

# Bayesian calibration of quasi-static field distortions in HARMONI

Gonzalo J. Carracedo Carballal <sup>1</sup>★, Javier Piqueras López,<sup>1</sup> Miguel Pereira Santaella <sup>2</sup>  
and Fraser Clarke<sup>3</sup>

<sup>1</sup>*Centro de Astrobiología (CAB, CSIC-INTA), Carretera de Ajalvir, Km 4, 28850 Torrejón de Ardoz, Spain*

<sup>2</sup>*Instituto de Física Fundamental, CSIC, Serrano 123, 28006 Madrid, Spain*

<sup>3</sup>*Department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, U.K.*

Accepted 2024 February 29. Received 2024 February 19; in original form 2023 December 12

## ABSTRACT

The high angular resolution monolithic optical and near-infrared Integral field spectrograph is the first light visible and near-infrared integral field spectrograph for the *Extremely Large Telescope*. To reach the diffraction limit of the telescope ( $\approx 10$  mas) and maintain an optimal image quality over long exposures, an accurate measurement of geometrical distortions in the instrument's guide star field is needed. Geometrical distortions affecting the guide stars map directly to pointing errors of the science field. The systematic contribution to the pointing error can be calibrated and removed by a corrective model. In this work, we propose a formulation of the corrective model that aims to calibrate the geometrical field distortions down to a given target residual, as well as reducing the time spent in calibrations. We also propose a calibration procedure that accounts for the uncertainties of the measurement process. We developed a tool named `harmoni-pm` to simulate the expected pointing error caused by geometrical distortions and validate the effectiveness of the proposed corrective model. We also relied on pseudo Zernike polynomials to formulate the model, and the Bayesian theoretical framework to study the propagation of uncertainties along the calibration process. Compared with the classical calibration algorithm, the Bayesian calibration algorithm was able to reduce the number of calibration points required to reach the same model residual. Additionally, we were able to optimize the hardware of the Geometrical Calibration Unit and reduce the time required to achieve the calibration goal.

**Key words:** Instrumentation – Algorithms – Bayesian Analysis – Integral Field Spectrography.

## 1. INTRODUCTION

The *Extremely Large Telescope* (ELT) is a visible to near-infrared (NIR) segmented telescope, currently under construction as part of the European Southern Observatory. At the time of its completion (expected near 2028), its 39-m primary mirror will allow access to a previously unexplored resolution regime of 10 mas near the diffraction limit.

High angular resolution monolithic optical and near-infrared integral field spectrograph (HARMONI) will be one of the first generation instruments of ELT, providing integral field spectroscopy with a variety of adaptive optics modes that enable diffraction-limited observations. HARMONI is conceived as a workhorse instrument and, as such, it must enable a wide range of science cases, ranging from the high-redshift universe to Solar system bodies (Thatte et al. 2021).

The performance of the instrument in the different science cases was validated by means of HARMONI's pipeline simulator HSIM (Zieleniewski et al. 2015; Pereira-Santaella et al. 2023). Simulations included the detection and classification of high-redshift Type Ia supernovae (Bounissou et al. 2018), observability studies of Population III stars (Grisdale et al. 2021), gas kinematic studies of

high-redshift galaxies (Richardson et al. 2020), multiwavelength observations of the circumgalactic medium emission (Augustin et al. 2019), supermassive black hole mass measurements (Nguyen et al. 2023), and determination of morpho-kinematics and star-formation rates of cosmic noon active galactic nuclei (García-Lorenzo et al. 2022) and other high-redshift galaxies (Kendrew et al. 2016; Grisdale et al. 2022).

While adaptive optics is necessary to achieve the diffraction limit, an equally accurate pointing is also critical. This represents a technical challenge not only from the opto-mechanical perspective, but also from the calibration process itself: an instrumental, quasi-systematic pointing error is expected, and must be removed up to a residual comparable to the diffraction limit. This is achieved by means of a corrective model, which predicts the quasi-systematic pointing error in all points of the instrument's focal plane.

In this paper, we contribute to the development of the corrective model of HARMONI in three ways. First, we introduce the formulation of a general corrective model as a linear combination of complex basis functions on the focal plane, along with a classical calibration strategy (Section 4.2). Next, we provide a conjugate Bayesian calibration algorithm that, in addition to being incremental and not requiring Monte-Carlo sampling, has the potential of reducing the number of measurements needed to reach the calibration goal (Section 4.3). We continue by exploring its practical application to HARMONI field distortions using pseudo Zernike polynomials

\* E-mail: [gcarracedo@cab.inta-csic.es](mailto:gcarracedo@cab.inta-csic.es)

(PZPs) as basis functions on the unit disc, and propose a geometrical calibration unit (GCU) mask pattern that reduces the impact of measurement noise in calibrations (Section 6). Finally, we discuss its potential application to the pointing model of a telescope, expressed as a multipole expansion of the pointing error (Section 8).

## 2. HARMONI

HARMONI is a slicer-based integral field spectrograph designed to operate in the 0.46- to 2.45- $\mu\text{m}$  range ( $3500 < R < 17000$ ) in a broad variety of scientific programs (Thatte et al. 2022). Its different spectral set-ups will enable observations in the NIR bands, including  $J$  (1.05–1.32  $\mu\text{m}$ ),  $H$  (1.44–1.82  $\mu\text{m}$ ),  $K$  (1.95–2.47  $\mu\text{m}$ ), and portions thereof. It will enable both adaptive optics (AO) [including laser-tomography adaptive optics (LTAO), single-conjugate adaptive optics (SCAO), and high-contrast adaptive optics (HCAO)] and non-AO observations, with fields of view (FoV) of 9 arcsec  $\times$  6 arcsec, 4 arcsec  $\times$  3 arcsec, 2 arcsec  $\times$  1.5 arcsec, and 0.8 arcsec  $\times$  0.6 arcsec. In its highest resolution configuration, it will be able to resolve objects around 10 mas, with a sampling of 4 mas per spaxel over a 0.8 arcsec  $\times$  0.6 arcsec FoV.

One of the design features of HARMONI is that guiding does not take place on the science target, in any AO mode other than HCAO (Thatte et al. 2022). Instead, the pointing of the science field in HARMONI is controlled by a guide probe measuring the position of a natural guide star (NGS). The basic control loop sends feedback to the telescope to ensure that the guide star is always centred on the guide probe. Any unknown term in the apparent distance between the NGS and the science target will therefore appear as a pointing error on the science target. If the error changes on the time-scale of an observation (effectively up to 1 h), it will also impact the image quality. This error would manifest as a *dragging* of the science object during the observation.

Of particular concern in HARMONI is optical distortion introduced by the Focal Plane Relay Subsystem (FPRS), as the fundamental architecture of the instrument means that the distortion pattern is seen to rotate as the instrument tracks the field rotation of the telescope. The instrument performance budgets—internal to the project—identify this as a major source of error if not properly characterized by modelling.

Different guide probes are used for different AO modes of HARMONI; a single probe on-axis for HCAO, a single probe within 15 arcsec of the science target for SCAO, a single guide probe within 60 arcsec for non-AO, and two guide probes within 60 arcsec for LTAO (Dohlen et al. 2022). The latter is the most challenging configuration for the geometric distortion, as it is the most demanding combination of spatial resolution and physical patrol field.

Ensuring that the field distortion contribution does not affect the LTAO image quality ( $\approx 10$  mas) means understanding the model at the level of  $\approx 2$  mas over 60 000 mas; i.e. 1 part in 30 000.

If we understand the guide probes as measurement devices of the positions of stars, we should expect them to be affected by random and quasi-systematic error contributions. While the former fall under the category of measurement noise and are—in internal project budgets—much smaller than the pointing accuracy, the latter are caused by slowly varying ( $\gg$  exposure time) opto-mechanical effects (e.g. optical distortions, mechanical non-linearities, etc.) and are much larger than the measurement noise. The overall instrument control concept therefore relies on a corrective *pointing model* to compensate for long-term quasi-systematic effects.

The calibration of the corrective model relies on HARMONI's GCU. The GCU consists of a deployable mask that can be inserted

in the focal plane of the telescope, and provides a pattern of sources whose locations are known beforehand. The difference between the nominal and measured locations is then used as training data for the corrective model.

## 3. METHODOLOGY

In order to validate the proposals made in this paper, two simulation codes were used: `harmoni-pm` and `bayescal`.

`harmoni-pm` (Carracedo-Carballal 2022) is a pointing error simulator written in Python 3 for HARMONI. It was used to characterize the error distribution of the measurements of the guide probes, identify the main drivers of the instrumental error, validate corrective models, and evaluate calibration patterns. `harmoni-pm` relies on an exhaustive, extensible instrument model in which contributions to the instrumental pointing error are described as composed  $\mathbb{R}^2 \rightarrow \mathbb{R}^2$  transforms. These transforms are applied to the coordinates of the points contained in the instrument focal plane, and each transform encodes the effect of a particular type of opto-mechanical distortion (e.g. relay optics, positioning errors of the GCU mask, etc.).

The transforms are parametrized by random variables, classified by the time at which they are sampled:

- (i) *Manufacturing time*. Used to represent manufacturing tolerances (e.g. the size of a particular mechanical element).
- (ii) *Calibration*. Used to model the repeatability of quantities that may change between two consecutive calibrations of the instrument.
- (iii) *Positioning*. Used to model quantities that change every time a measurement is performed.

`bayescal` (Carracedo-Carballal 2023) is a calibration process simulator written after the results of this characterization (Carracedo-Carballal et al. 2022). `bayescal` runs large batches of simulated calibrations, and it was used to produce the results of Section 7. It was designed to be small and to output easily reproducible results. It features a simplified instrumental model that captures the main traits of the instrumental pointing error as simulated by `harmoni-pm`.

This simplified instrumental model comprises the contribution of HARMONI's relay optics, positioning errors of the GCU mask, positioning accuracy of the guide probe, and angular errors of the instrument rotator. High-order distortions of the relay optics were included as a fully toleranced model of the FPRS mirrors. On the other hand, this instrumental model does not take into account systematic, mechanical effects of the guide probes.

## 4. PROBLEM FORMULATION

Let  $F \subseteq \mathbb{R}^2$  be the set of point coordinates contained in an idealized focal plane of the instrument. The points in this idealized focal plane are related to true sky coordinates by a simple transform involving the plate scale of the telescope. Let  $F' \subseteq F$  the set of point coordinates that can be actually observed by the instrument's detector, including the effects of the field distortion introduced by the different opto-mechanical elements in the light path. The instrumental distortion causes the observed locations  $\mathbf{x}' \in F'$  to be displaced with respect to their theoretical locations  $\mathbf{x} \in F$  as:

$$\mathbf{x}' = \mathbf{x} + \boldsymbol{\varepsilon}(\mathbf{x}). \quad (1)$$

The instrumental pointing error  $\boldsymbol{\varepsilon} \in \mathbb{R}^2$  can be seen as a function of  $\mathbf{x}$ , with  $\|\boldsymbol{\varepsilon}(\mathbf{x})\|_2$  a quantity much smaller than the focal plane dimensions. Requirements of the instrument impose smoothness conditions on  $\boldsymbol{\varepsilon}(\mathbf{x})$  and therefore  $\boldsymbol{\varepsilon}(\mathbf{x}) \approx \boldsymbol{\varepsilon}(\mathbf{x}')$  up to an error that is below the instrument's pointing accuracy.

The goal of the corrective model is, given certain maximum pointing error residual  $E > 0$ , to provide certain function  $\hat{\boldsymbol{\varepsilon}} : F \rightarrow \mathbb{R}^2$  such that  $\|\hat{\boldsymbol{\varepsilon}}(\mathbf{x}) - \boldsymbol{\varepsilon}(\mathbf{x})\|_2 < E, \forall \mathbf{x} \in F$ .

While the aforementioned maximum residual requirement applies for all points in  $F$ , the model still needs to be calibrated from a finite (and hopefully small) number of calibration points. These calibration points are provided by the GCU. During an instrument's calibration, the GCU deploys a mask with a pattern of illuminated points in the *ELT*'s focal plane. The position of these calibration points is assumed to be known with given accuracy.

The measurement devices for this task are the guide probes, that consist of two mirrors that can be placed anywhere in their corresponding half of the relayed focal plane, and image objects therein. The instrumental pointing error is simply the difference between the expected location of the reference object (in this case, the GCU's calibration points) and the location that is actually observed by the guide probe.

#### 4.1 The corrective model

We start by representing all two-dimensional (2D) vectors  $\mathbf{v} = (x, y) \in \mathbb{R}^2$  as complex quantities  $\hat{v} \in \mathbb{C}$  by means of the bijection:

$$\hat{v} = x + iy \quad (2)$$

where  $i$  is the imaginary unit. We therefore refer to the true instrumental pointing error in complex form as  $\hat{\boldsymbol{\varepsilon}}(\mathbf{x})$ . If  $\hat{\boldsymbol{\varepsilon}}(\mathbf{x})$  allows an asymptotic expansion of certain basis functions  $\hat{G}_j : F \rightarrow \mathbb{C}$  of the form:

$$\hat{\boldsymbol{\varepsilon}}(\mathbf{x}) = \sum_{j=0}^{\infty} \hat{G}_j(\mathbf{x}) \hat{\alpha}^j \quad (3)$$

we can formulate the corrective model as a truncated asymptotic expansion of order  $J$  that is fully parametrized by its complex coefficients  $\hat{\alpha}^j \in \mathbb{C}$ . A prediction of the pointing error  $\hat{\boldsymbol{\varepsilon}}(\mathbf{x})$  at the point  $\mathbf{x} \in F$  is obtained by evaluating

$$\hat{\boldsymbol{\varepsilon}}(\mathbf{x}) \approx \sum_{j=0}^{J-1} \hat{G}_j(\mathbf{x}) \hat{\alpha}^j. \quad (4)$$

#### 4.2 The classical calibration problem

The classical calibration problem exploits the linearity of equation (3) and ignores the effects of the measurement noise. Assuming that the distortion is well described by an expansion of the first  $J$  functions, the problem is posed as follows:

Let  $\mathbf{x}^i = (x, y)^\top \in F, 0 \leq i < J \leq K$  a set of  $K$  calibration points and  $\hat{\boldsymbol{\varepsilon}}^i = \hat{\boldsymbol{\varepsilon}}(\mathbf{x}^i) \in \mathbb{C}$  a set of  $K$  measurements of the pointing error in the calibration points in complex form. Find  $\hat{\alpha}^j \in \mathbb{C}, 0 \leq j < J$  such that the residual sum of squares

$$\text{RSS} = \sum_{i=0}^{K-1} \left| \hat{\boldsymbol{\varepsilon}}(\mathbf{x}^i) - \sum_{j=0}^{J-1} \hat{G}_j(\mathbf{x}^i) \hat{\alpha}^j \right|^2 \quad (5)$$

is minimal.

