Weierstraß-Institut

für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Technical Report ISSN 1618-7776

Calibration methods for gas turbine performance models

Jürgen Borchardt, Peter Mathé, Galina Printsypar

submitted: December 20, 2016

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mails: juergen.borchardt@wias-berlin.de
peter.mathe@wias-berlin.de
galina.printsypar@wias-berlin.de

No. 16

Berlin 2017

2010 *Mathematics Subject Classification.* 65-05.

Key words and phrases. model calibration, nonlinear least squares, Bayesian calibration, LASSO.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

ABSTRACT. The WIAS software package BOP is used to simulate gas turbine models. In order to make accurate predictions the underlying models need to be calibrated. This study compares different strategies of model calibration. These are the deterministic optimization tools as nonlinear least squares (MSO) and the sparsity promoting variant LASSO, but also the probabilistic (Bayesian) calibration. The latter allows for the quantification of the inherent uncertainty, and it gives rise to a surrogate uncertainty measure in the MSO tool. The implementation details are accompanied with a numerical case study, which highlights the advantages and drawbacks of each of the proposed calibration methods.

CONTENTS

1. INTRODUCTION

Engineering practice widely uses computer models. The accuracy of such computer models is usually defined by *model validation*, determining the degree to which the simulations reproduce the real behavior of the engineering systems. In particular, model updating matches model parameters with physical observations (measurements) utilizing mathematical means. Accurate predictions from simulations lie at the heart of the design, development, field support, and tendering activities. Specifically, we are concerned with the mathematical models of gas turbines (GT), which provide guarantee performance values to prospective customers. “Therefore, the engine performance model needs to be both predictive and stochastic. The uncertainty of the GT performance model has to be clearly derived based on all the available information and the guarantee process has to be based on a rigorous statistical analysis.”[1]

Gas turbine models are highly nonlinear, and the parameter calibration for these models is a complex problem that involves optimization in a high-dimensional space. Considerable computational efforts are needed to get a sufficient solution of this problem. Traditional approaches for GT model calibration do not consider uncertainties and do not provide uncertainty quantification of the estimated parameters. Several algorithms have been developed aiming to improve the traditional calibration methodology. The majority of these algorithms is probabilistic matching, e.g. weighted least squares, maximum likelihood estimates, Bayesian inference methods, we refer to [13] for a general account on GT simulation. While traditional approaches are deterministic, the need for quantification of the inherent uncertainties led to the use of Bayesian approaches for model calibration. The seminal work in this direction is [5], and the challenges

[1] see the introduction from [1] for more details.

in the calibration of computer models from the engineering point of view are highlighted in [14]. A comparison of deterministic and Bayesian methods for calibration is given in [16].

The present study aims to compare the deterministic and stochastic calibration techniques implemented in an in-house software package BOP for steady state as well as transient problems, cf. [2]. Both the deterministic and Bayesian approaches are implemented as add-ons of our process simulation tool BOP. This software package is especially tailored to the solution of large-scale problems from industrial process engineering, specifically the GT performance problem. Among other things, the simulation concept of BOP enables an efficient implementation of methods that require a large number of steady state solutions. These techniques offer a speedup for both parameter calibration add-ons, deterministic and Bayesian.

Special emphasis is on the uncertainty quantification obtained by these calibration methods. If deterministic approaches are used for parameter calibration, the quantification of the uncertainty of the calibrated gas turbine model is not clearly derived. However, a linearization of the model around the calibrated model parameters may be used to quantify the uncertainty. We shall highlight the relation of this linearization to the posterior covariance as obtained by the Bayesian approach.

In this report we compare the two calibration approaches. First, in Section 2 we derive the deterministic and probabilistic optimization methods, which use additional scaling parameters (weights) in order to tune the performance of the calibration. Implementation details for each method are given in Section 3. The main part concerns the simulation of a model with 5 calibration parameters and 21 experimental observations using the BOP tool (Section 4). In particular, we study how the additional tuning parameters influence the calibration results. Then, we compare the fits as obtained by either of the methods under consideration.

2. Calibration problem: methodology

In this section we introduce the calibration problem and we derive the three methods to be considered. The first one is a classical penalized least squares approach, the second one is a sparsity promoting modification, which is called LASSO, and the third one is a Bayesian calibration of the full model.

2.1. **Problem formulation.**

The gas turbine performance is modeled by a system of nonlinear equations. The model input consists of operating conditions $X \in \mathbb{R}^l$ and unknown model parameters $\theta \in \Theta \subset \mathbb{R}^n$. The model output is the experimental observations $Y \in \mathbb{R}^m$. Then, the nonlinear model for the gas turbine performance yields

$$Y = F(\theta, X) + \varepsilon, \tag{1}$$

where $F : \Theta \times \mathbb{R}^l \to \mathbb{R}^m$, $\varepsilon \in \mathbb{R}^m$ in the experimental error. The function F does not have an explicit representation and is given by means of the simulation tool BOP [2]. From now on we ignore the operating conditions X in the problem formulation (1) in order to simplify the notation.

The idea of the calibration problem is to adjust the model parameters θ such that the experimental observations Y are explained in the best way. Since the model output has some experimental error that is unknown, the common way to formulate the calibration problem is using the *least*

squares approach, thus the least squares fit is obtained by solving the minimization problem

$$\left\|F(\theta) - Y^0\right\|_{2,Y}^2 \longrightarrow \min_{\theta\in\Theta}. \tag{2}$$

Here, Y^0 is the measured output. However, this minimization does not take into account knowledge about the parameters θ. To account for this knowledge, the minimization problem (2) can be extended by a constraint optimization or by its Lagrange relaxation.

Remark 1. *In this report we are concerned with a single measurement point Y^0, i.e., with measurements from only one experiment. Experiments are usually performed several times with the same or different operational conditions X. Methods that are discussed below are easily generalized to the case of multiple measurement points. We would have $Y \in \mathbb{R}^{l\times m}$ corresponding to l experiments, and the minimization problem* (2) *would turn into*

$$\frac{1}{l}\sum_{k=1}^{l}\left\|F(X_k,\theta) - Y_k^0\right\|_{2,Y}^2 \to \min_{\theta\in\Theta}.$$

2.2. **Penalized least squares.** Here, we consider the following generalized minimization problem instead of (2)

$$\beta\left\|F(\theta) - Y^0\right\|_{2,Y}^2 + (1-\beta)\left\|\theta - \theta^0\right\|_{2,\theta}^2 \to \min_{\theta\in\Theta}. \tag{3}$$

Above, we use a weighted Euclidean norm

$$\left\|F(\theta) - Y^0\right\|_{2,Y}^2 = \left\langle \left(F(\theta) - Y^0\right), (m\Sigma_Y)^{-1}\left(F(\theta) - Y^0\right)\right\rangle, \tag{4}$$

where the matrix $\Sigma_Y = \operatorname{diag}\left(\sigma_{Y,i}^2\right)$ is the diagonal matrix with variances $\sigma_{Y,i}^2$ on the diagonal. Notice that the weights take into account the dimensionality of the response vector Y^0 in a normalizing way. Similarly, the Euclidean norm $\|\theta - \theta^0\|_{2,\theta}^2$ is defined, where θ^0 is the initial parameter set.

In addition, we introduce the weighting parameter $0 < \beta < 1$, which allows to control the importance of the fit. Having β close to 1 puts emphasis on the fit for the measurements, whereas β close to 0 puts emphasis on the initial parameter set θ^0. To have control over the calibration process, we introduce an additional scaling parameter T in the minimization problem (3). We thus arrive at

$$\frac{\beta}{T}\left\|F(\theta) - Y^0\right\|_{2,Y}^2 + \frac{1-\beta}{T}\left\|\theta - \theta^0\right\|_{2,\theta}^2 \to \min_{\theta\in\Theta}. \tag{5}$$

This scaling parameter T does not change the solution of the least square problem, but gains significance when interpreting (5) in a Bayesian context.

2.3. **LASSO.** The term *LASSO* refers to 'least absolute shrinkage and selection operator' [12]. The main idea of the method, motivated by the engineering point of view, is to change the initial parameter set θ^0 only in those coordinates where it is necessary in view of the given measurement Y^0. This approach is called *sparsity promoting minimization*, and the corresponding minimization problem is formulated as

$$\frac{\beta}{T}\left\|F(\theta) - Y^0\right\|_{2,Y}^2 + \frac{1-\beta}{T}\left\|\theta - \theta^0\right\|_{1,\theta} \to \min_{\theta\in\Theta}, \tag{6}$$

hence, the penalty is considered in L_1-norm instead of the Euclidean norm from (5).

