous proposons une procédure d'estimation basée sur une méthode de maximum de vraisemblance, prenant en compte cette incertitude.

[1] Institut de Mathématiques de Toulouse - EADS Innovation Works, 12 rue Pasteur, 92152 Suresnes
nabil.rachdi@eads.net
[2] MAP5, Université Paris Descartes SPC, 45 rue des saints pères, 75006 Paris
jean-claude.fort@parisdescartes.fr
[3] Institut de Mathématiques de Toulouse, 118 route de Narbonne F-31062 Toulouse
thierry.klein@math.univ-toulouse.fr

Contents

1. Introduction

One of engineering activities is to model real phenomena. Once a model is built (physical principles, state equations, etc.), some parameters have to be identified and some variables of the model may present some intrinsic variability. Hence, the identification of parameters should implicitly take into account the uncertainty of variables of the model. In this paper, we present a likelihood-based method to estimate aeronautic parameters in a *Fuel mass* model. We use an analytical model that can be viewed as a *black-box simulator*. From a data base $M_{fuel}^{*,1}, ..., M_{fuel}^{*,n}$ giving the mass of fuel consumed for n lines between two given cities with a specific commercial aircraft, we aim at identifying the *Specific Fuel Consumption* (SFC) which corresponds to a characteristic value of engines. The model we use depends in particular on the *cruise speed* (V) and on the *lift-to-drag ratio* (F). These variables present intrinsic variability: the cruise speed may depend on atmospheric conditions and the lift-to-drag ratio is also subject to variability potentially caused by turbulent phenomena. As a matter of fact, the identification of the parameter SFC should take into account the variability of the cruise speed and the lift-to-drag ratio.

In this paper, we propose an algorithm taken from the work of N. Rachdi *et al.* [7]. It allows a characterization of SFC from the observed data $M_{fuel}^{*,1}, ..., M_{fuel}^{*,n}$ and model simulations when the number of observations n is small.

This article is organized as follows. In Section 2 we describe the setting of the problem. In Section 3 we build the algorithm for the inverse problem with a Maximum-Likelihood based method. In Section 4 we apply the algorithm given in Section 3. In Section 5 we illustrate the effect of modeling

conditions, particularly the random modeling on SFC and the number of observed data. In Section 6 we establish the Theorem 6.2 providing an upper bound of the estimation error of the proposed algorithm. Section 7 is devoted to proving the Theorem 6.2.

We are grateful to Henri Yesilcimen for its contribution to the data generation and for fruitful discussions.

2. General setting

2.1. Observations

In our study, the data $M_{fuel}^{*,1}, ..., M_{fuel}^{*,n}$ were generated from an aeronautic software which simulates gas turbine configurations used for power generation. In particular, it can simulate the consumed mass of fuel at some configuration of engines, altitude, speed, atmospheric conditions, etc. (See Figure 1). This software is very complex and very much time consuming. In fact only 200 outputs from this software are available by choosing various atmospheric conditions. This sample constitutes our data reference.
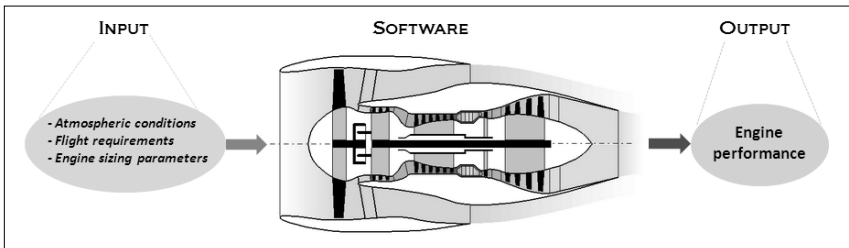
Figure 1. — Aeronautic software.

In a first time, we pick up a small sample of size $n = 32$ from this reference sample. The data are given in Table 1.

Table 1. — Simulated mass of fuel consumptions from aeronautic software.

Reference Fuel Masses [kg]							
7918	7671	7719	7839	7912	7963	7693	7815
7872	7679	8013	7935	7794	8045	7671	7985
7755	7658	7684	7658	7690	7700	7876	7769
8058	7710	7746	7698	7666	7749	7764	7667

Next, we will suppose that the observations $M_{fuel}^{*,1}, ..., M_{fuel}^{*,n}$ are drawn from an unknown probability distribution Q with associated Lebesgue density f with support

$$\mathcal{I} := [M_{inf} = 7600, M_{sup} = 8100].$$

The difference $M_{sup} - M_{inf} = 500\,kg$ have to be thought as an overconsumption of about 7%.

2.2. A simplified aeronautical model

We recall that we are interested in identifying the specific fuel consumption SFC. It is a significant factor determining the fuel efficiency of a particular engine. To handle this problem we introduce a classical simplified *Fuel mass* model given by the Bréguet formula:

$$M_{fuel} = (M_{empty} + M_{pload}) \left(e^{\frac{SFC \cdot g \cdot Ra}{V \cdot F} 10^{-3}} - 1 \right). \qquad (2.1)$$

The fixed variables are

- M_{empty} : *Empty weight* = basic weight of the aircraft (excluding fuel and passengers),
- M_{pload} : *Payload* = maximal carrying capacity of the aircraft,
- g : Gravitational constant,
- Ra : *Range* = distance traveled by the aircraft.

The uncertain variables mentioned in the introduction are

- V : *Cruise speed* = aircraft speed between ascent and descent phase,
- F : *Lift-to-drag ratio* = aerodynamic coefficient.

Table 2 gives the fixed variables values and the nominal values considered for random variables.

2.3. Noise modeling

As said in the introduction, we have to take into account the uncertainty of the cruise speed V and the lift-to-drag ratio F. Given the nominal value of each variable (see Table 2), an expert judgment can derive the uncertainty bounds (see Table 3).

Table 2. — Values of Fuel mass model inputs.

input	value or nominal value	unit
M_{empty}	42600	kg
M_{pload}	19900	kg
g	9.8	m/s^2
Ra	3000	km
V_{nom}	231	m/s
F_{nom}	19	–

Table 3. — Minimal and maximal values of uncertain variables.

variable	nominal value	min	max
V	**231**	226	234
F	**19**	18.7	19.05

The uncertainty on the cruise speed V represents a relative difference of arrival time of 8 minutes.

Moreover, specialists in turbine engineering propose to model the uncertainties as presented in Table 4.

Table 4. — Uncertainty modeling.

variable	distribution	parameter
V	*Uniform*	(V_{min}, V_{max})
F	*Beta*	$(7, 2, F_{min}, F_{max})$

The probability density function of a beta distribution on $[a, b]$ with shape parameters (α, β) is

$$g_{(\alpha,\beta,a,b)}(x) = \frac{(x-a)^{(\alpha-1)}(b-x)^{\beta-1}}{(b-a)^{\beta-1}B(\alpha,\beta)} 1\!\!1_{[a,b]}(x) ,$$

where $B(\cdot, \cdot)$ is the beta function.

Figure 2 shows the probability density functions of V and F.

In order to emphasize the "noisy" feature of the variables V and F, we will use the writing

- $V = V_{nom} + \epsilon_V$,
- $F = F_{nom} + \epsilon_F$,

where ϵ_V is a centered uniform random variable on the interval $[\epsilon_V^{min}, \epsilon_V^{max}]$ with

$$\epsilon_V^{min} = V_{min} - V_{nom} \quad and \quad \epsilon_V^{max} = V_{max} - V_{nom}\,.$$

Variable ϵ_F, supposed to be independent of ϵ_V, is a beta random variable on the interval $[\epsilon_F^{min}, \epsilon_F^{max}]$ with shape parameters $(7, 2)$ where

$$\epsilon_F^{min} = F_{min} - F_{nom} \quad and \quad \epsilon_F^{max} = F_{max} - F_{nom}\,.$$
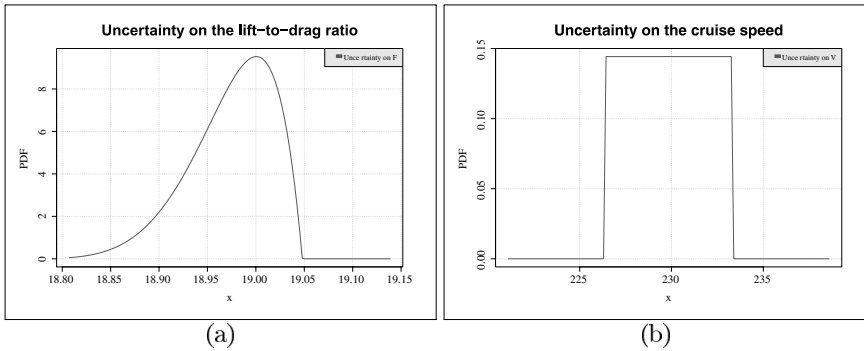
Figure 2. — (a) Uncertainty on F. (b) Uncertainty on V.

2.4. Robust identification of SFC

In our developments, we will not consider the parameter SFC to be deterministic but it will be supposed random. We indeed do not only need to compute SFC. We also want to take into account its own variability in order to have a robust characterization of this parameter.

ASSUMPTION 2.1. — Let us assume that the random variable SFC is **compactly** supported.

