Maximum-likelihood curve-fitting scheme for experiments with pulsed lasers subject to intensity fluctuations

Thomas Metz, Joachim Walewski, and Clemens F. Kaminski

Evaluation schemes, e.g., least-squares fitting, are not generally applicable to any types of experiments. If the evaluation schemes were not derived from a measurement model that properly described the experiment to be evaluated, poorer precision or accuracy than attainable from the measured data could result. We outline ways in which statistical data evaluation schemes should be derived for all types of experiment, and we demonstrate them for laser-spectroscopic experiments, in which pulse-to-pulse fluctuations of the laser power cause correlated variations of laser intensity and generated signal intensity. The method of maximum likelihood is demonstrated in the derivation of an appropriate fitting scheme for this type of experiment. Statistical data evaluation contains the following steps. First, one has to provide a measurement model that considers statistical variation of all enclosed variables. Second, an evaluation scheme applicable to this particular model has to be derived or provided. Third, the scheme has to be characterized in terms of accuracy and precision. A criterion for accepting an evaluation scheme is that it have accuracy and precision as close as possible to the theoretical limit. The fitting scheme derived for experiments with pulsed lasers is compared to well-established schemes in terms of fitting power and rational functions. The precision is found to be as much as three times better than for simple least-squares fitting. Our scheme also suppresses the bias on the estimated model parameters that other methods may exhibit if they are applied in an uncritical fashion. We focus on experiments in nonlinear spectroscopy, but the fitting scheme derived is applicable in many scientific disciplines.

OCIS codes: 000.5490, 000.4430, 300.6360, 120.3940.

1. Introduction

In an experiment in which one measures the functional dependence of output variables on certain input variables, this dependence is governed by physical quantities, some of which may be unknown (measurands) and whose estimation from the experimental data is desired.

We treat examples from laser spectroscopy for which a signal (the output) is generated by interactions of matter with radiation from pulsed laser sources (the input). Here, the measured values of input and output vary not only because of detector noise but also because of pulse-to-pulse fluctuations of the laser intensity. We derive an appropriate evaluation scheme for this class of experiment for situations in which the output depends linearly on the input as well as for nonlinear dependence.

The general approach to deriving an appropriate evaluation scheme consists of three steps: (a) establish a measurement model, (b) derive or choose estimators for the measurands, and (c) characterize the estimators. The measurement model is a formal, mathematical description of the experiment, usually in statistical terms, because of random variations in the measurement data. In step (b) one usually seeks to find optimal estimators with respect to certain criteria such as unbiasedness or maximum efficiency (least standard uncertainty). (Terms for the properties of estimators are explained elsewhere.[1–3]) For this purpose step (c), a characterization of the estimators in terms of expectations, variances, etc. is required. A characterization is also necessary for specifying the

T. Metz (thomas.metz@forbrf.lth.se) and J. Walewski are with the Division of Combustion Physics, Lund Institute of Technology, P.O. Box 118, SE-221 00 Lund, Sweden. C. F. Kaminski is with the Department of Chemical Engineering, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, UK.

Received 10 December 2001; revised manuscript received 3 November 2002.

0003-6935/03/091551-13$15.00/0

uncertainty of the estimates when one is presenting the results.

The finding of optimal estimators, as well as their characterization, depends in a crucial manner on the measurement model. Therefore it is important to establish a model that describes the actual experiment as accurately as possible. If an estimator is applied to an experiment that is not embraced by the measurement model for which the estimator originally was derived, the achieved precision or accuracy—or even both—may not be optimal. This is true, for example, for the simple least-squares fitting commonly applied to the class of experiment treated in this paper, as this scheme does not apply to data for which the output is statistically correlated to the input.

We derive an appropriate fitting scheme by modeling, in step (a), both the measurement errors (detector noise) and the laser intensity fluctuations as stochastic variables, whereas, in commonly applied least-squares schemes, only the measurement errors are modeled as such. The former description is more appropriate, as all the knowledge about the statistical nature of the random variability of the data is implemented in the evaluation. Therefore the resultant estimators can be expected to be of optimal precision.

To find optimal estimators in step (b) we apply the method of maximum likelihood (ML) to our model. Finally, in step (c) we characterize the obtained estimators by their expectations and variances, which we obtain from Monte Carlo simulations. In this way we can also compare our method and other fitting schemes.

The method of ML was proposed by Fisher.[4] It provides estimators that are not necessarily unbiased, but they are most efficient if most-efficient estimators exist.[3] Otherwise, they are at least asymptotically most efficient when the number of replicates is increased. Moreover, they are consistent. In many cases the method results in data adjustment or curve-fitting procedures, the most prominent of which are variants of the least-squares (LSQ) method.

The method of LSQ dates back to Legendre, Gauss, and Laplace[5] and has since been the subject of many publications. As long as the model function is linear in the unknowns, an analytical solution for the fitting exists, whereas the general, nonlinear case requires an iterative numerical procedure. A fairly general and elegant description of the nonlinear case has been known for several decades[6–10] but does not provide ML estimators for the type of experiment treated in this paper. For this reason, we present a more appropriate fitting scheme, to which we refer as covariant weighted LSQ (covariant WLSQ) fitting. This method provides estimates of unknown parameters with an excellent accuracy and with virtually the smallest standard uncertainty obtainable from the experimental data. Moreover, it can be used for the evaluation of a variety of experiments in physics and other sciences. The computer routines developed as part of this study are available free of charge from the authors (see Section 5 below).

The outline of this paper is as follows: In Subsection 2.A we begin with some general remarks about the importance of an appropriate measurement model, and in Subsection 2.B we describe our measurement model for laser spectroscopic experiments with pulsed lasers. Then we apply the method of ML to derive an expression that has to be minimized during the fitting (Subsection 2.C).

We proceed by demonstrating the method for different types of laser-spectroscopic experiment. We use Monte Carlo simulations to characterize and to compare the method with other fitting procedures, namely, simple LSQ and WLSQ fitting and a technique to which we refer as log–log fitting (Sections 3 and 4). Finally, in Section 5 we summarize and draw conclusions from our research.

A list of symbols used in this paper can be found in Appendix A. In Appendix B we provide short introductions to the fitting schemes discussed in this paper, and in Appendix C we describe the method of Monte Carlo simulations. Appendix D explains the data plotting technique that we use for presentation of our results in Section 4.

2. Theory

A. Statistical Inference

Here we give some general remarks about statistical inference and the importance of the measurement model.

The aim of a measurement is generally to infer the values of unknown quantities (measurands). The applied experimental apparatus, however, seldom provides the measurands directly but instead produces data related to them. From the data, the experimenter calculates estimates of the measurands, using some evaluation scheme. Because of the inevitable random variability in the data, the estimated values would show some random variation if the experiment were carried out several times or by different experimenters. This variability in the estimates implies an uncertainty about the true values of the measurands. Moreover, uncertainty can also arise from uncertainties of quantities used in the evaluation, e.g., experimental parameters or physical constants that are known only to a certain degree.

Both uncertain and randomly varying quantities are described mathematically by stochastic variables and so are the results of the evaluation, for which the associated stochastic variables are called estimators. The values that the estimators assume for a particular data set are the estimates. The evaluation scheme defines the estimators by mathematical functions and procedures through which the uncertainties and variations of the involved quantities propagate to the estimators, which determines their statistical properties, such as their distribution functions, variances, and expectations. From these the uncertainty of the estimates can be derived. Thus a characterization of the estimators in such terms is

necessary for description of the uncertainty of a measurement.

In addition, the experimenter usually seeks optimal estimators with respect to certain criteria such as unbiasedness, maximum efficiency, consistency, and robustness.[1–3] Also, for this purpose a characterization of the estimators is required.

For a characterization, the experiment has to be described in a formal way, namely, by a measurement model. This model comprises a physical model that describes the relations between physical quantities, and a statistical model that describes all uncertain or randomly varying quantities. The quantities that appear in the model are the unknown true values of the measurands, known quantities, uncertain quantities, and randomly varying quantities. The two former are modeled by numerical parameters, whereas the two latter are modeled by stochastic variables. Finally, the model contains the data as stochastic variables whose properties depend on the quantities mentioned above.

