

Article

Not peer-reviewed version

---

# Uncertainty Propagation using Dispersion Matrices Accounting for Systematic Error in Least-Squares Regression

---

[Rainer Feistel](#)<sup>\*</sup> and Jeremy Lovell-Smith

Posted Date: 30 November 2023

doi: 10.20944/preprints202311.1917.v1

Keywords: empirical equations; generalized least-squares regression; weight matrix; systematic errors; covariance propagation; projection matrix; residual structure



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# Uncertainty Propagation using Dispersion Matrices Accounting for Systematic Error in Least-Squares Regression

R Feistel <sup>1,\*</sup> and J W Lovell-Smith <sup>2</sup>

<sup>1</sup> Leibniz Institute for Baltic Sea Research (IOW), D-18119 Warnemünde, Germany

<sup>2</sup> Measurement Standards Laboratory (MSL), PO Box 31-310, Lower Hutt, New Zealand

\* Correspondence: rainer.feistel@io-warnemuende.de

**Abstract:** Empirical equations representing, interpolating and smoothing groups of measurement results by regression methods are in widespread use in metrology and other fields of science and engineering. Standard equations for the propagation of measurement uncertainties to values computed from such equations are available but suffer from a lack of general acceptance and are only infrequently applied in practice. One reason for the slow uptake is the lack of clear methods to account for systematic error. In this paper, uncertainty propagation equations in terms of covariance matrices are generalized to allow for systematic errors in least-squares regression, and effects of the resulting uncertainties are investigated analytically. A stochastic ensemble model of systematic error is proposed for the computation of the non-diagonal elements of the weight matrix of the Generalized Least-Squares method (GLS) from the measurement uncertainty. A GLS projector formalism is described which separates the effects of measurement scatter on values calculated by the equation from those on the related “residuals”, i.e., the residual errors in the fitted data. The same projectors act similarly on the associated uncertainties and covariance matrices and permit the effects of systematic errors on the simulation covariance to be quantified. It is demonstrated that, in order to include systematic errors in the uncertainty estimates of GLS, covariance matrices may be substituted by novel, specifically defined *dispersion matrices* that are not specified yet by the GUM<sup>1</sup>. Systematic measurement errors may be estimated from various sources; a particularly easy way suggested here involves analyzing the structure in the regression residuals. It is demonstrated that GLS projectors exhibit inherent features of “error blindness” also with respect to systematic errors.

**Keywords:** empirical equations; generalized least-squares regression; weight matrix; systematic errors; covariance propagation; projection matrix; residual structure

## 1. Introduction

Since Carl Friedrich Gauss had developed the mathematical method of least squares in 1795, empirical equations fitted to data have become widely used in metrology and various other branches of science and engineering. However, even though knowledge about uncertainties of values calculated from such equations is indispensable, comprehensive, systematic and generally accepted techniques for estimating those uncertainties are still not available, and related fundamental questions continue to be disputed (Feistel *et al* 2016). However, “the confidence level of a model’s predictions should be included in every modelling application” (Deletic *et al* 2012). In particular the handling of systematic errors in fitted data is insufficiently well known.

---

<sup>1</sup> GUM: Guide to the Expression of Uncertainty in Measurement,  
<http://www.bipm.org/en/publications/guides>

The motivation for developing the mathematical model presented in this paper arose from metrological problems encountered in the context of a new definition of relative humidity, RH (Lovell-Smith *et al* 2016, Feistel and Lovell-Smith 2017). A new, physically justified definition of RH, traceable to the International System of Units, SI, may rely on thermodynamic properties such as chemical potentials, activities, fugacities or virial coefficients in aqueous gaseous, liquid or solid mixtures. Equations for such properties are available from documents of IAPWS<sup>2</sup>, in particular as part of the international thermodynamic standard TEOS-10 (IOC *et al* 2010, IAPWS AN6-16 2016), while equations for the related uncertainty information are not or only incompletely reported for those quantities. For example, virial coefficients of water can be derived from the equation of state, but their uncertainties cannot (Feistel *et al* 2015). Systematic methods for estimating such uncertainties from the equations available are missing or lack of general acceptance. An attempt of applying established GUM (2008, 2011) covariance propagation methods to the IAPWS equations resulted in seemingly significantly underestimated uncertainties derived for the properties of interest (Feistel *et al* 2016, Lovell-Smith *et al* 2017). Among the most likely reasons for this apparent deficiency may be the unclear way of propagating systematic errors in the measurement results that were used for the construction of the fundamental IAPWS equations.

For proper metrological application of IAPWS equations, sufficiently complete and appropriate uncertainty information is indispensable. The method suggested in this paper aims at addressing associated problems that are unresolved by now. While the mathematical formalism proposed here is rather general and may be suitable also for other cases of Generalized Least-Squares (GLS) regression, the focus is here on empirical equations whose adjustable coefficients had been obtained from large sets of various thermodynamic properties of the target substance, in particular, of water, by means of Weighted Least-Squares (WLS) regression. The authors of the original measurement data are typically not involved in the construction of the equations, and any uncertainty or possibly covariance information related to those measurements is restricted to what those authors had provided along with their publications. The aim of this paper is making available lacking estimates for input data correlations, by analysing the combined regression of data belonging to different authors, quantities or experimental setups. This situation is radically different from cases where the experimenters are able include correlation information of their own measurements in uncertainty budgets, such as for the determination of the Boltzmann constant (Gaiser *et al* 2017). Populating non-diagonal elements of the least-squares weight matrix based on simple assumptions have not led to sufficiently improved uncertainty estimates (Feistel *et al* 2016, Lovell-Smith *et al* 2017). The approach developed in this paper goes clearly beyond the model of Davidson and MacKinnon (2004, equation 7.88 therein) which is, to the knowledge of the authors, the so-far most advanced GLS algorithm available.