2.4. **Bayesian calibration.** The above deterministic optimization algorithms yield minimizers, but the inherent uncertainty remains unclear. Therefore, we also consider Bayesian methods, which in addition to computing the minimizer allow to quantify the uncertainty. The functional in the minimization problem (5) has a Bayesian interpretation which is used to design a stochastic method to find the minimizer.

Suppose that the experimental observations Y and the parameters θ are random. Then, we consider the conditional probability $\mathbb{P}[Y|\theta]$ of Y given the parameters θ, which is implicitly given by (1). We are interested in $\mathbb{P}[\theta|Y]$ of θ, which is related to the sampling distribution $\mathbb{P}[Y|\theta]$ through the Bayesian formula as

$$\mathbb{P}(\theta|Y) = \frac{\mathbb{P}(Y|\theta)\,\pi(\theta)}{\mathbb{P}(Y)} \propto \mathbb{P}(Y|\theta)\,\pi(\theta), \tag{7}$$

where $\propto$ means "proportional to", and the probability π is called a prior distribution. If we now specify the sampling distribution that describes how we obtain the data Y in the model (1) as Gaussian, i.e., the error distribution ε is Gaussian, $\varepsilon \sim \mathcal{N}\left(0, \frac{Tm}{2\beta}\Sigma_Y\right)$, then the sampling distribution takes the form

$$\begin{aligned} f_{Y|\theta}(y) &\propto \exp\left(-\frac{1}{2}\left\langle (F(\theta)-y), \left(\frac{Tm}{2\beta}\Sigma_Y\right)^{-1}(F(\theta)-y)\right\rangle\right) \\ &= \exp\left(-\frac{\beta}{T}\left\|F(\theta)-y\right\|_{2,Y}^2\right). \end{aligned}$$

Similarly, we obtain for a Gaussian prior distribution of the form $\theta \sim \mathcal{N}\left(\theta^0, \frac{Tn}{2(1-\beta)}\Sigma_\theta\right)$ that

$$f(\theta) \propto \exp\left(-\frac{1-\beta}{T}\left\|\theta-\theta^0\right\|_{2,\theta}^2\right).$$

Thus, the product on the right in (7) takes the form of the exponential of the functional in (5) to be minimized. We draw two conclusions. First, within the Gaussian–Gaussian framework, the minimizer (5) maximizes the posterior distribution $f_{Y|\theta}f(\theta)$. Therefore, it is called the "maximum a posteriori" (MAP) estimator. Secondly, the parameters β and T have a probabilistic interpretation. Large values of T yield widely spread sampling and prior distributions, and hence the posterior will spread similarly. Small values of $\beta > 0$ will spread the sampling variance relative to the prior variance, which will have the desired impact as described in § 2.2.

Within the Bayesian approach the goal is to sample from the posterior distribution

$$f_{\theta|Y^0}(\theta) \propto f_{Y^0|\theta}(Y^0)f(\theta), \quad \theta \in \mathbb{R}^n. \tag{8}$$

The normalizing factor $\mathbb{P}(Y^0)$ in (7), which is the probability to see the measurement Y^0, is unknown. However, in case of a linear response function F in (1), an explicit description of the posterior distribution exists: the posterior distribution is again Gaussian, and its covariance can be computed exactly. Since the function F is linear, we represent $F(\theta)$ at any $\theta \in \Theta$ as $F(\theta) = F(\theta^0) + J(\theta-\theta^0)$ with J denoting the Jacobian matrix, and the posterior covariance

reads

$$\begin{aligned}\mathrm{cov}_{\mathrm{post}} &= \left(J^T \left(\frac{Tm}{2\beta} \Sigma_Y \right)^{-1} J + \left(\frac{Tn}{2(1-\beta)} \Sigma_\theta \right)^{-1} \right)^{-1} \\ &= T \left(\frac{2\beta}{m} J^T \Sigma_Y^{-1} J + \frac{2(1-\beta)}{n} \Sigma_\theta^{-1} \right)^{-1}. \end{aligned} \tag{9}$$

Moreover, the posterior mean equals the MAP estimator. On the other hand, in case of a nonlinear response function F, the posterior is no longer Gaussian, and the posterior mean may differ from the MAP. Therefore, the implementation of the Bayes algorithm must take into account that we need to sample from the posterior density (8) without knowing the normalizing constant.

The calibration techniques as described above are divided into deterministic and stochastic methods. These methods are implemented in the simulation tool BOP, see [2]. A preliminary discussion on the comparison between several calibration approaches in the context of gas turbine performance was given in [1].

3. Details on the calibration methods

In this section we discuss details of the implementation of the different calibration methods.

3.1. **Mean square optimization (MSO).** The deterministic method solves the least squares problem (5) to find optimal parameters θ for given measurement data Y. We use the damped Levenberg-Marquardt algorithm, which is an iterative method used to solve nonlinear least square problems. It originates in the studies [6, 7], and is described in [9]. First, we replace $F(\theta)$ by a linearized model and introduce a trust region with the radius $\Delta > 0$ (for more details see [9]):

$$\frac{\beta}{T} \left\| Js + r_Y^{k-1} \right\|_{2,Y}^2 + \frac{1-\beta}{T} \left\| s + r_\theta^{k-1} \right\|_{2,\theta}^2 \to \min_{\|s\| \in \Delta}, \tag{10}$$

where J is the Jacobian matrix, $r_Y^{k-1} = F(\theta^{k-1}) - Y^0$, $r_\theta^{k-1} = \theta^{k-1} - \theta^0$, s is the update at the current iteration $\theta^k = \theta^{k-1} + s$, and k is the iteration number.

The introduced linearization $F(\theta^k) \approx F(\theta^{k-1}) + Js$ provides a good approximation only when the update s is small. The trust region makes sure that the current update s is small enough for the approximation. Then, at each iteration step we apply a trust region around the current iterate θ^k by introducing an additional term in the minimization problem:

$$\frac{\beta}{T} \left\| Js + r_Y^{k-1} \right\|_{2,Y}^2 + \frac{1-\beta}{T} \left(\left\| s + r_\theta^{k-1} \right\|_{2,\theta}^2 + \lambda \left\| s \right\|_{2,\theta}^2 \right) \to \min, \tag{11}$$

where λ is the damping factor. Hence, the corresponding normal problem yields

$$\left[\frac{\beta}{m} J^T \Sigma_Y^{-1} J + \frac{(1-\beta)}{n} (1+\lambda) \Sigma_\theta^{-1} \right] s = \frac{\beta}{m} J^T \Sigma_Y^{-1} r_Y^{k-1} + \frac{(1-\beta)}{n} \Sigma_\theta^{-1} r_\theta^{k-1}, \tag{12}$$

$$\theta^k = \theta^{k-1} + s. \tag{13}$$

Large values of the damping factor λ stabilize the system (12), (13) and yield small values of the update s. We do not look explicitly for the damping factor λ corresponding to the trust region with

the radius Δ. Instead, we use the algorithm of Moré [15], which looks for the smallest values of λ which provide good approximation of the continuous problem.

Then, using the following normalization

$$\tilde{\theta} = \Sigma_\theta^{-\frac{1}{2}} \left(\theta - \theta^0\right), \qquad \tilde{Y} = \Sigma_Y^{-\frac{1}{2}} \left(Y - Y^0\right), \qquad \tilde{J} = \Sigma_Y^{-\frac{1}{2}} J \Sigma_\theta^{\frac{1}{2}}, \tag{14}$$

the system of equation (12) and (13) is dimensionless, and it can be written as

$$\left[\frac{\beta}{m}\tilde{J}^T\tilde{J} + \frac{(1-\beta)}{n}(1+\lambda)I\right]\tilde{s} = \frac{\beta}{m}\tilde{J}^T\tilde{Y}^{k-1} + \frac{(1-\beta)}{n}\tilde{\theta}^{k-1}, \tag{15}$$

$$\tilde{\theta}^k = \tilde{\theta}^{k-1} + \tilde{s}; \tag{16}$$

where $\tilde{Y}^{k-1}$ is evaluated using $Y^{k-1} = F(\tilde{\theta}^{k-1}\sigma_\theta + \theta^0)$ and the corresponding dimensionless variables in (14).

Given some $\epsilon > 0$, the stopping criterion for the iterative process is

$$\left\| \frac{\beta}{Tm}\tilde{J}^T\tilde{Y}^k + \frac{(1-\beta)}{Tn}\tilde{\theta}^k \right\| < \epsilon.$$

The deterministic optimization method cannot take into account uncertainty information, at a first glance. However, as was seen in § 2.4, for small changes in θ we may assume that the map F is near to linear, and the approximation of the covariance matrix can be computed using (9).