Of course, the measurement model has to describe the experimental situation as realistically as possible. Only if all experimenters agree on which model fulfills this requirement can the estimators be characterized properly, various estimators be compared, and the optimal one be found. Otherwise, estimators proposed by different experimenters are not mutually comparable, because an estimator that is optimal for one model might not be optimal for another one. So different experimenters will come to different conclusions about the quality of the estimators.

Once an appropriate measurement model is agreed on, estimators for the same measurand can be compared. Moreover, the adequacy of various models can be judged when existing evaluation schemes are applied based on models that are different from the one that describes the actual experiment. The criteria for such a judgment are, of course, the aforementioned properties of the resulting estimators in the light of the agreed model. For the types of experiment treated in this paper, we discuss such an alternative model.

The above considerations imply the following steps for each evaluation, which we follow here: (a) define the appropriate measurement model, (b) choose estimators for the measurands, and (c) characterize the estimators in light of the measurement model.

B. Measurement Model

In this and in Subsection 2.C we derive the covariant WLSQ fitting scheme with bias correction. Note that nothing more is done than the formulation of an appropriate measurement model for the experiment and consequent application of the method of ML with some natural approximations such as the use of the central-limit theorem or a Taylor expansion.

The measurement model used in this research is illustrated in Fig. 1. The output energy from a pulsed laser source is measured on Detector 1 before being passed into an interaction region. The signal

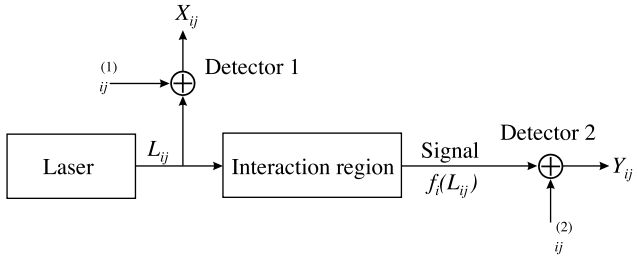

Fig. 1. Block diagram of the setup for a laser-spectroscopic experiment. The symbols shown in this figure are explained in Subsection 2.B.

generated in the interaction region is registered by Detector 2. The measurements are affected by additive noise, $E_{ij}^{(1)}$ and $E_{ij}^{(2)}$, for Detectors 1 and 2, respectively (The subscripts i and j are explained below.) Moreover, laser intensity L_{ij} varies from pulse to pulse even if the experimental setting is not changed. The statistical distribution of these variations as well as that of the noise are assumed to be stationary.

During the experiment we choose a number n of experimental settings with which to study the relationship among several physical quantities. For example, for the power dependence measurements described in Section 3 below, each setting corresponds to a mean intensity of the laser, which one can change, for instance, by placing filters into the laser beam. But even other parameters such as temperature and pressure in a gas cell, or the whole experimental setup, can be changed from one setting to another.

For each setting, several measurements (replicates) are recorded, from which one can estimate statistical properties of the observed quantities, such as means, variances, and covariances. For a specific setting $i \in \{1, \ldots, n\}$ a cluster containing m_i data points numbered $j \in \{1, \ldots, m_i\}$ is obtained. (For a list of symbols please refer to Appendix A.) L_{ij} is the true laser intensity, which corresponds to the jth data point within the ith cluster. Here the laser intensity is assumed to fluctuate normally about an expectation l_i with a variance σ_{Li}^2:

$$\forall j: \quad L_{ij} \in \mathbb{N}(l_i, \sigma_{Li}^2). \tag{1}$$

The observed laser intensity X_{ij} is assumed to contain additive, normally distributed detector noise $E_{ij}^{(1)}$ with zero mean and variance σ_{1i}^2:

$$X_{ij} = L_{ij} + E_{ij}^{(1)}, \tag{2}$$

$$\forall j: \quad E_{ij}^{(1)} \in \mathbb{N}(0, \sigma_{1i}^2). \tag{3}$$

The dependence of the generated signal on the laser intensity is described by function f_i, and detected signal Y_{ij} contains additive, normally distributed noise $E_{ij}^{(2)}$ with variance σ_{2i}^2:

$$Y_{ij} = f_i(L_{ij}) + E_{ij}^{(2)}, \tag{4}$$

$$\forall j: \quad E_{ij}^{(2)} \in \mathbb{N}(0, \sigma_{2i}^2). \tag{5}$$

Function f_i contains known and unknown parameters, the latter of which are the measurands to be estimated. It can be linear or nonlinear with respect to L_{ij}; the nonlinear case is the main interest of this research. Its dependence on setting index i allows not only for changes in the experimental parameters but also for totally different model functions for each setting. Thus the experimental setup can be different for each setting, and even the results from totally different experiments can be combined in this way.

Note the importance of distinguishing the various sources of random variation, namely, the fluctuations of laser intensity L_{ij} and detector noise $E_{ij}^{(1)}$ and $E_{ij}^{(2)}$. The first is common to the observed laser and signal intensities, X_{ij} and Y_{ij}, respectively, and their variations are therefore correlated. Uncertainties, however, are introduced only by detector noise. To understand this, assume, for a moment, data without noise, such that $X_{ij} = L_{ij}$ and $Y_{ij} = f_i(L_{ij})$, i.e., that all data points lie on the correct curve. This would lead to a perfect curve fit without any uncertainties in the parameters, although laser fluctuations were present.

The model presented here is not restricted to laser spectroscopic experiments but applies to a variety of experiments in which the observed input and output signals of a system are subject to statistical fluctuations of the true input signal.

Before proceeding with derivation of the covariant WLSQ fitting scheme we make some remarks about a conceivable alternative measurement model. Such a model is interesting to justify the application of the nonlinear LSQ scheme discussed elsewhere,[6–10] but we regard our measurement model as more realistic. In Section 3 below, we also discuss the numerical problems allied to applying LSQ and ML fitting to the alternative model. Furthermore, the alternative approach is not expected to provide estimates of higher fidelity than our fitting scheme does.

In this alternative fitting scheme the underlying measurement model assumes data given by some true values of physical quantities plus some random measurement error. The true values fulfill some algebraic relation to the unknowns. This is not the situation described by our model, in which the true laser intensities L_{ij} are stochastic variables characterized by their expectations l_i and their variances σ_{Li}^2, which are the unknowns. The relation between these unknowns and L_{ij} is not an algebraic one but a statistical one, because l_i and σ_{Li} determine not the values of L_{ij} but only their distribution function.

To apply the general LSQ scheme requires that the true laser intensities be modeled by quantities that are algebraically related to a set of unknowns. Because the laser intensities are random and cannot be obtained from the model parameters in a deterministic way, they have to be represented by unknowns themselves. Thus, for the application of the common LSQ scheme, one has to replace stochastic variables L_{ij} by unknown parameters, say, l_{ij}, such that l_i and σ_{Li} lose their meaning and are removed from the model.

The resulting measurement model treats all laser intensities equally and makes no assumptions about their possible values. In reality, their possible values are governed by some probability law and are consequently restricted, in the sense of a most-probable range of values. Moreover, in the experiment this probable range is the same for all points in a cluster but can be different for different clusters. By contrast, this knowledge is not incorporated in the alternative model.

The question remains: Will an evaluation scheme that is optimal for this measurement model also be optimal for the present experimental situation? Inasmuch as our model is an appropriate description of the experiment, there is no reason why such an evaluation scheme would provide more-accurate estimates than an optimal scheme based on our model.

A drawback of this model stems from the large number of unknowns, i.e., l_{ij}, σ_{1i}, and σ_{2i}, which have to be optimized during fitting. Compared with our method presented in Section 3 below, this approach is numerically highly demanding, and the fitting procedure hence is comparably slow.

C. Application of the Method of Maximum Likelihood

1. *Outline*

The most important step of the method of ML is the formulation of the correct likelihood function $\mathcal{L}$ of the experiment. This function contains the unknowns as parameters, for which the method of ML chooses those values that maximize the likelihood function. Often it is instead easier to maximize the log-likelihood function $\ln \mathcal{L}$ or to minimize a function χ^2, which is closely related to it. For the implementation of its minimization on a computer we use numerical standard algorithms in the MATLAB Optimization Toolbox, which is not explained further here.[11]

In the first part of this section linear relationships f_i are treated, as this can be done exactly, whereas the nonlinear case is cumbersome or requires some approximations. The resulting fitting scheme is called the covariant WLSQ method. Compared to simple LSQ and WLSQ, an improvement of the precision is already obtained for the linear case, which means that these fitting schemes are inappropriate not only for nonlinear relationships but also for linear ones if the input (laser intensity) and the output (signal) are statistically correlated to each other. The linear case is also interesting, because nonlinear relationships can often be approximated linearly within the relatively short range where the sample points of each cluster lie.