When an empirical thermodynamic equation of state is constructed for a certain substance such as H<sub>2</sub>O ice (Feistel and Wagner 2006), measurement results from various origins for various properties of ice are combined in a joint least-squares regression analysis. This way, all different thermodynamic properties become expressed as functions of a single common background set of regression coefficients. The corresponding close mutual mathematical relations between the properties, such as Maxwell relations, enforce internal consistency of the entire input data set; possible systematic errors of particular data groups typically result in conspicuous non-random pattern of their fitting residuals. This paper attempts to exploit this, otherwise commonly ignored pattern for estimating off-diagonal elements of the weight matrix. This approach is conceptionally consistent, as we believe, with recent suggestions of Willink (2013) and de Courtenay and Grégis (2017) for a “frequentist” method of including the systematic error of a single data set in the form of a statistical error revealed by the joint exploitation of related, alternative data sets. “A given systematic error appears like a realization drawn from a population of similar systematic errors of the same kind produced by all the laboratories that were accessible to the user and that were handling

---

<sup>2</sup> IAPWS: International Association for the Properties of Water and Steam, <http://www.iapws.org>

the same type of problem” (de Courtenay and Grégis 2017, p 9 therein). For the construction of an empirical thermodynamic potential, the naturally required combination of various thermodynamic properties measured by various methods and authors represents a most suitable option for the application of this concept. Typically, systematic errors remain unnoticed when a series of measurements is carried out with the same equipment and experimental procedure, while random errors usually disclose themselves by scattering values of repeated readings. Published data are often accompanied by uncertainty estimates that cover the scatter range, but hardly ever by covariance information reflecting concealed systematic offsets or trends. Comparison of results from different studies may reveal the existence of systematic deviations between them by discernable structures in graphical presentations.

It is clear that residual patterns may also result from inappropriate choices made for the regression functions, unrelated to possible systematic errors of the background data. However, there are no systematic and rigorous methods available for choosing the best set of regression functions. Here, we rely on the experience and careful data analysis carried out along with the regression procedure, so that trivial mistakes may largely be ruled out as possible reasons for residual structures of the fit. This assumption is reasonable for empirical equations endorsed by IAPWS and other published reference-quality equations.

Uncertainty of input data propagates to uncertainty of regression coefficients, which in turn propagates to any value computed from such an empirical equation. The mathematical framework for this propagation described by the GUM (2008, 2011) can be applied to regression methods in the form of the Generalized Least Squares (GLS) technique (Feistel *et al* 2016), where a covariance matrix associated with the regression coefficients is derived from the covariance matrix of the input data, whose diagonal elements (variances) express standard uncertainties, while the off-diagonal elements (covariances) may allow for effects of systematic errors (Draper and Smith 1998, Davidson and MacKinnon 1993, 2004, Lovell-Smith *et al* 2017) but are generally unavailable. Unfortunately, uncertainties of values computed from the empirical equation using this GLS covariance propagation method seem to be systematically and substantially underestimated, at least in part because the off-diagonal elements of the input covariance matrix (Feistel *et al* 2016) are unknown. This lack of knowledge is a major impediment to the ability to calculate and propagate uncertainty accurately using GLS. However, it may often be possible to populate the off-diagonal elements through analysis of structure in the residual error.

In this paper, an extended approach supporting the GLS method is suggested for including systematic errors in the regression procedure by generalizing the covariance matrix (expressing only random error) to a *dispersion matrix* (expressing random and systematic error). For this purpose, the stochastic ensemble of randomly varied input data, that was used previously to derive covariance propagation equations (Feistel 2011), is generalized here to include also systematic errors derived from additional sources such as structured residuals of a fit. While regression analyses generally act to reduce not just residual error and its structure, the new approach outlined here attempts to exploit the remaining structure, which represents readily available and relevant additional information.

In Section 2, the basic relations of the GLS method are summarized in vector-matrix notation. In Section 3, the term *simulation* is used for the regression results that reproduce (with certain deviations) the original measurements. The projector formalism of Aitken (1936) for the description of simulation results and related residuals by the GLS method is generalized here for the propagation of covariance and dispersion matrices. A list of mathematical properties of the projection matrices is reported, including inequalities for the simulation dispersion matrix. In Section 4, a stochastic ensemble is used as a model for the statistical description of potential errors, and introduced to the GLS weight matrix. For an ensemble describing random and systematic errors, an analytical formula of the weight matrix is given. In particular, to allow for systematic errors in the uncertainty propagation equations, a novel dispersion matrix will be defined as a suitable generalization of the conventional covariance matrix used for GLS regression. In Section 5, the application of the projector formalism of Section 3 to the error ensemble of Section 4 permits a description of the effect of systematic errors on the simulation dispersion matrix of the GLS method. In Section 6, a simple approach is suggested how systematic

errors may be estimated from structured residuals of a fit, and this way included in the dispersion and uncertainty propagation equations of the GLS method. In Section 7, it is demonstrated that the “error blindness” inherent to regression methods applies also to uncertainties derived from systematic errors by the new dispersion matrix method proposed in this paper. Section 8 concludes that allowing for systematic errors in the uncertainty propagation of an empirical thermodynamic potential does not necessarily raise the resulting uncertainties to the quantitative level given by the scatter range of the input data. This “error blindness” is immediately visible in a tutorial regression example presented in the Appendix.

## 2. Generalised Least-Squares Equations

Let a given set of  $M$  measurement results,  $\mathbf{y} = \{\mathbf{y}_i\}$ ,  $i = 1, \dots, M$ , be observed under the conditions  $\mathbf{X} = \{\mathbf{x}_i\}$ , at which the particular experiment,  $i$ , was conducted. For example,  $y_1$  may be the density of a given substance measured at the temperature  $T_1$  and the pressure  $p_1$ , i.e.,  $\mathbf{x}_1 = (T_1, p_1)^T$ , and  $y_2$  may be the heat capacity of the same substance measured at the temperature  $T_2$  and the pressure  $p_2$ , i.e.,  $\mathbf{x}_2 = (T_2, p_2)^T$ , etc. In the vector-matrix notation used here, a vector is conventionally a column vector, which is the transpose of a row vector, as indicated the superscript T. A combined standard uncertainty (GUM 2008),  $u_i = u_c(\mathbf{y}_i)$ , is associated with each measurement result, which is assumed to also include contributions from uncertainties of measuring the conditions  $\mathbf{x}_i$ . This approach is consistent with numerous experimental works where uncertainties are reported of the desired quantity but rarely of the conditions under which the measurement took place. The formalism for propagation of uncertainties of independent variables, if known, to the dependent regression variable is routinely available from standard software packages (Cox and Harris 2006, Saunders 2014).

The measured values are fitted by a set of related empirical functions,  $y_i \approx f_i(\mathbf{a}, \mathbf{x}_i)$ , each of which depends on the same set of  $N < M$  adjustable parameters,  $\mathbf{a} = \{\mathbf{a}_i\}$ ,  $i = 1, \dots, N$ , regarded as the *regression coefficients*. The coefficients  $\mathbf{a}$  are found by minimising a *generalised least-squares* (GLS) expression,  $p(\mathbf{a})$ , of the form (Davidson and MacKinnon 1993, 2004, Draper and Smith 1998, Strutz 2011, Feistel *et al* 2016, Lovell-Smith *et al* 2017),

$$p(\mathbf{a}) \equiv (\mathbf{y} - \mathbf{f})^T \mathbf{W} (\mathbf{y} - \mathbf{f}) = \text{Min}, \quad (1)$$

with a suitably specified symmetric and positive definite  $(M \times M)$  *weight matrix*  $\mathbf{W}$ , whose inverse,

$$\mathbf{U}_y \equiv \mathbf{W}^{-1}, \quad (2)$$

is known as the *covariance matrix* of the input values,  $\mathbf{y}$ , and  $(\mathbf{y} - \mathbf{f})$  is the vector of residual errors of the fit. Without relevant loss of generality, it is sufficient to consider in this paper only functions  $f$  that are linear functions of  $\mathbf{a}$  because the validity of linear covariance propagation equations exploited in the following is restricted to a small vicinity of a local minimum of equation (1).