If we write  $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}^0, \hat{\alpha}^1, \dots, \hat{\alpha}^{J-1})^\top \in \mathbb{C}^J$ ,  $\hat{\boldsymbol{\varepsilon}} = (\hat{\boldsymbol{\varepsilon}}^0, \hat{\boldsymbol{\varepsilon}}^1, \dots, \hat{\boldsymbol{\varepsilon}}^{K-1})^\top \in \mathbb{C}^K$ , and  $\hat{\mathbf{G}} \in \mathbb{C}^{K \times J}$  as a matrix such that its  $i$ th row and  $j$ th column is  $\hat{G}_j^i = \hat{G}_j(\mathbf{x}^i)$ , equation (5) can be written in matrix form as

$$\text{RSS} = \left\| \hat{\boldsymbol{\varepsilon}} - \hat{\mathbf{G}} \hat{\boldsymbol{\alpha}} \right\|^2. \quad (6)$$

Matrix  $\hat{\mathbf{G}}$  is also called collocation matrix, by similarity with other interpolation problems. If  $J = K$  and  $\det \hat{\mathbf{G}} \neq 0$ ,  $\hat{\mathbf{G}}$  is invertible, a solution with null residual can be found:

$$\hat{\boldsymbol{\alpha}} = \hat{\mathbf{G}}^{-1} \hat{\boldsymbol{\varepsilon}}. \quad (7)$$

How well this solution extends to the rest of points in  $F$  depends on the choice of  $\hat{G}_j$ , the measurement noise, and how uniformly the calibration points  $\mathbf{x}^i$  are distributed in  $F$ . In general, uneven distributions of calibration points translate to ill-conditioned collocation matrices, which amplify measurement noise. In the extreme case in which two calibration points are the same, equation (5) refers to an underdetermined system of equations and the collocation matrix becomes singular. A good calibration pattern is therefore not only determined by the number of calibration points, but also by their impact on the condition number of the collocation matrix.

While this approach does not take measurement noise into account, its effect can be arbitrarily reduced by providing more calibration points ( $K > J$ ) and ensuring that the rank of  $\hat{\mathbf{G}}$  is at least  $J$ . In this case, the system of equations becomes overdetermined and ordinary least squares can be used to minimize equation (6).

#### 4.3 The Bayesian calibration problem

Let us now assume a certain abstract model  $M$  that makes predictions  $\hat{\mathcal{D}} = (\hat{x}^0, \hat{x}^1, \dots)^\top$  of certain observable from a given set of model parameters  $\boldsymbol{\vartheta} = (\vartheta^0, \vartheta^1, \dots)^\top$ . We also assume that these parameters were fitted from previous observations  $\mathcal{D} = (x^0, x^1, \dots)^\top$ .

In the context of Bayesian inference, model parameters are treated as joint distributions  $p(\boldsymbol{\vartheta})$  that encode the degree of belief we have on them. If the probability of observing certain  $\mathcal{D}$  is known when the parameters are known to be *exactly*  $\boldsymbol{\vartheta}$ , the Bayes theorem can be used to find the degree of belief over the coefficients *after* providing additional evidence  $\mathcal{D}$ :

$$p(\boldsymbol{\vartheta}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})}{p(\mathcal{D})}. \quad (8)$$

The probability distribution  $p(\boldsymbol{\vartheta})$  is called prior probability distribution of the parameters, and  $p(\boldsymbol{\vartheta}|\mathcal{D})$  the posterior probability distribution of the parameters. Note that  $p(\boldsymbol{\vartheta}|\mathcal{D})$  is a function evaluated in  $\boldsymbol{\vartheta}$ , leaving  $\mathcal{D}$  fixed. While  $p(\mathcal{D}|\boldsymbol{\vartheta})$  has the form of the probability distribution of measuring certain observation vector  $\mathcal{D}$ , it is evaluated on  $\boldsymbol{\vartheta}$  and therefore it is not a proper probability distribution. We call this function the likelihood function, and we make this distinction explicit by the notation

$$\mathcal{L}(\boldsymbol{\vartheta}|\mathcal{D}) = p(\mathcal{D}|\boldsymbol{\vartheta}). \quad (9)$$

Finally,  $p(\mathcal{D})$  is the probability distribution of observing  $\mathcal{D}$  for all possible choices of  $\boldsymbol{\vartheta}$  and it is generally difficult to calculate. Since  $p(\mathcal{D})$  does not depend on  $\boldsymbol{\vartheta}$ , we can treat it as a normalization constant and write equation (8) as

$$p(\boldsymbol{\vartheta}|\mathcal{D}) \propto \mathcal{L}(\boldsymbol{\vartheta}|\mathcal{D})p(\boldsymbol{\vartheta}). \quad (10)$$

Posterior probability distributions can also be used as prior distributions for subsequent Bayesian updating steps, which enables iterative refining of the parameter knowledge as additional evidence is provided.

Identifying  $\boldsymbol{\vartheta}$  with the complex model coefficients  $\hat{\boldsymbol{\alpha}} \in \mathbb{C}^J$  and  $\mathcal{D}$  with both the calibration points in complex form  $\hat{\boldsymbol{p}} = (\hat{x}^0, \dots, \hat{x}^{K-1})^\top$  and its corresponding observations of the pointing error  $\hat{\boldsymbol{\varepsilon}} \in \mathbb{C}^K$ , we can formulate the Bayesian calibration problem as follows:

Given a set of observations of the pointing error  $(\hat{\boldsymbol{\varepsilon}}, \hat{\boldsymbol{p}})$  and prior knowledge of the model coefficients  $p(\hat{\boldsymbol{\alpha}})$ , find the posterior probability distribution of the parameters  $p(\hat{\boldsymbol{\alpha}}|\hat{\boldsymbol{\varepsilon}}, \hat{\boldsymbol{p}})$  as

$$p(\hat{\boldsymbol{\alpha}}|\hat{\boldsymbol{\varepsilon}}, \hat{\boldsymbol{p}}) = \frac{p(\hat{\boldsymbol{\varepsilon}}, \hat{\boldsymbol{p}}|\hat{\boldsymbol{\alpha}})p(\hat{\boldsymbol{\alpha}})}{p(\hat{\boldsymbol{\varepsilon}}, \hat{\boldsymbol{p}})}. \quad (11)$$

As an additional simplifying assumption, we will consider the calibration points  $\hat{\boldsymbol{p}}$  to be known up to arbitrary precision. This is a justifiable assumption, as the physical location of points is planned to be measured by means of microscopy, before the commissioning phase. Since the GCU mask is our ultimate source of truth, errors in this characterization will propagate directly as errors of the corrective model.

If we write all probabilities of the form  $p(\hat{\boldsymbol{\varepsilon}}, \hat{\boldsymbol{p}}|\hat{\boldsymbol{\alpha}})$  as  $p(\hat{\boldsymbol{\varepsilon}}|\hat{\boldsymbol{p}}, \hat{\boldsymbol{\alpha}})p(\hat{\boldsymbol{p}})$ , we formalise this assumption by modelling  $p(\hat{\boldsymbol{p}})$  as a Dirac delta distribution centred in the nominal calibration points. This enables us to marginalize equation (11) on  $\hat{\boldsymbol{p}}$ , from which we conclude that  $p(\hat{\boldsymbol{\varepsilon}}|\hat{\boldsymbol{\alpha}}) = p(\hat{\boldsymbol{\varepsilon}}|\hat{\boldsymbol{p}}, \hat{\boldsymbol{\alpha}})$ . Since this dependency on  $\hat{\boldsymbol{p}}$  appears in all probabilities involving  $\hat{\boldsymbol{\varepsilon}}$ , we will make it explicit only when further clarification is needed. Equation (11) can therefore be simplified as

$$p(\hat{\boldsymbol{\alpha}}|\hat{\boldsymbol{\varepsilon}}) = \frac{p(\hat{\boldsymbol{\varepsilon}}|\hat{\boldsymbol{\alpha}})p(\hat{\boldsymbol{\alpha}})}{p(\hat{\boldsymbol{\varepsilon}})}. \quad (12)$$

#### 4.4 Likelihood function

The most fundamental ingredient of Bayesian inference is an accurate description of the likelihood function. This method assumes a Gaussian likelihood (which is rather common for measurement devices with multiple contributions to the measurement error). One must note, however, that this is not true in all scenarios. The applicability of this method is tied to a proof of the Gaussianity of the measurement error.

For the sake of simplicity, we start by rewriting the complex arithmetic in equation (6) in terms of real-valued linear operations. This can be done by alternating real and imaginary parts of complex vectors, and encoding the complex arithmetic in a real-valued collocation matrix as described in Appendix A. In this form, the components of  $\boldsymbol{p}$  consist of alternating horizontal and vertical coordinates of the calibration points, and components of  $\boldsymbol{\varepsilon}$  of alternating horizontal and vertical projections of the pointing error.

Assuming Gaussianity, the probability of observing a pointing error  $\varepsilon^i$  in one of the two axes of certain measurement point is given by

$$\varepsilon^i \sim \mathcal{N} \left( \sum_{j=0}^{2J-1} G_j^i \alpha^j, \sigma_E^2 \right) \quad (13)$$

where  $G_j^i$  are the coefficients of the real-valued collocation matrix  $\boldsymbol{G} \in \mathbb{R}^{2K \times 2J}$ , and are calculated as described in equations (A13)–(A16). This implies that the coefficients  $G_j^i$  are indeed a function of  $p^j$ . The likelihood function  $\mathcal{L}(\boldsymbol{\alpha}|\varepsilon^i, p^i, \sigma_E^2)$  is simply the probability density  $p(\varepsilon^i|\boldsymbol{\alpha}, p^i, \sigma_E^2)$  evaluated in  $\boldsymbol{\alpha}$ .

If we can guarantee the independence of measurements, we can formulate the likelihood function with respect to the measurement vector  $\boldsymbol{\varepsilon}$  as simply the product of the likelihoods of the individual measurements:

$$\mathcal{L}(\boldsymbol{\alpha}|\boldsymbol{\varepsilon}, \boldsymbol{p}, \sigma_E^2) = \prod_{i=0}^{2N-1} \mathcal{L}(\boldsymbol{\alpha}|\varepsilon^i, p^i, \sigma_E^2). \quad (14)$$

The likelihood function can therefore be expressed as multivariate Gaussian distribution of mean  $\boldsymbol{G}\boldsymbol{\alpha}$  and covariance matrix  $\sigma_E^2 \mathbb{I}$ , where

$\mathbb{I}$  stands for the identity matrix. The probability density is therefore written as

$$\mathcal{L}(\boldsymbol{\alpha}|\boldsymbol{\varepsilon}, \boldsymbol{p}, \sigma_E^2) = \frac{1}{\sigma_E \sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma_E^2} (\boldsymbol{\varepsilon} - \boldsymbol{G}\boldsymbol{\alpha})^\top (\boldsymbol{\varepsilon} - \boldsymbol{G}\boldsymbol{\alpha}) \right]. \quad (15)$$

## 5. ANALYTIC SOLUTION

In many applications, posterior distributions can only be estimated by means of Monte-Carlo methods which are computationally expensive. None the less, there are particular instances in which the posterior distribution admits a closed-form expression that is of the same form as that of the prior distribution. This occurs for certain combinations of prior distributions and likelihood functions when they are of the exponential family (e.g. Gaussian, Gamma, exponential, etc.). We say in this case that the prior and posterior are conjugate probability distributions, and the prior  $p(\boldsymbol{\theta})$  is said to be a conjugate prior.

Simulated classical calibrations (Carracedo-Carballal et al. 2022) showed that the distribution of model coefficients is also well described by a multivariate Gaussian distribution of the form

$$\boldsymbol{\alpha} \sim \mathcal{MN}(\bar{\boldsymbol{\alpha}}, \boldsymbol{\Sigma}) \quad (16)$$

where  $\bar{\boldsymbol{\alpha}} \in \mathbb{R}^{2J}$  is the mean vector and  $\boldsymbol{\Sigma} \in \mathbb{R}^{2J \times 2J}$  the covariance matrix of the distribution. If these parameters refer to those of a prior or a posterior distribution, they are called hyperparameters.

If we write the prior distribution  $p(\boldsymbol{\alpha})$  at calibration the step  $n$  as

$$p(\boldsymbol{\alpha}) = \mathcal{MN}(\boldsymbol{\alpha}; \boldsymbol{\alpha}_n, \boldsymbol{\Sigma}_n) \quad (17)$$

and the corresponding posterior distribution after providing additional observations of the pointing error  $\boldsymbol{\varepsilon}$  as

$$p(\boldsymbol{\alpha}|\boldsymbol{\varepsilon}) = \mathcal{MN}(\boldsymbol{\alpha}; \boldsymbol{\alpha}_{n+1}, \boldsymbol{\Sigma}_{n+1}). \quad (18)$$

It can be proven (see Appendix B1) that the posterior hyperparameters can be calculated as

$$\boldsymbol{\Sigma}_{n+1} = (\boldsymbol{G}^\top \sigma_E^{-2} \boldsymbol{G} + \boldsymbol{\Sigma}_n^{-1})^{-1} \quad (19)$$

$$\boldsymbol{\alpha}_{n+1} = \boldsymbol{\Sigma}_{n+1} (\boldsymbol{G}^\top \sigma_E^{-2} \boldsymbol{\varepsilon} + \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\alpha}_n) \quad (20)$$

which reduces to the classical calibration when  $K = J$  and the prior hyperparameters describe an infinitely uninformative distribution.<sup>1</sup>

In a realistic application of this calibration technique, we could obtain a good representative of the model coefficients from the mean  $\boldsymbol{\alpha}_{n+1}$  and a measure of their uncertainty from their individual variances (which equal to the diagonal elements of  $\boldsymbol{\Sigma}_{n+1}$ ).

### 5.1 Prediction of model residuals

While the degree of uncertainty of the  $\boldsymbol{\alpha}$  is contained in  $\boldsymbol{\Sigma}_{n+1}$ , it is usually more relevant to understand how it translates to uncertainty in the pointing error. This result can be obtained from the variance of the posterior predictive distribution:

$$p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon}) = p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}|\boldsymbol{\varepsilon}) \quad (21)$$

where  $\tilde{\boldsymbol{\varepsilon}}$  denotes a prediction made by the corrective model, given that previous evidence  $\boldsymbol{\varepsilon}$  has been provided. Note that  $p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\alpha})$  has the form of the likelihood function and  $p(\boldsymbol{\alpha}|\boldsymbol{\varepsilon})$  is conjugate for this function. It can be proven (see Appendix B3) that  $p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon})$  is another multivariate Gaussian of the form

$$\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon} \sim \mathcal{MN}(\tilde{\boldsymbol{G}}\boldsymbol{\alpha}_{n+1}, \sigma_E^2 \mathbb{I} + \tilde{\boldsymbol{G}}\boldsymbol{\Sigma}_{n+1}\tilde{\boldsymbol{G}}^\top) \quad (22)$$

<sup>1</sup>Proof in Appendix B2.

where  $\tilde{\mathbf{G}}$  is the collocation matrix evaluated in the test points used for the prediction. The covariance matrix  $\sigma_{\mathbf{e}}^2 \mathbb{1} + \tilde{\mathbf{G}} \boldsymbol{\Sigma}_{n+1} \tilde{\mathbf{G}}^\top$  is hereinafter referred to as the model uncertainty covariance matrix and will be denoted by  $\tilde{\mathbf{U}}_{n+1}$ .

The variance of the predicted pointing error is simply calculated as

$$\sigma^2[\tilde{\boldsymbol{\epsilon}}|\boldsymbol{\epsilon}] = \text{diag}(\tilde{\mathbf{U}}_{n+1}). \quad (23)$$

We remark from this equation that the variance is lower bounded by the measurement noise variance  $\sigma_{\mathbf{e}}^2$ . This is expected, as the only way to access pointing error measurements is by means of a measurement device (the guide probe) with finite precision.

