Improving self-calibration

Torsten A. Enßlin, Henrik Junklewitz, Lars Winderling, Maksim Greiner, and Marco Selig
Max-Planck-Institut für Astrophysik, Karl-Schwarzschildstr. 1, 85748 Garching, Germany
and Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 Munich, Germany
(Received 4 December 2013; revised manuscript received 6 September 2014; published 1 October 2014)

Response calibration is the process of inferring how much the measured data depend on the signal one is interested in. It is essential for any quantitative signal estimation on the basis of the data. Here, we investigate self-calibration methods for linear signal measurements and linear dependence of the response on the calibration parameters. The common practice is to augment an external calibration solution using a known reference signal with an internal calibration on the unknown measurement signal itself. Contemporary self-calibration schemes try to find a self-consistent solution for signal and calibration by exploiting redundancies in the measurements. This can be understood in terms of maximizing the joint probability of signal and calibration. However, the full uncertainty structure of this joint probability around its maximum is thereby not taken into account by these schemes. Therefore, better schemes, in sense of minimal square error, can be designed by accounting for asymmetries in the uncertainty of signal and calibration. We argue that at least a systematic correction of the common self-calibration scheme should be applied in many measurement situations in order to properly treat uncertainties of the signal on which one calibrates. Otherwise, the calibration solutions suffer from a systematic bias, which consequently distorts the signal reconstruction. Furthermore, we argue that nonparametric, signal-to-noise filtered calibration should provide more accurate reconstructions than the common bin averages and provide a new, improved self-calibration scheme. We illustrate our findings with a simplistic numerical example.

DOI: 10.1103/PhysRevE.90.043301 PACS number(s): 89.70.Eg, 11.10.−z

I. INTRODUCTION

A. Motivation

Any measurement device needs a proper calibration, otherwise an accurate translation of the raw measurement data into a common system of units is impossible. Our ability to process, combine, communicate, and draw conclusions from the results of measurements depends critically on the achieved calibration accuracy.

The calibration problem is widespread across different fields. Knowing the amplifier gain factors and detector efficiencies of physical measurement apparatuses is necessary to analyze their data. In astronomy, the point spread function of a telescope observation might be unknown since it could depend on varying atmospheric influences. In analyzing sociological questionnaires, the reliability of people's answers might differ from topic to topic, but needs to be taken into account. In all those cases, the measurement response to the quantity of interest, our signal, needs to be known. This response expresses how the data react (on average) to changes in the signal. Only if one knows the response precisely, one can accurately recover the signal of interest correctly from data. The process of the response determination is called *calibration*, its result the *calibration solution*, *calibration reconstruction*, or just *calibration* for brevity.

Several kinds of calibration uncertainties appear in practice: offsets (or additive noise), gain uncertainties (or multiplicative noise), and nonlinearities (e.g., receiver saturation). This work deals with the first two kinds of problems: multiplicative and additive noise. Noise denotes here any influence of the data which is not due to the signal of interest, be it stochastic or just unknown in nature. Nonlinear signal responses complicate the signal inference considerably. If the nonlinearities are known, the generic insights about the calibration of additive and multiplicative noise derived in this work still apply. The calibration of unknown nonlinearities is beyond the scope of this paper, though.

The classical way to calibrate a measurement device is to apply it to a known reference signal, the calibrator. The obtained instrument response to this can then be used to gauge the instrument and to interpret the data obtained from measuring an unknown signal [1–6].

However, in many measurement situations, the response depends strongly on time, location, temperature, energy, frequency, or other dimensions. A simultaneous measurement of both calibrator and signal is often impossible. The external calibration needs then to be extrapolated within the time (space, energy, ...) domain of the signal measurement. Extrapolation in time is only possible if the calibration exhibits sufficient autocorrelation. This autocorrelation could be used to optimally filter out noise in the calibration solution. In practice, however, usually only averaging of the individual calibration solutions within suitably chosen intervals is performed.

A calibration obtained might be further improved by exploiting redundancies in the signal measurement. If the same aspects of the signal are measured repeatedly, but the data show significant deviations between the individual measurements, this indicates a change in the instrument's sensitivity. Thus, an external calibration can often be improved by an internal calibration using the unknown signal itself as an additional source of calibration information.

The usual internal calibration or self-calibration (*self-cal* [7–14]) scheme proceeds as follows. A coarse external calibration is obtained and then applied to the data to get a first signal reconstruction. This signal reconstruction is then used for a refinement of the calibration, which in turn helps to further improve the signal reconstruction. The reconstruction and calibration operations are repeated until some desired convergence criteria are met.

It is, in general, unclear whether such a procedure converges and whether the obtained solution is reasonable. There are *self-cal* schemes derived from minimizing an objective function [6] and convergence can be proven for them. However, for empirically designed *self-cal* schemes, as used, e.g., in radio interferometry, such a proof is often missing. It could well be that only a self-consistent solution of an incorrect signal and an incorrect calibration is obtained, although the joint fit to the data is perfect.

This problem is of generic nature. If a measured datum depends on two unknowns, the signal and the instrument sensitivity, these cannot unambiguously be reconstructed. The additional presence of measurement noise makes this inference problem even harder. External calibration is essential, but often relies on the ability to extrapolate it into domains in time or location, where, strictly speaking, it was not measured for.

In this work, we show that the classical *self-cal* scheme can be understood as a joint maximization of the joint posterior probability of signal and calibration given the data. This posterior represents all available information on signal and calibration. A stable fix point of the *self-cal* scheme is a maximum of this joint posterior. It therefore represents the most likely combination of signal and calibration, at least in some vicinity of the fix point.

However, such maximum *a posteriori* (MAP) estimators are known to be prone to overfitting the data. A posterior mean signal would be optimal with respect to an expected square error norm (e.g., [15]). In case of a symmetric posterior, mean and maximum coincide and the MAP estimator is also optimal in this sense. However, the presence of a nuisance parameter, here being the unknown calibration, can turn an originally symmetric problem into a skewed one. As a consequence, the maximum of such a skewed, nonsymmetric posterior is systematically biased away from the location of the posterior mean (e.g., [16]). Indeed, we will show in this work that using the joint MAP estimator of signal and calibration, as the *self-cal* scheme does, implies a systematical bias with respect to the more optimal posterior mean of signal and calibration.

B. Previous work

The previous work on calibration is vast, in particular the mathematical-statistical literature. It may be classified into whether it deals with univariate or multivariate calibration problems, concentrates on external or internal calibration, and uses frequentist or Bayesian methodologies. A review of various mathematical treatments of external calibration (univariant and multivariant as well as frequentistic and Bayesian) can be found in [3].

External calibration means that an external, high-quality data set is used to map out and reconstruct the response of a measurement device. This could be a single real function (univariate calibration, e.g., [1]) or a vector valued function (multivariate calibration, e.g., [2]). This calibrated response is then used in the interpretation of the following measurements. The main challenge in external calibration is to construct a noise suppressing reconstruction operation, which takes the often unknown noise variance properly into account. The function might be of parametric form [1], or nonparametric estimators might be used [4,17]. It was realized that in many situations, a calibration obtained at some instant is not accurate for subsequent measurements, as the instrument might have changed with time. An appropriate calibration transfer method should be employed that takes such uncertainties properly into account [5].

Internal calibration deals with the situation that an external calibration solution is not available, or is known to be inaccurate. For example, the instrument response might change on time scales comparable to the one needed to switch the instrument to the calibrator signal. This, for example, is a common problem in radio interferometry, where the rapidly changing Earth ionosphere can be regarded as part of the telescope optics. In such cases, the signal of interest has to serve as a calibration signal as well. The resulting *self-cal* schemes image the signal with an assumed calibration, calibrate on this signal reconstruction, and repeat these operations until convergence [7–14]. To the knowledge of the authors, an information theoretical investigation is lacking about under which conditions this leads to reliable results, and when it fails, although practitioners certainly have developed a good intuition on this.

A rigorous information theoretical treatment of the problem of unknown calibration should be built on the calibration marginalized likelihood since it contains all the available information from the data and on the measurement process. For a measurement with Gaussian noise, linear response, and linear calibration uncertainties with a Gaussian distribution of known covariance, this marginal likelihood can be calculated analytically [18] and is reproduced here in Eq. (43). This likelihood is a Gaussian probability density in the data, with a signal dependent covariance. Thus, the resulting signal posterior is very non-Gaussian. If the mean of this signal posterior can be calculated, all the available internal calibration information is taken implicitly into account, and there is no need for a determination of the calibration. In case of nonparametric measurement and calibration problems, the dimensionality of the problem is, however, often too large (virtually unbound) for the usual Monte Carlo methods to sample the posterior. To tackle such and other problems, information field theory [15,19,20] was developed. This exploits the mathematical and conceptual similarities of the nonparametric inference problem with statistical field theories well known in mathematical physics. For example, the reconstruction of Gaussian random fields with unknown covariance, as also needed for calibration, was successfully treated in this framework [16,17,21–23].

For the effective treatment of non-Gaussian posteriors, the method of minimal Gibbs free energy [24] (a thermodynamical incarnation of the variational Bayes approach) has proven to be useful and was applied to the calibration marginalized signal inference problem by Ref. [25]. However, due to the contrived structure of the marginal likelihood, relatively coarse approximations had to be used there to get to analytical formulas. For this reason, a more pragmatic approach shall be followed here in tackling the internal calibration problem.

C. Structure of this work

This work is organized as follows. In order to develop an intuitive understanding, we investigate an illustrative example from a frequentist and Bayesian perspective in Sec. II. Then,

we investigate in Sec. III the general theory of calibration of linear measurements with partly unknown response operators, in particular external calibration, classical *self-cal*, and a new, uncertainty corrected *self-cal* schemes. These different approaches are compared in Sec. IV via a numerical example that is based on the illustrative example of Sec. II. We conclude in Sec. V with a summary of our main findings and a brief perspective on what would be required to develop a full theory of calibration.

II. ILLUSTRATIVE EXAMPLE

It is the goal of this work to improve the present *self-cal* schemes such that the reconstructed signal is closer to the *a posteriori* mean. It turns out that this is a problem of high mathematical complexity even for linear responses. The most important correction we find can be understood intuitively, though. For this, we first turn to a simplistic example, which we investigate using a less formal language. A more general and rigorous treatment will be given in Sec. III, which is able to deal with the complex linear responses one can find in practice, like convolving telescope beams, etc. The illustrative example introduced here will be simulated in Sec. IV and is also the basis of the figures in this article.

A signal s should be observed with an instrument that has a sensitivity or gain g. In our illustrative example, which will be replaced by a more general case later on, the instrument's data

$$d = gs + n \tag{1}$$

is further corrupted by noise n. Here and in the following, any calibration offset in the data is regarded as part of the noise n.

Signal, noise, and gains should be independent stochastic processes so that their joint probability separates according to

$$\mathcal{P}(n, g, s) = \mathcal{P}(n)\,\mathcal{P}(g)\,\mathcal{P}(s). \tag{2}$$

At this stage, the problem of signal and gain reconstruction is symmetrically degenerate since we know as much about the signal as about the gain given the data. Typically, the gain is not completely unknown, for example, the sign of the instrument gain is usually known. For definiteness, let us assume it to be positive and actually $g = 1 + \gamma$, with 1 being the known part of the gain and γ denoting the unknown part that we need to calibrate. This could as well be positive as negative with the same probability. We will refer to any estimate of γ as a *calibration*.

A. Frequentist perspective

In frequentist data analysis, repeated instances of the data are assumed to exist. These permit us to perform data averages that can be tailored towards statistical averages. We adopt for a moment this perspective since it allows us to highlight the essence of the calibration problem.

