Linearity Characterization and Uncertainty Quantification of Spectroradiometers via Maximum Likelihood and the Non-parametric Bootstrap

Adam L. Pintar[1], **Zachary H. Levine**[2], **Howard W. Yoon**[3], **Stephen E. Maxwell**[3]

[1] Statistical Engineering Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8980 USA; apintar@nist.gov
[2] Quantum Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8441 USA; zlevine@nist.gov
[3] Sensor Science Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8441 USA; smaxwell@nist.gov

Abstract. A technique for characterizing and correcting the linearity of radiometric instruments is known by the names the "flux-addition method" and the "combinatorial technique". In this paper, we develop a rigorous uncertainty quantification method for use with this technique and illustrate its use with both synthetic data and experimental data from a "beam conjoiner" instrument. We present a probabilistic model that relates the instrument readout to a set of unknown fluxes via a set of polynomial coefficients. Maximum likelihood estimates (MLEs) of the unknown fluxes and polynomial coefficients are recommended, while a non-parametric bootstrap algorithm enables uncertainty quantification including standard errors and confidence intervals.

The synthetic data represent plausible outputs of a radiometric instrument and enable testing and validation of the method. The MLEs for these data are found to be approximately unbiased, and confidence intervals derived from the bootstrap replicates are found to be consistent with their target coverage of 95 %. For the polynomial coefficients, the observed coverages range from 91 % to 99 %. The experimental data set illustrates how a complete calibration with uncertainties can be achieved using the method plus one well-known flux level. The uncertainty contribution attributable to estimation of the instrument's nonlinear response is less than 0.025 % over most of its range.

Keywords: Satellite Based Measurements, Spectroradiometer, Calibration, Maximum Likelihood, Bootstrap

1. Introduction

Indicating instruments, such as spectroradiometers, pressure gauges, or scales, require calibrations so that their output may be considered a measurement result [1]. However, such instruments are often calibrated using only a small number of known stimuli, while the subsequent calibrations are expected to be valid over a dynamic range that can span several orders of magnitude. If an instrument's response is known to be perfectly linear in the stimulus, only a single known non-zero stimulus, combined with estimates of uncertainty, is required for calibration. Thus, assessing and correcting the linearity of an instrument's response is a critical component of instrument calibration.

One method for assessing linearity is the "flux-addition" method or the "combinatorial technique." The method is used in the calibration of many types of instruments [2]. The technique exploits the property of a linear function $f(a+b) = f(a) + f(b)$ to construct a system of equations that enables the recovery of the response function of an indicating instrument, up to an overall scaling factor, using a set of unknown stimuli.

A technique for recovering the response function involves assuming a polynomial form and setting up a system of equations that includes both the coefficients of the polynomial and the unknown stimuli. This system can be solved using either pseudoinverse methods or numerical optimization routines. However, because both the stimuli and the response function are unknown, additional constraints are needed to obtain a unique solution. We know of no published method to recover rigorous uncertainty statements, such as confidence intervals, about the estimated parameters. While repeating the experiment until a sufficient number of estimates of the parameters are generated to characterize the uncertainty to the desired level would work, it is not an efficient use of resources. Instead, here we develop a maximum likelihood inversion technique that, in concert with a bootstrap algorithm [3], may be used as part of a full uncertainty analysis when calibrating an indicating instrument. This algorithm is a novel contribution of this paper. We then apply this technique to two systems. The first system is synthetic data based on simulations of a light source and radiometer similar to that used in calibrations of Earth observing satellites, and the second system is NIST's optical Beam Conjoiner [4].

In the simulation, we make a direct comparison of the recovered parameters to the known inputs, enabling an evaluation of the algorithm's performance. We also explain how in the case of calibrations of imagers, a linear calibration may be achieved even in the presence of non-uniform illumination, a critical problem in the calibration of remote sensing instruments, which may have many pixels that observe different parts of a calibration standard. In the application to the Beam Conjoiner, we compare the results of the new method to the results obtained using the traditional inversion technique. We also show how uncertainty in the estimated nonlinear detector response can be incorporated into a calibration using the output from the bootstrap algorithm.

The remainder of the article is organized as follows: Section 2 describes the probabilistic model and techniques used to estimate the fluxes and polynomial coefficients as well as quantify uncertainty in those estimates. Section 3 presents the results of a simulation study using realistic assumptions based on NIST's experience calibrating earth observing satellites. Section 4 presents results for an experimental data set collected at NIST using the Beam Conjoiner and a commercial spectroradiometer. Finally, Section 5 summarizes the main contributions of the article.

2. Methods

Consider an indicating instrument output n in response to a stimulus Φ. In the absence of noise, an instrument has an ideal linear response when $n \propto \Phi$. We are concerned in this paper with instruments that can be described by a relation $n = h(\Phi)$, with h some well-behaved, monotonic, continuous function. In this case, it is of interest to find a "linearizing function" h^{-1} so that $h^{-1}(n) = \Phi$. This is a "linearizing function" in the sense that its output is proportional to the stimulus. For instruments with responses of the type considered here, the linearizing function h^{-1} can be approximated by a polynomial

$$h^{-1}(n) = \beta_0 + \sum_{m=1}^{p} \beta_m n^m,$$

with p chosen to achieve good accuracy. Overfitting is avoided through a constraint in our statistical model. Other functional forms, such as splines, may be substituted into the method presented here, but for concreteness we restrict the development to polynomial functions.

To use a single known non-zero stimulus to calibrate an indicating instrument, a first step is to estimate the shape of the linearizing function h^{-1}. We reduce this problem to estimating the above coefficients, up to an unknown scale factor, following the lead of Ref. [2]. The method is variously called the combinatorial technique, or in the case of radiometric calibrations, the flux-addition or Beam Conjoiner method. No prior knowledge of the magnitude of the stimuli is required, although they must be stable, repeatable, and chosen to sample the dynamic range of the instrument. The technique enables estimation of the magnitude of the observed stimuli and the coefficients of the the linearizing function, up to the aforementioned overall scale factor. Next, a bootstrap algorithm enables quantification of uncertainty in the parameters. Finally, measurement of a known stimulus then constrains the overall scaling of all measured parameters.

In the method presented here, all three steps originate from a generative probabilistic model. The model is generative in the sense that it could be used to generate data that mimic those from a real-world experiment. It is probabilistic in the sense that such data would be generated from a joint probability distribution, not deterministically. The model is described next.

2.1. Statistical Model

For concreteness, we develop our model in the context of calibration of a radiometric measurement device that is viewing a lamp illuminated integrating sphere with a set of $J > 2$ lamps that can be switched on and off. This choice was made because of an interest in optical satellite calibrations, but it does not restrict generalization of the model to other systems. Integrating spheres with two variable light sources have been considered for linearity characterization before [5]. Variable light sources can easily be accommodated in our model by considering each specified level to be a separate lamp and restricting the allowed configurations such that one light source cannot have more than one level on at once. An additional method, which we describe after this simpler version of the model, allows the introduction of correlations between these levels, as might be the case when using a variable aperture. The first model is described by the relations

$$\begin{aligned}
n_i &\sim \text{Normal}\left(\alpha_0 + \sum_{m=1}^{p} \alpha_m P_m\left(s(\Phi_i)\right),\ \sigma^2\right) \\
\Phi_i &= \sum_j x_{ij}\phi_j \\
\max_i\{\Phi_i\} &\sim \text{Normal}\left(\Phi_{\max}, \tau^2\right) \\
\alpha_m &\sim \text{Normal}\left(0, \gamma^2\right) \text{ for } m > 1 \\
\gamma &\sim \text{Exponential}\left(\lambda\right),
\end{aligned} \tag{1}$$

where $\sim$ indicates that the quantity on the left follows the distribution on the right, $\text{Normal}\left(\mu, \xi^2\right)$ is the normal distribution with mean μ and standard deviation ξ, $\text{Exponential}\left(\lambda\right)$ is the exponential distribution with mean λ, n_i is the instrument reading from data point i, Φ_i is the stimulus (here a flux) for data point i, s is a linear transformation of Φ_i to the interval $[-1, 1]$, P_m is the Legendre polynomial of order m [6], $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_p)$ is a vector of polynomial coefficients, ϕ_j is the flux of lamp j, x_{ij} takes the values 0 or 1 for lamp j being off or on in data point i (and the values are restricted to only physically allowed configurations in the case of variable light sources), σ represents noise, $\Phi_{\max}$ estimates the maximum flux, τ is the uncertainty in the estimate of the maximum flux, γ is a penalty parameter that pulls the regression coefficients towards expected values *a priori*, and λ is another penalty parameter that disallows very large values of γ. The noise represented by σ may be due both to instrumental sources and noise in the stimulus. The n_i, the x_{ij}, $\Phi_{\max}$, τ, and λ are taken to be fixed quantities and are inputs to the model. All other parameters, i.e., $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_J)$, $\boldsymbol{\alpha}$, γ, and σ are estimated by maximum likelihood and those estimates are called maximum likelihood estimates (MLEs)‡. If the parameter σ were known or estimated from a separate experiment, it could be treated as a known fixed quantity. The assumption of normally distributed homogeneous noise is not a unique possible choice, and the exact noise behavior of the model must be tuned to match the system under study (see Section 2.5 for an example).

