NUCLEAR SCIENCE AND ENGINEERING: **176**, 81–97 (2014)

Calibration and Improved Prediction of Computer Models by Universal Kriging

François Bachoc*

CEA-Saclay, DEN, DM2S, STMF, LGLS
F-91191 Gif-Sur-Yvette, France

and

Université Paris VII, Laboratoire de Probabilités et Modèles Aléatoires, Site Chevaleret
case 7012, 75205 Paris cedex 13, France

Guillaume Bois

CEA-Saclay, DEN, DM2S, STMF, LATF, F-91191 Gif-Sur-Yvette, France

Josselin Garnier

Université Paris VII, Laboratoire de Probabilités et Modèles Aléatoires & Laboratoire
Jacques-Louis Lions, Site Chevaleret, case 7012, 75205 Paris cedex 13, France

and

Jean-Marc Martinez

CEA-Saclay, DEN, DM2S, STMF, LGLS
F-91191 Gif-Sur-Yvette, France

Received June 22, 2012
Accepted March 12, 2013
doi:10.13182/NSE12-55

Abstract—*This paper addresses the use of experimental data for calibrating a computer model and improving its predictions of the underlying physical system. A global statistical approach is proposed in which the bias between the computer model and the physical system is modeled as a realization of a Gaussian process. The application of classical statistical inference to this statistical model yields a rigorous method for calibrating the computer model and for adding to its predictions a statistical correction based on experimental data. This statistical correction can substantially improve the calibrated computer model for predicting the physical system on new experimental conditions. Furthermore, a quantification of the uncertainty of this prediction is provided. Physical expertise on the calibration parameters can also be taken into account in a Bayesian framework. Finally, the method is applied to the thermal-hydraulic code FLICA 4, in a single-phase friction model framework. It allows significant improvement of the predictions of FLICA 4.*

*E-mail: francois.bachoc@gmail.com

I. INTRODUCTION

In physics and engineering, computer experiments have been used for a long time as surrogates to costly, or unpractical, real experiments. However, the impact of the substitution of a computer experiment for an actual experiment is not easy to assess. In general, this substitution induces a double error. The first error is that a computer experiment is the numerical implementation of a mathematical model, so that a bias always exists between this mathematical model and the computer model. The quantification and reduction of this first error is the field of model verification.[1] The second error is that the mathematical model itself may not represent perfectly the underlying physical phenomenon. This second error defines the field of model validation. A reference book on model validation is given in, for example, Ref. 2.

In the present work, and similar to several references on statistical analysis of the validation problem,[3] we assume that the computer model has already been verified. Hence, we focus on the validation problem, which is a very important issue in nuclear engineering.[4] Nevertheless, our objective is less to study the validity of the computer model than to improve the computer model predictions, and quantify the uncertainty obtained, by assimilating experimental results. A recent reference on demonstrating, or refuting, the validity of the actual computer model is Ref. 3.

In practice, the study of computer model predictions is complicated by the fact that a computer model often comes with fitting parameters. These parameters either allow modeling of a physical system more accurately or are a consequence of uncertainties with respect to some physical parameters. Hence, the analysis of the predictions of the computer model often needs to be carried out simultaneously with a calibration analysis. The goal of the calibration analysis is to quantify the uncertainty related to the fitting parameters. An example of a methodology carrying out calibration and prediction analysis globally is the best-estimate methodology.[5,6] This methodology proceeds through data assimilation and takes into account the different sources of uncertainties from the model parameters, the experimental errors, the numerical errors, and the limitations of the physical models.

In this work, a computer model is a function f_{mod} of the form $f_{mod}(\boldsymbol{x},\boldsymbol{\beta}) : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$. This computer model is a representation of a physical system that is a deterministic function taking the form $f_{real}(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}$.

The scalar output of the physical system is the physical quantity of interest. It is a deterministic function f_{real} of a vector $\boldsymbol{x}$ of input quantities, which we call experimental conditions. These components of the vector $\boldsymbol{x}$ can be divided into two categories. The first category contains the control variables. These variables define the physical system, independently of the environment in which the system is put. In engineering, for instance, geometric parameters of the system can often be placed in this category, as they remain fixed regardless of what happens to the system. The second category contains the environment variables. These variables are the input of the physical system whose values are not planned in the conception of the system. These variables are likely to be imposed beforehand by other systems. The distinction of the experimental conditions into these two categories is presented, for instance, in Ref. 7, Sec. 2.1. To give an illustration, in the system design phase, the environment variables are set by the future use of the system, while the control variables are the free parameters that may be set through an optimization phase.

The function f_{real} of the physical system cannot be evaluated for all the experimental conditions. Hence, this function is approximated by the computer model f_{mod}. This function shares the same input vector $\boldsymbol{x}$ as the physical system and provides the same scalar output. Furthermore, the function f_{mod} has a second kind of input, denoted by the vector $\boldsymbol{\beta}$. The components of this vector are the fitting parameters of the computer model f_{mod}. These parameters are unnecessary to carry out an experiment of the physical system, but they are needed to run the computer model. Hence, these quantities are seen as degrees of freedom for the computer model and allow it to give a good approximation of the physical system. In the sequel, these parameters are called model parameters.

The calibration problem is the problem of computing a value $\hat{\boldsymbol{\beta}}$ for the model parameter $\boldsymbol{\beta}$. This is done on the basis of a set of experimental results. A set of experimental results is a set $\{\boldsymbol{x}^{(1)}, f_{obs}(\boldsymbol{x}^{(1)}), \ldots, \boldsymbol{x}^{(n)}, f_{obs}(\boldsymbol{x}^{(n)})\}$, where $f_{obs}(\boldsymbol{x}^{(i)})$ is the output of the physical system observed at the experimental condition $\boldsymbol{x}^{(i)}$. Note that the quantity $f_{obs}(\boldsymbol{x}^{(i)})$ may be different from $f_{real}(\boldsymbol{x}^{(i)})$, for example, because of measurement errors. In practice, in nuclear engineering, calibration is an important issue because many computer models usually have model parameters that are totally or partially unknown.[8]

Before presenting the Gaussian process modeling of this work, we emphasize the potential limitations of uncertainty quantification methods that would address only calibration. The most classical example of these methods is the least-squares calibration, which consists of minimizing with respect to $\boldsymbol{\beta}$ a quadratic misfit between the experimental results and the values that the computer model predicts when parameterized by $\boldsymbol{\beta}$. Hence, the value of $\boldsymbol{\beta}$ obtained with the least-squares method is $\hat{\boldsymbol{\beta}}_{LS} \in \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(f_{mod}(\boldsymbol{x}^{(i)}, \boldsymbol{\beta}) - f_{obs}(\boldsymbol{x}^{(i)})\right)^2$. A method addressing only the calibration problem, such as the least-squares calibration method, is generally based on two hypotheses. The first hypothesis is that the computer model is capable of perfectly reproducing the physical system. That is, there is a model parameter $\boldsymbol{\beta}_0$ so that $\forall \boldsymbol{x}, f_{real}(\boldsymbol{x}) = f_{mod}(\boldsymbol{x}, \boldsymbol{\beta}_0)$. This hypothesis means that the physical knowledge that was put in the computer model is

sufficient to model the physical system perfectly. The second hypothesis is that the deviations $(f_{mod}(\boldsymbol{x}^{(i)},\boldsymbol{\beta}_0) - f_{obs}(\boldsymbol{x}^{(i)}))$ come from uncertainties related to the experiments. These uncertainties generally have two sources. First, the observations are affected by measurement errors. Second, there is a replicate uncertainty, meaning that the experimental conditions cannot be known exactly for a given experiment. The main limitation is the assumption that the deviations $f_{mod}(\boldsymbol{x}^{(i)}, \boldsymbol{\beta}) - f_{obs}(\boldsymbol{x}^{(i)})$ come only from uncertainties related to the experiments. Indeed, the order of magnitude of these uncertainties is known. Hence, when the errors $|f_{mod}(\boldsymbol{x}^{(i)}, \boldsymbol{\beta}) - f_{obs}(\boldsymbol{x}^{(i)})|$ are too large compared to this order of magnitude, it indicates that there is a problem with this assumption (this can be quantified by Monte Carlo methods). In this case, the fact that the computer model cannot represent the physical system perfectly needs to be taken into account.

The limitation discussed above gives motivations for reconsidering the assumption that there is a model parameter $\boldsymbol{\beta}_0$ so that $\forall \boldsymbol{x}, f_{real}(\boldsymbol{x}) = f_{mod}(\boldsymbol{x}, \boldsymbol{\beta}_0)$. In this paper, this is done by taking a model error into account.

In Sec. II, we present, in detail, the Gaussian process modeling of the model error and show how this modeling yields a framework for calibration and prediction. We also give a one-dimensional illustration of an analytical function. In Sec. III, we present an application case, relevant to nuclear engineering, on the thermal-hydraulic code FLICA 4. FLICA 4 is mainly dedicated to core thermal-hydraulic transient and steady-state analysis.[9]

II. THE GAUSSIAN PROCESS METHOD FOR CALIBRATION AND PREDICTION

II.A. Gaussian Process Modeling of the Model Error

We present a Gaussian process model that is the basis of the Gaussian process method for calibration and prediction. This statistical model is based on two main ideas:

1. The physical system $\boldsymbol{x} \rightarrow f_{real}(\boldsymbol{x})$ does not necessarily belong to the set of computer model functions $\{f_{mod}(\boldsymbol{x},\boldsymbol{\beta})\}$. We model the difference between the physical system and the correctly parameterized computer model by an error function that is called the model error. The notion of a correctly parameterized computer model is explained in the following.

2. The model error function is not observable everywhere and hence is unknown for the majority of the experimental conditions. This lack of knowledge is modeled by the introduction of a stochastic framework for this function; i.e., it is represented by a realization of the random process $Z(\omega,\boldsymbol{x})$. This probabilistic modeling is a Bayesian modeling of the uncertainty on the deterministic model error function. The reader may refer to Ref. 7, pp. 23–24, for a discussion of Bayesian modeling of deterministic functions. In this context, the particular interest of Gaussian processes is discussed in Ref. 10, p. 2. As it is the sum of the correctly parameterized computer model and the model error function, the physical system itself is a realization of a Gaussian process. Hence, we do not use the notation $f_{real}(\boldsymbol{x})$ anymore for the physical system. Instead, we denote it by the random process $Y_{real}(\omega, \boldsymbol{x})$.

