Scientific Research Publishing

Advances in Pure Mathematics, 2026, 16(5), 351-378
https://www.scirp.org/journal/apm
ISSN Online: 2160-0384
ISSN Print: 2160-0368

A Bayesian-Inspired Framework for Parameter Estimation and Error Quantification

Manfred Wiessner[1], Benoît Loridant[2], Paul Angerer[3], Martin Medebach[4], Ewald Werner[5,6], Ernst Gamsjäger[6*]

[1]Dual Analytics, Geistthal-Södingberg, Austria
[2]Chair of Mathematics and Statistics, Montanuniversität Leoben, Leoben, Austria
[3]Materials Center Leoben Forschung GmbH, Leoben, Austria
[4]Institute of Microwave and Photonic Engineering, Graz University of Technology, Graz, Austria
[5]Chair of Materials Science, Technical University Munich, Garching, Germany
[6]Chair of Mechanics, Montanuniversität Leoben, Leoben, Austria
Email: *ernst.gamsjaeger@unileoben.ac.at

How to cite this paper: Wiessner, M., Loridant, B., Angerer, P., Medebach, M., Werner, E. and Gamsjäger, E. (2026) A Bayesian-Inspired Framework for Parameter Estimation and Error Quantification. *Advances in Pure Mathematics*, **16**, 351-378.
https://doi.org/10.4236/apm.2026.165019

Received: April 3, 2026
Accepted: May 19, 2026
Published: May 22, 2026

Abstract

This work introduces a novel Bayesian inspired regression method for the simultaneous estimation of model parameters and data uncertainties. The key mathematical result of this framework is an extended least squares objective function. The conventional sum of squared residuals is expanded by adding the logarithms of the computed standard deviations. This approach is particularly useful in cases with strongly varying or parameter-dependent uncertainties. Through five examples, we demonstrate that our extended least squares analysis robustly estimates data uncertainties and quantifies model parameter correlations via the Hessian matrix of the objective function. Finally, in a sixth example from materials science, we applied the method to model the evolution of the dislocation density in martensite during annealing of chromium stainless steel using a Boltzmann function. The approach successfully estimates the values of the parameters, their uncertainties and their correlations. A key outcome of our uncertainty quantification is the derivation of a credible interval for the simulated dislocation densities.

Keywords

Regression Analysis, Bayesian Statistics, Credible Interval, Unknown Systematic Errors, Heteroskedasticity Errors

1. Introduction

Observations, phenomena, and processes are often quantitatively characterized by

analyzing experimental or empirical data. One of the most widely used methods for fitting model parameters to data sets is the least squares method. However, this method has several limitations: it is sensitive to outliers—often caused by measurement errors—and assumes known weighting of the data points and uncorrelated residuals, which may not hold in practice.

In addition, unknown systematic errors may occur due to inadequate models; compare [1]. In this context, [2] famously stated that "All models are wrong, but some are useful," highlighting the inevitable presence of unknown systematic errors and outliers in both models and measurement data. To address these challenges, various strategies have been developed:

- Robust regression techniques aim at down-weighting, as found in [3]-[6] or eliminating outliers, see [7], while outlier detection methods, described in [8], seek to identify and correct such anomalies in the data.
- Iteratively reweighted least squares (IRLS) is introduced in [9] and in the textbook from [10], and is now established as a standard method for regression models. This approach is particularly useful when the data weights are not known in advance but can be estimated from the residuals of a previous fit.
- By means of a Bayesian framework prior knowledge can be introduced into the regression analysis; details can be found in [11].

The above mentioned approaches (IRLS, prior knowledge from Bayesian framework) are further developed in the present work to address unknown uncertainties and varying weighting factors during regression analysis. In our method, the weighting factors are represented by functions with adjustable parameters, and the model parameters are estimated simultaneously. The underlying theory is presented in the following section.

2. Theory

2.1. A Comparison of Frequentist and Bayesian Paradigms

In the field of data analysis, measurements are primarily evaluated using two distinct schools of thought: frequentist or classical statistics and Bayesian statistics. This section provides a comparative overview of these two fundamental approaches.

According to [12], classical statistics is characterized by the following principles:

- At a given point x_i, frequent measurements y_i^j are carried out where j denotes the measurement series. Thereby, a probability distribution (e.g., a Gaussian distribution) for y_i can be estimated for each measurement point i.
- In classical statistics, a distinction is made between the true value of the model parameter and its estimated value obtained from sample data. While the true model parameter is treated as an a-priori unknown fixed constant, the estimator (the specific value derived from regression analysis) is a random variable with a distribution induced by the sampling process. Consequently, a confidence interval is determined based on the sampling distribution of this estimator, providing a plausible range for the true parameter.

A central method in classical regression analysis is the determination of a curve $y = f(x)$ that approximates n measurement points (x_i, y_i). Measurement uncertainties are considered by including weighting factors in the sum of squared residuals Z_{lsq}:

$$Z_{\text{lsq}} = \sum_{i=1}^{n} w_i \left[y_i - f\left(x_i, \vec{\xi}_1\right) \right]^2, \tag{1}$$

where w_i is the weighting factor of the i-th measurement point and $\vec{\xi}_1$ are the parameters of the model function $f\left(x_i, \vec{\xi}_1\right)$. The fitting parameters are obtained by minimizing Z_{lsq}, which represents the weighted classical least squares method. In contrast, Bayesian statistics interprets probabilities as measures of belief or uncertainty. In general, the probability $P(B \mid A)$ is the conditional probability of event B given that event A is true. Prior probabilities $P(A)$ for an event A express beliefs before observing data. This prior probability is updated by the likelihood $P(B \mid A)$ via Bayes' theorem (see, e.g., [12] and [13]) and yields the posterior probability $P(A \mid B)$, which represents updated beliefs after observing event B. The posterior probability is normalized by the evidence $P(B)$ (or marginal likelihood):

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}. \tag{2}$$

In this work, Bayes' theorem is used to infer the model parameters $\vec{\xi}$ by determining the conditional probability $P\left(\vec{\xi} \mid (x_i, y_i), M\right)$, where (x_i, y_i) are the observed data and M denotes the underlying model hypothesis. In classical statistics, model parameters are fixed, but unknown constants. In Bayesian statistics, in contrast, model parameters are described by probability density functions rather than being treated as fixed constants. Thus, the model curve is known within a credible interval. Whereas in classical statistics measurements are frequently repeated to determine the uncertainty of each measurement point, the measurement is assumed to be certain in conventional Bayesian statistics.

This paper will investigate the regression problem through a novel Bayesian approach employing a global minimization strategy. We find a measure for unknown systematic errors, regardless of whether they originate from the experiment or the model, by using the residuals in a Bayesian framework. This analysis will reveal that this method leads to a fundamental extension of the classical least-squares sum by introducing an additional significant term.

2.2. Determining Weighting Factors

The determination of weighting factors is a central topic in this study, as introduced above. In a common standard approach, often employed when repeated j measurements are available for each x_i, the weighting factor $w_i(x_i)$ for the data point i is determined as follows: Repeated measurements j produce a number of j values y_i at x_i, the sample variance $s_i^2(x_i)$ of these repeated measurements at x_i is calculated; the reciprocal of this variance is then used as

the weighting factor: $w_i(x_i) = \frac{1}{s_i^2(x_i)}$, see [10].

In the frequentist statistical approach it is implicitly assumed that the model is structurally perfect. The residuals, representing the deviations between the model and the measurements are solely dependent on the fluctuations in the data points, but do not account for systematic errors. In practice, it is often the case that different models describe the measurement data with varying degrees of accuracy, and it is usually not known a priori which model will be the most accurate. In addition to possibly occurring fluctuations in data points, unknown systematic errors may be caused by a too simple model, see [1]. These unknown systematic errors lead to significantly larger residuals than the variability of the data points alone would suggest. These considerations motivate us to assign lower weights to regions with higher residuals.

However, a more advanced regression analysis with observation weights not known a priori is known as iteratively reweighted least squares (IRLS) method [9]. The weights are iteratively optimized in each step. The regression is successful when the resulting parameter estimates do not differ within a certain threshold from those obtained in the previous iteration. This involves using the residuals from the weighted fit to re-estimate the variance (or standard deviation) function and update the weights accordingly. In practice, one or two iterations are usually sufficient to achieve stable estimates of the parameters. The IRLS method is a robust regression technique, but it has some limitations:

- The method can be prone to convergency problems and may end up at a local minimum instead of the global minimum.
- Since IRLS is primarily an optimization method, it also lacks a comprehensive statistical framework for determining the probability distributions of uncertainty parameters.

In order to resolve these limitations, we decided to develop a Bayesian informed extended least squares method. The weighting factors are modeled as functions dependent on a set of weight parameters $\vec{\xi}_2$, which are optimized simultaneously with the primary model parameters $\vec{\xi}_1$. This comes at the cost that the dimensionality of the optimization space is increased by the length of the vector $\vec{\xi}_2$ of the weight parameters. In contrast to the iteratively reweighted least squares (IRLS) method, our proposed procedure is straightforward and non-iterative. More details are provided in the following section.