The constraint $\max_i\{\Phi_i\} = \sum_j \phi_j \sim \text{Normal}\left(\Phi_{\max}, \tau^2\right)$, represents a measurement with all lamps on, and sets a scale for the solution. The value of τ should be no larger than σ, and choosing a value much much less than this can impose $\max_i\{\Phi_i\} = \Phi_{\max}$ as a hard constraint. The constraints $\alpha_m \sim \text{Normal}\left(0, \gamma^2\right)$ for $m > 1$ reduce overfitting for a given p. Without the first constraint, the estimation problem is ill-posed, and the second constraint helps to ease the burden of choosing a single optimal value of p (see Section 2.6). The higher order coefficients are forced toward zero by our imposed

‡ See for example Chapters 4 and 5 of [7]

constraints, due to our restriction to well-behaved instruments. Note that the n_i are described in terms of a forward model that is polynomial in the fluxes; a calibration is expected to result in coefficients that give the fluxes in terms of a polynomial in the n_i. This is discussed further in Section 2.2

The strength of the forcing of the polynomial coefficients, to balance prediction variance and bias, is a parameter that is optimized within the model and thus not arbitrary. This forcing is related to the technique known as ridge regression, introduced originally in [8], and the technique known as LASSO (for "Least Absolute Shrinkage and Selection Operator"), originally introduced in [9]. Since the Gaussian distribution is leveraged to achieve the shrinkage, the similarity to ridge regression is more striking. It also resembles standard Bayesian treatments of regression problems, see for example Section 14.8 of reference [10], which uses Gaussian prior distributions for regression coefficients, and [11] which uses Laplace prior distributions.

The statistical model assumes stability and repeatability of lamp flux levels over the duration of the acquisition of the n_i. That assumption is reflected in Equation (1) by the fact that the lamp fluxes, the ϕ_j, do not change across a data set (i.e., they don't depend on i). However, integrating spheres have variable throughput that depends on the average reflectance of the interior surface, and light sources change over time. In Section 3, the effects of variable throughput on the MLEs are examined, and we incorporate changing light sources into our synthetic data to investigate how this affects the utility of the method.

Up to constant terms, the log-likelihood function corresponding to Equation (1) is

$$\begin{aligned}\ell(\boldsymbol{\Phi}, \boldsymbol{\alpha}, \gamma, \sigma) = &-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{N}\left(n_i - \alpha_0 - \sum_{m=1}^{p}\alpha_m P_m(s(\Phi_i))\right)^2\right] - N\log\sigma \\ &-\frac{1}{2\tau^2}\left(\sum_{j=1}^{J}\phi_j - \Phi_{\max}\right)^2 \\ &-\frac{1}{2\gamma^2}\left[\sum_{m=2}^{p}(\alpha_m)^2\right] - (p-1)\log\gamma - \lambda\gamma. \qquad (2)\end{aligned}$$

Equation (2) is based on a joint probability distribution, so by definition, it is a generative model.

The integrating sphere used calibrate the Orbiting Carbon Observatory-2 [12] has 10 lamps, and a single lamp with a variable aperture. To make our model more like this, we make one of the lamps a variable aperture. Under the assumption that there are J lamps with lamp J having a variable aperture, we modify Φ_i from Equation (1) to

$$\Phi_i = \sum_{j=1}^{J-1} x_{ij}\phi_j + \sum_{k=1}^{N_v} x_{ik}^{(J)}\psi_k\phi_J \qquad (3)$$

where the first sum represents the lamps without a variable aperture and the second sum represents the lamp with the variable aperture. The ψ_k, which represent the fractional opening of the variable aperture, are constants in the interval [0,1] to be estimated. The $x_{ik}^{(J)}$ are indicator variables which take the values 0 and 1. At most one of $x_{ik}^{(J)}$ is 1, but they can all be 0. With J lamps including a single variable aperture lamp taking one of N_v levels if illuminated, there are $2^{J-1}(N_v+1)$ unique lamp settings.

A detail of integrating spheres is left out of our model for simplicity. An integrating sphere is a diffuse optical cavity whose throughput will change depending on how it is loaded by opening and closing shutters. Experience with the integrating sphere used in [13] indicates that this effect can be at the 0.1 % level, and it can be accommodated by multiplying the fluxes ϕ_j with experimentally determined configuration factors. The experimental determinations would be made by turning on a single lamp and measuring flux changes as other shutters are opened and closed.

2.2. Linearization Function

The model in Equation (1) is a forward model in the sense that for given values of the lamp fluxes, the polynomial coefficients, and the noise, the probability distribution for the instrument signal is fixed. However, the goal of calibration in this context is to predict the flux viewed by the detector corresponding to the instrument response. This means that to achieve a calibration, we must invert the polynomial equation $E[n] = \alpha_0 + \sum_{m=1}^{p} \alpha_m P_m(s(\Phi))$, where $E[n]$ is the expected instrument reading for the flux Φ. Because we are restricted to well-behaved, monotonic functions, the inverse function is also approximated by a polynomial,

$$\Phi = \beta_0 + \sum_{m=1}^{p} \beta_m E[n]^m. \tag{4}$$

To estimate the parameters $\beta_0, \dots, \beta_p$, from $\alpha_0, \dots, \alpha_p$, a regular sequence of values is constructed over the interval $[-1, 1]$ and transformed by s^{-1}, the inverse of the linear transformation s. Denote this sequence $\widehat{\Phi}_\ell$, with $\ell = 1, \dots, L$. The maximum likelihood estimates of $\alpha_0, \dots, \alpha_p$ are $\widehat{\alpha}_0, \dots, \widehat{\alpha}_p$, and we construct $\widehat{E[n_\ell]} = \widehat{\alpha}_0 + \sum_{m=1}^{p} \widehat{\alpha}_m P_m(s(\widehat{\Phi}_\ell))$. Finally, ordinary least squares is used to estimate $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)$ by minimizing $\sum_{\ell=1}^{L} \left(\widehat{\Phi}_\ell - \beta_0 - \sum_{m=1}^{p} \beta_m \widehat{E[n_\ell]}^m \right)^2$. The estimate is $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \dots, \widehat{\beta}_p)$.

2.3. Uncertainty

Maximizing Equation (2) yields the MLEs for $\boldsymbol{\phi}$, $\boldsymbol{\alpha}$, γ, σ, and $\boldsymbol{\psi}$ but does not yield a quantitative uncertainty for those estimates. To get this uncertainty, we use the statistical bootstrap. This is a more efficient use of data than an alternative approach of repeating the experiment and analysis a sufficient number of times so that uncertainties could be calculated from the distribution of results.

Uncertainty for all parameters is quantified by a bootstrapping pairs approach [see Section 7.2 and 9.5 in Ref. 3, for example]. A "pair" is a response n_i together with a configuration $\boldsymbol{x}_i$. The N pairs of $(n_i, \boldsymbol{x}_i)$ are sampled at random N times with no restriction on how many times a particular pair may be sampled. Some pairs in the original dataset will appear more than once and some will not appear at all in this new bootstrap replicate. Thus, each bootstrap replicate data set will differ from the original data set, and an MLE calculated from a bootstrap replicate data set will differ from the original MLE. The distribution of replicate MLEs approximates the true sampling distribution of the MLE, and so may be used to calculate standard errors and confidence intervals. The standard error is the standard deviation of the set of bootstrap replicates, while confidence intervals are determined by computing quantiles. For example, a 95 % confidence interval is calculated by selecting a range that includes the central 95 % of bootstrap replicates.

While the bootstrap algorithm accounts for sampling variability, it does not account for other sources of uncertainty like model uncertainty, i.e., the mismatch between the model and reality. One potential mismatch is that the model in Equation (1) assumes that the ϕ_j, $j = 1, \ldots, J$ are constant throughout the experiment. If the fluxes of the lamps drift during the experiment, this is not true, but it may be accounted for in the bootstrap algorithm by also randomly perturbing $\Phi_{\max}$ for each bootstrap sample. In the synthetic data in Section 3.1 two cases are considered, with lamps drifting independently of each other, and drifting together identically. These two cases represent possible extremes. Importantly, in the flux-addition method, the stimuli are applied in random order so that systematic drift in the lamps is converted into random noise.

To account for the possibility that the lamps drift during the experiment, $\Phi_{\max}$ in Equation (1) is perturbed by Gaussian white noise with variance $\mathrm{Var}\left[\max_i \{\Phi_i\}\right] = \mathrm{Var}\left[\sum_j \phi_j\right]$ for each bootstrap replicate. This can work well even if the individual lamps drift according to a distribution other than a Gaussian distribution because the assumption of Gaussian drift is placed on the sum of the lamp fluxes, which by the central limit theorem may be reasonable even if the individual lamps drift according to a different distribution. The value of $\mathrm{Var}\left[\sum_j \phi_j\right]$ is an input to the analysis procedure, and it is similar to a prior distribution in Bayesian inference. It must be based on additional data collected during or externally to the calibration experiment. For example, the experimenter could characterize the magnitude of the drift using a separate stable detector observing repeated light levels over time during the calibration experiment.

2.4. Bootstrap Sample Size Determination

The bootstrap, as a general purpose technique for calculating standard errors and confidence intervals, is justified asymptotically. That is, for small samples, it may not work well. A natural question is then, how small is too small? The authors of [14] assert that as long as the sample size is large enough that it is likely that all bootstrap replicates will be unique, the results of the procedure should be “reliable.” They go on to propose sample sizes such as 20, 30, and 50, drawing from the work of [15] and [16] for further support. In the experimental scenarios considered here, very small samples sizes are not expected. However, there are further considerations for applying a bootstrap algorithm to the cases we consider here, namely that each bootstrap replicate must sufficiently cover the whole response range $[0, \Phi_{\max}]$.