Motivated by these two ideas, the Gaussian process statistical model is defined by the following two equations:

$$Y_{real}(\omega, \boldsymbol{x}) = f_{mod}(\boldsymbol{x}, \boldsymbol{\beta}) + Z(\omega, \boldsymbol{x}) \tag{1}$$

and

$$Y_{obs}(\omega, \boldsymbol{x}) = Y_{real}(\omega, \boldsymbol{x}) + \epsilon(\omega, \boldsymbol{x}) , \tag{2}$$

where

ω = a value in a probability space Ω

$Y_{real}(\omega, \boldsymbol{x})$ = random process of the physical system

$Z(\omega, \boldsymbol{x})$ = model error process.

The random process Z is assumed to be Gaussian[7,11,12] and centered. The model error process Z is hence defined by its covariance function C_{mod}. In practice, this function belongs to a prescribed set of covariance functions (see, e.g., Ref. 13 for classical examples) and is defined up to a few hyperparameters that can be estimated from data.

Above, $\boldsymbol{\beta}$ is the correct parameter of the computer model. We call it the correct parameter because, with Z being centered, the computer model parameterized by $\boldsymbol{\beta}$ is the mean value of the physical system.

Also, $Y_{obs}(\omega,\boldsymbol{x})$ is the observed output of the physical system for the experimental conditions $\boldsymbol{x}$. This observation is the sum of the quantity of interest and of a measurement error $\epsilon(\omega,\boldsymbol{x})$. The error $\epsilon(\omega,\boldsymbol{x})$ follows a Gaussian centered law, and is independent from one experiment to another. The variance of ϵ is, in general, constant.

In the sequel, we do not write ω explicitly. One could model Z as a white noise process (with independent components for different $\boldsymbol{x}$), which would give a statistical model leading to the least-squares calibration of Sec. I. There are two reasons for not doing so, and we use instead a covariance function with a dependence structure:

1. The physical system is generally continuous with respect to the experimental conditions, and so is the numerical model. Hence, as a difference, the model error process Z must be a process with continuous trajectories. This is not the case for a white noise process.

2. Similarly, it is expected that if the computer model makes a certain error for a given experimental point, then

it will make a similar error for a nearby experimental point. This principle is taken into account by a covariance function with a dependence structure.

The statistical modeling also allows expert judgments to be taken into account for the model parameter $\boldsymbol{\beta}$. This is done within the Bayesian framework, modeling the constant but unknown correct model parameter $\boldsymbol{\beta}$ as a random vector. The law of this random vector is known and chosen according to the degree of knowledge one has about the model parameter $\boldsymbol{\beta}$. We use a Gaussian distribution for the Bayesian modeling of $\boldsymbol{\beta}$. Hence, we distinguish two cases:

1. *No prior information case:* $\boldsymbol{\beta}$ is a vector of unknown constants.

2. *Prior information case:* $\boldsymbol{\beta}$ is a random vector, with known mean vector $\boldsymbol{\beta}_{prior}$ and covariance matrix Q_{prior}.

In this paper, we work with a linear approximation of the computer model with respect to its model parameters (within the range of values that is under consideration). Hence, we consider computer models of the form $f_{mod}(\boldsymbol{x}, \boldsymbol{\beta}) = f_{mod}(\boldsymbol{x}, \boldsymbol{\beta}_{nom}) + \sum_{i=1}^{m} h_i(\boldsymbol{x})(\boldsymbol{\beta}_i - \boldsymbol{\beta}_{nom,\,i})$, where $\boldsymbol{\beta}_{nom}$ is the nominal vector around which the linear approximation is made. $\boldsymbol{\beta}_{nom}$ is generally chosen by expert judgment or by previous calibration studies. We choose, for simplicity reasons, to remove the perfectly known quantities $\boldsymbol{\beta}_{nom}$ and $f_{mod}(\boldsymbol{x}, \boldsymbol{\beta}_{nom})$. Indeed, up to a shift with respect to $\boldsymbol{\beta}$ and f_{mod}, we can consider that $\boldsymbol{\beta}_{nom} = 0$ and $f_{mod}(\boldsymbol{x}, \boldsymbol{\beta}_{nom}) = 0$. We then have

$$\forall \boldsymbol{x} : f_{mod}(\boldsymbol{x}, \boldsymbol{\beta}) = \sum_{i=1}^{m} h_i(\boldsymbol{x})\boldsymbol{\beta}_i \, . \tag{3}$$

The linear approximation is justified by a Taylor series expansion when the uncertainty concerning the correct parameter $\boldsymbol{\beta}$ is small. This linear approximation is frequently made, for example, in thermal hydraulics[6,14] or in neutron transport.[15] A thorough discussion on the validity of using the linear approximation in the nonlinear case is given in Sec. II.E.

Gaussian process modeling allows solution of the following two problems:

1. *Calibration:* It is the problem of estimating the correct model parameter $\boldsymbol{\beta}$ or equivalently finding the most accurate computer model function $\boldsymbol{x} \rightarrow f_{mod}(\boldsymbol{x}, \boldsymbol{\beta})$.

2. *Prediction:* For a new experimental condition $\boldsymbol{x}_{new}$, we want to predict the quantity of interest of the physical system and add a measure of uncertainty to this prediction. The main idea is that the quantity of interest is not predicted by the calibrated computer model because we are able to infer the value of the model error at $\boldsymbol{x}_{new}$.

The calibration and prediction are presented in Sec. II.C. They are obtained using classical linear algebra tools, as long as the covariance function C_{mod} of the model error is known. In fact, the function C_{mod} depends on a set of hyperparameters that are to be estimated from data. We present the estimation method in Sec. II.B.

Hence, in the most classical case, Gaussian process modeling is treated in two steps. In the first step, the hyperparameters of the covariance function are estimated so that this function is considered fixed in the second step, where linear algebra is used for the calibration and prediction. Note that there exist methods where these two steps are done simultaneously, for instance, if a Bayesian prior for the hyperparameters of the covariance function is used. These methods are more costly but can improve the quality of the Gaussian process modeling, as shown, e.g., in Ref. 16, in an optimization context.

Note that in our case, a third step of verification is necessary. This step consists of verifying that the modeling leads to calibration and prediction that give satisfying results. This step can notably be carried out by cross validation or by using a new set of experimental results that was never used before.

We now formulate the problem in vector-matrix form. Assume that n experiments are carried out at $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}$. We denote the following:

1. the $n \times m$ matrix $\mathbf{H}$ of partial derivatives of the computer model with respect to $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$; $\mathbf{H}$ is defined by $H_{i,\,j} = h_j(\boldsymbol{x}^{(i)})$
2. the random vectors $\mathbf{y}_{obs}$ of the n observations; $y_{obs,\,i}$ is the result of the i'th experiment.
3. the random vector ϵ of the measurement errors for the n experiments; we have $y_{obs,\,i} = Y_{real}(\boldsymbol{x}^{(i)}) + \epsilon_i$
4. the random vector z of the model error for the n experiments; $z = (Z(\boldsymbol{x}^{(1)}), \ldots, Z(\boldsymbol{x}^{(n)}))^t$
5. the covariance matrix of ϵ: $\mathbf{R}_{mes}$
6. the covariance matrix of z: $\mathbf{R}_{mod}$.

For the n experiments, Eqs. (1) and (2) become

$$\boldsymbol{y}_{obs} = \mathbf{H}\boldsymbol{\beta} + z + \epsilon \, . \tag{4}$$

Hence, we have a universal kriging model.[7] We denote by $\mathbf{R}$ the covariance matrix of the model and measurement error vector $z + \epsilon$:

$$\mathbf{R} : = \mathrm{cov}(z + \epsilon) = \mathbf{R}_{mod} + \mathbf{R}_{mes} \, . \tag{5}$$

The distinction between $\mathbf{R}_{mod}$ and $\mathbf{R}_{mes}$ is important because, usually, $\mathbf{R}_{mes}$ is at least partially known from knowledge of the experimental process, while $\mathbf{R}_{mod}$ generally does not benefit from physical knowledge. Indeed, physical knowledge is used in the conception of

the computer model, and hence, it may not help knowing the shape of the error of the computer model.

We can compute the a priori law of the vector of observations, i.e., the statistical distribution of the observations before carrying out the experiments, but given the hyperparameters. In the no prior information case we have, with $\boldsymbol{\beta}$ an unknown constant,

$$\boldsymbol{y}_{obs} \sim \mathcal{N}(\mathbf{H}\boldsymbol{\beta},\ \mathbf{R}) \ . \tag{6}$$

In the prior information case, we have, with $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_{prior},\ \mathbf{Q}_{prior})$,

$$\boldsymbol{y}_{obs} \sim \mathcal{N}(\mathbf{H}\boldsymbol{\beta}_{prior},\ \mathbf{H}\mathbf{Q}_{prior}\mathbf{H}^t + \mathbf{R}) \ , \tag{7}$$

where $\mathcal{N}(\boldsymbol{m}, \mathbf{R})$ stands for the multivariate Gaussian distribution with mean vector $\boldsymbol{m}$ and covariance matrix $\mathbf{R}$ and $\sim$ means "follows the distribution of."

II.B. Estimation of the Covariance of the Model Error

For the Gaussian process modeling defined in Eqs. (1) and (2) to be tractable with closed-form linear algebra formulas (Sec. II.C), it is necessary that the covariance functions of the model error Z and of the measurement error ϵ are known. We show here how to compute these covariance functions.

The covariance function of the measurement error can generally be specified from physical expertise. This is the case here. If it is not the case, this function can, for example, be estimated in the same way as the model error covariance function.

Generally, there is no expert judgment available concerning the model error covariance function C_{mod}, as has been discussed above. A specific structure is chosen for C_{mod}, with a limited number of degrees of freedom. Hence, we consider the family of covariance functions

$$\mathcal{C}_{mod} = \left\{ \sigma^2 C_{mod,\,\theta},\ \sigma > 0,\ \theta \in \Theta \right\} \ ,$$

where Θ is a subset of $\mathbb{R}^p$ and $C_{mod,\,\theta}$ is a stationary correlation function. We present classical correlation functions families, for which $p = d$, $\theta = (l_{c,\,1}, \ldots, l_{c,\,d})$, and $C_{mod,\,\theta}(\boldsymbol{x}^{(a)}, \boldsymbol{x}^{(b)}) = C_{mod,\,\theta}(\boldsymbol{h})$, with $\boldsymbol{h} = \boldsymbol{x}^{(a)} - \boldsymbol{x}^{(b)}$ (the component $l_{c,\,i}$ can be seen as a correlation length in the i'th dimension):

1. exponential correlation function:

$$C_{mod,\,\theta}(\boldsymbol{h}) = \exp\left(- \sum_{i=1}^{d} \frac{|h_i|}{l_{c,\,i}} \right) ;$$

2. Matérn $\nu = \frac{3}{2}$ correlation function, with $|\boldsymbol{h}|_\theta = \sqrt{\sum_{i=1}^{d} h_i^2 / l_{c,\,i}^2}$:

$$C_{mod,\,\theta}(\boldsymbol{h}) = (1 + \sqrt{6}|\boldsymbol{h}|_\theta)\exp(- \sqrt{6}|\boldsymbol{h}|_\theta) \ ;$$

3. Matérn $\nu = \frac{5}{2}$ correlation function:

$$C_{mod,\,\theta}(\boldsymbol{h}) = (1 + \sqrt{10}|\boldsymbol{h}|_\theta + \frac{10}{3}|\boldsymbol{h}|_\theta^2)\exp(- \sqrt{10}|\boldsymbol{h}|_\theta) \ ;$$

4. Gaussian correlation function:

$$C_{mod,\,\theta}(\boldsymbol{h}) = \exp(- |\boldsymbol{h}|_\theta^2) \ .$$

The correlation functions above yield sample functions of increasing regularity (see, e.g., Ref. 11). The importance of the regularity of the correlation function is presented in detail in Ref. 11.

