Correcting Systematic Errors by Hybrid 2D Correlation Loss Functions in nonlinear inverse modelling

Thomas Mayerhoefer (Thomas.Mayerhoefer@leibniz-ipht.de)
Leibniz Institute of Photonic Technology https://orcid.org/0000-0001-9396-7365
Isao Noda
University of Delaware
Susanne Pahlow
Leibniz Institute of Photonic Technology
Rainer Heintzmann
Leibniz Institute of Photonic Technology
Jürgen Popp
Leibniz Institute of Photonic Technology

Article

Keywords:

Posted Date: August 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1929442/v1

Correcting Systematic Errors by Hybrid 2D Correlation Loss Functions in nonlinear inverse modelling

Thomas G. Mayerhöfer[a,b,*], Isao Noda[c], Susanne Pahlow[a,b], Rainer Heintzmann[a,b], Jürgen Popp[a,b]
[a] *Leibniz Institute of Photonic Technology (IPHT), Jena, 07745, Albert-Einstein-Str. 9, Germany*
[b] *Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich Schiller University, Jena, 07743, Helmholtzweg 4, Germany*
[c] *University of Delaware, Newark, DE, 19716, USA*

ABSTRACT

Recently a new family of loss functions called *smart error sums* has been suggested. These loss functions account for correlations within experimental data and force modeled data to obey these correlations. As a result, multiplicative systematic errors of experimental data can be revealed and corrected. The *smart error sums* are based on 2D correlation analysis which is a comparably recent methodology for analyzing spectroscopic data that has found broad application. In this contribution we mathematically generalize and break down this methodology and the *smart error sums* to uncover the mathematic roots and simplify it to craft a general tool beyond spectroscopic modelling. This reduction also allows a simplified discussion about limits and prospects of this new method including one of its potential future uses as a sophisticated loss function in deep learning. To support its deployment, the work includes computer code to allow reproduction of the basic results.

Main

Systematic errors lead to non-accurate results with biases even if an experiment is repeated multiple times and the results are averaged to reduce the statistical random error. Systematic errors are often hard if not sometimes impossible to detect.[1,2] One reason is that the down- or upshift of the mean compared to the true value does not influence the distribution of results caused by random errors. Even though it is possible to remove or reduce random errors and obtain a seemingly consistent ("precise") result, these may nevertheless still be far from the underlying ground-truth.

In particular, in curve fitting, where experimental data points are fitted assuming that the points follow a mathematical or physical model, a good agreement between measured points and fitted curve may delude the operator into thinking that systematic errors are absent.[2,3] In general, it must be kept in mind that a good agreement of a fit is no indication of accuracy, neither in terms of the experimental data nor of the underlying model. That is data analysis of quantitative experiments is based upon the assumption that the measured or calculated independent and dependent variables are not subject to systematic errors.[4]

Curve fitting is often applied when nonlinear mathematical or physical problems are to be solved that cannot be linearized. When linear problems and those that can be linearized are to be solved, analytical solutions can be derived. Such linear models form the basis for many chemometric methods.[5] In contrast to linear problems, nonlinear problems require to iteratively improve the fitted curve by minimizing a measure of disagreement, to approach the curves of the experimental data. Potential applications of curve fitting are countless and encompass virtually all scientific disciplines. Examples include biosynthesis,[6] thermoluminescence,[7] solar energy,[8] materials science and technology,[9] agriculture,[10] cancer research,[11] kinetics,[12] thermal engineering,[13] transportation,[14] soil science,[15] remote sensing of ecosystems,[16] epidemiology,[17] power and energy engineering,[18] population growth[19] and spectroscopy,[20] to name just a few. The disagreement metrics to minimize during the fit depends on the properties of the noise and possibly on prior information on the parameters to fit. However, in most cases it is sensible to assume Gaussian statistics, which requires the minimization of the residual sum of squares (RSS) also called the sum of squared errors or its weighted version in case of non-uniform but known variances.

Recently, we derived alternative loss functions, which are based on 2D correlation analysis or spectroscopy (2D COS).[3,21] 2D correlation analysis has been introduced in the 1980s by one of us as a tool for infrared spectroscopy which found widespread use also in other spectroscopic methods, like Raman spectroscopy and mass spectrometry.[22] In principle, 2D correlation spectroscopy is based on acquiring a series of spectra under a systematic change of one parameter of the sample (the so-called perturbation), e.g., the stretch of a polymer, the temperature or the concentration of one compound in a mixture. The perturbation can certainly also be a parameter of an established physical model to describe how its alteration induces changes in the spectra, like the thickness of films. A variant of 2D correlation spectroscopy is hybrid 2D correlation spectroscopy, which allows the comparison of two different spectral series, e.g., the same compound under two different perturbations. In the sense of curve fitting, a variant would be to let one series consist of experimental data, whereas the second comprises modelled spectra.

In the original 2D correlation maps, half of the data points are redundant, due to symmetry relations between points separated by the diagonal from low to high values of the independent variable. For hybrid 2D correlation maps these relations do no longer hold, but the more the two series resemble each other, the smaller the deviations from this symmetry relations become. This property of hybrid 2D correlation maps can be exploited by formulating an alternative criterion for the resemblance between experimental and modelled data which includes the correlations in between the series, which we call *smart error sum* (SES). In contrast to the fits using the conventional RSS, the 2D correlation-based *smart error sum* does not force the modelled curve to agree point by point with experimental curves, but accounts for correlations between the latter and between the individual points of a curve. As a consequence, even when the experimental data are reduced by a (frequency dependent) factor, the data can still be analyzed in a meaningful way. In spectroscopy, such a situation is often encountered, for example, due to diffuse reflection caused by surface roughness, or measured data becomes larger than predictable by models which do not account for such experimental problems. This approach can not only detect, but also remove multiplicative systematic errors as has been demonstrated for infrared spectra of films on substrates with different thicknesses (additive systematic errors can also be treated after applying an exponential transformation).

Quite often, though, series of spectra are not available and it is only one data curve that is to be fitted. In this case inter spectral correlations cannot be exploited. However, it is still possible to use correlations between the individual data points based on a recent subtype of 2D correlation analysis. The so-called two-trace 2D correlation analysis or spectroscopy (2T2D COS)[23,24] requires only two sets of spectral data of which only one needs to be experimental. In this case the symmetry relations cannot be used as a criterion, but if experiment and model agreed perfectly, one of the maps would amount to become everywhere zero. The value of this idea has been proven employing the same physical system as the original *smart error sum*. As we will show, the 2T2D SES approach is similar to utilizing normalized cross-correlation (NCC) and zero mean normalized cross-correlation (ZNCC) as SES. NCC and ZNCC are related to 2D correlation analysis[22] and often used for signal analysis. Examples entail comparing image quality in competition with the conventional residual sum of squares,[25-28] tracking wavelength-shifts in Fiber-Bragg gratings.[29,30] and, recently, also least squares optimizations of images.[31]

While the application to real systems and problems helped to establish the validity of the approach, it also partially obscured the mathematical basis and the principal properties of the method. To enable broader application, we therefore here reduce it to its essential properties and demonstrate it based on a simple example in the following. In addition, we provide the 2T2D-based smart error sum in a form that scales, like the conventional sum of squared residuals, linearly with the number of points. Accordingly, the former can replace the latter in nonlinear curve-fitting applications that are prone to unknown systematic experimental errors. The code of the program which we used to obtain the results shown in the following is made available together with this work so that they can be easily reproduced. In addition, the code can effortlessly be modified to be used for other non-linear models.