Consider for pedagogy $\Phi_{\max} = 1$, and the following sequence of target fluxes $\Phi_i = 7.5 \times 10^{-5} \cdot 1.099^i, i = 1, \ldots, 100$. The sequence starts at about 8.2×10^{-5}, and ends at about 0.94. However, only 7 values (less than 10 %) are larger than 0.5. The sequence was chosen for illustrative purposes, but in the Beam Conjoiner application of Sections 2.5 and 4 a non-uniform sequence of fluxes with more small fluxes is observed. The non-uniformity is important because in each bootstrap replicate, the probability that any one observation will be left out of a bootstrap replicate is approximately e^{-1}, a little more than $\frac{1}{3}$. Thus, in the illustrative example 6 or 7 of the fluxes greater than 0.5 would be left out of about 1 out of every 100 bootstrap replicates, and 5 or more would be left out of about 1 out of every 16. For some bootstrap replicates, this will cause poor coverage of the flux range (0.5, 1.0), which will significantly degrade the performance of the bootstrap algorithm there. For example, when p is large, the bootstrap replicate of the non-linear detector response can fluctuate unrealistically in flux ranges with no data points. Fortunately, problems are easily diagnosed with graphical checks described in Section 4. In cases where poor coverage of the flux range may arise, we recommend either changing the experiment design to avoid a non-uniform sequence of flux values or to obtain repetitions.

2.5. Beam Conjoiner

In Section 4, we describe the results of an application of the methods here to data from the NIST Beam Conjoiner, an instrument which is described in detail in [4] and briefly here. The Beam Conjoiner comprises a light source (typically a single quartz tungsten halogen lamp) whose output is collimated and split along two different paths. Each beam passes through one neutral density filter wheel (filter 1 and filter 2) with four levels filters as well as a blocking insert. The beams are recombined and directed through a third, shared, neutral density filter wheel (filter 3) with five levels filters and a blocking insert before being viewed by the instrument. The forty filter combinations enable a dynamic range of signals of about 500 in a single measurement scan of all of the combinations. Additional ranges are achieved using an external neutral density filter wheel which then can be used to obtain a dynamic range of 14 orders of magnitude [17]. A schematic of the beam conjoiner is given in Figure 1.

To use the model in Equation (1), we make a slight adaptation for use with the beam conjoiner. There are different and equivalent ways to adapt Equation (1). The simplest is to label every combination of settings for filter 1 and filter 3 as one "lamp" and every combination of filter 2 and filter 3 as another "lamp". Then constraints on the x_{ij} are imposed to ensure that only physical combinations are enumerated. Thus if the x_{ij} are represented as an $N \times 40$ matrix, each of the first and last 20 columns would have at most a single nonzero value in each row, with each of these nonzero values representing a particular filter combination.

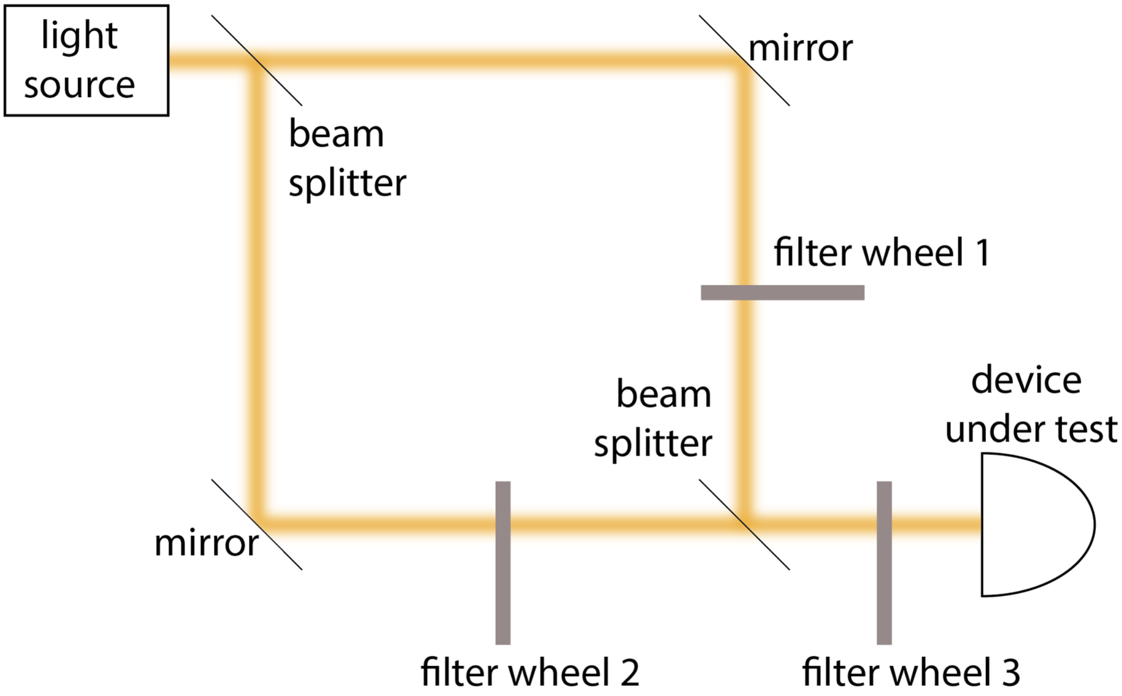

Figure 1: Schematic diagram of the Beam Conjoiner showing the primary elements. Filter wheels are tilted so that reflections are directed into the anti-reflection coated floor of the apparatus.

In our analysis of the data, we found a modification to the noise model to be necessary. We had made a simplifying assumption that instrumental noise would dominate, as represented by a single parameter σ that is constant across a data. However, in some systems, electronic noise will dominate at low signal levels and at high signal levels the light source and photon shot noise will dominate. As a result, σ from Equation (1) becomes σ_i, and for the Beam Conjoiner application takes the form

$$\sigma_i = \begin{cases} \sigma\,\Phi_i & \text{if } \Phi_i > \kappa_0\Phi_{\max} \\ \sigma\kappa_0\Phi_{\max} & \text{if } \Phi_i \leq \kappa_0\Phi_{\max}. \end{cases} \tag{5}$$

In Equation (5), the noise is assumed constant until the flux exceeds $\kappa_0\Phi_{\max}$, and after that it is assumed proportional to the flux, as would be appropriate in a case where fluctuations in the light source but not shot noise are dominant.

2.6. Polynomial Order

For real-world data, it is necessary to choose the polynomial degree p. The typical approach to choosing such a model parameter is to balance the trade-off between prediction variance and bias. As p increases, so too does the prediction variance, but prediction bias simultaneously decreases.

The likelihood in Equation (2), attempts to balance the bias-variance trade-off

ID	**identical lamps**	**linear drift**	**correlated drifts**
1	✓	0 %	NA
2	✓	0.5 %	✗
3	✓	0.5 %	✓
4	5 % range	0.5 %	✓

Table 1: The four conditions considered in the simulation study. NA means "not applicable."

automatically through the parameter γ. As γ approaches zero, the α_m for $m = 2, \ldots, p$ approach zero. In this limit, variance is minimized (zero variance), but bias is maximized. Conversely, as γ approaches ∞, bias is minimized, but variance is maximized. By estimating γ along with the other parameters, the two extremes are balanced. This implies that p should be chosen to be very large to allow bias and thus mean squared error to be minimized, and that minimization is part of the estimation process. If p is too small, the bias will be artificially inflated.

Practically, however, p cannot be chosen to be arbitrarily large. A range of values of p producing an optimal and stable MSE may be identified by K-fold cross validation (see for example Chapter 5 of [18]). Briefly, to carry out K-fold cross validation, the complete data set is partitioned at random into K parts. The MLEs are calculated based on $K-1$ parts, and those estimates are used to predict the instrument responses from the left out part. An MSE is calculated using the predicted and observed instrument responses. The process is repeated K times leaving each part out of the estimation once, resulting K MSEs.

3. Simulation Study

The methodology described in Section 2 was tested using synthetic data under four conditions of practical interest, listed in Table 1. For each condition, 100 sets of instrument readings, the n_i, were generated. For each set of n_i, the methodology of Section 2 was applied, and the results compared to the known truth. The comparisons are presented in Section 3.2. The generation of the sets of instrument readings are discussed next.

3.1. Signal Simulation

To generate one set of instrument readings n_i, the lamp configurations, x_{ij} and $x_{ik}^{(J)}$, are first defined. The number of lamps is taken to be $J = 7$. One lamp has a variable aperture and the number of open variable aperture positions $N_v = 4$ is used. These choices are arbitrary but similar to values on existing spheres used for calibration of remote sensing instruments [12, 19]. For the six lamps that do not have a variable aperture, there are $2^6 = 64$ possible off/on configurations. The variable aperture lamp can take on 5 possible states, off, on at full intensity, or on at one of the three intensities that are lower than full intensity. This gives $64 \times 5 = 320$ possible lamp configurations. Each simulated data set contains all 320 possible lamp configurations, plus five repetitions with all lamps off and five repetitions with all lamps on at full intensity, giving a total of $N = 330$ data points in each data set.