The likelihood function is basically the joint probability-density function (PDF) of all observed quantities in the experimental data set. If the random variations of different data points are independent, $\mathcal{L}$ is simply the product of the PDFs of the data points. However, the coordinates X_{ij} and Y_{ij} of one point are not statistically independent according to the measurement model in Subsection 2.B, and their

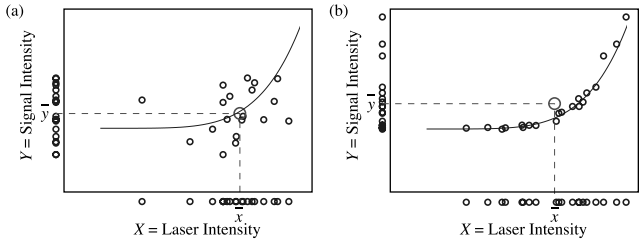

Fig. 2. Schematics of bias from nonlinearity. (a) When no laser fluctuations occur, the variations of observed laser and signal intensities are caused by detector noise. Here both variations are symmetrical, and a point $(\bar{x}, \bar{y})$ obtained from the mean intensities is expected to lie close to the curve that describes the nonlinear dependence. (b) Laser fluctuations contribute essentially to the variation. Although the laser intensities are distributed symmetrically, the signal intensities are not, because of the nonlinear dependence. Hence bias is introduced, as can be seen from point $(\bar{x}, \bar{y})$ lying systematically beside (here, above) the curve.

PDF has to be described by the general, two-dimensional normal distribution including the covariance. This is the main difference between covariant WLSQ and WLSQ fitting.

The nonlinear case becomes important when the laser fluctuations within each cluster are large. Then the distribution of the output Y_{ij} is not normal anymore because of the curvature of f_i and is cumbersome to describe exactly. Therefore we use another approach in which we average the observed input and output values for each cluster and consider the data points thus obtained, i.e., $(\bar{X}_i, \bar{Y}_i)$. According to the central-limit theorem of statistics, the means can approximately be described by a normal distribution, for which, however, the correct parameters must be used. Above all, the expectation of the averaged output, $\mathscr{E}(\bar{Y}_i)$, is not simply $f_i[\mathscr{E}(\bar{X}_i)]$, so the data points from averaging lie systematically beside the real curve that describes the relationship (see Fig. 2).

Because an exact expression for $\mathscr{E}(\bar{Y}_i)$ is not easy to find, we derive it from a second-order Taylor expansion of f_i. This turns out to be satisfactory for the examples discussed in Section 4 below, as bias on the estimates of the unknowns is minimized efficiently in this way. We refer to this treatment of $\mathscr{E}(\bar{Y}_i)$ as bias correction.

2. Linear Relationship

As mentioned above, likelihood function $\mathscr{L}$ is the product of the PDFs of the single, statistically independent data points. This product has to be calculated over all possible pairs of the cluster number (i) and the number of the data point within a cluster (j):

$$\mathscr{L} = \prod_{i=1}^{n} \prod_{j=1}^{m_i} p_{ij}(x_{ij}, y_{ij}), \tag{6}$$

where p_{ij} is the PDF for the data point (X_{ij}, Y_{ij}). With this function one can determine the probability of the data point's lying in a specific region of $\mathbb{R}^2$ (or the x_{ij}–y_{ij}-plane) by integrating p_{ij} over this region.

As follows from the model, i.e., formulas (1)–(5), X_{ij} is normally distributed and so is Y_{ij}, if the function f_i is linear. If X_{ij} and Y_{ij} were independent, p_{ij} could be written as the product of two normal distributions:

$$p_{ij}^{(\text{indep.})}(x_{ij}, y_{ij}) = \frac{1}{[2\pi\mathscr{V}(X_i)]^{1/2}} \exp\left\{-\frac{[x_{ij} - \mathscr{E}(X_i)]^2}{2\mathscr{V}(X_i)}\right\} \times \frac{1}{[2\pi\mathscr{V}(Y_i)]^{1/2}} \exp\left\{-\frac{[y_{ij} - \mathscr{E}(Y_i)]^2}{2\mathscr{V}(Y_i)}\right\}. \tag{7}$$

This would lead to the WLSQ fitting scheme (see Appendix B). Note that the expectations and variances of X_{ij} and Y_{ij} do not depend on j, as their distributions are the same for all j [see formulas (1), (3), and (5)]. To underline this fact we define (X_i, Y_i) as random variables with the same distributions as (X_{ij}, Y_{ij}) and use them to express the expectations, variances, etc.

In this study X_{ij} and Y_{ij} are correlated such that the general, two-dimensional normal distribution[6,12,13] must be used instead of Eq. (7):

$$p_{ij}(x_{ij}, y_{ij}) = \frac{1}{2\pi(\det \boldsymbol{\sigma}_i)^{1/2}} \exp\left(-\frac{1}{2}\, \boldsymbol{\delta}_{ij}^{\mathrm{T}} \boldsymbol{\sigma}_i^{-1} \boldsymbol{\delta}_{ij}\right), \tag{8}$$

$$\boldsymbol{\delta}_{ij} := \begin{bmatrix} x_{ij} - \mathscr{E}(X_i) \\ y_{ij} - \mathscr{E}(Y_i) \end{bmatrix}. \tag{9}$$

$\boldsymbol{\sigma}_i^{-1}$ is the inverse of the covariance matrix of X_{ij} and Y_{ij}, given by

$$\boldsymbol{\sigma}_i := \begin{bmatrix} \mathscr{V}(X_i) & \mathscr{C}(X_i, Y_i) \\ \mathscr{C}(X_i, Y_i) & \mathscr{V}(Y_i) \end{bmatrix}. \tag{10}$$

The readers may convince themselves that Eq. (8) becomes Eq. (7) if X_{ij} and Y_{ij} are uncorrelated, i.e., for $\mathscr{C}(X_i, Y_i) = 0$.

As $f_i(l)$ is assumed to be linear in l, expressions for $\mathscr{E}(X_i)$ and $\mathscr{E}(Y_i)$ can readily be derived from Eqs. (2) and (4):

$$\mathscr{E}(X_i) = l_i, \qquad \mathscr{E}(Y_i) = f_i(l_i). \tag{11}$$

The elements of $\boldsymbol{\sigma}_i$ can be expressed by the model quantities σ_{Li}, σ_{1i}, σ_{2i}, and l_i and the unknown parameters of the model function f_i, say, $a_1, a_2, \ldots$:

$$\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i(\sigma_{Li}, \sigma_{1i}, \sigma_{2i}, l_i, a_1, a_2, \ldots). \tag{12}$$

We do not present the mathematical expressions here, as we shall use another way of estimating $\boldsymbol{\sigma}_i$ (see below).

We now obtain the likelihood function by inserting Eqs. (9) and (11) as well as the expressions that correspond to Eq. (12) into Eq. (8), and then Eq. (8) into Eq. (6). Moreover, x_{ij} and y_{ij} in Eq. (9) have to be replaced by the experimental data. Then $\mathscr{L}$ is a function of the model parameters σ_{Li}, σ_{1i}, σ_{2i}, and l_i and the unknown parameters of f_i:

$$\mathscr{L} = \mathscr{L}(\sigma_{Li}, \sigma_{1i}, \sigma_{2i}, l_i, a_1, a_2, \ldots). \tag{13}$$

All these parameters have to be varied to maximize the likelihood function, $\mathcal{L}$. To reduce this large number of parameters we eliminate the parameters σ_{Li}, σ_{1i}, and σ_{2i} by replacing each of the covariance matrices $\boldsymbol{\sigma}_i$ by its sample estimate $\hat{\mathbf{u}}_i$. This matrix contains the sample estimates of the variances and the covariance[14,15] of the observed values x_{ij} and y_{ij} in the ith cluster. In this way, the so-called pseudo-likelihood is obtained, being a function of l_i and of a_1, a_2, . . . only:

$$\mathcal{L} = \mathcal{L}(l_i, a_1, a_2, \ldots). \tag{14}$$

To keep the computations simple we minimize only the sum of the negative exponents in Eq. (8):

$$\chi^2 := \sum_{i=1}^{n} \sum_{j=1}^{m_i} \boldsymbol{\delta}_{ij}^{\mathrm{T}} \hat{\mathbf{u}}_i^{-1} \boldsymbol{\delta}_{ij}, \tag{15}$$

which corresponds to the maximization of $\mathcal{L}$ as a result of the relationship $\ln \mathcal{L} = \text{constant} - \chi^2/2$. The constant arises from the coefficient in Eq. (8) and is not relevant for the optimization here. Note that it is its constancy that enables standard LSQ algorithms to be used for the implementation of our method; see Section 5 below.