2.3. The Extended Least Squares Method

The extended least squares method is derived by applying Bayes' theorem [14] to jointly infer the model and the weight parameters. The parameter set $\vec{\xi}$ is composed of the primary model parameters $\vec{\xi}_1$ and the weight parameters $\vec{\xi}_2$. The prior probability distribution $P(\vec{\xi} \mid M)$ contains prior knowledge of the parameter set $\vec{\xi}$, where M denotes the hypothesis space. Upon availability of a new measurement series, denoted as $D = \{(x_i, y_i) \mid i = 1, \cdots, n\}$, the prior distribution

is updated via the likelihood function $P\left(D\mid\vec{\xi},M\right)$. This likelihood quantifies how well the parameter vector $\vec{\xi}$ explains the observed data, *i.e.*, it reflects the goodness of fit between model and measurements. The joint probability of the data and parameters is given by the product of the likelihood function and the prior distribution, *i.e.* $P\left(D\mid\vec{\xi},M\right)\cdot P\left(\vec{\xi}\mid M\right)$. The evidence $P\left(D\mid M\right)$ serves as a normalizing constant for this joint probability. The posterior probability distribution $P\left(\vec{\xi}\mid D,M\right)$ is then obtained as:

$$P\left(\vec{\xi}\mid D,M\right)=\frac{P\left(D\mid\vec{\xi},M\right)\cdot P\left(\vec{\xi}\mid M\right)}{P\left(D\mid M\right)}, \tag{3}$$

which can be rewritten in logarithmic form:

$$\ln P\left(\vec{\xi}\mid D,M\right)=\ln P\left(D\mid\vec{\xi},M\right)+\ln P\left(\vec{\xi}\mid M\right)-\ln P\left(D\mid M\right). \tag{4}$$

Since the hypothesis space M of the model is assumed to be constant, the distribution $P\left(D\mid M\right)$, named evidence, also remains constant, see also [12]. In the case of a uniform prior distribution, a so-called "flat" prior, compare [12], where all admissible parameter values are considered equally probable, the prior simplifies to: $P\left(\vec{\xi}\mid M\right)=1$. Under this assumption, Equation (3) reduces to Equation (5), implying that the posterior distribution $P\left(\vec{\xi}\mid D,M\right)$ is proportional to the likelihood function $P\left(D\mid\vec{\xi},M\right)$ [15],

$$P\left(\vec{\xi}\mid D,M\right)\propto P\left(D\mid\vec{\xi},M\right). \tag{5}$$

Assuming a flat prior distribution, where all parameter values have equal prior probabilities, we can directly compare the likelihood of different parameter sets with the observed data. This allows us to focus on evaluating the goodness-of-fit and determining the most plausible parameter values based on the given data,

$$P\left(\vec{\xi}\mid D,M\right)=\frac{P\left(D\mid\vec{\xi},M\right)}{P\left(D\mid M\right)}. \tag{6}$$

In the following step, a model function y_{c} is defined which describes the data set D. Then, we define the likelihood as a Gaussian function G_i, where y_i, $y_{\mathrm{c},i}$ and σ_i are the measured y-value, the calculated regression value and the standard deviation at measurement point i, respectively:

$$G_i\left(y_{\mathrm{c},i},\sigma_i\right)=\frac{1}{\sigma_i\sqrt{2\pi}}\exp\left(-\frac{\left[y_i-y_{\mathrm{c},i}\right]^2}{2\sigma_i^2}\right). \tag{7}$$

The standard deviations σ_i are a priori unknowns and must be replaced by the estimated values $s_i=s_i\left(y_{\mathrm{c},i},\vec{\xi}_2\right)$ during regression. The calculated value at at each measurement point $D_i\left(x_i,y_i\right)$ is denoted as $y_{\mathrm{c},i}=y_{\mathrm{c}}\left(x_i,\vec{\xi}_1\right)$. In particular, we express the likelihood function $G_i\left(y_{\mathrm{c}}\left(x_i,\vec{\xi}_1\right),s_i\right)$ at each measurement point i:

$$G_i\left(y_{\mathrm{c}}\left(x_i,\vec{\xi}_1\right),s_i\right)=\frac{1}{s_i\sqrt{2\pi}}\exp\left(-\frac{\left[y_i-y_{\mathrm{c}}\left(x_i,\vec{\xi}_1\right)\right]^2}{2s_i^2}\right). \tag{8}$$

The regression values $y_{c,i} = y_c\left(x_i, \vec{\xi}_1\right)$ are computed from the model function at the positions x_i and depend on the primary model parameters $\vec{\xi}_1$.
In this approximation, we assume that all individual observations are independent. This means that the likelihood function $P\left(D \mid \vec{\xi}, M\right)$ for the full dataset $D = \left\{\left(x_i, y_i\right) \mid i = 1, \cdots, n\right\}$ can be written as a product of the individual likelihoods P_i at each point:

$$P\left(D \mid \left\{\vec{\xi}_1, \vec{\xi}_2\right\}, M\right) = \prod_{i=1}^{n} P_i\left(\left(x_i, y_i\right) \mid \left\{\vec{\xi}_1, \vec{\xi}_2\right\}, M\right). \tag{9}$$

In classical least squares regression, the s_i values usually reflect only the measurement uncertainty at each x_i, determined from repeated measurements in a frequentist framework. However, in this work, we take a different approach: we let s_i account not only for the measurement uncertainty of data point i, but also for discrepancies caused by an imperfect model. These additional contributions are derived from the residuals $\left(y_i - y_{c,i}\right)$.

The estimated standard deviations s_i are directly related to the classical weighting factors w_i through:

$$w_i = \frac{1}{s_i^2}. \tag{10}$$

The weighting factor w_i, as employed in our framework, aligns with its definition in the classical least squares method, as discussed by Wolberg (2006) [16]. However, in our approach the standard deviations of the measurement points i are a priori uncertain.

Instead of assigning an individual standard deviation s_i to each measurement point $\left(x_i, y_i\right)$, we introduce a single function $s_c\left(y_c, \vec{\xi}_2\right)$ that depends on a reduced set of fitting parameters $\vec{\xi}_2$. The individual variances s_i^2 are thus approximated by model-based variances:

$$s_i^2 \approx s_c^2\left(y_{c,i}, \vec{\xi}_2\right),$$

where $y_{c,i}$ is the model prediction at x_i.

Alternatively, the calculated value $y_{c,i}\left(\vec{\xi}_1\right)$ can be replaced by the measured value y_i, yielding:

$$s_i^2 \approx s_c^2\left(y_i, \vec{\xi}_2\right).$$

The parameters $\vec{\xi} = \left\{\vec{\xi}_1, \vec{\xi}_2\right\}$ are treated as random variables within the hypothesis space M. Here, $\vec{\xi}_1$ denotes the parameters of the model function, while $\vec{\xi}_2$ parameterizes the function s_c, which describes the (unknown) standard deviation associated with each data point. Both sets of parameters are estimated simultaneously within the Bayesian framework. It is worth noting that the fitted s_c term absorbs both measurement noise and model misspecification. These two components are in principle not separately identifiable from a single measurement series unless external information or replicated measurements are available.

The likelihood function

$$P\left(\left\{\left(x_i, y_i\right)\right\} \mid \left\{\vec{\xi}_1, \vec{\xi}_2\right\}, M\right) \tag{11}$$

is modified to account for the parameter-dependent variances. Assuming independent normally distributed residuals, it takes the form:

$$P\left(\left\{\left(x_i, y_i\right)\right\} \mid \left\{\vec{\xi}_1, \vec{\xi}_2\right\}, M\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} s_{\mathrm{c},i}\left(y_{\mathrm{c},i}, \vec{\xi}_2\right)} \exp\left(-\frac{\left[y_i - y_{\mathrm{c},i}\left(\vec{\xi}_1\right)\right]^2}{2 s_{\mathrm{c},i}^2\left(y_{\mathrm{c},i}, \vec{\xi}_2\right)}\right) \tag{12}$$

A direct numerical evaluation of Equation (12) is often numerically unstable. This is because the product of exponential terms can rapidly tend towards zero for large negative exponents, potentially leading to severe numerical underflow. To circumvent this, the logarithm of the likelihood function provides a more stable and computationally feasible formulation:

$$\begin{aligned} &\ln P\left(\left\{\left(x_i, y_i\right)\right\} \mid \left\{\vec{\xi}_1, \vec{\xi}_2\right\}, M\right) \\ &= -\frac{n}{2}\ln\left(2\pi\right) - \sum_{i=1}^{n}\left\{\ln\left(s_{\mathrm{c},i}\left(y_{\mathrm{c},i}, \vec{\xi}_2\right)\right) + \frac{\left(y_i - y_{\mathrm{c},i}\left(\vec{\xi}_1\right)\right)^2}{2 s_{\mathrm{c},i}^2\left(y_{\mathrm{c},i}, \vec{\xi}_2\right)}\right\}. \end{aligned} \tag{13}$$

Inserting Equation (13) into Bayes' theorem, Equation (4) yields the following expression for the posterior distribution:

$$\begin{aligned} &\ln P\left(\left\{\vec{\xi}_1, \vec{\xi}_2\right\} \mid \left\{\left(x_i, y_i\right)\right\}, M\right) \\ &= -\ln P\left(\left\{\left(x_i, y_i\right)\right\} \mid M\right) - \frac{n}{2}\ln\left(2\pi\right) \\ &\quad - \sum_{i=1}^{n}\left\{\ln\left[s_{c,i}\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right), \vec{\xi}_2\right)\right] + \frac{\left[y_i - y_{\mathrm{c},i}\left(\vec{\xi}_1\right)\right]^2}{2 s_{\mathrm{c},i}^2\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right), \vec{\xi}_2\right)}\right\} + \ln\left(P\left(\left\{\vec{\xi}_1, \vec{\xi}_2\right\} \mid M\right)\right) \end{aligned} \tag{14}$$

The posterior distribution for a flat prior is obtained by setting $P\left(\left\{\vec{\xi}_1, \vec{\xi}_2\right\} \mid M\right) = 1$ in Equation (14):

$$\begin{aligned} &\ln P\left(\left\{\vec{\xi}_1, \vec{\xi}_2\right\} \mid \left\{\left(x_i, y_i\right)\right\}, M\right) \\ &= -\ln\left[P\left(\left\{\left(x_i, y_i\right)\right\} \mid M\right)\right] - \frac{n}{2}\ln\left(2\pi\right) \\ &\quad - \sum_{i=1}^{n}\left\{\ln\left(s_{\mathrm{c},i}\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right), \vec{\xi}_2\right)\right) + \frac{\left(y_i - y_{\mathrm{c},i}\left(\vec{\xi}_1\right)\right)^2}{2 s_{\mathrm{c},i}^2\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right), \vec{\xi}_2\right)}\right\}. \end{aligned} \tag{15}$$

In order to apply our method to applications the following workflow can be used:

- The model function $y_{\mathrm{c},i}$ is described by the parameters $\vec{\xi}_1$ within the hypothesis space M.
- The weights w_i are approximated by the function for the inverse of the model based variances s_{c}^2 with the parameters $\vec{\xi}_2$, which are defined in the hypothesis space M.
- The prior knowledge is considered by a prior probability distribution.