This result can be used as part of a termination condition, in which new observations of the pointing error update the hyperparameters iteratively until the degree of certainty of the true pointing error reaches a threshold value (e.g.  $\|\sigma^2[\tilde{\boldsymbol{\epsilon}}|\boldsymbol{\epsilon}]\|_\infty < \sigma_{\max}^2$ , with  $\sigma_{\max}^2$  certain maximum variance of the pointing error in the test points).

We must take into account that the Bayesian updating step in equation (20) depends on the particular order in which new measurements have been presented to the model. This also implies that batched updatings (i.e. presenting  $K > 1$  measurements in the measurement vector  $\boldsymbol{\epsilon}$  at once) are not formally equivalent to sequential updatings (i.e. performing  $K$  updating steps with one measurement each).

This dependency on ordering is particularly pronounced in the first updatings with vague priors, as the first observations have a larger relative impact on the posterior than subsequent ones. If the prior is too vague, noisy measurements may introduce an undesired shrinkage that bias future updatings towards erroneous values.

In case uninformative priors are not available, this effect can be mitigated by reordering calibration points in a way that the first ones maximize their distance with respect to the next. The rationale of this strategy is to leverage the first steps to integrate distortion information from the whole field as much as possible. Taking points too close together may give excessive weight to certain correlations, shrinking future observations towards wrong directions.

### 5.1.1 Mean square error and root mean square error

In many practical applications, however, the mean square error (MSE) and the root mean square error (RMS) are the preferred metrics for model performance:

$$\text{RMS} = \sqrt{\frac{1}{K} \sum_{i=0}^{K-1} r_{2i}^2 + r_{2i+1}^2} = \sqrt{\text{MSE}} = \sqrt{\frac{\text{RSS}}{K}} \quad (24)$$

with  $r_i$  the per-axis real-valued residuals between the prediction and the measured pointing error. The uncertainties in the model parameters and the measurement noise can be used to model the residual vector  $\mathbf{r} = (r_0, r_1, \dots, r_{2K-1})^\top$  as a zero-mean multivariate Gaussian:

$$\mathbf{r} \sim \mathcal{MN}(\mathbf{0}, \tilde{\mathbf{U}}_{n+1}) \quad (25)$$

which can be used to provide upper bounds both on the mean and variance of the MSE and the RMS. If we consider the MSE a random variable calculated as

$$\text{MSE} = \frac{1}{K} \sum_{i=0}^{2K-1} r_i^2 \quad (26)$$

given that the components of  $\mathbf{r}$  come from a multivariate Gaussian, the distribution of the MSE is that of a generalized central  $\chi^2$

distribution, with no Gaussian contribution:

$$\text{MSE} \sim \chi^2 \left( w_i = \frac{\lambda_i}{K}, m = 0, s^2 = 0 \right) \quad (27)$$

where  $\lambda_i$  are the  $2K$  eigenvalues of  $\tilde{\mathbf{U}}_{n+1}$ . The expected value and variances of the MSE are given by

$$\begin{aligned} \mathbb{E}[\text{MSE}] &= \frac{1}{K} \sum_{i=0}^{2K-1} \lambda_i \\ \text{Var}[\text{MSE}] &= \frac{2}{K^2} \sum_{i=0}^{2K-1} \lambda_i^2. \end{aligned} \quad (28)$$

If the uncertainty is sufficiently big (e.g. in the earliest steps of the calibration), the expected value and variances of the RMS can be approximated by uncertainty propagation of the MSE as

$$\begin{aligned} \mathbb{E}[\text{RMS}] &\approx \sqrt{\mathbb{E}[\text{MSE}]} \\ \text{Var}[\text{RMS}] &\approx \frac{1}{4} \frac{\text{Var}[\text{MSE}]}{\mathbb{E}[\text{MSE}]}. \end{aligned} \quad (29)$$

We must remark, none the less, that this approach will introduce a slight amount of overestimation as it assumes that we are sampling  $\boldsymbol{\alpha}$  for every estimation and not taking its mean value  $\boldsymbol{\alpha}_{n+1}$  directly.

## 5.2 Repeatability priors

During the calibration procedure, posterior distributions obtained from the previous measurements are reinterpreted as prior distributions of the next. After the termination condition has been met, we can choose the mean  $\hat{\boldsymbol{\alpha}}_{n+1} \in \mathbb{C}^J$  as a good representative of the model coefficients. None the less, we still have to define the hyperparameters of the initial prior  $p(\boldsymbol{\alpha}) = p(\boldsymbol{\alpha}|\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0)$ , used before performing the first measurement.

A reasonable choice is to obtain them from the repeatability of classically calibrated model coefficients as observed after certain number  $M$  of calibration sessions:

$$\boldsymbol{\alpha}_0 = \bar{\boldsymbol{\alpha}} \quad (30)$$

$$\boldsymbol{\Sigma}_0 = \boldsymbol{\mathcal{Q}} \quad (31)$$

where  $\bar{\boldsymbol{\alpha}}$  is the sample mean of the coefficients and  $\boldsymbol{\mathcal{Q}}$  is the sample covariance matrix, both estimated from a sample of  $M$  model coefficient vectors.

This choice is motivated by the fact that, in some scenarios, an appropriate choice of basis functions in the asymptotic expansions can be used to distinguish between contributions of different origin and repeatability. This is particularly true for HARMONI, in which optical distortions (which are described by the higher order terms of the PZP expansion) have much smaller variability than mechanical distortions (which mostly affect the lowest order terms).

On the other hand, this choice implicitly assumes that the first  $M$  calibrations are good representatives of future calibrations, which is not necessarily true. We can overcome this limitation by reinterpreting  $\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0$  as random variables conditioned by observed calibrated coefficients  $\boldsymbol{\alpha}$ , providing us with a *Bayesianly justifiable* method (Rubin 1984) to update our knowledge of the repeatability of the distortion. From Bayes' theorem, it follows that

$$p(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0|\boldsymbol{\alpha}) = \frac{p(\boldsymbol{\alpha}|\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0) p(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0)}{p(\boldsymbol{\alpha})} \quad (32)$$

where the likelihood function  $p(\boldsymbol{\alpha}|\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0) = \mathcal{L}(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0|\boldsymbol{\alpha})$  is simply the prior distribution of the coefficients:

$$\mathcal{L}(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0|\boldsymbol{\alpha}) = \frac{1}{\sqrt{(2\pi)^{2J} \det \boldsymbol{\Sigma}_0}} \exp \left[ -\frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \right]. \quad (33)$$

We call the distribution  $p(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0)$  a *repeatability prior*, as it reflects the repeatability of measurements between two consecutive calibration sessions.

We have, in principle, certain freedom to choose the form of the prior  $p(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0)$ . If we wish to leverage the algebraic properties of conjugate prior distributions, we can model the joint probability of  $\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0$  as a Gaussian–Inverse–Wishart distribution:

$$\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0 \sim \text{GIW}(\boldsymbol{\mu}_m, \lambda_m, \boldsymbol{\Psi}_m, \nu_m) \quad (34)$$

which is a conjugate prior for the likelihood described in equation (33).  $\boldsymbol{\mu}_m \in \mathbb{R}^{2J}$  is the mean vector of  $\boldsymbol{\alpha}_0$ , and  $\boldsymbol{\Psi}_m \in \mathbb{R}^{2J \times 2J}$  is the inverse scale matrix (which is positive definite).  $\lambda_m > 0$  and  $\nu_m > 2J - 1$  are pseudo-counter parameters that weight the effect of new observations of  $\boldsymbol{\alpha}$  in the calculation of the posterior distribution.

The least informative prior has  $\lambda_0 = 1$  and  $\nu_0 = 2J + 1$ , and could be initialized from a zero mean and diagonal  $\boldsymbol{\Psi}_0$ . Unfortunately, this choice exhibits slow convergence to the true covariance matrix, depriving the algorithm of quickly leveraging the prior information from previous calibrations.

A faster, albeit more frequentist approach, is to rely on Parametric Empirical Bayes (PEB) for prior generation. Under PEB, prior hyperparameters  $\boldsymbol{\mu}_0$  and  $\boldsymbol{\Psi}_0$  are derived directly from maximum-likelihood estimators (MLEs) of previous observations (equations 30–31) as

$$\lambda_0 = M \quad (35)$$

$$\nu_0 = 2J + M \quad (36)$$

$$\boldsymbol{\mu}_0 = \bar{\boldsymbol{\alpha}} \quad (37)$$

$$\boldsymbol{\Psi}_0 = (\nu_0 - 2J - 1)\boldsymbol{Q} \quad (38)$$

in which we identified  $\boldsymbol{Q}$  with the mean variance. We also leveraged  $\nu_0$  and  $\lambda_0$  to account for the strength of the evidence used as prior ( $M$  observations).

The analytic formula for the Bayesian update step has been described by Gelman et al. (2021). If  $\boldsymbol{\alpha}^0, \dots, \boldsymbol{\alpha}^{n-1}$  are the calibrated coefficients obtained from  $n$  calibrations, the hyperparameters in equation (34) are updated as

$$\boldsymbol{\mu}_{m+1} = \frac{\lambda_m}{\lambda_m + n} \boldsymbol{\mu}_m + \frac{n}{\lambda_m + n} \bar{\boldsymbol{\alpha}} \quad (39)$$

$$\lambda_{m+1} = \lambda_m + n \quad (40)$$

$$\nu_{m+1} = \nu_m + n \quad (41)$$

$$\boldsymbol{\Psi}_{m+1} = \boldsymbol{\Psi}_m + S + \frac{\lambda_m n}{\lambda_m + n} (\bar{\boldsymbol{\alpha}} - \boldsymbol{\mu}_m)(\bar{\boldsymbol{\alpha}} - \boldsymbol{\mu}_m)^\top \quad (42)$$

where  $S$  is the sum of squares matrix about the sample mean:

$$S = \sum_{i=0}^{n-1} (\boldsymbol{\alpha}^i - \bar{\boldsymbol{\alpha}})(\boldsymbol{\alpha}^i - \bar{\boldsymbol{\alpha}})^\top. \quad (43)$$

Note that for the case of sequential updating ( $n = 1$ ),  $\bar{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$  and  $S$  is zero. While choosing sequential over batched updating should not affect the resulting priors, it can be desirable to benefit from the

most up-to-date repeatability prior before each calibration. In the next calibration, the hyperparameters  $\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0$  of the calibration priors would be drawn from the updated the Gaussian–Inverse–Wishart. The sampling procedure is documented by Gelman et al. (2021) and is performed in two steps: first, draw  $\boldsymbol{\Sigma}_0|\boldsymbol{\alpha} \sim \text{Inv-Wishart}(\boldsymbol{\Psi}_{m+1}, \nu_{m+1})$  and then draw  $\boldsymbol{\alpha}_0|\boldsymbol{\Sigma}_0, \boldsymbol{\alpha} \sim \mathcal{MN}(\boldsymbol{\mu}_{m+1}, \boldsymbol{\Sigma}_0/\lambda_{m+1})$ .

### 5.3 Calibration algorithm

We are now ready to formulate the full calibration algorithm. The algorithm assumes that the repeatability prior hyperparameters  $(\boldsymbol{\mu}_m, \boldsymbol{\Psi}_m, \lambda_m, \nu_m)$  can be saved and restored from permanent storage through certain routines named `load` and `save`. The algorithm also assumes that at least  $M$  previous classical calibrations were performed, from which  $\bar{\boldsymbol{\alpha}}$  and  $\boldsymbol{Q}$  were obtained.

---

#### Algorithm 1 Bayesian calibration algorithm

---

**Require:**  $\bar{\boldsymbol{\alpha}}$  and  $\boldsymbol{Q}$  obtained from the first  $M$  classical calibrations.  
**Require:**  $\sigma_E^2 > 0$   
**Require:**  $\sigma_{\max}^2 > 0$   
**Require:**  $N > 0$   
**Require:**  $\{p_0, \dots, p_{N-1}\} \subset F$

**if** first run **then**  
      $m \leftarrow 0$   
      $\mu_0 \leftarrow \bar{\boldsymbol{\alpha}}$   
      $\boldsymbol{\Psi}_0 \leftarrow \boldsymbol{Q}$   
      $\lambda_0 \leftarrow M$   
      $\nu_0 \leftarrow 2J + M$

**else**  
     `load` ( $m, \mu_m, \boldsymbol{\Psi}_m, \lambda_m, \nu_m$ )

**end if**  
 $n \leftarrow 0$   
draw  $(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0)$  from  $\text{GIW}(\mu_m, \boldsymbol{\Psi}_m, \lambda_m, \nu_m)$   
**while**  $\|\sigma^2[\bar{\boldsymbol{\epsilon}}|\boldsymbol{\varepsilon}]\|_\infty \geq \sigma_{\max}^2$  **do**  
     **if**  $n = N$  **then** ▷ Calibration failed  
         terminate with error condition  
     **end if**  
     measure  $\varepsilon_n$  in  $p_n$   
      $\boldsymbol{\Sigma}_{n+1} \leftarrow (G\sigma_E^{-2}G + \boldsymbol{\Sigma}_n^{-1})^{-1}$   
      $\boldsymbol{\alpha}_{n+1} \leftarrow \boldsymbol{\Sigma}_{n+1} (G\sigma_E^{-2}\varepsilon_n + \boldsymbol{\Sigma}_n^{-1}\boldsymbol{\alpha}_n)$   
      $n \leftarrow n + 1$

**end while**  
accept  $\boldsymbol{\alpha}_n$  as model coefficients ▷ Calibration successful  
draw  $\boldsymbol{\alpha}$  from  $\mathcal{MN}(\boldsymbol{\alpha}_n, \boldsymbol{\Sigma}_n)$   
 $\mu_{m+1} \leftarrow \frac{\lambda_m \mu_m + \boldsymbol{\alpha}}{\lambda_m + 1}$   
 $\lambda_{m+1} \leftarrow \lambda_m + 1$   
 $\nu_{m+1} \leftarrow \nu_m + 1$   
 $\boldsymbol{\Psi}_{m+1} \leftarrow \boldsymbol{\Psi}_m + \frac{\lambda_m}{\lambda_m + 1} (\boldsymbol{\alpha} - \mu_m)(\boldsymbol{\alpha} - \mu_m)$   
 $m \leftarrow m + 1$   
`save` ( $m, \mu_m, \boldsymbol{\Psi}_m, \lambda_m, \nu_m$ )

---

## 6. APPLICATION TO HARMONI'S FIELD DISTORTION

Before applying this method, one must provide evidence for the Gaussianity of the likelihood function. Simulations performed by the pointing simulator `harmoni-pm` (Carracedo-Carballal et al. 2022) showed that the measurement process is well described by a Gaussian with certain measurement noise variance  $\sigma_E^2$  for both axes (Appendix C), which also motivates a Gaussian likelihood.

Next, one must choose basis functions  $\hat{G}_j$  of the asymptotic expansion of the error and a good calibration pattern. While the latter is not as critical as the former, it has an impact on the number of points that are eventually required to calibrate the instrument up to the desired accuracy.