3. Nonlinear Relationship

If function $f_i(l)$ is nonlinear in l and the laser fluctuations are relatively large, the distributions of $f_i(L_{ij})$ and Y_{ij} will be distorted compared with normal distributions. An exact description of the distributions is cumbersome, which is why we use another approach. We consider data points $(\overline{X}_i, \overline{Y}_i)$ obtained from averaging the observed input and output in each cluster:

$$\overline{X}_i := \frac{1}{m_i} \sum_{j=1}^{m_i} X_{ij}, \quad \overline{Y}_i := \frac{1}{m_i} \sum_{j=1}^{m_i} Y_{ij}. \tag{16}$$

Each $\overline{X}_i$ is normally distributed, and so are $\overline{Y}_i$ if the number m_i of points per cluster is sufficiently large, because of the central-limit theorem.[3,13]

Because the number of data points used in the fitting is now reduced to n, the likelihood function is

$$\mathcal{L} = \prod_{i=1}^{n} p_i(\bar{x}_i, \bar{y}_i). \tag{17}$$

The PDF p_i of data point $(\overline{X}_i, \overline{Y}_i)$ can be approximated by a two-dimensional normal distribution similar to Eq. (8):

$$p_i(\bar{x}_i, \bar{y}_i) \approx \frac{1}{2\pi[\det(\boldsymbol{\sigma}_i/m_i)]^{1/2}} \exp\left[-\frac{1}{2} \boldsymbol{\delta}_i^{\mathrm{T}} \left(\frac{\boldsymbol{\sigma}_i}{m_i} \right)^{-1} \boldsymbol{\delta}_i \right], \tag{18}$$

$$\boldsymbol{\delta}_i := \begin{bmatrix} \bar{x}_i - \mathcal{E}(\overline{X}_i) \\ \bar{y}_i - \mathcal{E}(\overline{Y}_i) \end{bmatrix}. \tag{19}$$

Expression $\boldsymbol{\sigma}_i/m_i$ is the covariance matrix of the pair $(\overline{X}_i, \overline{Y}_i)$. The fitting consists of the minimization of

$$\chi^2 = \sum_{i=1}^{n} \boldsymbol{\delta}_i^{\mathrm{T}} \left(\frac{\hat{\mathbf{u}}_i}{m_i} \right)^{-1} \boldsymbol{\delta}_i. \tag{20}$$

As in the linear case, we estimate $\boldsymbol{\sigma}_i$ from sample covariance matrix $\hat{\mathbf{u}}_i$.

The expectation of $\overline{X}_i$ is

$$\mathcal{E}(\overline{X}_i) = \mathcal{E}(X_i) = l_i. \tag{21}$$

For $\mathcal{E}(\overline{Y}_i)$, however, an exact expression is not readily found, as such an expression might comprise integrals that cannot be solved exactly. A good approximation is obtained from a second-order Taylor expansion of f_i in the vicinity of l_i [cf. Eq. (4)]:

$$Y_{ij} \approx f_i(l_i) + f_i'(l_i)(L_{ij} - l_i) + \tfrac{1}{2} f_i''(l_i)(L_{ij} - l_i)^2 + E_{ij}^{(2)}. \tag{22}$$

From this the expectation of $\overline{Y}_i$ results in

$$\mathcal{E}(\overline{Y}_i) = \mathcal{E}(Y_i) \approx f_i(l_i) + \tfrac{1}{2} f_i''(l_i)\sigma_{Li}^2 \approx f_i(l_i) + \tfrac{1}{2} \frac{f_i''(l_i)}{f_i'(l_i)} \mathcal{C}(X_i, Y_i). \tag{23}$$

The expectation of the linear term $f_i'(l_i)(L_{ij} - l_i)$ is zero, because $\mathcal{E}(L_{ij}) = l_i$. In expression (23), σ_{Li}^2 was replaced according to

$$\mathcal{C}(X_i, Y_i) \approx f_i'(l_i)\sigma_{Li}^2, \tag{24}$$

and the covariance is estimated by the corresponding element in sample covariance matrix $\hat{\mathbf{u}}_i$.

Expression (24) results from the second-order Taylor expansion and the assumptions of the measurement model. In particular, the detector noise on Detector 1 should not be correlated to that of Detector 2, and the noise of both detectors should not be correlated to the laser fluctuations. Moreover, the assumed symmetry of the laser fluctuations (normal distribution) makes the third moments of L_{ij} vanish in expression (24). For unsymmetrical fluctuations a modification of this expression might therefore be necessary.

The additional term in expression (23) is caused by the curvature (f_i'') of function f_i and is proportional to the covariance of laser and signal intensity. It is the approximate amount by which the averaged data points are expected to lie above or under the correct curve f_i; see Fig. 2. Neglect of this term would result in bias on the estimates for the unknowns, which we demonstrate in Section 4 below. In this paper the term is referred to as bias correction.

For the linear case, it can be shown that minimizing Eq. (20) is equivalent to minimizing Eq. (15). Therefore and because of its smaller computational expense, we use Eq. (20) for both linear and nonlinear relationships, and we refer to the method as covariant WLSQ fitting with bias correction.

Finally we give a criterion that has to be fulfilled if

WLSQ fitting instead of covariant WLSQ fitting is used. It is obtained from the expectation value of χ^2 in Eq. (20), and it requires that all contributions to this expectation that contain the factors $\mathscr{C}(X_i, Y_i)$ be much less than the remaining ones:

$$\sum_{i=1}^{n} \frac{\mathscr{C}(X_i, Y_i)^2}{m_i \det \boldsymbol{\sigma}_i} \ll \sum_{i=1}^{n} \frac{\mathscr{V}(X_i)\mathscr{V}(Y_i)}{m_i \det \boldsymbol{\sigma}_i}. \tag{25}$$

3. Comparison of the Fitting Methods

We compared the performance of our fitting schemes with those of simple LSQ, WLSQ, and log–log fitting (for descriptions of these methods see Appendix B).

The examples are related to so-called power dependency measurements. Here one records the signal as a function of laser intensity with other experimental parameters held constant. The model functions f that are used in this context are described in Section 4.

The confidence interval of an estimated parameter of f can be obtained analytically.[14] However, for nonlinear model functions this entails systems of nonlinear equations that are difficult to solve. We chose the method of Monte Carlo simulation instead, which is mathematically simple but relies on the high processing speeds of modern computers. With this tool one can study in detail the influence of various quantities, such as detector noise, on the estimated parameters. Readers unfamiliar with this method are referred to Appendix C. The software used for the minimization in the various LSQ fitting approaches is described elsewhere.[11]

Fitting methods based on the alternative model mentioned in Subsection 2.B were not considered in this study for several reasons. The general nonlinear LSQ scheme[6–10] cannot be applied to this model because the scheme requires knowledge about σ_{1i} and σ_{2i} that is not available here. In contrast, our fitting method uses the covariance matrices $\boldsymbol{\sigma}_i$, which can be estimated by their sample counterparts $\hat{\mathbf{u}}_i$. Applying the method of ML to this model circumvents the issue of unknown σ_{1i} and σ_{2i}, and it provides a scheme that is exact apart from the numerical precision of the iteration algorithm and that does not need to handle covariance matrices. However, it requires much longer computational times than the scheme presented here. For instance, when an implementation that used optimization algorithms of the MATLAB Optimization Toolbox was applied to the type of data used in Section 4 below and noise was assumed on both detectors with constant and equal variance, the fitting took more than 100 times longer than with our scheme. The reason for this is the large number of unknown laser intensities l_{ij}, which have to be estimated by the fitting procedure just like the measurands. In the example we had 11 clusters containing 100 data points each, i.e., more than a thousand parameters.