This assumption will allow to apply the theoretical framework (presented in annex) which needs the noise to be compactly supported. Such assumption does not impact the numerical results.

As a first approach, let us assume that

$$SFC = \mu_{SFC} + \sigma_{SFC}\,\epsilon_{SFC}\,, \quad \epsilon_{SFC} \sim \mathcal{N}_\mathcal{T}(0,1)\,. \tag{2.2}$$

with unknown parameters μ_{SFC} and σ_{SFC}.

We give the following ranges of variation

$$\mu_{SFC} \in [15, 20] \quad \text{and} \quad \sigma_{SFC} \in [s, 1] \,,$$

for a small $s > 0$.

The distribution $\mathcal{N}_{\mathcal{T}}(0, 1)$ of ϵ_{SFC} is a symmetric truncated standard Gaussian on the interval $[-3, 3]$.

Now, our problem amounts to estimating the location parameter μ_{SFC} and the standard deviation σ_{SFC}.

Remark 2.2. — All the numerical results we will present have been generated without truncation, which will not have a significant effect. Indeed, the truncations are set to high quantiles (more than 99% for upper quantiles and less than 1% for lower quantiles).

2.5. Statistical modeling

Let us denote by $(\mathcal{E}, \mathbb{P}^\epsilon)$ the probability space associated to the *noise vector*

$$\boldsymbol{\epsilon} = (\epsilon_{SFC}, \epsilon_V, \epsilon_F)^T \,,$$

and denote the vector of parameters by

$$\boldsymbol{\theta} = (\mu_{SFC}, \sigma_{SFC})^T \,.$$

Then, we consider the analytical and simplified model of M_{fuel} the mass of consumed fuel, as the function

$$M_{fuel} = h(\boldsymbol{\epsilon}, \boldsymbol{\theta}) \,,$$

where $h : (\mathcal{E}, \mathbb{P}^\epsilon) \times \Theta \to \mathcal{I}_h$ is given by

$$h(\boldsymbol{\epsilon}, \boldsymbol{\theta}) = (M_{empty} + M_{pload})$$

$$\left(exp \left(\frac{(\mu_{SFC} + \sigma_{SFC} \, \epsilon_{SFC}) \cdot g \cdot Ra}{(V_{nom} + \epsilon_V) \cdot (F_{nom} + \epsilon_F)} \cdot 10^{-3} \right) - 1 \right) \quad (2.3)$$

with

$$\Theta = [15, 20] \times [s, 1]$$

and $\mathcal{I}_h$ is the interval

$$\mathcal{I}_h = h(\mathcal{E}, \Theta) = [M_{inf}^h, M_{sup}^h] \,. \quad (2.4)$$

We denote by $|\mathcal{I}_h|$ its length.

Remark 2.3. — We observe through simulations that

$$\mathcal{I} \subset \mathcal{I}_h \, ,$$

where $\mathcal{I} = [M_{inf}, M_{sup}]$ is the observation interval given above.

The purpose is now to estimate parameter $\boldsymbol{\theta} \in \Theta$ from the set of data $M_{fuel}^{*,1}, ..., M_{fuel}^{*,n}$. In the next section, we propose an estimation procedure taken from [7].

3. Parameter estimation

In our previous framework it is possible to apply the procedures developed in [7]. In particular, we choose to work with the $\log -$contrast which can be understood as a Maximum Likelihood based estimation.

The sample $M_{fuel}^{*,1}, ..., M_{fuel}^{*,n}$ is drawn from an unknown distribution Q. We will use the parametric family of distributions $\{Q_{\boldsymbol{\theta}} \, , \, \boldsymbol{\theta} \in \Theta\}$ where $Q_{\boldsymbol{\theta}}$ is the *pushforward measure* of $\mathbb{P}^{\boldsymbol{\epsilon}}$ by the (measurable) application $\mathbf{u} \mapsto h(\mathbf{u}, \boldsymbol{\theta})$. That means that we consider the models $h(\epsilon, \theta)$ to be a reasonable first approximation in order to obtain statistical information about Q.
Denoting $\rho_{\boldsymbol{\theta}}$ the Lebesgue density associated to the measure $Q_{\boldsymbol{\theta}}$, the maximum likelihood procedure is given by

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta}{\mathrm{Argmin}} \, -\frac{1}{n} \sum_{i=1}^{n} \log(\rho_{\boldsymbol{\theta}}(M_{fuel}^{*,i})) \, . \tag{3.1}$$

However, the above procedure is unfeasible because the density $\rho_{\boldsymbol{\theta}}$ is not analytically tractable. As suggested in [7], we replace $\rho_{\boldsymbol{\theta}}$ by an estimator denoted $\rho_{\boldsymbol{\theta}}^m$. There are many ways to estimate a density. We choose the simplest one, which is commonly used in industrial modeling.

Let $\epsilon_1, ..., \epsilon_m$ be m random variables i.i.d from $\mathbb{P}^{\boldsymbol{\epsilon}}$, we consider the kernel estimate of our density

$$\rho_{\boldsymbol{\theta}}^m(\cdot) \quad = \quad \frac{1}{m} \sum_{j=1}^{m} K_{b_{\boldsymbol{\theta}}^m} \left(\cdot - h(\epsilon_j, \boldsymbol{\theta}) \right) \, , \tag{3.2}$$

where $K_{b_{\boldsymbol{\theta}}^m}$ is the Gaussian kernel

$$K_{b_{\boldsymbol{\theta}}^m}(x) = \frac{1}{\sqrt{2 \, \pi} \, b_{\boldsymbol{\theta}}^m} e^{-\frac{x^2}{2 \, (b_{\boldsymbol{\theta}}^m)^2}} \, ,$$

and $b_{\boldsymbol{\theta}}^m$ is computed from the sample $h(\boldsymbol{\epsilon}_j, \boldsymbol{\theta})$, $j = 1, ..., m$ for $\boldsymbol{\theta} \in \Theta$, by Silverman's rule-of-thumb :

$$b_{\boldsymbol{\theta}}^m = \left(\frac{4}{3} \right)^{1/5} m^{-1/5} \, \hat{\sigma}_{\boldsymbol{\theta}} \, . \tag{3.3}$$

The quantity $\hat{\sigma}_{\boldsymbol{\theta}}$ is the empirical standard deviation of the sample $h(\boldsymbol{\epsilon}_j, \boldsymbol{\theta})$, $j = 1, ..., m$

$$\hat{\sigma}_{\boldsymbol{\theta}} = \frac{1}{m} \sum_{j=1}^{m} \left(h(\boldsymbol{\epsilon}_j, \boldsymbol{\theta}) - \frac{1}{m} \sum_{j=1}^{m} h(\boldsymbol{\epsilon}_j, \boldsymbol{\theta}) \right)^2 \, .$$

Other popular estimates are truncated projections on a suitable basis of functions (sine, cosine, wavelets, etc.). Here we do not discuss the optimization of the density estimation, but all what follows could be applied in the same way. The numerical results may be slightly different, but qualitatively the same.

Replacing $\rho_{\boldsymbol{\theta}}$ by $\rho_{\boldsymbol{\theta}}^m$ in (3.1) and simplifying by the multiplying constant $1/n$ yields the estimation procedure

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta}{\text{Argmin}} - \sum_{i=1}^{n} \log \left(\frac{1}{m} \sum_{j=1}^{m} K_{b_{\boldsymbol{\theta}}^m} \left(h(\boldsymbol{\epsilon}_j, \boldsymbol{\theta}) - M_{fuel}^{*,i} \right) \right) \, . \tag{3.4}$$

Hence, our problem is an inverse problem. More precisely, it is an inverse problem in presence of uncertainties, also called *probabilistic inverse problem* or *stochastic inverse problem*. This topic is often treated in the field of *uncertainty management*: the goal may for instance be to identify the intrinsic uncertainty of a system, see the PhD works [1] and [5]. Another reference is the paper of E. de Rocquigny and S. Cambier [3], where the purpose is to identify a parameter of interest which controls the vibration amplification of stream turbines. Our framework is different. The main difference lies in the absence of assumptions, in the present paper, on the distribution of the error between observation data and reference data. Thus, it differentiates the estimation procedures we propose from the ones developed in [3].

In the following section, we provide a numerical analysis using the algorithm given by (3.4). Theoretical aspects will be addressed in Section 6.

4. Numerical study : first approach

4.1. Estimation

Setting

$$J(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \log\left(\frac{1}{m}\sum_{j=1}^{m} K_{b_{\boldsymbol{\theta}}^m}\left(h(\boldsymbol{\epsilon}_j, \boldsymbol{\theta}) - M_{fuel}^{*,i}\right)\right) \text{ with } \boldsymbol{\theta} = (\mu_{SFC}, \sigma_{SFC})^T,$$

our problem is a minimization problem where we want to compute

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\mathrm{Argmin}}\, J(\boldsymbol{\theta})\,.$$

We recall that $n = 32$. The data $(M_{fuel}^{*,i})_{i=1,...,n}$ are provided by Table 1. We choose $m = 10000$ (the number of calls to the model given by (2.3)), and for $j = 1, ..., m$, $\boldsymbol{\epsilon}_j \sim \mathbb{P}^{\boldsymbol{\epsilon}}$ where

$$\mathbb{P}^{\boldsymbol{\epsilon}}(du, dv, dw) =$$

$$\frac{1}{\sqrt{2\pi}\,L}e^{-u^2/2}\, g_{(7,2,\epsilon_F^{min},\epsilon_F^{max})}(v)\,\frac{1}{\epsilon_V^{max} - \epsilon_V^{min}}\mathbb{1}_{[\epsilon_V^{min},\epsilon_V^{max}]}(w)\, du\, dv\, dw\,,$$

with $L = \Phi(3) - \Phi(-3)$ (Φ is the cumulative distribution function of a standard Gaussian random variable).