- The probability distributions of $\vec{\xi}_1$ and $\vec{\xi}_2$ are determined by using Equation (14).

In the subsequent section, an objective function is derived from the posterior distribution given by Equation (15). This objective function serves as the basis for parameter estimation in the proposed method. Subsequently, it is compared to the classical objective function used in conventional least squares regression.

2.4. Objective Function: Bayesian Statistics versus Classical Least Squares Method

Rearranging Equation (14) yields:

$$\begin{aligned}&\ln P\left(\left\{\vec{\xi}_1,\vec{\xi}_2\right\}\mid\left\{\left(x_i,y_i\right)\right\},M\right)+\ln P\left(\left\{\left(x_i,y_i\right)\right\}\mid M\right)+\frac{n}{2}\ln\left(2\pi\right)\\&=-\sum_{i=1}^{n}\left\{\ln\left[s_{\mathrm{c},i}\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right),\vec{\xi}_2\right)\right]+\frac{\left(y_i-y_{\mathrm{c},i}\left(\vec{\xi}_1\right)\right)^2}{2s_{\mathrm{c},i}^2\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right),\vec{\xi}_2\right)}\right\}+\ln P\left(\left\{\vec{\xi}_1,\vec{\xi}_2\right\}\mid M\right).\end{aligned}\tag{16}$$

The objective function Z is defined by multiplying both sides of Equation (16) by (−2), so that this objective function becomes closer to the classical least squares objective function. The Bayesian based problem of maximizing the log-posterior distribution is converted into a problem where the objective function Z is minimized.

$$\begin{aligned}Z&=-2\left[\ln P\left(\left\{\vec{\xi}_1,\vec{\xi}_2\right\}\mid\left\{\left(x_i,y_i\right)\right\},M\right)+\ln P\left(\left\{\left(x_i,y_i\right)\right\}\mid M\right)+\frac{n}{2}\ln\left(2\pi\right)\right]\\&=-2\left[-\sum_{i=1}^{n}\left\{\ln\left[s_{\mathrm{c},i}\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right),\vec{\xi}_2\right)\right]+\frac{\left[y_i-y_{\mathrm{c},i}\left(\vec{\xi}_1\right)\right]^2}{2s_{\mathrm{c},i}^2\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right),\vec{\xi}_2\right)}\right\}+\ln P\left(\left\{\vec{\xi}_1,\vec{\xi}_2\right\}\mid M\right)\right]\\&=\sum_{i=1}^{n}\left\{2\ln\left[s_{\mathrm{c},i}\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right),\vec{\xi}_2\right)\right]+\frac{\left[y_i-y_{\mathrm{c},i}\left(\vec{\xi}_1\right)\right]^2}{s_{\mathrm{c},i}^2\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right),\vec{\xi}_2\right)}\right\}-2\ln P\left(\left\{\vec{\xi}_1,\vec{\xi}_2\right\}\mid M\right)\end{aligned}\tag{17}$$

The final, simplified expression (the third line) is the formula that is numerically minimized to obtain the Maximum A Posteriori (MAP) estimates [17] for the parameters $\left\{\vec{\xi}_1,\vec{\xi}_2\right\}$.

For a non-informative or (flat) prior distribution, where the prior probability density $P\left(\vec{\xi}\mid M\right)=1$, its logarithm $\ln P\left(\vec{\xi}\mid M\right)$ becomes zero. Therefore, the last term in the last line of Equation (17) vanishes. The simplified objective function Z_1 thus reduces to (see also Equation (5)):

$$Z_1=\sum_{i=1}^{n}\left\{2\ln\left[s_{\mathrm{c},i}\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right),\vec{\xi}_2\right)\right]+\frac{\left[y_i-y_{\mathrm{c},i}\left(\vec{\xi}_1\right)\right]^2}{s_{\mathrm{c},i}^2\left(y_{\mathrm{c},i}\left(\vec{\xi}_1\right),\vec{\xi}_2\right)}\right\}\tag{18}$$

This objective function is equivalent to the negative log-likelihood of the model. Therefore, minimizing Z_1 is equivalent to finding the maximum likelihood estimate, which in this case also corresponds to the maximum a posteriori (MAP)

solution [18].

The model parameters $\vec{\xi}_1$ and the weight parameters $\vec{\xi}_2$ of the function $s_{c,i}$ are optimized by minimizing the objective function Z_1.

Let us recall that the objective function Z_{lsq} from the least squares method is given by

$$Z_{\text{lsq}} = \sum_{i=1}^{n} w_i \left[y_i - y_{c,i}\left(\vec{\xi}_1\right) \right]^2 = \sum_{i=1}^{n} \frac{\left[y_i - y_{c,i}\left(\vec{\xi}_1\right) \right]^2}{s_{c,i}^2\left(y_{c,i}\left(\vec{\xi}_1\right), \vec{\xi}_2\right)}. \tag{19}$$

Therefore, compared with Z_{lsq} (Equation (19)), Z_1 (Equation (18)) contains the additional term

$$2\sum_{i=1}^{n} \ln\left[s_{c,i}\left(y_{c,i}\left(\vec{\xi}_1\right), \vec{\xi}_2\right)\right].$$

This term penalizes high standard deviations $s_{c,i}$ by increasing the value of the objective function. It is important to note that when each $s_{c,i}$ is independent of the model parameters $\vec{\xi}_1$ and $\vec{\xi}_2$ (as typically assumed in classical weighted least squares, where uncertainties are fixed based on observed data only), this sum becomes a constant additive term in Z_1. In such a case, minimizing Z_{lsq} and Z_1 yields the identical optimized parameters $\vec{\xi}_1$. Thus, our Bayesian informed extended least squares algorithm simplifies to the classical least squares solution in the case of weights that do not depend on model $\vec{\xi}_1$ or the weight parameters $\vec{\xi}_2$.

This is precisely where the advantage of our method lies: the weights may depend on the parameters $\vec{\xi}_2$ of the weight function and on the parameters $\vec{\xi}_1$ of the model function and can be determined from the simultaneous optimization of $\vec{\xi}_1$ and $\vec{\xi}_2$.

2.5. Determination of Fit Parameter Errors—The Covariance Matrix

To determine the uncertainties of the fit parameters [19], the objective function Z can be locally approximated by a second-order Taylor series expansion around its minimum. The minimum of the objective function Z, Equation (17) is denoted as $\vec{\xi}_0$, in which all first partial derivatives are zero. The parameter vector $\vec{\xi}$ includes both the model parameters $\vec{\xi}_1$ and the parameters $\vec{\xi}_2$ of the function $s_{c,i}$. The second-order Taylor series expansion can thus be written as:

$$Z\left(\vec{\xi}\right) = Z\left(\vec{\xi}_{(0)}\right) + \frac{1}{2}\left(\vec{\xi} - \vec{\xi}_{(0)}\right)^{\mathrm{T}} \cdot \vec{\vec{H}} \cdot \left(\vec{\xi} - \vec{\xi}_{(0)}\right), \tag{20}$$

where $\vec{\vec{H}}$ is the Hessian matrix [11] with elements given by:

$$H_{ij} = \frac{\partial^2 Z\left(\vec{\xi}_{(0)}\right)}{\partial\xi_i \partial\xi_j}. \tag{21}$$

Consequently, the objective function Z, when approximated by a second-order Taylor expansion around its minimum (cf. Equation (20)), forms an N-dimensional paraboloid in the parameter space.

The covariance matrix $\ddot{\Sigma}$ between the element j of the vector $\vec{\xi}$ and the element k of the vector $\vec{\xi}$ is given by:

$$\Sigma_{jk} = \left(\ddot{H}\right)^{-1}_{jk}. \tag{22}$$

The off-diagonal elements $j \neq k$ are frequently defined as $\sigma_{jk} = \Sigma_{jk}$ (see e.g., [16]).

The diagonal elements Σ_{jj} of the covariance matrix $\ddot{\Sigma}$ are denoted as the variances σ_j^2:

$$\sigma_j^2 = \Sigma_{jj}. \tag{23}$$

The correlation between parameters is estimated from the 2nd order Taylor term according to [16]:

$$\rho_{jk} = \frac{\sigma_{jk}}{\sigma_j \ \sigma_k} \tag{24}$$

The correlation coefficient ρ_{jk} describes the degree of correlation between two parameters. The values of ρ_{jk} are between −1 to +1. If the value is about zero, the parameters are uncorrelated. For ρ_{jk} close to 1 the parameters are completely correlated, in case of ρ_{jk} close to −1, the parameters are completely anti-correlated. Values of exactly ±1 imply a perfect linear relationship.

