

Cascaded Calibration of Mechatronic Systems via Bayesian Inference

Max van Meer* Emre Deniz* Gert Witvoet*,**
Tom Oomen*,***

* Eindhoven University of Technology, Eindhoven, the Netherlands
(e-mail: m.v.meer@tue.nl).

** TNO, Dept. of Optomechatronics, Delft, the Netherlands.

*** Delft University of Technology, Delft, the Netherlands.

Abstract: Sensors in high-precision mechatronic systems require accurate calibration, which is achieved using test beds that, in turn, require even more accurate calibration. The aim of this paper is to develop a cascaded calibration method for position sensors of mechatronic systems while taking into account the variance of the calibration model of the test bed. The developed calibration method employs Gaussian Process regression to obtain a model of the position-dependent sensor inaccuracies by combining prior knowledge of the sensor with data using Bayesian inference. Monte Carlo simulations show that the developed calibration approach leads to significantly higher calibration accuracy when compared to alternative regression techniques, especially when the number of available calibration points is limited. The results indicate that more accurate calibration of position sensors is possible with fewer resources.

Keywords: Mechatronic systems, Bayesian methods, Calibration, Gaussian Process regression

1. INTRODUCTION

High-precision mechatronic systems rely on accurate position measurements to achieve high performance. At the same time, an increasing number of applications requires highly accurate position measurements in mass-produced systems, e.g., satellite swarms for optical communication (Gregory et al., 2010; Kramer et al., 2020), or segmented mirror telescopes (Nelson, 2006).

The accuracy of position measurements relates to their proximity to the actual positions, and precision refers to repeatability. This paper considers sensors that consistently exhibit position-dependent inaccuracies. These repeatable sensor inaccuracies can be measured using a test bed with a more accurate sensor, e.g., coordinate measurement machines (Takamasu et al., 1996) or optics-based test beds (Dresscher et al., 2019). When sensor inaccuracies are measured, a model is fitted to compensate for these inaccuracies through the process of calibration.

The test beds used to calibrate the position sensors of mechatronic systems require calibration themselves, to a standard regarded as an absolute measure of accuracy. This is done by a third party such as a metrology institute (Pendrell, 2009), or in-house using a highly accurate manual instrument, e.g., theodolite (Krishna, 1996) or laser tracking interferometers (Umetsu et al., 2005).

Due to this cascade of calibration steps, depicted schematically in Fig. 1, modeling errors in individual calibration steps can stack and limit the achieved accuracy of the sensor calibrated last. Two leading causes of modeling

errors are as follows. First, calibration on manual, external calibration instruments is time-consuming. While efforts have been made to partially automate the comparison of test bed sensor readings to accurate external readings (see, e.g., Wu and Wang (2013)), these methods introduce additional development cost and complexity. Consequently, the number of positions at which sensor readings are compared with those of more accurate sensors is limited. Second, some test bed locations may be unreachable to external instruments due to geometry constraints, further limiting the number of available calibration points.

In a parallel line of developments, the application of Gaussian Process regression to mechatronic systems has gained increased attention, see Poot et al. (2022); Rasmussen and Williams (2006), since it admits a highly flexible model structure while taking uncertainty into account using Bayes' theorem. By specifying a prior that imposes properties such as smoothness and learning hyperparameters from the data, a model is obtained that yields information not only of the expected function but also the variance of this function space.

Although regression techniques such as lookup tables can model individual functions well if the number of calibration points is large, it is shown in this paper that by taking into account the variance of individual calibration models using Bayesian inference, a significantly more accurate model is obtained, even if the number of calibration points is limited. Moreover, while sequential calibration using Bayesian inference has attained attention in the context of computer models (Jiang et al., 2020), the literature on the cascaded calibration of position sensors is sparse.

Therefore, the aim of this paper is to find an accurate mapping of position sensor readings to 'true' position values, i.e., sensor n in Fig. 1, while taking the uncertainty of the intermediate calibration model into account. The contributions of this paper are as follows:

* This work is part of the research programme VIDI with project number 15698, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO). In addition, this research has received funding from the ECSEL Joint Undertaking under grant agreement 101007311 (IMOCO4.E). The Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme.

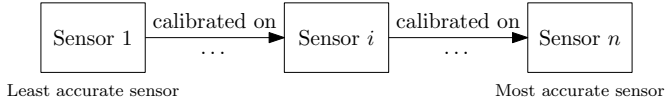


Fig. 1. Schematic overview of the cascaded calibration problem. An array of sensors is calibrated on each other. Consequently, any imperfection in the calibration of the more accurate sensors is propagated down the chain to the less accurate sensors.

- C1: A method for cascaded calibration of position sensors for mechatronic systems is developed. The approach takes the model uncertainty of the first calibration step into account to arrive at a more accurate estimation in the subsequent calibration step.
- C2: The effectiveness of the approach is demonstrated through Monte Carlo simulations on a reproducible case study, and it is shown that the developed calibration method yields significantly more accurate models of the sensor offsets than alternatives such as lookup tables. The results indicate that more accurate calibration of mass-produced mechatronic systems is possible with fewer resources.

This paper is structured as follows. First, the problem description is given in Section 2. Next, the developed approach to cascaded calibration is explained in Section 3. Subsequently, simulation results are presented in Section 4, and finally, conclusions are drawn in Section 5.

2. PROBLEM DESCRIPTION

In this section, the problem description is given. First, a motivating example is given. Subsequently, the calibration problem is described, and finally, the problem is formalized in terms of two regression problems.