If we would know the calibration, we could infer the signal by averaging over the data in a way that averages over noise realizations

$$\langle d\rangle_{(n|\gamma,s)} = (1+\gamma)\,s + \underbrace{\langle n\rangle_{(n)}}_{=0}. \tag{3}$$

Here, we assumed the noise to have a zero mean and denote averages over the probability of a given b by

$$\langle f(a)\rangle_{(a|b)} \equiv \int \mathcal{D}a\; f(a)\, P(a|b). \tag{4}$$

Here, $\int \mathcal{D}a$ denotes the phase space integral of a, at the moment a finite dimensional integral like $\int dn$, and later on also path integrals over functional spaces.

In case we do not know the calibration, we could still learn something about the signal, if we are able to average over the data in a way that averages over noise and calibration realizations. This reveals the signal since

$$\langle d\rangle_{(n,\gamma|s)} = (1 + \underbrace{\langle\gamma\rangle_{(\gamma)}}_{=0})\,s + \underbrace{\langle n\rangle_{(n)}}_{=0} = s. \tag{5}$$

This is less sensitive to the signal since we need more data to perform our averaging of two stochastic processes: noise and calibration. But, the point we want to make is that the signal can be estimated from a suitable linear data average, even without knowing the precise calibration, since there is a known and positive part of the response of the data to the signal.

Obtaining information on the unknown calibration γ, which would help us to get the signal more accurately, is more difficult. If we want to perform an analogous averaging to retrieve some information on γ, now over noise and signal realizations, we find

$$\langle d\rangle_{(n,s|\gamma)} = (1+\gamma)\,\underbrace{\langle s\rangle_{(s)}}_{=0} + \underbrace{\langle n\rangle_{(n)}}_{=0} = 0, \tag{6}$$

while assuming a zero mean for the signal as well. Thus, at linear order in the data, there is no calibration information available. We cannot proceed without some knowledge of the signal since the response of the data to our calibration could as well be positive (for $s > 0$) as negative (for $s < 0$). Furthermore, whenever the signal is close to zero, the data respond only poorly to the calibration.

In *self-cal* we obtain some information on the signal from the data, e.g., by using only the known part of the response, which then might be used to analyze the data for a better guess on the calibration. This means the data have to be used at least twice (first a rough signal reconstruction, then calibrating on this) and we end up with a scheme that is at least quadratically in the data. Indeed, if we investigate averages over squared data,

$$\langle d^2\rangle_{(n,s|\gamma)} = \langle s^2\rangle_{(s)}(1+\gamma)^2 + \langle n^2\rangle_{(n)}, \tag{7}$$

we find that this contains terms that are directly sensitive to the calibration and therefore calibration information is available.

It should be noted that the sensitivity of this squared data to the gains depends on the signal (and noise) variance, which we therefore would need to know. Any systematic error in its determination from data will lead to a systematic bias in the calibration. If such a biased calibration is used again for improving the signal variance in an attempt to iteratively improve the calibration solution, the bias is even increased. Without any external calibration constraints, the *self-cal* solution would easily drift far away from an initially

acceptable calibration.[1] Thus, a strong, but self-consistent bias can be present in the results of *self-cal*.

In practice, *self-cal* is rarely done using Eq. (7) as this requires too much data with comparable calibration coefficients for getting reliable averages to measure the calibration from the data variance. More direct and more sensitive calibration methods are used, e.g., the mentioned iteration of reconstruction and calibration steps. The *self-cal* instability exists there as well, in a slightly more subtle form. For the detailed information theoretical development and investigation of such methods, we switch now to a Bayesian perspective.

B. Bayesian perspective

In probabilistic logic (see, e.g., Refs. [26–28]), only a single realization of a data set needs to be available. All reasoning has to be done conditional to these data and averages over different data realizations are not part of the resulting data analysis method. Probabilities express the strength of believe in a certain possibility conditionally that some other statement is assumed to be true and not necessarily how often this possibility happens to be the case as in frequentist thinking. The data are regarded as a vector of values $d = (d_1, \ldots, d_{\mathtt{n}}) \in \mathbb{R}^{\mathtt{n}}$, $\mathtt{n} \in \mathbb{N}$, for which any datum d_i could be the result of a unique, nonreproducible measurement as, e.g., its gain $g_i = 1 + \gamma_i$ probably never takes exactly the same value again.[2]

The measurement equation of our illustrative example is still Eq. (1) if we read it as a vector equation with components

$$d_i = (1 + \gamma_i)\, s_i + n_i. \tag{8}$$

We might want to calculate the signal averaged over all unknowns, but conditioned to the data

$$m = \langle s \rangle_{(n,\gamma,s|d)} \tag{9}$$

since this is known (see, e.g., Ref. [15]) to minimize the expected square error

$$\langle (s - m)^2 \rangle_{(n,\gamma,s|d)}. \tag{10}$$

On linear order in the data, the optimal estimator of this mean is known to be given by (see, e.g., [29])

$$m = \langle s\, d^\dagger \rangle_{(n,\gamma,s)} \langle d\, d^\dagger \rangle^{-1}_{(n,\gamma,s)} d + \mathcal{O}(d^2) \tag{11}$$

with

$$\begin{aligned} \langle s\, d^\dagger \rangle_{(n,\gamma,s)} &= \underbrace{\langle s\, s^\dagger \rangle_{(s)}}_{\equiv S} \text{ and} \\ \langle d_i \overline{d_j} \rangle_{(n,\gamma,s)} &= (1 + \underbrace{\langle \gamma_i \overline{\gamma_j} \rangle_{(\gamma)}}_{\equiv \Gamma_{ij}})\, S_{ij} + \underbrace{\langle n_i \overline{n_j} \rangle_{(n)}}_{\equiv N_{ij}}, \end{aligned} \tag{12}$$

where the bar denotes complex conjugation. Here, we defined the matrices $S = \langle s\, s^\dagger \rangle_{(s)}$, $\Gamma = \langle \gamma\, \gamma^\dagger \rangle_{(\gamma)}$, and $N = \langle n\, n^\dagger \rangle_{(n)}$ that express the *a priori* uncertainty covariances in signal, calibration, and noise, as well as the notation $\dagger$ for the transpose of a vector (and its complex conjugate in case it is a complex number). This optimal linear estimator is known under many names, such as minimal square error (MSE) estimator, generalized Wiener filter [30], and others.

Using matrix notation and defining the component-wise matrix product $(S * \Gamma)_{ij} = S_{ij} \Gamma_{ij}$ (no summation), we get

$$m \approx \underbrace{S\, [S + S * \Gamma + N]^{-1}}_{F}\, d \tag{13}$$

and find that the reconstruction is a filtered version of the data. The filter F reduces the variance since its "denominator" $S + S * \Gamma + N$ is spectrally[3] larger than the "numerator" S. We can write $F < \mathbb{1}$ (spectrally). As larger the noise variance N is with respect to the signal variance S, as stronger the down-weighting of the data. Further down-weighting comes from the combined signal and calibration variation $S * \Gamma$. However, if $N \ll S$ and $\Gamma \ll \mathbb{1}$ (spectrally) the filter is close to the identity $\mathbb{1}$ and the signal estimate Fd is nearly unfiltered data.

In any case, the expected covariance of this reconstruction

$$\langle m m^\dagger \rangle_{(n,s,\gamma)} = S\, [S + S * \Gamma + N]^{-1} S = F\, S < S \tag{14}$$

is (spectrally) smaller than that of the signal S.

Using this linear data filter F [Eq. (13)] is in general not a bad idea since it is a conservative approach to signal reconstruction under calibration uncertainties. It adds the impact of the calibration uncertainty $S * \Gamma$ to the noise budget N of a generalized Wiener filter [30], which is then applied to the data. The disadvantage of this approach is that even in case the signal is so strong that the signal-to-noise ratio is excellent $S \gg N$ (spectrally), the calibration-noise covariance $S * \Gamma$ can still be substantial as it increases with increasing signal strength. For large calibration uncertainties, better, and necessarily nonlinear methods have to be used, since among all possible linear data filters, F is already the optimal one [in the sense of minimizing Eq. (10)].

The optimal (nonlinear) method can be constructed by rewriting (9) as

$$\begin{aligned} m &= \int \mathcal{D}n \int \mathcal{D}\gamma \int \mathcal{D}s\, \mathcal{P}(n,\gamma,s|d)\, s \\ &= \int \mathcal{D}\gamma \int \mathcal{D}s\, \mathcal{P}(\gamma,s|d)\, s \\ &= \int \mathcal{D}\gamma\, \mathcal{P}(\gamma|d) \int \mathcal{D}s\, \mathcal{P}(s|d,\gamma)\, s \\ &= \langle \langle s \rangle_{(s|d,\gamma)} \rangle_{(\gamma|d)}. \end{aligned} \tag{15}$$

Here, we have performed a noise marginalization[4] and have split $\mathcal{P}(s,\gamma|d)$ via the product rule into $\mathcal{P}(s|d,\gamma)\, \mathcal{P}(\gamma|d)$. The inner signal average $\langle s \rangle_{(s|d,\gamma)}$ assumes the calibration

[1] This instability is well known in radioastronomical interferometry. To suppress it, it is common practice to apply *self-cal* only to either the phases of the complex gain coefficients and to keep the gain amplitudes fixed or vice versa.

[2] Repeated measurements or measurements with different instruments can be combined into a single data vector by simple concatenation of the individual data vectors.

[3] Meaning that $\xi^\dagger (S + S * \Gamma + N)\, \xi > \xi^\dagger S\, \xi$ for $\forall\, \xi \in \mathbb{R}^{\mathtt{n}} \backslash 0$.

[4] This marginalization is trivial since the term $\mathcal{P}(d|n,\gamma,s) = \delta[d - (1 + \gamma)\, s - n]$ can be obtained using Bayes theorem, which cancels the noise phase space integral $\int \mathcal{D}n$.

to be known. However, the outer average goes over the unknown calibration while weighting each possible calibration according to its posterior probability given the data $\mathcal{P}(\gamma|d)$.

The inner signal average might in many situations be well dealt with by using the optimal linear estimator

$$\begin{aligned}\langle s\rangle_{(s|d,\gamma)} &\approx \langle s\, d^\dagger\rangle_{(n,s|\gamma)}\langle d\, d^\dagger\rangle^{-1}_{(n,s|\gamma)} d \\ &= \underbrace{S\, R^\dagger [R\, S\, R^\dagger + N]^{-1}}_{\equiv W} d, \end{aligned} \tag{16}$$

where $R = \mathrm{diag}(1+\gamma)$ is the calibration dependent response matrix of our measurement. The previously problematic signal suppression by the term $S * \Gamma$ in Eq. (13) became more specific since $R\, S\, R^\dagger = S + S * (\gamma\, \gamma^\dagger)$ and therefore $\Gamma \to \gamma\, \gamma^\dagger$. We therefore expect to obtain a higher fidelity signal recovery even when the subsequently applied calibration averaging in Eq. (15) might smooth out some of the features present in $\langle s\rangle_{(s|d,\gamma)}$ as the posterior average $\langle \gamma\, \gamma^\dagger\rangle_{(\gamma|d)}$ implies much less averaging than the prior average $\Gamma = \langle \gamma\, \gamma^\dagger\rangle_{(\gamma)}$.

It is common practice to use a single "best" calibration solution $\gamma^\star$, a so-called *point estimate*, instead of averaging over all possible calibrations. Thus, implicitly $\mathcal{P}(\gamma|d) \approx \delta(\gamma - \gamma^\star)$ is assumed. This is indeed often a good approximation, as we will argue in Sec. III G. The next order corrections that take into account the width of the distribution $\mathcal{P}(\gamma|d)$ are usually small. An imperfectly chosen $\gamma^\star$ has typically a larger impact on the reconstruction quality than these corrections and our focus should, therefore, be on how to calibrate most reliably.