Assume now that we have n experimental results $\boldsymbol{y}_{obs} = (y_{obs,\,1}, \ldots, y_{obs,\,n})$ and recall the notations of Eqs. (4) and (5). As we have seen, $\mathbf{R}_{mes}$ is fixed, and $\mathbf{R}_{mod}$ depends on (σ^2, θ), which are to be estimated. We use the notation $\mathbf{R}_{\sigma,\,\theta}$ for the global covariance matrix $\mathbf{R} = \mathbf{R}_{mod} + \mathbf{R}_{mes}$.

There are several methods that can be used to estimate the hyperparameters (σ^2, θ) from data $\boldsymbol{y}_{obs}$. The most widely used are maximum likelihood[17] and cross validation.[18,19]

In this work, we use the restricted maximum likelihood estimator (RMLE) of $(\sigma^2, \boldsymbol{\theta})$. This estimator is, for instance, presented in Ref. 20. The advantage of this estimator is that the estimation of (σ^2, θ) is independent of the estimation of $\boldsymbol{\beta}$. Furthermore, this method allows us to have the same estimation of (σ^2, θ) in both the prior and no prior information cases. Finally, let us notice that $n > m$ is required for the RMLE method; i.e., there are more experiments than model parameters. In thermal hydraulics, the field of the application case, this condition generally holds in practice. Nevertheless, in other fields of nuclear engineering, typically in neutron transport,[15] one may have $m \gg n$. In this case, if one wants to address the present model error modeling anyway, it is recommended to work in a fully Bayesian framework, both for the model parameters and the covariance hyperparameters as described in Ref. 7, Sec. 4.1.4. Indeed, the very large number of model parameters makes the uncertainty related to the hyperparameters of the model error covariance function too large to be neglected, as is done when these hyperparameters are fixed to their estimated values.

Let $\mathbf{W}$ be a $(n - m \times n)$ matrix of full rank so that $\mathbf{W}\mathbf{H} = 0$. Notice that if $\mathbf{H}$ is not of full rank, then m must be replaced by the rank of $\mathbf{H}$. Then,

$$\boldsymbol{w} : = \mathbf{W}\boldsymbol{y}_{obs} \sim \mathcal{N}(0,\ \mathbf{W}\mathbf{R}_{\sigma,\,\theta}\mathbf{W}^t) \ .$$

The law of $\boldsymbol{w}$ is independent of the value of β. Hence, the

RMLE $(\hat{\sigma}^2, \hat{\theta})$ is the maximum likelihood estimator on the transformed observations $\boldsymbol{w}$,

$$(\hat{\sigma}, \hat{\theta}) \in \arg\min_{(\sigma, \theta)} q(\sigma, \theta) , \tag{8}$$

with

$$q(\sigma, \theta) = \ln|\mathbf{W}\mathbf{R}_{\sigma, \theta}\mathbf{W}^t| + \boldsymbol{w}^t(\mathbf{W}\mathbf{R}_{\sigma, \theta}\mathbf{W}^t)^{-1}\boldsymbol{w} . \tag{9}$$

It is shown in Ref. 21 that changing $\mathbf{W}$ only adds a constant [with respect to (σ^2, θ)] term to Eq. (9). It is also shown in Ref. 21 how one can avoid a matrix product with $\mathbf{W}$. Indeed, for $\mathbf{W}$ so that $\mathbf{W}\mathbf{W}^t = \mathbf{I}_{n-m}$ and $\mathbf{W}^t\mathbf{W} = \mathbf{I}_n - \mathbf{H}(\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t$, we have

$$\begin{aligned} q(\sigma, \theta) = & -\ln|\mathbf{H}^t\mathbf{H}| + \ln|\mathbf{R}_{\sigma, \theta}| \\ & + \ln|\mathbf{H}^t\mathbf{R}_{\sigma, \theta}^{-1}\mathbf{H}| + \boldsymbol{y}_{obs}^t\boldsymbol{\Pi}_{\sigma, \theta}\boldsymbol{y}_{obs} , \end{aligned} \tag{10}$$

with

$$\boldsymbol{\Pi}_{\sigma, \theta} = \mathbf{R}_{\sigma, \theta}^{-1} - \mathbf{R}_{\sigma, \theta}^{-1}\mathbf{H}(\mathbf{H}^t\mathbf{R}_{\sigma, \theta}^{-1}\mathbf{H})^{-1}\mathbf{H}^t\mathbf{R}_{\sigma, \theta}^{-1} .$$

Let $\mathbf{U}$, $\mathbf{S}$, $\mathbf{V}$ be a singular value decomposition of $\mathbf{H}$, with $\mathbf{U}$ of size $n \times m$ so that $\mathbf{U}^t\mathbf{U} = \mathbf{I}_{m, m}$, $\mathbf{S}$ a diagonal matrix of size m, with nonnegative numbers on the diagonal, and $\mathbf{V}$ an orthogonal matrix of size m, so that $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^t$. Then, we can show that

$$\begin{aligned} q(\sigma, \theta) = & \ln\left|\mathbf{U}^t\mathbf{R}_{\sigma, \theta}^{-1}\mathbf{U}\right| + \ln|\mathbf{R}_{\sigma, \theta}| + \boldsymbol{y}_{obs}^t\mathbf{R}_{\sigma, \theta}^{-1}\boldsymbol{y}_{obs} \\ & - \boldsymbol{y}_{obs}^t\mathbf{R}_{\sigma, \theta}^{-1}\mathbf{U}(\mathbf{U}^t\mathbf{R}_{\sigma, \theta}^{-1}\mathbf{U})^{-1}\mathbf{U}^t\mathbf{R}_{\sigma, \theta}^{-1}\boldsymbol{y}_{obs}. \end{aligned} \tag{11}$$

Hence, it does not matter if $\mathbf{H}$ is ill-conditioned, or even singular, since its singular values are actually not used in the computation of the restricted likelihood. Using Eq. (11) allows both $n \times n$ matrix multiplications and numerical issues with respect to the condition number of $\mathbf{H}$ to be avoided.

II.C. Calibration and Prediction

Throughout this section, we assume that the covariance function C_{mod} of Z is estimated and fixed, and we use the classical kriging formulas to solve the calibration and prediction problems. The kriging formulas, in both the Bayesian and frequentist cases, can be found in Ref. 7. We will see that these formulas require $\mathbf{R}$ to be nonzero; i.e., there are model or measurement errors. For consistency, we first address the case $\mathbf{R} = 0$, which is actually straightforward. If there is a unique $\boldsymbol{\beta}$ so that $f_{mod}(\boldsymbol{x}, \boldsymbol{\beta})$ reproduces all the experiments, then $\boldsymbol{\beta}$ is the correct parameter with zero associated uncertainty. If there is more than one $\boldsymbol{\beta}$ so that $f_{mod}(\boldsymbol{x}, \boldsymbol{\beta})$ reproduces all the experiments, then the computer model is redundantly parameterized or the number of experiments is insufficient. If there is no $\boldsymbol{\beta}$, so that $f_{mod}(\boldsymbol{x}, \boldsymbol{\beta})$ reproduces all the experiments, then the assumption of no model error and no measurement error is invalidated. In the sequel, we consider $\mathbf{R}$ to be nonzero, and $\mathbf{R}$ invertible, which is the case for the classical covariance functions of Sec. II.B.

In the no prior information case, the calibration problem is the frequentist problem of estimating the unknown parameter $\boldsymbol{\beta}$. The maximum likelihood estimation of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^t\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^t\mathbf{R}^{-1}\boldsymbol{y}_{obs} . \tag{12}$$

This estimator is unbiased and has the covariance matrix

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{H}^t\mathbf{R}^{-1}\mathbf{H})^{-1} . \tag{13}$$

We see that if there is a $\boldsymbol{\beta}$ so that $\mathbf{H}\boldsymbol{\beta} = \boldsymbol{y}_{obs}$, then we have $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$. This means that, if we are in the favorable case when the computer model can perfectly reproduce the experiments, then the Gaussian process calibration of the computer model will achieve this perfect reproduction, as should be expected. Finally, as the random vector $\hat{\boldsymbol{\beta}}$ has Gaussian distribution, its covariance matrix is sufficient to yield confidence ellipsoids for $\boldsymbol{\beta}$.

In the prior information case, the posterior distribution of $\boldsymbol{\beta}$ given the observations $\boldsymbol{y}_{obs}$ is Gaussian with mean vector

$$\begin{aligned} \boldsymbol{\beta}_{post} = & \boldsymbol{\beta}_{prior} + \\ & (\mathbf{Q}_{prior}^{-1} + \mathbf{H}^t\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^t\mathbf{R}^{-1}(\boldsymbol{y}_{obs} - \mathbf{H}\boldsymbol{\beta}_{prior}) \end{aligned} \tag{14}$$

and covariance matrix

$$\mathbf{Q}_{post} = (\mathbf{Q}_{prior}^{-1} + \mathbf{H}^t\mathbf{R}^{-1}\mathbf{H})^{-1}. \tag{15}$$

We notice that, when $\mathbf{Q}_{prior}^{-1} \to 0$, then the prior information case calibration tends to the no prior information case calibration. This is an intuitive fact because $\mathbf{Q}_{prior}^{-1}$ small corresponds to a small a priori knowledge of $\boldsymbol{\beta}$ and hence should, in the limit case, correspond to an absence of knowledge.

Remark: The prior information case calibration of Eq. (14) is classically used in neutron transport,[15] when the linear approximation (3) of the computer model is also made. In Ref. 15, no model error is assumed, so that the physical system is predicted by the calibrated computer model only. In thermal hydraulics, which is the field of the case of application, this hypothesis is not justified. Indeed, computer models can rely on aggregation of correlation models that have no physical justification. We will see in the prediction formulas Eqs. (16) and (18), and in the application case of Sec. III, that modeling the model

error allows significant improvement of the predictions of a computer model that is only partially representative of the physical system.