2.1 Motivating example

A motivating example of cascaded encoder calibration is shown in Fig. 2. The angular position sensor S_1 of a mechatronic system requires calibration, but the system is too compact to be accessible by the manual calibration instrument S_3 , e.g., a theodolite or autocollimator (Gao et al., 2011). Hence, it is calibrated using a test bed with sensor S_2 , which is optically aligned with the mechatronic system, such that readings by S_2 of the mechatronic system can be compared with readings by S_1 of the mechatronic system. The test bed, in turn, is calibrated using the manual calibration instrument, see Fig. 3.

The manual calibration instrument S_3 cannot measure all locations of the test bed because its frame physically obstructs access. Moreover, accurate manual calibration is labor-intensive, especially if high accuracy is required over the entire 360° range of motion. Hence, the number of available calibration points of S_3 is limited.

Any imperfection in the calibration of the test bed to the manual instrument decreases the accuracy of the mechatronic system when it is calibrated on the test bed. This propagation of modeling errors motivates the need to take the uncertainty of the calibration model of the test bed into account when calibrating the mechatronic system.

2.2 Notation

The following notation is used. Sensor S_i , $i \in \{1, 2, 3\}$, is fixed to system i , where system 1 is the mechatronic

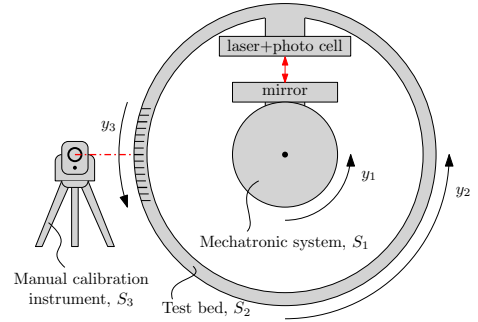


Fig. 2. Motivating example. A mechatronic system with angular position sensor S_1 is optically linked for calibration with a test bed, with its own sensor S_2 . The test bed itself is calibrated using a highly accurate manual measuring instrument S_3 . Sensors S_i yield different measurements y_i of the same actual position when they are aligned for calibration.

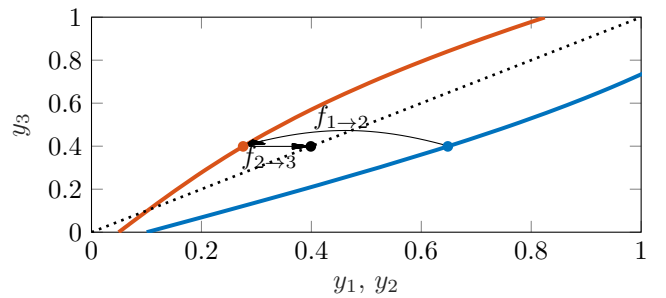


Fig. 3. Sensor S_1 (—) is calibrated on a test best with sensor S_2 (—), which, in turn, is calibrated on S_3 (.....). Since the systems are aligned during calibration, more accurate sensors can observe the sensor inaccuracies of less accurate sensors.

system, system 2 is the test bed, and system 3 is the manual calibration instrument. Sensor S_i yields measurement $y_i \in \mathbb{R}$. All positions are defined w.r.t. the same fixed reference frame. The true position of system i is denoted by y_i^* . When systems i and j are aligned for calibration (as detailed in Assumption 2), it holds that $y_i^* = y_j^*$, and hence, measurements y_i and y_j are two different measurements of the same true position.

2.3 Cascaded calibration of sensors

Sensor S_i is generally not perfectly accurate, i.e., $y_i \neq y_i^*$. Sensor inaccuracies can have multiple causes, e.g., manufacturing tolerances, mechanical wear, or imperfect assembly. These inaccuracies to the true position are position-dependent, expressed as follows:

$$y_i^* = f_{i \rightarrow i^*}(y_i), \quad (1)$$

where $f_{i \rightarrow i^*} : \mathbb{R} \rightarrow \mathbb{R}$ is a possibly nonlinear function, mapping inaccurate position measurements to true positions.

Assumption 1. Measurements y_i of y_i^* are corrupted by zero-mean additive Gaussian white noise ε_i with variance $\sigma_{n,i}^2$ assumed small compared to the sensor inaccuracies. Long-term temporal changes in sensor-inaccuracies are assumed negligible, and short-term changes are assumed zero-mean, such that temporal effects are averaged out over multiple observations at the same location, i.e., $f_{i \rightarrow i^*}(y_i, t) \approx f_{i \rightarrow i^*}(y_i)$.