For more details on the implemented algorithm we refer to the full description of the BOP add-on MSO [11], and the documentation [2].

3.2. **LASSO: sparse mean square optimization.** To enhance predictability and to select significant variables in the mean square optimization we also implement the $L1$-norm regularization as given in (6), as described in [4]. There, the weighted $L1$-norm is reduced to the $L2$-norm as follows

$$\|x\|_{1,x} = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i|}{\sigma_{x_i}^2} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i}{\sigma_{x_i}|x_i|^{1/2}}\right)^2 = \left\|M_x^{1/2}x\right\|_{2,x}^2, \tag{17}$$

where $M_x = \operatorname{diag}\left\{|x_i|^{-1}\right\}_{i=1\ldots n}$. Hence, the minimization problem (6) is equivalent to

$$\frac{\beta}{T}\left\|F(\theta) - Y^0\right\|_{2,Y}^2 + \frac{1-\beta}{T}\left\|M_{\theta-\theta^0}^{1/2}(\theta-\theta^0)\right\|_{2,\theta}^2 \to \min_{\theta\in\Theta}. \tag{18}$$

The iterative process, with $r_\theta^{k-1} = \theta^{k-1} - \theta^0$ and $r_Y^{k-1} = F(\theta^{k-1}) - Y^0$, is given by

$$\left[\frac{\beta}{m}J^\top\Sigma_Y^{-1}J + \frac{(1-\beta)}{n}(M_{r_\theta^{k-1}} + \lambda)\Sigma_\theta^{-1}\right]s = \frac{\beta}{m}J^\top\Sigma_Y^{-1}r_Y^{k-1} + \frac{(1-\beta)}{n}M_{r_\theta^{k-1}}\Sigma_\theta^{-1}r_\theta^{k-1}, \tag{19}$$

$$\theta^k = \theta^{k-1} + s. \tag{20}$$

If, at some iteration k and for some component i, the deviation $\left|\left(r_\theta^k\right)_i\right|$ is smaller than a tolerance δ, then we set $\theta_i^l = \theta_i^0$ for all $l > k$ and exclude this component in further iterations. Then,

we define for $i, j = 1 \dots n$

$$\left(I_\theta^k\right)_{ij} = \begin{cases} 1, \text{ if } i = j \text{ and } \left|\left(r_\theta^k\right)_i\right| \geq \delta; \\ 0, \text{ otherwise.} \end{cases}$$

When the algorithm converges, the following condition is satisfied

$$\left\| I_\theta^k \left(\frac{\beta}{Tm} J^T \Sigma_Y^{-1} r_Y^k + \frac{(1-\beta)}{Tn} \Sigma_\theta^{-1} \text{sgn}\left(r_\theta^k\right) \right) \right\| < \epsilon.$$

3.3. **Bayesian approach.** Since in the model (1) for the gas turbine performance the function F is nonlinear, we must use an asymptotic sampling scheme, i.e., a Markov Chain Monte Carlo (MCMC) sampling method. Here we describe the *Metropolis-Hastings Markov chain algorithm*, which simulates an acceptance-rejection scheme. The Metropolis-Hastings Markov chain computes a sequence of random vectors $\theta^0 \dots \theta^N$, which are asymptotically (for large N) distributed with given posterior density (8). To account for the *burn-in* of the Markov chain we skip the first $N_0 - 1$ computed values to ensure that the remaining samples $\theta^{N_0} \dots \theta^N$ are close to the limiting distribution, which is the desired posterior distribution.

A sketch of the Metropolis-Hastings Markov chain algorithm is shown below (see Algorithm 1). At each step, we make a new proposal $\tilde{\theta}^k$, and then, we either accept or reject it based on the Metropolis-Hastings criterion. In case of a rejection, the next sample equals the previous one.

Data: Choose starting vector θ^0 in hyper-rectangle and sample size N
for $k = 1 \dots N$ **do**
  Propose new vector $\tilde{\theta}^k$ from proposal distribution $Q(\tilde{\theta}^k | \theta^{k-1})$;
  Compute $\mathbb{P}\left(\tilde{\theta}^k\right)$, $\mathbb{P}\left(\theta^{k-1}\right)$, and set $\alpha^k \leftarrow \mathbb{P}\left(\tilde{\theta}^k\right) / \mathbb{P}\left(\theta^{k-1}\right)$;
  if $\alpha^k \geq 1$ **then**
    accept $\theta^k \leftarrow \tilde{\theta}^k$;
  else
    accept $\theta^k \leftarrow \tilde{\theta}^k$ with probability α^k;
    otherwise **reject** and set $\theta^k \leftarrow \theta^{k-1}$;
  end
end

Algorithm 1: Metropolis-Hastings sampling

The described Markov chain in Algorithm 1 is determined by the choice of *a proposal distribution* Q. Common choices for the proposal distributions are uniform, Gaussian, or triangular. We use a Gaussian proposal distribution centered around the previously accepted sample θ^{k-1} with a variance controlled by *a fraction parameter* δ. During the acceptance-rejection iterations, we adapt the fraction δ to meet an asymptotically optimal averaged acceptance ratio $\alpha \sim 0.25$ (see [10]). For more details and discussions on the Bayesian algorithm see [8] and the documentation [2].

4. Numerical experiment with five calibration parameters

The numerical experiments are based on a specific GT model, which is first briefly outlined, cf. § 4.1. In particular, we indicate how a normalization is used to make the representation of the

results easy. We also discuss how the LASSO achieves a reasonable calibration accuracy but changes the model parameters only in a sparse manner, cf. § 4.2. Then, we study the impact of the additional tuning parameters: the weight β, which balances between the importance of the measurement accuracy and the model confidence, and the 'temperature' T, which influences the efficiency of the Bayesian approach. Since the MSO and Bayesian methods solve the same minimization problem (5), the impact of these parameters is similar. Thus, to avoid repetitions and overload of the numerical simulation, we present the dependence study for the weight β for the MSO method, cf. § 4.3, and for the scaling T for the Bayesian method, cf. § 4.4. Finally, we compare the MSO and Bayesian methods for some selected parameters, cf. § 4.5.

4.1. **Setup.** Using the mathematical model for the gas turbine simulation, we set up a calibration problem with $n = 5$ model parameters and an experimental measurement for $m = 21$ output parameters. For the model parameters θ_i, $i = 1 \dots n$, we provide initial values θ_i^0 and intervals $[\theta_i^0 - d_{\theta_i}; \theta_i^0 + d_{\theta_i}]$ where they are allowed to vary (Table 1). For the experimental observations Y_i, $i = 1 \dots m$, we provide measured values Y_i^0 with their standard deviations $\sigma_{Y,i}$ (Table 2).

In order to be able to use the model for different environmental conditions, so called ANSYN factors are introduced in the model. In our case these are parameters "A", "B", and "E" in Table 1. "The ANalysis by SYNthesis (ANSYN) technique is a standard performance analysis method widely used in industry. It is currently used to evaluate engine performance from tests and to derive correcting factors (i.e. ANSYN factors) that modify certain parameters of the components' characteristics in order to allow the reproduction of real engine behavior by means of a synthesis calculation" [3].

#	*Name*	*Initial value* θ_0	*Interval*
A	COMP.MREDFX_NREDS_VGV	1	$[0.99; 1.01]$
B	GLOBAL.LPT_MRED_ANSYNX	1	$[0.99; 1.01]$
C	GLOBAL.THG1_X	1550	$[1538; 1562]$
D	GLOBAL.THG2_X	1700	$[1692; 1708]$
E	TSTG4.ETA_ANSYNX	0	$[-0.001; 0.001]$

TABLE 1. Input data for model parameters

#	*Name*	*Measured value* Y_0	*Standard deviation* σ_Y
A	CD06.PT_IN	3.1615	0.05
B	CINT.DPTS	0.0028	0.05
C	EV.DPT	1.1662	0.05
D	EVP.PT_IN	32.3097	0.15
F	GENO.PGROSS	262095.7062	262.096
G	GLOBAL.DPSSEV_MEA	0.0827	0.001
N	GLOBAL.M_FUEL_TOTAL	14.2132	0.03
U	TDIF.GASOUT.TT	887.8054	2.0

TABLE 2. Input data for selected experimental observations

To have an easy comparison between the parameters after the calibration is performed, we introduce the following normalization. For the model parameters θ_i we are given initial parameters θ_i^0, and interval length $2d_{\theta_i}$, and hence we normalize[2] as follows

$$\tilde{\theta}_i = \sqrt{3}\frac{\theta_i - \theta_i^0}{d_{\theta_i}}, \quad i = 1 \dots n,$$

where $2d_{\theta_i}$ is the length of the intervals given in Table 1. Similarly, the normalization is done for the experimental observations by

$$\tilde{Y}_i = \frac{Y_i - Y_i^0}{\sigma_{Y,i}}, \quad i = 1 \dots m,$$

where $\sigma_{Y,i}$ is the standard deviation from Table 2.