For functions  $f$  linear in  $\mathbf{a}$ , and a regular matrix  $\mathbf{A}$ , a unique solution of the problem (1) exists for the regression coefficients (Aitken 1936, Davidson and MacKinnon 1993),

$$\mathbf{a} = \mathbf{A}^{-1} \mathbf{J}_f \mathbf{W} \mathbf{y}. \quad (3)$$

Here,  $\mathbf{A}$  is the symmetric  $(N \times N)$  *Gaussian normal matrix* of the problem (1),

$$\mathbf{A} \equiv \mathbf{J}_f \mathbf{W} \mathbf{J}_f^T, \quad (4)$$

that is assumed here to be regular,  $\det(\mathbf{A}) \neq 0$ , and

$$\mathbf{J}_f \equiv \frac{\partial \mathbf{f}}{\partial \mathbf{a}} \quad (5)$$

is the  $(N \times M)$  *Jacobian* (or *sensitivity matrix*) of the empirical functions  $f$  with respect to the regression coefficients,  $\mathbf{a}$ , evaluated at the matrix of the independent variables,  $\mathbf{X}$ , of the input vector,  $\mathbf{y}$ . For functions  $f$  linear in  $\mathbf{a}$ , the Jacobian  $\mathbf{J}_f$  is a matrix independent of the regression coefficients,  $\mathbf{a}$ . Without loss of generality, any intercept of  $f$  at  $\mathbf{a} = 0$  may have been subtracted from  $\mathbf{y}$  in advance, so that for simplicity of the equations we may consider only intercept-free functions  $f$  of the form



$$\mathbf{f} = \mathbf{J}_f^T \mathbf{a}. \quad (6)$$

Following the GUM (2011) covariance propagation method, the covariance matrix of the regression coefficients  $\mathbf{a}$  is (Feistel *et al* 2016)

$$\mathbf{U}_a = \mathbf{A}^{-1}. \quad (7)$$

The minimum which equation (1) takes is known as the *goodness of fit*,  $p(\mathbf{a})$ , which can be expressed by means of equation (3),

$$p(\mathbf{a}) = (\mathbf{y} - \mathbf{J}_f^T \mathbf{a})^T \mathbf{W} (\mathbf{y} - \mathbf{J}_f^T \mathbf{a}) = \mathbf{y}^T (\mathbf{W} - \mathbf{W} \mathbf{J}_f^T \mathbf{A}^{-1} \mathbf{J}_f \mathbf{W}) \mathbf{y}. \quad (8)$$

While the goodness of fit is a clear and easily available indicator for a suitable choice of the regression function and of the weights used for the data points, it only insufficiently reflects possible systematic structures within the distribution of the residuals. Checking the value (8) is therefore only a first necessary step before a more thorough analysis may be performed, such as by graphical inspection of the data scatter.

The purpose of the determined set of empirical regression coefficients  $\mathbf{a}$  is that certain functions of interest,  $\mathbf{h}(\mathbf{a}, \mathbf{x})$  may be *predicted* for certain (interpolated or extrapolated) independent values,  $\mathbf{x}$ , such as the entropy of the substance under study at special  $T$ - $p$  conditions where no measurements had been made. Following the GUM (2011) covariance propagation method, the symmetric  $(L \times L)$  covariance matrix,  $\mathbf{U}_h$ , of a set of  $L$  predicted quantities,  $\mathbf{h}$ , is computed from the covariance matrix,  $\mathbf{U}_a$ , equation (7), of the regression coefficients by (Feistel *et al* 2016),

$$\mathbf{U}_h \equiv \mathbf{J}_h^T \mathbf{U}_a \mathbf{J}_h, \quad (9)$$

where

$$\mathbf{J}_h \equiv \frac{\partial \mathbf{h}^T}{\partial \mathbf{a}} \quad (10)$$

is the  $(N \times L)$  *Jacobian* (or *sensitivity matrix*) of the prediction functions  $\mathbf{h}$  with respect to the regression coefficients,  $\mathbf{a}$ , evaluated at the independent variables,  $\mathbf{x}$ , of the output vector of interest. Note that the functions  $\mathbf{h}$  may or may not include some of the empirical functions  $\mathbf{f}$ , and the independent variables  $\mathbf{x}$  may or may not include some of the points where the input values  $\mathbf{y}$  were measured. There is no restriction to the choice of the number  $L$ . The main diagonal elements of the matrix  $\mathbf{U}_h$  represent the squares of the propagated combined uncertainties,  $u_c(\mathbf{h})$ , of the predicted values,  $\mathbf{h}$ .

As the covariance propagation equations are linear in the covariance matrices, equations (4), (7), (9), any coverage factor applied to input uncertainties is accordingly returned by the output uncertainties. "An *expanded uncertainty*  $U$  is obtained by multiplying the combined standard uncertainty  $u_c$  by a *coverage factor*  $k$ . ... The choice of the factor  $k$ , which is usually in the range 2 to 3, is based on the coverage probability or level of confidence required of the interval" (GUM 2008 section 3.3.7 therein). Coverage factors may be applied to the standard uncertainty regardless of whether the underlying statistics is Gaussian or not.

From equation (9) it is clear that providing the covariance matrix along with the regression coefficients of an empirical equation will empower users of the equation to determine uncertainties for any properties calculated from that equation, without needing the background data used in the equation's construction. However, in order to allow for systematic errors in the corresponding propagation equations reported above, more general dispersion matrices will be defined and suggested in section 4, to be used in place of covariance matrices that allow for random errors only.