We now present the prediction formulas. In the same way as the computer model, the goal of the prediction is to give the most probable value of the physical system, for a new experimental condition, without doing a real experiment. However, this most probable value is not necessarily given by the output of the calibrated computer model because the model error is also inferred. We now give the notations that we use for a new experimental condition $\boldsymbol{x}_{new}$:

1. the random value of the physical system at $\boldsymbol{x}_{new}$: $Y_{real}(\boldsymbol{x}_{new})$
2. the vector of derivatives of the computer model with respect to $\beta_1, \ldots, \beta_m$ at $\boldsymbol{x}_{new}$: $\boldsymbol{h}(\boldsymbol{x}_{new})$. Hence, we have $(\boldsymbol{h}(\boldsymbol{x}_{new}))_i = h_i(\boldsymbol{x}_{new})$.
3. the model error at $\boldsymbol{x}_{new}$: z_{new}
4. the covariance vector of the model error between $(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)})$ and $\boldsymbol{x}_{new}$: $\boldsymbol{r}_{mod}(\boldsymbol{x}_{new})$ given by $(\boldsymbol{r}_{mod}(\boldsymbol{x}_{new}))_i := \mathrm{cov}(z_i, z_{new})$.

In the no prior information case, the best linear unbiased predictor (BLUP) of $Y_{real}(\boldsymbol{x}_{new})$ with respect to the vector of observations $\boldsymbol{y}_{obs}$ is

$$\hat{y}(\boldsymbol{x}_{new}) = \underbrace{(\boldsymbol{h}(\boldsymbol{x}_{new}))^t\hat{\boldsymbol{\beta}}}_{\text{calibrated computer model}} + \underbrace{(\boldsymbol{r}_{mod}(\boldsymbol{x}_{new}))^t\mathbf{R}^{-1}(\mathbf{y}_{obs} - \mathbf{H}\hat{\boldsymbol{\beta}})}_{\text{inferred model error}}. \quad (16)$$

We refer to Ref. 22 for a detailed definition of the BLUP and the computation of the predictor in Eq. (16). This predictor is composed of the calibrated computer model and the inferred model error. By inspection of Eq. (16), the inferred model error has the following properties:

1. With $\boldsymbol{x}_{new}$ being fixed, this term is large when the errors $\boldsymbol{y}_{obs} - \mathbf{H}\boldsymbol{\beta}$ between the experimental results and the calibrated computer model are large.

2. With the observations being fixed, this term is a linear combination of the components of $\boldsymbol{r}_{mod}(\boldsymbol{x}_{new})$. These elements are usually a decreasing function of the distance between $\boldsymbol{x}_{new}$ and the experimental conditions $\boldsymbol{x}^{(i)}$. Hence, if $\boldsymbol{x}_{new}$ is far from an experimental condition $\boldsymbol{x}^{(i)}$, then the weight of this experimental result is small in the combination. Hence, the prediction of $Y_{real}(\boldsymbol{x}_{new})$ is almost only composed of the calibrated computer model when $\boldsymbol{x}_{new}$ is far from any available experimental condition, while the model error inference term is significant when $\boldsymbol{x}_{new}$ is in the neighborhood of an available experimental condition (the neighborhood is defined in terms of the correlation lengths).

The mean square error of the BLUP is

$$\hat{\sigma}^2(\boldsymbol{x}_{new}) = C_{mod}(\boldsymbol{x}_{new}, \boldsymbol{x}_{new}) - \boldsymbol{r}_{mod}(\boldsymbol{x}_{new})^t\mathbf{R}^{-1}\boldsymbol{r}_{mod}(\boldsymbol{x}_{new}) + (\boldsymbol{h}(\boldsymbol{x}_{new}) - \mathbf{H}^t\mathbf{R}^{-1}\boldsymbol{r}_{mod}(\boldsymbol{x}_{new}))^t(\mathbf{H}^t\mathbf{R}^{-1}\mathbf{H})^{-1}(\boldsymbol{h}(\boldsymbol{x}_{new}) - \mathbf{H}^t\mathbf{R}^{-1}\boldsymbol{r}_{mod}(\boldsymbol{x}_{new})). \quad (17)$$

Since only linear combinations have been used, the BLUP has a Gaussian distribution, and the mean square error allows confidence intervals to be built.

In the prior information case, the posterior distribution of $Y_{real}(\boldsymbol{x}_{new})$ given the observations $\boldsymbol{y}_{obs}$ is Gaussian with mean

$$\hat{y}(\boldsymbol{x}_{new}) = \underbrace{(\boldsymbol{h}(\boldsymbol{x}_{new}))^t\boldsymbol{\beta}_{post}}_{\text{calibrated computer model}} + \underbrace{(\boldsymbol{r}_{mod}(\boldsymbol{x}_{new}))^t\mathbf{R}^{-1}(\boldsymbol{y}_{obs} - \mathbf{H}\boldsymbol{\beta}_{post})}_{\text{inferred model error}} \quad (18)$$

and variance

$$\hat{\sigma}^2(\boldsymbol{x}_{new}) = C_{mod}(\boldsymbol{x}_{new}, \boldsymbol{x}_{new}) - \mathbf{r}_{mod}(\boldsymbol{x}_{new})^t\mathbf{R}^{-1}\boldsymbol{r}_{mod}(\boldsymbol{x}_{new}) + (\boldsymbol{h}(\boldsymbol{x}_{new}) - \mathbf{H}^t\mathbf{R}^{-1}\boldsymbol{r}_{mod}(\mathbf{x}_{new}))^t(\mathbf{H}^t\mathbf{R}^{-1}\mathbf{H} + \mathbf{Q}_{prior}^{-1})^{-1} \times (\boldsymbol{h}(\boldsymbol{x}_{new}) - \mathbf{H}^t\mathbf{R}^{-1}\boldsymbol{r}_{mod}(\boldsymbol{x}_{new})). \quad (19)$$

We can make the same remarks as for Eq. (16). Similar to calibration, the limit when $\mathbf{Q}_{prior}^{-1} \to 0$ of the prediction in the prior information case is the prediction in the no prior information case.

II.D. Illustration of an Analytical Test Case

We illustrate the calibration and the prediction on an analytical test case. This test case is academic and allows the most important features of Gaussian process modeling to be understood.

We study the case in which the physical system is the function $x \to x^2$ on $[0, 1]$. The computer model is $f_{mod}(x, \boldsymbol{\beta}) = \beta_0 + \beta_1 x$. We assume that the covariance function of the model error is known and has the Gaussian form $C_{mod}(x - y) = \sigma^2\exp(-|x - y|^2/l_c^2)$, with $\sigma = 0.3$ and $l_c = 0.5$. There are three observations, noiseless, for experimental points 0.2, 0.5, and 0.8.

The results in the no prior information case are presented in Fig. 1. We first see that there is a negative correlation in the estimation of $\boldsymbol{\beta}$. This correlation can be interpreted. Indeed, if β_0, the value at 0 of the line $x \to \beta_0 + \beta_1 x$ is increased, then, for the line to remain close to the parabola $x \to x^2$, the slope of the line (β_1) must be decreased. Furthermore an important remark is that the

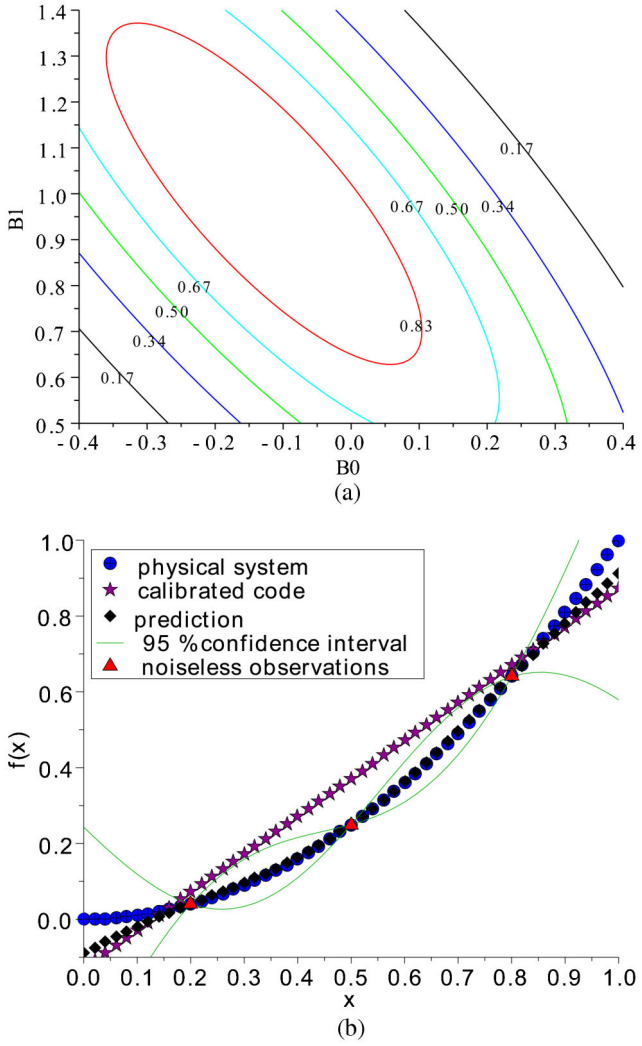

Fig. 1. Calibration and prediction in the no prior information case. (a) Iso-density curves of the probability density function for the estimation of $\boldsymbol{\beta}$, given by Eqs. (12) and (13). (b) Calibrated line [Eq. (12)], real parabola, prediction [Eq. (16)], and 95% confidence intervals of the form $[\hat{y}(\boldsymbol{x}_{new}) - 1.96\hat{\sigma}(\boldsymbol{x}_{new}), \hat{y}(\boldsymbol{x}_{new}) + 1.96\hat{\sigma}(\boldsymbol{x}_{new})]$.

calibrated line is above and does not go through the three observation points. This is surprising at first sight, all the more since the least-squares estimator of Sec. I would go through the three points. This is because, as shown in Eq. (16), the calibrated line is not intended to constitute a predictive model of the parabola. Indeed, it is completed by the inferred model error from the three observation points. We see in Fig. 1 that the prediction curve approximates almost perfectly the parabola. Let us also notice that in the extrapolation region ($0 \le x \le 0.2$ and $0.8 \le x \le 1$), the calibrated line approximates better the parabola than a line that would go between the three observation points.