This optimization procedure can be solved by Quasi-Newton methods. We present the results in Table 5. In Figure 3 we show the resulting probability density function of SFC given by $\mathcal{N}_{\mathcal{T}}(\widehat{\mu}_{SFC}, \widehat{\sigma}_{SFC})$.

Table 5. — SFC characterization parameters.

estimator	value of $J(\widehat{\boldsymbol{\theta}})$	estimated SFC location	estimated SFC dispersion
$\widehat{\boldsymbol{\theta}}$	$J(\widehat{\boldsymbol{\theta}}) = 199.465$	$\widehat{\boldsymbol{\mu}}_{\mathbf{SFC}} = \mathbf{17.397}$	$\widehat{\boldsymbol{\sigma}}_{\mathbf{SFC}} = \mathbf{0.201}$

Figure 4 provides profile views of the criterion function $\boldsymbol{\theta} = (\mu_{SFC}, \sigma_{SFC}) \mapsto J(\boldsymbol{\theta})$, first at $\sigma_{SFC} = \widehat{\sigma}_{SFC}$ (Figure 4(a), we show $\log(J(\boldsymbol{\theta}))$) and then at $\mu_{SFC} = \widehat{\mu}_{SFC}$ (Figure 4(b)).

We notice that the minimum $\widehat{\boldsymbol{\theta}} = (\widehat{\mu}_{SFC}, \widehat{\sigma}_{SFC})$ is correctly located.
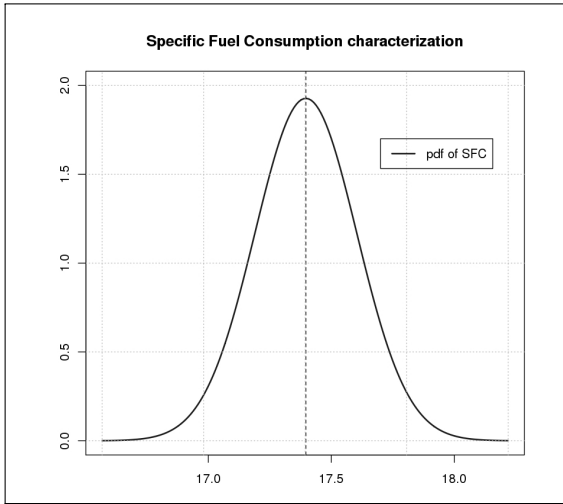
Figure 3. — Estimated Specific Fuel Consumption distribution.
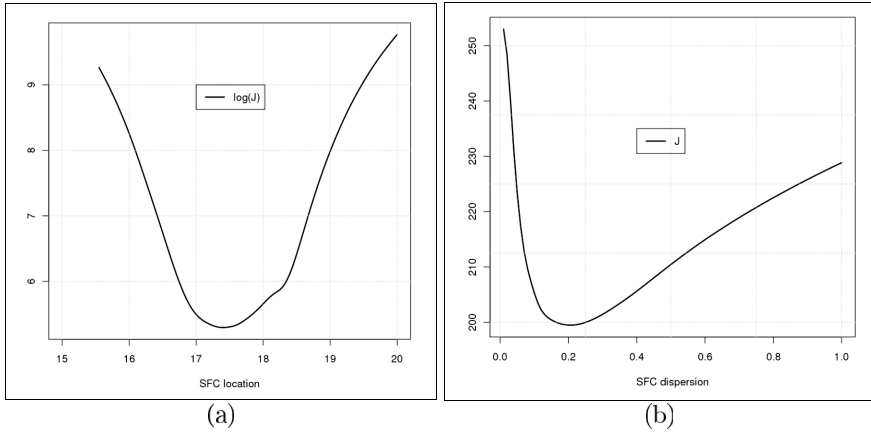
Figure 4. — (a) Profile view of $\log(J)$ at $\sigma_{SFC} = \widehat{\sigma}_{SFC}$.
(b) Profile view of J at $\mu_{SFC} = \widehat{\mu}_{SFC}$.

4.2. Comparison with reference sample

In order to analyse the results obtained in the previous subsection, we need some reference sample of SFC values at the same simulation conditions. We take for reference the sample of SFC values of size 200 described in the introduction, provided by the aeronautic software. The characteristics of this sample are given in Table 6.

Table 6. — Reference sample characteristics.

	Mean	Stand. dev.
Reference sample	**17.49**	**0.57**

The data in Table 6 have to be compared with those in Table 5 where the mean and standard deviation are $\widehat{\mu}_{SFC} = 17.397$ and $\widehat{\sigma}_{SFC} = 0.201$, respectively. Table 7 provides the associated relative errors. Figure 5 shows the histogram of the reference sample and the estimated distribution of SFC obtained in Figure 3.

Table 7. — Relative errors of the mean and deviation between reference SFC sample and the estimated model (3).

	Reference sample	Estimated SFC (3)	Relative error
Mean	**17.49**	**17.397**	**0.5 %**
Stand. dev.	**0.57**	**0.201**	**60.6 %**

It appears that the location of the variable of interest SFC is well reached whereas the standard deviation estimation provides an error of 60%. The "error" has roughly two origins:

Statistical error: this error is mainly due to the limited number ($n = 32$) of data from the observed masses of fuel $M_{fuel}^{*,i}$. It is also due to the error induced by the kernel approximation of ρ_θ. Yet the choice of $m = 10000$ calls to the analytical model garantees that the error on ρ_θ is small .

Model error: this error is relative to the use of Fuel mass model (2.1) with uncertain variables V and F (Figure 2), and includes the Gaussian hypothesis for SFC (2.2). Thus, the model error can be separated into 2 parts: *physical* model error and *uncertainty modeling* error.

We observe on Figure 5 that the SFC parameter does not behave like a Gaussian variable. This can be qualified as *model error*.

However, if one just wants to estimate the mean value of SFC, the Gaussian hypothesis does not have a significative impact (0.5% of error). On the other hand, if one wants more information about SFC, other modeling tools are needed to allow a robust characterization approach.

In the next subsection, we will discuss the *uncertainty modeling*, more precisely, the Gaussian hypothesis for SFC given by (2.2).
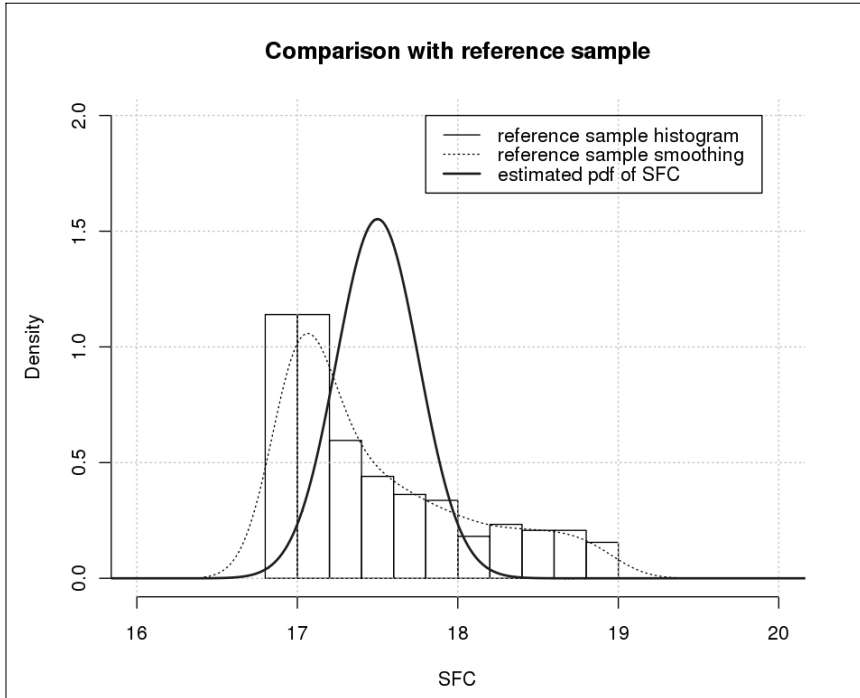
Figure 5. — Reference and estimated SFC distributions.

5. On the probabilistic modeling of SFC

5.1. Considering Wiener-Hermite representation in the previous analysis

The characterization of a random variable by the mean and the standard deviation only could be too approximative in order to study the whole behavior of the variable. In this study, we have made an *a priori* (a model) on the variable of interest SFC. In (2.2) we supposed that

$$SFC \sim \mathcal{N}_{\mathcal{T}}(\mu_{SFC}, \sigma^2_{SFC}),$$

which we now rewrite

$$SFC = \mu_{SFC} + \sigma_{SFC}\, \xi_{\mathcal{T}}, \quad \xi_{\mathcal{T}} \sim \mathcal{N}_{\mathcal{T}}(0,1). \tag{5.1}$$

We will see that this (truncated) Gaussian hypothesis on SFC is a particular case of a more general representation.