### 3. Simulation Covariance Equations

The uncertainties of the GLS input data are related to those of the associated GLS results by a linear projector whose properties are derived in this section. For the computation of the residuals, the set of prediction functions  $\mathbf{h}$  is chosen to match the empirical functions  $\mathbf{f}$ , both evaluated at the same set of independent variables. This set of functions  $\mathbf{f}$  is regarded here as the *simulation* of the original measurements  $\mathbf{y}$  by the empirical equation. In this case, the Jacobians  $\mathbf{J}_f$ , equation (5), and  $\mathbf{J}_h$ , equation (10) are identical matrices,

$$\mathbf{J}_h = \mathbf{J}_f = \mathbf{J}. \quad (11)$$

For the related  $(M \times M)$  simulation covariance matrix,  $\mathbf{U}_f = \mathbf{U}_h$ , of the predicted values, we obtain from equations (7) and (9)

$$\mathbf{U}_f = \mathbf{J}^\top (\mathbf{J} \mathbf{W} \mathbf{J}^\top)^{-1} \mathbf{J} = \mathbf{J}^\top \mathbf{A}^{-1} \mathbf{J}. \quad (12)$$

Note that  $\mathbf{U}_f$  is symmetric

$$\mathbf{U}_f = \mathbf{U}_f^\top. \quad (13)$$

We define the  $(M \times M)$  GLS covariance propagation matrix,

$$\mathbf{P} \equiv \mathbf{U}_f \mathbf{U}_y^{-1} = \mathbf{U}_f \mathbf{W} = \mathbf{J}^\top (\mathbf{J} \mathbf{W} \mathbf{J}^\top)^{-1} \mathbf{J} \mathbf{W}, \quad (14)$$

which expresses the “ratio” of the output simulation covariance to the input measurement covariance, as mediated by the regression functions. This projector turns out to be equal to the projector of the simulation itself,  $\mathbf{f} = \mathbf{P} \mathbf{y}$ , originally introduced by Aitken (1936), see item (x) below.

Some general properties related to the matrix  $\mathbf{P}$  are summarized below. Further properties are reported in Appendix A.

- (i)  $\mathbf{P}$  is of rank<sup>3</sup>  $N$  and is singular. (15)

*Proof:*  $\mathbf{U}_f = \mathbf{P} \mathbf{U}_y$  is of rank  $N$ , see equation (A.5), while  $\mathbf{U}_y$  is of rank  $M$  with  $M > N$ , see equation (A.4). The assumption  $\text{rank}(\mathbf{P}) < N$  would result in  $\text{rank}(\mathbf{U}_f) < N$ , in contradiction to the first sentence. The assumption  $\text{rank}(\mathbf{P}) > N$  is in contradiction to equation (14).

So,  $\text{rank}(\mathbf{P}) = N$ .

- (ii)  $\mathbf{P}$  is idempotent,  $\mathbf{P}^2 = \mathbf{P}$ , (16)

that is,  $\mathbf{P}$  is a *projection matrix*, or *projector* (Marsaglia and Styan 1974).

- (iii)  $(\mathbf{I} - \mathbf{P})$  is idempotent,  $(\mathbf{I} - \mathbf{P})^2 = (\mathbf{I} - \mathbf{P})$ , that is,  $(\mathbf{I} - \mathbf{P})$  is a projection matrix and positive semi-definite. Here,  $\mathbf{I}$  is the  $(M \times M)$  unity matrix. (17)

- (iv)  $\mathbf{P}$  and  $(\mathbf{I} - \mathbf{P})$  are disjoint projectors,  $(\mathbf{I} - \mathbf{P})\mathbf{P} = \mathbf{P}(\mathbf{I} - \mathbf{P}) = \mathbf{0}$ . (18)

*Note:* Columns of  $\mathbf{P}$  are right eigenvectors of  $(\mathbf{I} - \mathbf{P})$  to the eigenvalue 0, as the rows of  $\mathbf{P}$  are such left eigenvectors. Similarly, columns of  $(\mathbf{I} - \mathbf{P})$  are right eigenvectors of  $\mathbf{P}$  to the eigenvalue 0, as the rows of  $(\mathbf{I} - \mathbf{P})$  are such left eigenvectors.

- (v) The columns of  $\mathbf{J}^\top$  are right eigenvectors of  $\mathbf{P}$  to the eigenvalue 1,  $\mathbf{P} \mathbf{J}^\top = \mathbf{J}^\top$ . (19)

- (vi) The rows of the  $(N \times M)$  matrix  $\mathbf{K}$ ,

$$\mathbf{K} \equiv (\mathbf{J} \mathbf{W} \mathbf{J}^\top)^{-1} \mathbf{J} \mathbf{W}, \quad (20)$$

are left eigenvectors of  $\mathbf{P}$  to the eigenvalue 1,  $\mathbf{K} \mathbf{P} = \mathbf{K}$ .

*Proof:* By the definition (14) of  $\mathbf{P}$ ,

$$\mathbf{K} \mathbf{P} = (\mathbf{J} \mathbf{W} \mathbf{J}^\top)^{-1} \mathbf{J} \mathbf{W} \mathbf{J}^\top (\mathbf{J} \mathbf{W} \mathbf{J}^\top)^{-1} \mathbf{J} \mathbf{W} = (\mathbf{J} \mathbf{W} \mathbf{J}^\top)^{-1} \mathbf{J} \mathbf{W} = \mathbf{K}. \quad (21)$$

- (vii) The eigenvectors given by  $\mathbf{J}$  and  $\mathbf{K}$  are pairwise mutually orthonormal,  $\mathbf{K} \mathbf{J}^\top = \mathbf{J} \mathbf{K}^\top = \mathbf{I}$ , where  $\mathbf{I}$  is the  $(N \times N)$  unity matrix. (22)

*Proof:* By the definition of  $\mathbf{J}$  and  $\mathbf{K}$ ,

$$\mathbf{K} \mathbf{J}^\top = (\mathbf{J} \mathbf{W} \mathbf{J}^\top)^{-1} \mathbf{J} \mathbf{W} \mathbf{J}^\top = \mathbf{I}, \quad (23)$$

and

$$\mathbf{J} \mathbf{K}^\top = \mathbf{J} \mathbf{W} \mathbf{J}^\top (\mathbf{J} \mathbf{W} \mathbf{J}^\top)^{-1} = \mathbf{I}. \quad (24)$$

*Note:* This general property of conjugate left and right eigenvectors is known as the *principle of biorthogonality* (Horn and Johnson 1984, §1.4.7 therein).

- (viii) The equation  $\mathbf{U}_f = \mathbf{P} \mathbf{U}_y$  can be rewritten in explicitly symmetric form,

$$\mathbf{U}_f = \mathbf{P} \mathbf{U}_y \mathbf{P}^\top. \quad (25)$$

---

<sup>3</sup> The rank of a matrix is the maximum number of linearly independent column vectors (or equivalently, row vectors) that form the matrix (Horn and Johnson 1984).

*Proof:* Multiplying  $\mathbf{U}_f = \mathbf{P}\mathbf{U}_y$  by  $\mathbf{P}^T$  from the right and applying equation (A.24) to the left gives equation (25).