Methods

The following is based on the matrix algebra employed for 2D-COS as used by Noda and Osaki.[22] Since the formalism was originally developed for spectroscopy, we have to slightly reformulate it. However, to allow the reader to connect the following to the original literature, we will try to adhere to the original terminology as closely as possible. We assume that we have a function $\tilde{y}_k = \tilde{y}_k(x,t)$ of two variables x and t of which we call the former the shaping variable and the latter the perturbation. A number of different data points located on m different curves which differ with regard to t, shall be represented by employing discrete values x_i and t_j according to $\tilde{y}_k = \tilde{y}_k(x_i,t_j)$. These curves will be called a set of dynamic spectra $\tilde{y}_k = \tilde{y}_k(x_i,t_j)$

The dynamic spectra are arranged in a matrix $\mathbf{Y}_k$ in the following way:

$$\mathbf{Y}_k = \begin{bmatrix} \tilde{y}_k(x_1,t_1) & \tilde{y}_k(x_2,t_1) & \dots & \tilde{y}_k(x_n,t_1) \\ \tilde{y}_k(x_1,t_2) & \tilde{y}_k(x_2,t_2) & \dots & \tilde{y}_k(x_n,t_2) \\ \dots & \dots & \dots & \dots \\ \tilde{y}_k(x_1,t_m) & \tilde{y}_k(x_2,t_m) & \dots & \tilde{y}_k(x_n,t_m) \end{bmatrix}. \quad (1)$$

The index $k \in \{1,2\}$ and indicates if the set of dynamic spectra consists of either the set of "measured" ($k=1$) or the set of simulated spectra ($k=2$). One may think that it is advantageous to mean-center the dynamic spectra, i.e., subtracting the mean spectrum of the series from each individual measured spectrum. However, such mean-centering or, more general, referencing is often not only unnecessary,[32] but sometimes even detrimental. Yet, if the array of curves or dynamic spectra share a common offset, this offset needs to be removed prior to application, otherwise not only the 2D correlation maps,[33] but also the *smart error sums* are ill-defined.

From the matrices $\mathbf{Y}_k$ the variance-covariance matrices $\mathbf{\Phi}_{xx}$ can be generated by:

$$\mathbf{\Phi}_{xx} = \frac{1}{m-1}\mathbf{Y}_1^T\mathbf{Y}_2 . \quad (2)$$

If $\mathbf{Y}_1 = \mathbf{Y}_2$, then we speak of conventional 2D-COS, whereas the case $\mathbf{Y}_1 \neq \mathbf{Y}_2$ leads to a so-called hybrid correlation analysis. In case of the conventional *smart error sum* $\mathbf{Y}_1$ is formed from the "measured" and $\mathbf{Y}_2$ from the corresponding simulated curves. As pointed out in ref. [22], each element of the variance-covariance

matrix expresses the similarity between a specific pair of intensity variations at different x_j. If $\mathbf{Y}_1 = \mathbf{Y}_2$, the diagonal elements are the autocorrelation functions of the intensity variations along t at a given x_j.

The variance-covariance matrix is identical to the synchronous 2D correlation map/spectrum. In order to compute the asynchronous 2D correlation map/spectrum, the Hilbert–Noda transformation matrix **N** must be calculated first. The elements of **N** are given by:

$$N_{ij} = \begin{cases} 0 & \text{if } i = j \\ \dfrac{1}{\pi(j-i)} & \text{otherwise} \end{cases} . \tag{3}$$

The elements of the asynchronous 2D correlation map/spectrum can then be calculated according to,

$$\mathbf{\Psi}_{xx} = \frac{1}{m-1} \mathbf{Y}_1^T \mathbf{N} \mathbf{Y}_2 , \tag{4}$$

where we again distinguish between the conventional case (I) and hybrid 2D-COS (II). As already mentioned, in the introduction for the conventional ($\mathbf{Y}_1 = \mathbf{Y}_2$) 2D-correlation analysis for synchronous and asynchronous spectra/maps certain symmetry relationships hold. Accordingly, the synchronous spectra are always symmetric relative to the diagonal elements $\Phi(x_i, x_i)$. This condition can be expressed as,

$$\Phi(x_j, x_k) = \Phi(x_k, x_j) . \tag{5}$$

Accordingly, the diagonal from small to large x values is a mirror plane that relates each point above the diagonal to its mirror image below it. From eqn. (5) it follows that the sum of differences of all variances and covariances above the diagonal and their counterparts below the diagonal are zero:

$$\sum_{k=1}^{l} \sum_{j=k}^{l} \left[\Phi(x_j, x_k) - \Phi(x_k, x_j) \right] = 0 . \tag{6}$$

For hybrid 2D-COS, the synchronous maps are not necessarily obeying the above condition. The residuals of the differences of the elements $\Phi(x_j, x_k)$ and $\Phi(x_k, x_j)$ is a measure of dissimilarity, which can be generally written as,

$$D_S^p = \sum_{k=1}^{l} \sum_{j=k}^{l} \left[\Phi(x_j, x_k) - \Phi(x_k, x_j) \right]^p , \tag{7}$$

with D_S, the so-called Minkowski distance, which is called the Euclidian distance for $p = 2$. Therefore, hybrid 2D correlation maps allow a derivation of these quantities simply from their symmetry relations.[3] For asynchronous 2D correlation maps, a similar relationship can be derived. Accordingly, similarly to eqn. (5), we find from the condition that the conventional asynchronous 2D correlation maps are always antisymmetric with respect to the diagonal the following relation: [3]

$$\begin{aligned} &\Psi\left(x_j,x_k\right)=-\Psi\left(x_k,x_j\right) \\ &\rightarrow \sum_{k=1}^{l}\sum_{j=k}^{l}\left[\Psi\left(x_j,x_k\right)+\Psi\left(x_k,x_j\right)\right]=0 \end{aligned}. \tag{8}$$

This relation leads for the hybrid 2D-correlation asynchronous map to:

$$D_A^p=\sum_{k=1}^{l}\sum_{j=k}^{l}\left[\Psi\left(x_j,x_k\right)+\Psi\left(x_k,x_j\right)\right]^p . \tag{9}$$

Note that for both, eqs. (7) and (9), we can include the diagonal since the terms with $j = k$ are zero.