Next, values are chosen for the ϕ_j, $j = 1, \ldots, 7$, ψ_k, $k = 1, \ldots, 4$, and $\beta_0, \ldots, \beta_p$. In all simulations, $p = 3$, $\psi_1 = 0.25$, $\psi_2 = 0.5$, $\psi_3 = 0.75$, $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.022$, and $\beta_3 = -0.008$. Note that all quantities here are chosen to be dimensionless, as our procedure for recovering the inputs is used on normalized data. For simulation 1, where the lamps are identical and do not drift, $\phi_j = \frac{1}{7}$, $j = 1, \ldots, 7$. To represent drift in the simulations 2 through 4, each initial $\phi_j = \frac{1}{7}$ is multiplied by a random number between 0.995 and 1.005 times the number in the sequence of data points divided by N. In simulation 2 with "uncorrelated drifts", a different random number is used for each lamp in each data set. For simulations 3 and 4, which have "correlated drifts", the same random number is used for all lamps, with a different number used in each data set. For simulation 4, the ϕ_j are randomly selected to be within approximately 2.5 % of $\frac{1}{7}$, subject to a sum to unity constraint. This is summarized in Table 1.

Given values for all of the x_{ij}, $x_{ik}^{(J)}$, ϕ_j, and ψ_k, $\Phi_i = \sum_j^{J-1} x_{ij}\phi_j + \sum_k^{N_v} x_{ik}^{(J)}\psi_k\phi_J$ (for readability, the drifts are omitted here). Then, Φ_i is perturbed by Gaussian white noise with standard deviation $1.1 \times 10^{-4}\sqrt{\Phi_i}$ (to represent shot noise) yielding $\widetilde{\Phi}_i$. This value is for illustrative purposes. The magnitude of the shot noise selected here is in only slight conflict with the homogeneous noise model of Equation (1), but a real-world application may require tuning of the noise model, as is done in Section 2.5. We found improved numerical convergence when using the homogeneous model, and it did not result in bias, as evidenced in the results section below.

Given values for all of the β_m, $\widetilde{n}_i$ is obtained by solving the equation

$$\widetilde{\Phi}_i = \beta_0 + \sum_{m=1}^{p} \beta_m \widetilde{n}_i^m.$$

Because we restrict our analyis to well-behaved instruments with monotonic responses, there is a unique real solution. Finally, Gaussian white noise (electrical noise) with standard deviation 10^{-3} is added to the $\widetilde{n}_i$ to yield n_i. Because of the arbitrarily chosen nonzero value for $\beta_0 = -0.5$, $-0.5 \lessapprox n_i \lessapprox 0.5$.

3.2. Results

To apply our method, values for $\Phi_{\max}$, τ, and λ must be selected. Since for each simulation, the true values of the ϕ_j $j = 1, \ldots, 7$ are exactly $\frac{1}{7}$, or on average $\frac{1}{7}$ in the case of the simulations 2, 3, and 4, $\Phi_{\max}$ is set to unity. The values for τ and λ are 0.001 and 1, respectively. For the bootstrap algorithm, it is also necessary to specify $\mathrm{Var}\left[\max_i\{\Phi_i\}\right] = \mathrm{Var}\left[\sum_j \phi_j\right]$, as described in Section 2.3. For simulation 1, it is set to zero since the lamps do not not drift. For simulations 2 and 3, it is set to 0.00055, and for simulation 4, 0.0014. The values for simulations 2, 3, and 4 are based on the known distribution of the amount of drift for the lamps.

Figure 2 shows the recovered MLEs for β_0, β_1, β_2, and β_3 for 100 simulated data sets for each of the 4 simulation scenarios described in Table 1. The points are the MLEs ordered from smallest to largest, and the shading represents 95 % confidence intervals. The black vertical lines are the true values, and the numbers to the right are the proportion of confidence intervals that bracket the true value, the target being 0.95. From Figure 2, it can be seen that the true values of β_0, β_1, β_2, and β_3, lie around the center of the 100 maximum likelihood estimates for each scenario. This implies that the estimation procedure is approximately unbiased. To put it another way, the true values are approximately equal to the estimates, on average. For β_0 and β_1, for the 100 simulations pictured, the observed relative bias is less than 0.1 %. For β_2 it is less than 1 %, and for β_3, it is less than 4 %.

Also from Figure 2, it may be seen that the bootstrap procedure described in Section 2.3 achieves its nominal coverage of 95 % for β_0, β_1, β_2, and β_3. This claim is tenable even though only one of the values on the right side of Figure 2 is exactly 0.95. Because of the finite number (namely, 100) of data sets for each simulation, 1 of the 16 proportions in Figure 2 is expected to lie outside of the interval $(0.91, 0.99)$ even if the true coverage is exactly 95 %, none do. Another important observation from Figure 2 is that for β_0 and β_1 the confidence intervals are shortest for simulation scenario 1, and longest for scenarios 3 and 4. This occurs because of the extra uncertainty introduced into the bootstrap algorithm that is intended to account for lamp drift during the experiment. The widths of the confidence intervals for scenario 2 are in between the two extremes of perfectly correlated drift between the lamps in scenarios 3 and 4 and no drift in scenario 1. The widths of the confidence intervals for β_2 and β_3 do not vary systematically between the simulations, implying that uncertainty for those higher order coefficients is primarily due to sampling variability not lamp drift.

Figure 3 shows similar results as in Figure 2, but for ψ_1, ψ_2, and ψ_3. Again, the true values are centered around the 100 maximum likelihood estimates, implying that the estimators are unbiased. The observed relative biases are less than 0.1 %. Further, the coverage proportion of the confidence intervals is consistent with the nominal 95 %. In Figure 3, it is also observed that the widths of the confidence intervals between simulations are similar. As with β_2 and β_3, this implies that the dominant component of uncertainty for ψ_1, ψ_2, and ψ_3 is sampling variability, not lamp drift.

Figure 4 is similar to Figures 2 and 3, but for $\phi_1, \ldots, \phi_7$. Simulation scenario 4 is not plotted because lamps are not assumed to be nominally identical as in scenarios 1, 2, and 3. Again, the true average lamp fluxes are centered about the 100 maximum likelihood estimates, implying no bias. The observed relative biases are less than 0.1 %. On the other hand, for simulations 2 and 3, the coverage proportions of the confidence intervals are not consistent with the nominal 95 %. This occurs because the lamp fluxes are not exactly $\frac{1}{7}$, but instead drift around $\frac{1}{7}$ during the experiment. The deviation of the coverage proportions from the nominal coverage for $\phi_1, \ldots, \phi_7$ when the lamps drift

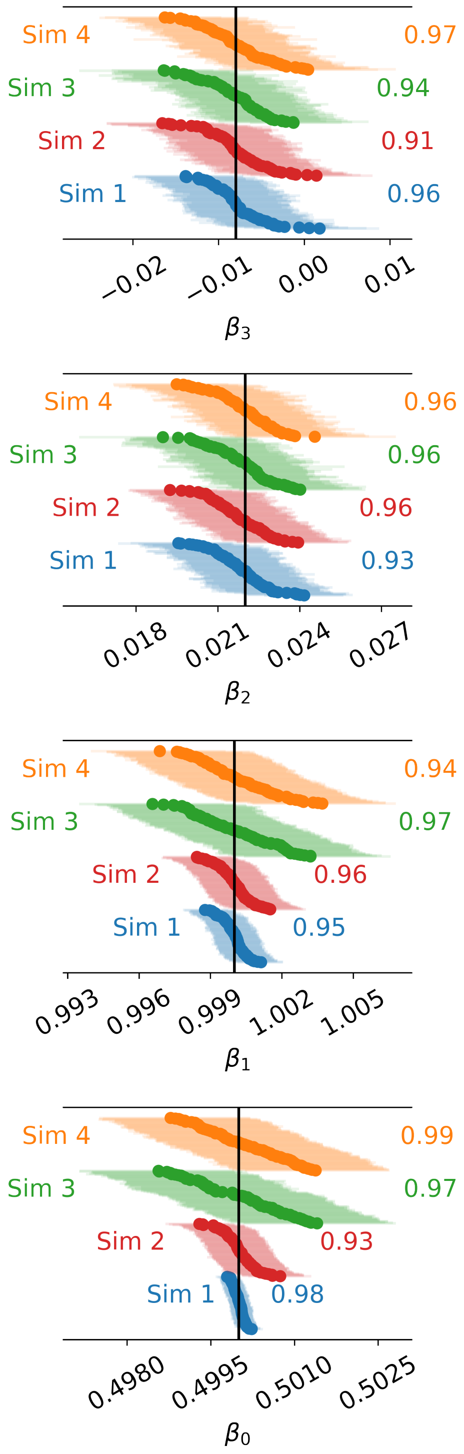

Figure 2: Results for β_0, β_1, β_2, and β_3 for 100 simulated data sets for each of the 4 simulation scenarios described in Table 1. The points are the MLEs ordered from smallest to largest, and the shading represents 95 % confidence intervals. The vertical axis indexes the data set within a simulation, 1 to 100. The black vertical lines are the true values, and the numbers to the right are the proportion of confidence intervals that bracket the true value.