- (ix) The regression coefficients,  $\mathbf{a}$ , are obtained from application of the matrix  $\mathbf{K}$ , equation (20), to the measurement vector,  $\mathbf{a} = \mathbf{K}\mathbf{y}$ . (26)

*Proof:* According to equation (20),  $\mathbf{a} = \mathbf{K}\mathbf{y} = (\mathbf{J}\mathbf{W}\mathbf{J}^T)^{-1} \mathbf{J}\mathbf{W}\mathbf{y}$  equals equation (3).

- (x) The simulated values,  $\mathbf{f}$ , are obtained from application of the projection matrix  $\mathbf{P}$ , equation (14), to the measurement vector (Aitken 1936, Davidson and MacKinnon 2004),  $\mathbf{f} = \mathbf{P}\mathbf{y}$  (27)

*Note:* The related covariance matrices propagate by the same projector,  $\mathbf{U}_f = \mathbf{P}\mathbf{U}_y$ . The projector  $\mathbf{P}$  will be termed *simulation projector* because it maps the input vector  $\mathbf{y}$  to the simulation vector  $\mathbf{f}$ .

*Proof:* By definition of  $\mathbf{f}$ , equation (6), and using equations (26), (A.9), it follows that  $\mathbf{f} = \mathbf{J}^T \mathbf{a} = \mathbf{J}^T \mathbf{K}\mathbf{y} = \mathbf{P}\mathbf{y}$ .

- (xi) The residuals,  $(\mathbf{y} - \mathbf{f})$ , satisfy the equation

$$\mathbf{K}(\mathbf{y} - \mathbf{f}) = \mathbf{0}. \quad (28)$$

*Proof:* Multiplying (27) by  $\mathbf{K}$  results in  $\mathbf{K}\mathbf{f} = \mathbf{K}\mathbf{P}\mathbf{y} = \mathbf{K}\mathbf{y}$ , because of (20).

- (xii) The residuals,  $(\mathbf{y} - \mathbf{f})$ , are obtained from application of the conjugate projection matrix,  $(\mathbf{I} - \mathbf{P})$ , equation (17), to the measurement vector,  $(\mathbf{y} - \mathbf{f}) = (\mathbf{I} - \mathbf{P})\mathbf{y}$ . (29)

*Note:* The difference between the related covariance matrices propagates by the same projector,  $\mathbf{U}_y - \mathbf{U}_f = (\mathbf{I} - \mathbf{P})\mathbf{U}_y$ . The projector  $(\mathbf{I} - \mathbf{P})$  will be termed *residual projector* because it maps the input vector  $\mathbf{y}$  to the residual vector  $(\mathbf{y} - \mathbf{f})$ .

*Note:* Using equations (A.10), (14), the goodness of fit, equation (8), takes the form

$$\rho(\mathbf{a}) = \mathbf{y}^T \mathbf{W} (\mathbf{I} - \mathbf{P}) \mathbf{y} = \mathbf{y}^T \mathbf{W} (\mathbf{U}_y - \mathbf{U}_f) \mathbf{W} \mathbf{y}. \quad (30)$$

This equation directly relates the covariance matrices to the residual scatter.

#### 4. Stochastic Ensemble Equations

In this section, an analytical formula for the GLS weight matrix is derived in terms of random and systematic errors of the measurements. The metrological concepts of uncertainty and covariance have probabilistic backgrounds. Particular measured and calculated quantities may be understood as randomly selected samples drawn from a virtual large set of alternative values that is characterized by corresponding statistical properties (de Courtenay and Grégis 2017). Such sets are considered here, and may be termed *stochastic ensembles* in this context (Feistel 2011).

“It is understood that the result of the measurement is the best estimate of the value of the measurand, and that all components of uncertainty, including those arising from systematic effects ... contribute to the dispersion” (GUM 2008, annex B.2.18 therein). Let a potential measurement vector,  $\mathbf{m}$ , be associated with the original measurement results,  $\mathbf{y}$ , by

$$\mathbf{m} = \mathbf{y} + \delta\mathbf{m}, \quad (31)$$

where  $\delta\mathbf{m}$  is termed the *error* of the potential value with respect to “the best estimate of the value of the measurand”, so that the statistical properties of the error distribution define the related dispersion. “Traditionally, an error is viewed as having two components, namely, a **random** component and a **systematic** component” (GUM 2008, section 3.2.1 therein), where the “systematic error is equal to error minus random error” (GUM 2008, annex B.2.22 therein), that is,

$$\delta\mathbf{m} = \boldsymbol{\mu} + \boldsymbol{\xi}. \quad (32)$$

Here, the *systematic error*,  $\boldsymbol{\mu}$ , is the “mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand” (GUM 2008, annex B.2.22 therein), that is,

$$\boldsymbol{\mu} = \langle \mathbf{m} \rangle - \mathbf{y} = \langle \delta\mathbf{m} \rangle. \quad (33)$$

Here, the brackets  $\langle \dots \rangle$  indicate the statistical average over such „an infinite number of measurements“. The *random error*,  $\boldsymbol{\xi}$ , is the “result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions” (GUM 2008, annex B.2.21 therein), that is,



$$\xi = \mathbf{m} - \langle \mathbf{m} \rangle = \delta \mathbf{m} - \langle \delta \mathbf{m} \rangle. \quad (34)$$

Separately for each vector,  $\mathbf{m}$ , of repeated potential measurements, a regression analysis can be performed with respect to the same regression functions, resulting in a corresponding statistical ensemble of regression coefficients. Modified by the small error,  $\delta \mathbf{m}$ , the associated error of the particular regression coefficients,  $\mathbf{a} + \delta \mathbf{a}$ , results from equation (3),

$$\delta \mathbf{a} = \mathbf{U}_a \mathbf{J}_f \mathbf{W} \delta \mathbf{m}. \quad (35)$$

In turn, by means of the Jacobian (10), the error of  $\mathbf{a}$  propagates linearly (GUM 2008, annex E.3.2 therein) to errors of quantities  $\mathbf{h}$  derived from the empirical equation,

$$\delta \mathbf{h} = \mathbf{J}_h \delta \mathbf{a}. \quad (36)$$

In contrast to the errors considered so far, GUM (2008, annex E.3.7 therein) “does not classify components of uncertainty as either ‘random’ or ‘systematic’”. GUM (2008, section 3.3.3 therein) “groups uncertainty components into two categories based on their method of evaluation, ‘A’ and ‘B’... These categories apply to *uncertainty* and are not substitutes for the words ‘random’ and ‘systematic’”. “Uncertainty of measurement is thus an expression of the fact that, for a given measurand and a given result of measurement of it, there is not one value but an infinite number of values dispersed about the result that are consistent with all of the observations and data and one’s knowledge of the physical world, and that with varying degrees of credibility can be attributed to the measurand” (GUM 2008, annex D.5.2 therein).