Again, we regard the posterior mean as a good estimate and choose it as a starting point

$$\gamma^\star = \langle\gamma\rangle_{(s,\gamma|d)} = \langle\langle\gamma\rangle_{(\gamma|d,s)}\rangle_{(s|d)}. \tag{17}$$

If we would know the signal close enough, calibration would be simple, as we could ignore the outer averaging over $\mathcal{P}(s|d)$. We would form signal subtracted data $d' = d - s = \gamma\, s + n$ from which we could construct the optimal linear estimator of the calibration

$$\begin{aligned}\gamma^\star &\approx \langle\gamma\rangle_{(\gamma|d',s)} \approx \langle\gamma\, d'^\dagger\rangle_{(n,\gamma|s)}\langle d' d'^\dagger\rangle^{-1}_{(n,\gamma|s)} d' \\ &= \Gamma\, R'^\dagger [R'\, \Gamma\, R'^\dagger + N]^{-1}(d - s), \end{aligned} \tag{18}$$

with $R' = \mathrm{diag}(s)$ being the response of the data to the calibration parameters γ.

Iterating the linear estimators for signal and calibration [(16) and (18)], while assuming $s = Wd$ and $\gamma = \gamma^\star$, is then a plausible *self-cal* scheme. It ignores, however, the uncertainties in signal reconstruction and calibration and therefore might suffer from a bias similar to the one discussed before using frequentist arguments. In particular, the outer averaging in Eq. (17) over $\mathcal{P}(s|d)$ is crucial. If we ignore for a moment, for simplicity, the signal dependence of the "denominator" in Eq. (18), we see that our calibration estimator

$$\begin{aligned}\gamma^\star &= \langle \Gamma\, \underline{R'^\dagger}[\ldots]^{-1}(d - \underline{s})\rangle_{(s|d)} \\ &\approx \mathcal{O}(\langle s\rangle_{(s|d)}) - \mathcal{O}(\langle s\, s^\dagger\rangle_{(s|d)}) \end{aligned} \tag{19}$$

requires the knowledge of the *a posteriori* signal mean $m = \langle s\rangle_{(s|d)}$ and variance $\langle s\, s^\dagger\rangle_{(s|d)}$ since the two underlined "numerator" terms both contain the unknown signal. In classical *self-cal* schemes $\langle s\, s^\dagger\rangle_{(s|d)}$ is approximated by $mm^\dagger$. The latter has, however, less variance than the former if m is a filtered version of s as assumed here.[5] This means that a systematic bias is present in such schemes, as the $\mathcal{O}(\langle s\, s^\dagger\rangle_{(s|d)})$ term is systematically underestimated by $mm^\dagger$ leading to an overestimation of $\gamma^\star$. The most important result of this work is to show how to correct for this bias.

Such a bias was not present in case of the signal estimation using a point estimator $\gamma^\star$ for the calibration (instead of the calibration averaging). The difference lies in the fact that for the chosen illustrative data model [Eq. (8)] (as well as for many realistic measurement situations), the symmetry of signal and gain as suggested by Eq. (1) is broken since s varies around zero and g around a known nonzero value. Thus, signal estimators can be built on a more reliably nonzero gain than calibration estimators, which have to exploit opportunistically any sufficiently nonzero signal fluctuation suitable for calibration.

III. THEORY OF CALIBRATION

A. Generic problem

A more rigorous and more abstract treatment of the calibration problem should be addressed now. The signal and data domain are not necessarily the same anymore as signals live typically in continuous domains (time, position, or spectral spaces) and data sets are always finite. For dealing with probabilities over spaces of continuous functions (fields in physical language), we use the formalism of information field theory [15,19,20].

A generic, linear response that maps the signal into data domain will be assumed. This covers many realistic measurement situations. Unknown properties of this response are to be calibrated. The unknown signal, calibration, and noise components are all assumed to fluctuate around zero with known individual covariances, but no cross correlations between them. As we do not assume any higher order statistics of these components to be known, the maximum entropy criterion [27,28,31,32] suggests we should model our *a priori* knowledge states as Gaussian distributions. This does not imply that our analysis is only valid for Gaussian statistics. If signal, noise, or calibration follow non-Gaussian distributions and those are known, the here derived methods still produce sensible results. Just more efficient methods might be constructed that exploit the additional statistical knowledge.

We assume that a signal $s = (s_x)_x$ over some continuous domain (parametrized by x) was targeted by a linear measurement device that produced the finite dimensional data $d = (d_i)_i$ with signal independent Gaussian noise $n = (n_i)_i$ that includes also calibration offsets

$$d = Rs + n. \tag{20}$$

The signal response $R = R^\gamma = (R^\gamma_{i\,x})_{i\,x}$ depends on the unknown calibration parameters $\gamma = (\gamma_a)_a$ as well as on the signal and data domain coordinates, here x and i, between

[5] See Eq. (14), which is valid also here if we set $S * \Gamma \to 0$ there. If m resulted from naive data averaging, noise remnants might be significantly present and can lead to an overestimation of the posterior signal variance and therefore also to a systematically biased calibration.

which it translates via $(R^\gamma s)_i = \int dx\, R^\gamma_{i\,x} s_x$. This is the general form for any linear signal response. It not only embraces the illustrative example of the previous section, where $R^\gamma_{ix} = \delta(i-x)(1+\gamma_i)$, but also a convolution with a calibration dependent kernel $R^\gamma_{ix} = f(i-x,\gamma)$, Fourier transformations $R^\gamma_{kx} = \exp(ikx)$, and more complex measurement situations.

In general, the response can depend in a very complicated way on the unknown parameters γ. We suppose that a first order Taylor expansion captures the most relevant dependence

$$R^\gamma = R^0 + \sum_a \gamma_a R^\gamma_{,\gamma_a}|_{\gamma=0} + \mathcal{O}(\gamma^2), \tag{21}$$

with $R^0 = R^\gamma|_{\gamma=0}$ being the well calibrated part of the response and $R^\gamma_{,\gamma_a} = \partial R^\gamma/\partial\gamma_a$ its linear dependence on the calibration parameter. Thereby, we ignore second order corrections in γ.

To have a compact notation, we define scalar products for the continuous u-dimensional signal domain and its Fourier space as

$$j^\dagger s = \int dx^u\, \overline{j_x} s_x = \int \frac{dk^u}{(2\pi)^u}\, \overline{j_k} s_k \tag{22}$$

for the discrete data domain as

$$n^\dagger d = \sum_i \overline{n_i}\, d_i, \tag{23}$$

and for the calibration parameter domain something analog to (23) or (22), depending on whether the calibration parameters form a discrete set or a continuous function. Discrete calibration parameters are instrument gains since there are at most a finite number of parameters per data value, so that the calibration domain can be mapped onto the data domain (the set of data indices). A continuous set of calibration parameters would be the spatial sensitivity map of a telescope, the so-called telescope beam, for which the domain in which the calibration parameters reside (the sphere $\mathcal{S}^2$ of directions in the telescope frame) can often be mapped onto the signal domain (positions in the sky, also $\mathcal{S}^2$).

In order to have an illustrative case, we assume further that the signal obeys *a priori* a Gaussian distribution

$$\mathcal{P}(s) = \mathcal{G}(s,S) \equiv \frac{1}{|2\pi S|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} s^\dagger S^{-1} s\right), \tag{24}$$

with known covariance $S = \langle s\, s^\dagger\rangle_{(s)} = \int \mathcal{D}s\, s\, s^\dagger\, \mathcal{P}(s)$. This and other covariances are assumed here to be known either from similar previous measurements or on theoretical grounds. In practice, they might need to be determined from the data themselves. This is often well possible, as shown in Refs. [16,21,23,24], and explained in Sec. III C. The extension to non-Gaussian cases can be treated in future studies along the lines sketched in Refs. [15,16,21,24].

The noise covariance $N = \langle n\, n^\dagger\rangle_{(n)}$ is assumed to be known as well, leading to the likelihood

$$\mathcal{P}(d|s,\gamma) = \mathcal{P}(n = d - R^\gamma s|s) = \mathcal{G}(d - R^\gamma s, N). \tag{25}$$

Likelihood and prior can be combined into the joint probability of data and signal $\mathcal{P}(d,s|\gamma) = \mathcal{P}(d|s,\gamma)\,\mathcal{P}(s|\gamma) = \mathcal{P}(d|s,\gamma)\,\mathcal{P}(s)$ [see Eq. (2) for the last step], from which the signal posterior for known calibration can be obtained via Bayes theorem,

$$\mathcal{P}(s|d,\gamma) = \frac{\mathcal{P}(d,s|\gamma)}{\mathcal{P}(d|\gamma)} = \frac{e^{-\mathcal{H}(d,s|\gamma)}}{\mathcal{Z}(d|\gamma)}.$$

Here, we have introduced the information Hamiltonian $\mathcal{H}(d,s|\gamma) \equiv -\ln \mathcal{P}(d,s|\gamma)$, and its partition function

$$\mathcal{Z}(d|\gamma) \equiv \int \mathcal{D}s\, e^{-\mathcal{H}(d,s|\gamma)} = \int \mathcal{D}s\, \mathcal{P}(d,s|\gamma) = \mathcal{P}(d|\gamma), \tag{26}$$

in order to exploit the mathematical and conceptual analogies of Bayesian inference and thermodynamics.

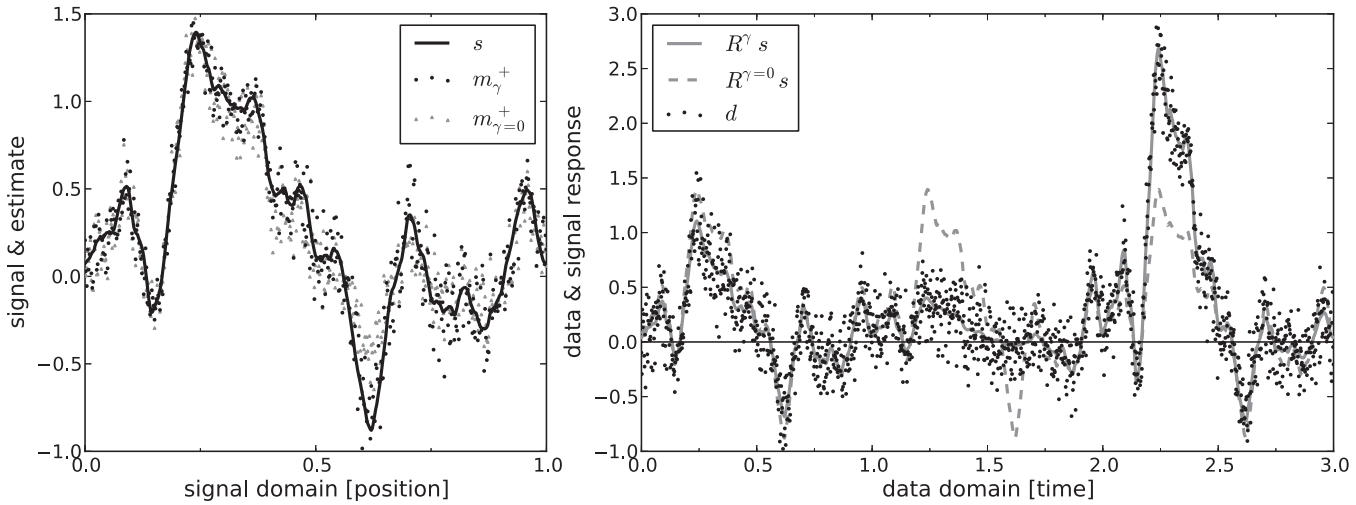

FIG. 1. Simulated signal [according to Eqs. (24) and (58)] and data realization from observing the signal three times [according to Eqs. (20), (25) and (57)]. Left: Signal (line), its prior-free and therefore noisy estimation by Eq. (32) using the correct calibration (γ, black dots), and no calibration ($\gamma = 0$, gray triangles). Right: Data (dots), the signal response $R^\gamma s$ [gray line, Eqs. (57) and (58)] and the response in case of zero calibration $R^{\gamma=0}s$ (dashed, gray line, basically three repetitions of the signal pattern). The difference between these two lines contains information on the calibration. The corresponding gain curve is shown in Fig. 2.