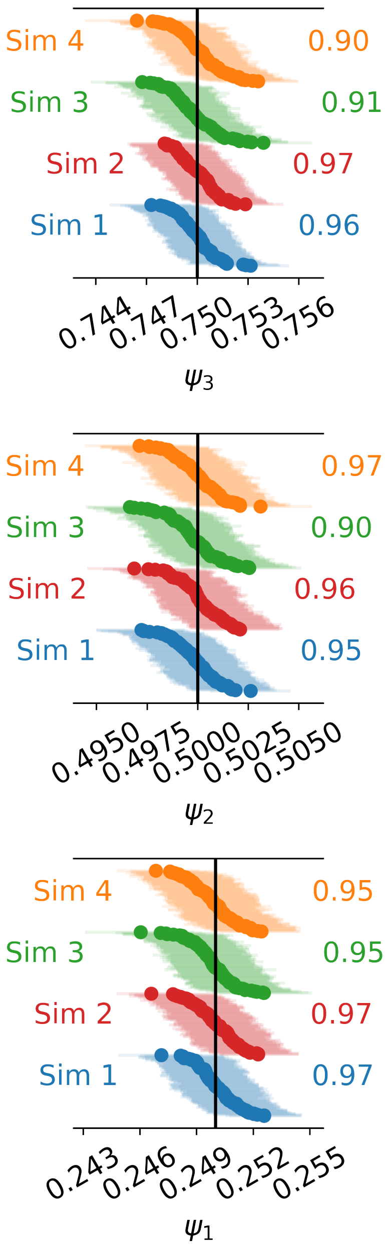

Figure 3: Results for ψ_1, ψ_2, and ψ_3 for 100 simulated data sets for each of the 4 simulation scenarios described in Table 1. The points are the MLEs ordered from smallest to largest, and the shading represents 95 % confidence intervals. The vertical axis indexes the data set within a simulation, 1 to 100. The black vertical lines are the true values, and the numbers to the right are the proportion of confidence intervals that bracket the true value, the target being 0.95.

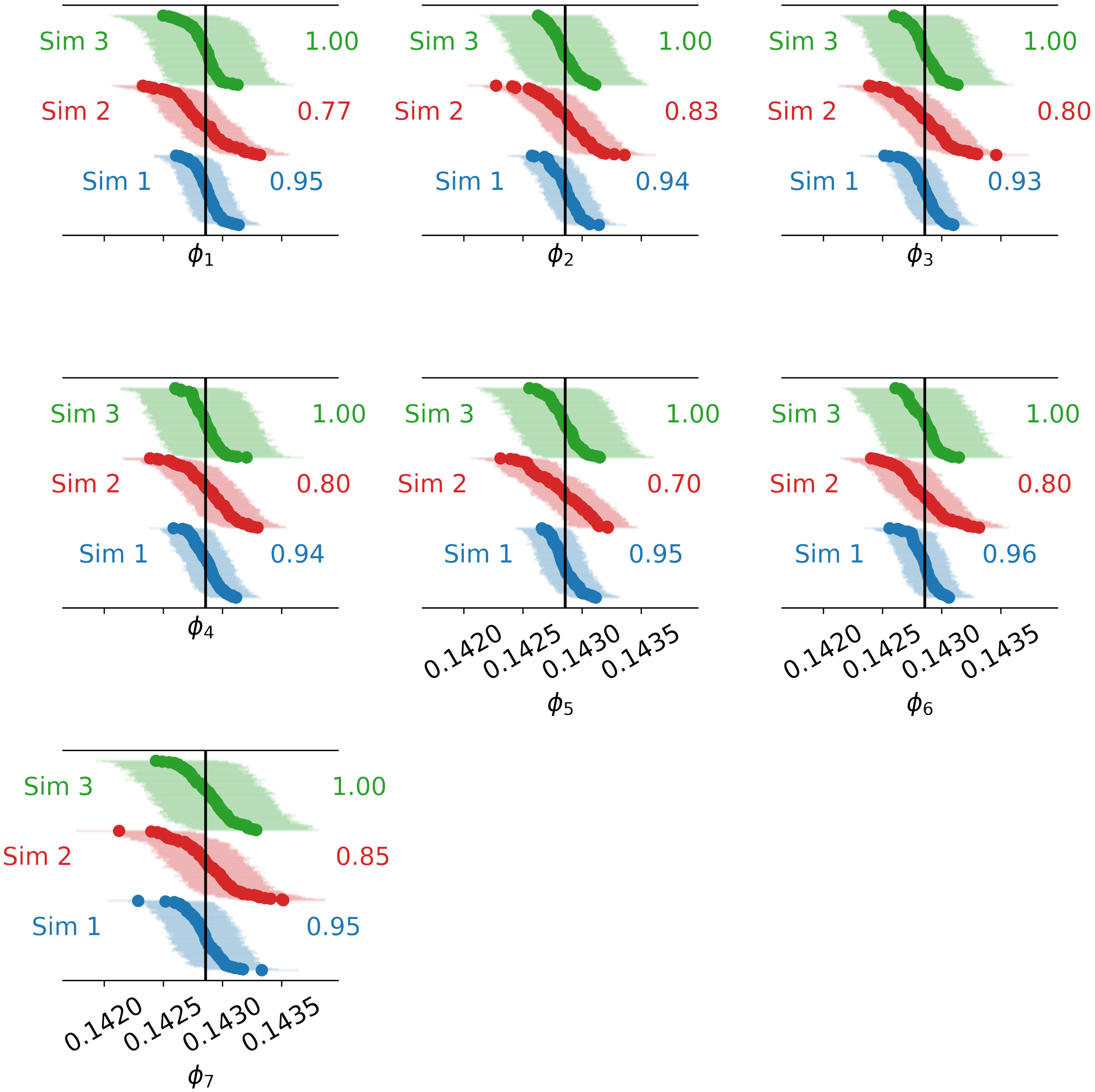

Figure 4: Results for $\phi_1, \ldots, \phi_7$ for 100 simulated data sets for the first 3 of 4 simulation scenarios described in Table 1. The points are the MLEs ordered from smallest to largest, and the shading represents 95 % confidence intervals. The vertical axis indexes the data set within a simulation, 1 to 100. The black vertical lines are the true values, and the numbers to the right are the proportion of confidence intervals that bracket the true value, the target being 0.95. The 4^{th} simulation scenario is not presented because there is not a single true value. Tick marks refer to the same values in all graphs.

is not concerning since the modified bootstrap algorithm accounts for the variation due to drift through $\Phi_{\max}$, which affects the individual ϕ_j only indirectly through their sum.

The simulation scenarios considered here show that the maximum likelihood estimation procedure and the non-parametric bootstrap algorithm work well for estimating nonlinear detector response as well as variable lamp intensities. For the simulations where the lamps do not drift during the experiment, the procedures also work well for estimating the true lamp intensities. The only cases for which the bootstrap algorithm did not convey good estimates of the uncertainty were simulation scenarios 2 and 3 (and 4, although not presented) for the true lamp intensities, $\phi_1, \ldots, \phi_7$. In those simulations, the lamps are assumed to drift during the experiment, and so in reality there is no single true value to compare the estimated lamp intensity to. Thus, we do not find it concerning that the coverage proportions for the bootstrap confidence intervals are not consistent with the target 95 %. However, even in the case of drifting lamps, the maximum likelihood estimates center around $\frac{1}{7}$ because, over the 100 simulations, it is the average lamp intensity.

4. Beam Conjoiner

A dataset to evaluate the linearity of a spectroradiometer belonging to NIST was taken on the beam conjoiner. A total of four runs through each of the 150 filter wheel positions is included in our dataset. For the present work we average across a small set of pixels near the peak of the spectrum and use our methods to extract the nonlinearity.

Some practical matters must be addressed before applying our methods to the Beam Conjoiner data. First, we study the required polynomial order, per Section 2.6. The results of this study are shown in in Figure 5. For $8 \leq p \leq 15$, the distributions of points (blue circles) and means (orange circles) are nearly identical; although, the minimum mean is achieved for $p = 10$. Based on Figure 5, any p in the range of $8 \leq p \leq 15$ is a reasonable choice. The values of MSE in Figure 5 may be interpreted in the following way: when p is small, prediction bias dominates the MSE. As p increases, prediction bias decreases, and the γ parameter in Equation (2) automatically balances the trade off between prediction bias and variance, which produces a range of p for which the MSE is optimal and stable.

The quantity κ_0, representing the flux level at which noise from the light source begins to dominate, from Equation (5) must also be chosen. In this case it was done by inspecting residuals after fitting the constant noise model, and $\kappa_0 = 0.2$ was selected. It is also necessary to select $\Phi_{\max}$, τ, λ. As in the simulations, $\Phi_{\max}$ and λ are set to unity and $\tau = 0.0001$. Since we set $\Phi_{\max} = 1$, the n_i are also scaled to the interval $[0, 1]$ before the analysis.

The results of this section do not expand uncertainty by perturbing $\Phi_{\max}$ in the Monte Carlo bootstrap algorithm, but in practice, it may be necessary, depending on the deployment conditions. This makes it analogous to the conditions in simulation scenario 1.