2.6. Approximation of the Posterior Probability Distribution from the Objective Function

The relationship between the posterior probability distribution from Bayes' theorem and the objective function of the extended least squares method follows from the first line of Equation (17):

$$\ln P\left(\left\{\vec{\xi}_1, \vec{\xi}_2\right\} \mid \left\{\left(x_i, y_i\right)\right\}, M\right) = -\frac{1}{2} Z - \ln P\left(\left\{\left(x_i, y_i\right)\right\} \mid M\right) - \frac{n}{2} \ln\left(2\pi\right). \tag{25}$$

The posterior probability distribution P can be approximated by an n-dimensional Gaussian function. This is achieved by approximating the objective function $Z\left(\vec{\xi}\right)$ by a second-order Taylor series expansion.

$$\begin{aligned} \ln P\left(\left\{\vec{\xi}_1, \vec{\xi}_2\right\} \mid \left\{\left(x_i, y_i\right)\right\}, M\right) \approx & -\frac{1}{2}\left[Z\left(\vec{\xi}_{(0)}\right) + \frac{1}{2}\left(\vec{\xi} - \vec{\xi}_{(0)}\right)^{\mathrm{T}} \cdot \ddot{H} \cdot \left(\vec{\xi} - \vec{\xi}_{(0)}\right)\right] \\ & - \ln P\left(\left\{\left(x_i, y_i\right)\right\} \mid M\right) - \frac{n}{2} \ln\left(2\pi\right) \end{aligned} \tag{26}$$

and

$$P\left(\vec{\xi} \mid \left\{\left(x_i, y_i\right)\right\}, M\right) \approx \frac{\exp\left(-\frac{1}{2} Z\left(\vec{\xi}_{(0)}\right)\right) \cdot \exp\left(-\frac{1}{4}\left(\vec{\xi} - \vec{\xi}_{(0)}\right)^{\mathrm{T}} \cdot \ddot{H} \cdot \left(\vec{\xi} - \vec{\xi}_{(0)}\right)\right)}{P\left(\left\{\left(x_i, y_i\right)\right\} \mid M\right) \cdot \left(2\pi\right)^{n/2}}. \tag{27}$$

The corresponding probability density function P in Equation (27) is represented by an n-dimensional Gaussian function. It is clear that the Gaussian function can only approximate the solution, since the probability distributions of variances σ_k^2 follow a scaled χ^2-distribution (a special case of the Γ-distribution)

and thus deviate from the Gaussian shape.

3. Results and Discussion

In this section, the probability densities of the parameters $\left\{\vec{\xi}_1, \vec{\xi}_2\right\}$ are calculated based on the theory described above for several applications. To ensure full transparency and reproducibility, detailed calculations are provided for each example in the accompanying GitHub repository [20].

3.1. Example 1: Determination of the Standard Deviation

In this first example, we test the algorithm on a synthetic dataset of eight elements, indexed $i = 1, 2, \cdots, 8$. The dataset and detailed calculations are provided in the accompanying GitHub repository [20]. The noise-free values, y_i^0, are set to a constant mean of $\mu = 1.0$. A normally distributed noise component, y_i', is then superimposed using a random number generator, yielding the "experimental" data y_i:

$$y_i = y_i^0 + \varepsilon_i \tag{28}$$

The true standard deviation of this noise is set to $\sigma = 0.2$ for all data points. Consequently, the observed data y_i are drawn from a normal distribution $y_i \sim \mathcal{N}\left(\mu = 1, \sigma^2 = 0.04\right)$. The generated data are presented in **Figure 1**.

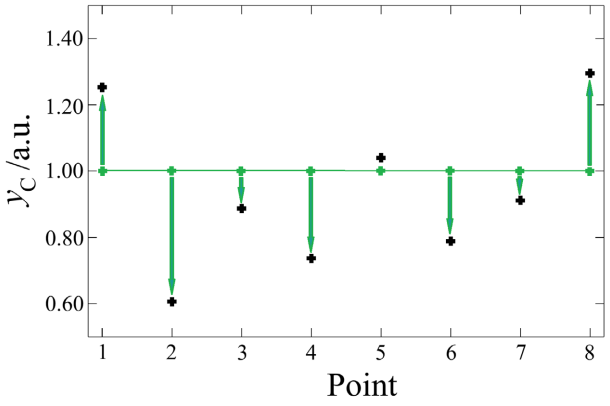

Figure 1. Synthetic data y_i with superimposed artificial noise ($\sigma = 0.2$). The green crosses represent the noise-free (true) values, y_i^0. The arrows illustrate the magnitude and sign of the residuals $\left(y_i - y_i^0\right)$.

Given the dataset depicted in **Figure 1**, it is assumed that the true (noise-free) value of each signal is $y_i^0 = 1$. The model function $y_c\left(x_i\right) = 1$ is chosen. Since the arguments of this function always equal the constant 1, the model parameter vector $\vec{\xi}_1$ contains no adjustable parameters for this specific example (*i.e.*, $\vec{\xi}_1 = \{\}$).

The objective is to estimate the standard deviation s from the "true", but a-priori unknown standard deviation σ of the noise from the data. Within our Bayesian framework, this is achieved by determining the posterior probability distribu-

tion $P(s \mid D, M)$ for the estimated standard deviation s (cf. Equation (15)).

It is assumed that the prior distribution is the same for all $s > 0$, and therefore the logarithm of the prior distribution vanishes in this function $\ln P(s \mid D, M)$. The probability density $P(s \mid D, M)$ is calculated for the unknown estimated standard deviation $s_{c,i}$ in the hypothesis space $M = \mathbb{R}^+$, with $\vec{\xi}_2 = \{s\}$ and $s_{c,i} = s \in M$:

$$\ln P(\{s\} \mid \{y_1, y_2, \cdots, y_8\}, M) = -\ln K - \frac{8}{2}\ln(2\pi) - \sum_{i=1}^{8}\left\{\ln(s) + \frac{(y_i - 1)^2}{2s^2}\right\}$$

with the evidence $K = P(\{(x_1, y_1), (x_2, y_2), \cdots, (x_8, y_8)\} \mid M)$ as a normalization factor.

The posterior probability density P is plotted versus the parameter s as can be seen in Figure 2. The normalization constant K is computed numerically, so that the integral of the posterior probability density over all possible values of s equals 1. In this example, $K = 0.2074$.

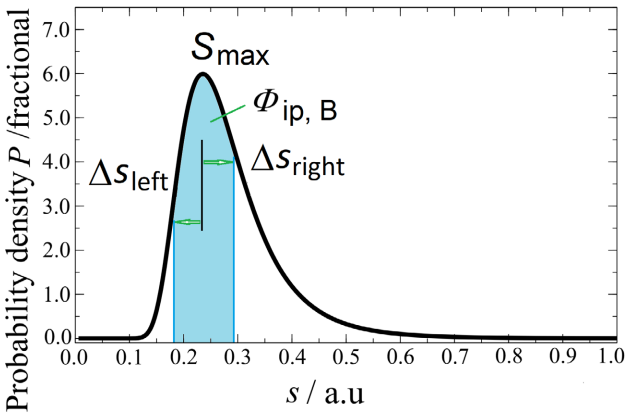

Figure 2. Directly calculated probability density P as function of the parameter s.

The maximum of the probability density P can be found at $s_{max} = 0.235$. This value s_{max} is the estimated standard deviation based on the Bayesian approach.

The directly calculated probability density P is now compared to the results obtained from classical statistics. The standard deviation $s_{classic}$ from classical statistics ([16]) follows from

$$s_{classic} = \sqrt{\frac{\sum_{i=1}^{8}(y_i - 1)^2}{8}} = 0.235. \tag{29}$$

This result $s_{classic} = 0.235$ is identical to the value $s_{max} = 0.235$ obtained from the Bayesian analysis. Since in this case the average μ is known and equals unity, all $n = 8$ degrees of freedom are maintained, and no degree of freedom is used to calculate the average. Consequently, the denominator of Equation (29) uses $n = 8$ instead of the typical $n - 1 = 7$ degrees of freedom. The inflection point on the left side of the maximum s_{max} of the probability density P curve is located at $s_{left} = 0.179$, the inflection point on the right side of the maximum

lies at $s_{\text{right}} = 0.291$ as depicted in **Figure 2**.

The "left error" of the estimated value $s_{\max}$ is determined to:

$\Delta s_{\text{left}} = 0.235 - 0.179 = 0.056$

and its "right error" is calculated to: $\Delta s_{\text{right}} = 0.291 - 0.235 = 0.056$. As the result of the Bayesian approach, the probability $\Phi_{\text{ip,B}}$ between the two inflection points $s = 0.179$ and $s = 0.291$ (*i.e.* the integral over the probability density function P, blue area in **Figure 2**) is calculated to: $\Phi_{\text{ip,B}} = 0.577$.

We approximate $\ln P(s)$ by a truncated Taylor series at $s = s_{\max}$ as expansion point by considering the 2nd derivative, since the 1st derivative is zero at $s_{\max}$. The resulting approximation of $P(s)$ is the Gaussian function depicted in **Figure 3** (red curve). For comparison, the directly calculated probability density function is also shown in the figure as black curve. The probability $\Phi_{\text{ip,B}} = 0.577$ following from the directly calculated distribution (**Figure 2**) is compared to the probability $\Phi_{\text{ip,G}} = 0.682$ between the inflection points, in the case that the standard deviation is approximated by a Gaussian distribution (red curve in **Figure 3**).

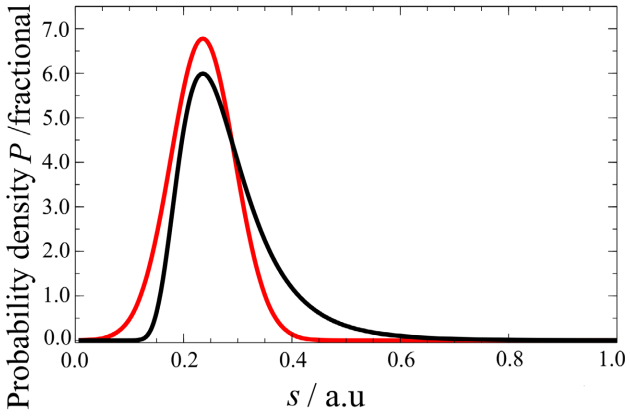

Figure 3. Approximated probability density of the truncated Taylor series (red curve) in comparison with the originally calculated probability P as a function of s (black curve).

The error estimation via Bayesian statistics described above is now compared to the approximation in classical error estimation using the extended objective function, see also Equation (13):

$$Z_1 = -2\ln P\left(\{s\} \mid \{y_1, y_2, \cdots, y_8\}, M\right) - n\ln(2\pi).$$

The estimated error Δs of the standard deviation $s_{\max}$ is deduced from the inverse of the Hessian matrix $\ddot{H}$. In this example $H_{1,1}$ is the only non-zero second derivative of the objective function and the estimated error follows as $\Delta s = \pm 0.042$. Comparing Gaussian Δs with the error $\Delta s = \pm 0.056$ obtained from the Bayes' treatment—for both inflection points, the left and the right one—shows that the values are rather different, however, in the same order of magnitude.