4.2. **Deterministic calibration: $L1$- vs. $L2$-norm regularization.** We solve the calibration problem described above using the deterministic MSO method and compare $L1$- and $L2$-norm regularizations. For the MSO method we use given intervals and standard deviations from Tables 1 and 2 to compute the weighted Euclidean norms in (4) and (17). We use $T = 1$ and $\beta = 0.5$, which results in equal weights for both terms in the minimization problems (5) and (6).

Figure 1 shows the normalized model parameters after solving the calibration problem using the MSO method with $L1$- and $L2$-norm regularizations *on the left* and *on the right*, respectively. Since the normalized variables have standard deviation one, we mark the intervals $[-1; 1]$ by *vertical dashed blue lines*. The *red vertical lines* show the normalized model parameters after the optimization. For $L2$-norm regularization, *light gray bars* represent the approximated posterior standard deviation $\sigma_{\theta,i}$ and they have width $2\sigma_{\theta,i}$, while *dark gray bars* have width $4\sigma_{\theta,i}$ for $i = 1 \dots n$.

Remark 2. *For the L1-optimization (LASSO) the posterior standard deviation is actually not justified (by mathematical reasoning). Therefore, no artificial confidence regions are given for the L1-optimization results.*

Figure 1 (*upper panel*) shows that all five model parameters change using the MSO method with $L2$-norm regularization, but the most significant shift is observed only in the parameter “A”. On the other hand, the MSO method with $L1$-norm regularization selects as significant and optimizes only parameter “A”.

Figure 1 (*lower panel*) also shows the corrected values of the normalized experimental observations after the optimization using MSO method with $L1$- and $L2$-norm regularizations, and the plots exhibit comparable accuracy.

4.3. **MSO method: dependence on the weight β.** We introduced the weight β in the minimization problem (3) to calibrate the knowledge about the accuracy of the experimental observations Y^0 and about the values of the model parameters θ^0. In § 4.2 we used $\beta = 0.5$, which means that we do not give any preference to the experimental observation or to the model parameters. In this section we present results for the same numerical experiment with five calibration parameters, cf. § 4.1, but using the β values 0.1 and 0.9.

[2]The scaling is performed such that after normalization the new variables have mean zero and standard deviation equal to one. For uniformly distributed θ_i on some interval $[a, b]$ the standard deviation equals $(b - a)/\sqrt{12}$.

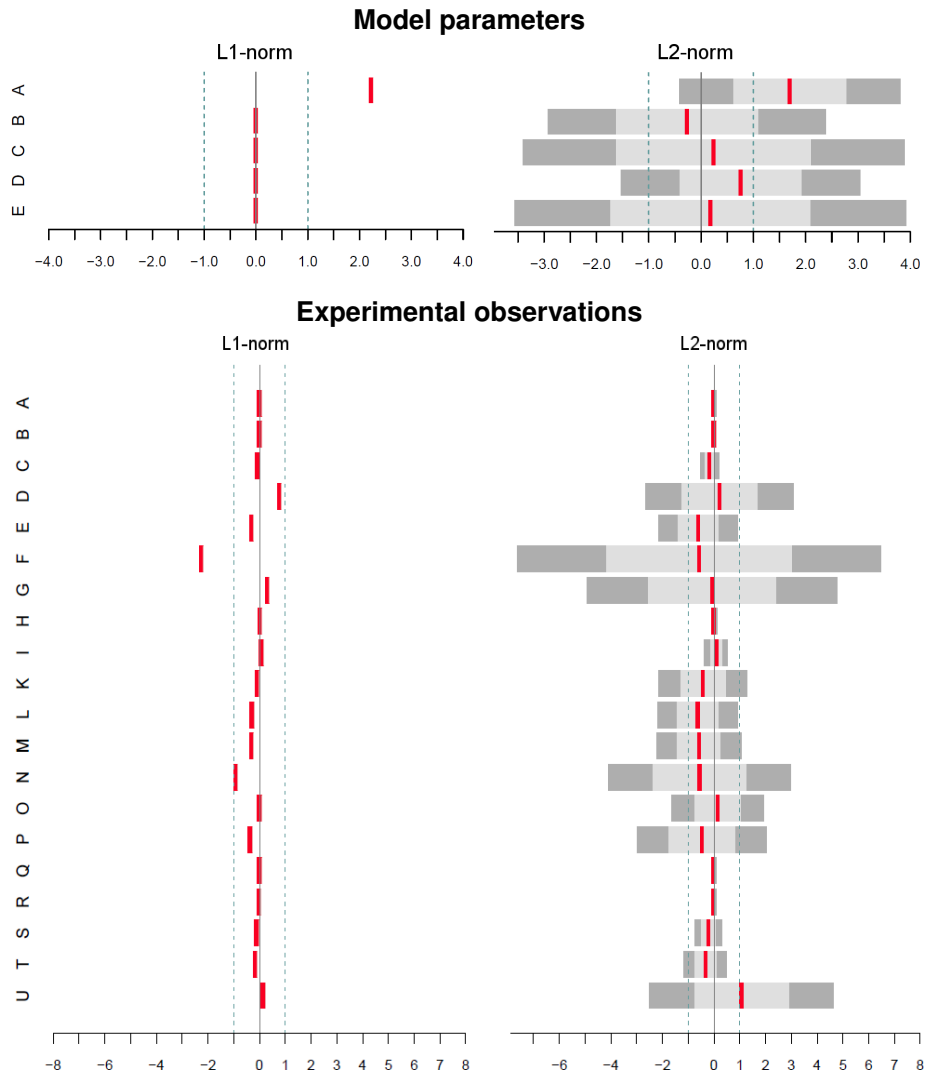

FIGURE 1. Calibration results with $\beta = 0.5$ using MSO method: $L1$- and $L2$-norm regularizations *on the left* and *on the right*, respectively. *Upper panel* shows fit for normalized model parameters $\tilde{\theta}$, *lower panel* shows the calibration for the normalized experimental observations $\tilde{Y}$.

The upper panels in Figures 2 and 3 show the normalized model parameters after solving the calibration problem with $\beta = 0.1$ and 0.9, respectively. If $\beta = 0.1$, i.e., if we trust our knowledge about the values of the model parameters θ^0 more than experimental observations Y^0, then the MSO method with $L1$-norm regularization (*the left of* Figure 2) shows that only the parameter "A" needs to be calibrated. Moreover, the resulting value of parameter "A" using $\beta = 0.1$ is closer to the initial one in comparison with the resulting value with $\beta = 0.5$. If $\beta = 0.9$, i.e., if we trust the experimental observations Y^0 and are not sure about our knowledge of the model parameters θ^0, then the MSO method with $L1$-norm (*the left of* Figure 3) selects parameters "A" and "D" for the calibration. For the MSO method with $L2$-norm regularization (*the right of* Figures 2 and 3), the model parameters after the optimization have larger shifts from their initial values for $\beta = 0.9$ than for $\beta = 0.1$. This behavior is consistent with the values of β.