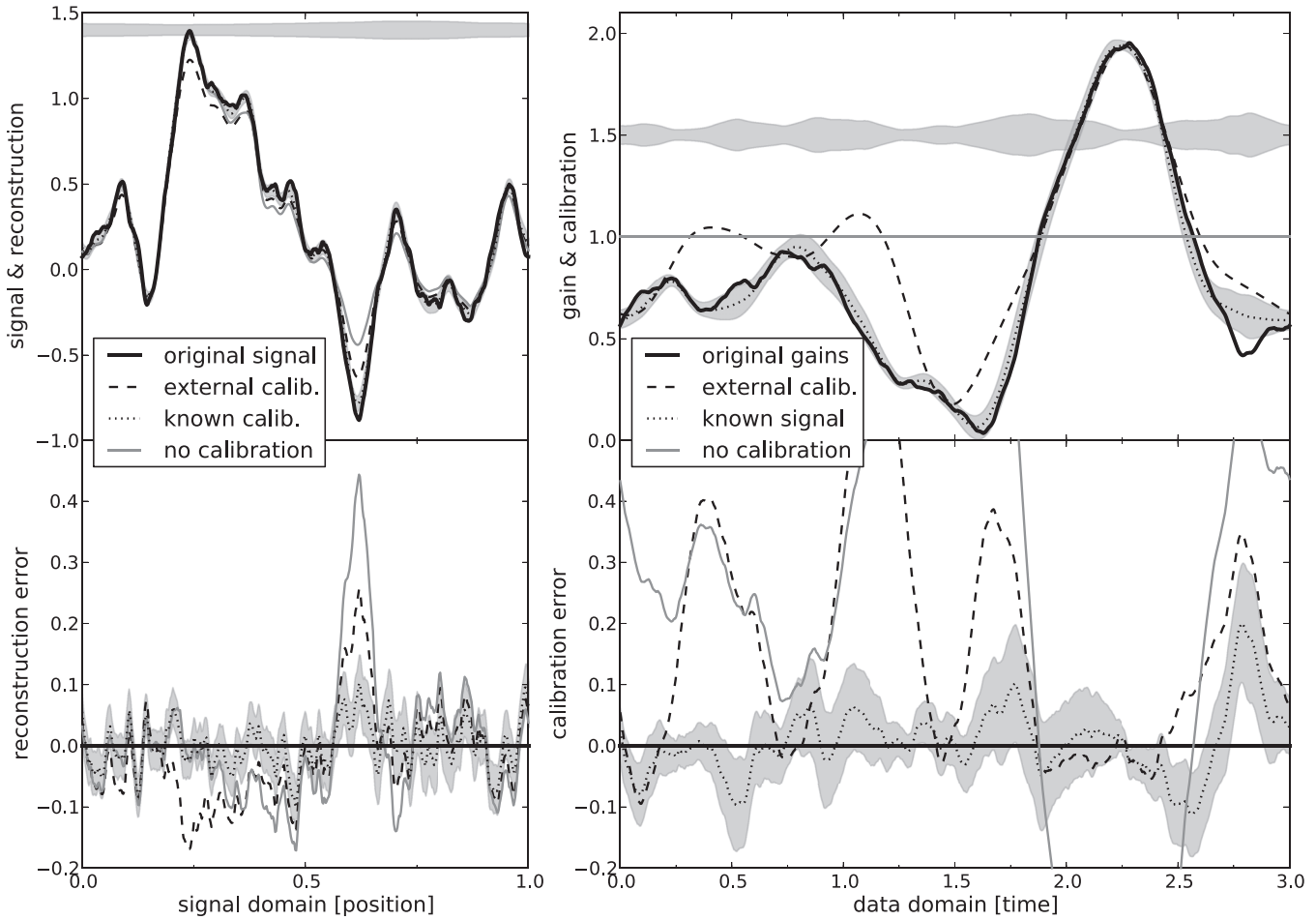

FIG. 2. Signal reconstruction and calibration without *self-cal*. Left: Original signal (as in Fig. 1), and signal reconstructions using only the external calibration data [according to Eq. (39) with external calibrator c at only four moments as described in Sec. IV], the correct gains [$g = 1 + \gamma$, Eq. (27)], and no calibration [$g = 1, \gamma = 0$ in Eq. (27)]. Right: Original gains [according to Eqs. (35) and (58)] and their calibration reconstruction using only the external calibration data (as on the right-hand side), calibrating on the correct signal [Eq. (39) with $c = s$], and assuming no calibration ($g = 1, \gamma = 0$). The gray areas in the left and right panels show the one sigma posterior uncertainties of the signal and calibration reconstructions using the correct calibration and signal, respectively. These are the accuracies of the best achievable reconstructions and show that recovering the calibration accurately is more difficult than recovering the signal. In the top panels, these uncertainties are shown twice, once around the signal/calibration reconstructions and once at an arbitrary location for better visual inspection of their structures.

B. Wiener filter

Under these conditions, the optimal signal reconstruction for a given calibration γ is known to be the Wiener filter (e.g., see Ref. [15])

$$m^\gamma = \langle s \rangle_{(s|d\,\gamma)} = D^\gamma\, j^\gamma, \tag{27}$$

where

$$D^\gamma = (S^{-1} + R^{\gamma\dagger} N^{-1} R^\gamma)^{-1}, \tag{28}$$

$$j^\gamma = R^{\gamma\dagger} N^{-1} d \tag{29}$$

are the information propagator (or Wiener variance) and information source, respectively [15]. This formula is equivalent to the data space centric formula for Wiener filtering [see Eq. (16)] we had argued to be the optimal linear filter [minimizing Eq. (10)]. The remaining *a posteriori* uncertainty of the signal is given by the Wiener variance

$$\langle (s - m^\gamma)(s - m^\gamma)^\dagger \rangle_{(s|d,\,\gamma)} = D^\gamma. \tag{30}$$

Since the signal posterior for known calibration is a Gaussian (for this case composed of a Gaussian prior and likelihood, and a linear response), it must be

$$\mathcal{P}(s|d, \gamma) = \mathcal{G}(s - m^\gamma, D^\gamma), \tag{31}$$

as can also be verified by a direct calculation.

The often used so-called *prior-free* or maximum likelihood reconstruction can as well be reproduced by taking the limit of $S \to \infty$ or $S^{-1} \to 0$, which removes any prior contribution to the filter formula, and interpreting the matrix inversion in Eq. (28) as a pseudoinverse,[6] so that

$$m^+_\gamma = (R^{\gamma\dagger} N^{-1} R^\gamma)^+ R^{\gamma\dagger} N^{-1} d. \tag{32}$$

[6]We define the pseudoinverse of a Hermitian matrix $A = \sum_i a_i a_i^\dagger \lambda_i$ with eigenvalues λ_i and normalized eigenvectors a_i as

$$A^+ = \sum_i a_i a_i^\dagger \begin{cases} \lambda_i^{-1}, & \lambda_i \neq 0 \\ 0, & \lambda_i = 0. \end{cases}$$

This prior-free signal estimator is very noisy, as can be seen from Fig. 1. There, a simulated signal, the resulting data, and the corresponding prior-free signal estimator are shown. The latter exhibits a lot of noise[7] compared to the reconstructions exploiting the knowledge on covariances shown in Fig. 2.

In the following, we often suppress the γ dependence of R, D, j, m and other quantities for notational compactness, as we also do not note explicitly that m is a function of the data d.

C. Power spectrum estimation

A problem in setting up the Wiener filter is often that the signal and noise covariances are not known precisely or might even be completely unknown. Thus, these need to be inferred from the same data used for imaging. The proper way is to formulate hyperpriors on these spectra, and to solve the combined problem of simultaneous signal and spectra recovery. A suitable, however, numerically expensive method for this is Gibbs sampling [33,34], here introduced in Sec. III F.

An approximative, but numerically cheaper approach of iteratively analyzing a reconstruction for its covariance and using this covariance in improved reconstructions was developed in [16,21,23,24]. The basic idea works for a statistical stationary signal, for which the signal covariance is diagonal in Fourier space, with the power spectrum $P_s(k) = \langle |s_k|^2 \rangle_{(s)}$ on the diagonal [here, quantities with the index k denote Fourier transformed quantities such as $s_k = \int dx\, s_x \exp(ikx)$]. In case of a Jeffreys prior on the power spectra, a uniform distribution on logarithmic scale, the formula to get a point estimate for the spectrum is

$$P_s(k) \approx |m_k|^2 + D_{kk}, \tag{33}$$

where D_{kk} corrects for the missing variance in the Wiener filter reconstruction m. Equation (27) for m, Eq. (28) for D, and Eq. (33) for $P_s(k)$ have to be iterated until convergence. The accuracy of this spectral estimate can be improved by averaging Fourier modes with similar spectrum and by exploiting available prior information on the spectral values and their smoothness as a function of wave vector [16,23]. The method can even be extended to estimate simultaneously the signal and noise covariance [21] and be combined with nonlinear signal estimators [24,35,36].

Thus, an unknown covariance can be dealt with in principle. In order to be able to concentrate on the essentials of the calibration problem, we assume in the following known covariances as well as Gaussian prior distributions for signal, noise, and unknown calibration parameters.

Although we showed that methods exist to obtain estimates of these covariances from the data themselves, we should investigate how sensitive a reconstruction is to inaccuracies in those estimates. For this, we consider the special case of a signal and data space being identical, the response being the identity matrix, and signal and noise being statistically homogeneous processes. In this case, their covariances are diagonal in Fourier space, with the corresponding power spectra $P_s(k)$ and $P_n(k)$ on the diagonals. The Wiener filter in Fourier space is then

$$m_k = \frac{d_k}{1 + P_n(k)/P_s(k)}. \tag{34}$$

Thus, high signal-to-noise (S/N) modes with $P_n(k)/P_s(k) \ll 1$ are unmodified by the filter $m_k \approx d_k$, whereas low S/N modes with $P_n(k)/P_s(k) \gg 1$ are strongly suppressed by the filter $m_k \to 0$. Only for S/N ratios around one, the precise value of the spectra matters. Overestimation or underestimation of the S/N ratio leads to too much noise in the reconstruction or an unnecessary strong signal suppression, respectively. However, this effect is mainly relevant for the modes with a S/N around unity. Therefore, moderate inaccuracies in the power spectra or covariances lead only to a minor degradation of the reconstruction fidelity [16]. This usually also holds for more complex measurement situations than used in this argumentation [22].

D. External calibration

Somehow, the calibration parameters γ have to be measured in order that Eq. (27) can be used to determine the *a posteriori* mean of the signal. The simplest strategy is to use an external calibrator signal as a known reference from which the calibration can be determined.

In case the calibration parameters are constant in time, they can be determined using a known calibration signal c and then be transferred to the measurement of interest. The calibrator $c = (c_x)_x$ is just a signal, which ideally is known before the measurement, which is strong enough to have a signal response $R^\gamma c$ dominating over the noise, and which is sufficiently complex to probe the relevant calibration uncertainties. The last requirement means that the calibrator response should depend on the calibration parameters such that $\partial R^\gamma c/\partial \gamma_a \equiv R^\gamma_{,\gamma_a} c$ is significantly nonzero for any relevant γ_a.

For the calibration parameters we assume here and in the following a Gaussian, zero-centered prior[8]

$$\mathcal{P}(\gamma) = \mathcal{G}(\gamma, \Gamma), \tag{35}$$

with known uncertainty covariance $\Gamma = \langle \gamma\, \gamma^\dagger \rangle_{(\gamma)}$. The knowledge on Γ can come from theoretical considerations, previous measurements, or might be obtained from the data themselves. Prior and likelihood [Eq. (25)] form the joint probability

[7] This noise could be reduced by binning, averaging, or smoothing. This requires that an averaging length scale has to be specified. The optimal averaging length scale should be a tradeoff between suppressing noise and keeping signal features. However, the Wiener filter [see Eq. (27)] performs already this averaging in an optimal way [minimizing Eq. (10)] with an averaging length that depends on the local signal-response-to-noise ratio and therefore can vary with position. We therefore use in the following the Wiener filter method and regard binning and averaging scheme applied in practice as approximative realizations thereof.