To circumvent the large number of unknowns, one could replace the unknown laser intensities l_{ij} by the measured laser intensities x_{ij} and apply the common LSQ scheme then. This approach reduces the number of fitting parameters substantially, but it results in biased estimates,[16,17] so this attempt would not overcome the numerical drawbacks.

Furthermore, in a ML scheme based on the alternative model, σ_{1i} and σ_{2i} appear as unknowns and have to be varied during the iteration just like the other unknowns $l_{ij}, a_1, a_2, \ldots$, which would increase the computational expense even more. Besides, σ_{1i} and σ_{2i} appear in the pre-exponential factors of the likelihood function of the alternative model, and these factors would therefore vary during the iteration. Hence the evaluation scheme would not be of the LSQ type, and standard LSQ algorithms could not be used for the implementation of the fitting as in this research; cf. Section 5 below.

To avoid σ_{1i} and σ_{2i} as unknowns, they could be obtained from sample covariance matrices by solution of Eq. (12). However, the resulting expressions depend on the parameters $a_1, a_2, \ldots$ and have to be recalculated in each iteration step, again increasing the required computational time and leading the scheme's not being of the LSQ type.

Because of the long computational times, Monte Carlo simulations for such a scheme would have taken too long, which is why a ML scheme based on the alternative model was not included in the comparison here.

4. Results and Discussion

To investigate the accuracy and precision of the different estimators we generated synthetic measurement sets with predefined model parameters (see Appendix C). Synthetic data sets—as opposed to measured sets—permit the calculation of bias of the various estimators, as the correct model parameters are known. Walewski *et al*. obtained realistic values for standard deviations σ_{Li}, σ_{1i}, and σ_{2i} (Subsection 4.B), laser intensities l_i (in the same subsection), and the model parameters by mimicking real measurement data.[18] Each cluster contains $m_i = 100$ measurement points (replicates). Thousand of such synthetic data sets were generated; fitting schemes discussed in Subsection 4.C and Appendix B were applied to each set. Two classes of model function have been studied in this research: power functions and rational functions.

To keep the following discussion easy to grasp, we show only the attained distributions of the estimated parameters. Anyhow, for a thorough characterization of estimators their mutual correlation, i.e., their covariances, also has to be considered, especially when the estimates are applied for prediction of measurement data (see the ISO guide[15,19]). These covariances can easily be estimated from sets of estimated parameters obtained from the Monte Carlo simulation. For a linear dependence on the varying l the covariance matrix of the estimates can also be derived analytically with a formula presented by Celmiņš.[8]

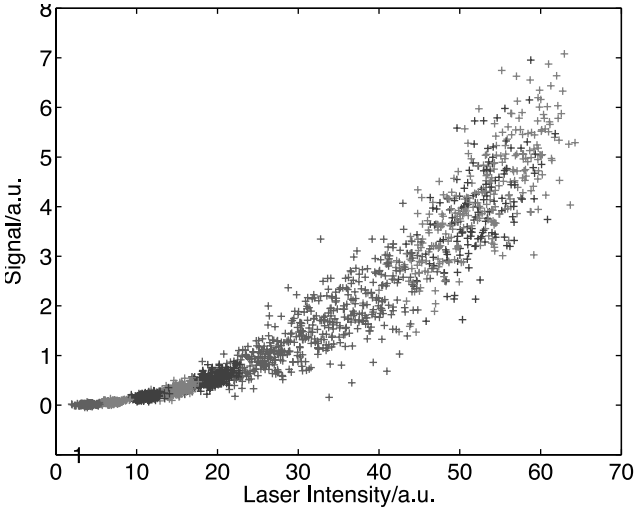

Fig. 3. Example of a synthetic data set generated with parameters for the laser intensities and the uncertainties that were obtained from a fit to measurement data by Walewski *et al.*[18] There are 11 clusters, with $m_i = 100$ data points each. The clusters are distinguished by different gray levels of the symbols. Thousand of such sets were produced by Monte Carlo simulation and evaluated with fitting schemes described in Subsection 2.C and Appendix B.

A. Power Function

Power functions can be found nearly everywhere in the field of physics. Examples in laser spectroscopy are unsaturated multiphoton processes such as multiphoton laser-induced fluorescence and higher-harmonic generation.[20] The model function is written as follows:

$$f(l, a, b) = al^b. \tag{26}$$

The parameters to be estimated are a, b, and expected laser intensities l_i. For analyzing such data, experimenters previously almost universally fitted a straight line to the logarithms of the mean values of all measurement clusters. We refer to this fitting scheme as log–log fitting and compare it with the other methods.

The model parameters chosen were $a_0 = 1.36 \times 10^{-3}$ and $b_0 = 2$. In Fig. 3 an example of such a synthetic measurement set is shown. The distributions of the estimated parameters that we obtained are shown in Figs. 4 and 5 as box-and-whiskers plots (for an explanation of this plotting technique see Appendix D). The estimated parameters are displayed as the relative deviation from preset values a_0 and b_0, respectively. Because ML estimators are asymptotically normally distributed when the number of replicates is increased, we choose, for the examples discussed in this paper, to estimate the expectation of the estimated parameters by taking the value of the median, which indeed is an appropriate procedure for symmetrically distributed variables. The vertical length of the box and the whiskers together is a measure for the standard deviation of the estimated parameter. Notice that smaller standard deviation implies higher precision.

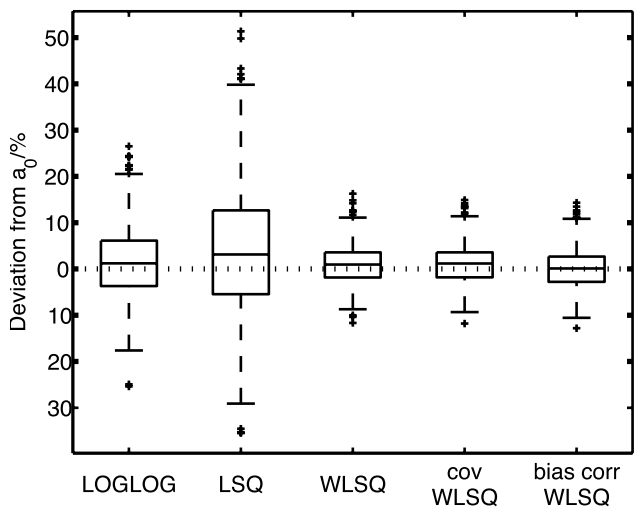

Fig. 4. Estimates of parameter a for a power function [Eq. (26)] for the various fitting schemes (see Appendix B). The results are shown as the relative deviation from initial model parameter a_0. One thousand estimates of the parameter obtained from Monte Carlo simulations were used for each fitting scheme for calculating the distributions. On the ordinate the fitting schemes, log–log, simple LSQ, WLSQ, covariant WLSQ without bias correction (cov WLSQ) and covariant WLSQ with bias correction (bias corr WLSQ) are marked. The interpretation of this plot and the meaning of the crosses (outliers) are explained in Appendix D. Note that the simple LSQ fitting scheme shows a more than three-times-worse precision than the WLSQ scheme.

An evident result for both model parameters a and b is that the precision increases when one is going from simple LSQ to WLSQ or one of the two covariant WLSQ estimators, WLSQ or covariant WLSQ. Both simple LSQ and these other estimators had been obtained with the method of ML but based on different statistical models. The underlying statistical model for simple LSQ does not properly describe the evaluated data, which resulted in poor precision, in this

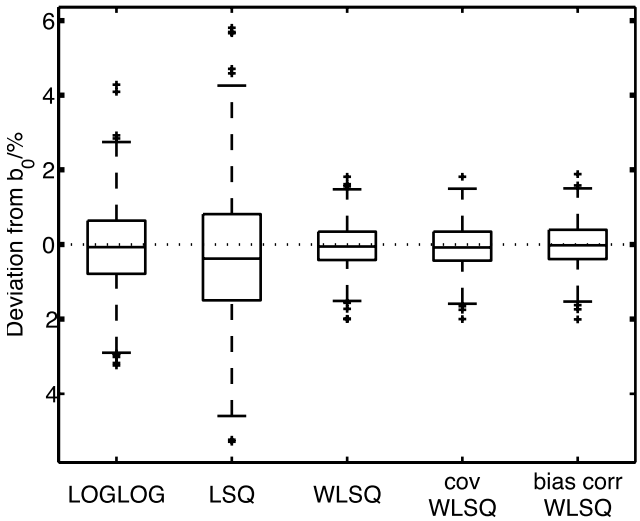

Fig. 5. Estimates of parameter b for a power function [Eq. (26)] for the various fitting schemes (see Appendix B). The results are shown as relative deviations from initial model parameter b_0.

case more than three times lower than that of the WLSQ fitting scheme. Simple LSQ differs from WLSQ fitting, among other things, in assuming that the standard deviation of Y_i is independent of subscript i (see Appendix B). As one can see from Fig. 3 this is a faulty assumption for the evaluated data sets.