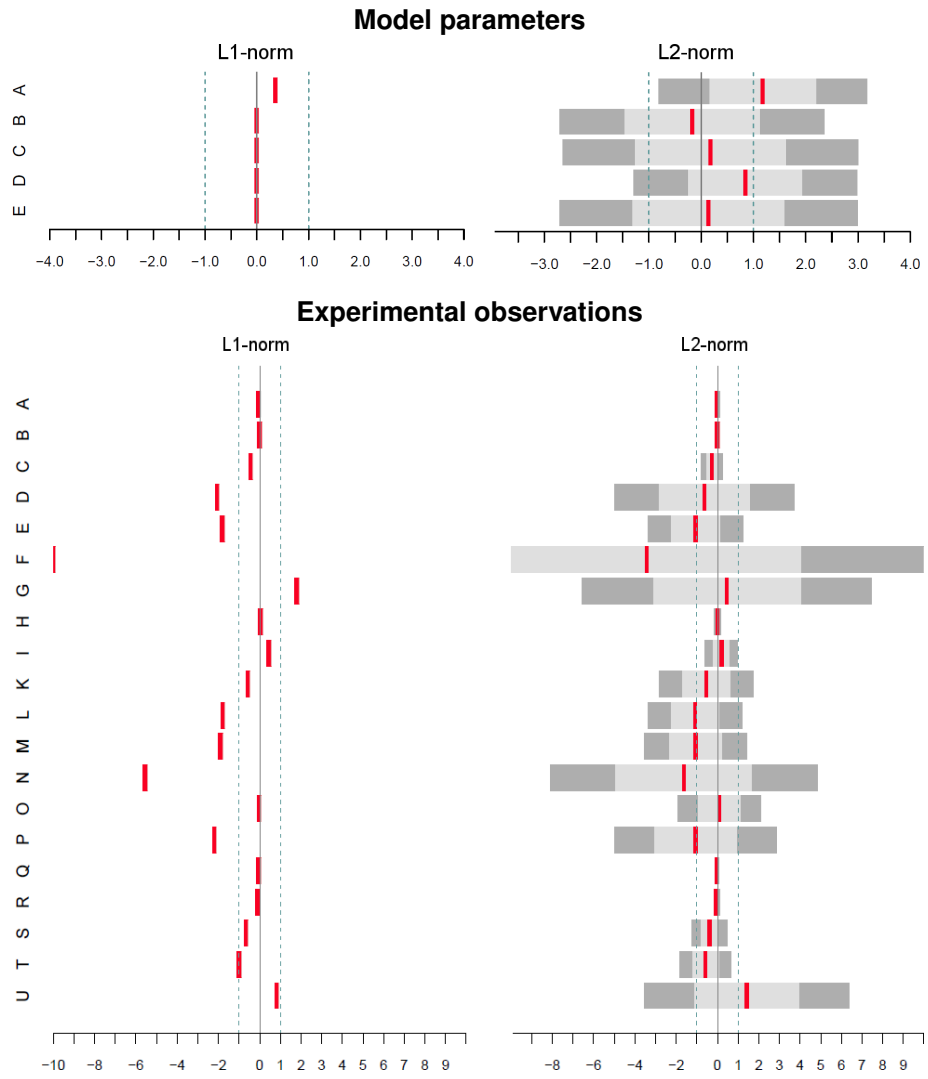

FIGURE 2. Calibration results with $\beta = 0.1$ using MSO method: $L1$- and $L2$-norm regularizations *on the left* and *on the right*, respectively. *Upper panel* shows fit for normalized model parameters $\tilde{\theta}$, *lower panel* shows the calibration for the normalized experimental observations $\tilde{Y}$.

The lower panels in Figures 2 and 3 show the corrected values of the normalized experimental observations after the optimization using the MSO method with $\beta = 0.1$ and 0.9, respectively. The weight $\beta = 0.9$ provides a better match to the experimental observations than the value 0.1, which is explained by the contribution of the terms in the minimization problem given the weight values.

4.4. **Bayesian approach: dependence on the scaling T.** Unlike the MSO method, the Bayesian calibration approach accounts for the stochastic information for the model parameters and for the experimental observations and hence provides a stochastic solution (sample) for the calibration problem. Here, we solve the calibration problem presented in § 4.1 accounting for the randomness of the input parameters. Similarly to § 4.2, we use the weight $\beta = 0.5$ for this example.

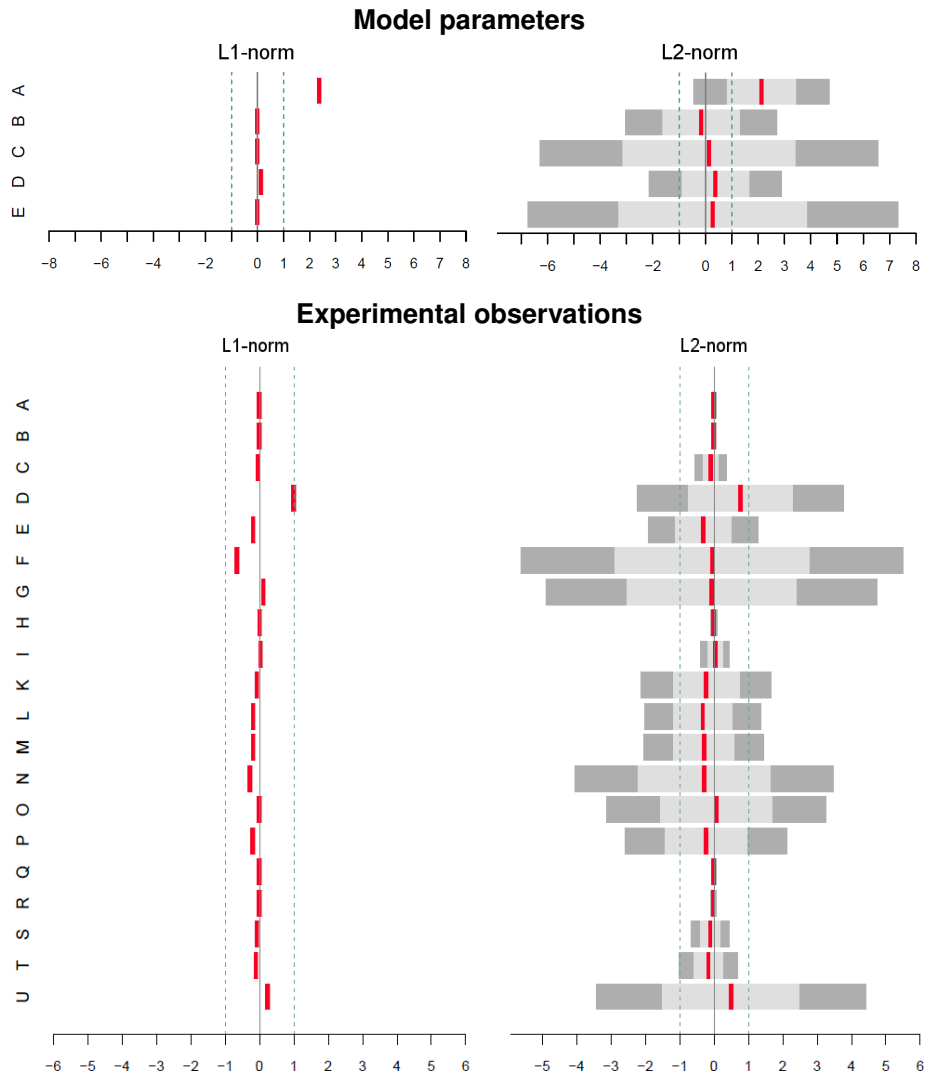

FIGURE 3. Calibration results with $\beta = 0.9$ using the MSO method: $L1$- and $L2$-norm regularizations *on the left* and *on the right*, respectively. *Upper panel* shows fit for normalized model parameters $\tilde{\theta}$, *lower panel* shows the calibration for the normalized experimental observations $\tilde{Y}$.

First, we solve the calibration problem with the default scaling factor $T = 1$, and we observe that the algorithm cannot find an appropriate fraction parameter that yields a 'good' acceptance ratio, cf. § 3.3. The 'bad' acceptance ratio is too far from the asymptotically optimal one and yields too wide posterior distributions, which sometimes include values outside of the subset Θ from (3). Hence, we decrease the scaling factor to $T = 0.1$ and apply again the Bayesian calibration approach, which yields a 'good' solution.

The achieved acceptance ratios and the used fraction parameters for both cases are shown in Table 3. For both cases, i.e., $T = 0.1$ and 1, Figure 4 shows posterior distributions of the model parameters (*dark gray line*) in comparison with their prior distributions (*light blue line*), as well as the MAP values (*dashed lines*) and the posterior mean (*solid lines*). Similarly to the MSO approach, we observe the most significant shift in the parameter "A". For the case with $\beta = 0.5$ and $T = 0.1$, Figure 5 shows the posterior densities of the selected experimental observations,

	$T = 0.1$	$T = 1$
Acceptance ratio	0.239	0.739
Fraction	0.438	0.3

TABLE 3. Used fraction values and achieved acceptance ratios using Bayesian approach with $\beta = 0.5$.

i.e., parameters “A”, “B”, “C”, “F”, “N”, and “U”. Parameter names written in red indicate that the MAP estimate is outside of the prior rectangle.

It is clear from (5) that the scaling parameter T does not change the solution of the least square problem, and hence the location of the MAP estimate in the Bayesian approach. However, it influences the posterior covariance. Larger values of T yield larger covariance. This can be seen from (9), but this is also evident in the numerical simulations (Figure 4). In addition, if the input intervals for the model parameters and the input standard deviations for the experimental observations are too large, then the calibration algorithms encounter difficulties to obtain a reliable solution. Then, we can use the scaling parameter T to improve the posterior distributions. T equals 1 by default and is set less than 1 when the algorithm encounters difficulties.