As a third error estimation, the error Δs of the estimated standard deviation $s_{\max}$ follows from the approximation suggested by Lane [21]:

$$\Delta s = \frac{0.71}{\sqrt{n}} s. \tag{30}$$

For our example, this yields:

$$\Delta s = \frac{0.71}{\sqrt{8}} \cdot 0.235 \approx 0.059.$$

This estimated error $\Delta s = 0.059$ is compared to the values determined from the inflection points ($\Delta s = 0.056$) and from the second derivative ($\Delta s = 0.042$), highlighting the differences between these different methods of uncertainty estimation. The calculation using Maple version 18 [22] can be found in GitHub.

The classical Gaussian distribution is inherently a symmetric distribution with non-zero values in the negative s -range. However, the propability density must be zero for values of $s < 0$ since a negative standard deviation is not possible. This is the case for the probability density function used in our approach, which can be seen as an advantage compared to the classical usage of a Gaussian distribution.

3.2. Example 2: Determination of the Average and the Standard Deviation

The same y_i values from example 1 are used as measurement points in example 2, see **Figure 1**. In contrast to example 1, where the model curve was known to be $y_{c,i} = 1$, the model for this example is described by a constant but a priori unknown parameter, $y_{c,i} = y_{\text{average}}$. Both this value y_{average} and the standard deviation $s_{c,i} = s$ are estimated simultaneously by analyzing the measurement points y_i.

The hypothesis space M is defined as

$M = \left\{ \left(y_{\text{average}}, s \right); y_{\text{average}} \in \mathbb{R}, s > 0 \right\} = \mathbb{R} \times \mathbb{R}^+$.

The posterior probability density P of the a priori unknown value $\vec{\xi}_1 = \left\{ y_{\text{average}} \right\}$ and its unknown standard deviation $\vec{\xi}_2 = \{ s \}$ are determined from Equation (15):

$$\begin{aligned} &\ln P\left(\left\{ s, y_{\text{average}} \right\} \mid \{ y_1, \cdots, y_n \}, M \right) \\ &= -\ln K - \frac{n}{2} \ln(2\pi) - \sum_{i=1}^{n} \left\{ \ln(s) + \frac{\left(y_i - y_{\text{average}} \right)^2}{2s^2} \right\}. \end{aligned}$$

In this example, $n = 8$. The evidence $K = P\left(\{ y_i \}_{i=1}^{n} \mid M \right)$ serves as a normalization factor, and its value is found to be $K = 0.0632$. The resulting probability density, which depends on both y_{average} and s, is presented in **Figure 4**. The calculation is done using Maple version 18 [22]. The Maple code can be found in the accompanying GitHub repository [20].

The maximum of the posterior probability density is reached at $y_{\text{average}} = 0.949$ and at a standard deviation $s = 0.231$. These values are the maximum a posteriori (MAP) estimates for the parameters, as depicted in **Figure 4**.

The value $y_{\text{average,classic}}$ obtained using classical statistics is the sample mean:

$$y_{\text{average,classic}} = \frac{\sum_{i=1}^{n} y_i}{n} = 0.941$$

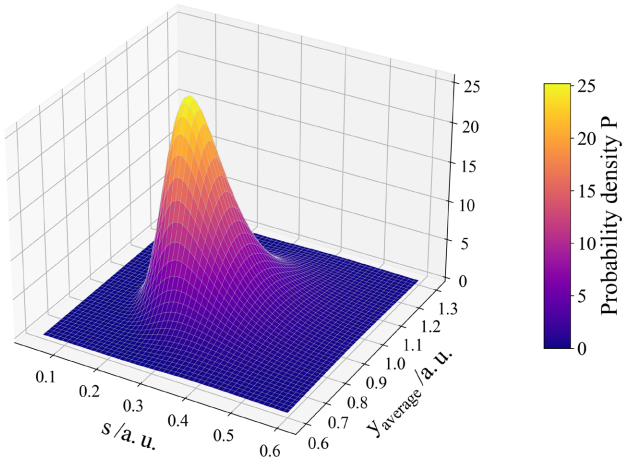

Figure 4. Probability density P as a function of the average y_{average} and standard deviation s.

In this example, the standard deviation s_{classic} from classical statistics is given by the sample standard deviation formula:

$$s_{\text{classic}} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y_{\text{average,classic}}\right)^2}{n-1}} = 0.244 \tag{31}$$

The use of $(n-1)$ degrees of freedom in the denominator is necessary because the value of y_{average} was estimated from the data itself, using one degree of freedom.

The objective function Z_1 from Equation (18) applied to example 2 is

$$Z_1 = Z_1\left(y_{\text{average}}, s\right) = 2n\ln s + \frac{1}{s^2}\sum_{i=1}^{n}\left(y_i - y_{\text{average}}\right)^2.$$

The Hessian matrix $\ddot{H}$ of the objective function Z_1 (Equation (17)) is computed with respect to the parameters y_{average} and s:

$$\ddot{H} = \begin{bmatrix} 300.9 & -20.2 \\ -20.2 & 582.9 \end{bmatrix}$$

The covariance matrix $\ddot{\Sigma}$, which is the inverse of the Hessian matrix $\ddot{H}$, is given by:

$$\ddot{\Sigma} = \begin{bmatrix} 3.33\times10^{-3} & 1.15\times10^{-4} \\ 1.15\times10^{-4} & 1.72\times10^{-3} \end{bmatrix}$$

The error of the average is $\Delta y_{\text{average}} = \sqrt{s_1^2} = s_1 = 0.058$; and the error of the standard deviation is $\Delta s = \sqrt{s_2^2} = s_2 = 0.041$. The correlation $\rho_{y_{\text{average}},s}$ between the estimated average y_{average} and the estimated standard deviation s is then calculated as:

$$\rho_{y_{\text{average}},s} = \frac{s_{1,2}}{s_1 s_2} = 0.048$$

This value of $\rho_{y_{\text{average}},s} = 0.045$, being close to zero, indicates that no significant linear correlation exists between the two parameters.

In summary, the estimated average is $y_{\text{average}} = 0.949 \pm 0.058$, and the standard deviation is $s = 0.231 \pm 0.048$.

3.3. Example 3: Addressing Weighting Challenges in Polynomial Fits (Model 1)

The regression of datasets with unknown uncertainties of the individual measurement points is a challenging problem. In our example 3, polynomial functions are used to fit data over a wide range. It is inherit in this problem that the values of the a-priori unknown weights of the data points can vary over several orders of magnitude. It illustrates how our extended least squares technique provides a robust uncertainty estimation even under conditions of strongly varying weighting factors. It becomes clear that constant weighting factors would not properly capture the regression problem.

The data set used in this third example are synthetically generated. We use the following linear model:

$$y_i = 1 + 0.5 x_i + \varepsilon_i. \tag{32}$$

Here, x_i spans the range $[-10, 85]$ with a step size of 5, resulting in $n = 20$ generated values. The measurement error ε_i assumed follows a centered Gaussian distribution with a standard deviation σ_i that deliberately depends on the expected measurement value, given by:

$$\sigma_i = 2 + 0.05 \cdot |x_i|. \tag{33}$$

This choice for σ_i intentionally introduces a heteroskedastic error structure (*i.e.*, a variance of measurement errors that is not constant across all measurements), where uncertainties increase with the magnitude of the measured value. This is a characteristic often observed in real-world experimental data and is a crucial aspect because standard least squares methods often assume homoskedasticity (*i.e.*, constant error variance across all measurements), which can lead to inefficient parameter estimates and unreliable uncertainty quantification when applied to such heteroskedastic data.

This generated set of n data pairs (x_i, y_i) is fitted using a linear function $y_{c,i} = a_0 + a_1 x$, where $y_{c,i}$ denotes the regression value at $x = x_i$:

$$y_{c,i} = a_0 + a_1 \cdot x_i. \tag{34}$$

As already mentioned a key aspect of our approach is that the parameters of the linear fit (a_0, a_1) and the parameters of the uncertainty model (s_0, s_1) are estimated simultaneously within a single optimization process. The standard deviations $s_{c,i}$ of the measured data points y_i are modeled as a function of the fitted values $y_{c,i}$ according to:

$$s_{c,i} = s_0 + s_1 \cdot |y_{c,i}|. \tag{35}$$

Here, s_0 is the intercept and represents the estimated standard deviation at $y_{c,i} = 0$. The parameter s_1 is the slope and characterizes the linear increase of the estimated standard deviation as the absolute value of $y_{c,i}$ increases. Both s_0

and s_1 are positive parameters, *i.e.*, $s_0, s_1 \in \mathbb{R}^+$, which are obtained during the optimization process as described in Equation (18). The overall hypothesis space M for this fit is thus defined by: $a_0, a_1 \in \mathbb{R}$ and $s_0, s_1 \in \mathbb{R}^+$. The formulations for $y_{c,i}$, Equation (34), and $s_{c,i}$, Equation (35), are denoted as model 1.

It is worth noting that the standard deviation parameter s_0 from Equation (35) must be positive and does not go to zero. This ensures that both the weighting factors and the natural logarithm $\ln(s)$ of the standard deviation remain finite. The parameters a_0, a_1, s_0, s_1 (which define the models in Equations (34) and (35) are determined by minimizing the objective function Z_1 (Equation (18)), as summarized in Table 1. This function represents the negative logarithm of the joint posterior probability density of the parameters, making the minimization equivalent to finding MAP estimation. It is important to note that even though a linear regression model is used, the objective function Z_1 is nonlinear due to the simultaneous estimation of the uncertainty parameters. The estimated parameter values a_0, a_1, s_0, and s_1 result from this optimization procedure, and their uncertainties are derived from the diagonal entries of the inverse Hessian matrix, Equation (23).

Table 1. Estimated parameter values and one-σ uncertainties for model 1.