The so called *Wiener Chaos Expansion*, developed in the 30's by Wiener
[11], gives a representation of any second-order random variable Z:

$$Z = \sum_{l=0}^{\infty} z_l \Upsilon_l((\xi_k)_{k \geqslant 1}), \quad \text{(with convergence in } L_2(\mathbb{P}) \text{)} \tag{5.2}$$

where $(\xi_k)_{k \geqslant 1}$ is a (infinite) sequence of independent standard normal ran-
dom variables and the Υ_l's are the multivariate Hermite polynomials. This
expansion is also called *Wiener-Hermite expansion*.

In practice, we have to consider a finite sequence $(\xi_1, ..., \xi_M)$ where M
is called the *order* of the expansion, and the sum in (5.2) is truncated at p
which is the *degree* of the expansion.

Hence, considering all M-dimensional Hermite polynomials of degree
lower than p, the representation (5.2) is approximated by

$$Z \simeq Z^{p,M} = \sum_{l=0}^{P-1} z_l \Upsilon_l(\boldsymbol{\xi}), \quad \boldsymbol{\xi} = (\xi_1, ..., \xi_M), \tag{5.3}$$

where

$$P = \frac{(M+p)!}{M!\, p!}.$$

The integer P corresponds to the number of coefficients to be estimated.
For our purpose, by Assumption 2.1, we will consider the following approx-
imation

$$Z \simeq Z^{p,M} = \sum_{l=0}^{P-1} z_l \Upsilon_l(\boldsymbol{\xi}_{\mathcal{T}}), \quad \boldsymbol{\xi}_{\mathcal{T}} = (\xi_{\mathcal{T}}^1, ..., \xi_{\mathcal{T}}^M), \tag{5.4}$$

where $\boldsymbol{\xi}_{\mathcal{T}}$ is a vector of truncated standard Gaussian variables.

Moreover, one can notice that by orthogonality arguments in (5.3), we
have without truncation

$$\mathbb{E}(Z^{p,M}) = z_0 = \mathbb{E}(Z) \tag{5.5}$$

and

$$\text{Var}(Z^{p,M}) = \sum_{l=1}^{P-1} z_l^2. \tag{5.6}$$

Let us notice that by the decomposition

$$Z^{p,M} = Z + \left(Z^{p,M} - Z\right),$$

each choice of p and M will induce a *model error*

$$mod_{err} := Z^{p,M} - Z.$$

We illustrate this aspect concerning SFC in the next subsection.

5.2. Application to the Specific Fuel Consumption

In our purpose, if we suppose that $\mathbb{E}(SFC^2) < \infty$ (it is implicitly supposed in the Gaussian hypothesis), we can set the following modeling

$$SFC^{p,M} = \sum_{l=0}^{P-1} z_l \Upsilon_l(\boldsymbol{\xi}_{\mathcal{T}}), \quad \boldsymbol{\xi}_{\mathcal{T}} = (\xi_{\mathcal{T}}^1, ..., \xi_{\mathcal{T}}^M), \quad M, p \geqslant 1$$

which we rewrite by (5.5)

$$SFC^{p,M} = \mu_{SFC} + \sum_{l=1}^{P-1} z_l \Upsilon_l(\boldsymbol{\xi}_{\mathcal{T}}), \quad \boldsymbol{\xi}_{\mathcal{T}} = (\xi_{\mathcal{T}}^1, ..., \xi_{\mathcal{T}}^M), \quad M, p \geqslant 1. \quad (5.7)$$

It appears now that the Gaussian representation (5.1) is the particular case of (5.7) with $p = 1$ and $M = 1$. Moreover, in view of the Wiener representation (5.2), the Gaussian one (5.1) may lead to a rough approximation (if SFC is not Gaussian) and thus contributes to a non negligible model error. It is clearly observed in Figure 5 where the reference data does not seem to be drawn from a Gaussian distribution.

As a matter of fact, one can hope to reduce the model error (described in the previous subsection), at least the error corresponding to SFC modeling, by considering a less restrictive representation (5.7) with some appropriate order $M \geqslant 1$ and degree $p \geqslant 1$.

Let us consider the Wiener-Hermite expansion of order $M = 2$ and degree $p = 2$

$$SFC^{2,2} = \mu_{SFC} + \sum_{l=1}^{5} \theta_l \Upsilon_l(\boldsymbol{\xi}_{\mathcal{T}}), \quad \boldsymbol{\xi}_{\mathcal{T}} = (\xi_{\mathcal{T}}^1, \xi_{\mathcal{T}}^2)$$

or

$$SFC^{2,2} = \mu_{SFC} + \theta_1 \xi_{\mathcal{T}}^1 + \theta_2 \xi_{\mathcal{T}}^2 + \theta_3 \xi_{\mathcal{T}}^1 \xi_{\mathcal{T}}^2 + \theta_4 ((\xi_{\mathcal{T}}^1)^2 - 1) + \theta_5 ((\xi_{\mathcal{T}}^2)^2 - 1), \quad (5.8)$$

that leads to estimate $P = \frac{(2+2)!}{2!2!} = 6$ coefficients. Table 8 shows the result obtained by the algorithm developed in the previous section where we change the function $h(\boldsymbol{\epsilon}, \boldsymbol{\theta})$ in (2.3) replacing $\sigma_{SFC}\epsilon_{SFC}$ by $\theta_1 \xi_{\mathcal{T}}^1 + \theta_2 \xi_{\mathcal{T}}^2 + \theta_3 \xi_{\mathcal{T}}^1 \xi_{\mathcal{T}}^2 + \theta_4 ((\xi_{\mathcal{T}}^1)^2 - 1) + \theta_5 ((\xi_{\mathcal{T}}^2)^2 - 1)$, with $\boldsymbol{\epsilon} = (\xi_{\mathcal{T}}^1, \xi_{\mathcal{T}}^2, \epsilon_V, \epsilon_F)$ and $\boldsymbol{\theta} = (\mu_{SFC}, \theta_1, ..., \theta_5)$.

Table 8. — SFC characterization parameters.

	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5
$SFC^{2,2}$	17.470	0.047	0.054	0.182	0.103	0.063

Table 9. — Relative errors of the mean and standard deviation with $SFC^{2,2}$.

	Reference sample	from $SFC^{2,2}$	Relative error
Mean	17.49	17.470	0.11 %
Stand. dev.	0.57	0.230	59.65 %

Let us compare the relative errors of the first two statistical moments by considering $SFC^{1,1}$ (i.e the Gaussian hypothesis Table 7) and $SFC^{2,2}$ (Table 9).

The Wiener-Hermite modeling seems to improve the mean estimation of SFC whereas the standard deviation is poorly estimated in the two cases with an error of about 60%. There is no significative difference between the two methods regarding the first two moments. However, the behavior of density functions corresponding to $SFC^{1,1}$ (see Figure 6) and $SFC^{2,2}$ is clearly not the same. We present in Figure 6 the result obtained when SFC is modeled by a Wiener expansion of order $M = 2$ and degree $p = 2$.

The distribution of SFC given by the Wiener expansion in Figure 6 seems to have a behavior close to the reference sample one, despite the fact that there is a non negligible bias. As mentioned in the previous subsection, this is due to the statistical and model errors. Indeed, let us recall that we have at disposal $n = 32$ reference fuel masses from which we characterize the SFC parameter. It would be interesting to see what happens when adding reference fuel masses, i.e by reducing the statistical error.

5.3. Wiener-Hermite analysis with augmented reference fuel mass sample

We present here numerical results obtained by adding 50 new samples from the data basis built with the complex software with the same initial conditions. Figure 7 shows the characterization of SFC obtained by a Wiener expansion of order $M = 2$ and degree $p = 2$ from the augmented reference fuel mass sample of size $n = 82$.

Hence, by adding reference data we improve significatively the characterization of SFC on the first two statistical moments as well as on the whole probability density function of SFC.
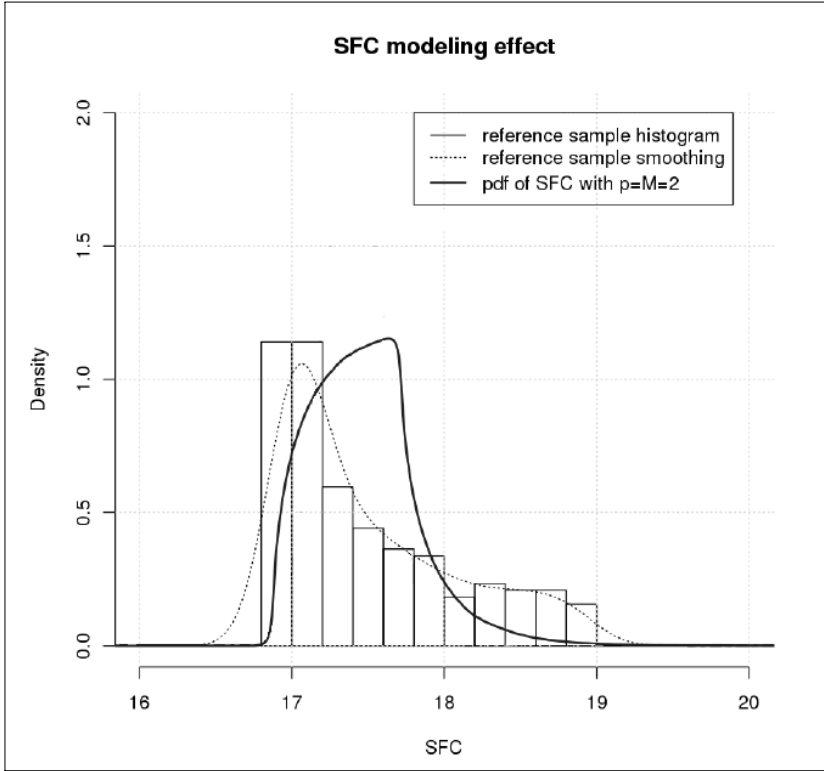
Figure 6. — Estimations of SFC probability density with a Wiener Expansion $p = M = 2$.