4.5. **MSO method vs. Bayesian approach.** To compare the deterministic and stochastic optimization methods, we perform simulations for the problem discussed in § 4.1 with $\beta = 0.9$ and $T = 0.1$ using the MSO with $L2$-norm regularization and Bayesian methods. Table 4 compares the MAP estimates obtained for five model parameters and five selected experimental observations obtained with these two methods. Additionally, we present the posterior mean values for the Bayesian approach. We observe that both methods produce well comparable values for the calibrated parameters. Moreover, the last two columns of Table 4 show how the standard deviation, which is estimated in the deterministic MSO method, matches the one obtained with the Bayesian approach.

Figure 6 shows the calibration results for the model parameters using the MSO method (*the left* column) and using the Bayesian approach (*the right* column) with sample size 5000. The distributions shown in *light blue* color are the prior distributions that are assumed to be Gaussian. In case of the Bayesian approach, the *dark gray* lines show the obtained posterior distributions. The MAP estimate, the posterior mean, and the standard deviations in Table 4 are computed from the obtained posterior distributions in this case. The MSO method outputs, for each model parameter and each experimental observation, a minimizer and an approximated standard deviation, computed from (9). Then, in the MSO plots Gaussian distributions shown in *dark gray* are drawn with these means and covariances.

Remark 3. *We make the following important notice. The primary goal of the Bayesian analysis is to provide samples from which one can detect the (posterior) mean, the (posterior) standard deviation, and other quantities of interest (QOI). The determination of the MAP may be seen as a by-product. As mentioned in § 2.4, if the functional dependence in (1) would be linear, and if the error distribution as well as the chosen prior distribution were assumed to be Gaussian, then the posterior would again be Gaussian. Its (theoretical population) mean would then coincide with the MAP estimate. If this is not the case, then the Bayesian approach delivers information about the posterior mean rather than the MAP, because the samples concentrate around its center. Thus, for the Bayesian approach the posterior mean is the QOI. If the focus is on the*

MAP, then other types of simulation must be considered. As the simulation results from Table 4 highlight, there is a good agreement between MAP and posterior mean, which indicates that the function F in (1) is 'linear' around the posterior mean.

Figure 7 shows the posterior densities of the selected experimental observations, i.e., parameters "C", "D", "F", "G", and "N", using the MSO method (*the left* column) and using the Bayesian approach (*the right* column). The *vertical blue dashed lines* in the MSO results represent intervals for the experimental observation, which are computed as $\left[Y^0 - \sqrt{3}\sigma_Y, Y^0 + \sqrt{3}\sigma_Y\right]$ to have a fair comparison between Bayesian and MSO methods. The same intervals in the Bayesian results are shown by boxes filled with inclined lines in *light blue color*.

In general, the calibration results compare well for both methods. The standard deviations for all model parameters match the standard deviations obtained with the Bayesian approach (Table 4). For all experimental observations, the MSO method underestimated the standard deviations in comparison with the Bayesian approach. The latter effect might be explained by the error introduced by the numerical differentiation, i.e., by the computation of the Jacobian, in the numerical estimation of the covariance (see the first term on the right hand side of (9)).

		MAP		*Posterior mean*	*STD*	
	#	*MSO*	*Bayesian*	*Bayesian*	*MSO*	*Bayesian*
Model param.	A	1.0124	1.0172	1.0122	0.0024	0.0025
	B	0.9991	1.0076	0.9987	0.0027	0.0032
	C	1550.9851	1553.5717	1552.2116	7.1960	6.8659
	D	1701.8083	1697.3673	1702.0322	1.8858	1.9936
	E	0.00017	-0.0012	0.0001	0.00066	0.0007
Exp. observ.	C	1.1622	1.1676	1.1621	0.0037	0.0052
	D	32.4269	32.5305	32.4384	0.0727	0.1143
	F	262085.1877	261265.0415	262070.6743	235.1104	874.3147
	G	0.0827	0.0853	0.0825	0.00078	0.0012
	N	14.2051	14.2228	14.2076	0.0183	0.0433

TABLE 4. MAP estimates, posterior mean values, and standard deviations (STD) for the model parameters and selected experimental observations using the MSO and Bayesian approaches with $\beta = 0.9$ and $T = 0.1$

Table 5 shows the computational times for the MSO method using $L1$- and $L2$-norm regularizations and for the Bayesian method with sample sizes 1000, 2000, and 5000. In general, the MSO is one or two orders of magnitude faster than the Bayesian approach. The MSO method with $L1$-norm regularization converges slower than the MSO method with $L2$-norm regularization. The computational time for the Bayesian method depends approximately linearly on the number of samples. From our experience, 1000 samples give a good enough approximation for this particular problem, but to get precise posterior distributions for comparison purposes, we used bigger sample size, i.e. 5000.

	MSO regularization		*Bayesian* number of samples		
	$L1$-norm	$L2$-norm	1000	2000	5000
Time, $[sec]$	59.96	20.7	419.69	892.62	2070.03

TABLE 5. Computational time for the MSO and Bayesian approaches with $\beta = 0.9$ and $T = 0.1$

5. DISCUSSION

We are concerned with calibration methods for mathematical models of gas turbine performance as these are implemented in the simulation tool BOP. Specifically, we discuss in detail the following calibration methods: the classical penalized least square method (MSO), the sparsity promoting minimization, i.e., LASSO, and the Bayesian method.

The penalized least squares method, called MSO within BOP, is a versatile calibration method. Based on our numerical experiment, it works well and fast but has some disadvantages. First, as a result of the calibration all model parameters are changed, which from an engineering point of view, is not desirable, as the engineers trust in their model, and they want to correct only those parameters which are significant for the reproduction of the real gas turbine behavior. Another disadvantage is the lack of sensitivity information (uncertainty quantification) at the minimizer. Both these drawbacks can partially be resolved. The LASSO method calibrates the model parameters selectively (sparsely), only those parameters get optimized that are important for the minimization problem of interest, see Figures 1.

To retrieve sensitivity information the engineer has several choices. First, a full Bayesian approach is implemented. This gives the desired uncertainty quantification if the MAP, which corresponds to the minimizer in the MSO method, is close to the posterior mean. This is the case in the simulation study. However, this method is time consuming (see Table 5). Alternatively, by linearizing the mapping F from (1) around the minimizer, the explicit form of the Bayesian posterior covariance as this is given in (9) can be used to get the sensitivity information. Again, the comparison in Table 4 highlights that this is a reliable alternative. The increased computational time is the price that has to be paid for the fully reliable solution. At present no Bayesian alternative to the LASSO is implemented, and therefore, no quantification of the uncertainty inherent in the obtained solution is available.

Another aspect of this study is the choice of the tuning parameters in the minimizer, namely weights $0 < \beta < 1$ and $T > 0$. Their role was discussed in § 4.3 and 4.4, respectively. From engineering point of view the effect of the weight β is the following. When setting up a calibration problem, if we are sure about our prior knowledge of the model parameters, then we choose $\beta < 0.5$. Then, less priority is given to the matching of the experimental observations than to the matching of the initial values of the model parameters. On the contrary, if we are sure that the experimental measurements are accurate and we do not have so much knowledge about the model parameters, then we choose $\beta > 0.5$. Finally, if we are equally confident about the initial values of the model parameters and about the experimental observations, then we choose $\beta = 0.5$.

The scaling parameter T is introduced to control the posterior covariances and, in general, does not change the solution of the minimization problem. It equals one by default and should not be changed unless required to solve the calibration problem. As we saw in § 4.4, smaller values of T might be needed to obtain a 'good' acceptance ratio and a reliable solution using the Bayesian method if the posterior covariance is too large.

Based on these observations we draw the following conclusions. Practically, to solve a calibration problem, the Bayesian approach should first be used to gain some intuition about the model's and the parameters' behavior. Then, in order to validate the accuracy of the MSO, one should perform the comparison of the calibration results obtained with the MSO and the Bayesian method for the given minimization problem. If the comparison is successful then one can proceed using the MSO method for further fast simulations for the given minimization problem. In order to use the sparse minimization (LASSO), a sparse Bayesian method should be implemented.

REFERENCES

[1] M. Arias-Chao, D. S. Lilley, P. Mathé, and V. Schloßhauer. Calibration and uncertainty quantification of gas turbine performance models. In *ASME Turbo Expo 2015: Turbine Technical Conference and Exposition*, pages V07AT29A001–V07AT29A001. American Society of Mechanical Engineers, 2015.