For $p = 2$, D_S^2 and D_A^2 are special residual sums of squares, which we call the synchronous and the asynchronous residual sum of squares, SRSS and ARSS. SRSS and ARSS can be combined ad hoc to the *smart error sum* (*SES*) according to:

$$\ln\left(SRSS\right)+\ln(ARSS)=SES . \tag{10}$$

Obviously, in cases where only a single measured spectrum is available for curve fitting, the *smart error sum* cannot be used, simply because it is not possible to generate a 2D correlation map from a single spectrum. For this case, we have introduced an alternative *smart error sum* based on hybrid 2T2D-COS, with one measured curve, while the other is the simulated one. Synchronous and asynchronous 2D correlation spectrum/map are then calculated by,[23,24]

$$\begin{aligned} \Phi\left(x_j,x_k\right)&=\frac{1}{2}\left[s\left(x_j\right)\cdot s\left(x_k\right)+m\left(x_j\right)\cdot m\left(x_k\right)\right] \\ \Psi\left(x_j,x_k\right)&=\frac{1}{2}\left[s\left(x_j\right)\cdot m\left(x_k\right)-s\left(x_j\right)\cdot m\left(x_k\right)\right] \end{aligned}, \tag{11}$$

wherein $m\left(x_j\right)=\mathbf{Y}_1\left(x_j,t_l\right)$ is the measured and $m\left(s_j\right)=\mathbf{Y}_2\left(x_j,t_l\right)$ the simulated curve with an arbitrary t_l . Based on eqn. (11), the hybrid synchronous spectrum is always symmetric and the asynchronous spectrum

is always antisymmetric relative to the diagonal. Therefore, the underlying principle of the *smart error sum*, introduced for series of curves, namely to increase the symmetry of the hybrid synchronous 2D correlation map and the antisymmetry of the hybrid asynchronous 2D correlation map by varying the fit parameters, cannot be employed. As an alternative we can use that the asynchronous 2T2D-Correlation map Ψ_{xx} vanishes if both the given and the modelled curve are linearly dependent. Put in concrete terms, in this case the given and the modelled curve can also, as in case of the conventional *smart error sum*, differ by a simple scalar multiplication factor. Accordingly, the 2T2D *smart error sum* is given by,[21]

$$D_{A2T}^{p} = \sum_{k=1}^{l}\sum_{j=k}^{l}\left[\Psi\left(x_j, x_k\right)\right]^{p}, \tag{12}$$

where we set $p = 2$. Therefore, a corresponding algorithm would minimize D_{A2T}^{2} by finding optimized values for the fit parameters. In eqn. (12), all points below the diagonal need not to be considered, which follows from the asynchronous map being perfectly antisymmetric:

$$\left[\Psi\left(\tilde{\nu}_j, \tilde{\nu}_k\right)\right]^{p} = (-1)^{p}\left[\Psi\left(\tilde{\nu}_k, \tilde{\nu}_j\right)\right]^{p}. \tag{13}$$

On the other hand, if set $p = 2$, then there is a possibility to significantly simplify eqn. (12), if we let both sums run from 1 to l. In this case,

$$\begin{aligned}
D_{A2T}^{2} &= \sum_{k=1}^{l}\sum_{j=1}^{l}\left[\Psi\left(\tilde{\nu}_j, \tilde{\nu}_k\right)\right]^{2} \\
&= \frac{1}{4}\sum_{k=1}^{l}\sum_{j=1}^{l}\left[s\left(\tilde{\nu}_j\right)s\left(\tilde{\nu}_j\right)\cdot m\left(\tilde{\nu}_k\right)m\left(\tilde{\nu}_k\right) + s\left(\tilde{\nu}_k\right)s\left(\tilde{\nu}_k\right)\cdot m\left(\tilde{\nu}_j\right)m\left(\tilde{\nu}_j\right) - 2\cdot s\left(\tilde{\nu}_j\right)\cdot m\left(\tilde{\nu}_j\right)s\left(\tilde{\nu}_k\right)\cdot m\left(\tilde{\nu}_k\right)\right] \\
&= \frac{1}{4}\left(\sum_{j=1}^{l}s\left(\tilde{\nu}_j\right)s\left(\tilde{\nu}_j\right)\sum_{k=1}^{l}m\left(\tilde{\nu}_k\right)m\left(\tilde{\nu}_k\right) + \sum_{k=1}^{l}s\left(\tilde{\nu}_k\right)s\left(\tilde{\nu}_k\right)\sum_{j=1}^{l}m\left(\tilde{\nu}_j\right)m\left(\tilde{\nu}_j\right) - 2\sum_{j=1}^{l}m\left(\tilde{\nu}_j\right)s\left(\tilde{\nu}_j\right)\sum_{k=1}^{l}m\left(\tilde{\nu}_k\right)s\left(\tilde{\nu}_k\right)\right) \\
&= \frac{1}{2}\left(\sum_{j=1}^{l}s\left(\tilde{\nu}_j\right)^{2}\sum_{k=1}^{l}m\left(\tilde{\nu}_k\right)^{2} - 2\left(\sum_{j=1}^{l}m\left(\tilde{\nu}_j\right)s\left(\tilde{\nu}_j\right)\right)^{2}\right)
\end{aligned} \tag{14}$$

which scales with $\mathcal{O}(N)$ like the conventional residual sum of squares instead of $\mathcal{O}(N^2)$ like the other *smart error sums* based on 2D correlation analysis. The advantage of the *smart error sums* in comparison

with the conventional residual sum of squares as minimalization criterion is that for the former experimental and simulated curve are not forced to agree by all means but can be different by an individual factor, the optimum of which is determined by maximum correlation. In other words, not only the best agreement between original and simulated values determines the fit parameters, but also the correlation of the curves in a series or within a curve. From a mathematical point of view, the additional degree of freedom can be understood in terms of the phase angle:

$$\Theta(x_1, x_2) = \arctan\left\{\frac{\Psi(x_1, x_2)}{\Phi(x_1, x_2)}\right\}. \tag{15}$$

The term phase angle is used, because $\mathbf{\Phi}_{xx}$ and $\mathbf{\Psi}_{xx}$ are linearly independent and can be described formally as a complex function:

$$\mathbf{X}_{xx} = \mathbf{\Phi}_{xx} + i\mathbf{\Psi}_{xx}. \tag{16}$$

$\Theta(x_1, x_2)$ is then derived from the polar form. For hybrid 2T2D correlation analysis, $\Psi(x_1, x_2)$ becomes zero if the original curve and its best fit agree within a multiplication factor, which means that $\Theta(x_1, x_2) = 0$. This situation means that the two curves are linearly dependent or even identical if systematic errors are absent. Accordingly, an alternative form for the 2T2D-based *smart error sum* is given by:[21]

$$D_{A2T}^{p} = \sum_{k=1}^{l}\sum_{j=k}^{l}\left[\Theta(x_k, x_j)\right]^{p}. \tag{17}$$