### 6.1 Basis functions

Due to the circular geometry of HARMONI's focal plane, we propose a corrective model consisting of a truncated expansion of PZP of the pointing error. PZPs are enumerated by two indices: the radial order  $n$  and the azimuthal order  $m$ . Biesheuvel et al. (2018) formulate these polynomials in polar coordinates as

$$\hat{Z}_n^m(\rho, \theta) = R_n^{|m|}(\rho)e^{im\theta}, \forall \rho \in [0, 1], \theta \in [-\pi, \pi] \quad (44)$$

where  $\rho = |\hat{v}| = (x^2 + y^2)^{\frac{1}{2}}$  and  $\theta = \arg(\hat{v}) = \arctan\frac{y}{x}$ . The radial polynomial  $R_n^{|m|}$  is a real function of  $\rho$  that can be written in terms of the Jacobi polynomials  $P_n^{(\alpha, \beta)}(x)$  as

$$R_n^{|m|}(\rho) = P_{\frac{n-|m|}{2}}^{(0, |m|)}(2\rho^2 - 1)\rho^{|m|}. \quad (45)$$

Our method requires the basis functions to be enumerated by a single index  $0 \leq j < J$ . While multiple indexations are possible, in this article we will refer to the so-called OSA/ANSI indices (Thibos et al. 2002), which relate  $m$ ,  $n$ , and  $j$  by

$$j = \frac{n(n+2) + m}{2}. \quad (46)$$

This choice has the advantage of sorting PZPs in increasing radial order. This ensures that, as long as  $J$  is a triangular number, all azimuthal orders up to certain radial order are contained in the expansion. In particular, first order distortions are encoded by the first three PZPs:  $\hat{Z}_0$  describes a displacement in the focal plane,  $\hat{Z}_1$  a change in the aspect ratio, and  $\hat{Z}_2$  any combination of a rotation and a rescaling of the focal plane.

As PZPs are defined in the unit disc, our basis functions must be evaluated in normalized focal plane as

$$\hat{G}_j(\mathbf{x}) = \hat{Z}_j\left(\frac{\mathbf{x}}{R}\right). \quad (47)$$

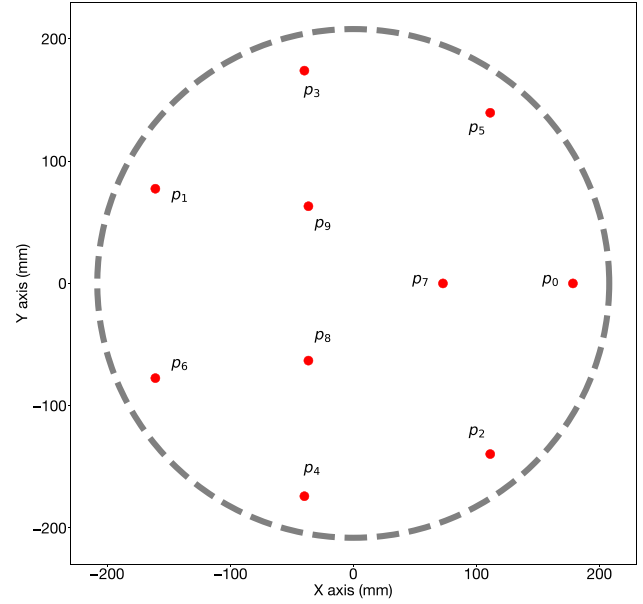
### 6.2 Calibration pattern

The choice of the calibration pattern is not only given by the basis functions, but also by practical constraints imposed by the instrument design. The performance of several calibration patterns was evaluated in a previous work (Carracedo-Carballal et al. 2022): dense square grid of points (of around 400 points) had the best performance, at the cost of a long calibration time. Spiral-like patterns behave well in the  $J = K$  case for low-order realizations of the model, but tend to leave highly undersampled regions in higher orders.

The best trade-off between performance and number of points was given by the Optimal Concentric Sampling (OCS) pattern. This pattern, originally described by Ramos-López et al. (2016), optimizes the condition number of the collocation matrices of the Zernike expansion, and it is parametrized by the truncation order of the expansion. For a corrective model with  $J = 10$  complex coefficients, the corresponding pattern takes the form shown in Fig. 1.

## 7. RESULTS

The performance of the algorithm was evaluated against simulated distortions between the *ELT*'s focal plane and the relayed focal plane, using *bayescal*. The goals of these simulations were:



**Figure 1.** OCS pattern with 10 calibration points. The sub-indices represent the order in which the calibration points are traversed.

- (i) Validate the ability of the model to produce calibration coefficients meeting certain accuracy goal.
- (ii) Compare the predicted uncertainties of the model with the model residuals after calibration.
- (iii) Study the impact of the number of classical calibrations used as prior evidence of the repeatability prior.
- (iv) Compare the performance of the Bayesian calibration algorithm with respect to the classical calibration technique.

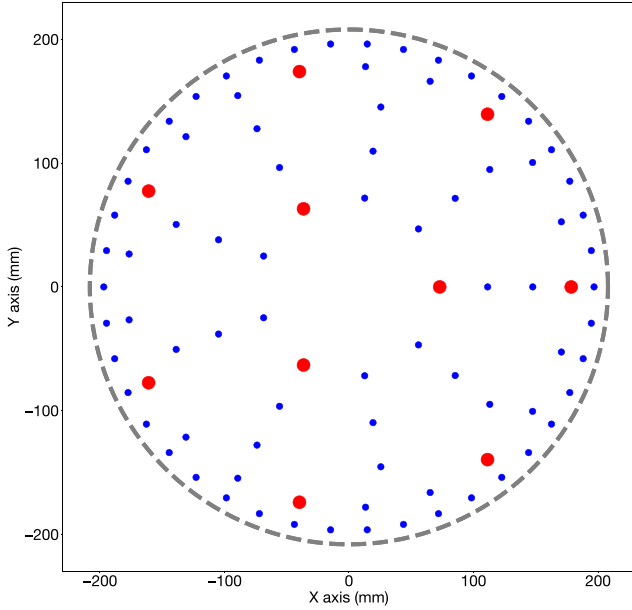
*bayescal* models the total distortion as a composition of three  $\mathbb{R}^2 \rightarrow \mathbb{R}^2$  coordinate transforms, representing how 2D coordinates are displaced between focal planes:

$$\mathbf{x}' = T_\phi \circ T_{\Delta\mathbf{x}} \circ T_{\text{fprs}}[\mathbf{x}]. \quad (48)$$

These transforms can be split into a fixed part and a variable part. The fixed part corresponds to  $T_{\text{fprs}}$  and was modelled after a Finite Element Method (FEM) analysis of FPRS distortions up to the 3rd radial order (introducing pointing errors of order 100  $\mu\text{m}/30$  mas). The variable part is used to represent the repeatability of measurements between calibrations, and consists of a random 2D displacement  $T_{\Delta\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$  composed with a random 2D rotation  $T_\phi = R(\phi)\mathbf{x}$ .

In order to highlight the ability of the model to distinguish between fixed and variable contributions to the pointing error, the effect of the latter was artificially increased to surpass the former. In particular, coordinates of the displacement  $\Delta\mathbf{x} = (\Delta x, \Delta y)^\top$  are drawn from zero-mean Gaussian distributions, with  $\sigma_{\Delta x} = 1$  mm and  $\sigma_{\Delta y} = 2$  mm. The angle of the rotation is also drawn from a zero-mean Gaussian distribution, with  $\sigma_\phi = 30$  arcsec.

The measurement noise refers to the uncertainty of the technique used to locate the guide star's centre of the imaged star. It depends on contributions like the detector's resolution, fitting algorithm, or signal-to-noise ratio of the acquired image. We chose a realistic, isotropic Gaussian measurement noise of  $\sigma_x = \sigma_y = \sigma_E = 0.2$   $\mu\text{m}$ , resulting in a total measurement noise variance of  $\sigma_n^2 = \sigma_x^2 + \sigma_y^2 \approx 8 \times 10^{-2}$   $\mu\text{m}^2$ . The calibration goal was set to  $\sigma_{\text{max}} = 6$   $\mu\text{m}$ , corre-



**Figure 2.** Original OCS pattern (larger dots) augmented with redundant points (smaller dots). This redundancy was achieved by multiplying the density of points in the azimuthal and radial directions by 3.

sponding to the tracking accuracy requirement of the guide probes (Estrada et al. 2022).

Since calibrations assumed a model with  $J = 10$ , the same 10-point OCS calibration pattern shown in Fig. 1 was used. Additionally, in a real-world scenario, we should expect that some of the optical fibres composing the points of the GCU mask may be damaged during the life time of the instrument. As a consequence, the definitive GCU mask pattern must contain redundant points that can replace the failing ones.

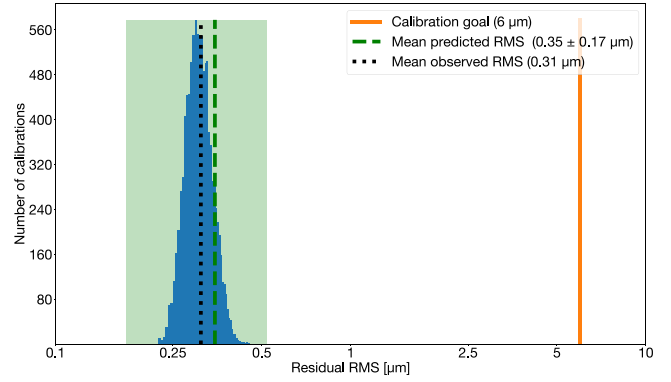
Validations were performed against a redundant version of the same pattern with  $K = 90$  calibration points in total, distributed in the azimuthal and radial directions (Fig. 2). Calibration points were ordered to maximize their distance with respect to the previous, improving the quality of the sampling during the first steps of the calibration. The resulting order is reflected by the sub-indices of the points in Fig. 1.

### 7.1 Accuracy testing

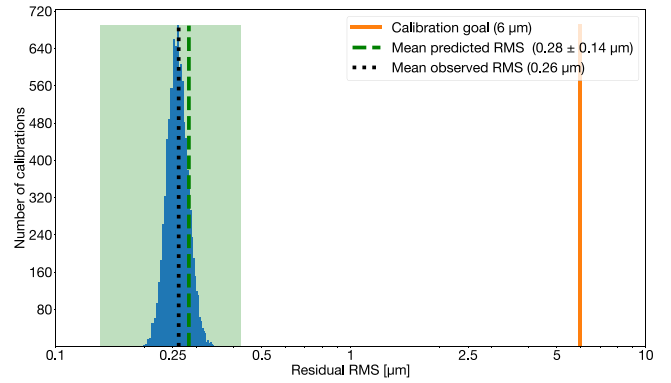
For accuracy testing, we forced a non-informative prior for  $\alpha_0$ ,  $\Sigma_0$  in all calibrations, with  $\mu_0 = \mathbf{0}$  and  $\Psi_0 = (100 \mu\text{m})^2 \mathbb{I}$ . This resulted in the algorithm requiring  $J$  measurements prior to having a good description of the field distortion.

A total of  $10^4$  calibrations were performed, with different realizations of the distortion each. The performance was evaluated as the RMS measured in all 90 calibration points, and compared against the RMS predictions described in equation (29).

The result of this test is reflected in Fig. 3. We see not only that the calibration meets the goal, but also that the predicted RMS ( $\approx 0.35 \pm 0.17 \mu\text{m}$ ) overestimates the observed mean RMS ( $\approx 0.31 \mu\text{m}$ ) by small amount. We connect this excess uncertainty to the vague priors used for testing: increasing the number of measurements per point, we observe a reduction in the uncertainties that pushes the prediction closer to the observed values (Fig. 4).



**Figure 3.** Histogram of the RMS of the residual after  $10^4$  simulated calibrations. All 10 calibration points, one measurement per calibration point, no prior information. The dashed line represents the mean expectation of the predicted RMS. The box represents the standard deviation of the uncertainty of the RMS prediction.



**Figure 4.** Histogram of the RMS of the residual after  $10^4$  simulated calibrations. All 10 calibration points, two measurements per calibration point, no prior information.

### 7.2 Uncertainty prediction

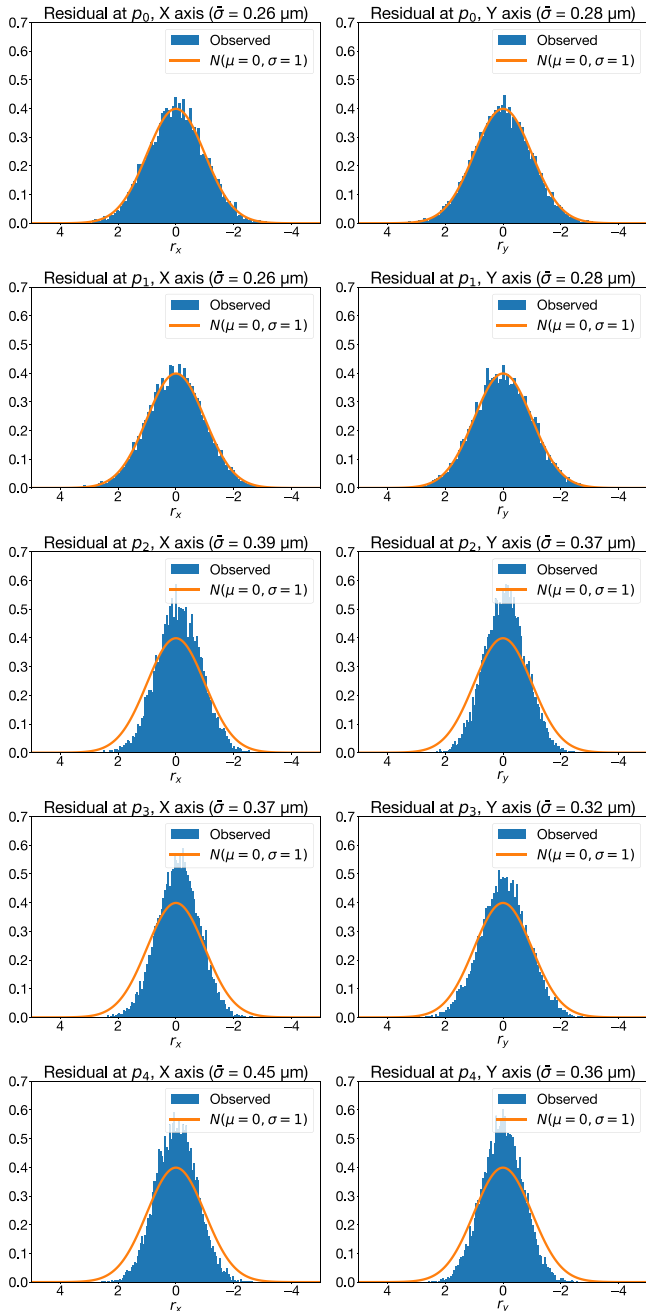
A critical feature of this algorithm is its ability to make informed predictions of its own uncertainty. For this analysis, we trained the model with the result of  $M = 100$  classical calibrations, starting with  $\mu_0 = \mathbf{0}$  and  $\Psi_0 = (100 \mu\text{m})^2 \mathbb{I}$  and updating the repeatability priors as usual, iterating the algorithm along  $10^4$  simulated calibrations. Let  $\varepsilon_x, \varepsilon_y$  be the horizontal and vertical components of the measured pointing error, and  $\tilde{\varepsilon}_x, \tilde{\varepsilon}_y$  the random variables predicted by the pointing model. Since the uncertainty decreases as the number of accepted calibrations increases, we normalize the measured pointing errors by

$$r_x(\mathbf{p}_i) = \frac{\varepsilon_x(\mathbf{p}_i) - \mathbb{E}[\tilde{\varepsilon}_x(\mathbf{p}_i)]}{\sigma[\tilde{\varepsilon}_x(\mathbf{p}_i)]} \quad (49)$$

$$r_y(\mathbf{p}_i) = \frac{\varepsilon_y(\mathbf{p}_i) - \mathbb{E}[\tilde{\varepsilon}_y(\mathbf{p}_i)]}{\sigma[\tilde{\varepsilon}_y(\mathbf{p}_i)]} \quad (50)$$

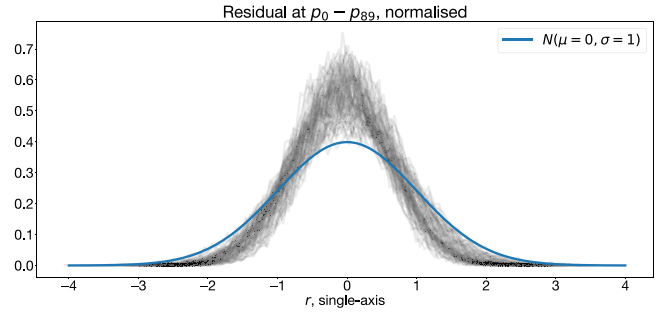
which, for a *perfect* calibration with a good approximation of its uncertainty, should result in both  $r_x$  and  $r_y$  behaving as Gaussian distributions of zero mean and unitary variance.