[2] J. Borchardt, G. Blättermann, P. Mathé, and V. Schloßhauer. *Process Simulator BOP (Version 3.2)*. Weierstrass-Institute for Applied Analysis and Stochastics, Mohrenstraße 39, D-10117 Berlin, 2016.

[3] X. Canalias, F. Köpf, and P. Sahm. Generation of physically based analysis factors to improve synthesis models of jet engines. In *ASME Turbo Expo 2005: Power for Land, Sea, and Air*, pages 673–681. American Society of Mechanical Engineers, 2005.

[4] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

[5] M. Kennedy and A. O'Hagan. Bayesian calibration of computer model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

[6] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math.*, 2:164–168, 1944.

[7] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.*, 11:431–441, 1963.

[8] P. Mathé. Bayesian approach to model calibration. *Weierstrass Institute for Applied Analysis and Stochastics*, 2014.

[9] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.

[10] G. Roberts, A.Gelman, and W. Gilks. Weak convergence and optimal scaling of random walk metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997.

[11] V. Schloßhauer. BOP MSO - a software package for mean square error optitmization. *Weierstrass Institute for Applied Analysis and Stochastics*, 2014.

[12] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996.

[13] P. Walsh and P. Fletcher. *Gas Turbine Performance*. Blackwell Science, 2004.

[14] L. Wang. Uncertainty quantification and model validation. In *ASME V&V Conference*, Las Vegas, 2012.

[15] G. A. Watson. *Numerical analysis*, chapter The Levenberg-Marquardt algorithm: implementation and theory by J.J. Moré, pages 105–116. Lecture Notes in Mathematics, Vol. 630. Springer-Verlag, Berlin-New York, 1978. Edited by G. A. Watson.

[16] Y. Xiong, W. Chen, K.-L. Tsui, and D. W. Apley. A better understanding of model updating strategies in validating engineering models. *Computer Methods in Applied Mechanics and Engineering*, 198(15–16):1327 – 1337, 2009.

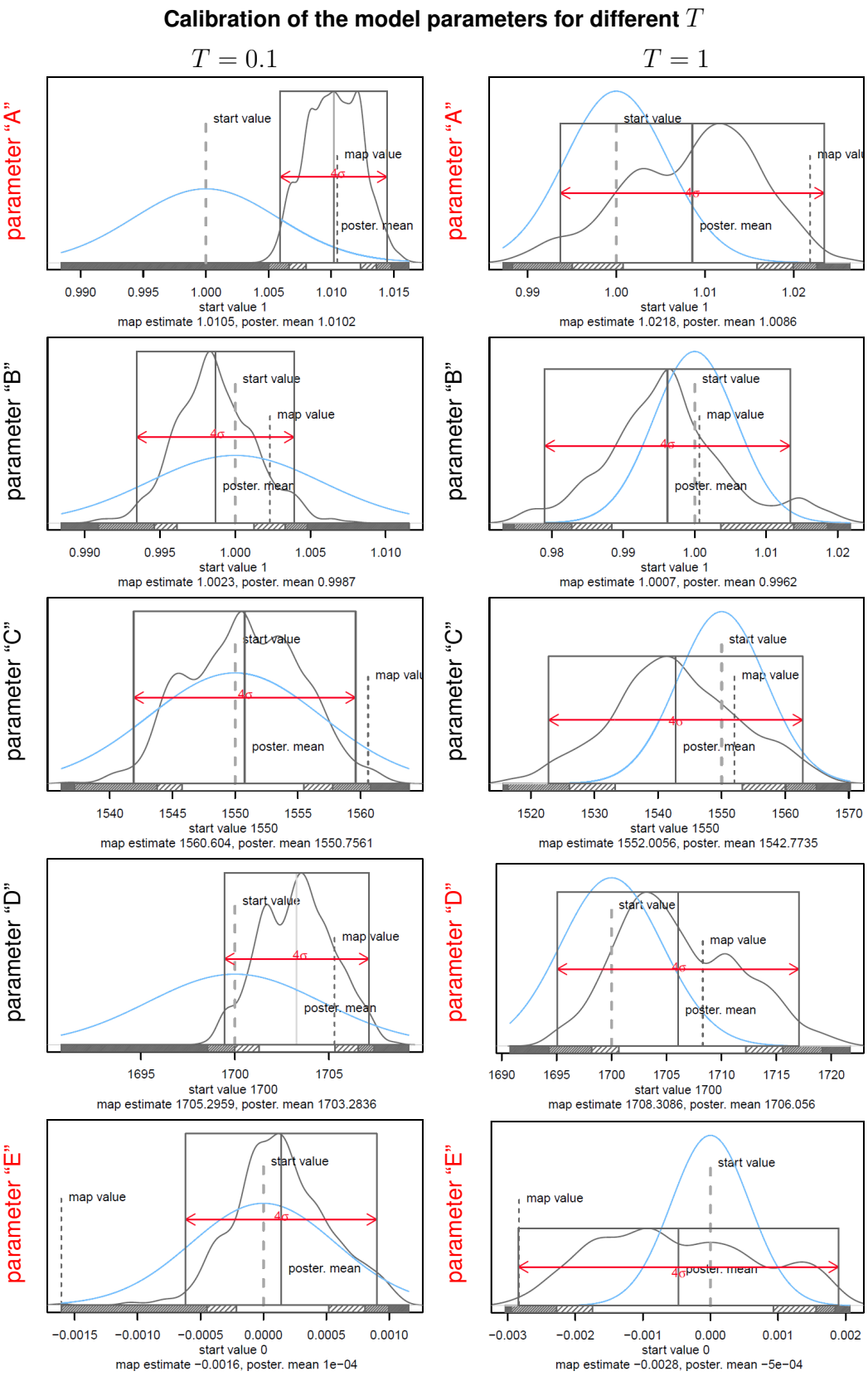

FIGURE 4. Calibration results with $\beta = 0.5$ and with $T = 0.1$ and 1 for the model parameters using the Bayesian algorithm.

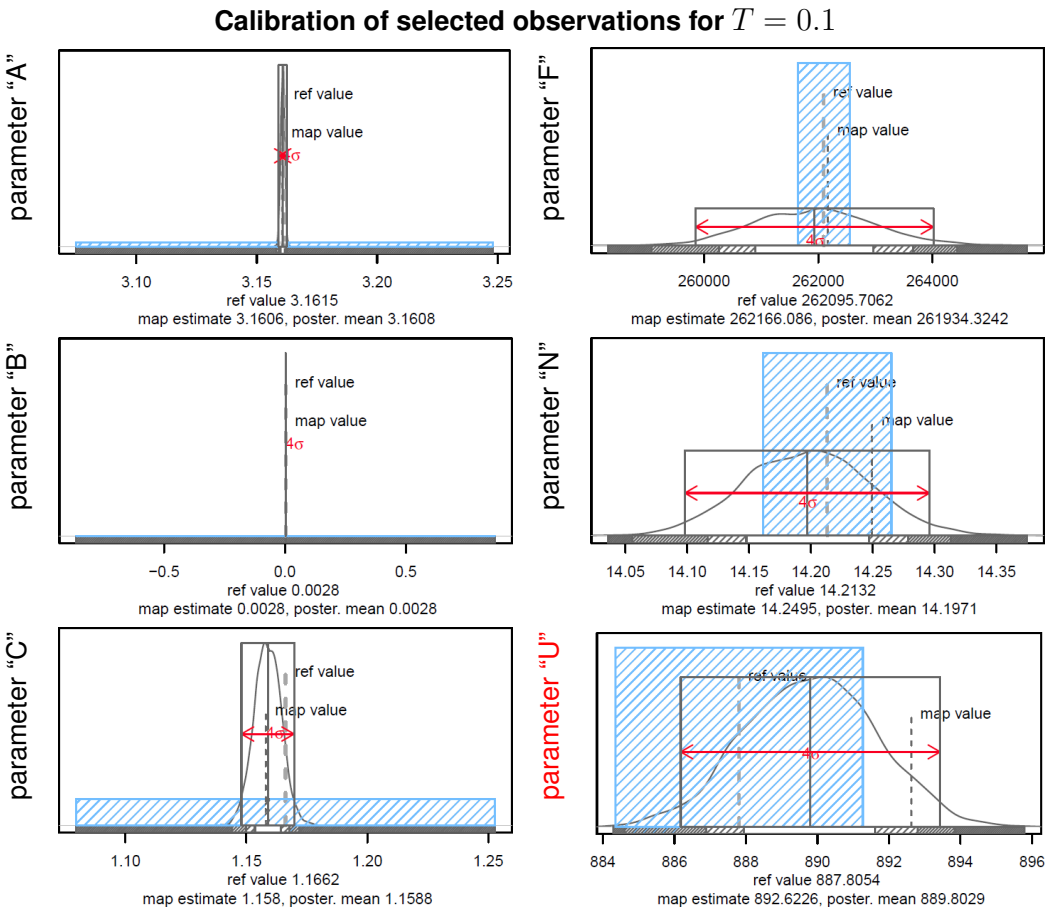

FIGURE 5. Calibration results with $\beta = 0.5$ and with $T = 0.1$ for selected experimental observations, i.e. parameters "A", "B", "C", "F", "N", and "U", using the Bayesian algorithm.