If sensor S_i of system i is not perfectly accurate, then these inaccuracies are measured by a second system j , provided

Procedure 1 Cascaded calibration of S_1 on S_3 via S_2

- 1: Align systems 2 and 3 and obtain a data-set $\mathcal{D}_2 = \{\bar{y}_{2,k}, \bar{y}_{3,k}\}_{k=1}^{N_2}$. Use these observations of $\bar{y}_3 = f_{2 \rightarrow 3}(\bar{y}_2)$ to fit a function $y_3 = \hat{f}_{2 \rightarrow 3}(y_2)$.
 - 2: Align systems 1 and 2 and obtain a data-set $\mathcal{D}_1 = \{\bar{y}_{1,k}, \bar{y}_{2,k}\}_{k=1}^{N_1}$. Then construct $\mathcal{D}'_1 = \{\bar{y}_{1,k}, \hat{f}_{2 \rightarrow 3}(\bar{y}_{2,k})\}_{k=1}^{N_1}$ using the model of Step 1, and use these ‘observations’ of $\hat{f}_{2 \rightarrow 3}(\bar{y}_2) = f_{1 \rightarrow 3}(\bar{y}_1)$ to fit a function $y_3 = \hat{f}_{1 \rightarrow 3}(y_1)$.
-

that sensor S_j is more accurate, i.e.,

$$\int_{y_{j \min}}^{y_{j \max}} |y_j(y_j^*) - y_j^*| dy_j^* < \int_{y_{i \min}}^{y_{i \max}} |y_i(y_i^*) - y_i^*| dy_i^*. \quad (2)$$

System j can only be used to measure the sensor inaccuracies of S_i if systems i and j are aligned, i.e., $y_i^* = y_j^*$, because only then do they attempt to measure the same true position. Therefore, the following assumption is made:

Assumption 2. When a pair (y_i, y_j) of sensor readings is measured at a fixed point in time, it is assumed that misalignment errors are negligible w.r.t. sensor inaccuracies:

$$|y_i^* - y_j^*| \ll |y_i^* - y_i|. \quad (3)$$

Hence, both measurements (y_i, y_j) are assumed to describe the same true position $y_i^* \approx y_j^*$ during calibration.

The manual calibration instrument S_3 is the most accurate sensor available, and therefore, y_3 can effectively be used as a proxy for y_3^* . The following function is then defined, assuming systems 3 and i are aligned:

$$y_3^* = f_{i \rightarrow 3}(y_i), \quad (4)$$

i.e., $f_{i \rightarrow 3}$ describes the relationship between a sensor reading y_i and its ‘true’ position $y_3^* \approx y_3$.

Assumption 3. Each $f_{i \rightarrow 3}(y_i)$, $i \in \{1, 2\}$, is bijective, i.e., any value of y_i maps to one value of y_3^* and vice versa.

The aim is to obtain a model $\hat{f}_{1 \rightarrow 3}$ of $f_{1 \rightarrow 3}$, describing the sensor inaccuracy of S_1 , but importantly, $f_{1 \rightarrow 3}$ is never observed directly. In other words, the manual calibration instrument is not aligned with the mechatronic system for calibration, for two reasons: (i) it is economically not viable to perform manual calibration on many different mechatronic systems with sensor S_1 , and (ii) the mechatronic system may be built too compactly to be physically accessible by a manual calibration instrument.

To prevent the need to calibrate S_1 on S_3 directly, the test bed with S_2 is first calibrated on S_3 . Subsequently, S_2 can calibrate many different mechatronic systems, each with its own sensor S_1 . These two steps are detailed in Procedure 1. When $\hat{f}_{1 \rightarrow 3}$ is modeled offline through Procedure 1, it is used online to transform inaccurate position measurements y_1 to corrected measurements $\hat{f}_{1 \rightarrow 3}(y_1)$.

Importantly, cascaded calibration requires making a fit on another fit. Since measurements of S_3 are labor-intensive and time-consuming, the first model $\hat{f}_{2 \rightarrow 3}$ might be based on a limited amount of data ($N_2 \ll N_1$), and consequently, it may have a large variance. This potentially deteriorates the accuracy of $\hat{f}_{1 \rightarrow 3}$ w.r.t. the true $f_{1 \rightarrow 3}$.

In the next section, it is explained how the construction of $\hat{f}_{1 \rightarrow 3}(y_1)$ from data through Steps 1-2 of Procedure 1 is framed as a series of regression problems.

2.4 Cascaded calibration through regression

To obtain a model $\hat{f}_{1 \rightarrow 3}$ of the sensor inaccuracies of S_1 that can be used for calibration, the following cascade of regression problems is defined.

Problem 1. Consider Procedure 1, and suppose that data-sets \mathcal{D}_1 and \mathcal{D}_2 are available. Parameterize the models as $y_3 = \hat{f}_{1 \rightarrow 3}(y_1, \alpha)$ and $y_3 = \hat{f}_{2 \rightarrow 3}(y_2, \beta)$, respectively, and let their structures be fully determined by Hilbert spaces \mathcal{K}_1 and \mathcal{K}_2 (Wegman, 2006). The aim is to obtain the best possible fit of $f_{1 \rightarrow 3}$, even though $f_{1 \rightarrow 3}$ is not measured directly, but instead by solving two sub-problems:

$$\min_{\mathcal{K}_1, \mathcal{K}_2} \mathcal{J} = \left(\frac{\int_{y_1^{\min}}^{y_1^{\max}} [\hat{f}_{1 \rightarrow 3}(y_1, \alpha^*) - f_{1 \rightarrow 3}(y_1)]^2 dy_1}{y_1^{\max} - y_1^{\min}} \right)^{\frac{1}{2}}$$

subject to

$$\beta^* = \operatorname{argmin} \left\| \hat{f}_{2 \rightarrow 3}(\bar{y}_2, \beta) - f_{2 \rightarrow 3}(\bar{y}_2) \right\|_2^2 + \left\| \hat{f}_{2 \rightarrow 3} \right\|_{\mathcal{K}_2}^2, \\ \alpha^* = \operatorname{argmin} \left\| \hat{f}_{1 \rightarrow 3}(\bar{y}_1, \alpha) - \hat{f}_{2 \rightarrow 3}(\bar{y}_2, \beta^*) \right\|_2^2 + \left\| \hat{f}_{1 \rightarrow 3} \right\|_{\mathcal{K}_1}^2, \quad (5)$$

where y_1^{\min} and y_1^{\max} specify a range of positions where a good model of the sensor inaccuracies of S_1 is desired. Note that solving for β^* and for α^* amounts to constructing the fits of Steps 1 and 2 in Procedure 1, respectively.