For series-based hybrid 2D correlation analysis, the ratios $\Psi_1(x_1, x_2)/\Phi_1(x_1, x_2)$ for the set of the given curves and $\Psi_2(x_1, x_2)/\Phi_2(x_1, x_2)$ for the fitted curves are equal if the correlations within both sets of curves agree. An alternative form of the original smart error sum is therefore,[21]

$$D_{SES}^{p} = \sum_{k=1}^{l}\sum_{j=k}^{l}\left[\Theta_{ex}(x_k, x_j) - \Theta_{sim}(x_k, x_j)\right]^{p}. \tag{18}$$

where $\Theta_{ex}(x_k, x_j)$ are the phase angles of the original data and $\Theta_{sim}(x_k, x_j)$ are those of the simulated curves. This form, for $p = 2$ and without consideration of its symmetry properties, has originally been introduced by Shinzawa et al.[34,35] and used exclusively for the method of alternating least squares (ALS). In

this form, a theoretical problem of eqn. (10) is avoided, which occurs if either SRSS or ARSS becomes zero, which in practice unlikely happens due to numerical errors related to the conversion of numbers to the binary system. To be on the safe side, a dummy regularizing positive constant ε can be incorporated into eqs. (6) and (7), e.g. $\varepsilon = 10^{-10}$ (this value is small enough to have no effect on the actual computation and may be viewed as a predictable substitute for random bit noise). Similar consideration may apply to the calculation of the phase angle defined in eqn. (15) and it may be advantageous to regularize the denominator, even though the chance for the intensity of synchronous spectrum becomes exactly zero might be slim (there are chances that this can happen near the zero-crossing area. The sign of the regularization constant has to be the same as the sign of the synchronous spectrum intensity. The primary reason for regularizing the ratio between asynchronous and synchronous intensities is to avoid the ambiguity of the zero-divided-by-zero situation where the dynamic spectrum remains zero). Note that for a typical arctangent function routine, the direction of a vector in the phase plain is confined to the first and fourth quadrants. In other words, the phase angle calculated by a computer is automatically assumed to take the value between –π/2 and +π /2. In a practical physically expected situation, a phase vector can point to the direction outside of this artificial confinement. Therefore, it is generally advantageous to assume that the phase angle should be confined between -π/4 and +3π/4, and to add π whenever the calculated value lies between - π/2 and - π/4.

According to eqn. (18), individual 2D maps can differ by multiplication factors, even though their phase angles are equal. In the absence of systematic errors, the factor becomes unity. It might not be obvious, but the normalized cross correlation *NCC* coefficient can be derived from the 2T2D SES:

$$\sqrt{-\frac{D_{A2T}^2}{\sum_{j=1}^{l} s\left(\tilde{\nu}_j\right)^2 \sum_{k=1}^{l} m\left(\tilde{\nu}_k\right)^2} - \frac{1}{2}} = \frac{\sum_{j=1}^{l} m\left(\tilde{\nu}_j\right) s\left(\tilde{\nu}_j\right)}{\sqrt{\sum_{j=1}^{l} s\left(\tilde{\nu}_j\right)^2 \sum_{k=1}^{l} m\left(\tilde{\nu}_k\right)^2}} = NCC\,. \tag{19}$$

As long as $m = C \cdot s$, with an arbitrary factor *C*, $NCC = 1$, otherwise $-1 < NCC < 1$. In other words, $-NCC$ can also be employed as a loss function and then shares the property of the 2T2D *smart error sum* that the experimental and the simulated spectrum can differ by a factor. Quite often the NCC is used in a form that

is mean-centered, to be more precise zero mean-centered, which is then called zero-mean normalized cross correlation $ZNCC$,

$$ZNCC = \frac{1}{l+1} \frac{\sum_{j=1}^{l} \left(m\left(\tilde{\nu}_j\right) - \mu_m \right) \left(s\left(\tilde{\nu}_j\right) - \mu_s \right)}{\sigma_m \sigma_s}, \tag{20}$$

where μ_i are the mean spectral intensities of m and s and σ_i are their standard deviations. In contrast to NCC and the 2T2D *smart error sum*, the $ZNCC$ is additionally immune to offsets O: $m = C \cdot s + O$. To show the main features of the *smart error sums* we use Cauchy-type model distributions of the general form,

$$f(x,t) = t(1+at) \frac{\gamma}{(x - x_0 - bt)^2 + \gamma^2}, \tag{21}$$

to generate curves that we fit with the same type of functions. If the constants a and b are small, the function does not deviate noticeably from the Cauchy-distribution that is depicted in Fig. 1. The larger b is, the more the maximum shifts to smaller x-values for increasing t. The parameter a induces a non-linear increase of the amplitude.

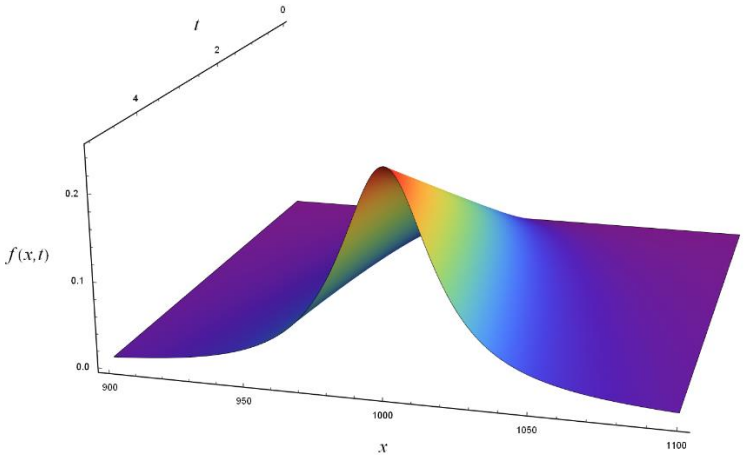

Fig. 1: Cauchy-type function that was used throughout this work. The curve depicted was generated with $x_0 = 1000$, $\gamma = 30$, $a = b = 0$.

The fits were performed by a corresponding custom-made program based on Wolfram Mathematica® 10.3 which is provided as supporting information together with a variant programmed in Julia.

Results and discussion

The crucial point for applying the *smart error sum* is the appearance or nature of the synchronous and asynchronous map or, alternatively, the map of the phase angles. For the *smart error sum* based on a series of curves for different perturbations *t* (eqn. (10), methods), it is pivotal that the hybrid synchronous map is not already symmetrical to the diagonal from low to high *x*, as it is for non-hybrid maps, and that the hybrid asynchronous map is not zero. Unfortunately, this case is not uncommon, e.g., for the Cauchy-type functions if *b* equals zero (cf. eqn. (21), methods). This may come as a surprise, because if you are familiar with 2D correlation spectroscopy, then you know that the asynchronous map is supposed to be nonzero if "the dynamic spectrum behaves nonlinearly with respect to the external variable", i.e. the perturbation.[22] We can introduce such a nonlinearity by setting $a \neq 0$ in eqn. (21). But, as long as $b = 0$, the asynchronous map will remain zero everywhere, which does not change even if we multiply all curves with a constant factor to cause $\mathbf{Y}_1 \neq \mathbf{Y}_2$. In fact, it seems that it is not a nonlinear change of $f(x,t)$ in *t* that results in a non-zero asynchronous map, but the change must be disproportionate. Such a change can be induced by setting $b \neq 0$, because then the maximum of the distribution downshifts increasingly if *t* increases.