Hence, the inference of the model error improves the prediction capability of the calibrated computer model. This is all the more true as the physical system is predicted closer to the experiments. In extrapolation, the model error cannot be precisely inferred from the available observations, and the inferred model error in Eq. (16) is hence very close to zero. Hence, in extrapolation, the prediction is made using the calibrated computer model only. This is as expected because when one cannot statistically improve the prediction of the computer model, a conservative choice is to rely only on physical knowledge. Finally, we see that the confidence intervals [whose lengths are four times the standard deviations Eq. (17), which correspond to 95% confidence] have length zero at the points where the noiseless observations are done, and that this length increases when one moves away from observation points. This shape of the confidence intervals is classical in kriging with noiseless observations.

We also consider the prior information case with

$$\boldsymbol{\beta}_{prior} = \begin{pmatrix} 0.2 \\ 1 \end{pmatrix} \text{and} \quad \mathbf{Q}_{prior} = \begin{pmatrix} 0.09 & 0 \\ 0 & 0.09 \end{pmatrix}.$$

The results for this case are shown in Fig. 2. Looking at Figs. 2c and 2d, we can see that, from the prior $\boldsymbol{\beta}$ to the posterior $\boldsymbol{\beta}$, the line goes substantially closer to the three observation points. Nevertheless, it is not as close as in the no prior information case. This is a classical case in the prior information case (as well as in Bayesian statistics), when the observations and the prior judgment are in disagreement, and the posterior $\boldsymbol{\beta}$ is a compromise between the observations and the prior judgment. Looking at Figs. 2a and 2b, we see that a negative correlation between the two components of $\boldsymbol{\beta}$ appears in the posterior law of $\boldsymbol{\beta}$.

Finally, the prediction of the physical system, and the confidence intervals, are similar to the no prior information case.

To conclude the illustration of analytical functions, we see that Gaussian process modeling has the potential to both improve the prediction capability of the computer model and correctly assess the resulting uncertainty. In Sec. III, we confirm this on the computer model FLICA 4 (Ref. 9), a thermal-hydraulic code relevant to core thermal-hydraulic transient and steady-state analysis. Before this, we give general practical recommendations concerning the use of Gaussian process modeling.

II.E. General Recommendations for Gaussian Process Modeling

The first important point is that, as stated in Sec. I, the method presented here does not address the complex field of code verification. As a consequence, discretization or numerical parameters, such as the length or volume of a node, shall not be considered as model parameters or treated by the present method.

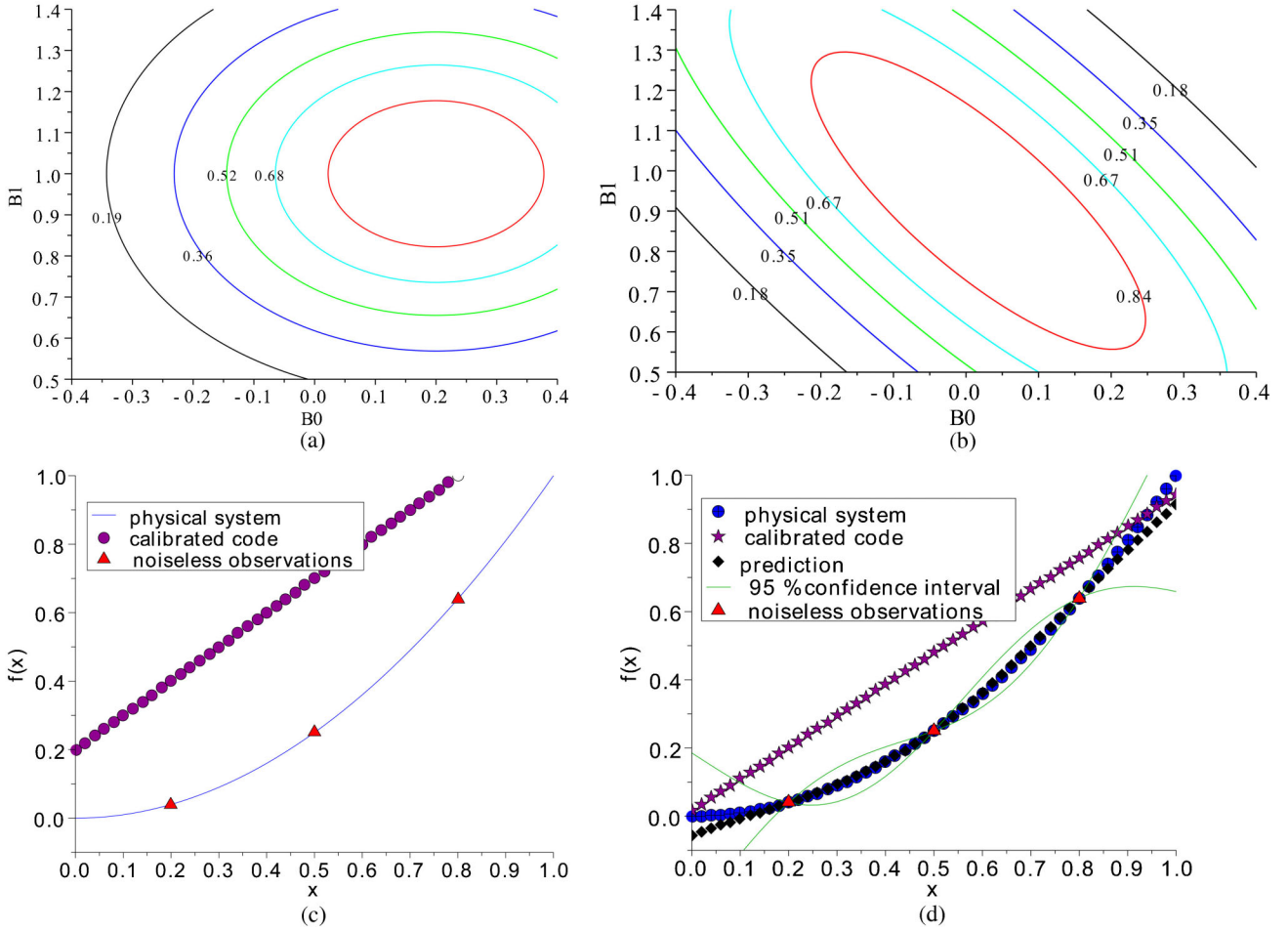

Fig. 2. Calibration and prediction in the prior information case. (a) Iso-density curves of the prior probability density function of $\boldsymbol{\beta}$. (b) Iso-density curves of the posterior probability density function of $\boldsymbol{\beta}$ given by Eqs. (14) and (15). (c) and (d) Calibrated line with (c) the prior and (d) the posterior [Eq. (14)] mean values for $\boldsymbol{\beta}$, real parabola, prediction [Eq. (18)], and 95% confidence intervals [Eq. (19)] of the form $[\hat{y}(\boldsymbol{x}_{new}) - 1.96\hat{\sigma}(\boldsymbol{x}_{new}), \hat{y}(\boldsymbol{x}_{new}) + 1.96\hat{\sigma}(\boldsymbol{x}_{new})]$.

Another important point is the linear approximation of Eq. (3). If the main objective is to achieve a precise enough prediction of the physical system, and not to calibrate the computer model, then it is not a problem if the computer model is not linear with respect to its model parameters. Indeed, the linear approximation boils down to modeling the Gaussian process Z in Eq. (1) as the model error of the linearized computer model in Eq. (3). In the prediction formulas (16) and (18), we see that the statistical correction can compensate for the linear approximation error of the code. This fact is confirmed in Sec. III for FLICA 4. The linear approximation yields a much cheaper method than similar nonlinear methods,[23,24] which may need to use Markov chain Monte Carlo methods, and possibly to approximate the computer model by a surrogate model in both the $\boldsymbol{x}$ and $\boldsymbol{\beta}$ domains. Now, if calibration in itself is one of the main objectives, one should act with caution with respect to the linear approximation. In this case, we advise running a sensitivity analysis first to check the linearity assumption (e.g., the Morris method[25]). If the linearity assumption is invalidated, then we recommend proceeding in two steps. First, a nonlinear calibration should be carried out, such as the least-squares calibration or a Bayesian calibration.[24] Then, the model parameters should be fixed to their calibrated values, or a very narrow prior, centered around these values, should be used, before using the present method.

Concerning the computation of the derivatives with respect to the model parameters $\boldsymbol{\beta}$, two cases are possible. First, the code can already provide them, by means of the adjoint sensitivity method.[2] Similarly, automatic differentiation methods can be used on the source file of the code and yield to a differentiated code.[26] If these kinds of methods are not available, finite differences are necessary to approximate the derivatives. Our main advice here is not to use a too small variation step. Indeed, on the one hand, if the code is approximately linear with respect to the model parameters, a too large variation step will

provide a good estimate of the derivatives anyway, whereas a too small variation step can yield numerical errors. On the other hand, if the code is not approximately linear, the linear approximation should not be used for calibration. For prediction, the model error compensates for the linear approximation error as well as for the error in calculating the derivatives.

The fourth important point is that extrapolation is not recommended. This is a general advice for all kriging models. The experimental results should be made in the prediction domain of interest. Hence, for example, kriging methods are not advisable to address scaling issues that intrinsically ask to extrapolate experimental results from one scale to another.

When dealing with more complex systems than the one of Sec. III, such as system–thermal hydraulics, one may deal with high-dimensional problems, either with respect to the number of experimental conditions (dimension of $\boldsymbol{x}$) or to the number of model parameters (dimension of $\boldsymbol{\beta}$). The dimension of $\boldsymbol{x}$ is a potential problem. A common rule of thumb for kriging models is that one should have $n \geq 10 \dim(\boldsymbol{x})$. Note that screening methods exist and allow selection of only the most impacting experimental conditions.[27] If the number of experiments is really too small compared to the number of experimental conditions, our opinion is that it is not possible to take into account the model error correctly, so that only the calibration should be carried out. If $\boldsymbol{\beta}$ is high-dimensional, we advise either to use a full Bayesian framework, as described in Sec. II.B, or to select only the most important model parameters (from physical expertise) and to fix the other model parameters at their nominal values. For example, in Sec. III, the less important parameters a_l, C_f, n, and d are fixed to their nominal values. In this case, the process modeling of the model error also compensates for the error made by freezing these parameters.

III. APPLICATION TO THE THERMAL-HYDRAULIC CODE FLICA 4

III.A. Presentation of FLICA 4 and the Experimental Results

The experiment consists of measuring the pressure drop in an ascending pressurized flow of liquid water through a tube that can be electrically heated. This paper focuses on the frictional pressure drop (ΔP_{fric}) in a single-phase flow.

III.A.1. Thermal-Hydraulic Code FLICA 4

The mathematical model for ΔP_{fric} is given by the local equation

$$\Delta P_{fric} = \frac{H}{2\rho D_h} G^2 f_{iso} f_h \ , \tag{20}$$

where

H = friction height

ρ = density

D_h = hydraulic diameter

G = flow rate

f_{iso}, f_h = friction coefficients, respectively, in the isothermal and heated flow regimes.