In all cases, the calibration goal was achieved after taking measurements at the two first calibration points ( $\mathbf{p}_0, \mathbf{p}_1$ ) once. This is not surprising, as the variable part of the pointing error consists of a displacement and a rotation.

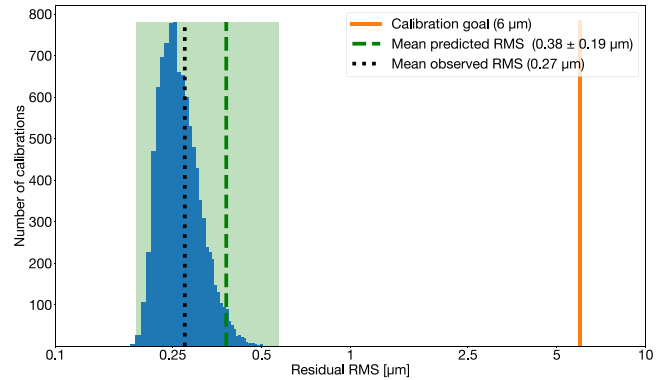


**Figure 5.** Histogram of the normalized residuals with respect to the predicted residuals, one measurement per calibration point. While the pointing error at the two first calibration points  $p_0, p_1$  seems to be perfectly described by the model, we observe extra uncertainty in the rest. We also observe small offsets in the non-visited points of the calibration patterns with respect to their predicted means, all of them smaller than  $1\sigma$ .

The results for the first five calibration points are summarized in Fig. 5. We can see that, in all cases, the observed residuals are consistent with the predicted uncertainties. We can also see that the histogram of observed residuals at  $p_0, p_1$  is well described by the uncertainty predicted by the model. Unsurprisingly, the model overestimates the uncertainty around  $p_{2-4}$ , as information of the distortion in those points is obtained more indirectly from the prior distribution. We observe a similar behaviour by overlaying the normalized residual histogram in both axes and all 90 calibration



**Figure 6.** Histograms of the normalized residual, using prior information.  $10^4$  calibrations, two calibration points, one measurement per calibration point.



**Figure 7.** Histograms of the residual RMS, using prior information.  $10^4$  calibrations, two calibration points, one measurement per calibration point.

points (Fig. 6). In terms of RMS, we observe similar performance as providing 10 points with a very uninformative prior (Fig. 7).

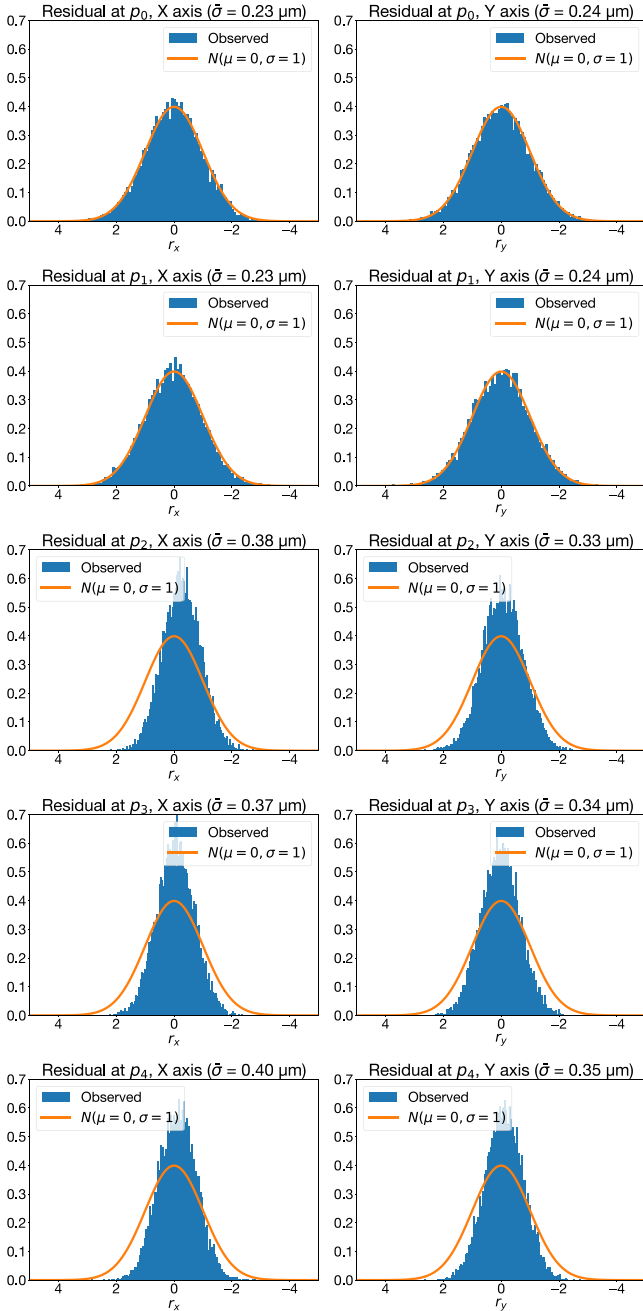
By repeating the test duplicating the number of measurements, the model still needs two different calibration points to achieve the calibration goal. Apart from a lower uncertainty in the sampled points, the impact in the performance is negligible (Figs 9, 8, and 10).

### 7.3 Impact of the number of classical calibrations

In the two previous tests, we made the choice of  $M = 100$  calibrations on somewhat vague assumptions. In order to optimize this choice, we repeated the simulations using prior information from  $M = 25, 50, 100,$  and  $200$  classical calibrations.

Results of this tests are summarized in Fig. 11. In all cases, the calibration goal was met after measuring two calibration points. In the  $M = 25$  case, the tails of the RMS histogram were considerably larger than those of the classical calibration. None the less, the performance of the algorithm can be made comparable with the previous calibrations simply by taking one extra measurement in each point (Fig. 12).

It should be noted that the smallest  $M$  is not necessarily the best choice in all cases. In particular, for  $M = 25$ , initial Bayesian calibrations required two measurements per calibration point to surpass the performance of the classical calibration, but only one measurement per calibration point in future calibrations. This evolution in the number of measurements may not be desirable in contexts in which a repeatable behaviour of the uncertainty is desired.

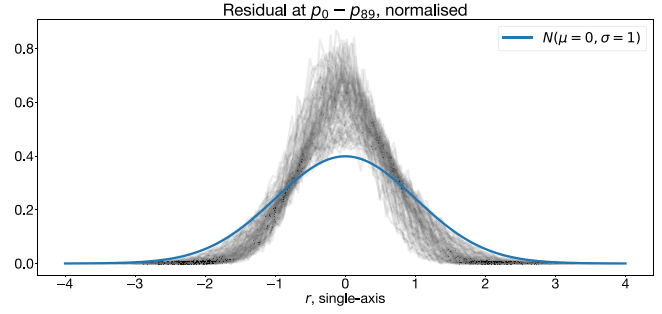


**Figure 8.** Histogram of the normalized residuals with respect to the predicted residuals, two measurements per calibration point.

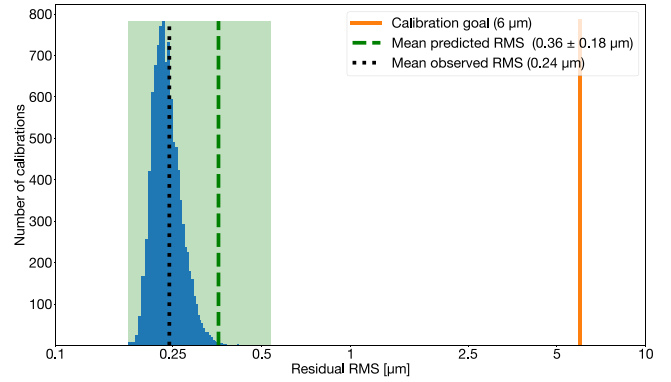
In our case,  $M = 100$  ensures both a reasonable number of previous calibrations (if we assumed one geometrical calibration per day, this number would be achieved in approximately 3 month) and a stable number of required measurements.

#### 7.4 Performance with respect to the classical calibration

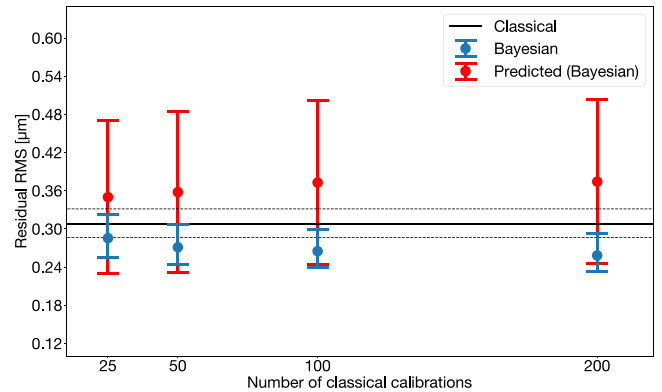
We finish by comparing the performance of the classical calibration with respect to the Bayesian calibration. We do this by performing  $10^4$  simulations of each, and calculating the RMS of the residuals as measured in the 90 locations defined by the redundant OCS pattern. For the Bayesian calibration, we considered two cases: the



**Figure 9.** Histograms of the normalized residual, using prior information.  $10^4$  calibrations, two calibration points, two measurements per calibration point.



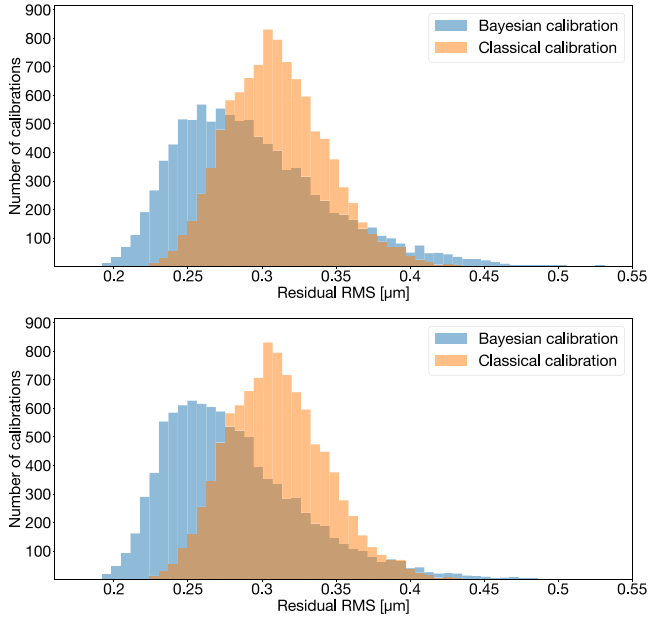
**Figure 10.** Histograms of the residual RMS, using prior information.  $10^4$  calibrations, two calibration points, two measurements per calibration point.



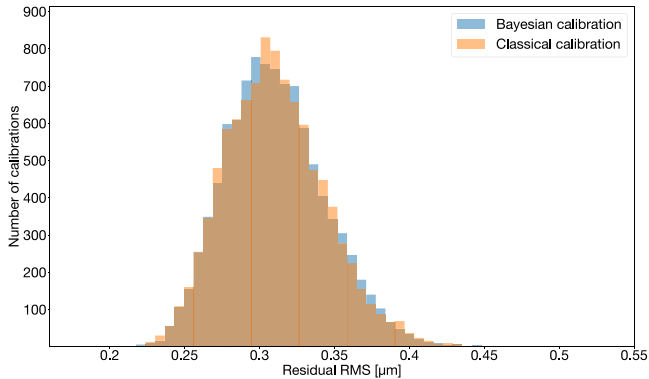
**Figure 11.** Quartiles of the RMS after  $10^4$  simulations, compared with the predicted uncertainty (two measurements needed in total). Horizontal lines represent the quartiles of the RMS of a classical calibration using all 10 points.

informative case (as described in algorithm 1) and the uninformative case (forcing  $\alpha_0 = \mathbf{0}$  and  $\Sigma_0 = (200 \mu\text{m})^2 \mathbb{I}$ ).

In the uninformative case, calibrations were carried out by performing one simulated measurement of the distortion in each of the 10 calibration points of the non-redundant OCS pattern. Then, we calculated the histograms of the residual RMS and overlaid them in Fig. 13. From these results, it is evident that the performance of the uninformative Bayesian calibration is equivalent to that of the classical calibration. This is an expected consequence of the removal of the shrinkage introduced by the prior distributions: in the limiting



**Figure 12.** RMS histogram with  $M = 25$  and  $10^4$  simulations. Bayesian calibrations were performed with one measurement per point (top panel) and two measurements per point (bottom panel). Classical calibrations were performed with all 10 calibration points, one measurement per calibration point.

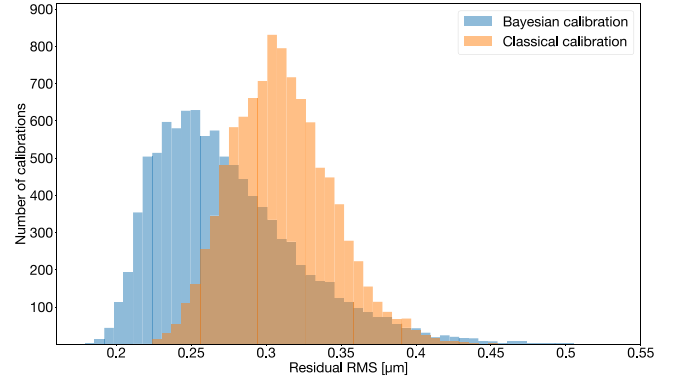


**Figure 13.** Residual RMS histograms for the classical and Bayesian calibration, no prior information,  $10^4$  calibrations. Both Bayesian and classical calibrations were performed with all 10 calibration points, one measurement per calibration point.