[8] The mean can always be subtracted by a redefinition of the calibration parameters. As it is known, it should be part of the known part of the response, whereas the calibration parameter should only affect the unknown part.

$\mathcal{P}(d,\gamma|\,c) = \mathcal{P}(d|\gamma,c)\,\mathcal{P}(\gamma)$ that contains all available information on the calibration.

In general, the calibration inference from this is a nonlinear and nontrivial problem. In many cases, the MAP approximation provides a reasonable estimate for γ. This is obtained by minimizing the corresponding Hamiltonian

$$\begin{aligned}\mathcal{H}(d,\gamma|c) &= -\ln\mathcal{P}(d,\gamma|c)\\ &= \tfrac{1}{2}\gamma^\dagger\Gamma\gamma + \tfrac{1}{2}(d-R^\gamma c)^\dagger N^{-1}(d-R^\gamma c)\\ &\quad + \text{const.}\end{aligned} \tag{36}$$

The gradient of this Hamiltonian,

$$\frac{\partial\mathcal{H}(d,\gamma|c)}{\partial\gamma} = \Gamma^{-1}\gamma - c^\dagger R^{\gamma\dagger}_{,\gamma} N^{-1}(d-R\,c), \tag{37}$$

should then be followed (downhill) until it is zero and the Hamiltonian minimal. Here, $(R^\gamma_{,\gamma})_a \equiv R_{,a} \equiv \partial R^\gamma/\partial\gamma_a$ denotes the derivative of the response with respect to the calibration. It is apparent that the discrepancy of the data from the calibration signal response $d - R^\gamma c$ drives the calibration solution away from the default value $\gamma = 0$ preferred by the prior of the calibration $\mathcal{P}(\gamma) = \mathcal{G}(\gamma,\Gamma)$.

In case the calibration parameters enter only linearly,

$$R^\gamma = B^0 + \sum_a \gamma_a B^a, \tag{38}$$

with B^0 and B^a known and γ independent, we have $R_{,a} = R^\gamma_{,\gamma_a} = B^a$ and the minimum of the Hamiltonian is at

$$\begin{aligned}\gamma^\star &= \Delta\,h, \text{ with}\\ \Delta^{-1}_{ab} &= \Gamma^{-1}_{ab} + c^\dagger B^{a\dagger} N^{-1} B^b c,\\ h_b &= c^\dagger B^{b\dagger} N^{-1}(d - B^0 c).\end{aligned} \tag{39}$$

This MAP estimator for the calibration $\gamma^\star$ is actually also the calibration posterior mean $\langle\gamma\rangle_{(\gamma|d,\,c)}$ since this particular posterior is a Gaussian for which mean and maximum coincide. This Gaussian calibration posterior is

$$\mathcal{P}(\gamma|d,c) = \mathcal{G}(\gamma-\gamma^\star,\Delta), \tag{40}$$

with the uncertainty covariance $\Delta = \langle(\gamma-\gamma^\star)(\gamma-\gamma^\star)^\dagger\rangle_{(\gamma|d,\,c)}$ given in Eq. (39).

In this specific linear calibration case, external calibration is Wiener filtering. This can be seen by comparing Eq. (39) with the Wiener filter equations for the signal [Eqs. (27) and (28)], while recognizing that the roles of the following terms correspond to each other: $\gamma^\star \leftrightarrow m$, $\Gamma \leftrightarrow S$, $N \leftrightarrow N$, $B^0 c \leftrightarrow R$, $\Delta \leftrightarrow D$, $d - B^0 c \leftrightarrow d$, and $h \leftrightarrow j$.

There is, however, an interesting difference. The signal information source $j = R^\dagger N^{-1} d \approx R^\dagger N^{-1} R\,s = [B^0 N^{-1} B^0 + \mathcal{O}(\gamma)]s$ contains a calibration-independent term $B^0 N^{-1} B^0\,s$, which reacts to s even when $\gamma = 0$, whereas the calibration information source $h_a \approx c^\dagger B^{a\dagger} N^{-1}(R - B^0)\,c = \sum_b (c^\dagger B^{a\dagger} N^{-1} B^b\,c)\,\gamma_b = \sum_a Q^{ab}\,\gamma_b$ is a quadratic function of the calibration signal c, which vanishes for locations with vanishing c. The quadratic dependence of $Q^{ab} = c^\dagger B^{a\dagger} N^{-1} B^b\,c$ on the calibration signal strength will become important again later on, when we investigate *self-cal*, the attempt to calibrate on an unknown signal.

E. Calibration binning

It should be noted that the usage of an *a priori* calibration covariance $\Gamma = \langle\gamma\gamma^\dagger\rangle_{(\gamma)}$ to suppress the calibration estimation noise is not standard practice. Instead, bin averaging and interpolation is often performed on χ^2 or maximum likelihood calibration estimators.

There is, however, no consensus on the question as to how to choose the bin size and interpolation scheme. The optimal bin size should, on the one hand, be sufficiently large to average down the noise, and on the other hand, be sufficiently small in order not to iron out existing small scale (spatial or temporal) variations in the gain parameters. Therefore, the optimal bin choice depends on the interplay of expected calibration variations as encoded in Γ, the noise level N, and the strength of the calibrator signal in data space $R^\gamma\,c$. Since all these elements are part of the MAP gain estimator [cf. Eq. (37)] that reduces to the Wiener filter solution for linear calibration problems [Eq. (39)], we expect the latter to implement (nearly) an optimal averaging and interpolation scheme. The optimal bin size could be read off from this scheme (it should be of the order of the correlation length of Δ), or, even better, the binning and averaging be replaced with the more accurate calibration solution given by Eq. (37) or (39).

In the following, we use the unbinned, nonparametric Wiener filter solution since it is optimal or close to optimal. We believe that binning schemes used in practice and chosen with experience can come sufficiently close to the Wiener filter performance as that the difference does not matter much for our discussion. When they matter, an adoption of the here proposed nonparametric Wiener filter calibration methodology would be beneficial and highly recommended for the application.

F. Gibbs sampling

The signal and the calibration are the two unknowns. Their joint posterior probability distribution $\mathcal{P}(s,\gamma|d)$ can be probed via Gibbs sampling in case it is possible to draw samples from $\mathcal{P}(s|d,\gamma)$ and $\mathcal{P}(\gamma|d,s)$ [34,37]. These are Gaussian distributions in our case, given by Eqs. (31) and (40) (with $c = s$), respectively, from which it is well possible to draw samples. The Gibbs sampling procedure is then to update a combined signal and calibration probe $p^{(i)} = (s^{(i)},\gamma^{(i)}) \to p^{(i+1)} = (s^{(i+1)},\gamma^{(i+1)})$ via

$$\begin{aligned}s^{(i+1)} &\hookleftarrow \mathcal{P}(s^{(i+1)}|d,\gamma^{(i)}),\\ \gamma^{(i+1)} &\hookleftarrow \mathcal{P}(\gamma^{(i+1)}|d,s^{(i+1)}).\end{aligned} \tag{41}$$

If this updating is an ergodic process for the combined p space, as it is in our case of Gaussian probabilities, the sample distribution can be shown to converge towards $\mathcal{P}(s,\gamma|d)$.

Marginalization with respect to s or γ to obtain $\mathcal{P}(\gamma|d)$ and $\mathcal{P}(s|d)$, respectively, can be obtained from the samples by forgetting the corresponding marginal variable. Any posterior average, such as $\langle s\rangle_{(s|d)}$, is given by the corresponding sample averages. The Gibbs sampling provides therefore a route to calculate any desired estimate from the full posterior, without invoking approximations, except for replacing the posterior integration by finite sampling and therefore getting some shot noise. Beating down this shot noise by generating a large number of samples can become computationally expensive,

why it makes sense also to investigate analytical alternatives as we do in the following. Analytical investigations also provide deeper insight into the structure of the problem, which is less easily obtained from the sampling machinery.

Anyhow, we provide Gibbs sampling results as an optimal benchmark for the different *self-cal* schemes implemented in Sec. IV.

G. Calibration marginalized imaging

All relevant information on the signal is contained in the calibration marginalized posterior

$$\mathcal{P}(s|d) = \int \mathcal{D}\gamma\, P(s,\gamma|d). \tag{42}$$

In the case of linear calibration coefficients $R^\gamma = B^0 + \sum_a \gamma_a B^a$ [see Eq. (38)], the calibration marginalized likelihood, from which this posterior can be constructed, can be calculated analytically:

$$\begin{aligned}\mathcal{P}(d|s) &= \int \mathcal{D}\gamma\, \mathcal{P}(d|s,\gamma)\,\mathcal{P}(\gamma)\\ &= \int \mathcal{D}\gamma\, \mathcal{G}(d - R^\gamma s, N)\,\mathcal{G}(\gamma,\Gamma)\\ &= \mathcal{G}\left(d - B^0 s, N + \sum_{ab} B^a s \Gamma_{ab} s^\dagger B^{b\dagger}\right).\end{aligned} \tag{43}$$

This result can be found in Ref. [18].[9]

The resulting posterior $\mathcal{P}(s|d) = \mathcal{P}(d|s)\,\mathcal{P}(s)/\mathcal{P}(d)$ is non-Gaussian, as the signal field appears as part of the calibration marginalized effective noise $N + \sum_{ab} B^a s \Gamma_{ab} s^\dagger B^{b\dagger}$. Ideally, the mean of this signal posterior is calculated since this gives the optimal signal estimate.

However, integrating over this non-Gaussian function is often infeasible. The quadratic dependence of the effective noise on the unknown signal inhibits that this can be calculated via a simple Gaussian integration. In high-dimensional settings, Monte Carlo methods used to estimate phase space integrals might become too expensive. In such cases, approximative strategies are needed. One is to use the MAP estimator for this posterior. However, due to the skewness of the distribution, this can be expected to give biased results. It is better to characterize the calibration posterior $\mathcal{P}(\gamma|d)$ by its mean $\gamma^\star$ and uncertainty covariance Δ and to use them to construct an approximative signal estimation.

Let us assume we managed somehow to estimate the calibration as $\gamma^\star$ with some uncertainty covariance Δ and that we can well approximate[10]

$$\mathcal{P}(\gamma|d) \approx \mathcal{G}(\gamma - \gamma^\star, \Delta). \tag{44}$$

For a Gaussian signal field with $\mathcal{P}(s) = \mathcal{G}(s, S)$ we could simply use $\gamma = \gamma^\star$ in the Wiener filter formula (27). However, this is suboptimal if the calibration uncertainty is significant. In that case, correction terms might become important, which we calculate now to first order in the calibration uncertainty Δ.