Figure 6 plots the residuals $n_i - \widehat{\alpha}_0 - \sum_{m=1}^{p} \widehat{\alpha}_m P_m(s(\widehat{\Phi}_i))$ for $p = 10$ on the vertical axis versus the estimated fluxes $\widehat{\Phi}_i$ on the horizontal axis. As noted previously, and

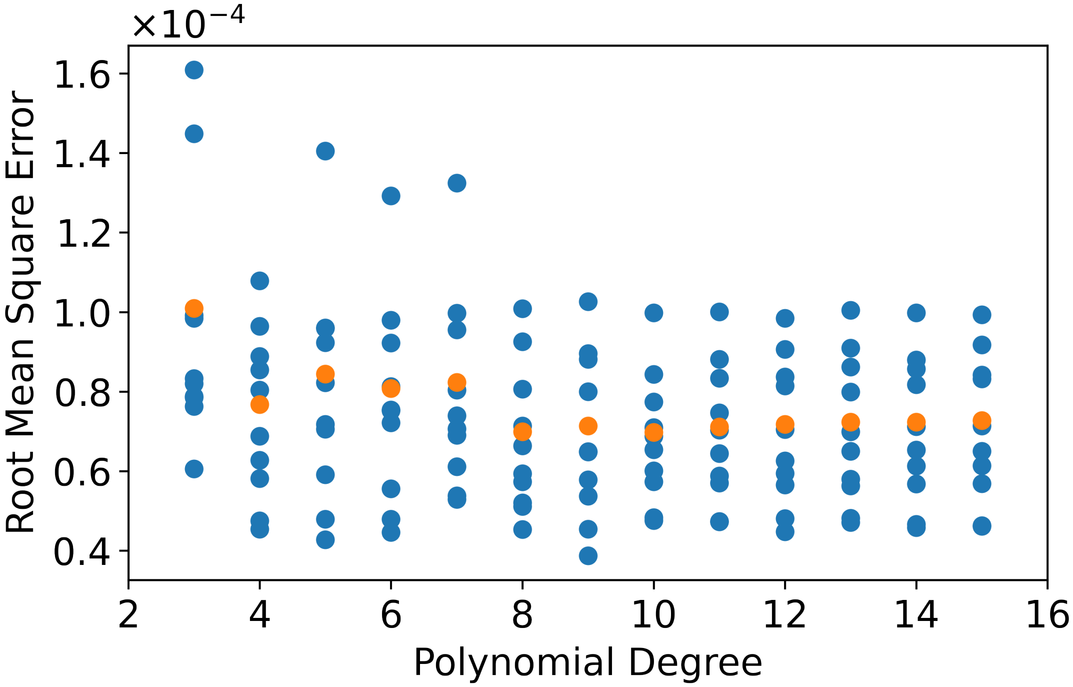

Figure 5: The 10-fold cross validation error plotted as a function of the polynomial degree p for the Beam Conjoiner data discussed in Section 2.5. For each polynomial degree, the 10 estimates of root mean squared prediction error (MSE) are plotted (blue circles) as well as the square root of the average MSE (orange circles).

accounted for by Equation (5), the increasing noise with flux is apparent in Figure 6. The residuals at each estimated flux level are symmetrically distributed around zero. This implies that the model described in Equation (1), with the modifications detailed in Section 2.5, represents the Beam Conjoiner data well. The light red band in Figure 6 depicts 95 % pointwise prediction bounds for the residuals. The target 95 % coverage of the prediction bounds is achieved. This implies that the bootstrapping pairs algorithm properly accounts for uncertainty in the estimated model parameters and that the model for the noise in Equation (5) is appropriate. The light grey band in Figure 6 depicts 95 % pointwise confidence bounds for the mean residual. By construction the band contains zero, but at higher flux values and those where there are fewer observed data, the bands widen indicating more uncertainty in the average residual.

Figure 7 shows the estimated nonlinear detector response where the horizontal axis represents flux and the vertical axis the detector response; the dashed green line is the diagonal, the black curve is the estimated mean detector response, and the red points are instrument readings, n_i. Note that in Figure 7 the deviation between the diagonal and the estimated mean detector response has been exaggerated; otherwise, visually, the curves are indistinguishable. Figure 8 shows the estimated non-linear response in a different way, depicting differences between the estimated expected detector response

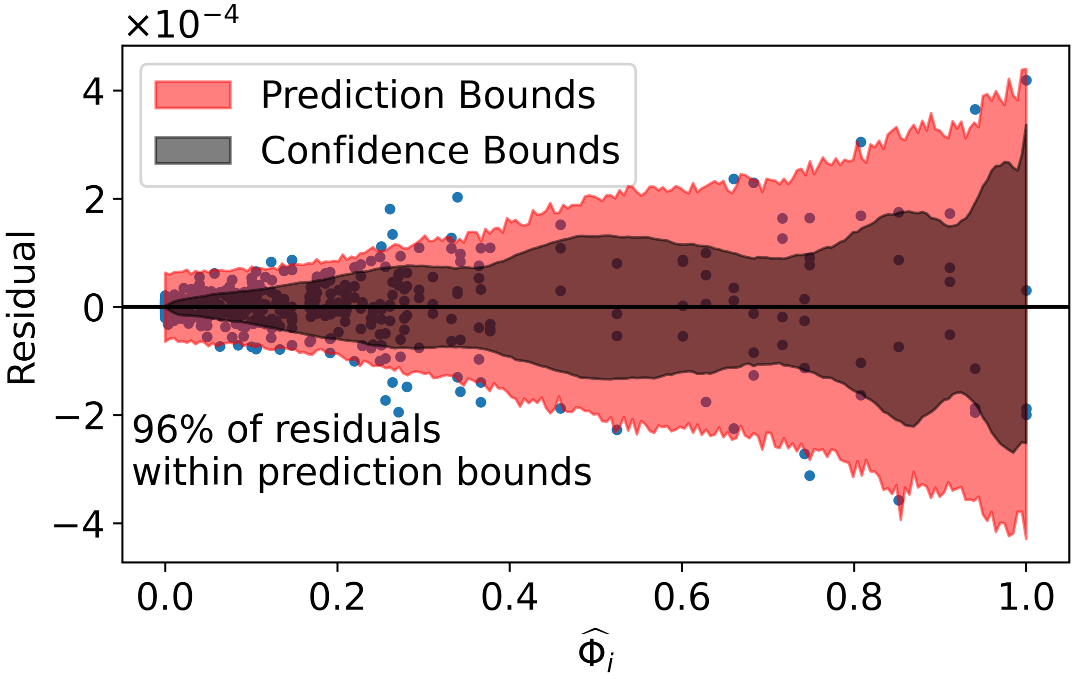

Figure 6: Residuals versus the estimated fluxes (blue points), 95 % pointwise prediction bounds for the residuals (light red band), and 95 % pointwise confidence bounds for the mean residual.

and the diagonal (black curve), bootstrap replicates of that difference (blue curves), and differences between the the observed detector responses and the diagonal (red points); the diagonal in Figure 7 becomes a horizontal line at zero in Figure 8. No exaggeration for visual effect is needed in Figure 8. There are 986 bootstrap replicates shown because 14 of the attempted 1000 bootstrap replicates encountered failures in optimization of the likelihood function. The Beam Conjoiner data exhibit statistically significant nonlinear behavior as seen in Figure 8, where the bootstrap replicates do not envelop zero. The orange curve in Figure 8 is discussed below in Section 4.1.

We recommend that users graphically check the fit of the model to the data. At a minimum, figures such as 6 and 8 should be examined. For example, in Figure 6, if the residuals were not symmetrically distributed around zero for each value of $\widehat{\Phi}$, p should be increased. Also in Figure 6, the 95 % prediction bounds should envelop about 95 % of the residuals, and if they do not, the noise model should be modified. In Figure 8, both the estimated non-linearity and the bootstrap replicates of it should be examined for unrealistic fluctuations. As explained in Section 2.4, by random chance, some bootstrap replicate data sets may cover the flux range of interest poorly, which can lead to unreasonably large oscillations of the estimated non-linearity for those replicates. Potential remedies are suggested in Section 2.4.