The cost \mathcal{J} cannot be evaluated in practice because S_1 can only be compared to S_2 , and S_2 to S_3 , but not S_1 to S_3 . On the other hand, \mathcal{J} can be evaluated in simulation, when $f_{1 \rightarrow 3}$ is known. In the following sections, it is shown that for a specific choice of the structure of \hat{f} through \mathcal{K}_1 and \mathcal{K}_2 , the cost \mathcal{J} is significantly reduced, when compared to conventional regression methods, indicating that more accurate calibration is achievable.

3. CASCADED CALIBRATION VIA BAYESIAN INFERENCE

In this section, the developed solution to Problem 1 is explained. The key idea is to recognize that the two sub-problems in (5) need to be posed in a coupled fashion. If these problems were solved independently, then any inaccuracy in $\hat{f}_{2 \rightarrow 3}(y_2)$ that follows from the fact that y_3 is only available at a limited number of positions is overconfidently carried over to $\hat{f}_{1 \rightarrow 2}(y_1)$.

Instead, the developed approach takes into account the uncertainty of $\hat{f}_{2 \rightarrow 3}(y_2)$ at locations y_2 where no data of y_3 is available, through Bayes’ rule.

3.1 Calibration of S_2

First, sensor S_2 needs to be calibrated on sensor S_3 ; see the first sub-problem in (5). Given a limited number of observations of pairs (y_2, y_3) showing the relation $y_3 = f_{2 \rightarrow 3}(y_2)$, there is uncertainty in $\hat{f}_{2 \rightarrow 3}(y_2, \beta)$ for values of y_2 that are far from calibrated locations. It is explained next how this model uncertainty, or variance, is computed explicitly so that it can be used for more accurate regression in the next section. To this end, a probabilistic viewpoint is adopted.

The model structure of $\hat{f}_{2 \rightarrow 3}(y_2, \beta)$ is assumed to be

$$\hat{f}_{2 \rightarrow 3}(y_2, \beta) = \phi_2(y_2)^\top \beta, \quad (6)$$

where $\phi_2 : \mathbb{R} \rightarrow \mathbb{R}^D$ maps any y_2 into some D -dimensional feature space, with weights β . A Gaussian prior is assumed on β , i.e.,

$$p(\beta) = \mathcal{N}(\beta_0, \Sigma_{2,p}), \quad (7)$$

with mean β_0 and prior variance $\Sigma_{2,p} \in \mathbb{R}^{D \times D}$. With this model, the likelihood $p(\bar{Y}_3 | \bar{Y}_2, \beta)$, or the probability density of the observations given the parameters, is given by

$$p(\bar{Y}_3 | \bar{Y}_2, \beta) = \mathcal{N}(\Phi_2(\bar{Y}_2)^\top \beta, \Sigma_{\bar{Y}_3}), \quad (8)$$

where

$$\begin{aligned} \bar{Y}_i &:= [\bar{y}_{i,1}, \dots, \bar{y}_{i,N_i}]^\top, & i \in \{1, 2\}, \\ \Phi_i(\bar{Y}_i) &:= [\phi_i(\bar{y}_{i,1}), \dots, \phi_i(\bar{y}_{i,N_i})]^\top, & i \in \{1, 2\}, \end{aligned} \quad (9)$$

and the variance of the observations is given by

$$\Sigma_{\bar{Y}_3} = \sigma_{n,3}^2 I. \quad (10)$$

The prior on β is conditioned on the data \mathcal{D}_2 to obtain the posterior distribution, i.e., the probability of the parameters given the data. From Bayes' rule, it is known that

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}}, \quad (11)$$

or specifically,

$$p(\beta | \bar{Y}_3, \bar{Y}_2) = \frac{p(\bar{Y}_3 | \bar{Y}_2, \beta)p(\beta)}{p(\bar{Y}_3 | \bar{Y}_2)}. \quad (12)$$

The expression (12) describes the posterior distribution of the parameters. For regression, rather, the predictive distribution $p(Y_3 | \bar{Y}_2, \bar{Y}_3) = p(\hat{f}_{2 \rightarrow 3}(Y_2) | \bar{Y}_2, \bar{Y}_3)$ is of interest. For arbitrary sensor positions $Y_2 \in \mathbb{R}^{M_2}$, this predictive distribution is computed as

$$p(\hat{f}_{2 \rightarrow 3}(Y_2) | \bar{Y}_2, \bar{Y}_3) = \int p(Y_3 | Y_2, \beta)p(\beta | \bar{Y}_2, \bar{Y}_3)d\beta, \quad (13)$$