The covariant WLSQ fitting scheme does not provide significantly improved precision for the estimated parameters compared with WLSQ fitting, because the correlation of signal and laser intensity is too small compared with their variances. We evaluated inequality (25) in Section 3 and obtained 3.9 for the left-hand side and 15 for the right-hand side, which supports our observation.

The widely used log–log fitting scheme yields estimates with higher precision than the simple LSQ scheme, because its statistical model assumes standard deviations of Y_i approximately proportional to their expectations $\mathcal{E}(Y_i)$. This is a somewhat better description of the data sets in this example (see Fig. 3). The log–log fitting scheme hence shows a higher precision than simple LSQ. Anyhow, its precision is worse than that for WLSQ estimation because the latter takes the adequate standard deviations for both Y and X into consideration.

For all fitting schemes the precision of estimator $\hat{B}$ is much larger than that of $\hat{A}$ because power function f is more sensitive to changes in b than to changes in a.

The bias of the estimated measurands is small and is readily observed only for simple LSQ. Anyhow, all fitting schemes except bias-corrected covariant WLSQ result in bias on both measurands. The discussion of the bias is taken up in more detail in Subsection 3.B.

B. Rational Function

Rational functions have been shown to describe the dependence on laser intensity of various nonlinear spectroscopic methods such as saturated laser-induced fluorescence, degenerate four-wave mixing, and polarization spectroscopy.[18,21] For our study we chose the following model function:

$$f(l, a, l_{\mathrm{sat}}) = \frac{a l^3}{(1 + l/l_{\mathrm{sat}})^2}. \tag{27}$$

This equation describes the laser intensity dependence of polarization spectroscopy signals.[18] Model parameter l_{sat} is the so-called polarization spectroscopy saturation intensity.

The model parameters chosen for the Monte Carlo simulations were $a_0 = 1.92 \times 10^{-4}$ and $l_{\mathrm{sat},0} = 31.8$, ensuring data sets with shapes similar to that of Fig. 3. The results are shown in Figs. 6 and 7. Here again we notice the increasing precision of the estimators when we go from simple LSQ to covariant WLSQ estimation. In contrast to that of the power function in Subsection 4.A, the fitting of Eq. (27) is obviously more sensitive to skew signal distributions because of the curvature of the function, which leads

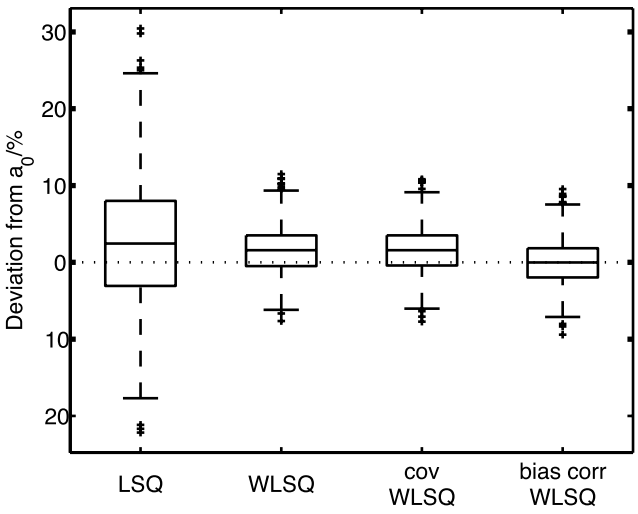

Fig. 6. Estimates of parameter a for a rational function [Eq. (27)] for the various fitting schemes (see Appendix B). The results are shown as relative deviations from initial model parameter a_0.

to a larger relative bias for the fitting schemes without bias correction. As discussed in Subsection 4.A, exponent b in power function equation (26) is only weakly dependent on the scatter of the measurement values, and the same is true for its bias. However, in Eq. (27) the model parameters enter the function not as exponents but as linear scaling parameters for the amplitude of the fraction and as a scaling parameter for the laser intensity (in the denominator). These parameters possess higher sensitivity to skew signal distributions. The simple LSQ, the WLSQ estimator, and the covariant WLSQ estimator without bias correction yield biased distributions of the estimated model parameters, whereas the bias disappears when the covariant WLSQ estimator with bias correction is used. This is a practical proof that the bias correction is appropriate.

Anyhow, the observed bias in Figs. 6 and 7 is still

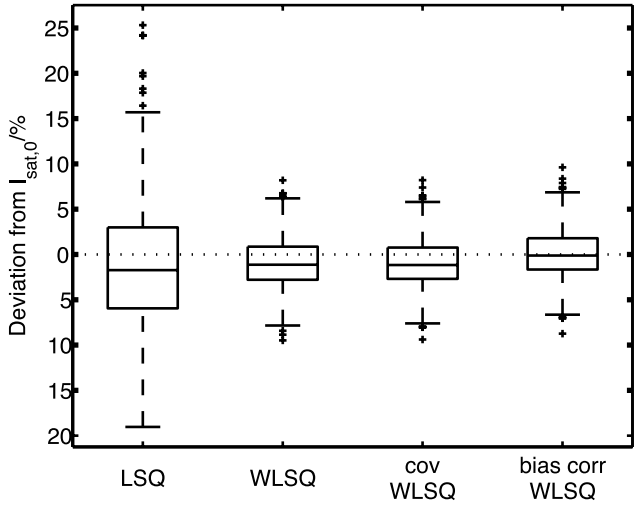

Fig. 7. Estimates of parameter l_{sat} for a rational function [Eq. (27)] for various fitting schemes (see Appendix B). The results are shown as relative deviations from initial model parameter $l_{\mathrm{sat},0}$.

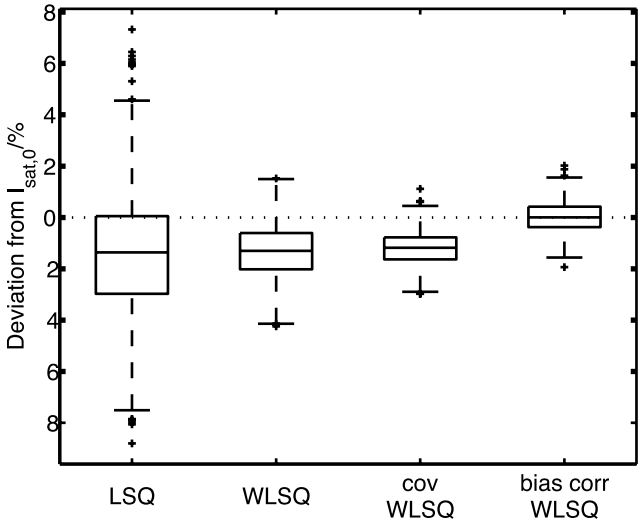

Fig. 8. Estimates of parameter l_{sat} for a rational function [Eq. (27)] for several fitting schemes (see Appendix B). The results are shown as relative deviations from initial model parameter $l_{sat,0}$. The σ_{2i} values for this Monte Carlo simulation are ten times smaller than for Fig. 7.

much smaller than the standard deviation of the estimators, which is due to the large detector noise on the signal, i.e., large quotients σ_{2i}/σ_{Li} ($\sim$4 in this case). To test the sensitivity of σ_{2i}/σ_{Li} we repeated the Monte Carlo simulations with ten-times smaller σ_{2i}, i.e., $\sigma_{2i}/\sigma_{Li} \sim 0.4$, which corresponds to measurements made with detectors of low noise levels. For a discussion of detectors and the noise that they introduce, see Dereniak and Growe.[22]

The results of our simulations are shown in Fig. 8 (l_{sat} only). The bias of the simple LSQ, the WLSQ, and the covariant WLSQ estimator without bias correction are approximately the same as in Fig. 7, but the precision is now so high that the bias becomes comparable to the standard deviation of the WLSQ and the covariant WLSQ estimator without bias correction. Note that still no bias is revealed for the covariant WLSQ fitting scheme with bias correction.