It looks like the presence of such a disproportionate change with increasing *t* is the criterion that must be fulfilled for the *smart error sums* to work, including the ones that are based on 2T2D-correlation as well as NCC (eqn. (19)) and ZNCC (eqn. (20)). Accordingly, the synchronous map presents the proportionate changes and the asynchronous map the disproportionate changes of $f(x,t)$ with *t*. The different synchronous and asynchronous maps for the Cauchy-type functions are displayed in Fig. 2. The synchronous maps are for all investigated functions quite similar. In contrast, the asynchronous maps stay zero for linear and quadratic ($a = 1$) proportionate changes, while the map becomes non-zero for disproportionate changes ($b = 1$) and shows a typical pair of cross-peaks indicating a shift of the peak maximum. The same holds true for the asynchronous 2T2D maps, except that two cases have to be distinguished for $b = 1$. These are the case $t_1 = t_2$, for which the asynchronous map is still zero, whereas it is similar to the conventional

asynchronous map for the second case for which $t_1 \neq t_2$. This agrees with the definition of the *smart error sum* in this case, since *t* itself can certainly be one of the fit parameters.

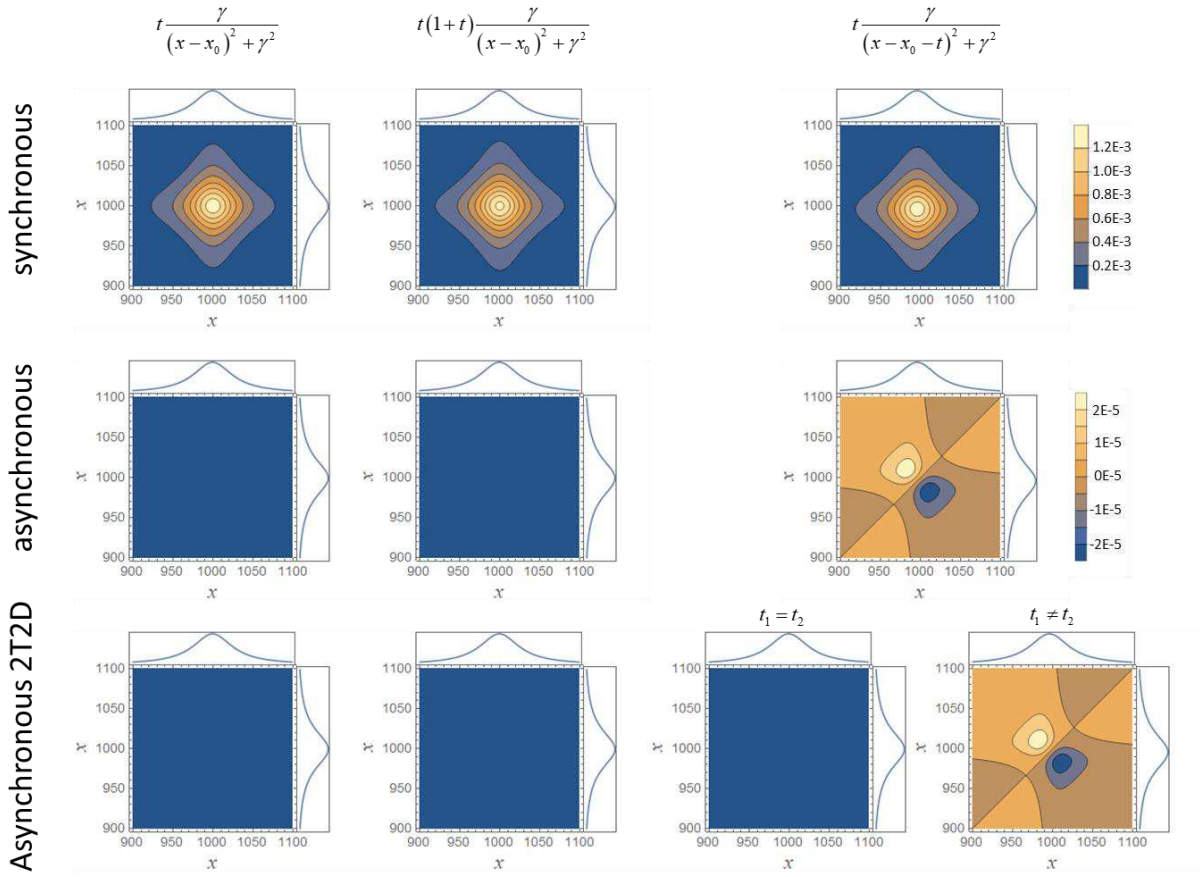

Fig. 2: Comparison of the hybrid synchronous and asynchronous maps of the Cauchy-type functions used in this work ($\mathbf{Y}_1 = \mathbf{Y}_2$).

In fact, the example that we want to showcase in the following is about fitting *t*. For a reader familiar with 2D correlation spectroscopy this may seem surprising, since *t* is usually assumed to simply increase systematically, which is important as otherwise semiquantitative deductions about the relative order of spectral changes are not possible. Note that 2D-COS can also be used in case of unevenly spaced increments of the perturbation, a corresponding extension of eqs. (2)-(4) for such increments has been provided.[36] If the sequence of the dynamic spectra is unknown, a computation of the asynchronous spectrum is not meaningful in contrast to the synchronous spectrum.[37] For applications of the smart error sum, however, neither equidistance nor the order of *t* values is of importance, which is why *t* can also be a fit parameter.

We nevertheless generated 5 curves by assuming $t \in \{1,2,3,4,5\}$, $x_0 = 1000$ and $\gamma = 30$ in the range of $900 < x < 1100$ and added 10 % systematic error by multiplying the curves by 1.1. Note that this is an oversimplification of a real situation, where the factor would depend on x – otherwise it would simply be possible to correct the spectra by introducing this factor as an additional fit parameter, but here we focus on demonstrating the method in a simple setting. The erroneous data were subsequently fitted by employing the conventional sum of squares and with the different *smart error sums*. The results are depicted in Fig. 3 and in Table 1. Obviously, apparently perfect fits are possible employing the conventional sum of squares by adjusting the parameter t.

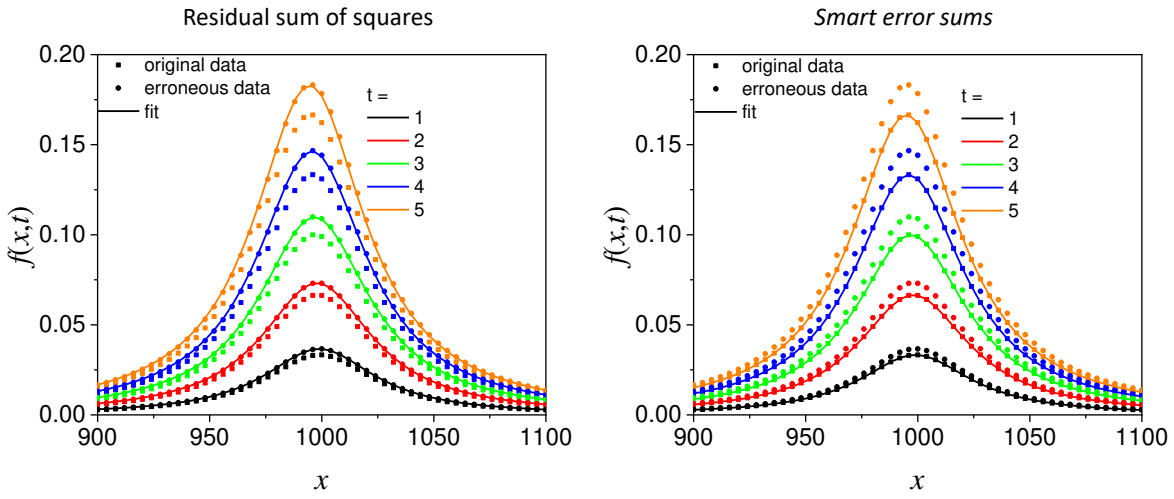

Fig. 3: Fit of the erroneous data with the conventional residual sum of squares (left panel) and with the *smart error sums* (right panel, the results obtained with the different error sums agree within line thickness except for eqn. (18) if $t \in \{1,3\}$).