which is again a Gaussian with mean and variance

$$\begin{aligned} \mathbb{E}[\hat{f}_{2 \rightarrow 3}(Y_2)] &= m_2(Y_2) + K_2(Y_2, \bar{Y}_2) \\ &\quad [K_2(\bar{Y}_2, \bar{Y}_2) + \Sigma_{\bar{Y}_3}]^{-1} (\bar{Y}_3 - m_2(\bar{Y}_2)), \\ \text{cov}(\hat{f}_{2 \rightarrow 3}(Y_2)) &= K_2(Y_2, Y_2) - K_2(Y_2, \bar{Y}_2) \\ &\quad [K_2(\bar{Y}_2, \bar{Y}_2) + \Sigma_{\bar{Y}_3}]^{-1} K_2(\bar{Y}_2, Y_2), \end{aligned} \quad (14)$$

and the elements of $K_i(\bar{Y}_i, Y_i) = K_i(Y_i, \bar{Y}_i)^\top \in \mathbb{R}^{N_i \times M_i}$, $K_i(\bar{Y}_i, \bar{Y}_i) \in \mathbb{R}^{N_i \times N_i}$ and $K_i(Y_i, Y_i) \in \mathbb{R}^{M_i \times M_i}$ are obtained from evaluating a kernel function $k_i(y_A, y_B)$ for the corresponding values of arbitrary positions Y_i and measurements \bar{Y}_i . This kernel has the property that $K_i = \Phi_i^\top \Sigma_{i,p} \Phi_i$, and thus relates to the chosen model structure or prior. The prior mean is denoted by $m_i : \mathbb{R}^{M_i} \rightarrow \mathbb{R}^{M_i}$. See Section 3.3 for details on the choice of k_i and m_i .

Crucially, (14) provides an analytic expression of the covariance of $\hat{f}_{2 \rightarrow 3}(Y_2)$, i.e., the uncertainty of the model, illustrated in light purple in the left of Fig. 4. This covariance is instrumental to obtaining a more accurate estimate $\hat{f}_{1 \rightarrow 3}$, as explained in the next section.

3.2 Calibration of S_1

In this section, a model $\hat{f}_{1 \rightarrow 3}$ is made, based on the model of $\hat{f}_{2 \rightarrow 3}$ obtained in the previous section. Since sensor readings S_1 can only be experimentally compared

with S_2 , but not to S_3 , the model of $\hat{f}_{2 \rightarrow 3}$ is applied to $\mathcal{D}_1 = \{\bar{y}_{1,k}, \bar{y}_{2,k}\}_{k=1}^{N_1}$ to obtain

$$\mathcal{D}'_1 = \{\bar{y}_{1,k}, \hat{y}_{3,k}\}_{k=1}^{N_1}, \quad (15)$$

where

$$\begin{aligned} \hat{y}_{3,n} &:= \hat{f}_{2 \rightarrow 3}(\bar{y}_{2,n}), \\ \hat{Y}_3 &:= [\hat{f}_{2 \rightarrow 3}(\bar{y}_{2,1}), \dots, \hat{f}_{2 \rightarrow 3}(\bar{y}_{2,N_1})]^\top. \end{aligned} \quad (16)$$

The key insight that distinguishes the regression approach in this paper from traditional methods is that the variance of the 'observations' $\hat{y}_3 = \hat{f}_{2 \rightarrow 3}(\bar{y}_2) \in \mathcal{D}'_1$, i.e., the prediction of the model created in the previous section evaluated at measurements \bar{y}_2 , is affected by the uncertainty of the model. The covariance matrix corresponding to \hat{Y}_3 follows directly from this model uncertainty and is given by

$$\Sigma_{\hat{Y}_3} := \text{cov}(\hat{f}_{2 \rightarrow 3}(\bar{Y}_2)), \quad (17)$$

which is computed directly through (14). By assuming a Gaussian prior on $\hat{f}_{1 \rightarrow 3}$ as before and conditioning on \mathcal{D}'_1 , the predictive distribution $p(\hat{f}_{1 \rightarrow 3})$ is a Gaussian with

$$\begin{aligned} \mathbb{E}[\hat{f}_{1 \rightarrow 3}(Y_1)] &= m_1(Y_1) + K_1(Y_1, \bar{Y}_1)^\top \\ &\quad [K_1(\bar{Y}_1, \bar{Y}_1) + \Sigma_{\hat{Y}_3}]^{-1} (\hat{Y}_3 - m_1(\bar{Y}_1)), \\ \text{cov}(\hat{f}_{1 \rightarrow 3}(Y_1)) &= K_1(Y_1, Y_1) - K_1(Y_1, \bar{Y}_1) \\ &\quad [K_1(\bar{Y}_1, \bar{Y}_1) + \Sigma_{\hat{Y}_3}]^{-1} K_1(\bar{Y}_1, Y_1). \end{aligned} \quad (18)$$

Indeed, the posterior mean $\mathbb{E}[\hat{f}_{1 \rightarrow 3}]$, shown in dashed blue in Fig. 4, is a function of $\Sigma_{\hat{Y}_3} = \text{cov}(\hat{f}_{2 \rightarrow 3})$, shown in light purple. Clearly, by taking into account the uncertainty of the model $\hat{f}_{2 \rightarrow 3}$, the model $\hat{f}_{1 \rightarrow 3}$ is affected. It is shown in Section 4 that this choice results in a more accurate model $\hat{f}_{1 \rightarrow 3}$ than when the uncertainty is ignored. This concept is shown visually in Fig. 4, where the key observation is that $\hat{f}_{1 \rightarrow 3}$ (dashed blue) does not rely on \hat{y}_3 (purple dots) at locations where $\hat{f}_{2 \rightarrow 3}$ has high variance (light purple), but instead relies more on its prior $\mathbb{E}[y_3] = y_1$.