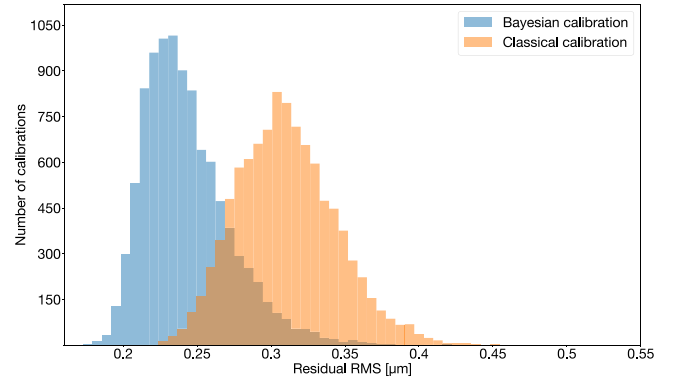
case in which the priors are absolutely flat,  $\alpha_1$  tends to the MLE for  $\alpha$ .

The informative case was performed by estimating  $\bar{\alpha}$  and  $\mathcal{Q}$  from  $M = 100$  previously simulated classical calibrations, and allowing  $\alpha_0$ ,  $\Sigma_0$  to be drawn from the repeatability priors. In this case, the calibration goal was achieved after measuring only two calibration points. The residual RMS histograms are shown in Fig. 14. Even though the original estimate of the RMS remained at  $0.38 \mu\text{m}$ , we observe that the Bayesian calibration now outperforms the classical calibration. This illustrates the ability of the model to capture the repeatability of the highest order distortion and shrink measurements towards the mean.

If we further increase the number of measurements per point to be two, we observe an additional reduction of the RMS towards the measurement noise (Fig. 15). This, once again, illustrates the ability



**Figure 14.** Residual RMS histograms for the classical and Bayesian calibration, using prior information,  $10^4$  calibrations. Bayesian calibrations used two calibration points, one measurement per calibration point. Classical calibrations were performed with all 10 calibration points, one measurement per calibration point.



**Figure 15.** Residual RMS histograms for the classical and Bayesian calibration, using prior information,  $10^4$  calibrations. Bayesian calibrations used two calibration points, two measurements per calibration point. Classical calibrations were performed with all 10 calibration points, one measurement per calibration point.

of the model to acknowledge the increase of measurement precision by repeating the same measurements twice.

## 8. OTHER APPLICATIONS

The freedom of choice of basis functions allows this method to be applied beyond the calibration of geometrical distortions of an instrument like HARMONI. In particular, this method could be applied to the calibration of a pointing model of a ground-based telescope.

For instance, we can identify  $F$  with the set of azimuthal coordinates of calibration stars in the sky and  $F'$  with the coordinates to which the telescope must be commanded to observe these stars in the centre of their science field. The fact that sky coordinates are a particular case of spherical coordinates motivates a multipole expansion of the pointing error as

$$\hat{\varepsilon}(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=+l} Y_l^m(\theta, \phi) \hat{\alpha}_l^m \quad (51)$$

with  $Y_l^m(\theta, \phi)$  being Laplace's spherical harmonics, and also the basis functions of this pointing model. This series can be truncated at  $l = d$ , with  $d$  being the order of the resulting model. For a  $d$ th order

spherical harmonic model, one must keep the first  $(d + 1)^2$  spherical harmonics with  $0 \leq l \leq d$ .

A similar argument used in the derivation of OCS can be made for a calibration pattern with spherical harmonics. In this case, additional constraints exist regarding the availability of stars near the calibration points and the time required to complete the calibration. Dette et al. (2005) discusses a family sampling patterns for spherical harmonic expansions that satisfy different optimality criteria (Kiefer 1974). Unlike OCS, these criteria minimize a property of the inverse of the information matrix of a given sampling pattern, like its determinant, its trace, or its greatest eigenvalue.

The proposed patterns consist of the Cartesian products of two samplings in azimuth and elevation. While the azimuth sampling is uniform, the elevation sampling is calculated from the nodes of quadrature formulas with uniform weights.

## 9. CONCLUSIONS AND FUTURE WORK

(i) We proposed a general corrective model that is able to describe the quasi-static geometric distortion in the focal plane of an instrument.

(ii) The Bayesian formulation of the calibration problem enables incremental calibration, informed updating of the uncertainties in the model parameters, and informative predictions of the distortion.

(iii) Given that proper priors are provided, the Bayesian calibration algorithm is able to outperform the classical calibration algorithm with fewer measurements.

(iv) We applied this model to HARMONI's field distortion under certain simplifying assumptions like highly repeatable GCU mask distortion, no bias in guide probe positioning, and isotropic and homogeneous measurement noise.

(v) The classical formulation of the calibration problem was used to reduce the number of points in the GCU mask while keeping a good immunity to measurement noise. As a result of this trade-off, the redundant OCS pattern was proposed.

(vi) The current formulation of the corrective model can be easily augmented to include anisotropic pointing error measurement, as well as position-dependent measurement errors of the guide probes.

For future work, we plan to extend the instrumental model to incorporate the effect of the systematic positioning errors of the guide probe mechanisms. As these errors are of mechanical origin, we expect them to be one of the biggest contributions to the total instrumental error.

We also intend to couple the corrective model to distortions of the GCU mask itself, due to thermal expansion of the supporting structure. Additionally, while the current estimates of the RMS can be used as *sufficiently good* termination conditions for calibration, they still overestimate the distribution of observed RMS. The cause of this overestimation can be tied to the crude approximation of the RMS distribution (approximation of the MSE distribution as a Gaussian plus a first order expansion of the square root) and to the choice of the calibration coefficients (we are not sampling  $\alpha$  but picking  $\alpha_{n+1}$ , which is known with less uncertainty). In this regard, we plan to develop the estimation of the RMS further to produce more realistic predictions of the observed RMS.

## ACKNOWLEDGEMENTS

GCC and JPL acknowledge support from grants PID2019-105423GA-I00 and PID2022-140483NB-C22 funded by AEI

(10.13039/501100011033) and BDC 20221289 funded by MCI through the Recovery, Transformation and Resilience Plan from the Spanish State, and by NextGenerationEU from the European Union through the Recovery and Resilience Facility. MPS acknowledges funding support from the Ramón y Cajal program of the Spanish Ministerio de Ciencia e Innovación (RYC2021-033094-I).

## DATA AVAILABILITY

Source code of the simulation scripts, data products, and Jupyter Lab notebooks are freely accessible from the GitHub repository at <https://github.com/BatchDrake/bayescal>, under the terms of the 3-clause BSD license.

## REFERENCES

- Augustin R. et al., 2019, *MNRAS*, 489, 2417
- Biesheuvel R. S., Janssen A. J. E. M., Pozzi P., Pereira S. F., 2018, *OSA Continuum*, 1, 581
- Bounissou S., Thatte N., Zieleniewski S., Houghton R. C. W., Tecza M., Hook I., Neichel B., Fusco T., 2018, *MNRAS*, 478, 3189
- Carracedo-Carballeda G. J., 2022, harmoni-pm GitHub repository. GitHub. Available at: <https://github.com/BatchDrake/harmoni-pm>
- Carracedo-Carballeda G. J., 2023, bayescal GitHub repository. GitHub. Available at: <https://github.com/BatchDrake/bayescal>
- Carracedo-Carballeda G. J., Piqueras-López J., Clarke F., 2022, in Ibsen J., Chiozzi G., eds, Proc. SPIE Conf. Ser. Vol. 12189, Software and Cyberinfrastructure for Astronomy VII. SPIE, Bellingham, p. 121892G
- Dette H., Melas V. B., Pepelyshev A., 2005, *Ann. Stat.*, 33, 2758
- Dohlen K. et al., 2022, Proc. SPIE Conf. Ser. Vol. 12187, Modeling, Systems Engineering, and Project Management for Astronomy X. SPIE, Bellingham, p. 121871D
- Estrada A. et al., 2022, in Navarro R., Geyl R., eds, Proc. SPIE Conf. Ser. Vol. 12188, Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation V. SPIE, Bellingham, p. 121883O
- García-Lorenzo, B., Monreal-Ibero, A., Pereira-Santaella, M., Thatte N., Ramos Almeida C., Galbany L., Mediavilla E., 2022, *A&A*, 659, A79
- Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B., 2021, Bayesian Data Analysis, 3rd edn. CRC Press, Boca Raton, FL. <http://www.stat.columbia.edu/gelman/book/BDA3.pdf>
- Grisdale K., Thatte N., Devriendt J., Pereira-Santaella M., Slyz A., Kimm T., Dubois Y., Yi S. K., 2021, *MNRAS*, 501, 5517
- Grisdale K. et al., 2022, *MNRAS*, 513, 3906
- Kendrew S. et al., 2016, *MNRAS*, 458, 2405
- Kiefer J., 1974, *Ann. Stat.*, 2, 849
- Nguyen D. D., Cappellari M., Pereira-Santaella M., 2023, *MNRAS*, 526, 3548
- Pereira-Santaella M., Routledge L., Zieleniewski S., Kendrew S., 2023, HSI3 GitHub repository. GitHub. Available at: <https://github.com/HARMONI-ELT/HSIM>
- Ramos-López D., Sánchez-Granero M., Fernández-Martínez M., Martínez-Finkelshtein A., 2016, *Appl. Math. Comput.*, 274, 247
- Richardson M. L. A., Routledge L., Thatte N., Tecza M., Houghton R. C. W., Pereira-Santaella M., Rigopoulou D., 2020, *MNRAS*, 498, 1891
- Rubin D. B., 1984, *Ann. Stat.*, 12, 1151
- Thatte N. et al., 2021, *The Messenger*, 182, 7
- Thatte N. et al., 2022a, in Proc. SPIE Conf. Ser. Vol. 12184, Ground-based and Airborne Instrumentation for Astronomy IX. SPIE, Bellingham, p. 1218420
- Thibos L. N., Applegate R. A., Schwiegerling J. T., Webb R., 2002, *J. Refract. Surg.*, 18, S652
- Zieleniewski S., Thatte N., Kendrew S., Houghton R., Swinbank A., Tecza M., Clarke F., Fusco T., 2015, *MNRAS*, 453, 3755

## APPENDIX A: REAL-VALUED COLLOCATION MATRIX

Let  $\hat{\mathbf{p}} = (\hat{p}^0, \hat{p}^1, \dots, \hat{p}^{K-1})^\top \in \mathbb{C}^K$  be a complex vector of evaluation points in the relayed focal plane,  $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}^0, \hat{\alpha}^1, \dots, \hat{\alpha}^{J-1})^\top \in \mathbb{C}^J$  the complex vector of coefficients of the corrective model, and  $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}^0, \hat{\varepsilon}^1, \dots, \hat{\varepsilon}^{K-1})^\top \in \mathbb{C}^K$  the vector of pointing errors evaluated in  $\hat{\mathbf{p}}$  in complex form. Predictions of the pointing error can be made by evaluating

$$\hat{\boldsymbol{\varepsilon}} = \hat{\mathbf{G}}\hat{\boldsymbol{\alpha}} \quad (\text{A1})$$

in which the coefficients  $\hat{G}_j^k$  of the complex collocation matrix  $\hat{\mathbf{G}}$  are obtained from evaluations of the first  $J$  basis functions  $\hat{G}_j = \hat{G}_j(\hat{p}^k)$ .

In order to obtain a real-valued expression for the pointing error in both axes, we need to decompose all complex quantities into their real and imaginary parts:

$$\hat{\varepsilon}^k = \varepsilon_x^k + i\varepsilon_y^k \quad (\text{A2})$$

$$\hat{\alpha}^j = \alpha_x^j + i\alpha_y^j \quad (\text{A3})$$

$$\hat{G}_j^k = X_j^k + iY_j^k. \quad (\text{A4})$$

Now, equation (A1) can be expanded to

$$\varepsilon_x^k + i\varepsilon_y^k = \sum_{j=0}^{J-1} X_j^k \alpha_x^j + iX_j^k \alpha_y^j + iY_j^k \alpha_x^j - Y_j^k \alpha_y^j. \quad (\text{A5})$$

By grouping real and imaginary quantities, we arrive to the following set of real-valued equations:

$$\varepsilon_x^k = \sum_{j=0}^{J-1} X_j^k \alpha_x^j - Y_j^k \alpha_y^j \quad (\text{A6})$$

$$\varepsilon_y^k = \sum_{j=0}^{J-1} X_j^k \alpha_y^j + Y_j^k \alpha_x^j. \quad (\text{A7})$$

The quantities  $\varepsilon_d^k$ ,  $d \in \{x, y\}$  are still enumerated by two indices. Since single-indexed vectors are more convenient from the implementation perspective, we can rely on the following interlaced representation of complex vectors:

$$v_x^k = v^{2k} \quad (\text{A8})$$

$$v_y^k = v^{2k+1} \quad (\text{A9})$$

which assigns even indices to the real parts and odd indices to the imaginary parts. Under this representation, equation (A6) can be written as

$$\varepsilon^{2k} = \sum_{j=0}^{J-1} X_j^k \alpha^{2j} - Y_j^k \alpha^{2j+1} \quad (\text{A10})$$

$$\varepsilon^{2k+1} = \sum_{j=0}^{J-1} X_j^k \alpha^{2j+1} + Y_j^k \alpha^{2j} \quad (\text{A11})$$

or, in a more compact form

$$\boldsymbol{\varepsilon} = \mathbf{G}\boldsymbol{\alpha} \quad (\text{A12})$$

where  $\boldsymbol{\varepsilon} \in \mathbb{R}^{2K}$ ,  $\boldsymbol{\alpha} \in \mathbb{R}^{2J}$ , and  $\mathbf{G} \in \mathbb{R}^{2K \times 2J}$ . The coefficients of the real collocation matrix  $G_p^q$  can be expressed in terms of the real and imaginary parts of the coefficients of the complex collocation matrix  $\hat{G}_j^k$  as

$$G_{2j}^{2k} = +X_j^k = +\Re[\hat{G}_j^k] \quad (\text{A13})$$

$$G_{2j+1}^{2k} = -Y_j^k = -\Im[\hat{G}_j^k] \quad (\text{A14})$$

$$G_{2j}^{2k+1} = +Y_j^k = +\Im[\hat{G}_j^k] \quad (\text{A15})$$

$$G_{2j+1}^{2k+1} = +X_j^k = +\Re[\hat{G}_j^k]. \quad (\text{A16})$$

## APPENDIX B: POSTERIOR DISTRIBUTIONS

This paper introduces Bayesian updating formulae for two distributions: the calibration-time coefficient distribution and the model repeatability distributions. These distributions have been chosen to be conjugate with respect to their corresponding likelihood functions, so that the posterior distribution is analytic and of the same form as the prior distribution. This section describes how the Bayesian updating formulae have been obtained in each case.

### B1 Model coefficient distribution

Let  $\boldsymbol{\alpha} = (\alpha^0, \alpha^1, \dots, \alpha^{2J-1})^\top \in \mathbb{R}^{2J}$  the real and imaginary parts of the model coefficients and  $\sigma_E^2$  the variance of the measurement noise,  $\boldsymbol{\varepsilon} = (\varepsilon^0, \varepsilon^1, \dots, \varepsilon^{2K-1})^\top \in \mathbb{R}^{2K}$  the real vector constructed from measurements of the pointing error in the horizontal and vertical axes of certain  $K$  measurement points, and  $\mathbf{p} = (p^0, p^1, \dots, p^{2K-1})^\top$  the real vector constructed from alternating horizontal and vertical coordinates of the  $K$  measurement points.