Note that for the 2T2D *smart error sums* following eqs. (12) and (17), for which the latter is based on the phase angle, the convergence is fast enough so that the original t values are virtually recovered. While eqn. (14) allows a much faster fit, Mathematica returns a slightly worse result. The convergence is much slower for the series based *smart error sums*. In principle, instead of adding the logarithm of SRSS and ARSS, an alternative for connecting both residual sums would be to use the product (in case of an addition, a weighting would be necessary, since the ARSS usually is several orders of magnitude smaller). However, we found

that the convergence would be much worse, which is why we prefer eqn. (10) over using the product of SRSS and ARSS (the use of either alone further deteriorates the result, in particular if SRSS is used).

Table 1: Results of the fits based on the different errors sums and relative time needed.

Eqn.	t_1	t_2	t_3	t_4	t_5	Rel. timing
(1)	1.09993	2.19944	3.29812	4.39557	5.4914	1
(10)	1.00067	1.99862	3.00092	3.99983	4.99997	105.9
(12)	1	2	3	4	5	29.5
(14)	0.99983	2.00001	3.00001	4	4.99999	2.3
(17)	1	2	3	4	5	46.4
(18)	0.956611	2.02517	2.95784	4.00158	4.99879	110.2
(19)	0.999966	2.00002	3.00002	4	5	11.0
(20)	0.999999	1.99999	3	4	5	62.5

In particular the phase angle-based *smart error sum*, eqn. (18), has a comparably slow convergence, so that the fitted values differ considerably from the original values with the default settings of Mathematica's *NMinimize*, even though we added the condition that the solutions must be in the interval of ±1 of the original value (cf. Table 1). Obviously, although it points to the correct values, the phase angle and series based *smart error sum* has by far the worst convergence independent of which of Mathematica's built-in methods is chosen ("Nelder-Mead",[38] "DifferentialEvolution" "SimulatedAnnealing" and "RandomSearch"). The conventional correlation-based *smart error sums* NCC (eqn. (19)) and ZNCC (eqn. (20)) are slower than the faster 2T2D *smart error sum*. Not only do they show a somewhat inferior convergence, in real life applications where the error does not consist of a multiplicative error that is independent of x, they show also an inferior performance due to normalization. In particular ZNCC does not show any advantage compared to an also zero mean-centered and normalized residual sum of squares (not shown), which is in line with its poorer performance compared to NCC for pattern recognition in image analysis.[39]

For the conventional residual sum of squares, the *t* values completely reflect the error of the data (cf. Table 1), but the nearly perfect adaption to the altered data belies about the failure and can cheat the user into believing that the parameters obtained from the fit are errorless. Not only that, it can even mislead the user into believing that the model that is applied is correct. As mentioned above, if $b \neq 0$, then the maximum of the curves is increasingly shifted to lower *x* with increasing *t*. Nevertheless, it is also possible to fit the curves under the assumption $b = 0$ if x_0 is allowed to be one of the fit parameters – in this case the fit of the erroneous curve is perfect for $x_0 = \{999, 998, 997, 996, 995\}$ and $t = \{1.1, 2.2, 3.3, 4.4, 5.5\}$, although the employed fit function is wrong.

If *smart error sums* are employed, this erroneous situation cannot occur, because in this case the fit cannot converge to a result. Hybrid 2D correlation analysis reveals the reason for this "failure" as can be seen in Fig. 4. For mixtures of different functions, the asynchronous maps do not show the expected form, i.e., the values above the diagonal from small to large *x* do not have in general the inverted sign of the values below the diagonal, even though for both series the parameters, with the exception of the constants *a* and *b*, were the same. As a consequence, fits employing one of the *smart error sums* must fail, since the antisymmetry of the asynchronous map with respect to the diagonal can never be reached. In other words, if the use of a *smart error sum* for the fit of a physical problem leads to convergence, the used theoretical model must be adequate for the problem. On the other hand, if a fit fails because the asynchronous map does not show the expected distribution of signs, the applied model is not adequate for the problem at hand.

To present a concrete and practical example, Cauchy-functions are assumed in large parts of the spectroscopic community to describe the absorption of light, i.e., absorbance peaks. In fact, Lorentz derived based on dispersion theory that such profiles (therefore they are often called Lorentz-profiles) are good approximations for weak oscillators.[40] In the Lorentz-profile, however, the band position is decoupled from the peak intensity, in contrast to dispersion theory (the coupling results from the polarization of matter by light). If a single band is assumed it can be shown that $b = 1/3$.[41] If the conventional residual sum of squares