The reduction of σ_{2i} also entails demonstrably better precision when the covariant WLSQ instead of the WLSQ estimator, is used; i.e., the correlation of Y_i and X_i is now so pronounced that consideration of this correlation in the fitting scheme produces a distinct effect. For condition (25) (Section 3) we got 124 for the left-hand side and 136 for the right, i.e., almost similar values, a result that supports our observation.

Note that the bias of the uncorrected fitting schemes becomes even comparable to or larger than the standard deviation for the case of large detector noise if the number of replicates is increased. As the variance of the estimates converges toward zero for $m_i \to \infty$, whereas the bias does not depend on m_i [cf. expression (23)], there are always a certain number of replicates for which the bias becomes significant.

The covariant WLSQ fitting scheme with bias correction requires large numbers of replicates according to the central-limit theorem. Anyhow, this need not be a limiting requirement, as we investigated by rerunning the Monte Carlo simulations for this model function with $m_i = 3$ and $m_i = 10$ replicates, respectively. We observed the following: For three replicates the covariant WLSQ fitting scheme with bias correction did not succeed in suppressing the bias, but for ten replicates the bias of the estimates was already barely noticeable.

5. Summary and Conclusion

We have presented a statistical model for laser-spectroscopic experiments that possesses the following features: The dependence of the signal on the laser intensity is described by a well-known model function. The laser intensity is subject to stationary fluctuations, and both variables are measured with additive stationary noise. With aid of this model we derived maximum-likelihood estimators of the model parameters for the case when the signal depends linearly on the input. This method is referred to as covariant weighted least-squares fitting (*without* bias correction). We have also presented an approximate ML scheme for nonlinear model functions, and the method is referred to as covariant weighted least-squares fitting *with* bias correction.

We also described an alternative measurement model that treats the true laser intensities as unknowns and not as stochastic variables. However, the general, nonlinear LSQ scheme[6–10] was not applied to this model, as the variances of the detector noise are unknown. ML estimators based on this model were not examined either, because their calculation requires computational times more than 100 times longer than for the above schemes. Additionally, because the alternative model does not give a better description of the experiment, such a scheme is not expected to provide more-precise estimators.

We elucidated the precision and accuracy of our method by applying it to two classes of nonlinear model function. Hereby, the precision of the model parameters could be compared when common methods (simple LSQ, WLSQ, and log–log) and the method derived here were applied. The precision increased as much as three times for the examples described in this paper when we went from the simple LSQ by way of WLSQ fitting and our methods. The biases of all estimators except the covariant WLSQ fitting with bias correction were found to depend on the class of the model function. It was also found that the bias may approach the same order of magnitude as the uncertainty that is due to random variations if one measures both signal and laser intensity with low-noise detectors. Note that for any magnitude of the statistical variations the bias always becomes significant when the number of replicates is increased over a certain value. When the covariant WLSQ fitting scheme with bias correction was applied to these measurements, the bias was virtually reduced to zero. The applied method of bias correction was shown to be successful for as few replicates as ten for every cluster.

The covariant WLSQ fitting scheme presented here is highly versatile because one can use standard LSQ

fitting routines for maximization of the likelihood function when an appropriate coordinate transformation is applied. Another advantageous feature is that the minimized quantity χ^2 is approximately χ^2-distributed and the goodness of an obtained fit can hence be judged.

In conclusion, we showed that the method of ML is a versatile and efficient method for measurement evaluation if it is applied to a statistical model that is adequate for the experiment. The method is simple and offers the opportunity of treating a large variety of similar experiments in the fields of physics, chemistry, and biology and other disciplines.

For the calculations in this research, a program interface for MATLAB was developed. It allows one to carry out covariant WLSQ estimations (with bias correction) by using the LSQ routines of the MATLAB Optimization Toolbox. In addition, it performs Monte Carlo simulations for the estimation of confidence intervals of the model parameters. This program can be freely obtained from the World-Wide Web (http://www.forbrf.lth.se/WFIT/) or by an inquiry to T. Metz.

This research was financed by the Swedish Research Council, the Swedish National Energy Administration, and Sydkraft AB. Thanks are due to Jan and Ulla Holst and especially to Axel Franke for enlightening comments on the manuscript.

Appendix A: Symbols Used in This Paper

1. General Designations

- A Stochastic variable
- A_{ij} Stochastic variable for the ith cluster (setting); jth value (replicate)
- a The name of a physical quantity or
  A concrete number such as
  an observation of a stochastic variable, e.g., x_{ij};
  a parameter of the distribution of a stochastic variable, e.g., l_i, σ_{Li}, σ_{1i}, σ_{2i};
  a parameter in function f, *e.g.*, a, b;
  a variable of a PDF, e.g., x_{ij}, $\bar{y}_i$
- a_0 Predefined value of a in the Monte Carlo simulations
- $\mathscr{E}(A)$ Expectation of A
- $\mathscr{V}(A)$ Variance of A
- $\mathscr{C}(A, B)$ Covariance of A and B
- $\hat{A}$ Estimator of a
- $\mathbf{a}^{\mathrm{T}}$ Transpose of vector or matrix $\mathbf{a}$

2. Specific Symbols

- $E^{(k)}$ Measurement noise from detector k
- f_i Generic function describing the dependence of the generated signal on the laser intensity for setting number i
- L Laser intensity
- $\mathscr{L}$ Likelihood function
- m_i Number of points in cluster i
- n Number of clusters
- $\mathbb{N}(\mu, \sigma^2)$ Normal distribution with expectation μ and variance σ^2
- p PDF
- $\boldsymbol{\sigma}_i$ Covariance matrix of X_{ij} and Y_{ij}; cf. Eq. (10). The distributions of X_{ij} and Y_{ij} do not depend on j
- $\hat{\mathbf{U}}_i$ Estimator for $\boldsymbol{\sigma}_i$ (not used in the text)
- $\hat{\mathbf{u}}_i$ Sample covariance matrix, estimate for $\boldsymbol{\sigma}_i$ (i.e., observation of $\hat{\mathbf{U}}_i$)
- X Observed laser intensity
- Y Observed signal intensity

Appendix B: Fitting Schemes

In curve fitting there are several well-established schemes, which provide ML estimators only if certain conditions are fulfilled.

In WLSQ fitting, statistical independence is assumed for different data points but also for the measured quantities X_{ij} and Y_{ij} (for denotations see Section 2). Furthermore, the statistical variations of X_{ij} and Y_{ij} are assumed to be normally distributed, with variances σ_{xi}^2 and σ_{yi}^2, respectively. For these conditions the method of ML leads to the minimization of

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{(x_{ij} - l_i)^2}{\sigma_{xi}^2} + \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{[y_{ij} - f(l_i)]^2}{\sigma_{yi}^2}. \quad \text{(B1)}$$

Here x_{ij} and y_{ij} are experimental data, whereas l_i and the unknown parameters of f are to be optimized. The σ_{xi} and σ_{yi} are assumed to be known. WLSQ fitting is a special case of covariant WLSQ fitting [compare Eq. (B1) and Eq. (15) for $\mathscr{C}(X_i, Y_i) = 0$]. This variant of WLSQ fitting is used in Section 4.

The unweighted LSQ fitting scheme is obtained when all σ_{xi} and σ_{yi} are equal; say, σ_x and σ_y, respectively. Then the factor $1/\sigma_x^2$ can be removed outside the sums and we obtain ML estimators from minimizing the following expression:

$$\chi^2 = \frac{1}{\sigma_y^2} \left\{ \lambda^2 \sum_{i=1}^{n} \sum_{j=1}^{m_i} (x_{ij} - l_i)^2 + \sum_{i=1}^{n} \sum_{j=1}^{m_i} [y_{ij} - f(l_i)]^2 \right\}, \quad \text{(B2)}$$

$$\lambda := \frac{\sigma_y}{\sigma_x}. \quad \text{(B3)}$$

Note that only the ratio λ has to be known, because one only has to minimize the expression in braces.