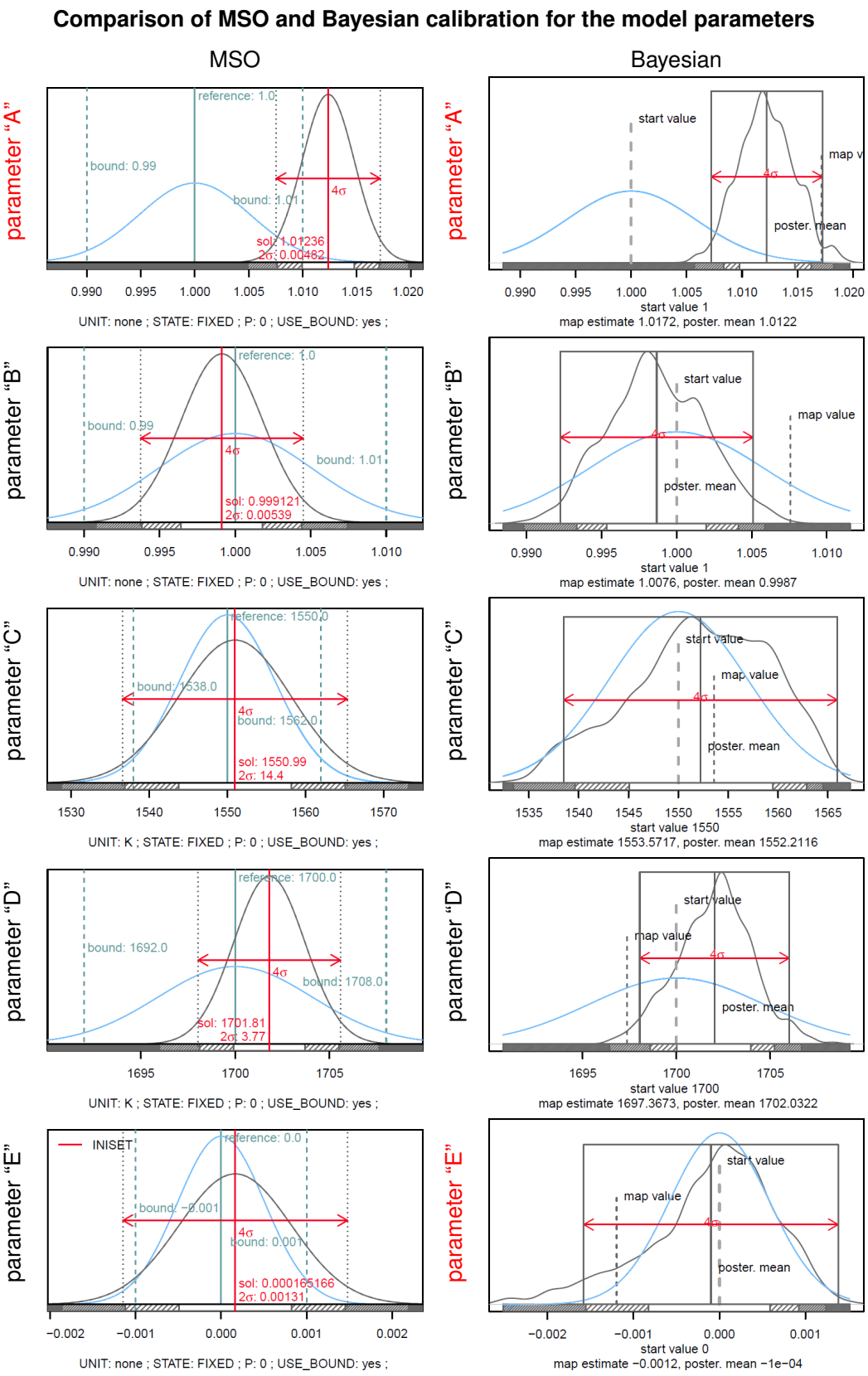

FIGURE 6. Calibration results with $\beta = 0.9$ and $T = 0.1$ for the model parameters using the Bayesian and MSO algorithms.

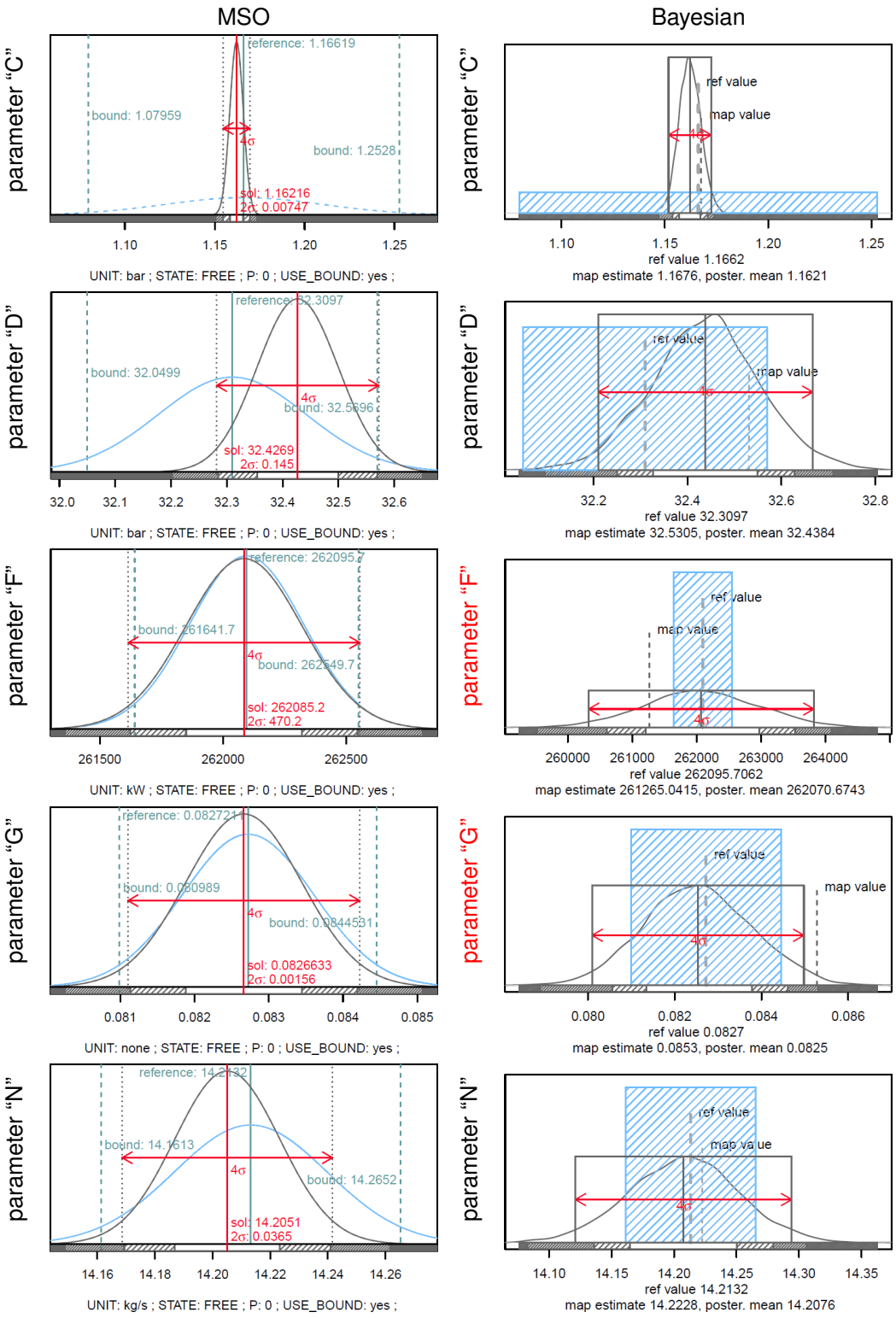

FIGURE 7. Calibration results with $\beta = 0.9$ and $T = 0.1$ for selected experimental observations, i.e. parameters "C", "D", "F", "G", and "N", using the Bayesian and MSO algorithms.