In Eq. (2), each quantity is local, and the equation is hence numerically integrated in space and time by FLICA 4. The isothermal regime is defined by the temperature of the liquid being uniformly equal to the wall temperature. On the other hand, the heated flow regime is characterized by a heat flux imposed on the test section and thus a varying liquid temperature. In this work, we focus on the single-phase case, and we study the isothermal and heated flow subcases.

The friction coefficient in the isothermal regime is

$$f_{iso} = \begin{cases} \dfrac{a_l}{\mathrm{Re}} & \text{if } \mathrm{Re} < \mathrm{Re_l} \ , \\ \dfrac{a_t}{\mathrm{Re}^{b_t}} & \text{if } \mathrm{Re_t} < \mathrm{Re} \ , \\ \dfrac{a_l}{\mathrm{Re}} \dfrac{\mathrm{Re}_t - \mathrm{Re}}{\mathrm{Re}_t - \mathrm{Re}_l} + \dfrac{a_t}{\mathrm{Re}^{b_t}} \dfrac{\mathrm{Re} - \mathrm{Re}_l}{\mathrm{Re}_t - \mathrm{Re}_l} & \text{if } \mathrm{Re_l} < \mathrm{Re} < \mathrm{Re_t} \ , \end{cases} \tag{21}$$

where

$\mathrm{Re} = GD_h/\mu$ = Reynolds number

μ = viscosity

$\mathrm{Re}_l, \mathrm{Re}_t$ = limiting values for the Reynolds numbers that are defined according to the literature and represent the limits of the transition regime between laminar and turbulent flows

a_l, a_t, b_t = parts of the model parameters of FLICA 4. They are the three components of the vector $\boldsymbol{\beta}$ of model parameters in the isothermal regime.

The friction coefficient in the heated flow regime is a correction factor expressed as

$$f_h = 1 - \frac{P_h}{P_w} \frac{C_f (T_w - T_b)}{1 + d\left(\frac{T_w + T_b}{2T_0}\right)^n} \ , \tag{22}$$

where

P_h, P_w = heated and wetted perimeters, respectively

T_w = wall temperature

T_b = bulk temperature

$T_0 = 100°C$ = normalization temperature

C_f, n, d = three components of the vector $\boldsymbol{\beta}$ of model parameters in the heated flow case.

Finally, note that tests with no heat flux (isothermal tests) result in $T_w = T_b$, therefore, the correction factor f_h is equal to 1, as expected.

III.A.2. Experimental Results

Several experimental tests have been conducted in order to calibrate the FLICA 4 friction model. These tests have been used in previous calibration studies. The database is composed of n_i measurements under isothermal conditions, and n_h measurements for heated tests. An experimental condition $\boldsymbol{x}$ consists of geometrical data (the channel width e, the hydraulic diameter D_h, and the friction height H_f) and thermal-hydraulic conditions (the outlet pressure P_o, the flow rate G_i, the wall heat flux ϕ_w, the inlet liquid enthalpy h_i^l, the thermodynamic title X_{th}^i, and the inlet temperature T_i). For each test, the pressure drop due to friction ΔP_{fric} is measured.

III.B. Settings for the Study

III.B.1. Objectives

We carry out the Gaussian process modeling method on FLICA 4 in the isothermal and heated flow regimes. We limit the calibration part of the study to the parameters a_t and b_t. That is, we enforce the parameter a_l of the isothermal model, and the parameters C_f, n, and d of the heat-correction model to their nominal values, computed in previous calibration studies. Indeed, the parameters a_t and b_t are the most influential parameters for FLICA 4.

We work in the prior information case [calibration given by Eq. (14)]. From previous calibration studies, we have $\boldsymbol{\beta}_{prior} = (0.22, 0.21)^t$. Parameter $\mathbf{Q}_{prior}$ corresponds to a 50% uncertainty and is chosen to be diagonal with diagonal vector $(0.11^2, 0.105^2)^t$. Hence, this prior is rather large, so that the calibration essentially depends on the experimental results.

An important point is that the two categories of experimental conditions (control and environment variables; see Sec. I) are not equally represented in the experimental results. The category of the control variables consists of the channel width e, the hydraulic diameter D_h, and the friction height H_f. The category of the environment variables consists of the outlet pressure P_o, the flow rate G_i, the wall heat flux ϕ_w, the liquid enthalpy h_i^l, the thermodynamic title X_{th}^i, and the inlet temperature T_i. The $n_i + n_h$ experiments are divided into eight campaigns. Within a campaign, the control variables remain constant, while the environment variables vary. Hence, we only dispose of eight different control variables triplet. This means that, from the point of view of the prediction given by the Gaussian process model Eq. (18), it is very unlikely that the prediction of the calibrated code is significantly improved when considering new control variables. We experienced that, when predicting for new control variables, the Gaussian process method does not damage the predictions given by the nominal calibration of FLICA 4, but it does not significantly improve it. However, as we see next, we can give significantly improved predictions for observed control variables and new environmental variables.

To conclude, this study follows the double objective of calibration and prediction, in the prior information case for the parameters a_t and b_t. Concerning the prediction, the objective is to predict for experienced control variables and new environment variables.

III.B.2. On the Different Covariance Functions

The environment and control variables listed above are not independent. Hence, it would be redundant to incorporate all of them into the covariance function. One possible minimal set of environment and control variables is the set $(G_i, P_o, h_i^l, \phi_\omega, H_f, D_h)$. For this set, we will use the covariance function C, with C being one of the four covariance functions of Sec. II.B.

To summarize, we represent the experimental conditions of an experiment by $\boldsymbol{x} = (G_i, \phi_w, h_i^l, P_o, H_f, D_h)$. The covariance function is $C_{mod}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = \sigma^2 C\left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\right)$, with C being either the exponential, the Matérn $\frac{3}{2}$, the Matérn $\frac{5}{2}$, or the Gaussian correlation functions of Sec. II.B. The hyperparameters to be estimated are the variance σ^2 and the six correlation lengths $l_{c,1}, \ldots, l_{c,6}$.

Finally, we consider that the covariance matrix of the measurement error process is $\mathbf{R}_{mes} = \sigma_{mes}^2 \mathbf{I}_n$, with $\sigma_{mes} = 150$ Pa provided by the experimentalists.

III.B.3. Cross Validation

It is well known, in the general framework of statistical prediction, that the quality of a predictor should not be evaluated on the data that helped to build it (Ref. 28, chap. 7). This is particularly true for the Gaussian process model, since it is based on the kriging equations, which yield an interpolation of the observations when there is no measurement error. When a rather limited number of observations is available, as is the case here, cross validation is a very natural method to assess the predictive capability of a prediction model. In our case, we are interested in the following two quality criteria,

root-mean-square error (RMSE) and interval of confidence (IC), for the Gaussian process predictor:

$$\mathrm{RMSE}^2 = \frac{1}{n}\sum_{i_c=1}^{n_c}\sum_{\boldsymbol{x}\in C_{i_c}}\left(\hat{y}_{\overline{C}_{i_c}}(\boldsymbol{x}) - y_{obs}(\boldsymbol{x})\right)^2 \tag{23}$$

and

$$\mathrm{IC} = \frac{1}{n}\sum_{i_c=1}^{n_c}\sum_{\boldsymbol{x}\in C_{i_c}} 1_{\left|\hat{y}_{\overline{C}_{i_c}}(\boldsymbol{x}) - y_{obs}(\boldsymbol{x})\right| \le 1.64\left(\hat{\sigma}(\boldsymbol{x})\right)_{\overline{C}_{i_c}}}, \tag{24}$$

where

$\overline{C}_{i_c}$ = set of experimental conditions and observations that is the union of the subsets $C_1, \ldots, C_{i_c-1}, C_{i_c+1}, \ldots, C_{n_c}$

$\hat{y}_{\overline{C}_{i_c}}(\boldsymbol{x})$, $(\hat{\sigma}(\boldsymbol{x}))_{\overline{C}_{i_c}}$ = posterior mean and standard deviation, respectively, of the predicted output at $\boldsymbol{x}$ given the experimental data in $\overline{C}_{i_c}$

and where

$$\left[\hat{y}_{\overline{C}_{i_c}}(\boldsymbol{x}) - 1.64\left(\hat{\sigma}(\boldsymbol{x})\right)_{\overline{C}_{i_c}}, \hat{y}_{\overline{C}_{i_c}}(\boldsymbol{x}) + 1.64\left(\hat{\sigma}(\boldsymbol{x})\right)_{\overline{C}_{i_c}}\right]$$

corresponds to a 90% confidence interval.

In Eqs. (23) and (24), we use a K-fold cross-validation procedure, with $K = 10$. To do this, we partition the set of n experiments into $n_c = 10$ subsets $C_1, \ldots, C_{n_c}$, with each subset being well distributed in the experimental domain. It is emphasized that at step i_c of the cross validation, the Gaussian process model is built without using the experimental results of the class C_{i_c}. Hence, the important point is that, in the computation of the posterior mean and variance of the observed value at $\boldsymbol{x}$, this observed value is unused for the estimation of the hyperparameters as well as for the prediction formula.

The cross validation presented here can yield a high computational cost because one has to repeat the hyperparameter estimation procedure n_c times. When these estimations are too costly, a simplified but approximate cross-validation (CV) procedure is possible in which the hyperparameters are estimated only once for all the CV steps. For this simplified version, cross validation is carried out only with respect to the prediction formulas of Eqs. (18) and (19). Let us note that, in this context, there exist formulas[29] that allow calculation of the result of the cross-validation procedure without actually calculating K times the prediction formulas Eqs. (16) through (19). These virtual CV formulas reduce even more the CV computational cost. Nevertheless, in our case, we are able to estimate the hyperparameters at each step of the cross validation. Indeed, we have a rather limited number n of experimental results [the computation of the restricted likelihood is $O(n^3)$].

III.C. Results

III.C.1. Results in the Isothermal Regime

In a first step, we consider the results in the isothermal and turbulent flow regime only. That is, the regime when $f_h = 1$ in Eq. (20) and when $\mathrm{Re} > \mathrm{Re}_t$ in Eq. (21). We have n_{it} experimental results.

The isothermal regime is characterized by no wall heat flux, $\phi_w = 0$. Hence, it is useless to include it in the covariance function because it is uniformly zero for all the experimental conditions. So, we only have five correlation lengths out of six to estimate, which are $l_{c,1}$, $l_{c,3}$, $l_{c,4}$, $l_{c,5}$, and $l_{c,6}$, corresponding to G_i, h_i^l, P_o, D_h, and H_f.