The optimal, calibration marginalized signal estimator is

$$\begin{aligned} m = \langle s\rangle_{(s|d)} &= \langle\langle s\rangle_{(s|d,\,\gamma)}\rangle_{(\gamma|d)} = \int \mathcal{D}\gamma\, \mathcal{P}(\gamma|d)\,\langle s\rangle_{(s|d,\,\gamma)}\\ &\approx \int \mathcal{D}\gamma\, \mathcal{G}(\gamma - \gamma^\star, \Delta)\, m^\gamma\\ &\approx D\,\Big\{ j + \frac{1}{2}\sum_{ab}[\Delta_{ab}\,(j_{,ba} - M_{,ba}\,D\,j\\ &\qquad + 2\,M_{,b}\,D\,M_{,a}\,D\,j - 2\,M_{,b}\,D\,j_{,a})]\Big\}_{\gamma=\gamma^\star}. \end{aligned} \tag{45}$$

In the last step, we Taylor expanded $m^\gamma = D^\gamma\, j^\gamma$ up to second order in $\gamma - \gamma^\star$, performed the Gaussian integration, exploited the Hermitian symmetry of Δ, suppressed in the notation the dependence of all calibration dependent terms on γ, and introduced further the notations

$$\begin{aligned} M &= R^\dagger N^{-1} R, \quad M_{,a} = R^\dagger_{,\gamma_a} N^{-1} R + R^\dagger N^{-1} R_{,\gamma_a},\\ M_{,ab} &= R^\dagger_{,\gamma_a} N^{-1} R_{,\gamma_b} + R^\dagger_{,\gamma_b} N^{-1} R_{,\gamma_a}\\ &\quad + R^\dagger_{,\gamma_a\gamma_b} N^{-1} R + R^\dagger N^{-1} R_{,\gamma_a\gamma_b},\\ j &= R^\dagger N^{-1} d,\\ j_{,a} &= R^\dagger_{,\gamma_a} N^{-1} d,\\ j_{,ab} &= R^\dagger_{,\gamma_a\gamma_b} N^{-1} d. \end{aligned} \tag{46}$$

In case of only linear calibration parameters as in Eq. (38), $R = B^0 + \sum_a \gamma_a B^a$, the derivatives simplify to

$$\begin{aligned} M_{,a} &= B^{a\dagger} N^{-1} R + R^\dagger N^{-1} B^a,\\ M_{,ab} &= B^{a\dagger} N^{-1} B^b + B^{b\dagger} N^{-1} B^a,\\ j_{,a} &= B^{a\dagger} N^{-1} d, \text{ and } j_{,ab} = 0. \end{aligned} \tag{47}$$

From Eq. (45) it becomes apparent that the optimal signal reconstruction in the presence of calibration uncertainties should contain a correction to $m^\star = D^{\gamma^\star}\, j^{\gamma^\star}$, which corrects for the possibility that certain structures in the data might well be due to a miscalibration rather than being caused by real signal structures. Thus, the reconstruction will be less prone to overfitting calibration errors.

The expected level of this correction can, however, be expected to be moderate in typical situations. The individual correction terms in Eq. (45) can be paired into similar ones with opposite signs which partly, but not fully, balance each other. As a consequence, we expect only a moderate net correction by them. For the sake of clarity of the following discussion, we will therefore neglect these corrections and only work with the lowest order signal estimator $m^\star = D^{\gamma^\star}\, j^{\gamma^\star}$. The accuracy of this depends, however, crucially on the quality of the calibration, which should therefore be our focus.

[9] One can also simply calculate the first two moments of the data given the signal averaged over noise and calibration realizations $\bar{d} = \langle d\rangle_{(n,\gamma|s)} = B^0 s$, $\langle (d-\bar{d})\,(d-\bar{d})^\dagger\rangle_{(n,\gamma|s)} = N + \sum_{ab} B^a s \Gamma_{ab} s^\dagger B^{b\dagger}$, and realize that the calibration marginalized likelihood has to be a Gaussian with this mean and variance since both noise and calibration uncertainty just add Gaussian variance to the data.

[10] In the case of an external calibration of only linear calibration parameters [Eq. (38)], we had shown in Eq. (40) this to be an exact result.

H. Self-calibration

1. Motivation

In many situations, only insufficient external calibration measurements are available. In this case, the signal s of scientific interest has also to serve as a calibration signal. Some *self-cal* procedure has to be applied in which signal and calibration parameters have to be determined simultaneously from the same data.

Furthermore, the case of a perfectly known external calibration is rarely met in practice. Usually, the calibration signal c was measured with another imperfect reference instrument as well as with the scientific instrument that is also used to observe the science signal s. We can now regard the combined measurements (c with reference instrument, c with scientific instrument, and s with scientific instrument) as a single measurement, with combined signals, responses, noises, and calibration parameter sets.

In our mathematical description, we can combine these individual measurements into a single measurement of a multicomponent signal $s' = (c, s)^{\mathrm{t}}$ by a multicomponent instrument delivering the combined data $d' = (d_c^{\mathrm{r}}, d_c^{\mathrm{s}}, d_s^{\mathrm{s}})^{\mathrm{t}}$. Here, the data d_c^{r} result from the measurement of the calibration signal c with the reference instrument r, the data d_c^{s} from the calibration measurement of c with the scientific instrument s, and data d_s^{s} from the science signal s measurement with the scientific instrument s. The combined measurement equation reads as

$$\begin{pmatrix} d_c^{\mathrm{r}} \\ d_c^{\mathrm{s}} \\ d_s^{\mathrm{s}} \end{pmatrix} = \begin{pmatrix} R_c^{\mathrm{r}} & 0 \\ R_c^{\mathrm{s}} & 0 \\ 0 & R_s^{\mathrm{s}} \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} + \begin{pmatrix} n_c^{\mathrm{r}} \\ n_c^{\mathrm{s}} \\ n_s^{\mathrm{s}} \end{pmatrix} \text{ or } d' = R' \quad s' + n',$$

with the combined noise vector n', and the combined response R' of the three original measurements. In order that the calibration measurement provides any benefit, the calibration parameters of the last two measurements with the scientific instrument need to be identical or at least sufficiently correlated with each other.

Since for this combined measurement no external calibration exists (we have incorporated all external measurements), it should as well be reconstructed with a *self-cal* scheme.

2. Practice

Self-cal usually consists of repeatedly reconstructing the signal, assuming a calibration to be correct, and determining the calibration, while assuming the signal to be given. These steps are repeated until signal and calibration estimates have converged sufficiently. However, a proof that this converges and the meaning of the fix point seem are often missing in the *self-cal* literature.

Using simultaneously MAP estimators for the signal inference and the calibration actually means that the joint posterior of signal and calibration parameters is extremized in both unknowns. This is equivalent to the minimum of the information Hamiltonian

$$\begin{aligned} \mathcal{H}(d, \gamma, s) &= -\ln \mathcal{P}(d, \gamma, s) \\ &= \tfrac{1}{2}(d - R^{\gamma} s)^{\dagger} N^{-1} (d - R^{\gamma} s) \\ &\quad + \tfrac{1}{2} \gamma^{\dagger} \Gamma \gamma + \tfrac{1}{2} s^{\dagger} S^{-1} s + \text{const}, \end{aligned} \tag{48}$$

which is as given by

$$\begin{aligned} 0 &= \frac{\partial \mathcal{H}(d, \gamma, s)}{\partial s} = D^{-1} s - j\big|_{\gamma}, \\ 0 &= \frac{\partial \mathcal{H}(d, \gamma, s)}{\partial \gamma} = \Gamma^{-1} \gamma - s^{\dagger} R_{,\gamma}^{\dagger} N^{-1} (d - R s). \end{aligned} \tag{49}$$

The resulting formula is identical to the Wiener filter signal reconstruction [Eq. (27)] and the calibration on this signal [Eq. (37)]. Thus, the joint MAP *self-cal*scheme is equivalent or at least similar to the usual practice of iterating signal and calibration estimation.

It has been noticed, e.g. by Ref. [16], that using a joint MAP solution simultaneously for signal and nuisance parameters (here the unknown calibration, in [16] the unknown signal covariance) can be suboptimal. It is better to use the signal marginalized posterior to determine the calibration parameters and then to use the resulting parameters in the signal reconstruction. This approximation is also known under the term *empirical Bayes* (e.g., [38]).

I. Signal marginalized calibration

The signal marginalized Hamiltonian

$$\begin{aligned} \mathcal{H}(d, \gamma) &= -\ln \int \mathcal{D}s \, \mathcal{P}(d, \gamma, s) \\ &= \frac{1}{2} \left[\gamma^{\dagger} \Gamma^{-1} \gamma - \mathrm{Tr}\,(\ln D) - j^{\dagger} D \, j \right] + \text{const} \end{aligned} \tag{50}$$

can be minimized with respect to γ to find the MAP calibration solution $\gamma^{\star}$. The gradient and Hessian of this Hamiltonian are

$$\begin{aligned} \frac{\partial \mathcal{H}(d, \gamma)}{\partial \gamma_a} &= (\Gamma^{-1} \gamma)_a + \frac{1}{2} \mathrm{Tr}(D \, M_{,a}) - j^{\dagger} D \, j_{,a} \\ &\quad + \frac{1}{2} j^{\dagger} D \, M_{,a} D \, j, \end{aligned} \tag{51}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{H}(d, \gamma)}{\partial \gamma_a \partial \gamma_b} &= \Gamma_{ab}^{-1} + \frac{1}{2} \mathrm{Tr}(D \, M_{,ab} - D \, M_{,a} D \, M_{,b}) \\ &\quad + \frac{1}{2} j^{\dagger} D \, M_{,ab} D \, j + j^{\dagger} D \, M_{,a} D \, j_{,b} \\ &\quad + j^{\dagger} D \, M_{,b} D \, j_{,a} - j_{,a}^{\dagger} D \, j_{,b} - j^{\dagger} D \, j_{,ab} \\ &\quad - j^{\dagger} D \, M_{,a} D \, M_{,b} D \, j. \end{aligned} \tag{52}$$

The Hessian can be used to construct an approximative calibration uncertainty covariance matrix via

$$\Delta_{ab}^{-1} \approx \left. \frac{\partial^2 \mathcal{H}(d, \gamma)}{\partial \gamma_a \partial \gamma_b} \right|_{\gamma = \gamma^{\star}} \tag{53}$$

so that a Gaussian approximation of the calibration posterior [Eq. (44)] as well as a calibration marginalized signal reconstruction [Eq. (45)] can be obtained.

It is instructive to compare the classical formula used for external calibration (37) to the one of *self-cal* (51). For this we have to identify $m = m^{\gamma} = D^{\gamma} j^{\gamma}$ in Eq. (51), which reads

now as

$$\frac{\partial \mathcal{H}(d,\gamma)}{\partial \gamma_a} = (\Gamma^{-1}\gamma)_a + \frac{1}{2}\,\mathrm{Tr}(D\,M_{,a}) - m^\dagger R_{,a} N^{-1}(d - R\,m), \tag{54}$$

with c in Eq. (37). We see that the only change is the additional term $\frac{1}{2}\,\mathrm{Tr}(D\,M_{,a})$, which ensures that the signal uncertainty is taken into account in the calibration.

In case of only linear calibration parameters as in Eq. (38), $R^\gamma = B^0 + \sum_a \gamma_a B^a$, a nearly closed calibration formula can be given:

$$\begin{aligned} \gamma^\star &= \Delta' h, \text{ with} \\ \Delta'^{-1}_{ab} &= \Gamma^{-1}_{ab} + \mathrm{Tr}[(m\,m^\dagger + D)B^{a\dagger}N^{-1}B^b], \\ h_b &= m^\dagger B^{b\dagger} N^{-1} d - \mathrm{Tr}[(m\,m^\dagger + D)B^{0\dagger}N^{-1}B^b]. \end{aligned} \tag{55}$$

This formula is not exactly closed since $m = m^{\gamma^\star}$ and $D = D^{\gamma^\star}$ are still calibration dependent. However, iterations as performed usually in *self-cal* schemes should converge to a fixed point. In practice, one might prefer to use a gradient scheme based on Eq. (54) rather than to iterate Eq. (55) since the latter contains nested matrix inversions that are numerically expensive.

The apparent calibration covariance Δ' is also not exactly identical to Δ obtained from the inverse Hessian (52) since precisely the calibration dependence of m and D were ignored in the identification of Δ'. It should, however, be a useful approximation with lower computationally complexity than Eq. (52).

A comparison of the calibration formulas (39) and (55) while identifying c with m reveals the main effect of the signal marginalization. This inserts an additional signal uncertainty covariance D wherever a term $m\,m^\dagger$ appears. As we had seen in the case of the external calibration, the quantity determining how sensitive the calibration information h reacts to γ, $Q^{ab} = s^\dagger B^{a\dagger} N^{-1} B^b\, s = \mathrm{Tr}(s\,s^\dagger B^{a\dagger} N^{-1} B^b)$ in $h_a \approx \sum_b Q^{ab}\gamma_b$ (neglecting the noise impact), depends quadratically on the unknown signal s. Using $m\,m^\dagger$ as an estimator for the quadratic signal $ss^\dagger$ underestimates the variance of the latter since m is a filtered version of s with less power. The correct *a posteriori* expectation value for $ss^\dagger$,

$$\langle ss^\dagger\rangle_{(s|d,\gamma)} = m\,m^\dagger + D, \tag{56}$$

contains the signal uncertainty covariance D in order to correct for this bias. This is therefore the appropriate term to be used in Q^{ab}.