Next, the choice of k_i and m_i , which determine the prior of $\hat{f}_{i \rightarrow j}$, is explained.

3.3 Selection of the model structure and hyper-parameters

In this section, the chosen model structure, or prior, is elaborated, and it is shown how the hyper-parameters are chosen automatically using data with empirical Bayes.

First, the choice of the prior is explained. It follows from Equations (14), (18) that the model structure of $\hat{f}_{i \rightarrow j}$ is uniquely determined by choice of kernel function k_i , defining the prior variance, and the prior mean m_i .

For a function $y_j = \hat{f}_{i \rightarrow j}(y_i)$, the prior mean is

$$m_i(y_i) := \mathbb{E}[y_j]. \quad (19)$$

An intuitive choice is to pick $m_i(y_i) = y_i$, which assumes that in the absence of observations, sensor S_j is expected to yield identical measurements as S_i . If prior information on the sensor inaccuracy is available, it can also be incorporated into m_i .

The prior variance, i.e., the range of possible functions that $\hat{f}_{i \rightarrow j}$ can take, is determined by the kernel function. For an overview of possible kernel functions, including, e.g.,

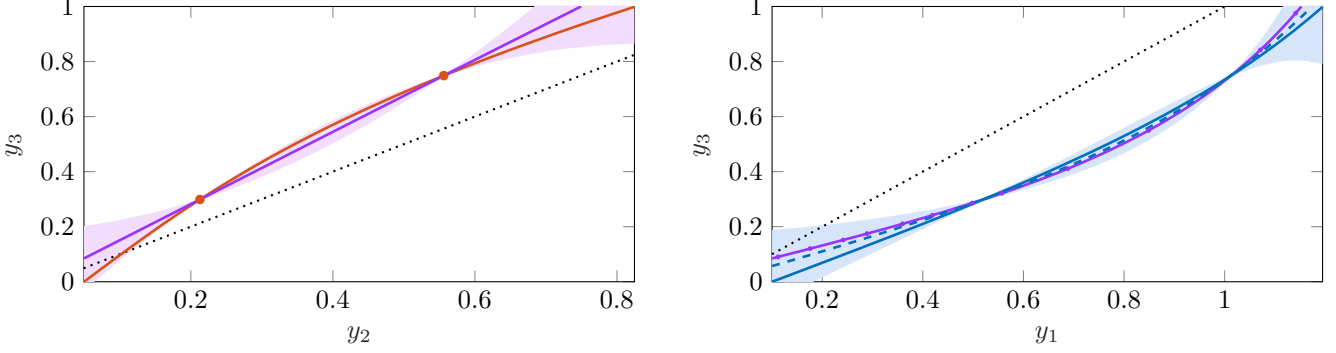


Fig. 4. Example of cascaded calibration via Bayesian inference. Sensor readings y_2 (—) and y_1 (—) have different position-dependent inaccuracies w.r.t. $y_3 \approx y^*$ (.....) while the systems are physically aligned to y^* . \bar{y}_3 is measured at two locations (●) and a model $\hat{f}_{2 \rightarrow 3}(y_2)$ (—, ■) is fitted through these points. Next, this model is used to obtain $\hat{y}_3 = \hat{f}_{2 \rightarrow 3}(\bar{y}_2)$ (—, right) for many measurements \bar{y}_2 . Finally, a model $\hat{f}_{1 \rightarrow 3}$ (—, ■) is fitted on \hat{y}_3 . Because this model takes the uncertainty of $\hat{f}_{2 \rightarrow 3}$ into account, the model $\hat{f}_{1 \rightarrow 3}$ (—) partially disregards the data at locations with high uncertainty, and instead relies more on its prior $\mathbb{E}[y_3] = y_1$, leading to a more accurate fit to $f_{1 \rightarrow 3}$ (—).

polynomial or sinusoidal model structures, see Duvenaud (2014). In this paper, the attention is restricted to squared exponential (SE) kernel functions of the form

$$k_i(y_A, y_B) = \sigma_{f,i}^2 \exp\left(-\frac{1}{2\ell_i^2}(y_A - y_B)^2\right), \quad (20)$$

where hyper-parameters ℓ_i and $\sigma_{f,i}^2$ are the characteristic length scale and the magnitude of the prior variance, respectively. This model structure can be interpreted as imposing smoothness on $\hat{f}_{i \rightarrow j}$.

Hyper-parameters $\Theta_i = \{\ell_i, \sigma_{f,i}^2, \sigma_{n,i}^2\}$ can be chosen from prior knowledge of the smoothness and magnitude of the sensor inaccuracies and noise. Alternatively, these can be learned from the data \mathcal{D}_i , also known as empirical Bayes, by maximizing the log marginal likelihood, given by

$$\log p(Y_j | Y_i, \Theta_i) = -\frac{1}{2} Y_j^\top \tilde{K}_i^{-1} Y_j - \frac{1}{2} \log |\tilde{K}_i| - \frac{N_i}{2} \log 2\pi, \quad (21)$$

with $\tilde{K}_i = K_i(Y_i, Y_i) + \Sigma_{Y_i}$. This expression is maximized with respect to Θ using an optimization algorithm for non-convex problems, to find optimal hyper-parameters.