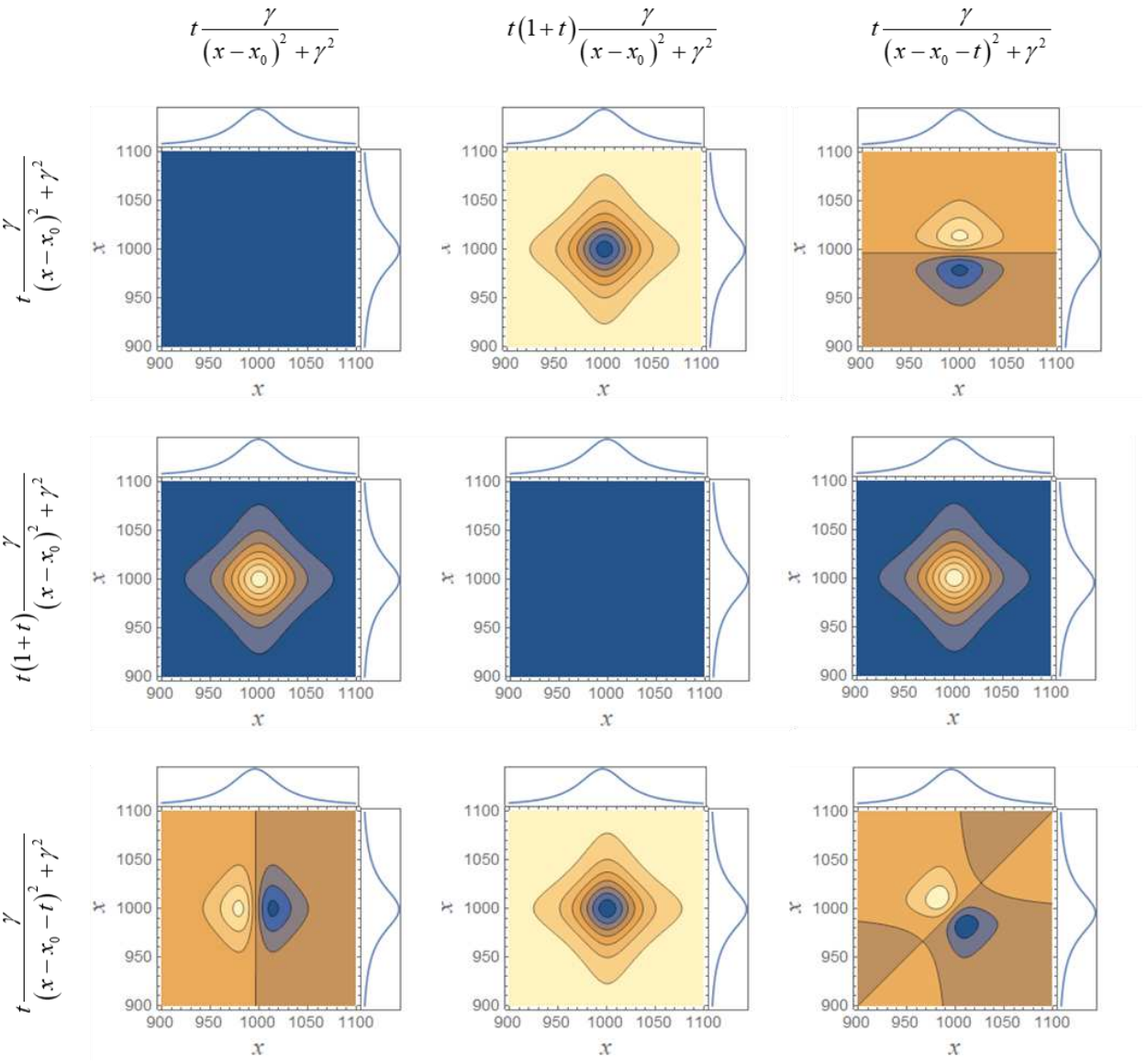

Fig. 4: Asynchronous 2D correlation maps for the mixture of different functions. For all parameters (except the constants *a* and *b*) the same values were chosen for both series.

is used, the Lorentz-profile can nevertheless be employed to fit the bands – as long as no series is fitted, which is the usual case, a band shift is simply compensated by changing the peak position as in the example discussed above. In contrast, a fit employing one of the *smart error sums* cannot succeed. The simple reason is that the asynchronous map, and, with it, the map of the phase angles, is not antisymmetric in the sense that the values above the diagonal have in general the opposite value of those below the diagonal, if the

model that was used to generate the original curves is different from the one the simulation is based on. Note that 2T2D maps behave differently and cannot be used to evaluate functional relations.

Again, it must be stated that this property can only be advantageously exploited for complex underlying laws that lead to disproportionate changes due to the perturbation. On the other hand, all less-complex relationships can be linearized and treated with the method of linear least squares, which provides analytical solutions. Therefore, curve fitting is not required. More importantly, such problems also do not require to employ neural networks/deep learning, methods which are specifically suitable to solve nonlinear modelling problems. For deep learning, on the other hand, it should be helpful to use loss functions based on *smart error sums*, because the latter obviously help to develop the correct functional relationships. In this stage this remains speculative, but we think that such loss functions do also support the training of neural networks as they can be effective measures to prevent underfitting and overfitting.

Acknowledgment

Financial support from the EU, the "Thüringer Ministerium für Wirtschaft, Wissenschaft und Digitale Gesellschaft", the "Thüringer Aufbaubank", the Federal Ministry of Education and Research, Germany (BMBF), the German Science Foundation, the "Fonds der Chemischen Industrie" and the Carl-Zeiss Foundation is gratefully acknowledged.

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

References

1 Strutz, T. *Data Fitting and Uncertainty: A practical introduction to weighted least squares and beyond*. (Springer Fachmedien Wiesbaden, 2015).

2 Zielesny, A. *From Curve Fitting to Machine Learning: An Illustrative Guide to Scientific Data Analysis and Computational Intelligence*. (Springer International Publishing, 2016).

3 Mayerhöfer, T. G. *et al.* Hybrid 2D Correlation-Based Loss Function for the Correction of Systematic Errors. *Anal. Chem.* **94**, 695-703 (2022). https://doi.org:10.1021/acs.analchem.1c02830

4 Wolberg, J. *Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments*. (Springer Berlin Heidelberg, 2006).

5 Otto, M. *Chemometrics: Statistics and Computer Application in Analytical Chemistry*. (Wiley, 2016).

6 Rastogi, A., Sahoo, S., Bandyopadhyay, T. K., Mukherjee, R. & Banerjee, R. Detailed morphological and kinetic studies of cellulose biosynthesis from Leifsonia soli. *Polymer* **242**, 124568 (2022). https://doi.org:https://doi.org/10.1016/j.polymer.2022.124568

7 Wazir-ud-Din, M. *et al.* Computerized glow curve deconvolution (CGCD): A comparison using asymptotic vs rational approximation in thermoluminescence kinetic models. *Appl. Radiat. Isot.* **179**, 110014 (2022). https://doi.org:https://doi.org/10.1016/j.apradiso.2021.110014

8 Mugi, V. R. & V.P, C. Comparison of drying kinetics, thermal and performance parameters during drying guava slices in natural and forced convection indirect solar dryers. *Solar Energy* **234**, 319-329 (2022). https://doi.org:https://doi.org/10.1016/j.solener.2022.02.012

9 Zhang, W. *et al.* Grain growth kinetics and densification mechanism of (TiZrHfVNbTa)C high-entropy ceramic under pressureless sintering. *Journal of Materials Science & Technology* **110**, 57-64 (2022). https://doi.org:https://doi.org/10.1016/j.jmst.2021.08.070

10 Sun, Z., Hu, D., Xie, L. & Ying, Y. Detection of early stage bruise in apples using optical property mapping. *Computers and Electronics in Agriculture* **194**, 106725 (2022). https://doi.org:https://doi.org/10.1016/j.compag.2022.106725

11 Zhang, Q. *et al.* Quantitative transport mapping (QTM) for differentiating benign and malignant breast lesion: Comparison with traditional kinetics modeling and semi-quantitative enhancement curve characteristics. *Magn Reson Imaging* **86**, 86-93 (2022). https://doi.org:10.1016/j.mri.2021.10.039

12 Meleties, M. *et al.* High-Throughput Microrheology for the Assessment of Protein Gelation Kinetics. *Macromolecules* **55**, 1239-1247 (2022). https://doi.org:10.1021/acs.macromol.1c02281

13 Hassan, M. *et al.* Transport pattern of Non-Newtonian mass and thermal energy under two diverse flow conditions by using modified models for thermodynamics properties. *Case Studies in Thermal Engineering* **29**, 101714 (2022). https://doi.org:https://doi.org/10.1016/j.csite.2021.101714