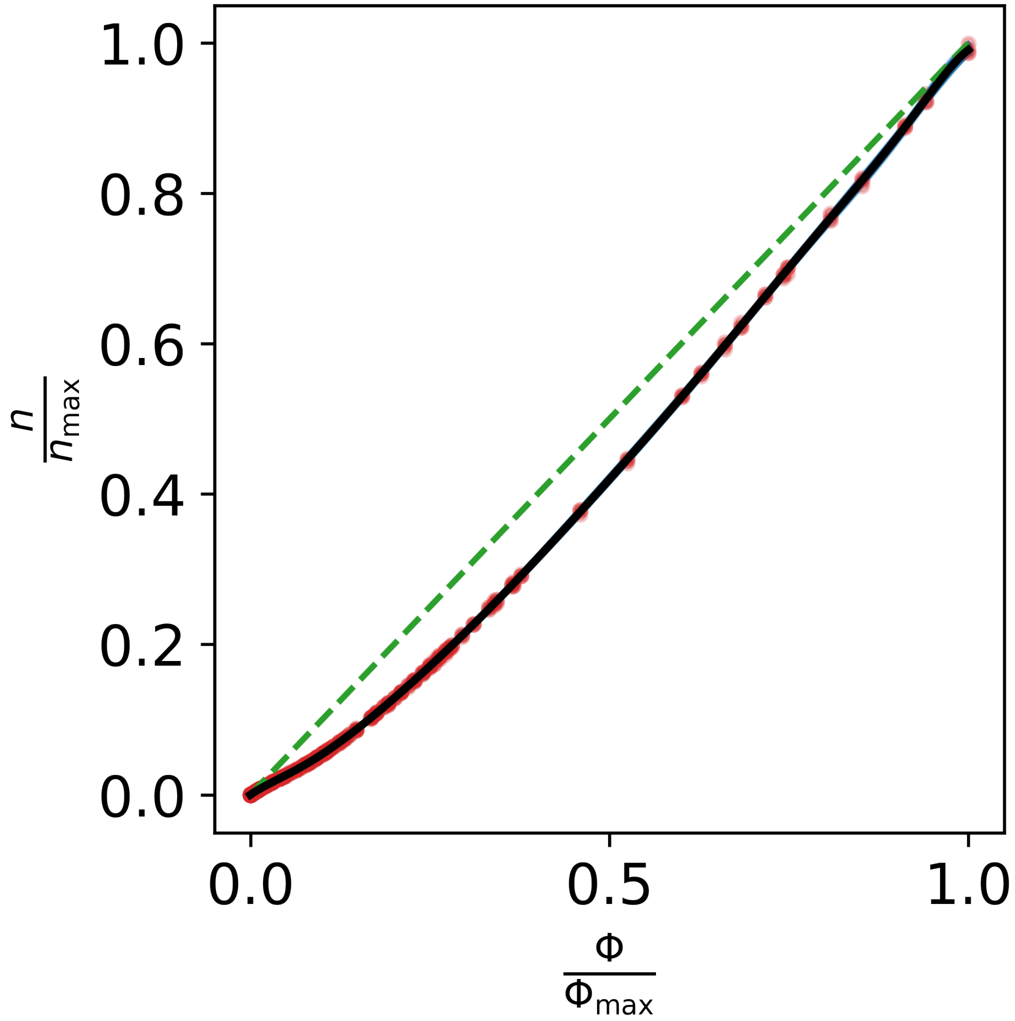

Figure 7: The horizontal axis depicts the flux, and the vertical axis depicts the instrument reading. The black curve is the MLE of the expected instrument reading. The red points depict the observed n_i. The deviation between the expected detector response and the (green dashed) 1:1 diagonal are exaggerated by a factor of 25; otherwise, the curves would be visually indistinguishable.

4.1. Comparison to an Existing Approach

The method for estimating $\beta_0, \ldots, \beta_p$, $\phi_{11} \ldots \phi_{1J_1}$, and $\phi_{21}, \ldots, \phi_{2J_2}$ from Ref. [4] is based on optimizing a least-squares-like objective function. Using similar notation Equation (2), one form of the objective function is

$$\mathcal{O} = \left[\Phi_{\max} - \max_i \{\Phi_i\}\right]^2 + \sum_{i=1}^{N} \left[\left(\beta_0 + \sum_{m=1}^{p} \beta_m n_i^m\right) - \Phi_i\right]^2 . \tag{6}$$

where the first term fixes the normalization of the problem. An alternative to this normalization, as is done in [4], is to set the coefficient of the linear term to unity. The largest difference between the use of this objective function and our model is the firm statistical justification on which our model lies. While the objective function often "works," it is challenging to understand where its limitations may lie. Mathematically, the objective function, unlike our model, is not derived from a generative probability model as is the likelihood function in Equation (2). We consider it to be a best practice when making statistical inferences from data to begin with a generative probability model, and indeed it is a requirement in two major paradigms of statistical inference: maximum likelihood and Bayesian inference. Further, the model presented in this paper has been adapted to deal with differing spread in the residuals at different signal levels apparent in Figure 6. This adaptation has not been done with the objective function.

For comparison with our method, Figure 8 includes an estimate of detector nonlinearity based on minimization of the objective function (orange curve). The two procedures produce similar estimates. The largest differences are found at high values of flux where there is more distinction between the approaches since the maximum likelihood approach of this work accounts for increasing variability in the detector response as flux increases. The estimate of nonlinearity based on Equation (6) is also within the range of the bootstrap replicates of the estimated nonlinearity based on Equations (1) and (2) (light blue curves). These observations provide further confidence in the proposed approach.

To recap, both approaches produce similar estimates of the nonlinear behavior of the detector, but the approach proposed herein adds flexibility, a mechanism for choosing the degree of the polynomial, p, and rigorous confidence and prediction intervals that may be used to express uncertainty in the results.

4.2. Calibration

A final step in producing a calibration using the Beam Conjoiner is to scale the results using a single, known flux level. To isolate the uncertainty in a calibration due to nonlinearity, we perform an analysis assuming we have a single known flux level Φ_{ref} with zero uncertainty. Also, the average detector response, $E[n_{\text{ref}}]$ is considered to have been measured a sufficient number of times to have negligible uncertainty. Extending the result to nonzero uncertainty in the Φ_{ref} may be performed using Monte Carlo error propagation techniques, though we emphasize that what we show here is uncertainty due to estimation of the linearization function, which is not affected by uncertainty in the reference measurements.

Uncertainty estimation is done by applying the calibration to every bootstrap replicate individually, so that each predict a flux of exactly Φ_{ref} when $n = E[n_{\text{ref}}]$. Each bootstrap replicate represents a plausible shape for the response of the instrument, and a calibration scales and offsets these so that each give the same result at zero signal and at the single calibration level. The uncertainty attributable to empirical estimation

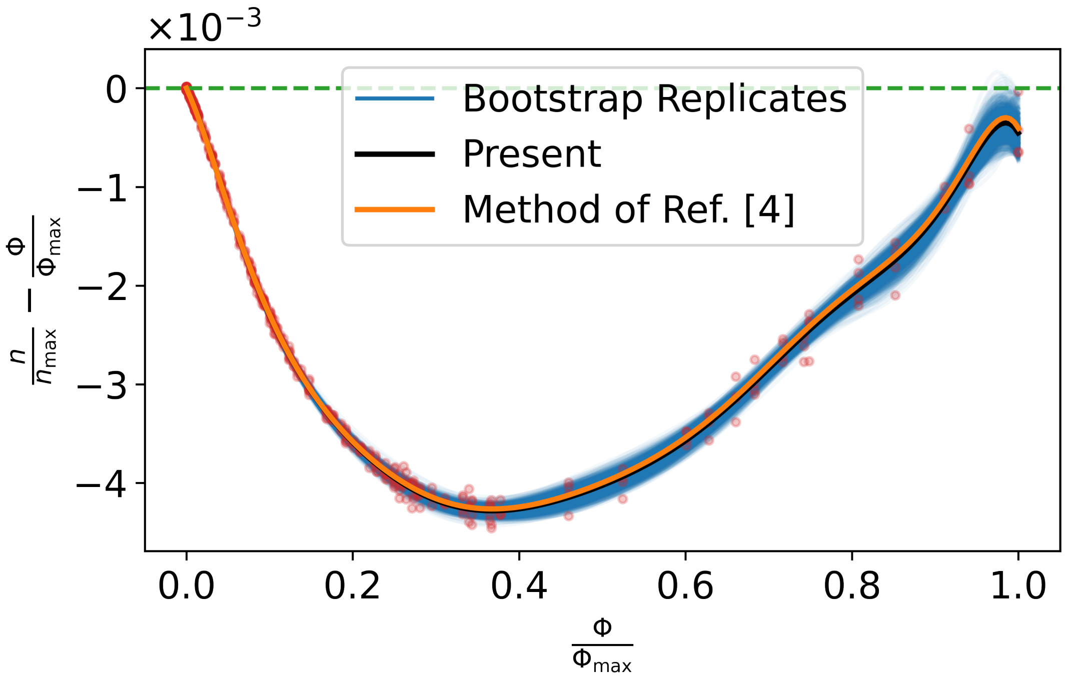

Figure 8: The horizontal axis depicts the flux, and the vertical axis depicts the instrument reading minus the flux. The black curve is the estimated expected instrument reading minus the flux where the estimated expected instrument reading is calculated by the methods described in the present work. The blue curves are bootstrap replicates (as described in the present work) of the estimated expected instrument reading minus the flux. The red points depict n_i minus the estimated flux for data point i, $\widehat{\Phi}_i$ where the $\widehat{\Phi}_i$ are calculated by the methods described in the present work. The orange curve is an estimate of the expected instrument reading minus the flux based on Equation (6) and [4].

of the linearization function is then represented by the spread in the scaled bootstrap curves as illustrated in figure 9a, where we have taken $E[n_{\text{ref}}] = 0.5$ and $\Phi_{\text{ref}} = 0.5$, which are chosen because indicating instruments are generally designed to perform well at midrange. Figure 9b represents these data as a relative uncertainty. It may be seen directly in Figure 9b, that for most of the range of $\widehat{\Phi}_{\text{cal}}$, the bootstrap replicates are within 0.025 % of the calibrated maximum likelihood result.