An *uncertainty matrix* is a matrix “containing on its diagonal the squares of the standard uncertainties associated with estimates of the components of a ... vector quantity, and in its off-diagonal positions the covariances associated with pairs of estimates. ... An uncertainty matrix is also known as a *covariance matrix* or *variance-covariance matrix*” (GUM 2011, section 3.11 therein). The covariance of random variables  $y$  and  $z$  is defined by (GUM 2008, annex C.3.4 therein)

$$\text{cov}(y, z) = \langle (y - \langle y \rangle)(z - \langle z \rangle) \rangle = \langle yz \rangle - \langle y \rangle \langle z \rangle. \quad (37)$$

Applied to the measurements  $\mathbf{m}$ , the related covariance matrix is

$$\text{cov}(\mathbf{m}, \mathbf{m}) = \langle (\mathbf{m} - \langle \mathbf{m} \rangle)(\mathbf{m} - \langle \mathbf{m} \rangle)^T \rangle = \langle \xi \xi^T \rangle. \quad (38)$$

However, this covariance matrix allows for random errors only and ignores systematic errors. For this reason, deviating from the GUM recommendation (38), we shall consider here the *dispersion matrix*,

$$\mathbf{U}_y = \langle (\mathbf{m} - \mathbf{y})(\mathbf{m} - \mathbf{y})^T \rangle = \langle \delta \mathbf{m} \delta \mathbf{m}^T \rangle = \boldsymbol{\mu} \boldsymbol{\mu}^T + \langle \xi \xi^T \rangle, \quad (39)$$

as a more suitable expression for the combined uncertainties of the measurements  $\mathbf{m}$ . The dispersion matrix is a generalization of the covariance matrix and reduces to the latter when systematic errors are absent. Non-diagonal elements, if any, of the covariance matrix, equation (38), result from correlations between random errors rather than from systematic errors, equation (39).

A manifest reason for the preferred presentation of propagation equations for covariance matrices, equation (38), rather than dispersion matrices, (39), within the GUM lies possibly in the fact that measurements are commonly corrected for known systematic errors and only random errors need to be propagated to properties derived from the original measurements. In the context of multiple GLS regression, however, a-posteriori corrections of published measurement results are excluded, and systematic errors detected during the regression procedure need to be included in suitable mathematical tools for the uncertainty evaluation of the regression result. In this sense, GUM (2008, section 6.3.1 therein) notes that “occasionally, one may find that a known correction for a systematic effect has not been applied to the reported result of a measurement, but instead an attempt is made to take the effect into account by enlarging the ‘uncertainty’ assigned to the result. This should be avoided; only in very special circumstances should corrections for known significant systematic effects not be applied to the result of a measurement. Evaluating the uncertainty of a measurement result should not be confused with assigning a *safety limit to some quantity*.” GLS regression seems to be, as we assume here, such a “special circumstance”.

Taking the average,  $\langle \dots \rangle$ , over a stochastic ensemble of errors given by equation (35), the dispersion matrix of the regression coefficients evaluates to

$$\langle \delta \mathbf{a} \delta \mathbf{a}^T \rangle = \mathbf{U}_a \mathbf{J}_f \mathbf{W} \langle \delta \mathbf{m} \delta \mathbf{m}^T \rangle \mathbf{W} \mathbf{J}_f^T \mathbf{U}_a. \quad (40)$$

This equation (40) is consistent with the GUM covariance propagation equations (2), (7), (4) if the original covariance matrix,  $\mathbf{U}_a$ , is now understood as the dispersion matrix,  $\mathbf{U}_a = \langle \delta \mathbf{a} \delta \mathbf{a}^T \rangle$ , and if the errors scatter as given by the dispersion matrix

$$\langle \delta \mathbf{m} \delta \mathbf{m}^T \rangle = \mathbf{U}_y = \mathbf{W}^{-1}. \quad (41)$$

This relation defines the weight matrix of equation (1) in terms of the stochastic ensemble considered, or vice versa. In turn, as a result of the error propagation equation (36), the ultimately desired dispersion matrices  $\mathbf{U}_h = \langle \delta \mathbf{h} \delta \mathbf{h}^T \rangle$ , of a set of derived quantities,  $\mathbf{h}$ , and  $\mathbf{U}_f = \langle \delta \mathbf{f} \delta \mathbf{f}^T \rangle$ , of the prediction functions,  $\mathbf{f}$ , respectively, are computed from the propagation equations (9) and (12), starting from the dispersion matrix  $\mathbf{U}_a$ .

Note that only the second moments of the error distribution are relevant here, regardless of any other details of the actual shape, as long as the linearization (e33) is justified by a negligible amount of large errors. This linearization approximation permits general analytical solutions for infinite statistical ensembles to be derived, avoiding the need to obtain particular numerical solutions from finite-size Monte-Carlo sampling on probability distributions that may not be known, as is often required for other uncertainty analyses (Deletic *et al* 2012).

A simple stochastic ensemble is that of uncorrelated random noise,

$$\langle \delta m_i \rangle = 0, \quad \langle \delta m_i \delta m_j \rangle = \sigma_i^2 \delta_{ij}. \quad (42)$$

Here,  $\delta_{ij}$  is the Kronecker symbol that equals 1 if  $i = j$  and 0 otherwise. From equation (41) and (42), the weight matrix is found to be diagonal (Feistel 2011),

$$\langle \delta \mathbf{m} \delta \mathbf{m}^T \rangle = \mathbf{U}_y = \mathbf{W}^{-1} = \langle \boldsymbol{\xi} \boldsymbol{\xi}^T \rangle = \text{diag}(\boldsymbol{\sigma}^2). \quad (43)$$

In this special case, the GLS problem (1) reduces to weighted least squares (WLS). The variance of this ensemble may be identified with the measurement uncertainty,  $\sigma_i^2 = \omega_i^2$ .