Let  $p(\boldsymbol{\alpha})$  be the prior probability density of the coefficient vector, which encodes our knowledge on the model coefficients. According to the Bayes theorem, the presence of a new observation  $\boldsymbol{\varepsilon}$  updates the probability density  $p(\boldsymbol{\alpha})$  as

$$p(\boldsymbol{\alpha}|\boldsymbol{\varepsilon}) = \frac{p(\boldsymbol{\varepsilon}|\boldsymbol{\alpha}, \mathbf{p}, \sigma_E^2)}{p(\boldsymbol{\varepsilon})} p(\boldsymbol{\alpha}) \quad (\text{B1})$$

where  $p(\boldsymbol{\varepsilon}|\boldsymbol{\alpha}, \mathbf{p}, \sigma_E^2)$  is just the likelihood function introduced in equation (15) and  $p(\boldsymbol{\varepsilon})$  acts as a mere normalization constant. Equation (B1) can therefore be written as

$$p(\boldsymbol{\alpha}|\boldsymbol{\varepsilon}) \propto \mathcal{L}(\boldsymbol{\alpha}|\boldsymbol{\varepsilon}, \mathbf{p}, \sigma_E^2) p(\boldsymbol{\alpha}). \quad (\text{B2})$$

If  $p(\boldsymbol{\alpha})$  is modelled after a multivariate Gaussian with a mean vector  $\boldsymbol{\alpha}_n$  and a covariance matrix  $\boldsymbol{\Sigma}_n$

$$\boldsymbol{\alpha} \sim \mathcal{MN}(\boldsymbol{\alpha}_n, \boldsymbol{\Sigma}_n) \quad (\text{B3})$$

the posterior  $p(\boldsymbol{\alpha}|\boldsymbol{\varepsilon})$  will also be multivariate Gaussian, as it is conjugate with respect to the likelihood function (15):

$$\boldsymbol{\alpha}|\boldsymbol{\varepsilon}_i \sim \mathcal{MN}(\boldsymbol{\alpha}_{n+1}, \boldsymbol{\Sigma}_{n+1}) \quad (\text{B4})$$

where the sub-indices  $n$  and  $n+1$  have been conveniently chosen to distinguish between prior hyperparameters and posterior hyperparameters.  $p(\boldsymbol{\alpha})$  can be written as

$$p(\boldsymbol{\alpha}) = \frac{1}{\sqrt{(2\pi)^{2J} \det \boldsymbol{\Sigma}_n}} \exp \left[ -\frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_n)^\top \boldsymbol{\Sigma}_n^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_n) \right]. \quad (\text{B5})$$

Since all distributions in equation (B2) are of the exponential family, we can take the logarithms in both sides of the equation and remove the common factors:

$$\begin{aligned} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{n+1})^\top \boldsymbol{\Sigma}_{n+1}^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{n+1}) &= \sigma_E^{-2} (\boldsymbol{\varepsilon} - \mathbf{G}\boldsymbol{\alpha})^\top (\boldsymbol{\varepsilon} - \mathbf{G}\boldsymbol{\alpha}) \\ &\quad + (\boldsymbol{\alpha} - \boldsymbol{\alpha}_n)^\top \boldsymbol{\Sigma}_n^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_n) + S \end{aligned} \quad (\text{B6})$$

where  $S$  is a normalization term that does not depend on  $\boldsymbol{\alpha}$ . Both sides of the identity (B6) are therefore second-degree polynomials on  $\boldsymbol{\alpha}$ .

We introduce now the precision matrix  $\boldsymbol{\Lambda}_n = \boldsymbol{\Sigma}_n^{-1}$ . By definition, all covariance matrices  $\boldsymbol{\Sigma}_n$  are symmetric and positive definite, hence the corresponding precision matrices  $\boldsymbol{\Lambda}_n$  exist and must also be symmetric and positive definite. This implies that  $\mathbf{u}^\top \boldsymbol{\Lambda}_n \mathbf{v} = \mathbf{v}^\top \boldsymbol{\Lambda}_n \mathbf{u}$  for any  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{2J}$ . By exploiting this property, we can further expand (B6) into

$$\begin{aligned} & \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_{n+1} \boldsymbol{\alpha} + \boldsymbol{\alpha}_{n+1}^\top \boldsymbol{\Lambda}_{n+1} \boldsymbol{\alpha}_{n+1} - 2\boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_{n+1} \boldsymbol{\alpha}_{n+1} - S = \\ & \boldsymbol{\varepsilon}^\top \sigma_E^{-2} \boldsymbol{\varepsilon} + \boldsymbol{\alpha}^\top \mathbf{G}^\top \sigma_E^{-2} \mathbf{G} \boldsymbol{\alpha} - 2\boldsymbol{\alpha}^\top \mathbf{G}^\top \sigma_E^{-2} \boldsymbol{\varepsilon} \\ & + \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_n \boldsymbol{\alpha} + \boldsymbol{\alpha}_n^\top \boldsymbol{\Lambda}_n \boldsymbol{\alpha}_n - 2\boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_n \boldsymbol{\alpha}_n. \end{aligned} \quad (\text{B7})$$

By grouping the quadratic terms on  $\boldsymbol{\alpha}$ , we see that  $\boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_{n+1} \boldsymbol{\alpha} = \boldsymbol{\alpha}^\top \mathbf{G}^\top \sigma_E^{-2} \mathbf{G} \boldsymbol{\alpha} + \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_n \boldsymbol{\alpha} = \boldsymbol{\alpha}^\top (\mathbf{G}^\top \sigma_E^{-2} \mathbf{G} + \boldsymbol{\Lambda}_n) \boldsymbol{\alpha}$ . This allows us to conclude that

$$\boldsymbol{\Lambda}_{n+1} = \mathbf{G}^\top \sigma_E^{-2} \mathbf{G} + \boldsymbol{\Lambda}_n \quad (\text{B8})$$

which can be interpreted as the Bayesian updating formula for the inverse of the covariance matrix  $\boldsymbol{\Sigma}_n$ .

By repeating the process on the linear terms in  $\boldsymbol{\alpha}$  and removing common factors, we see that  $\boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_{n+1} \boldsymbol{\alpha}_{n+1} = \boldsymbol{\alpha}^\top \mathbf{G}^\top \sigma_E^{-2} \boldsymbol{\varepsilon} + \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_n \boldsymbol{\alpha}_n$ . Since  $\boldsymbol{\alpha}^\top$  is a common factor in this identity, we can rewrite it as

$$\boldsymbol{\alpha}^\top (\boldsymbol{\Lambda}_{n+1} \boldsymbol{\alpha}_{n+1} - \mathbf{G}^\top \sigma_E^{-2} \boldsymbol{\varepsilon} - \boldsymbol{\Lambda}_n \boldsymbol{\alpha}_n) = 0 \quad (\text{B9})$$

which must hold for any  $\boldsymbol{\alpha}$  and therefore  $\boldsymbol{\Lambda}_{n+1} \boldsymbol{\alpha}_{n+1} = \mathbf{G}^\top \sigma_E^{-2} \boldsymbol{\varepsilon} + \boldsymbol{\Lambda}_n \boldsymbol{\alpha}_n$ . By solving now for  $\boldsymbol{\alpha}_{n+1}$

$$\boldsymbol{\alpha}_{n+1} = \boldsymbol{\Lambda}_{n+1}^{-1} (\mathbf{G}^\top \sigma_E^{-2} \boldsymbol{\varepsilon} + \boldsymbol{\Lambda}_n \boldsymbol{\alpha}_n) \quad (\text{B10})$$

which can be interpreted as the Bayesian updating formula for the mean vector  $\boldsymbol{\alpha}_n$ . The full Bayesian updating formula for the model coefficient distribution is therefore written as

$$\boldsymbol{\Sigma}_{n+1} = (\mathbf{G}^\top \sigma_E^{-2} \mathbf{G} + \boldsymbol{\Sigma}_n^{-1})^{-1} \quad (\text{B11})$$

$$\boldsymbol{\alpha}_{n+1} = \boldsymbol{\Sigma}_{n+1} (\mathbf{G}^\top \sigma_E^{-2} \boldsymbol{\varepsilon} + \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\alpha}_n). \quad (\text{B12})$$

## B2 Connection to the classical calibration

When  $K = J$ , a classical solution to the calibration problem can be obtained by the formula

$$\boldsymbol{\alpha} = \mathbf{G}^{-1} \boldsymbol{\varepsilon} \quad (\text{B13})$$

with  $\boldsymbol{\alpha}, \boldsymbol{\varepsilon} \in \mathbb{R}^{2J}$  and  $\mathbf{G} \in \mathbb{R}^{2J \times 2J}$ . This solution can be connected to a limiting case of the Bayesian update when a batch of  $J$  measurements has been used as evidence and the prior  $p(\boldsymbol{\alpha})$  tends to an infinitely uninformative distribution. From equations (B11)–(B12), we obtain

$$\boldsymbol{\Sigma}_1 = (\mathbf{G}^\top \sigma_E^{-2} \mathbf{G} + \boldsymbol{\Sigma}_0^{-1})^{-1} \quad (\text{B14})$$

$$\boldsymbol{\alpha}_1 = \boldsymbol{\Sigma}_1 (\mathbf{G}^\top \sigma_E^{-2} \boldsymbol{\varepsilon} + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\alpha}_0) \quad (\text{B15})$$

where  $\mathbf{G} \in \mathbb{R}^{2J \times 2J}$  is now the classical real collocation matrix.

If we introduce a prior for the covariance matrix of the form  $\boldsymbol{\Sigma}_0 = \sigma_0^2 \mathbb{I}$  (i.e. same variance  $\sigma_0^2$  for all coefficients and no correlation between them), we can analyse the limiting case of an infinitely uninformative prior distribution by taking the limits of  $\boldsymbol{\Sigma}_1$  and  $\boldsymbol{\alpha}_1$  when  $\sigma_0^2 \rightarrow \infty$ :

$$\lim_{\sigma_0^2 \rightarrow \infty} \boldsymbol{\Sigma}_1 = (\mathbf{G}^\top \sigma_E^{-2} \mathbf{G})^{-1} = \sigma_E^2 \mathbf{G}^{-1} \mathbf{G}^{\top -1} \quad (\text{B16})$$

$$\lim_{\sigma_0^2 \rightarrow \infty} \boldsymbol{\alpha}_1 = \boldsymbol{\Sigma}_1 \mathbf{G}^\top \sigma_E^{-2} \boldsymbol{\varepsilon} = \mathbf{G}^{-1} \boldsymbol{\varepsilon} \quad (\text{B17})$$

from which we see that  $\boldsymbol{\alpha}_1$  equals to the classical solution as written in equation (B13).

We should remark that, as per equation (B11), sequential and batched updates have the same effect on the covariance matrix. This is no longer true for the mean vector  $\boldsymbol{\alpha}_n$ , in which updates from previous observations condition the probability of future observations.

On the other hand, since equation (B16) still applies to sequential updates, we should expect a significant decrease of the total variance of  $\boldsymbol{\Sigma}_J$  with respect to  $\boldsymbol{\Sigma}_{J-1}$  (as it now contains information of all distortion modes).

## B3 Posterior predictive distribution

The choice of conjugate priors also ensures the existence of a closed-form expression for the posterior predictive distribution. In order to find this expression, we consider a collocation matrix evaluated in  $N$  distinct test points  $\tilde{\mathbf{G}} \in \mathbb{R}^{2N \times 2J}$ , with  $N < J$ . We start by remarking that  $p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon})$  can be obtained from the following marginalization:

$$p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon}) = \int \cdots \int p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\alpha}, \sigma_E^2 \mathbb{I}) p(\boldsymbol{\alpha}|\boldsymbol{\varepsilon}) d^{2J} \boldsymbol{\alpha}. \quad (\text{B18})$$

Since  $p(\boldsymbol{\alpha}|\boldsymbol{\varepsilon}) = p(\boldsymbol{\alpha}|\boldsymbol{\alpha}_{n+1}, \boldsymbol{\Sigma}_{n+1})$ , it can also be written as

$$p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon}) = \int \cdots \int p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\alpha}, \sigma_E^2 \mathbb{I}) p(\boldsymbol{\alpha}|\boldsymbol{\alpha}_{n+1}, \boldsymbol{\Sigma}_{n+1}) d^{2J} \boldsymbol{\alpha}. \quad (\text{B19})$$

Since  $p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\alpha}, \sigma_E^2 \mathbb{I})$  depends on  $\boldsymbol{\alpha}$  only to calculate the mean  $\tilde{\mathbf{G}}\boldsymbol{\alpha}$ , we can tentatively introduce the change of variable  $\mathbf{m} = \tilde{\mathbf{G}}\boldsymbol{\alpha}$ . Unfortunately,  $\tilde{\mathbf{G}}$  is a rectangular projection matrix, and any change of variable requires the transformation to be bijective.

We can overcome this difficulty by remarking that  $\tilde{\mathbf{G}}$  consists on  $2N$  linearly independent vectors  $\mathbf{G}_i \in \mathbb{R}^{2J}$ . Let  $G = \text{span}\{\mathbf{G}_0, \dots, \mathbf{G}_{2N-1}\}$  the vector sub-space spanned by the rows of  $\tilde{\mathbf{G}}$ . It is clear that  $G \subset \mathbb{R}^{2J}$  as  $\dim G = 2N < 2J$ .