If the uncertainties of the x values are negligible, the first sums in Eqs. (B1) and (B2) disappear, as does the necessity to optimize l_i, and χ^2 becomes simply

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{[y_{ij} - f(x_{ij})]^2}{\sigma_{yi}^2}, \quad \text{(B4)}$$

or

$$\chi^2 = \frac{1}{\sigma_y^2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} [y_{ij} - f(x_{ij})]^2 \quad \text{(B5)}$$

Table 1. Fitting Schemes[a]

Fitting Scheme Condition	LSQ		WLSQ		Covariant WLSQ	
					No BC	BC
Equation	(B2)	(B5)	(B1)	(B4)	(15)	(20)
All standard deviations equal	x	x				
Variations/uncertainties of x and y independent	x	x	x	x		
Normally distributed errors[b]	(x)	(x)	(x)	(x)	(x)	(x)
Variation in x negligible		x		x		
f linear in l					x	

[a]Crosses mark the necessary conditions for which a fitting scheme provides ML estimators for the unknown parameters. BC, bias correction.

[b]Parentheses mean that the normality assumption is not mandatory if the number of data points per cluster is large. See also Subsection 3.C.

for WLSQ or unweighted LSQ fitting, respectively. The latter is the variant of LSQ to which we refer in this paper as simple LSQ.

When the relationship between x and y follows a power function such as $y = ax^b$, one frequently fits—in the LSQ sense—a straight line to the logarithms of the data, as $\ln y = \ln a + b \ln x$. In this paper we call this method the log–log fitting scheme. This way of fitting is much easier to perform, because the optimization problem is linear in $\ln a$ and b. For statistical reasons the log–log method can be doubtful, because the logarithms of the data have to fulfill the conditions necessary for LSQ to provide ML estimators, in particular, normally distributed errors and constant standard deviations. These deviations of the logarithms correspond to constant relative errors [e.g., $\sigma_{yi}/\mathcal{E}(Y_i) \approx$ constant]. The validity of these assumptions has, of course, to be confirmed before one uses the method. Moreover, the logarithms of the observed intensities are biased estimators for the logarithms of the true intensities, which yield bias on the fitting parameters.[23] In general, one should be careful when fitting treated data (here, the logarithms of x and y), and use of the original data should be preferred.

In Table 1 several fitting schemes and the conditions for which they provide ML estimators are summarized.

Appendix C: Method of Monte Carlo Simulation

The method of Monte Carlo simulation can be summarized as follows: First, one defines a model function f, the parameters of which one derives to be estimated. One proceeds by choosing typical experimental values for the input variables of the statistical model, in our case laser intensities l_i as well as the model parameters in f and the standard deviations (σ_{Li}, σ_{1i}, and σ_{2i}) in the statistical model. All these values can be estimated by analysis of a measurement sample or by use of typical values from the literature.

With the aid of a random-number generator and formulas (1)–(5), one produces data sets (x_{ij}, y_{ij}). (For the simulations in this paper a value of $m_i = 100$ laser shots was chosen for all i.) Then one uses the fitting schemes listed in Appendix B to calculate estimated values for the parameters in f from each simulated data set.

The generation and fitting are repeated several times, and the estimated parameters are stored. For the results presented in this paper a number of thousand runs was chosen, which ensures sufficient precision for median and variance of the estimated parameters. The distribution of the estimated parameters provides a measure of their precision and accuracy.

Appendix D: Box-and-Whiskers Plots

The box-and-whiskers plot was invented by Tukey[24] to illustrate the statistical distribution of the values in a sample. Compared to plotting all the single sample values directly, the box-and-whiskers plot shows more clearly the important features of the distribution.

In a box-and-whiskers plot the data are represented by a geometrical figure, as shown, e.g., in Fig. 8. This figure indicates the following properties of the distribution: The lowest 25% of the sample values lie under the lower edge of the box, and the highest 25% lie above the upper edge. Thus 50% of all values lie within the box. The values that correspond to the edges of the box are called the quartiles; the height of the box, the interquartile range.

The horizontal line that divides the box into two parts indicates the position of the median. Above this line lies 50% of the sample; below it, the other 50%. Thus the upper and the lower parts of the box contain 25% of the sample values each.

The vertical lines above and below the box are the so-called whiskers. Each of them extends to the greatest and least data value whose distance from the box is at the most 1.5 times the height of the box. For a normal distribution the maximum range that can be spanned by the box and the whiskers together is expected to contain 99.3% of the sample values. Data points lying beyond the edges of the whiskers are considered outliers and are plotted directly.

In the experiments reported here the box-and-whiskers plot is used mainly to compare the scattering and the median of estimated parameters from Monte Carlo simulations. In this way the precision

and accuracy of the corresponding estimators can be compared.

References

1. P. H. Garthwaite, I. T. Jolliffe, and B. Jones, *Statistical Inference* (Prentice-Hall, London, 1995).
2. M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, 3rd ed. (Griffin, London, 1973), Vol. 2.
3. J. F. Kenney and E. S. Keeping, *Mathematics of Statistics*, 2nd ed. (Van Nostrand, Princeton, N.J., 1951), Part II.
4. R. A. Fisher, "On an absolute criterion for fitting frequency curves," Mess. Math. **41,** 155–160 (1912).
5. P. J. Mohr and B. N. Taylor, "CODATA recommended values of the fundamental physical constants: 1998," Rev. Mod. Phys. **72,** 351–495 (2000).
6. S. Brandt, *Statistical and Computational Methods in Data Analysis* (North-Holland, Amsterdam, 1970).
7. H. I. Britt and R. H. Luecke, "The estimation of parameters in nonlinear implicit models," Technometrics **15,** 233–247 (1973).
8. A. Celmiņš, "Least squares adjustment with finite residuals for non-linear constraints and partially correlated data," in *Nineteenth Conference of Army Mathematicians* (U.S. Army Research Office, Washington, D.C., 1973), AROD-73-3-PT-2, pp. 809–858.
9. W. H. Jefferys, "On the method of least squares," Astron. J. **85,** 177–181 (1980).
10. W. H. Jefferys, "On the method of least squares. II," Astron. J. **86,** 149–155 (1981).
11. MATLAB Optimization Toolbox, V. 2.2. Mathworks: http://www.mathworks.com/access/helpdesk/help/toolbox/optim/optim.shtml.
12. D. Taupin, *Probabilities, Data Reduction and Error Analysis in the Physical Sciences* (Éditions de Physique, Les Ulis, France, 1988).
13. W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1971), Vol. II.
14. P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 2nd ed. (McGraw-Hill, Boston, Mass., 1992).
15. International Organization for Standardization, *Guide to Expression of Uncertainty in Measurement*, 2nd ed. (ISO, Geneva, 1995).
16. W. A. Fuller, *Measurement Error Models* (Wiley, New York, 1987).
17. R. J. Caroll, D. Ruppert, and L. A. Stefanski, *Measurement Error in Nonlinear Models*, Vol. 63 of Monographs on Statistics and Applied Probability (Chapman & Hall, London, 1995).
18. J. Walewski, C. F. Kaminski, S. F. Hanna, and R. P. Lucht, "Dependence of partially saturated polarization spectroscopy signals on pump intensity and collision rate," Phys. Rev. A **64,** 063816 (2001).
19. B. N. Taylor and C. E. Kuyatt, "Guidelines for evaluating and expressing the uncertainty of NIST measurement results," Tech. Rep. (National Institute of Standards and Technology, Gaithersburg, Md., 1994); http://physics.nist.gov/pubs/guidelines/contents.html.
20. W. Demtröder, *Laser Spectroscopy* (Springer-Verlag, Berlin, 1996).
21. A. C. Eckbreth, *Laser Diagnostics for Combustion Temperature and Species* (Overseas Publishers, Amsterdam, 1996).
22. E. L. Dereniak and D. G. Growe, *Optical Radiation Detectors* (Wiley, New York, 1984).
23. W. J. Thompson and J. R. Macdonald, "Correcting parameter bias caused by taking logs of exponential data," Am. J. Phys. **59,** 854–856 (1991).
24. J. W. Tukey, *Exploratory Data Analysis* (Addison-Wesley, Reading, Mass., 1977), pp. 39–41, 44.