The Table 10 gives the coefficients corresponding to this simulation.

Table 10. — SFC characterization parameters from augmented fuel mass sample.

	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5
$SFC^{2,2}$	17.50	0.281	0.008	0.012	0.191	0.219

Table 11. — Relative errors of the mean and standard deviation with $SFC^{2,2}$ from augmented fuel mass sample.

	Reference sample	from $SFC^{2,2}$	Relative error
Mean	17.49	17.50	0.06 %
Stand. dev.	0.57	0.404	29.12 %

In the next section we introduce some knowledge on the SFC modeling through an expert judgment inducing a new statistical modeling that will prove to be better.
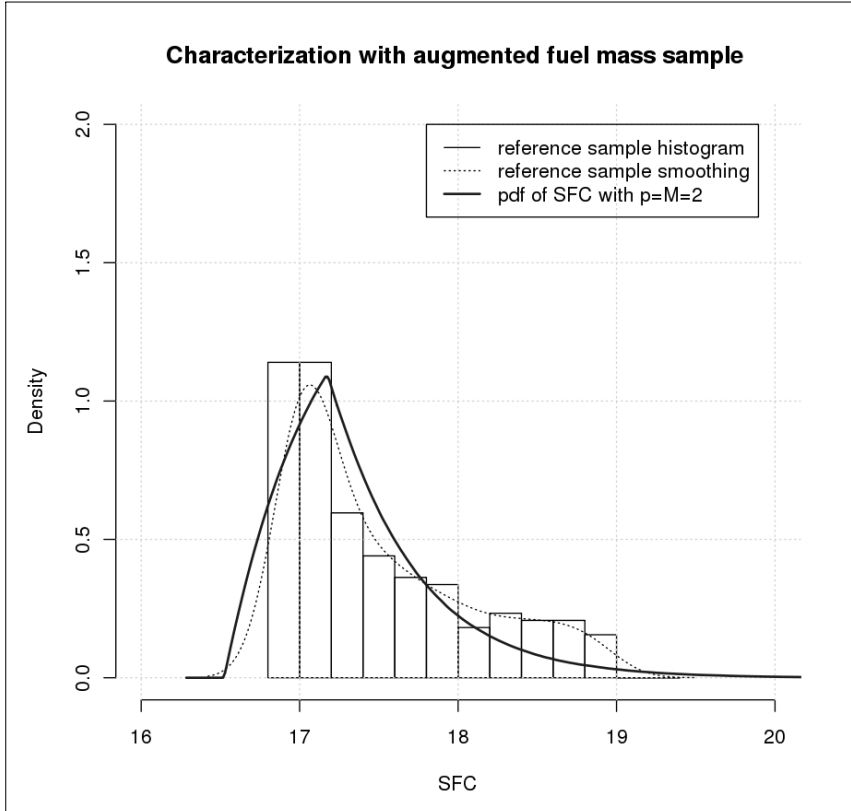
Figure 7. — Characterization of SFC with an augmented sample of fuel mass.

5.4. Analysis with a "good" a priori knowledge

In the previous analyses, we only consider truncated Wiener-Hermite expansions. This is more of a mathematical hypothesis than a knowledge brought to the modeling. Suppose now that an expert judgment says that the distribution of the SFC is of exponential form. Mathematically, it is equivalent to supposing that the probability density of SFC belongs to the family

$$\left\{ p(u; \boldsymbol{\theta}) = \theta_2 \, e^{-\theta_2(u-\theta_1)} \, \mathbb{1}_{[\theta_1, +\infty[} \, , \quad \boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}_+ \times \mathbb{R}_+^* \right\} .$$

One can check that this suggestion induces the modeling

$$SFC^{exp} = \theta_1 - \frac{1}{\theta_2} \log(\xi), \quad \xi \sim \mathcal{U}([0,1]), \qquad (5.9)$$

where $\mathcal{U}([0,1])$ is the uniform distribution on the interval $[0,1]$.
In order to satisfy Assumption 2.1, we consider the truncated version

$$SFC^{exp} = \theta_1 - \frac{1}{\theta_2} \log(\xi_\mathcal{T}), \quad \xi_\mathcal{T} \sim \mathcal{U}([c,1]), \qquad (5.10)$$

for some small $c > 0$.

Remark 5.1. — Representation (5.10) seems quite different from the one provided by the Wiener-Hermite expansions (see (5.1) and (5.8)). Yet, as the random variable SFC^{exp} has finite variance, the modeling (5.10) could be seen as a practical alternative to a Wiener expansion (5.2). Such a Wiener expansion would be given by choosing a high order M^{exp} and a high degree p^{exp} in (5.4).

In what follows, we present the results of the numerical analysis corresponding to $n = 32$ and $n = 82$.

Table 12. — Estimation of $\boldsymbol{\theta} = (\theta_1, \theta_2)$ when $n = 32$.

	θ_1	θ_2
SFC^{exp}	**17.23**	**3.45**

Table 13. — Relative errors of the mean and standard deviation between reference SFC sample and SFC^{exp} when $n = 32$.

	Reference sample	from SFC^{exp} $(n = 32)$	Relative error
Mean	**17.49**	**17.52**	**0.17 %**
Stand. dev.	**0.57**	**0.29**	**49.12 %**

Table 14. — Estimation of $\boldsymbol{\theta} = (\theta_1, \theta_2)$ when $n = 82$.

	θ_1	θ_2
SFC^{exp}	**16.95**	**2**

Table 15. — Relative errors of the mean and standard deviation between reference SFC sample and SFC^{exp} when $n = 82$.

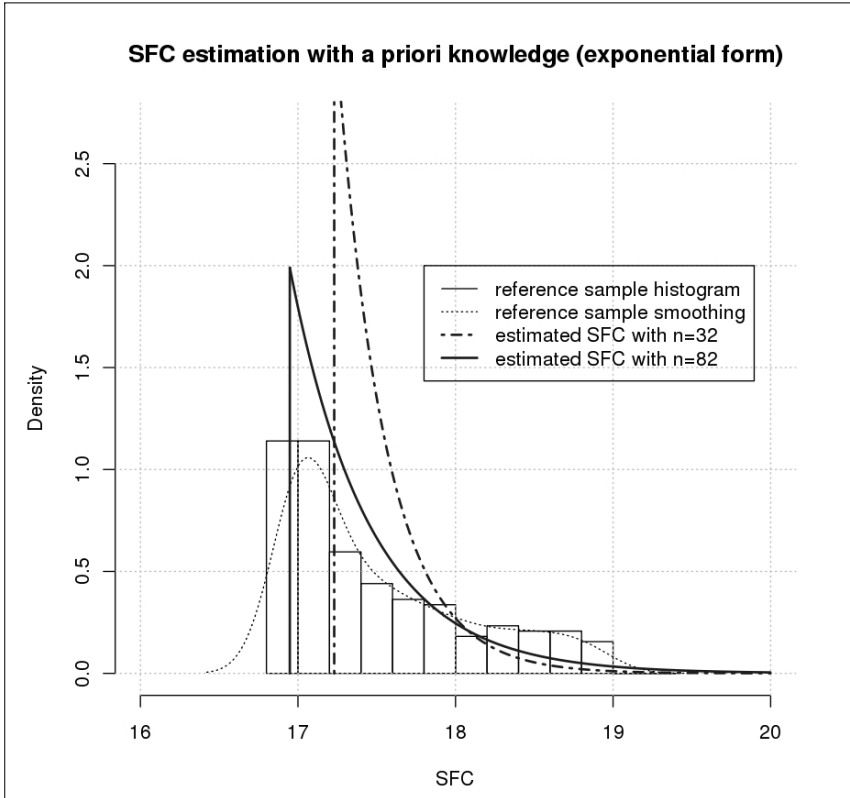|	Reference sample	from SFC^{exp} $(n = 82)$	Relative error
Mean	**17.49**	**17.45**	**0.23 %**
Stand. dev.	**0.57**	**0.501**	**12.1 %**

Figure 8. — Characterization of SFC with an exponential hypothesis.

We clearly see that the informative knowledge contributes to improving significatively the characterization of the Specific Fuel Consumption. With $n = 82$ fuel mass data, the results are satisfying as shown in Figure 8 and Table 15.