Parameter	Value
a_0	1.77 ± 0.50
a_1	0.468 ± 0.017
s_0	1.72 ± 0.42
s_1	0.087 ± 0.026

The fitted linear model (red dashed line) and the generated test data set (black dots) show good agreement, as illustrated in Figure 5.

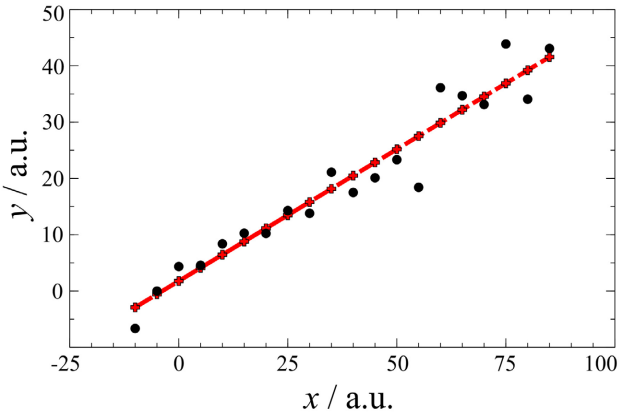

Figure 5. The generated test dataset (black points) is compared to the fitted linear model (red dashed line) from model 1, which represents the most probable course of the data based on our analysis.

The correlations ρ_{ij} between the model parameters are calculated by means of the covariance matrix $\vec{\Sigma}$ according to Equation (24) and are listed in Table 2.

Table 2. Correlation matrix ρ_{ij} between model parameters a_0, a_1 and the Gaussian uncertainties s_0, s_1 of model 1.

	a_0	a_1	s_0	s_1
a_0	1	−0.522	−0.043	0.031
a_1	−0.522	1	−0.068	−0.020
s_0	−0.043	−0.068	1	−0.647
s_1	0.031	−0.020	−0.647	1

The regression parameters for the intercept a_0 and the slope a_1 are significantly negatively correlated with each other ($\rho_{a_0,a_1} = -0.522$), as are the uncertainty parameters s_0 and s_1 ($\rho_{s_0,s_1} = -0.647$).

The uncertainty of the standard deviation curve, Δs_c is obtained by propagating the parameter uncertainties, which are derived from the inverse Hessian matrix [23]. In the Bayesian context of our analysis, this propagation provides the credible intervals for the model curve. This approach, while mathematically similar to error propagation in classical statistics, allows for a direct interpretation of the parameter uncertainties as probability distributions,

$$\begin{aligned}\Delta s_c^2 &= \left(\frac{\partial s(s_0,s_1)}{\partial s_0}\right)^2 \Delta s_0^2 + \left(\frac{\partial s(s_0,s_1)}{\partial s_1}\right)^2 \Delta s_1^2 \\ &\quad + 2\left(\frac{\partial s(s_0,s_1)}{\partial s_0}\right)\left(\frac{\partial s(s_0,s_1)}{\partial s_1}\right)\cdot \rho_{s_0,s_1}\cdot \Delta s_0 \cdot \Delta s_1 \\ &= \Delta s_0^2 + y_c^2 (\Delta s_1)^2 + 2|y_c|\cdot \rho_{s_0,s_1}\cdot \Delta s_0 \cdot \Delta s_1.\end{aligned} \tag{36}$$

It is shown in **Figure 6** that the determined residuals are consistent with the modeled standard deviation curve $\pm s_c$, *i.e.* with the highest probability density for the errors (see red curves in **Figure 6**). The uncertainty of the highest probability density of the errors $\pm s_c$ is depicted by a red shaded regions.

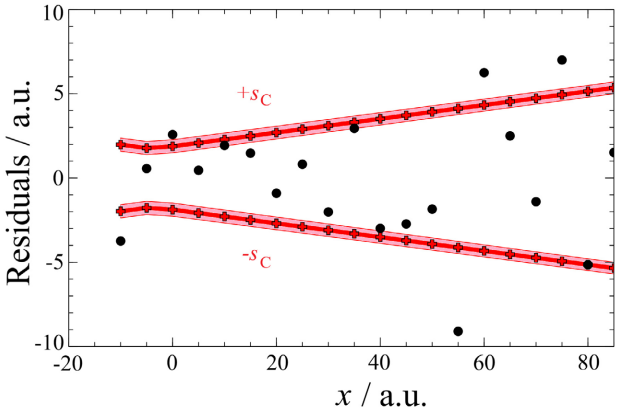

Figure 6. The residuals of model 1 are plotted versus the independent variable x. The red line represents the most probable course of the modeled standard deviation s_c, which is the MAP estimation. The shaded red regions represent the credible interval (or $\pm 1\sigma$) for the standard deviation curve, illustrating its uncertainty.

The large relative uncertainty of the intercept parameter, a_0, combined with the strong anti-correlations between several model parameters, suggests that the four-parameter model is over-parameterized. This indicates a potential for over-fitting, where the model is too complex for the given data. Consequently, while the model has a high explanatory power for the training data, this comes at the cost of a significant loss of predictive power for new, unseen data.

A simpler model with fewer optimization parameters may therefore be more appropriate. Since the uncertainty parameter s_0 must be retained to prevent a singularity in the weighting coefficient w_i (Equation (10)) for $y_c = 0$, the model's intercept parameter, a_0, is the logical choice to be excluded. In example 4, we investigate the implications of this model deficiency by applying a fitting function without the constant intercept term to the same test data. This new fitting approach intentionally deviates from the created data set, introducing a known systematic error due to the model's simplified form.

3.4. Example 4: Polynomial Fit with an Unknown Systematic Error Due to Model Simplification (Model 2)

The effect of an unknown systematic error is examined in example 4 (model 2). For this purpose the test data from example 3 are used (black points in Equation (5)). The systematic error is caused by skipping the parameter a_0 (see Table 3) of example 3. Thus, the following model function, Equation (37) is applied to the test data:

$$y_{c,i} = a_1 \cdot x_i. \tag{37}$$

The Gaussian uncertainty $s_{c,i}$ of the data points is estimated with the same equation as for example 3, *i.e.* Equation (38):

$$s_{c,i} = s_0 + s_1 \cdot \left| y_{c,i} \right|. \tag{38}$$

The hypothesis space M is defined as: $a_1 \in \mathbb{R}$, s_0 and $s_1 \in \mathbb{R}^+$. The formulations for $y_{c,i}$, Equation (37), and $s_{c,i}$, Equation (38), are denoted as model 2.

The fit procedure described here results in the following values for the model parameters and their uncertainties (Table 3):

Table 3. Estimated parameter values and one-σ uncertainties for model 2.

Parameter	Value
a_1	0.499 ± 0.015
s_0	2.55 ± 0.52
s_1	0.058 ± 0.027

The residuals are calculated as the difference of the values from model 2 with the fit parameters from Table 3 and the test data. The results are depicted in Figure 7. It is worth mentioning that the residuals are higher than those obtained in example 3 (compare Figure 6) in particular for small x-values. However, only one primary model parameter is engaged in this model (example 4), *i.e.* the slope a_1.

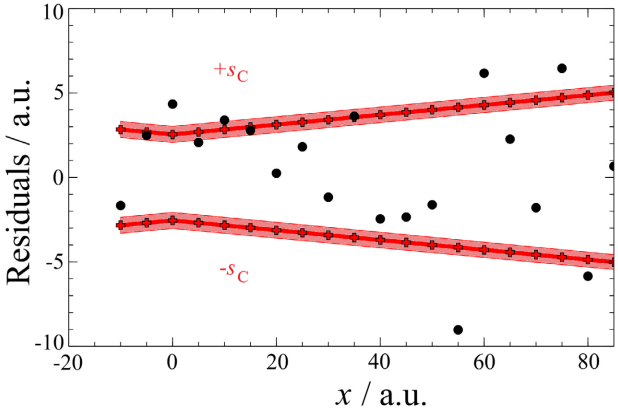

Figure 7. The residuals of model 1 are plotted versus x with an unknown systematic error due to model simplification and the course of the modeled standard deviation is represented by red pluses connected with a red line (the red regions represents the uncertainties Δs_i).

The residuals in Figure 7 are influenced by two components: the inherent noise of the measurement and a systematic error introduced by not considering a possible non-zero axis intercept within the model. As expected, the sum of the unweighted residual squares for model 2:

$$\sum_{i=1}^{n}\left[y_i - y_{c,i}\left(\vec{\xi}_1\right)\right]^2 = 62.2$$

is higher than for model 1, where a value of 57.9 is obtained. Consequently, the curve of the calculated standard deviation s_c deviates more strongly from zero over the entire range for model 2 than for model 1, as shown in Figure 6 and Figure 7.

The comparison of model 1 and model 2 highlights a key challenge in model selection: choosing the most appropriate model. When choosing between models, a good strategy is to apply Occam's razor, using the simplest possible model that fits the data adequately. This principle suggests that the three-parameter model 2 might be a better choice than the more complex four-parameter model 1, especially if the advantages of its simplicity, such as being less prone to overfitting, compensate for the introduced systematic error.

The correlations ρ_{ij} between the model parameters are calculated by means of the covariance matrix $\ddot{\Sigma}$ according to Equation (24) and are provided in Table 4.

Table 4. Correlation matrix ρ_{ij} between the model parameter a_1 and the Gaussian uncertainties s_0 and s_1 of model 2.

	a_1	s_0	s_1
a_1	1	−0.086	0.036
s_0	−0.086	1	−0.681
s_1	0.036	−0.681	1

The regression parameter a_1 shows no significant correlation with the uncertainty parameters s_0 and s_1 ($\rho_{a_1,s_0} = -0.086$, $\rho_{a_1,s_1} = 0.036$). This indicates that the values of a_1 are largely independent of the uncertainty parameters. A strong negative correlation between s_0 and s_1 is still present, however, similar to model 1.