The matrix  $\mathbf{W}$  of equation (43) can be calculated analytically,

$$\mathbf{U}_y^{-1} = \mathbf{W} = \text{diag}(\boldsymbol{\sigma}^{-2}), \quad (44)$$

and the associated dispersion matrix of the regression coefficients, equation (7), then equals the covariance matrix

$$\mathbf{U}_a = \{ \mathbf{J}_f \text{diag}(\boldsymbol{\sigma}^{-2}) \mathbf{J}_f^T \}^{-1}. \quad (45)$$

Extending this model, in order to allow systematic errors to be included in the stochastic ensemble, the definition (42) may be generalized to the model of equation (39),

$$\langle \delta m_i \rangle = \mu_i, \quad \langle \delta m_i \delta m_j \rangle = \mu_i \mu_j + \sigma_i^2 \delta_{ij}, \quad (46)$$

where each potential error value has some systematic mean offset,  $\mu_i$ , that is unaffected by taking the ensemble average, plus a random error. While the average over the random error is zero, see equation (34), the mean value  $\langle \delta m_i \rangle = \mu_i$  results from the non-random error contribution. The related combined uncertainties are given by the diagonal elements of  $\mathbf{U}_y$ ,

$$\omega_i^2 = \langle \delta m_i \delta m_i \rangle = \mu_i^2 + \sigma_i^2. \quad (47)$$

Note that neglecting the random-error part of equation (46) would result in a singular matrix  $\mathbf{U}_y$  for which the inverse  $\mathbf{W}$  does not exist. A model similar to equation (46) is discussed by Davidson and MacKinnon (1993, p324 therein, and 2004, equation 7.88 therein).

Related to equation (46), the GLS weight matrix follows from equation (41),

$$\mathbf{W}^{-1} = \boldsymbol{\mu} \boldsymbol{\mu}^T + \text{diag}(\boldsymbol{\sigma}^2), \quad (48)$$

where the systematic errors give rise to non-vanishing off-diagonal elements. The inverse matrix of equation (48),  $\mathbf{U}_y^{-1} = \mathbf{W}$ , can be calculated analytically by the Sherman-Morrison-Woodbury formula (Seber and Lee 2003, equation A.9.4 therein). It has the elements

$$W_{ij} = \frac{\delta_{ij}}{\sigma_i^2} - \frac{\mu_i \mu_j}{\sigma_i^2 \sigma_j^2} \frac{1}{1 + \sum_{k=1}^M (\mu_k / \sigma_k)^2}. \quad (49)$$

This formula avoids the need for numerical inversion of the typically very large ( $M \times M$ ) matrix (48). It permits the explicit formulation of the GLS problem, equation (1), and the analytical computation of the related normal matrix,  $\mathbf{A} = \mathbf{J}_f \mathbf{W} \mathbf{J}_f^T$ , for the systematic-error model (46).

Using the matrix (49), it is instructive to express the least-squares condition, equation (1), in the form

$$\rho(\mathbf{a}) = (\mathbf{r}^T \mathbf{r}) \frac{1 + (\mathbf{s}^T \mathbf{s}) \sin^2 \alpha}{1 + (\mathbf{s}^T \mathbf{s})} = \text{Min}. \quad (50)$$

Here,  $\alpha(\mathbf{a})$  is the angle<sup>4</sup> between the weighted residual vector,

$$\mathbf{r}(\mathbf{a}) \equiv \text{diag}(\boldsymbol{\sigma}^{-1})(\mathbf{y} - \mathbf{f}) = \left\{ \frac{(y_i - f_i)}{\sigma_i} \right\}, \quad (51)$$

and the vector of weighted systematic errors,

$$\mathbf{s} \equiv \text{diag}(\boldsymbol{\sigma}^{-1})\boldsymbol{\mu} = \left\{ \frac{\mu_i}{\sigma_i} \right\}. \quad (52)$$

The structure of equation (50) reveals that the GLS fit of the model (48) results in a compromise between minimizing the WLS sum of squared residuals,  $|\mathbf{r}|^2$ , that is, adjusting the simulation,  $\mathbf{f}$ , close to the measurement,  $\mathbf{y}$ , and minimizing the angle  $\alpha$ , that is, adjusting the residuals  $\mathbf{r}$  close to the errors  $\mathbf{s}$ , or, in other words, the simulation,  $\mathbf{f}$ , close to the “corrected” measurement,  $(\mathbf{y} - \boldsymbol{\mu})$ .

The dispersion matrix of the regression coefficients, which is suggested to be reported in conjunction with any empirical equation and its coefficients, is given by equations (7), (4),

$$\mathbf{U}_a = (\mathbf{J}_f \mathbf{W} \mathbf{J}_f^T)^{-1}. \quad (53)$$

Related to the special weight matrix  $\mathbf{W}$  given by equation (49), the requisite ( $N \times N$ ) inverted matrix  $\mathbf{U}_a$  is not available in analytical form. As this matrix contains only numerical constants associated with the given equation, the numerical inversion is required only once.

There may be various forms of, or reasons for systematic errors. As well, more complicated statistical properties than those given by equation (46) may be implemented by suitable stochastic ensembles. A key question is how such a matrix may be estimated for practical regression analysis, at least for a model like that of equation (46).

## 5. Uncertainty of Simulation

Making use of the stochastic ensemble, equation (48), the simulation dispersion matrix is

$$\mathbf{U}_f = \mathbf{P} \mathbf{U}_y = \mathbf{P} \langle \delta \mathbf{m} \delta \mathbf{m}^T \rangle = \mathbf{P} \boldsymbol{\mu} \boldsymbol{\mu}^T + \mathbf{P} \text{diag}(\boldsymbol{\sigma}^2). \quad (54)$$

The uncertainty of the reproduced values consists of two additive contributions, one resulting from the random error, and one from the systematic error. Note that by its definition, equation (14), the projection matrix  $\mathbf{P}$  itself is not independent of the stochastic ensemble chosen, namely

$$\mathbf{P} = \mathbf{J}^T \left( \mathbf{J} \langle \delta \mathbf{m} \delta \mathbf{m}^T \rangle^{-1} \mathbf{J}^T \right)^{-1} \mathbf{J} \langle \delta \mathbf{m} \delta \mathbf{m}^T \rangle^{-1}. \quad (55)$$

Because of equation (25), equation (54) can be reformulated in an explicitly symmetric form,

$$\mathbf{U}_f = \mathbf{P} \mathbf{U}_y \mathbf{P}^T = \mathbf{P} \langle \delta \mathbf{m} \delta \mathbf{m}^T \rangle \mathbf{P}^T = \mathbf{P} \boldsymbol{\mu} \boldsymbol{\mu}^T \mathbf{P}^T + \mathbf{P} \text{diag}(\boldsymbol{\sigma}^2) \mathbf{P}^T. \quad (56)$$

Let  $\mathbf{v}$  be a normalized vector,  $\mathbf{v}^T \mathbf{v} = 1$ . Consider the quadratic form

$$\mathbf{v}^T \mathbf{U}_f \mathbf{v} = \mathbf{v}^T \mathbf{P} \boldsymbol{\mu} \boldsymbol{\mu}^T \mathbf{P}^T \mathbf{v} + \mathbf{v}^T \mathbf{P} \text{diag}(\boldsymbol{\sigma}^2) \mathbf{P}^T \mathbf{v}. \quad (57)$$