5.5. Conclusion

Section 5 was dedicated to illustrating the effect of the modeling conditions for SFC characterization. In particular, we showed the impact of a "model error" through the modeling of the random variable SFC. We also illustrated how the statistical error, through the number of fuel mass data, appears in the performance of the estimation.

In all cases, we computed a parameter $\widehat{\boldsymbol{\theta}}$. If we supposed that there is no model error, that is $Q \in \{Q_\theta, \theta \in \Theta\}$ (where Q is the "true" distribution of M_{fuel}^*), the error is only due to the limited number of data. So it makes sense to investigate the difference $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|$, where $\boldsymbol{\theta}^*$ is defined as $Q = Q_{\theta^*}$. If this is not the case it gives an insight on the statistical error part.

It is the topic of the following section.

6. Theoretical result

In this paper, the study of the procedure performance (3.4) will be non-asymptotic, i.e for a fixed number of observations $M_{fuel}^{*,i}$ (n) and a fixed number of variables ϵ_j (m). The asymptotic study is let to a forthcoming work.

The quality of such estimation procedure can be investigated by giving an upper bound of the distance between the *reachable parameter* $\widehat{\boldsymbol{\theta}}$ and the *best parameter* $\boldsymbol{\theta}^*$ (unknown). The latter can be seen as the parameter obtained if one has an infinite number of observations $M_{fuel}^{*,i}$ and variables ϵ_j. More precisely,

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{Argmin}} \, \mathbb{E}_Q \log \left(\rho_{\boldsymbol{\theta}}(M_{fuel}^*) \right) , \qquad (6.1)$$

where $\mathbb{E}_Q \log \left(\rho_{\boldsymbol{\theta}}(M_{fuel}^*) \right)$ can be seen as the "limit" of the quantity

$$\frac{1}{n} \sum_{i=1}^{n} \log \left(\frac{1}{m} \sum_{j=1}^{m} K_{b_{\boldsymbol{w}}^m} \left(h(\boldsymbol{\epsilon}_j, \boldsymbol{\theta}) - M_{fuel}^{*,i} \right) \right) \qquad (6.2)$$

in (3.4) when n and m go to infinity.

The Maximum Likelihood equation (6.1) turns out to be the minimization of the Kullback-Leibler divergence between Q and the family $\{Q_\theta, \theta \in \Theta\}$, while equation (6.2) is a smoothed empirical counterpart.

We consider the model $h(\boldsymbol{\epsilon}, \boldsymbol{\theta})$ given in (2.3), but what follows can be generalized to any other one.

Then, denoting by $\|\cdot\|$ the Euclidian norm in $\mathbb{R}^2$, it makes sense to bound the quantity

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 .$$

Let us denote by

$$\mathcal{R}(\boldsymbol{\theta}) := \mathbb{E}_Q \log \left(\rho_{\boldsymbol{\theta}}(M^*_{fuel}) \right) ,$$

and by f the Lebesgue density associated to the measure Q.

ASSUMPTION 6.1. — *Let us consider the following assumptions.*

– **A1** *The map* $\boldsymbol{\theta} \mapsto \mathcal{R}(\boldsymbol{\theta})$ *is twice differentiable with*

$$\nabla \mathcal{R}(\boldsymbol{\theta}^*) = 0$$

and has a symmetric positive definite Hessian matrix $\nabla^2 \mathcal{R}$. *Let us denote by* $\lambda_{min} > 0$ *the smallest eigenvalue of the set of matrices* $\{\nabla^2 \mathcal{R}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$.

– **A2** *It exists* $\eta > 0$ *such that for all* $\boldsymbol{\theta} \in \Theta$, *the density probability of* $h(\boldsymbol{\epsilon}, \boldsymbol{\theta})$ *we noted* $\rho_{\boldsymbol{\theta}}$, *satisfies*

$$\rho_{\boldsymbol{\theta}} > \eta .$$

– **A3** *For all* $\boldsymbol{\theta} \in \Theta$, *the second derivative of* $\rho_{\boldsymbol{\theta}}$, *we note* $\rho_{\boldsymbol{\theta}}^{''}$, *exists and*

$$C := \sup_{\boldsymbol{\theta} \in \Theta} \|\rho_{\boldsymbol{\theta}}^{''}\|_2 < +\infty .$$

– **A4** *We suppose that*

$$0 < \delta < \inf_{\boldsymbol{\theta} \in \Theta} \hat{\sigma}_{\boldsymbol{\theta}} \quad \text{and} \quad \sup_{\boldsymbol{\theta} \in \Theta} \hat{\sigma}_{\boldsymbol{\theta}} < \sigma < +\infty ,$$

where $\hat{\sigma}_{\boldsymbol{\theta}}$ *is defined in (3.3).*

We prove the following consistency theorem:

THEOREM 6.2. — *Let us consider the estimator* $\widehat{\boldsymbol{\theta}}$ *in (3.4) and the Assumptions (6.1).*

Then, for all $0 < \tau < 1/2$, *with probability at least* $1 - 2\tau$

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \leqslant c_1 \sqrt{\frac{\log(a_1 \tau^{-1})}{n}} + \frac{c_2 \sqrt{\log(a_2 \tau^{-1})} + c_3 \, m^{1/10}}{\sqrt{m}} ,$$

for some constants c_1, c_2, c_3, a_1 *and* a_2.

The risk bound of this theorem seems surprising since we obtain a rate of $n^{1/4}$, whereas one expects a rate close to $\sqrt{n}$ for the treated parametric problem. This theorem is a consistency result. Consequently, it does not give information about the rate of convergence. The obtained bound can be explained by the fact that, by Assumption A1, we have

$$\mathcal{R}(\widehat{\boldsymbol{\theta}}) - \mathcal{R}(\boldsymbol{\theta}^*) \approx \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \,.$$

Indeed, if $\mathcal{R}(\widehat{\boldsymbol{\theta}}) - \mathcal{R}(\boldsymbol{\theta}^*) \approx 1/\sqrt{n}$ (bound given in [7]) then obviously $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \approx n^{-1/4}$.

However, the $\sqrt{n}$-rate can be reached by considering the approach of Corollary 5.53 (p. 77) in [9] where an additional assumption is made on the risk function $\boldsymbol{\theta} \mapsto \mathcal{R}(\boldsymbol{\theta})$. More precisely, this assumption relies on the function $\boldsymbol{\theta} \mapsto \log(\rho_{\boldsymbol{\theta}})$ which is supposed to satisfy a Lipschitz condition. This work is let to a forthcoming paper which will deal with a central limit theorem for the parameter $\widehat{\boldsymbol{\theta}}$ (the rate of convergence will therefore be reachable).

7. Proof of Theorem 6.2

We give a general proof of Theorem 6.2.

By Assumption A1, we have the Taylor-Lagrange formula

$$\mathcal{R}(\widehat{\boldsymbol{\theta}}) = \mathcal{R}(\boldsymbol{\theta}^*) + \frac{1}{2} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T \nabla^2 \mathcal{R}(\xi) (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \,, \tag{7.1}$$

for some $\xi \in \Theta$.

Then, we will use the following lemma

LEMMA 7.1 (Rayleigh's quotient). — *Let H be a real symmetric matrix $p \times p$ and denote by $\lambda_1 < ... < \lambda_p$ the ordered eigenvalues of H.*

It holds that for all $\mathbf{x} \in \mathbb{R}^p - \{0\}$

$$\lambda_1 \leqslant \frac{\mathbf{x}^T H \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \leqslant \lambda_p \,.$$

Now, applying this lemma with $H = \nabla^2 \mathcal{R}(\xi)$ and $\mathbf{x} = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ yields

$$\lambda_{min} \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \leqslant (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T \nabla^2 \mathcal{R}(\xi) (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \,,$$

where $\lambda_{min} > 0$ is the smallest eigenvalue of the set of matrices $\{\nabla^2 \mathcal{R}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$.

Then, using this last inequality with equality (7.1) gives

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \leqslant \frac{2}{\lambda_{min}} \left(\mathcal{R}(\widehat{\boldsymbol{\theta}}) - \mathcal{R}(\boldsymbol{\theta}^*)\right) . \tag{7.2}$$

The problem turns to bound the positive quantity $\mathcal{R}(\widehat{\boldsymbol{\theta}}) - \mathcal{R}(\boldsymbol{\theta}^*)$, where $\widehat{\boldsymbol{\theta}}$ is given by (3.4). Such bound can be investigated by applying Theorem 4.1 in [7], which is a general result. We will aim at computing constants K_1^τ and K_2^τ such that, with high probability

$$\mathcal{R}(\widehat{\boldsymbol{\theta}}) - \mathcal{R}(\boldsymbol{\theta}^*) \leqslant \frac{2\|f\|_2}{\eta} \left(\frac{1}{\sqrt{n}} \frac{\eta}{2\delta^2 \|f\|_2} \gamma K_1^\tau + \frac{1}{\sqrt{m}} \frac{1}{\sqrt{2\pi\delta}} K_2^\tau + \frac{1}{m^{2/5}} \frac{C(1.06\sigma)^2}{\sqrt{3}}\right) . \tag{7.3}$$

In our framework, the main work is to compute the concentration constants K_1^τ and K_2^τ derived from [7] in the following particular case.