In Fig. 3, we plot, for the tenfold cross validation, the $n_c = 10$ posterior mean values of a_t and b_t for the four covariance functions of Sec. II.B. The conclusions are that the Gaussian process calibration does not change significantly the nominal values $a_t = 0.22$ and $b_t = 0.21$. Furthermore, we do not notice significant differences concerning the choice of the covariance function for the calibration. Finally, we can observe a high correlation in the posterior means of a_t and b_t. This is confirmed in the n_c posterior covariance matrix, where the correlation coefficient is larger than 0.95.

Concerning the prediction, we first compute the RMSE and IC criteria for the four covariance functions. Results are presented in Table I. The first comment is that the predictive variances of Eq. (19) are reliable because they yield rather precise 90% confidence intervals. This is also observed for kriging, e.g., Ref. 12. The second comment is that there is no significant difference between the different covariance functions. This may be due to the amplitude of the measurement error, which makes the problem of the regularity of the covariance function insignificant. It is shown in Ref. 11, Sec. 3.7, that in a particular asymptotic context, even a small measurement error can have a significant effect on prediction errors.

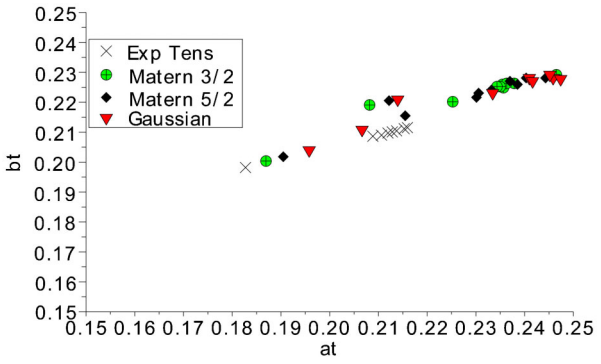

Fig. 3. Calibration in the isothermal regime: tenfold cross validation. Plot of the $n_c = 10$ posterior means [Eq. (14)] of a_t and b_t for the exponential, Matérn 3/2, Matérn 5/2, and Gaussian covariance functions of Sec. II.B.

TABLE I

Prediction Results in the Isothermal Regime*

Covariance Function	RMSE (Pa)	IC
Exponential	289.5	0.93
Matérn 3/2	296.2	0.92
Matérn 5/2	302.7	0.89
Gaussian	310.8	0.88

*RMSE and IC criteria of Eqs. (23) and (24) obtained with a tenfold CV procedure, for the covariance functions presented in Sec. II.B.

We now present more detailed results for the Matérn $\frac{3}{2}$ covariance function. We first compare the Gaussian process predictions with the predictions given by the calibrated code alone. With the same CV procedure, the RMSE criterion for the calibrated code alone is RMSE = 741 Pa. This is to be compared with an RMSE of ~300 Pa for the Gaussian process method. Hence, the inference of the model error process significantly improves the predictions of the code. We illustrate this in Fig. 4, where we plot, for each of the n_{it} observations, the predicted values and confidence intervals with the tenfold CV method. The plots are done with respect to the experiment index. This index has physical meaning because two experiments with successive indices are similar (for instance, the experiences of a given campaign have successive indices). We first see that Gaussian process modeling significantly reduces the prediction errors, and that the confidence intervals are reliable. Then, we observe a regularity in the plot of the prediction error for the calibrated code, especially for the largest indices. This regularity is not present anymore in the error of the Gaussian process method. The conclusion is that the Gaussian process method detects a regularity in the error of the calibrated code, and uses it to significantly improve its predictions.

In Table II, we show the $n_c = 10$ different estimations of $(\sigma^2, l_{c,1}, l_{c,3}, l_{c,4}, l_{c,5}, l_{c,6})$, for the different steps of the cross validation. The first conclusion is the singularity at steps 5 and 6 of the cross validation. The explanation is that, among the n_{it} experimental results, there are two singular points that have very similar experimental conditions but substantially different values for the quantity of interest. These two points are in CV classes 5 and 6. Hence, the estimation of the hyperparameters in the CV steps 1, 2, 3, 4, 7, 8, 9, 10, where this singularity is present in the data used for the estimation, is different from steps 5 and 6, where the singularity is absent. In Fig. 4, these two singular points yield the two largest prediction errors for the Gaussian process method. Indeed, when one of them is in the test group, the other is in the learning group. As the Gaussian process modeling principle is to assume a correlated model error, the quantity of interest of the singular point of the test group is (up to the measurement error) predicted by the quantity of interest of the singular point of the learning group.

The correlation lengths in Table II correspond to normalized experimental conditions varying between 0 and 1. Hence, the second conclusion is that the estimated correlation lengths are rather large, corresponding to rather large scales of variations of the model error, as discussed for Fig. 4. When an estimated correlation length is very large (larger than 10), it is equivalent to assuming that the model error is independent of the corresponding experimental condition. The third conclusion is that the estimations of the hyperparameters can vary moderately among the CV steps. This is an argument in favor of reestimating the hyperparameters at each step of the cross validation because this takes into account these variations.

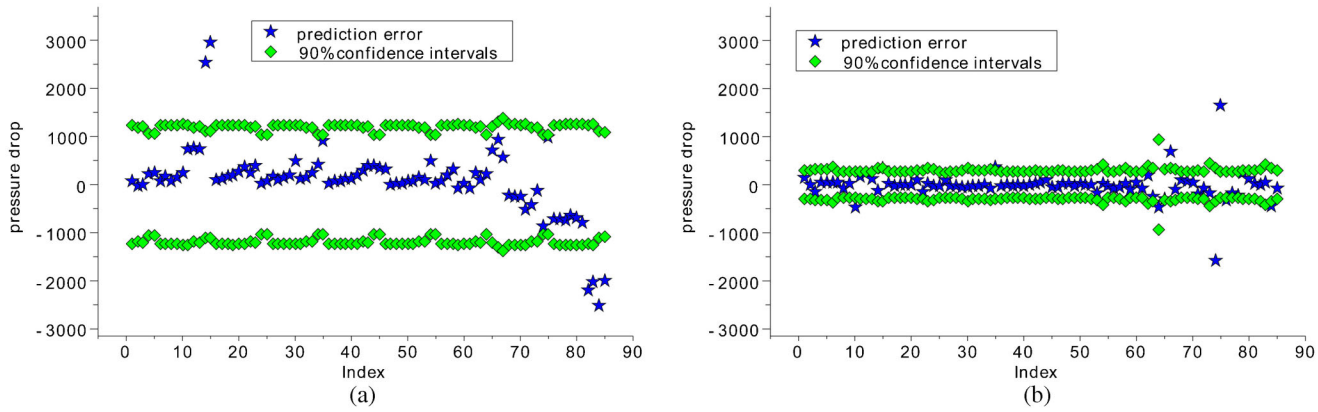

Fig. 4. Prediction errors [observed values minus predicted values, Eq. (18)] and 90% confidence intervals for these prediction errors, derived by (a) the calibrated thermal-hydraulic code FLICA 4 and (b) the Gaussian process method. The 90% confidence intervals are of the form $[-1.65\hat{\sigma}(\boldsymbol{x}), 1.65\hat{\sigma}(\boldsymbol{x})]$, with $\hat{\sigma}(\boldsymbol{x})$ given by Eq. (19). Plot with respect to the index of experiment.

TABLE II

Estimated Hyperparameters in the Isothermal Regime*

Cross-Validation Step	$\sqrt{\sigma^2}$	$l_{c,1}$	$l_{c,3}$	$l_{c,4}$	$l_{c,5}$	$l_{c,6}$
1	2220	2.3	4.0	100	0.40	53
2	2100	2.2	3.5	100	0.40	100
3	2088	2.1	3.8	100	0.39	100
4	2266	2.3	2.0	100	0.50	100
5	4491	3.4	100	24	1.36	100
6	1953	1.6	15	3.4	7.7	0.6
7	2385	2.4	4.6	100	0.44	100
8	2436	2.4	4.8	100	0.45	99
9	2331	2.4	4.2	100	0.43	100
10	2294	2.4	3.8	100	0.42	100

*Estimated correlation lengths for the Matérn $3/2$ covariance function of Sec. II.B, for the tenfold CV procedure.

Note that for the Gaussian process model to be used for new experimental conditions, the hyperparameters are to be reestimated with all the observations.

III.C.2. Results in the Single-Phase Regime

We now use all the experiments of the single-phase regime (isothermal and heated flow regimes), i.e., $n = n_i + n_h$ experiments. Hence, we estimate six correlation lengths for the six environment and control variables G_i, ϕ_w, h_i^l, P_o, D_h, and H_f.

Concerning the prediction, we first compute the RMSE and IC criteria for the four covariance functions. Results are presented in Table III. As in the isothermal case, we see that the predictive variances are reliable and that there is no significant difference among the four covariance functions. As for the isothermal regime, we present in more detail, the results for the Matérn $3/2$ covariance function.

With the same CV procedure, the RMSE criterion for the calibrated code alone is RMSE = 567 Pa. This is to be compared with an RMSE of $\sim$200 Pa of the Gaussian process method. Hence, the inference of the model error process significantly improves the predictions of the code, in the same way as in the isothermal regime. We illustrate this in Fig. 5, where we plot the same quantities as in Fig. 4. We obtain the same conclusion: The Gaussian process model detects a regularity in the error of the calibrated code, and uses it to improve its predictions.

III.C.3. Influence of the Linear Approximation

All the results above are obtained using the linear approximation of FLICA 4 with respect to a_t and b_t. We have implemented the equivalent of the calibration and prediction formulas of Eqs. (14) and (18), when FLICA 4 is not considered linear with respect to a_t and b_t (Ref. 3). Integrals in the a_t, b_t domain were calculated on a 5×5 grid, which, to avoid bias, was also used when the linear approximation of FLICA 4 was used. Using the same tenfold CV procedure as before, in the single-phase regime, we obtain RMSE = 197.8 with the linear approximation and RMSE = 196.9 without the linear approximation ($<1\%$ relative difference). The posterior means of a_t and b_t, along the different CV steps have a root-mean-square difference of 0.025 ($>10\%$ relative difference) between the cases where the linear approximation was made or not. Hence, this is an illustration of the general remark of Sec. II.E: If the computer model is nonlinear with respect to its calibration parameters, it is the model error with respect to the linearized computer model that is inferred. Thus, the predictions of the physical system are similar, whether or not the linear approximation is made.

TABLE III

Prediction Results in the Single-Phase Regime*

Covariance Function	RMSE (Pa)	IC
Exponential	202.2	0.95
Matérn $3/2$	196.2	0.95
Matérn $5/2$	196.9	0.95
Gaussian	199.5	0.94

*Same settings as in Table I.