The calibration propagator Δ also gets a similar term $Q^{ab} = \mathrm{Tr}[B^{a\dagger}N^{-1}B^b(m\,m^\dagger + D)]$ that ensures that a good guess for the signal variance is used in the term describing the calibration measurement precision. This additional positive term due to the D correction in Δ^{-1} decreases Δ and makes therefore the calibration reconstruction $\gamma^\star = \Delta h$ less reactive to variations in the data. This prevents an overcalibration on data features that might be caused by noise. Furthermore, the new *self-cal* scheme corrects a systematic bias of classical *self-cal* towards delivering higher calibration values.[11] We therefore suspect the signal marginalized calibration procedure to provide a more accurate calibration and signal reconstruction than the classical joint MAP calibration procedure. Whether this is indeed the case, we investigate numerically.

IV. NUMERICAL EXAMPLE

A. Gain uncertainties

As an illustrative case to compare the performance of the different calibration schemes we investigate a simple one-dimensional measurement problem with gain fluctuations in the spirit of the simplistic example of Sec. II.

A signal field $s = (s_x)_x$ over the periodic domain $\Omega = \{x\}_x = [0,1) \subset \mathbb{R}$ is observed $u = 3$ times by a scanning instrument. The instrument has a perfect pointlike response at scanning location $x_t = t \mod 1$ at time t but a time varying gain $g_t = 1 + \gamma_t$. The instrument samples with a period $\tau = 2^{-9} \approx 2 \times 10^{-3}$ so that the ith data point is at location $x_{i\tau} = (i\tau) \mod 1$. It is convenient to regard the data as a function of time (which is discrete with period τ, so that $t \in \{0, \tau, 2\tau, \ldots, u\}$) and to exploit the fact that the spatial and temporal coordinates are well aligned (except that the temporal domain is u times larger than the spatial domain).

The response operator

$$R_{tx} = (1 + \gamma_t)\,\delta(x - x_t) \tag{57}$$

is of the linear calibration parameter form $R^\gamma = B^0 + \sum_a \gamma_a B^a$ [Eq. (38)] with $B^0_{tx} = \delta_{xx_t}$ and $B^a_{tx} = \delta_{at}\delta_{xx_t}$ so that $R_{tx,t'} = (R^\gamma_{tx})_{,\gamma_{t'}} = \delta_{tt'}\delta_{xx_t}$ and $R_{tx,t't''} = (R_{tx})_{,\gamma_{t'}\gamma_{t''}} = 0$.[12]

The Gaussian signal, noise, and calibration covariances are assumed to be known and to be described by power spectra in Fourier space. In our concrete example, we use

$$\begin{aligned} P_s(k) &= \frac{a_s}{[1 + (k/k_s)^2]^2}, \\ P_\gamma(\omega) &= \frac{a_\gamma}{[1 + (\omega/\omega_\gamma)^2]^2}, \quad P_n(\omega) = a_n, \end{aligned} \tag{58}$$

respectively. We express the amplitudes as $a_s = \sigma_s^2\lambda_s$, $a_\gamma = \sigma_\gamma^2\tau_\gamma u$, and $a_n = \sigma_n^2\tau_n$ in terms of their respective variances $\sigma_s^2 = \langle s_x^2\rangle_{(s)}$, $\sigma_\gamma^2 = \langle \gamma_t^2\rangle_{(\gamma)}$, and $\sigma_n^2 = \langle n_t^2\rangle_{(n)}$ and correlation lengths $\lambda_s = 4/k_s$, $\tau_\gamma = 4/\omega_\gamma$, and $\tau_n = \tau$. We choose $\sigma_s = 1$,

[11] This is valid in the here discussed case in which the signal is obtained via noise suppressing filtering, otherwise the bias could even be opposite in cases, in which noise remnants add spurious variance to the reconstruction.

[12] As a consequence of this simple response and noise structure while assuming white noise with $N_{tt'} = \sigma_n^2\delta_{tt'}$, we get

$$\begin{aligned} M_{xy} &= \delta_{xy}\sum_t \delta_{xx_t}(1+\gamma_t)^2\sigma_n^{-2}, \\ M_{xy,t} &= 2\delta_{xy}\delta_{xx_t}(1+\gamma_t)\sigma_n^{-2}, \\ M_{xy,tt'} &= 2\delta_{xy}\delta_{xx_t}\delta_{tt'}\sigma_n^{-2}, \\ j_x &= \sum_t (1+\gamma_t)\,\delta_{xx_t} d_t \sigma_n^{-2}, \\ j_{x,t} &= \delta_{xx_t} d_t \sigma_n^{-2}, \text{ and } j_{x,tt'} = 0. \end{aligned}$$

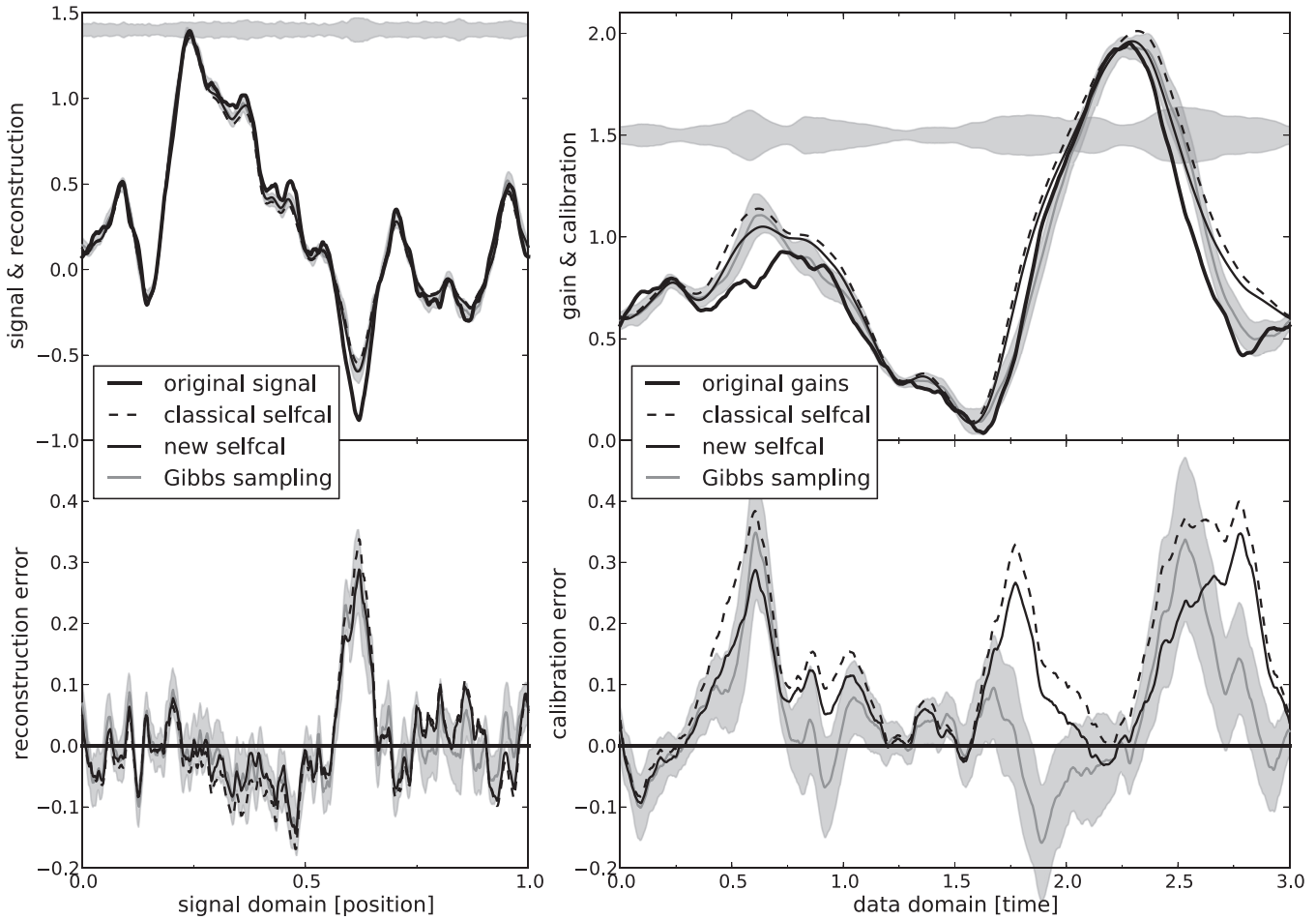

FIG. 3. Signal reconstruction and calibration using *self-cal*. Left: Signal (as in previous figures) and its reconstruction using *classicalself-cal* [iterating Eq. (27) with $\gamma = \gamma^*$ to get m and Eq. (39) with $c = m$ to get γ^*; precisely Eq. (59) is used with $T = 0$], *new self-cal* [Eq. (59) with $T = 1$], and using Gibbs sampling (see Sec. III F). Right: The gain curve and its reconstructions using the *classical* [Eq. (59) with $T = 0$] and the *new self-cal* [Eq. (59) with $T = 1$] scheme as well as using Gibbs sampling (Sec. III F). The uncertainty estimates of the Gibbs sampling are shown as gray bands in all panels. In the top panels, it is shown twice, once around the Gibbs sampling mean and once at an arbitrary location for better visual inspection of its structure.

$\sigma_\gamma = 0.75$, and $\sigma_n = 0.2$ and correlation lengths $\lambda_s = 0.3$ and $\tau_\gamma = 1.5$. This way, we have a unit variance signal, a 75% calibration uncertainty, and 20% white noise per measurement (in terms of typical signal strength). The noise is white, the signal short correlated (with about three correlation regions within the signal domain Ω), and the gain correlates over a slightly larger region (a bit more than the size of the signal domain Ω). The gains are only slightly correlated between subsequent passages over the same position ($\tau_\gamma = 1.5$).

Any systematic difference in the data resulting from identical signal positions should be due to gain variations. A decent *self-cal* scheme should be able to exploit this redundancy to estimate the gains and therefore the signal.

However, a global degeneracy of the data with respect to its variations being caused by signal and gain variations can only partly be broken by the three redundant scans over the signal domain. The data $d_t \approx (1 + \gamma_t)\, s_{x_t}$ only report a product of signal and response and one of those can be traded for the other. Therefore, a few external calibration measurements are essential to break the degeneracy globally.

To fix this degeneracy, we assume that four additional external calibration measurements of the gain value have been performed at certain times $t_j \in \{0, 0.75, 1.5, 2.25\}$, with $d'_j = (1 + \gamma_{t_j})\, c + n'_j$ by momentarily switching the observation to a strong calibration source with a known strength of $c = 4$. We assume that the noise during these calibration measurements is as before, $n'_j \hookleftarrow \mathcal{G}(n'_j, \sigma_n^2)$.[13]

The *self-cal* equations become

$$\begin{aligned} \gamma^\star &= \Delta\, h, \text{ with} \\ \Delta^{-1}_{tt'} &= \Gamma^{-1}_{tt'} + \delta_{tt'}\left(q_t + c^2 \sum_j \delta_{tt_j}\right)\sigma_n^{-2}, \\ h_t &= \left[d_t m_{x_t} - q_t + c^2 \sum_j \delta_{tt_j} d'_j\right]\sigma_n^{-2}, \\ q_t &= m^2_{x_t} + T\, D_{x_t x_t}. \end{aligned} \tag{59}$$

[13]For mental and notational convenience, we ignore that during the external calibration measurement usually no science signal data can be taken by real instruments. However, this idealization is inessential and has only a negligible impact on the results.