7.1. On concentration constants K_1^τ and K_2^τ

Since we apply Theorem 4.1 in [7], one has to guarantee that the constants K_1^τ and K_2^τ are finite.
For this, let us first recall some definitions and notations relative to empirical processes.

DEFINITION 7.2. — **Empirical process**. *Let W be some probability measure on some space T and let us suppose given a k i.i.d sample $\xi_1, ..., \xi_k$ drawn from W. Let us denote by W_k the empirical measure*

$$W_k := \frac{1}{k} \sum_{i=1}^{k} \delta_{\xi_i}$$

and $\mathcal{G}$ some class of real valued functions $g : T \to \mathbb{R}$.

We call W-empirical process indexed by $\mathcal{G}$ the following application

$$\begin{aligned} \mathbb{G}_k : \mathcal{G} &\longrightarrow \mathbb{R} \\ g &\longmapsto \mathbb{G}_k := \sqrt{k} \int_T g(t) \, (W_k - W) \, (dt), \end{aligned}$$

also written

$$\mathbb{G}_k\, g := \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \left(g(\xi_k) - \mathbb{E}_W\left(g(\xi)\right) \right).$$

We denote the supremum of an empirical process by

$$\|\mathbb{G}_k\|_{\mathcal{G}} := \sup_{g \in \mathcal{G}} |\mathbb{G}_k\, g| \,.$$

Following the proof lines of Theorem 2.1, Table 2 p.11 in [7] (giving classes of functions) and considering the inequality (7.3), it is easy to verify that K_1^τ is defined as

$$\text{for all } n \geqslant 1, \quad \mathbb{P}(\|\mathbb{U}_n\|_{\mathcal{A}} \leqslant K_1^\tau) \geqslant 1 - \tau \tag{7.4}$$

where $\mathbb{U}_n$ is the Q-empirical process $(Q_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{M_{fuel}^{*,i}})$ indexed by the class of functions

$$\mathcal{A} = \{y \in \mathcal{I} \longmapsto (y-\lambda)^2, \, \lambda \in \mathcal{I}_h\} \tag{7.5}$$

where we recall

$$\mathcal{I} = [M_{inf}, M_{sup}] \quad \text{and} \quad \mathcal{I}_h = h(\mathcal{E}, \Theta) \,.$$

Similarly, the constant K_2^τ is defined as follows

$$\text{for all } m \geqslant 1, \quad \mathbb{P}(\|\mathbb{V}_m\|_{\mathcal{B}} \leqslant K_2^\tau) \geqslant 1 - \tau \tag{7.6}$$

where $\mathbb{V}_m$ is the $P^{\boldsymbol{\epsilon}}$-empirical process $(P_m^{\boldsymbol{\epsilon}} = \frac{1}{m}\sum_{j=1}^{m} \delta_{\boldsymbol{\epsilon}_j})$ indexed by the class of functions

$$\mathcal{B} = \{\mathbf{x} \in \mathcal{E} \longmapsto e^{-(h(\mathbf{x},\boldsymbol{\theta})-\lambda)^2/2\,b^2}, \quad (\boldsymbol{\theta}, \lambda, b) \in \Theta \times \mathcal{I}_h \times [\delta, \sigma]\} \,. \tag{7.7}$$

By the writings (7.4) and (7.6), the constants K_1^τ and K_2^τ arise from the "*concentration of the measure phenomenon*" (see [6], [2]). More precisely, these constants characterize the *tightness* of the sequences of random variables $\|\mathbb{U}_n\|_{\mathcal{A}}$ (which is $(M_{fuel}^{*,i})_{i=1,\ldots,n}$ dependent) and $\|\mathbb{V}_m\|_{\mathcal{B}}$ (which is $(\boldsymbol{\epsilon}_j)_{j=1,\ldots,m}$ dependent).

Now, we aim at computing (upper bound of) these constants using concentration inequalities where the classes of functions $\mathcal{A}$ and $\mathcal{B}$ will play a crucial role. In particular, we will apply the following theorem which is Theorem 2.14.9 in [10].

Before, let us recall the definition of the *bracketing numbers* (taken from [10] p. 83-85).

DEFINITION 7.3. — **Bracketing numbers.** *Let $\mathcal{G}$ be some class of functions on T and denote by W a probability measure on T.*

Given two functions l, u, the bracket $[l, u]$ is the set of all functions g with $l \leqslant g \leqslant u$. An ε-bracket is a bracket $[l, u]$ with $||u - l||_{2,W} < \varepsilon$. The bracketing number $N_{[]}(\varepsilon, \mathcal{G}, L_2(W))$ is the minimum number of ε-brackets needed to cover the class of functions $\mathcal{G}$.

The entropy with bracketing is the logarithm of the bracketing number.

Remark 7.4. — The bracketing numbers measure the "size", the complexity of a class of functions.

THEOREM 7.5. — *Let $\mathcal{G}$ be a uniformly bounded class of (measurable) functions $g : T \to [0, 1]$ and denote by W a probability measure on T. If the class $\mathcal{G}$ satisfies, for some constants K and L*

$$N_{[]}(\varepsilon, \mathcal{G}, L_2(W)) \leqslant \left(\frac{K}{\varepsilon} \right)^L \quad \text{for every } 0 < \varepsilon < K. \tag{7.8}$$

Then, for every $t > 0$,

$$\mathbb{P}(\|\mathbb{G}_k\|_{\mathcal{G}} > t) \leqslant \left(\frac{D\,t}{\sqrt{L}} \right)^L e^{-2t^2},$$

for a constant D that only depends on K.

The proof of this theorem can be found in [8].

Now, let K^τ be a constant (to determine) which satisfies

$$\mathbb{P}(\|\mathbb{G}_k\|_{\mathcal{G}} \leqslant K^\tau) \geqslant 1 - \tau.$$

This is equivalent to

$$\mathbb{P}(\|\mathbb{G}_k\|_{\mathcal{G}} > K^\tau) \leqslant \tau. \tag{7.9}$$

By Theorem 7.5, applied with $t = K^\tau$, we have

$$\mathbb{P}(\|\mathbb{G}_k\|_{\mathcal{G}} > K^\tau) \leqslant \left(\frac{D\,K^\tau}{\sqrt{L}} \right)^L e^{-2(K^\tau)^2}. \tag{7.10}$$

Hence, the constant K^τ can be taken such that

$$\left(\frac{D\,K^\tau}{\sqrt{L}} \right)^L e^{-2(K^\tau)^2} \leqslant \tau,$$

which is similar to

$$(K^\tau)^2 - \frac{L}{2} \log(K^\tau) \geqslant \frac{\log(a_{L,D}\, \tau^{-1})}{2}\,, \quad \text{with } a_{L,D} = \left(\frac{D}{\sqrt{L}}\right)^L. \quad (7.11)$$

Then, for small enough $\tau > 0$, let us consider the constant

$$K^\tau = \sqrt{\frac{\log(a_{L,D}\, \tau^{-1})}{2}} \quad (7.12)$$

which satisfies (7.11).

Finally, we see that the constant K^τ can be characterized (only) by the class of functions $\mathcal{G}$ through the constants D and L provided by Theorem 7.5.

In our purpose, the classes of interest are $\mathcal{A}$ and $\mathcal{B}$ defined in (7.5) and (7.7), respectively. Next, one can easily check that these classes are uniformly bounded and it is suitable to work with normalized classes

$$\bar{\mathcal{A}} = \alpha_{\mathcal{A}} + \frac{1}{\beta_{\mathcal{A}}}\mathcal{A}\,, \quad (7.13)$$

$$\bar{\mathcal{B}} = \alpha_{\mathcal{B}} + \frac{1}{\beta_{\mathcal{B}}}\mathcal{B}\,, \quad (7.14)$$

such that all the functions take values in $[0, 1]$.

Now, we have to prove that the classes $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ have polynomial bracketing numbers following (7.8). This will give the constants $L_{\bar{\mathcal{A}}}$, $D_{\bar{\mathcal{A}}}$ and $L_{\bar{\mathcal{B}}}$, $D_{\bar{\mathcal{B}}}$ needed to identify the key constants K_1^τ and K_2^τ defined in (7.4) and (7.6), respectively.

7.2. Characterization of $L_{\bar{\mathcal{A}}}$, $D_{\bar{\mathcal{A}}}$, $L_{\bar{\mathcal{B}}}$, $D_{\bar{\mathcal{B}}}$

We consider the Theorem 2.7.11 in [10] (p. 164) which deals with classes that are Lipschitz in a parameter. It reads:

THEOREM 7.6. — *Let $\mathcal{G} = \{t \in T \mapsto g_s(t)\,, s \in S\}$ be a class of functions satisfying*

$$\text{for all } t \in T\,, s, s' \in S\,, \quad |g_s(t) - g_{s'}(t)| \leqslant d(s, s')\, G(t)\,,$$

for some metric d on S and some function $G : t \mapsto G(t)$.
Then, for any norm

$$N_{[\,]}(2\varepsilon\, \|G\|, \mathcal{G}, \|\cdot\|) \leqslant N(\varepsilon, S, d)\,,$$

where $N(\varepsilon, S, d)$ is the minimal number of balls $\{r\,,\, d(r,s) < \varepsilon\}$ of radius ε needed to cover the set S.