Let  $W \subset \mathbb{R}^{2J}$  such that  $W \perp G$  and  $W \cup G = \mathbb{R}^{2J}$ . If  $\mathbf{w}_i \in W$  are a set of  $2(J - N)$  basis vectors of  $W$ , we can complete the matrix  $\tilde{\mathbf{G}}$  as follows:

$$\mathbf{A} = \begin{bmatrix} \mathbf{G}_0^\top \\ \vdots \\ \mathbf{G}_{2N-1}^\top \\ \mathbf{w}_0^\top \\ \vdots \\ \mathbf{w}_{2(J-N)-1}^\top \end{bmatrix}. \quad (\text{B20})$$

Since by construction the matrix  $\mathbf{A} \in \mathbb{R}^{2J \times 2J}$  is full rank and its rows span all  $\mathbb{R}^{2J}$ ,  $\mathbf{A}^{-1}$  is guaranteed to exist. We can now introduce the following change of variable:

$$\mathbf{m} = \mathbf{A}\boldsymbol{\alpha} \iff \boldsymbol{\alpha} = \mathbf{A}^{-1}\mathbf{m} \quad (\text{B21})$$

with  $\mathbf{m} = (m_0, \dots, m_{2J-1})^\top \in \mathbb{R}^{2J}$ . This enables the rewrite of equation (B19) as

$$p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon}) \propto \int \cdots \int p(\tilde{\boldsymbol{\varepsilon}}|\mathbf{m}, \sigma_E^2 \mathbb{I}) p(\mathbf{A}^{-1}\mathbf{m}|\boldsymbol{\alpha}_{n+1}, \boldsymbol{\Sigma}_{n+1}) d^{2J} \mathbf{m}. \quad (\text{B22})$$

We further note that  $p(\tilde{\boldsymbol{\varepsilon}}|\mathbf{m}, \sigma_E^2 \mathbb{I})$  does not depend on the extra dimensions we added to  $\tilde{\mathbf{G}}$ . This allows us to marginalize only in

the first  $2N$  components of  $\mathbf{m}$ :

$$p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon}) = p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon}, m_{2N}, \dots, m_{2J-1}) \propto \int \dots \int p(\tilde{\boldsymbol{\varepsilon}}|\mathbf{m}, \sigma_{\mathbb{E}}^2) p(\mathbf{A}^{-1}\mathbf{m}|\boldsymbol{\alpha}_{n+1}, \boldsymbol{\Sigma}_{n+1}) d^{2N}m \quad (\text{B23})$$

where we kept the dependency  $m_{2N}, \dots, m_{2J-1}$  as it is still necessary to compute  $\mathbf{A}^{-1}\mathbf{m}$ . Note that  $p(\mathbf{A}^{-1}\mathbf{m}|\boldsymbol{\alpha}_{n+1}, \boldsymbol{\Sigma}_{n+1})$  is also a multivariate Gaussian probability distribution. The logarithmic probability density takes the form

$$\ln p(\mathbf{A}^{-1}\mathbf{m}|\boldsymbol{\alpha}_{n+1}, \boldsymbol{\Sigma}_{n+1}) = -\frac{1}{2}(\mathbf{A}^{-1}\mathbf{m} - \boldsymbol{\alpha}_{n+1})^\top \boldsymbol{\Sigma}_{n+1}^{-1} (\mathbf{A}^{-1}\mathbf{m} - \boldsymbol{\alpha}_{n+1}) + K. \quad (\text{B24})$$

By writing  $\boldsymbol{\Sigma}_{n+1}^{-1} = (\mathbf{A}^{-1}\mathbf{A})^\top \boldsymbol{\Sigma}_{n+1}^{-1} \mathbf{A}^{-1}\mathbf{A}$ , equation (B24) becomes

$$\ln p(\mathbf{A}^{-1}\mathbf{m}|\boldsymbol{\alpha}_{n+1}, \boldsymbol{\Sigma}_{n+1}) = -\frac{1}{2}(\mathbf{m} - \mathbf{A}\boldsymbol{\alpha}_{n+1})^\top \mathbf{A}^{-1\top} \boldsymbol{\Sigma}_{n+1}^{-1} \mathbf{A}^{-1} (\mathbf{m} - \mathbf{A}\boldsymbol{\alpha}_{n+1}) + K \quad (\text{B25})$$

from which we identify the exponent of a multivariate Gaussian with mean  $\mathbf{A}\boldsymbol{\alpha}_{n+1}$  and covariance matrix  $\mathbf{A}^\top \boldsymbol{\Sigma}_{n+1} \mathbf{A}$ :

$$p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon}, m_{2N}, \dots, m_{2J-1}) \propto \int \dots \int p(\tilde{\boldsymbol{\varepsilon}}|\mathbf{m}, \sigma_{\mathbb{E}}^2) p(\mathbf{m}|\mathbf{A}\boldsymbol{\alpha}_{n+1}, \mathbf{A}^\top \boldsymbol{\Sigma}_{n+1} \mathbf{A}) d^{2N}m. \quad (\text{B26})$$

The multivariate Gaussian density  $p(\tilde{\boldsymbol{\varepsilon}}|\mathbf{m}, \sigma_{\mathbb{E}}^2)$  can also be written as  $p(\tilde{\boldsymbol{\varepsilon}} - \mathbf{m}'|\mathbf{0}, \sigma_{\mathbb{E}}^2)$  with  $\mathbf{m}' = (m_0, \dots, m_{2N-1})^\top$ , as it does not depend on the extra dimensions we added to construct  $\mathbf{A}$ . We are now ready to remove the dummy dependency on  $m_{2N}, \dots, m_{2J-1}$  along with the corresponding rows and columns in the mean expression  $\mathbf{A}\boldsymbol{\alpha}_{n+1}$  (which now reduces to  $\tilde{\mathbf{G}}\boldsymbol{\alpha}$ ) and the covariance matrix  $\mathbf{A}^\top \boldsymbol{\Sigma}_{n+1} \mathbf{A}$  (which now reduces to  $\tilde{\mathbf{G}}^\top \boldsymbol{\Sigma}_{n+1} \tilde{\mathbf{G}}$ ):

$$p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon}) \propto \int \dots \int p(\tilde{\boldsymbol{\varepsilon}} - \mathbf{m}'|\mathbf{0}, \sigma_{\mathbb{E}}^2) p(\mathbf{m}'|\mathbf{G}\boldsymbol{\alpha}_{n+1}, \tilde{\mathbf{G}}^\top \boldsymbol{\Sigma}_{n+1} \tilde{\mathbf{G}}) d^{2N}m. \quad (\text{B27})$$

We finish by remarking that equation (B27) is actually the convolution integral of two multivariate Gaussians distributions:

$$p(\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon}) = p(\mathbf{x}|\mathbf{0}, \sigma_{\mathbb{E}}^2) * p(\mathbf{x}|\tilde{\mathbf{G}}\boldsymbol{\alpha}_{n+1}, \tilde{\mathbf{G}}^\top \boldsymbol{\Sigma}_{n+1} \tilde{\mathbf{G}}) \quad (\text{B28})$$

from which it follows that  $\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon}$  is also a multivariate Gaussian variable of the form

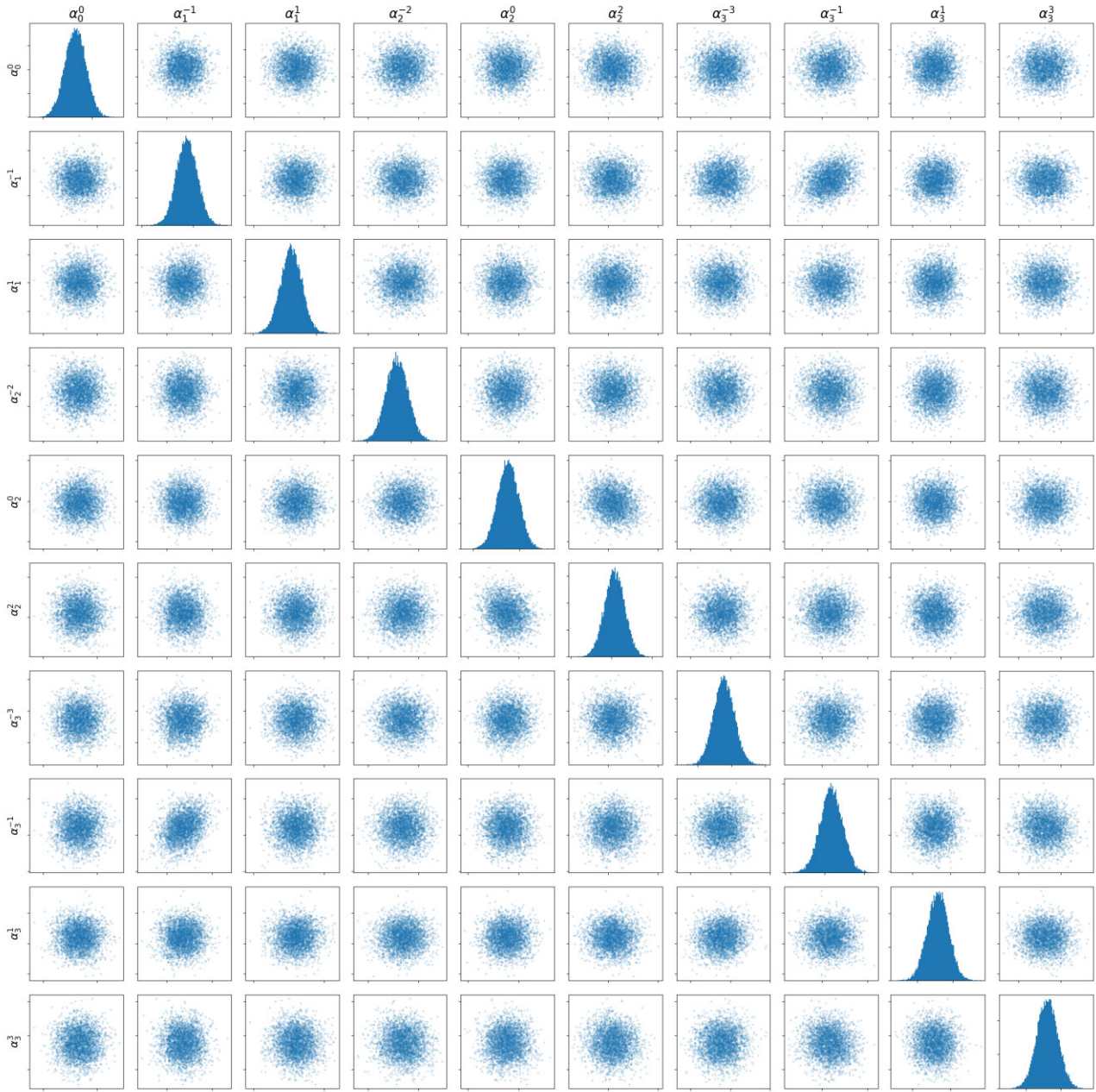
$$\tilde{\boldsymbol{\varepsilon}}|\boldsymbol{\varepsilon} \sim \mathcal{MN}(\tilde{\mathbf{G}}\boldsymbol{\alpha}_{n+1}, \sigma_{\mathbb{E}}^2 \mathbb{I} + \tilde{\mathbf{G}}^\top \boldsymbol{\Sigma}_{n+1} \tilde{\mathbf{G}}). \quad (\text{B29})$$

## APPENDIX C: SCATTER PLOT MATRICES

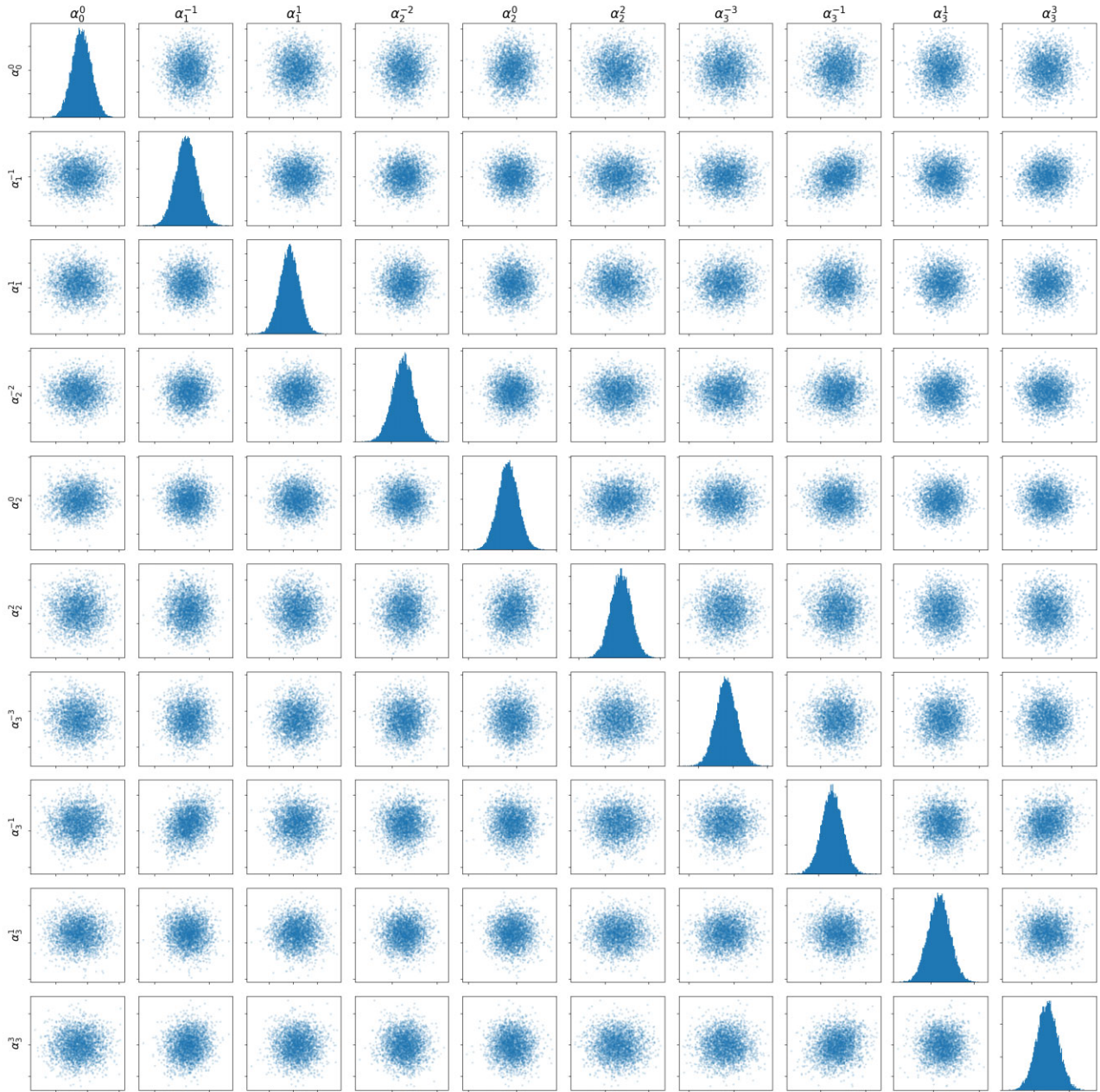
In order to understand the stochastic behaviour of the corrective model coefficients, we simulated a total of 20 000 classical calibrations. Simulations were performed by means of HARMONI's pointing simulator `harmoni-pm`, and assumed a corrective model with  $J = 10$  complex coefficients. Each calibration was fed from 553 simulated measurements of the pointing error along a uniform grid of points inside the surface of the technical field. Each simulated measurement took into account contributions like manufacturing tolerances, optical distortions, and mechanical instabilities, and produced a  $\mathbb{C}^{10}$  vector of calibrated coefficients.

In order to represent distribution and correlation between coefficients, a representation in the form of two scatter plot matrices was used, one for the real part of the coefficients (Fig. C1) and another for the imaginary part of the coefficients (Fig. C2). The plot in the  $i$ th row and  $j$ th column (with  $i \neq j$ ) is a 2D scatter plot of the pair formed by the real (imaginary) parts of the coefficients  $\hat{\alpha}^i$  and  $\hat{\alpha}^j$ . On the other hand, the  $i$ th plot in the diagonal represents the histogram of the real (imaginary) part of the coefficient  $\hat{\alpha}^i$ . Since the purpose of these plots was to represent the shape of the underlying coefficient distribution, the simulated real (imaginary) parts of the coefficients were normalized by their respective sample means and standard deviations.

The obtained scatter plot matrices are compatible with a multivariate Gaussian random variable, with slight correlations between certain coefficients.



**Figure C1.** Scatter plot matrix of the real part of the complex coefficient vector  $\hat{\alpha}$ . In this figure,  $\alpha_n^m$  represents the real part of the complex coefficient multiplying the polynomial  $\hat{G}_n^m$ . Rows and columns are sorted by increasing OSA/ASI index, as per equation (46). Plot axes are normalized by the sample mean and standard deviation of the real parts of the respective coefficients.



**Figure C2.** Scatter plot matrix of the imaginary part of the complex coefficient vector  $\hat{\alpha}$ . In this figure,  $\alpha_n^m$  represents the imaginary part of the complex coefficient multiplying the polynomial  $\hat{G}_n^m$ . Rows and columns are sorted by increasing OSA/ASI index, as per equation (46). Plot axes are normalized by the sample mean and standard deviation of the imaginary parts of the respective coefficients.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.