TABLE I. Reconstruction error of the different reconstruction for the example shown in Figs. 1–3.

Reconstruction method	ε_s	ε_γ
Wiener filter using known gains/signal	0.037	0.056
Expected uncertainties of above	0.040	0.063
Gibbs sampling	0.073	0.116
Expected uncertainty of above	0.042	0.076
No calibration, unit gains	0.110	0.533
Only external calibration	0.081[a]	0.246
Classical *self-cal*	0.089	0.192
New *self-cal*	0.073	0.141

[a]In this particular realization of signal, gain, and data, despite the external calibration being relatively poor it has coincidentally provided a better signal reconstruction than classical *self-cal*.

Here, we introduced the expected posterior variance of the signal realization as constrained by the data $q_t = \langle s^2_{x_t} \rangle_{(s|d,\gamma)} = m^2_{x_t} + D_{x_t x_t}$. Furthermore, we introduced T as a parameter that switches between classical *self-cal* ($T = 0$) and the new signal marginalized *self-cal* ($T = 1$).

B. Calibration comparison

A simulated signal, gain, and resulting data realization using the above specifications, as well as their reconstructions using different information, assumptions, and approximations can be seen in Figs. 1, 2, and 3. These were generated using the generic signal inference framework NIFTY[14] [39].

We quantify the signal and gain reconstructions in Table I in terms of their average squared errors $\varepsilon_s^2 = (m - s)^\dagger (m - s)$ and $\varepsilon_\gamma^2 = \frac{1}{u}(\gamma^\star - \gamma)^\dagger (\gamma^\star - \gamma)$, respectively. Their expectation values, in case of known Gaussian statistics, are given by

$$\begin{aligned} \left\langle \varepsilon_s^2 \right\rangle_{(d,s|\gamma)} &= \int_0^1 dx\, D_{xx}, \\ \left\langle \varepsilon_\gamma^2 \right\rangle_{(d,\gamma|s)} &= \frac{1}{3} \int_0^3 dt\, \Delta_{tt}. \end{aligned} \tag{60}$$

The best results are of course obtained when signal or calibration are known. These Wiener filter solutions are optimal (dotted lines in Fig. 2) and their uncertainty estimates are reliable (gray regions in Fig. 2).

The worst signal reconstruction is the one obtained while assuming unit gains (thin gray lines in Fig. 2). Using only the four external calibration measurements gives slightly better results (dashed lines in Fig. 2). The classical *self-cal* provides more accurate calibration (dashed lines in Fig. 3), which is further improved by the uncertainty corrections included in the new *self-cal* scheme (solid lines in Fig. 3). The best *self-cal* solutions are provided by the Gibbs sampling. Despite some numerical noise in the results, which can only be suppressed by investing a large number of samples, these are optimal and therefore provide a good benchmark for comparison. The *new self-cal*scheme obviously does not fully reach the accuracy of the Gibbs sampling. Nevertheless, it is a significant improvement over the *classical self-cal* as its solutions are visibly closer to the optimal Gibbs sampling results.

These numbers and also the bottom panels of Fig. 3 show further that the uncertainties in the calibration are systematically larger than those of the signal. This is due to the fact that the *self-cal* has to rely on the signal being significantly nonzero, which is not the case for many locations, whereas the signal reconstruction is data driven for all positions except some rare points where the gain $g = 1 + \gamma$ happens to vanish.

Since the signal uncertainty correction of the calibration removed a systematic bias of the classical scheme, which had led to overestimated gain solutions, the corresponding reconstructed signal shows more variation as the one without this correction. This is visible by careful inspection of the top left panels of Fig. 3.

V. CONCLUSIONS

We investigated the calibration problem of signal reconstruction from data. Although we concentrated on simplified cases, approximating all uncertainties in signal, calibration, and noise to be Gaussian distributed, we believe that the gained qualitative insights are also valid in many other circumstances.

In case a perfect or sufficient external calibration measurement is missing, the signal to be measured has also to serve as a calibrator. This is usually done by *self-cal* schemes, which reconstruct the signal assuming some calibration, calibrate on the reconstructed signal, and iterate this until convergence or other termination criteria are met. We have shown that such *self-cal* schemes arise naturally from trying to maximize the joint posterior of signal and calibration. We therefore demonstrated that any fixed point of such *self-cal* iterations must be a maximum of this posterior. There is, however, no guarantee that the obtained maximum is a global one.

The joint MAP estimator is not necessarily optimal in the sense of a minimal expected square error. Due to the interwoven coupling between signal and calibration in the data, this maximum is indeed not optimal. In order to obtain improved signal and calibration schemes, we worked out the calibration marginalized signal posterior and the signal marginalized calibration posterior and the resulting maximum *a posteriori* estimators. Both contain correction terms taking into account the remaining uncertainties of calibration and signal, respectively.

For the canonical situation that the signal is a quantity that varies around zero, whereas the signal response has a known nonzero part, we argue that the calibration corrections due to signal uncertainties are more essential than the signal reconstruction corrections due to calibration uncertainties. The reason is that, in this case, the information source of the data on the unknown signal contains a calibration independent term, whereas the information source for the calibration requires signal information. This is reflected by the observation that the calibration uncertainty corrections for the signal as given by Eq. (45) contain pairs of mutually nearly canceling terms.

In contrast to this, the calibration correction for signal uncertainties as given by Eq. (55) is of a systematic nature. It reduces, on average and for positive known part of the response, the values of the inferred calibration solution. This leads to a more pronounced and thereby more accurate signal

[14]To be found at www.mpa-garching.mpg.de/ift/nifty

reconstruction as more of the data variance can be assigned to the signal. We have illustrated this with a simplistic numerical example.

The proposed improvement of *self-cal* schemes should not be regarded as the ultimate theory of calibration. A number of approximations have been incorporated in order to limit the computational complexity. In particular, the mutual dependence of signal and calibration uncertainties are not fully taken into account and only the dominant influence of the uncertainties on the posterior means of signal and calibration were calculated. A comparison with a numerically expensive, but asymptotically exact, Gibbs sampling scheme shows that the corrections are indeed a good step in the right direction. However, they also show that there is still space for further improvements.

Thus, we believe that these corrections can help to refine the contemporary art of calibration and thereby improve measurement results in many areas of science and technology.

ACKNOWLEDGMENTS

We gratefully thank V. Böhm, S. Dorn, and N. Opermann for discussions, feedback, and comments. The calculations were performed using the generic signal inference framework NIFTY to be found at www.mpa-garching.mpg.de/ift/nifty [39].

[1] H. Scheffé, Ann. Stat. **1**, 1 (1973).
[2] P. J. Brown, J. R. Stat. Soc., Ser. B **44**, 287 (1982).
[3] C. Osborne, Int. Stat. Rev. **59**, 309 (1991).
[4] M.-A. Gruet, Ann. Stat. **24**, 1474 (1996).
[5] R. N. Feudale, N. A. Woody, H. Tan, A. J. Myles, S. D. Brown, and J. Ferré, Chemom. Intell. Lab. Syst. **64**, 181 (2002).
[6] N. Padmanabhan, D. J. Schlegel, D. P. Finkbeiner, J. C. Barentine, M. R. Blanton, H. J. Brewington, J. E. Gunn, M. Harvanek, D. W. Hogg, Ž. Ivezić, D. Johnston, S. M. Kent, S. J. Kleinman, G. R. Knapp, J. Krzesinski, D. Long, E. H. Neilsen, Jr., A. Nitta, C. Loomis, R. H. Lupton, S. Roweis, S. A. Snedden, M. A. Strauss, and D. L. Tucker, Astrophys. J. **674**, 1217 (2008).
[7] J. F. Jones, J. Spacecr. Rockets **4**, 554 (1967).
[8] F. R. Schwab, in *1980 International Optical Computing Conference I*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 231, edited by W. T. Rhodes (SPIE, Bellingham, WA, 1980), pp. 18–25.
[9] T. J. Cornwell and P. N. Wilkinson, Mon. Not. R. Astron. Soc. **196**, 1067 (1981).
[10] T. Cornwell, in *Synthesis Mapping*, edited by A. R. Thompson and L. R. D'Addario (National Radio Astronomy Observatory, Green Bank, West Virginia, 1982), p. 13.
[11] T. J. Pearson and A. C. S. Readhead, Astron. Astrophys. **22**, 97 (1984).
[12] R. D. Ekers, in *Serendipitous Discoveries in Radio Astronomy*, edited by K. I. Kellermann and B. Sheets (National Radio Astronomy Observatory, Green Bank, West Virginia, 1984), p. 154.
[13] J. E. Noordam and O. M. Smirnov, Astron. Astrophys. **524**, A61 (2010).
[14] A. Lannes and J.-L. Prieur, Astron. Nachr. **332**, 759 (2011).
[15] T. A. Enßlin, M. Frommert, and F. S. Kitaura, Phys. Rev. D **80**, 105005 (2009).
[16] T. A. Enßlin and M. Frommert, Phys. Rev. D **83**, 105014 (2011).
[17] J. H. Davis, C. Bœhm, N. Oppermann, T. Enßlin, and T. Lacroix, Phys. Rev. D **86**, 015027 (2012).
[18] S. L. Bridle, R. Crittenden, A. Melchiorri, M. P. Hobson, R. Kneissl, and A. N. Lasenby, Mon. Not. R. Astron. Soc. **335**, 1193 (2002).
[19] J. C. Lemm, *Bayesian Field Theory* (Johns Hopkins University Press, Baltimore, MD, 2003), arXiv:physics/9912005.
[20] T. Enßlin, in *Proceedings of the 32nd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, American Institute of Physics Conference Series Vol. 1553, edited by U. von Toussaint (AIP, Melville, NY, 2013), pp. 184–191.
[21] N. Oppermann, G. Robbers, and T. A. Enßlin, Phys. Rev. E **84**, 041118 (2011).
[22] N. Oppermann *et al.*, Astron. Astrophys. **542**, A93 (2012).
[23] N. Oppermann, M. Selig, M. R. Bell, and T. A. Enßlin, Phys. Rev. E **87**, 032136 (2013).
[24] T. A. Enßlin and C. Weig, Phys. Rev. E **82**, 051112 (2010).
[25] L. Winderling, Master's thesis, Ludwig-Maximilians-Universität, München, 2012.
[26] R. T. Cox, Am. J. Phys. **31**, 66 (1963).
[27] E. T. Jaynes, in *Probability Theory*, edited by G. Larry Bretthorst (Cambridge University Press, Cambridge, UK, 2003), p. 758.
[28] A. Caticha, in *Proceedings of the 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, AIP Conf. Proc. No. 1073 (AIP, New York, 2008), arXiv:0808.0012.
[29] S. Zaroubi, Y. Hoffman, K. B. Fisher, and O. Lahav, Astrophys. J. **449**, 446 (1995).
[30] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (Wiley, New York, 1949).
[31] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[32] E. T. Jaynes, Proc. IEEE **70**, 939 (1982).
[33] S. Geman and D. Geman, IEEE Trans. Pattern Anal. Machine Intell. **PAMI-6**, 721 (1984).
[34] B. D. Wandelt, D. L. Larson, and A. Lakshminarayanan, Phys. Rev. D **70**, 083511 (2004).
[35] M. Selig and T. Enßlin, arXiv:1311.1888.
[36] H. Junklewitz, M. R. Bell, M. Selig, and T. A. Enßlin, arXiv:1311.5282.
[37] J. Jasche, F. S. Kitaura, B. D. Wandelt, and T. A. Enßlin, Mon. Not. R. Astron. Soc. **406**, 60 (2010).
[38] S. Petrone, J. Rousseau, and C. Scricciolo, arXiv:1204.1470.
[39] M. Selig, M. R. Bell, H. Junklewitz, N. Oppermann, M. Reinecke, M. Greiner, C. Pachajoa, and T. A. Enßlin, Astron. Astrophys. **554**, A26 (2013).