In what follows, we detail the case of the class $\bar{\mathcal{A}}$. The case of the class $\bar{\mathcal{B}}$ is exactly in the same spirit.

Let us recall that Q is the probability measure considered on $\mathcal{I}$ (observation space) and that we have

$$\bar{\mathcal{A}} = \{f_\lambda \,:\, y \in \mathcal{I} \longmapsto \alpha_{\mathcal{B}} + \frac{1}{\beta_{\mathcal{A}}}(y - \lambda)^2,\, \lambda \in \mathcal{I}_h\}\,,$$

where $\mathcal{I} = [M_{inf}, M_{sup}]$ and $\mathcal{I}_h = [M^h_{inf}, M^h_{sup}]$ (with $\mathcal{I} \subset \mathcal{I}_h$). So

$$|f_\lambda(y) - f_\lambda(y)| = \frac{1}{\beta_{\mathcal{A}}}|(y - \lambda_1)^2 - (y - \lambda_2)^2| \leqslant |\lambda_1 - \lambda_2|\, F(y)\,,$$

with $F(y) = \frac{2}{\beta_{\mathcal{A}}}(y + M^h_{sup})$, and by Theorem 7.6 applied with $\|\cdot\| = \|\cdot\|_{2,Q}$, it holds that

$$N_{[]}(\epsilon,\, \bar{\mathcal{A}},\, L_2(Q)) \leqslant N\left(\frac{\epsilon}{2\,\|F\|_{2,Q}}, \mathcal{I}_h, |\cdot|\right)\,.$$

Moreover, since

$$\|F\|_{2,Q} \leqslant \sup_{y \in \mathcal{I}} F(y)\, \|f\|_2$$

where f is the density associated to the measure Q, and using the fact that $\mathcal{I} \subset \mathcal{I}_h$, we obtain that

$$\|F\|_{2,Q} \leqslant \frac{4}{\beta_{\mathcal{A}}} M^h_{sup}\, \|f\|_2\,.$$

This last inequality yields

$$N\left(\frac{\varepsilon}{2\,\|F\|_{2,Q}}, \mathcal{I}_h, |\cdot|\right) \leqslant N\left(\frac{\beta_{\mathcal{A}}\,\varepsilon}{8\,M^h_{sup}\,\|f\|_2}, \mathcal{I}_h, |\cdot|\right)\,.$$

Since $\mathcal{I}_h = [M^h_{inf}, M^h_{sup}]$, the quantity (covering number) in the right member is bounded by

$$\frac{8\,|\mathcal{I}_h|\,M^h_{sup}\,\|f\|_2}{\beta_{\mathcal{A}}\,\varepsilon}\,,\quad |\mathcal{I}_h| = M^h_{sup} - M^h_{inf}.$$

We finally get

$$N_{[]}(\varepsilon,\, \bar{\mathcal{A}},\, L_2(Q)) \leqslant \frac{8\,|\mathcal{I}_h|\,M^h_{sup}\,\|f\|_2}{\beta_{\mathcal{A}}\,\varepsilon}\,,$$

that is

$$N_{[]}(\varepsilon, \, \bar{\mathcal{A}}, \, L_2(Q)) \leqslant \left(\frac{K_{\bar{\mathcal{A}}}}{\varepsilon} \right)^{L_{\bar{\mathcal{A}}}},$$

with

$$L_{\bar{\mathcal{A}}} = 1$$

and

$$K_{\bar{\mathcal{A}}} = \frac{8 \, |\mathcal{I}_h| \, M_{sup}^h \, \|f\|_2}{\beta_{\mathcal{A}}} \quad \text{that determines } D_{\bar{\mathcal{A}}} \text{ by } [8] \, .$$

A similar work gives the constant $L_{\bar{\mathcal{B}}} = 1$ and a constant $D_{\bar{\mathcal{B}}}$.

7.3. End of the proof

By the previous subsection, we get the constants $K_{\bar{\mathcal{A}}}^\tau$ and $K_{\bar{\mathcal{B}}}^\tau$ given by (7.12) with associated constants L and D:

$$K_{\bar{\mathcal{A}}}^\tau = \sqrt{\frac{\log(a_1 \, \tau^{-1})}{2}}, \quad a_1 = a_{L_{\bar{\mathcal{A}}}, D_{\bar{\mathcal{A}}}} = D_{\bar{\mathcal{A}}} \qquad (7.15)$$

$$K_{\bar{\mathcal{B}}}^\tau = \sqrt{\frac{\log(a_2 \, \tau^{-1})}{2}}, \quad a_2 = a_{L_{\bar{\mathcal{B}}}, D_{\bar{\mathcal{B}}}} = D_{\bar{\mathcal{B}}} \qquad (7.16)$$

where initially $a_{L,D} = \left(\frac{D}{\sqrt{L}} \right)^L$ (by (7.11)).

But, the constants of interest K_1^τ and K_2^τ defined in (7.4) and (7.6) are relative to non normalized classes $\mathcal{A}$ and $\mathcal{B}$. Let us remark that if $\bar{\mathcal{G}} = \alpha + \frac{1}{\beta}\mathcal{G}$, then

$$\|\mathbb{G}_k\|_{\bar{\mathcal{G}}} = \frac{1}{\beta} \, \|\mathbb{G}_k\|_{\mathcal{G}} \, . \qquad (7.17)$$

Now, let us denote by $K_{\bar{\mathcal{G}}}^\tau$ the constant that satisfies

$$\mathbb{P}(\|\mathbb{G}_k\|_{\bar{\mathcal{G}}} \leqslant K_{\bar{\mathcal{G}}}^\tau) \geqslant 1 - \tau \, ,$$

and denote by $K_{\mathcal{G}}^\tau$ the constant that satisfies

$$\mathbb{P}(\|\mathbb{G}_k\|_{\mathcal{G}} \leqslant K_{\mathcal{G}}^\tau) \geqslant 1 - \tau \, .$$

By (7.17), it is easy to check that we can take

$$K_{\mathcal{G}}^\tau = \beta \, K_{\bar{\mathcal{G}}}^\tau \, .$$

We deduce that

$$K_1^\tau = \beta_{\mathcal{A}} \, K_{\bar{\mathcal{A}}}^\tau$$

and

$$K_2^\tau = \beta_\mathcal{B}\, K_{\bar{\mathcal{B}}}^\tau\,.$$

Finally, by (7.2) and (7.3) we have with probability $1 - 2\tau$

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \leqslant \frac{4\|f\|_2}{\lambda_{min}\eta}\left(\frac{1}{\sqrt{n}}\frac{\eta}{2\delta^2\|f\|_2}\gamma K_1^\tau + \frac{1}{\sqrt{m}}\frac{1}{\sqrt{2\pi}\delta}K_2^\tau + \frac{1}{m^{2/5}}\frac{C(1.06\sigma)^2}{\sqrt{3}}\right)$$

which we rewrite

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 \leqslant \frac{\sqrt{2}\,c_1}{\sqrt{n}}K_{\bar{\mathcal{A}}}^\tau + \frac{\sqrt{2}\,c_2}{\sqrt{m}}K_{\bar{\mathcal{B}}}^\tau + \frac{c_3}{m^{1/5}}$$

with corresponding constants c_1, c_2 and c_3 and $K_{\bar{\mathcal{A}}}^\tau, K_{\bar{\mathcal{B}}}^\tau$ are given by (7.15) and (7.16).

That concludes the proof of Theorem 6.2.

Bibliography

[1]   BARBILLON (P.). — Méthodes d'interpolation à noyaux pour l'approximation de fonctions type boîte noire coûteuses. Thèse de doctorat, Université de Paris-Sud (2010).

[2]   BILLINGSLEY (P.). — Convergence of probability measures. Wiley New York (1968).

[3]   DE ROCQUIGNY (E.) and CAMBIER (S.). — Inverse probabilistic modelling of the sources of uncertainty: a non-parametric simulated-likelihood method with application to an industrial turbine vibration assessment. Inverse Problems in Science and Engineering, 17(7), p. 937-959 (2009).

[4]   DE ROCQUIGNY (E.), DEVICTOR (N.) and TARANTOLA (S.) editors. — Uncertainty in industrial practice. John Wiley.

[5]   KUHN (E.). — Estimation par maximum de vraisemblance dans des problèmes inverses non linéaires. Thèse de doctorat, Université de Paris-Sud (2003).

[6]   LEDOUX (N.). — The concentration of measure phenomenon. AMS (2001).

[7]   RACHDI (N.), FORT (J.C.) and KLEIN (T.). — Risk bounds for new M-estimation problems. http://hal.archives-ouvertes.fr/hal-00537236_v2/ (submitted) (2012).

[8]   TALAGRAND (M.). — Sharper bounds for Gaussian and empirical processes. The Annals of Probability, 22(1), p. 28-76 (1994).

[9]   VAN DER VAART (A.W.). — Asymptotic statistics. Cambridge University Press (2000).

[10]   VAN DER VAART (A.W.) and WELLNER (J.A.). — Weak Convergence and Empirical Processes. Springer Series in Statistics (1996).

[11]   WIENER (N.). — The homogeneous chaos. American Journal of Mathematics, 60(4), p. 897-936 (1938).