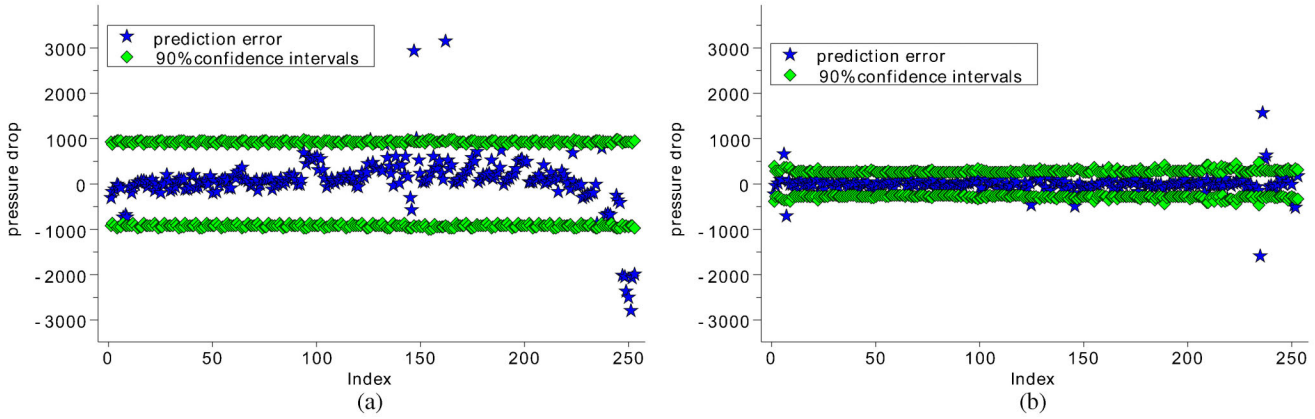

Fig. 5. Same settings as in Fig. 4, but in the single-phase regime.

IV. CONCLUSION

In this work, a Gaussian process modeling method has been presented for computer model calibration and improved prediction of the underlying physical system. It is based on modeling of the model error, which is the bias between the computer model and the physical system. A set of experimental results on the physical system is used, which enables inference of the model error for each new potential experimental point. As a result, an improved prediction for the value of the physical system, and an associated confidence interval, are provided.

The Gaussian process modeling method is carried out in two steps. In a first step, the covariance function of the model error is estimated, based on a comparison between experimental results and the computer model. In this paper, the estimation is carried out with the restricted maximum likelihood method, although the possibility of using other methods is discussed. This estimation step yields the main computational cost, since one needs to minimize a function involving a matrix inversion. The size of this matrix is equal to the number of experimental results.

Once the estimation is done, calibration and prediction can be carried out with closed-form matrix-vector formulas. The calibration is the computation of the best parameters for the computer model. Physical knowledge of the calibration parameters of the physical model can be taken into account in a Bayesian framework. The prediction is the computation of a predicted value and an associated confidence interval for each new potential experimental point. The predicted value is the sum of the calibrated computer model and a Gaussian inference of the model error. Hence, the calibrated computer model is completed by a statistical term. This statistical term is based on the experimental results and can significantly improve the predictions of the computer model. The closed-form linear algebra formulas for calibration and prediction rely on a linearization of the computer model, with respect to the model parameters, around a reference parameter. These formulas can still be used when the linear approximation does not hold, in which case, the calibration will be carried out on the linearized computer model, and the model error will incorporate the linear approximation error. It is shown that the linear approximation has no consequence on the prediction but shall be treated carefully if calibration is one of the main objectives. Gaussian process modeling of the model error can be carried out without linearization of the computer model,[23,24] but this yields a much more costly computation.

The method is applied to the friction model of FLICA 4, for which the data of several experimental campaigns are available. We evaluate the prediction capability of either the calibrated code alone or the Gaussian process modeling method. This evaluation is done rigorously using a tenfold cross validation on the experimental results. It is shown that the error of FLICA 4 can be divided by a factor between 2 and 3. We also study different covariance functions for the model error and come to the conclusion that, because of the measurement errors, the choice of the covariance function does not have a significant influence on the prediction capability in this case.

Based on this case study, we believe Gaussian process modeling of the model error to be promising in the field of computer model validation for nuclear engineering, by its ability to complete a computer model with a statistical inference of the model error.

An interesting area of research is the implementation of this method to functional output computer models, arising, for example, in the case of time-dependent problems. In the general context of kriging with functional output, two kinds of methods exist. The first solution is to consider a joint covariance structure, with respect to the inputs, and with respect to the functional output time or space parameter.[30] The second solution is to use a low-

dimensional representation of functional outputs, such as principal component analysis or wavelets, and to build a kriging model for each of the coefficients of the representation. The adaptation of these methods to Gaussian process modeling of the model error, in a nuclear engineering context, may motivate further research.

ACKNOWLEDGMENTS

We thank the two anonymous reviewers for their comments and suggestions, which helped to improve the quality of the manuscript.

REFERENCES

1. P. KNUPP and K. SALARI, *Verification of Computer Codes in Computational Science and Engineering*, Chapman & Hall/CRC, Boca Raton, Florida (2003).

2. D. G. CACUCI, *Sensitivity and Uncertainty Analysis. Theory*, Chapman & Hall/CRC, Boca Raton, Florida (2003).

3. S. WANG, W. CHEN, and K.-L. TSUI, "Bayesian Validation of Computer Models," *Technometrics*, **51**, 439 (2009); http://dx.doi.org/10.1198/TECH.2009.07011.

4. A. PETRUZZI and F. D'AURIA, "Thermal-Hydraulic System Codes in Nuclear Reactor Safety and Qualification Procedures," *Science and Technology of Nuclear Installations*, Vol. 2008 (2008).

5. D. G. CACUCI and M. IONESCU-BUJOR, "Best-Estimate Model Calibration and Prediction Through Experimental Data Assimilation—I: Mathematical Framework," *Nucl. Sci. Eng.*, **165**, 18 (2010).

6. A. PETRUZZI, D. G. CACUCI, and F. D'AURIA, "Best-Estimate Model Calibration and Prediction Through Experimental Data Assimilation—II: Application to a Blowdown Benchmark Experiment," *Nucl. Sci. Eng.*, **165**, 45 (2010).

7. T. J. SANTNER, B. J. WILLIAMS, and W. I. NOTZ, *The Design and Analysis of Computer Experiments*, Springer, New York (2003).

8. M. KENNEDY and A. O'HAGAN, "Bayesian Calibration of Computer Models," *J. R. Stat. Soc.: Ser. B (Stat. Method.)*, **63**, 425 (2001); http://dx.doi.org/10.1111/1467-9868.00294.

9. I. TOUMIA et al, "FLICA-4: A Three-Dimensional Two-Phase Flow Computer Code with Advanced Numerical Methods for Nuclear Applications," *Nucl. Eng. Des.*, **200**, 139 (2000); http://dx.doi.org/10.1016/S0029-5493(99)00332-5.

10. C. E. RASMUSSEN and C. K. I. WILLIAMS, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Massachusetts (2006).

11. M. L. STEIN, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York (1999).

12. B. A. LOCKWOOD and M. ANITESCU, "Gradient-Enhanced Universal Kriging for Uncertainty Propagation," *Nucl. Sci. Eng.*, **170**, 168 (2012).

13. P. ABRAHAMSEN, "A Review of Gaussian Random Fields and Correlation Functions," Technical Report, Norwegian Computing Center (1997).

14. A. DE CRÉCY, "Determination of the Uncertainties of the Constitutive Relationships of the Cathare 2 Code," *Proc. Int. Mtg. Mathematical Methods for Nuclear Applications (M&C)*, Salt Lake City, Utah, September 2001, American Nuclear Society (2001).

15. T. KAWANO et al, "Evaluation and Propagation of the ^{239}Pu Fission Cross-Section Uncertainties Using a Monte Carlo Technique," *Nucl. Sci. Eng.*, **153**, 1 (2006).

16. R. BENASSI, J. BECT, and E. VAZQUEZ, "Bayesian Optimization using Sequential Monte Carlo," *Proc. Learning and Intelligent Optimization Conf. (LION6)*, Paris, France, January 16–20, 2012.

17. K. V. MARDIA and R. J. MARSHALL, "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression," *Biometrika*, **71**, 135 (1984); http://dx.doi.org/10.1093/biomet/71.1.135.

18. H. ZHANG and Y. WANG, "Kriging and Cross Validation for Massive Spatial Data," *Environmetrics*, **21**, 290 (2010); http://dx.doi.org/10.1002/env.1023.

19. S. SUNDARARAJAN and S. S. KEERTHI, "Predictive Approaches for Choosing Hyperparameters in Gaussian Processes," *Neural Computation*, **13**, 1103 (2001); http://dx.doi.org/10.1162/08997660151134343.

20. N. CRESSIE and S. N. LAHIRI, "The Asymptotic Distribution of REML Estimators," *J. Multivariate Anal.*, **45**, 217 (1993); http://dx.doi.org/10.1006/jmva.1993.1034.

21. D. A. HARVILLE, "Bayesian Inference for Variant Components Using Only Error Contrasts," *Biometrika*, **61**, 383 (1974); http://dx.doi.org/10.1093/biomet/61.2.383.

22. J. SACKS et al, "Design and Analysis of Computer Experiments," *Stat. Sci.*, **4**, 409 (1989); http://dx.doi.org/10.1214/ss/1177012413.

23. M. J. BAYARRI et al, "A Framework for Validation of Computer Models," *Technometrics*, **49**, *2*, 138 (2007); http://dx.doi.org/10.1198/004017007000000092.

24. D. HIGDON et al, "Combining Field Data and Computer Simulations for Calibration and Prediction," *SIAM J. Sci. Comput.*, **26**, 448 (2004); http://dx.doi.org/10.1137/S1064827503426693.

25. M. D. MORRIS, "Factorial Sampling Plans for Preliminary Computational Experiments," *Technometrics*, **33**, *2*, 161 (1991); http://dx.doi.org/10.1080/00401706.1991.10484804.

26. L. HASCOËT and V. PASCUAL, "Tapenade 2.1 User's Guide," Technical Report 0300, INRIA (2004).

27. A. MARREL et al, "An Efficient Methodology for Modeling Complex Computer Codes with Gaussian Processes," *Comput. Stat. Data Anal.*, **52**, 4731 (2008); http://dx.doi.org/10.1016/j.csda.2008.03.026.

28. T. HASTIE, R. TIBSHIRANI, and J. FRIEDMAN, *The Elements of Statistical Learning*, Springer, New York (2008).

29. O. DUBRULE, "Cross Validation of Kriging in a Unique Neighborhood," *Math. Geol.*, **15**, 687 (1983); http://dx.doi.org/10.1007/BF01033232.

30. J. ROUGIER, Efficient Emulators for Multivariate Deterministic Functions," *J. Comput. Graph. Stat.*, **17**, 827 (2008); http://dx.doi.org/10.1198/106186008X384032.