3.4 Summary

The complete algorithm to obtain estimates of y_3 for arbitrary measurements y_1 is summarized in Algorithm 1. Note that after following steps 1-5, step 6 can be repeated cheaply for any Y_1 , since the computation of $\mathbb{E}[Y_3] = \mathbb{E}[\hat{f}_{1 \rightarrow 3}(Y_1)]$ through (18) is simply a matrix-vector multiplication once the inverted matrix in (18) is stored for future use.

Remark 1. The posterior means in (14) and (18) are identical to the solutions of the sub-problems in (5), see Rasmussen and Williams (2006, Section 6.2) for details.

4. RESULTS

In this section, the effectiveness of the developed calibration approach is demonstrated through Monte Carlo simulations. The simulation set-up is given first, and subsequently, the results are presented.

Algorithm 1 Cascaded calibration via Bayesian inference

- Require:** Data-sets $\mathcal{D}_1, \mathcal{D}_2$, test points $Y_1 \in \mathbb{R}^{M_1}$.
- 1: Specify kernel functions k_1, k_2 with initial hyper-parameters $\Theta_{i,0}$, see Section (3.3).
 - 2: Find optimal hyper-parameters Θ_2^* by maximization of $\log p(\bar{Y}_3 | \bar{Y}_2, \Theta_2)$, see (21).
 - 3: Compute $\mathbb{E}[\hat{Y}_3] = \mathbb{E}[\hat{f}_{2 \rightarrow 3}(\bar{Y}_2)] \in \mathbb{R}^{N_1}$ and $\text{cov}(\hat{Y}_3) = \text{cov}(\hat{f}_{2 \rightarrow 3}(\bar{Y}_2)) \in \mathbb{R}^{N_1 \times N_1}$ with (14).
 - 4: Find optimal hyper-parameters Θ_1^* by maximization of $\log p(\hat{Y}_3 | \bar{Y}_1, \Theta_1)$, see (21).
 - 5: Compute $\mathbb{E}[Y_3] = \mathbb{E}[\hat{f}_{1 \rightarrow 3}(Y_1)] \in \mathbb{R}^{M_1}$ with (18).
 - 6: **return** $\mathbb{E}[Y_3]$

4.1 Monte Carlo simulation set-up

Suppose the sensor measurements y_i obtained from sensor S_i during alignment ($y_1^* = y_2^* = y_3^* = y^*$) are described by

$$\begin{aligned} y_3 &= y^* + \varepsilon, \\ y_2 &= y^* + \sum_{k=1}^{N_s} a_k \sin(\omega_{1,k} y^*) + b_k \cos(\omega_{1,k} y^*) + \varepsilon, \\ y_1 &= y^* + \sum_{k=1}^{N_s} c_k \sin(\omega_{2,k} y^*) + d_k \cos(\omega_{2,k} y^*) + \varepsilon, \end{aligned} \quad (22)$$

with $N_s = 10$ and $\varepsilon \sim \mathcal{N}(0, 10^{-8})$. The range of interest is $y^* \in [0, 1]$ m, and hence, the values y_1^{\min} and y_1^{\max} in (5) follow from (22).

For the Monte Carlo simulations, $N = 12000$ different pairs of functions (22) are generated with $a_k, b_k, c_k, d_k \sim \mathcal{N}(0, 10^{-4})$ and $\omega_{i,k} \sim \mathcal{N}(0, 6)$. The data sets are collected as follows. Data-set \mathcal{D}_1 is obtained by observing y_2 for an equally spaced grid of $N_1 = 100$ values of y_1 . Subsequently, y_2 is observed for an equally spaced grid of 100 values of y_3 , but then 10% of the data on either edge and 20% of the data in the center is removed, leading to $N_2 = 64$. This represents a scenario where S_3 cannot measure the test bed everywhere because it is physically obstructed.

Algorithm 1 is followed for all N functions. The accuracy of the resulting model $\hat{f}_{1 \rightarrow 3}$ is then assessed by its cost \mathcal{J} in (5) and compared with two alternative techniques:

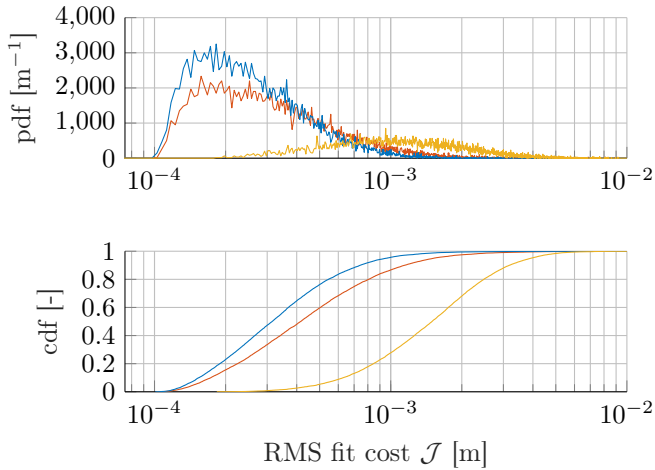


Fig. 5. Normalized empirical probability density functions (top) and cumulative density functions (bottom) of \mathcal{J} ($N=12000$). Algorithm 1 (—) leads to a more accurate fit $\hat{f}_{1\rightarrow 3}$ than Alternative 1 (—), in which the variance of the first model is ignored. Both methods perform significantly better than Alternative 2 (—), a lookup table with linear interpolation.

Alternative 1: Algorithm 1 is followed, except that uncertainty of the model $\hat{f}_{2\rightarrow 3}$ is approximated as $\Sigma_{\hat{Y}_3} = \sigma_{n,3}^2 I$, where $\sigma_{n,3}^2$ is found by maximization of (21).