5. Summary

This article investigates the problem of quantifying uncertainty when applying the radiometry technique known as the flux-addtion method. There are four primary novel contributions:

(i) A generative probability model applicable to the calibration of radiometeters is

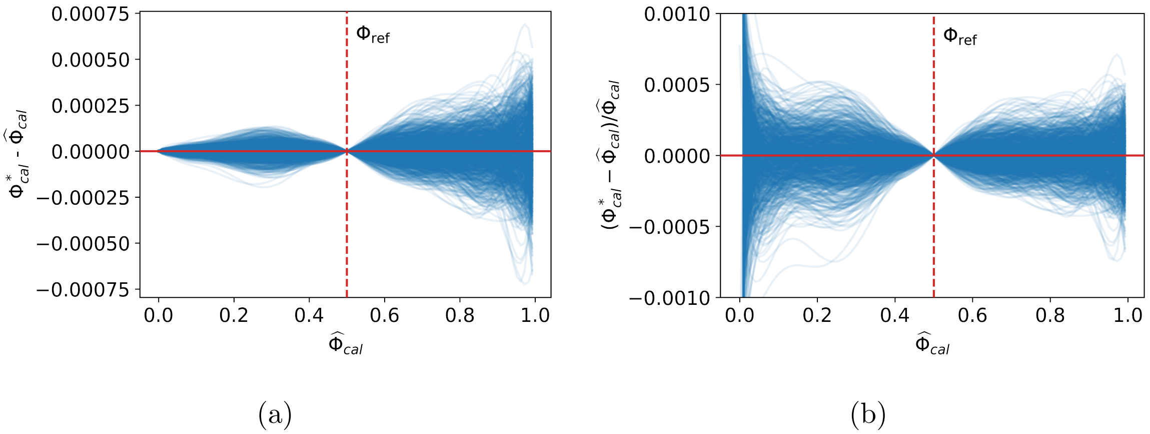

Figure 9: In both figures, the horizontal axis depicts the MLEs of calibrated fluxes, $\widehat{\Phi}_{\mathrm{cal}}$. In (a) the vertical axis depicts bootstrap replicates of MLEs of calibrated fluxes minus MLEs of calibrated fluxes. Each curve represents a single bootstrap replicate. The vertical axis in (b) is similar, but the relative difference is plotted.

introduced. This permits the application of maximum likelihood.

(ii) A non-parametric bootstrap algorithm is described for the purpose of making uncertainty statements about estimates of the parameters of the probability model.

(iii) The coverage probabilities of confidence intervals derived from the bootstrap algorithm are assessed via simulation.

(iv) The probability model requires the choice of a polynomial degree. A cross-validation-based strategy is proposed to help make that choice.

The probabilistic model and bootstrap algorithm were tested in a simulation study. In the simulation study, the MLEs for the polynomial coefficients of the linearizing function were found to be approximately unbiased, meaning when estimates are averaged across many data sets, the average value is approximately the true value. The relative bias observed in the simulation study for the polynomial coefficients defining the linearization function was less than 4 %; although, for the intercept, linear, and quadratic coefficients, it was less than 1 %. The simulation study also shows that bootstrap percentile confidence intervals for the same polynomial coefficients are consistent with their nominal 95 % coverage. In 100 simulations, the observed coverage proportions ranged from 91 % to 99 %. Similar results were obtained for the other model parameters, e.g., fractional flux factors for variable aperture lamps.

In simulation configurations 2, 3, and 4 (Table 1), where the lamps drifted, we found that confidence intervals were too short when we failed to account for this additional source of variability. To get confidence intervals that were consistent with their target coverage (95 %), additional noise was injected into the bootstrap algorithm, comeasurate with the size of the drift. The real world implication is that to use this technique with a light source that drifts on a level comparable or greater than other sources of noise, the experimenter must characterize the magnitude of the drift using a separate stable detector observing repeated light levels over time. Using the methodology on a real satellite instrument requires careful planning to ensure that the measurements, the model, and the bootstrap algorithm for uncertainty quantification are well-matched.

The probabilistic model and bootstrap algorithm were also tested on an experimental data set from the NIST Beam Conjoiner. The model was deemed to perform well for these data because pointwise 95 % prediction intervals for the residuals enveloped approximately 95 % of them. The probabilistic model was also compared to a conventional technique. The two approaches matched well for estimating the linearization function.

After a further calibration step, for the Beam Conjoiner data set, the component of uncertainty in a calibrated value, attributable to estimation of the linearizing function, was found to be less than 0.025 % for most of the range of the instrument. The method provides unbiased estimates for nonlinear corrections required in high-precision calibrations along with accurate uncertainty statements.

Acknowledgements

The authors thank Robert Rosenberg of NASA Jet Propulsion Laboratory for helpful conversations and feedback. Stephen Maxwell acknowledges the ongoing support of the NIST Greenhouse Gas Measurements Program.

Disclaimers

Certain commercial products or company names are identified here to describe our study adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products or names identified are necessarily the best available for the purpose. The manuscript is a product of the United States Government and not subject to copyright in the United States.

Appendix A. Table of Symbols

Table A1: Table of symbols. If a symbol appears in the *Integrating Sphere* section, it is not repeated in the following sections, e.g., Φ_i. With the exception of symbols appearing in the calibration section, accented symbols, such as $\widehat{\alpha}_0$, are not included because the accent does not change the underlying meaning of the symbol. For $\widehat{\alpha}_0$, the accent designates that it is the MLE of the unaccented symbol. The meaning of all accents are discussed at first use. In the calibration section, some symbols are originally defined with accents.

Integrating Sphere

n_i	Instrument reading for data point $i = 1, \dots, N$.
n	General symbol denoting an instrument reading.
N	Number of instrument readings.
α_m	Legendre polynomial coefficients; $m = 1, \dots, p$.
p	Maximum polynomial degree.
$P_m(\cdot)$	Legendre polynomial of order m.
$s(\cdot)$	Function that shifts and scales inputs to the interval $[-1, 1]$.
Φ_i	Stimulus or flux for data point i.
Φ	General symbol denoting a stimulus or flux.
σ	Standard deviation of noise.
ϕ_j	Flux for lamp $j = 1, \dots, J$.
x_{ij}	Indication of lamp j being 0 (off) or 1 (on) for data point i.
Φ_{max}	Estimate of maximum flux.
τ	Uncertainty in estimated maximum flux expressed as a standard deviation.
γ	Parameter used to force the Legendre polynomial coefficients.
λ	Prior expected value of the parameter used to force the polynomial coefficients.
ψ_k	Scale in $[0, 1]$ for variable aperture setting $k = 1, \dots, N_v$.
$x_{ik}^{(J)}$	Indication of variable aperture state k on lamp J for data point i.
N_v	Number of variable aperture settings.
β_m	Polynomial coefficients defining the linearization function h^{-1}.

Beam Conjoiner

σ_i	Standard deviation of noise for data point i.
κ_0	The cutoff below which noise is assumed constant is $\kappa_0 \Phi_{\text{max}}$.
σ	The standard deviation of the noise is assumed to be proportional to the flux with proportionality constant σ.

Calibration

Φ_{ref}	Reference flux.
$E[n_{\text{ref}}]$	Expected instrument response at Φ_{ref}.
$\widehat{\Phi}_{\text{cal}}$	Calibrated flux (MLE).
Φ^*_{cal}	Calibrated flux (From bootstrap replicate).

References

[1] Joint Committee for Guides in Metrology (JCGM) 2012 *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM)* (International Bureau of Weights and Measures (BIPM)) URL `https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf`

[2] White D R, Clarkson M T, Saunders P and Yoon H W 2008 *Metrologia* **45** 199–210 ISSN 0026-1394 URL `https://doi.org/10.1088%2F0026-1394%2F45%2F2%2F009`

[3] Efron B and Tibshirani R J 1993 *An Introduction to the Bootstrap* (CRC press)

[4] Thompson A K and Chen H M 1994 *J. Res. Natl. Inst. Stand. and Technol.* **99** 751–755

[5] Shin D J, Park S, Jeong K L, Park S N and Lee D H 2013 *Metrologia* **51** 25–32

[6] Legendre and related functions, Chap. 14 in the *NIST Digital Library of Mathematical Functions* http://dlmf.nist.gov/, Release 1.1.3 of 2021-09-15, F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. URL `http://dlmf.nist.gov/14`

[7] Wood S 2015 *Core Statistics* (Cambridge University Press) ISBN 9781107071056

[8] Hoerl A E and Kennard R W 1970 *Technometrics* **12** 55–67

[9] Tibshirani R 1996 *Journal of the Royal Statistical Society: Series B (Methodological)* **58** 267–288

[10] Gelman A, Carlin J, Stern H, Dunson D, Vehtari A and Rubin D 2013 *Bayesian Data Analysis, Third Edition* Chapman & Hall/CRC Texts in Statistical Science (Taylor & Francis) ISBN 9781439840955

[11] Park T and Casella G 2008 *Journal of the American Statistical Association* **103** 681–686

[12] Rosenberg R, Maxwell S, Johnson B C, Chapsky L, Lee R A M and Pollock R 2017 *IEEE Transactions on Geoscience and Remote Sensing* **55** 1994–2006 ISSN 0196-2892

[13] Rosenberg R, Maxwell S, Johnson B C, Chapsky L, Lee R A and Pollock R 2017 *IEEE Transactions on Geoscience and Remote Sensing* **55** 1994–2006

[14] Stoudt S, Pintar A L and Possolo A 2020 *Metrologia* **58**

[15] Chernick M R 2011 *Bootstrap Methods: A Guide for Practitioners and Researchers* (John Wiley & Sons)

[16] Hall P 2013 *The Bootstrap and Edgeworth Expansion* (Springer Science & Business Media)

[17] Eppeldauer G and Hardis J E 1991 *Applied Optics* **30** 3091–3099

[18] James G, Witten D, Hastie T and Tibshirani R 2013 *An Introduction to Statistical Learning: with Applications in R* Springer Texts in Statistics (Springer New York) ISBN 9781461471387

[19] Markham B, Barsi J, Kvaran G, Ong L, Kaita E, Biggar S, Czapla-Myers J, Mishra N and Helder D 2014 *Remote Sensing* **6** 12275–12308 ISSN 2072-4292 URL `https://www.mdpi.com/2072-4292/6/12/12275`