3.5. Example 5: Addressing the High Parameter Correlation of Example 4 (Model 3)

The results of example 4 demonstrate that the uncertainty of the parameter s_1 is in the same order of magnitude as the parameter itself. In addition a strong anti-correlation between s_0 and s_1 occurs, $\rho_{s_0,s_1} = -0.681$. Thus, the number of fit parameters has to be further reduced. The same model function, Equation (39) as used above is again applied to the test data:

$$y_{c,i} = a_1 \cdot x_i. \tag{39}$$

However, the Gaussian uncertainty s of the data points is now the same constant for all measuring points *i.e.*

$$s_{c,i} = s_0. \tag{40}$$

The hypothesis space M is defined as: $a_1 \in \mathbb{R}$ and $s_0 \in \mathbb{R}^+$.

The optimized fitting parameters are summarized in Table 5. The standard deviations $\pm s_c$ are constant for the whole measurement region, see Figure 8. Thus, the uncertainty of the standard deviation is also constant (red shaded regions in Figure 8).

Table 5. Estimated parameter values and one-σ uncertainties for example 5.

Parameter	Value
a_1	0.497 ± 0.013
s_0	3.79 ± 0.42

The correlation coefficient ρ_{a_1,s_0} is now close to zero, $\rho_{a_1,s_0} = 4.61 \cdot 10^{-10}$, *i.e.* there is no correlation between a_1 and s_0 (see Table 6).

Table 6. Correlation matrix ρ_{ij} between the model parameter a_1 and the gaussian uncertainty s_0 of example 5.

	a_1	s_0
a_1	1	4.61×10^{-10}
s_0	4.61×10^{-10}	1

The residuals and the course of the standard deviation for example 5 are plotted in Figure 8. Since both, the model function and the model of the Gaussian uncertainty are strongly simplified (the remaining parameters are s_0 and a_1 only), systematic errors occur. The strong simplification of the model to only two pa-

rameters (a_1 and s_0) introduces systematic errors, leading to higher residuals. However, these drawbacks are acceptable as the primary objective, the reduction of parameter correlation, is successfully achieved. The result is a robust model in which the standard deviation is sufficiently described and the correlation between the parameters is negligibly small.

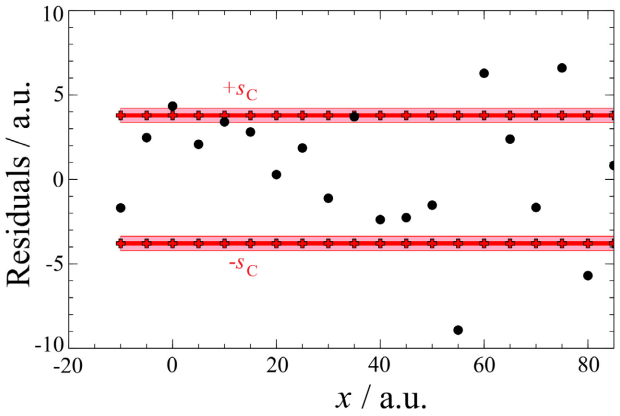

Figure 8. The residuals of example 5 are plotted against the independent variable x . The course of the modeled standard deviation is represented by red pluses connected with a dashed line, and the red regions represent the uncertainties Δs_i .

In classical statistics the weighting factors are calculated from a multitude of measurements. In our approach the weighting factors are estimated from the residuals. Thereby, measurement uncertainties and errors due to the arbitrary choice of the model are both taken into account.

3.6. Example 6: Microstructural Changes during the Heat Treatment of Chromium Stainless Steels

As a final example, we demonstrate the application of our extended least squares method to a materials science problem. We examine the temperature-dependent change in dislocation density ρ in martensite during heating for the Böhler steel grade, N404 [24]. The dislocation density ρ in martensite during heat treatment is evaluated from in-situ X-ray diffraction data by means of a Rietveld method by using the Double-Voigt model [25] considering the peak broadening caused by the instrument itself and by lattice strain. The data for each heat treatment are monitored in a single in-situ experiment and each point of the dislocation density ρ characterized by time and temperature corresponds to a diffractogram. Details with respect to the evaluation of ρ can be found in [26].

Its chemical composition is provided in **Table 7**. To ensure full reproducibility, the complete Python implementation for model fitting is available in the accompanying GitHub repository [20].

The mechanical properties of this steel, particularly tensile strength and ductility, are strongly dependent on the dislocation density ρ , which is influenced by the final heat treatment. In-situ XRD measurements were conducted to investigate

different heat treatments [26], from which the dislocation density ρ in martensite was determined, as shown in **Figure 9**.

Table 7. Chemical composition of the chromium stainless steel, showing mass fractions (wt%) of the main alloying elements. The remainder is iron (Fe).

Component	C	Si	Mn	Cr	Mo	Ni	N	Fe
Mass fraction (wt%)	0.04	0.40	0.40	15.4	0.90	5.30	0.04	bal.

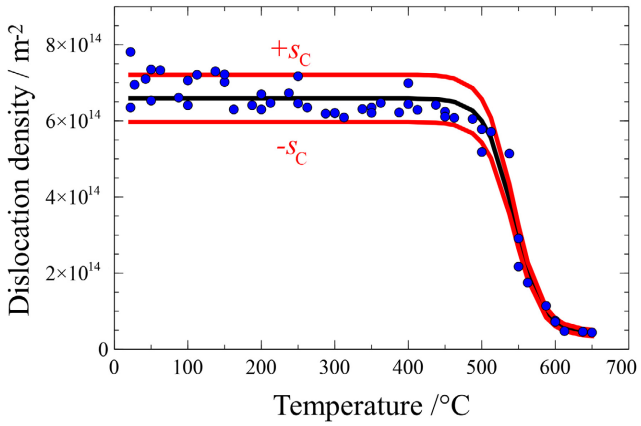

Figure 9. Experimentally determined and modeled dislocation density during the heating in a tempering process. Uncertainties ($\pm 1 s_c$) are estimated by the linear function yielding only positive values for the standard deviation.

In crystalline materials, the reduction of dislocation density during heat treatment is controlled by thermally activated mechanisms such as annihilation, which are described by Arrhenius-type behavior. As temperature increases, the mobility of dislocations rises exponentially, leading to a rapid decrease in dislocation density once a characteristic activation temperature is reached.

Such processes typically exhibit three regimes:

- a low-temperature regime with negligible changes.
- a transition region with a strong temperature dependence.
- a high-temperature regime where the system approaches a saturated state.

The sigmoidal Boltzmann function was selected based on physical considerations of thermally activated processes governing dislocation annihilation.

The sigmoidal Boltzmann function is given by:

$$\rho(\vartheta) = \rho_2 + \frac{\rho_1 - \rho_2}{1 + \exp\dfrac{\vartheta - \vartheta_0}{\Delta\vartheta}}, \tag{41}$$

with ρ_1 and ρ_2 representing the initial and final dislocation density, respectively. The parameters ϑ_0 and $\Delta\vartheta$ are the center temperature and temperature change. The slope $\left(\mathrm{d}\rho/\mathrm{d}\vartheta\right)\big|_{\vartheta_0}$ at the center temperature ϑ_0 is given by:

$$\left(\frac{\mathrm{d}\rho}{\mathrm{d}\vartheta}\right)_{\vartheta_0} = \frac{\rho_2 - \rho_1}{4\Delta\vartheta}. \tag{42}$$

However, it should be noted that the underlying microstructural evolution may involve multiple mechanisms (e.g., recovery, rearrangement, and annihilation of dislocations), each characterized by different activation energies and kinetics. As a result, the actual process cannot be fully described by a single activation energy or a single rate-controlling mechanism.

The adopted sigmoidal function should therefore be interpreted as an effective, phenomenological description that captures the overall transition behavior. In particular, the parameter $\Delta\vartheta$ can be understood as reflecting a distribution of activation energies or overlapping processes occurring over a finite temperature range.

The Boltzmann function seems to provide a physically motivated and sufficiently flexible approximation of the experimentally observed behavior.

The choice of the uncertainty function is primarily guided by statistical considerations. In particular, the residuals normalized by the standard error, *i.e.*, $\left(\rho_i - \rho\left(\vartheta_i\right)\right)/s_c\left(\vartheta_i\right)$, should be approximately Gaussian distributed with zero mean and unit variance. This requirement ensures the consistency of the underlying Gaussian likelihood assumption and avoids systematic over- or underestimation of the uncertainty.

In addition, the dislocation density must be a non-negative quantity. Therefore, the model for its uncertainty must ensure that the corresponding credible intervals do not extend into physically meaningless, negative values.

We evaluated several functions for the standard error s_c. A simple approach is equal weighting:

$$s_c = s_0. \tag{43}$$

However, this function can yield unphysical, negative values for the dislocation density within the credible interval at high temperatures (e.g., 650˚C). As an alternative, we evaluated a function where the standard error is related to the square root of the dislocation density:

$$s_c = s_0 + s_1 \cdot \sqrt{\rho\left(\vartheta\right)}. \tag{44}$$

This approach also yielded possible negative values for ρ within the credible interval at high temperatures.

As a third alternative, a linear relationship between the standard error and the dislocation density was assumed. This function ensures that the credible interval for ρ remains positive:

$$s_c = s_0 + s_1 \cdot \rho\left(\vartheta\right), \tag{45}$$

where the densities ρ are positive within s_c, see **Figure 9**.

For this Gaussian uncertainty function $\pm s_c$ and the model parameters, the hypothesis space M is defined as: $\rho_1 \in \mathbb{R}^+$, $\rho_2 \in \mathbb{R}^+$, $\vartheta_0 \in \mathbb{R}^+$, $\Delta\vartheta \in \mathbb{R}^+$, $s_0 \in \mathbb{R}^+$, and $s_1 \in \mathbb{R}^+$.

The model parameters, the parameters of the uncertainty function and their errors are summarized in **Table 8**. It should be mentioned that the uncertainty of the parameter s_0 is in the range of the parameter itself. However, the Gaussian

uncertainty function seems to be a reasonable choice as can be seen in Figure 9 by comparing the experimental data with the $\pm s_c$ lines of the Gaussian uncertainty function. The width of the red curves corresponds to the uncertainties of the Gaussian uncertainty functions $\pm 1 s_c$ calculated from Equation (45).