Alternative 2: A lookup table is made of $\hat{f}_{2\rightarrow 3}$ using \mathcal{D}_2 . Subsequently, a lookup table $\hat{f}_{1\rightarrow 3}$ is made using \mathcal{D}'_1 , see (15). Linear interpolation is used between entries in the lookup table.

4.2 Simulation results

The computational times of Algorithm 1 and Alternative 1 were less than five seconds each on a personal computer for each of the $N = 12000$ simulations. The results of the Monte Carlo simulations are shown in Fig. 5. The empirical probability distribution functions are normalized to have area 1. It is clear from the Fig. 5 that taking the model uncertainty of $\hat{f}_{2\rightarrow 3}$ into account through Algorithm 1 leads to a considerably better fit than when the uncertainty is ignored (Alternative 1). Both kernel-based methods result in more accurate models than a lookup table with linear interpolation (Alternative 2).

The results indicate that Algorithm 1 is a suitable solution to Problem 5, leading to better models of sensor inaccuracies than alternative approaches. With better models of sensor inaccuracies, more accurate calibration is achieved.

5. CONCLUSION AND RECOMMENDATIONS

A cascaded calibration method is developed to accurately model position-dependent inaccuracies of position sensors, enabling more accurate calibration of mass-produced systems in less time. By taking into account the uncertainty resulting from the first regression step using Bayesian inference, more accurate calibration is achieved than conventional methods such as lookup tables. The approach is especially advantageous when the number of calibration points is limited. Moreover, since the model hyperparameters are tuned automatically using the data, the procedure is convenient to implement in practice.

Future work is required to deal with cases when sensor readings are not fully repeatable. In the current framework, this is done by storing only average readings in the data sets, but a proper Bayesian treatment of this spread might further increase the achieved accuracy. Moreover, future efforts will be aimed at experimental validation of the method on the motivating example.

REFERENCES

- Dresscher, M., Human, J.D., Witvoet, G., Van Der Heiden, N., Den Breeje, R., Kuiper, S., Fritz, E.C., Korevaar, C.W., Van Der Valk, N.C., De Lange, T.J., Saathof, R., Doelman, N., Crowcombe, W.E., Duque, C.M., and De Man, H. (2019). Key Challenges and Results in the Design of Cubesat Laser Terminals, Optical Heads and Coarse Pointing Assemblies. *2019 IEEE Int. Conf. on Space Optical Systems and Applications, ICSOS 2019*.
- Duvenaud, D. (2014). *Automatic model construction with Gaussian processes*. Ph.D. thesis, University of Cambridge.
- Gao, W., Saito, Y., Muto, H., Arai, Y., and Shimizu, Y. (2011). A three-axis autocollimator for detection of angular error motions of a precision stage. *CIRP Annals - Manufacturing Technology*, 60(1), 515–518.
- Gregory, M., Heine, F., Kämpfner, H., Meyer, R., Fields, R., and Lunde, C. (2010). TESAT laser communication terminal performance results on 5.6Gbit coherent inter satellite and satellite to ground links. In N. Kadowaki (ed.), *International Conference on Space Optics - ICSO 2010*, vol. 10565, 37. SPIE.
- Jiang, C., Hu, Z., Liu, Y., Mourelatos, Z.P., Gorsich, D., and Jayakumar, P. (2020). A sequential calibration and validation framework for model uncertainty quantification and reduction. *Computer Methods in Applied Mechanics and Engineering*, 368, 113172.
- Kramer, L., Peters, J., Voorhoeve, R., Witvoet, G., and Kuiper, S. (2020). Novel motorization axis for a Coarse Pointing Assembly in Optical Communication Systems. In *IFAC PapersOnLine*, vol. 53, 8426–8431. Elsevier Ltd.
- Krishna, R. (1996). Improved pointing accuracy using high-precision theodolite measurements. In E.R. Washwell (ed.), *GOES-8 and Beyond*, vol. 2812, 199 – 209. International Society for Optics and Photonics, SPIE.
- Nelson, J. (2006). Segmented mirror telescopes. In *Optics in Astrophysics*, 61–72. Springer.
- Pendrill, L.R. (2009). EURAMET: European Assoc. of National Metrology Inst. *NCSLI Measure*, 4(4), 40–44.
- Poot, M., Portegies, J., Mooren, N., van Haren, M., van Meer, M., and Oomen, T. (2022). Gaussian Processes for Advanced Motion Control. *IEEEJ Journal of Industry Applications*, 21011492.
- Rasmussen, C. and Williams, C. (2006). *Gaussian processes for machine learning*. London, England.
- Takamasu, K., Ozawa, S., and Asano, T. (1996). Basic concepts of nano-CMM. *The Japan-China Bilateral Symposium on Advanced Manufacturing Engineering*, 155–158.
- Umetsu, K., Furutnani, R., Osawa, S., Takatsuji, T., and Kurosawa, T. (2005). Geometric calibration of a coordinate measuring machine using a laser tracking system. *Meas. Science and Tech.*, 16(12), 2466–2472.
- Wegman, E.J. (2006). Reproducing Kernel Hilbert Spaces. In *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Inc., Hoboken, NJ, USA.
- Wu, B. and Wang, B. (2013). Automatic measurement in large-scale space with the laser theodolite and vision guiding technology. *Advances in Mech. Eng.*, 2013.