14 Cao, L. Road condition judgment system of railway transportation based on artificial intelligence recognition technology. *International Journal of System Assurance Engineering and Management* (2022). https://doi.org:10.1007/s13198-021-01509-w

15 Bagheri, M. & Rezania, M. Effect of Soil Moisture Evaporation Rate on Dynamic Measurement of Water Retention Curve with High-Capacity Tensiometer. *International*

Journal of Geomechanics **22**, 04021301 (2022). https://doi.org:doi:10.1061/(ASCE)GM.1943-5622.0002291

16 Filipponi, F., Smiraglia, D. & Agrillo, E. Earth Observation for Phenological Metrics (EO4PM): Temporal Discriminant to Characterize Forest Ecosystems. *Remote Sensing* **14**, 721 (2022). https://doi.org:10.3390/rs14030721

17 Alemu, A. Curve Fitting and Least Square Analysis to Extrapolate for the Case of COVID-19 Status in Ethiopia. *Advances in Infectious Diseases* **10**, 143-159 (2020). https://doi.org:10.4236/aid.2020.103015

18 Cui, L.-L., Li, Y.-F. & Long, P. Study on Coal Consumption Curve Fitting of the Thermal Power Based on Genetic Algorithm. *Journal of Power and Energy Engineering* **03**, 431-437 (2015). https://doi.org:10.4236/jpee.2015.34058

19 Ullah, M. S., Mostafa, G., Jahan, N. & Khan, M. Analyzing and Projection of Future Bangladesh Population Using Logistic Growth Model. *International Journal of Modern Nonlinear Theory and Application* **08**, 53-61 (2019). https://doi.org:10.4236/ijmnta.2019.83004

20 Major, G. H. *et al.* Practical guide for curve fitting in x-ray photoelectron spectroscopy. *Journal of Vacuum Science & Technology A* **38**, 061203 (2020). https://doi.org:10.1116/6.0000377

21 Mayerhöfer, T. G., Richard-Lacroix, M., Pahlow, S., Hübner, U. & Popp, J. Smart Error Sum Based on Hybrid Two-Trace Two-Dimensional (2T2D) Correlation Analysis. *Appl. Spectrosc.* **0**, 00037028221077310 (2022). https://doi.org:10.1177/00037028221077310

22 Noda, I. & Ozaki, Y. *Two-Dimensional Correlation Spectroscopy: Applications in Vibrational and Optical Spectroscopy*. (Wiley, 2005).

23 Noda, I. Two-trace two-dimensional (2T2D) correlation spectroscopy – A method for extracting useful information from a pair of spectra. *J. Mol. Struct.* **1160**, 471-478 (2018). https://doi.org:https://doi.org/10.1016/j.molstruc.2018.01.091

24 Noda, I. Closer examination of two-trace two-dimensional (2T2D) correlation spectroscopy. *J. Mol. Struct.* **1213**, 128194 (2020). https://doi.org:https://doi.org/10.1016/j.molstruc.2020.128194

25 Hisham, M. B., Yaakob, S. N., Raof, R. A. A., Nazren, A. B. A. & Wafi, N. M. in *2015 IEEE Student Conference on Research and Development (SCOReD)*. 100-104.

26 Cui, Z., Qi, W. & Liu, Y. A Fast Image Template Matching Algorithm Based on Normalized Cross Correlation. *Journal of Physics: Conference Series* **1693**, 012163 (2020). https://doi.org:10.1088/1742-6596/1693/1/012163

27 Xu, J., He, X. & Ji, W. Mechanical System and Template-Matching-Based Position-Measuring Method for Automatic Spool Positioning and Loading in Welding Wire Winding. *Applied Sciences* **10**, 3762 (2020).

28 Hamada, M., Dérian, P., Mauzey, C. F. & Mayor, S. D. Optimization of the Cross-Correlation Algorithm for Two-Component Wind Field Estimation from Single Aerosol Lidar Data and Comparison with Doppler Lidar. *Journal of Atmospheric and Oceanic Technology* **33**, 81-101 (2016). https://doi.org:10.1175/jtech-d-15-0009.1

29 Liu, Q., Tokunaga, T. & He, Z. Realization of nano static strain sensing with fiber Bragg gratings interrogated by narrow linewidth tunable lasers. *Opt. Express* **19**, 20214-20223 (2011). https://doi.org:10.1364/OE.19.020214

30 Theodosiou, A., Komodromos, M. & Kalli, K. Improvements on the cross-correlation algorithm used for tracking fractional Bragg grating wavelength shifts in multimode

fibres. *Optical Fiber Technology* **46**, 36-42 (2018). https://doi.org:https://doi.org/10.1016/j.yofte.2018.09.008

31 Woodford, O. Using Normalized Cross Correlation in Least Squares Optimizations. (2018). <https://doi.org/10.48550/arxiv.1810.04320>.

32 Czarnecki, M. A. Two-Dimensional Correlation Spectroscopy: Effect of Reference Spectrum on Noise-Free and Noisy Spectra. *Appl. Spectrosc.* **57**, 991-995 (2003). https://doi.org:10.1366/000370203322258968

33 Mayerhöfer, T. G., Ilchenko, O., Kutsyk, A. & Popp, J. Beyond Beer's Law: Quasi-Ideal Binary Liquid Mixtures. *Appl. Spectrosc.* **76**, 92-104 (2022). https://doi.org:10.1177/00037028211056293

34 Shinzawa, H., Jiang, J.-H., Iwahashi, M., Noda, I. & Ozaki, Y. Self-modeling curve resolution (SMCR) by particle swarm optimization (PSO). *Anal. Chim. Acta* **595**, 275-281 (2007). https://doi.org:https://doi.org/10.1016/j.aca.2006.12.004

35 Shinzawa, H., Iwahashi, M., Noda, I. & Ozaki, Y. A convergence criterion in alternating least squares (ALS) by global phase angle. *J. Mol. Struct.* **883-884**, 73-78 (2008). https://doi.org:https://doi.org/10.1016/j.molstruc.2008.01.033

36 Isao, N. Two-Dimensional Correlation Analysis of Unevenly Spaced Spectral Data. *Appl. Spectrosc.* **57**, 1049--1051 (2003).

37 Noda, I. Two-dimensional correlation analysis of spectra collected without knowing sampling order. *J. Mol. Struct.* **1156**, 418-423 (2018). https://doi.org:https://doi.org/10.1016/j.molstruc.2017.11.085

38 Nelder, J. A. & Mead, R. A SIMPLEX-METHOD FOR FUNCTION MINIMIZATION. *Computer Journal* **7**, 308-313 (1965).

39 Martin, J. & Crowley, J. in *Proceedings of the Intelligent Autonomous System* (IOS Press, 1995).

40 Lorentz, H. A. The absoption and emission lines of gaseous bodies. *Koninkl. Ned. Akad. Wetenschap. Proc.* **8**, 591-611 (1906).

41 Mayerhöfer, T. G. *Wave Optics in Infrared Spectroscopy*. (DOI: 10.13140/RG.2.2.14293.55520, 2021).