Table 8. Estimated parameter values and one-σ uncertainties for example 6.

Parameter	Value
ρ_1 / m^{-2}	(6.59 ± 0.06) 10^{14}
ρ_2 / m^{-2}	(0.40 ± 0.04) 10^{14}
ϑ_0 / °C	543.2 ± 1.4
$\Delta\vartheta$ / K	19.2 ± 1.3
s_0 / m^{-2}	(0.04 ± 0.05) 10^{14}
s_1	0.078 ± 0.015

The errors from Table 8 are determined during the calculation of the covariance matrix $\ddot{\Sigma}$. Finally, the correlation $\ddot{C}$ between the model parameters is calculated using Equation (24) from the covariance matrix $\ddot{\Sigma}$ and is shown in Table 9.

Table 9. Correlation matrix $\ddot{C}$ of the model and uncertainty parameters for example 6.

	ρ_1	ρ_2	ϑ_0	$\Delta\vartheta$	s_0	s_1
ρ_1	**1.00**	−0.06	−0.34	0.14	0.06	−0.09
ρ_2	−0.06	**1.00**	−0.10	−0.75	0.33	−0.31
ϑ_0	−0.34	−0.10	**1.00**	−0.29	−0.13	0.12
$\Delta\vartheta$	0.14	−0.75	−0.29	**1.00**	−0.42	0.40
s_0	0.06	0.33	−0.13	−0.42	**1.00**	−0.92
s_1	−0.09	−0.31	0.12	0.40	−0.92	**1.00**

Some parameters in Table 9 exhibit notable correlations, the parameters $\Delta\vartheta$ and ρ_2 are anti-correlated with a value of −0.75, that of $\Delta\vartheta$ and ϑ_0 with −0.29. In addition, the parameters of the uncertainty function s_0 and s_1 are also anti-correlated with a value of −0.92.

It is worth mentioning that the absolute error of ρ_c spans several orders of magnitude, ranging from low to high dislocation densities. This means that the choice of the uncertainty function requires care, and a linear function, Equation (45), seems to be appropriate in this case.

4. Conclusions

In this work, a Bayesian informed extended least squares algorithm is presented. It is possible to find an estimate for the combination of measurement uncertainty and a priori unknown systematic errors using this approach. The standard devia-

tions of all measurement points can evolve during regression so that measurement points with large residuals are penalized.

This task is carried out by adding a term to the least squares method. In the limiting case of constant standard deviations of all measurement points, our approach transitions to the classical least squares method.

The theoretical framework presented in this work was applied to five model examples and one materials science problem, demonstrating the following key achievements:

- Examples 1 and 2: A realistic non-symmetric probability density of the standard deviation and the average value follows from our extended least squares method, **Figure 4**.
- Example 3: From a single measurement series, we obtain the course of highest probability and a realistic progression of the standard deviation, **Figure 6**. This statistical treatment is particularly useful when there is only one data set. In addition, experimental efforts can be reduced because of better model prediction.
- Examples 4 and 5: Even in the case of inadequate models, where the model has a small number of fitting parameters, we obtain a realistic course of the standard deviation, **Figure 7** and **Figure 8**. This optimization strategy allows for monitoring the residuals and the correlations between the parameters.
- Example 6: The dislocation density ρ changes by several orders of magnitude during annealing of chromium stainless steel. With the presented method, realistic trends are obtained for the modeled dislocation density and its standard deviations.

The enhanced least squares method presented in this work is a powerful tool for simultaneously calculating model parameters and the standard deviations of model predictions. The method is broadly applicable to a wide range of optimization problems. It is particularly useful when dealing with limited measurement data and in cases where strong fluctuations in the weighting factors occur.

5. Outlook

The presented approach may give rise to further research and applications in various fields:

- While our current approach assumes Gaussian-distributed residuals, it can be extended to handle non-Gaussian distributed residuals, such as Poisson or binomially distributed residuals.
- The objective function proposed in this work allows for analytical solutions to the regression problems, as demonstrated in most of our examples. For applications involving large datasets, the method can be efficiently implemented using numerical algorithms, such as Monte Carlo methods, to address complex real-world problems.
- Our approach provides a robust framework for optimizing data collection strategies. It can be utilized in experimental design to identify the most in-

formative measurements, thereby facilitating more accurate and efficient parameter estimation.

Code and Data Availability

We have implemented the first five problem sets in [22] (examples 1 - 5) and problem set 6 is solved by a Python optimizer. The codes are publicly available on GitHub, and can be found via [20]. Using these codes, it is possible to reproduce the content of the tables and figures.

Author Contribution

Manfred Wiessner, Ernst Gamsjäger, and Benoît Loridant conceived the study and developed the software. The original draft of the manuscript was written by Manfred Wiessner, Ernst Gamsjäger, and Benoît Loridant. Martin Medebach contributed additional ideas. All authors critically reviewed the manuscript and approved the final version.

Conflicts of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] Gagin, A. and Levin, I. (2015) Accounting for Unknown Systematic Errors in Rietveld Refinements: A Bayesian Statistics Approach. *Journal of Applied Crystallography*, **48**, 1201-1211. https://doi.org/10.1107/s1600576715011322

[2] Box, G.E.P. (1979) Robustness in the Strategy of Scientific Model Building. In: Launer, R.L. and Wilkinson, G.N., Eds., *Robustness in Statistics*, Elsevier, 201-236. https://doi.org/10.1016/b978-0-12-438150-6.50018-2

[3] Dumouchel, W. and O'Brien, F. (1989) Integrating a Robust Option into a Multiple Regression Computing Environment. In: Buja, A. and Tukey, P.A., Eds., *Computer Science and Statistics*: *Proceedings of the* 21*st Symposium on the Interface*, Springer-Verlag, American Statistical Association, 297-302.

[4] Holland, P.W. and Welsch, R.E. (1977) Robust Regression Using Iteratively Reweighted Least-Squares. *Communications in Statistics- Theory and Methods*, **6**, 813-827. https://doi.org/10.1080/03610927708827533

[5] Huber, P.J. (1981) Robust Statistics. Wiley. https://doi.org/10.1002/0471725250

[6] Street, J.O., Carroll, R.J. and Ruppert, D. (1988) A Note on Computing Robust Regression Estimates via Iteratively Reweighted Least Squares. *The American Statistician*, **42**, 152-154. https://doi.org/10.1080/00031305.1988.10475548

[7] Singh, K. and Upadhyaya, S. (2012) Outlier Detection: Applications and Techniques. *International Journal of Computer Science Issues*, **9**, 307-323.

[8] Hodge, V. and Austin, J. (2004) A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, **22**, 85-126. https://doi.org/10.1023/b:aire.0000045502.10941.a9

[9] Green, P.J. (1984) Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and Some Robust and Resistant Alternatives. *Journal of the Royal Statistical*

Society Series B: Statistical Methodology, **46**, 149-170. https://doi.org/10.1111/j.2517-6161.1984.tb01288.x

[10] Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W. (2005) Applied Linear Statistical Models. 5th Edition, McGraw-Hill/Irwin.

[11] Sivia, D. and Skilling, J. (2006) Data Analysis: A Bayesian Tutorial. Oxford University Press.

[12] MacKay, D.J.C. (2005) Information Theory, Inference, and Learning Algorithms. Cambridge University Press.

[13] Bolstad, W.M. and Curran, J.M. (2016) Introduction to Bayesian Statistics. 3rd Edition, Wiley. https://doi.org/10.1002/9781118593165

[14] Bayes, T. (1763) LII. An Essay Towards Solving a Problem in the Doctrine of Chances. by the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, **53**, 370-418. https://doi.org/10.1098/rstl.1763.0053

[15] Edwards, A.W.F. (1990) Likelihood. In: Eatwell, J., Milgate, M. and Newman, P., Eds., *Time Series and Statistics*, Palgrave Macmillan, 126-129. https://doi.org/10.1007/978-1-349-20865-4_16

[16] Wolberg, J. (2006) Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments. Springer.

[17] Berger, J.O. (2013) Statistical Decision Theory and Bayesian Analysis. Springer.

[18] Bassett, R. and Deride, J. (2019) Maximum a Posteriori Estimators as a Limit of Bayes Estimators. *Mathematical Programming*, **174**, 129-144. https://doi.org/10.1007/s10107-018-1241-0

[19] Clifford, A.A. (1973) Multivariate Error Analysis: A Handbook of Error Propagation and Calculation in Many-Parameter Systems. Springer.

[20] https://github.com/ManfredWiessner/Bayesian_inspired_framework

[21] Lane, D. (2018) HyperStat Online Statistics Textbook. http://davidmlane.com/hyperstat/A19196.html

[22] Maplesoft (2022) Maplesoft. https://www.maplesoft.com

[23] Ku, H.H. (1966) Notes on the Use of Propagation of Error Formulas. *Journal of Research of the National Bureau of Standards, Section C: Engineering and Instrumentation*, **70**, 263-273. https://doi.org/10.6028/jres.070c.025

[24] Böhler Edelstahl GmbH & Co KG, Product N404. https://www.bohler-edelstahl.com/en/products/n404/

[25] Baizar, D. and Ledbetter, H. (1994) Accurate Modeling of Size and Strain Broadening in the Rietveld Refinement: The "Double-Voigt" Approach. *Advances in X-Ray Analysis*, **38**, 397-404. https://doi.org/10.1154/s0376030800018048

[26] Wiessner, M., Gamsjäger, E., van der Zwaag, S. and Angerer, P. (2017) Effect of Reverted Austenite on Tensile and Impact Strength in a Martensitic Stainless Steel—An *In-Situ* X-Ray Diffraction Study. *Materials Science and Engineering: A*, **682**, 117-125. https://doi.org/10.1016/j.msea.2016.11.039