The second term is the squared vector norm,

<sup>4</sup> The angle  $\alpha$  between two vectors  $\mathbf{r}$  and  $\mathbf{s}$  is defined by their scalar product,  $\mathbf{r}^T \mathbf{s} = |\mathbf{r}| |\mathbf{s}| \cos \alpha$

$$\mathbf{v}^T \mathbf{P} \text{diag}(\boldsymbol{\sigma}^2) \mathbf{P}^T \mathbf{v} = (\text{diag}(\boldsymbol{\sigma}) \mathbf{P}^T \mathbf{v})^T (\text{diag}(\boldsymbol{\sigma}) \mathbf{P}^T \mathbf{v}) = |\mathbf{v}^T \mathbf{P} \text{diag}(\boldsymbol{\sigma})|^2 \geq 0. \quad (58)$$

In the absence of systematic errors, the main-diagonal elements of the matrix  $\mathbf{A}$ , equation (4), consist of sums of  $M$  quadratic terms and will increase with each additional input data point. Consequently, the elements of the matrices  $\mathbf{U}_a$ , equation (7), and  $\mathbf{U}_f$ , equation (12), are expected to decrease like  $1/M$  with an increasing number of data points,  $M$  (Davidson and MacKinnon 2004, GUM 2008 annex E.3.5 therein). The prediction uncertainties related to random errors may take values that become the smaller the more input data are included because the simulation covariance is proportional to the inverse normal matrix,

$$\mathbf{v}^T \mathbf{U}_f \mathbf{v} = \mathbf{v}^T \mathbf{J}^T \mathbf{A}^{-1} \mathbf{J} \mathbf{v}. \quad (59)$$

The first term of equation (57) is just the square of a scalar,

$$\mathbf{v}^T \mathbf{P} \boldsymbol{\mu} \boldsymbol{\mu}^T \mathbf{P}^T \mathbf{v} = (\mathbf{v}^T \mathbf{P} \boldsymbol{\mu})^2 \geq 0. \quad (60)$$

The quadratic form, equation (57), is bounded from below by the systematic errors,

$$\mathbf{v}^T \mathbf{U}_f \mathbf{v} \geq (\mathbf{v}^T \mathbf{P} \boldsymbol{\mu})^2. \quad (61)$$

If, in particular,  $\mathbf{v}$  is chosen to have only a single non-zero component, the quadratic form (57) is reduced to the squared uncertainty of a predicted (reproduced) value given by the related main-diagonal element of  $\mathbf{U}_f$ . So, the uncertainties of simulated values cannot be less than the limit given by the systematic errors in the form of the inequality (61).

The inequality (61) is the main result of this section. Its meaning may be discussed in slightly more detail here. As described in Section 3, the matrix  $\mathbf{P}$  is a projector that maps any  $M$ -dimensional measurement vector to an  $N$ -dimensional subspace spanned by the  $N$  eigenvectors of  $\mathbf{P}$  that form the sensitivity matrix  $\mathbf{J}$  of the  $M$  predicted values,  $f$ , with respect to the  $N$  regression coefficients,  $\mathbf{a}$ . The product  $\mathbf{P}\boldsymbol{\mu}$  is the projection of all  $M$  systematic errors  $\boldsymbol{\mu}$  of the measurements to that subspace, expressing the partial effect of the systematic errors that is transferred by the  $N$  regression coefficients to the  $M$  predicted values.

Let

$$\boldsymbol{\mu} = \sum_{i=1}^N c_i \mathbf{J}_i + \sum_{i=N+1}^M c_i \mathbf{J}'_i \quad (62)$$

be the representation of the systematic-error vector  $\boldsymbol{\mu}$  in terms of eigenvectors of  $\mathbf{P}$ . Here,  $\mathbf{J}_i$ ,  $i = 1, \dots, N$ , are the eigenvectors belonging to the eigenvalue 1, see equation (19), given by the sensitivity matrix  $\mathbf{J}$ , and  $\mathbf{J}'_i$ ,  $i = N+1, \dots, M$ , are the eigenvectors of  $\mathbf{P}$  with respect to the eigenvalue 0, see equation (18), which are eigenvectors to the eigenvalue 1 of the conjugate projector,  $(\mathbf{I} - \mathbf{P})$ . The  $c_i$  are the corresponding expansion coefficients. Projection by  $\mathbf{P}$  reduces the sum (62) to its first  $N$  terms, thus rendering irrelevant the other eigenvectors,  $\mathbf{J}'_i$ ,

$$\mathbf{P}\boldsymbol{\mu} = \sum_{i=1}^N c_i \mathbf{J}_i \quad (63)$$

This way, a certain part of the systematic error is suppressed by the smoothing effect of the regression procedure, mathematically represented by the simulation projector  $\mathbf{P}$ , and only the part affecting the regression coefficients is carried on. Introducing a vector  $\mathbf{c}$  of those  $N$  expansion coefficients, the sum in equation (63) may be written in matrix notation,

$$\mathbf{P}\boldsymbol{\mu} = \mathbf{J}^T \mathbf{c}. \quad (64)$$

Making use of the orthogonality, equation (23), and of equation (21), the vector  $\mathbf{c}$  in equation (64) can explicitly be calculated from  $\boldsymbol{\mu}$  by the relation,

$$\mathbf{c} = \mathbf{K} \mathbf{J}^T \boldsymbol{\mu} = \mathbf{K} \mathbf{P} \boldsymbol{\mu} = \mathbf{K} \boldsymbol{\mu}. \quad (65)$$

The matrix  $\mathbf{K}$  is available from the computation of the regression coefficients in terms of the input measurements, equation (26),  $\mathbf{a} = \mathbf{K} \mathbf{y}$ . Because  $\boldsymbol{\mu}$  is the systematic error of the measurement vector  $\mathbf{y}$ , the analogy to equation (65),  $\mathbf{c} = \mathbf{K} \boldsymbol{\mu}$ , indicates that the vector  $\mathbf{c}$  can be understood as the systematic error of the regression coefficients,  $\mathbf{a}$ . Similarly, since  $\mathbf{f} = \mathbf{P} \mathbf{y}$  is the simulation vector,  $\mathbf{P}\boldsymbol{\mu}$  in equations (61) and (64) can be understood as the systematic error of the simulation. The quadratic form, equation (61), can be expressed in terms of  $\mathbf{c}$ ,

$$\mathbf{v}^T \mathbf{U}_f \mathbf{v} \geq (\mathbf{v}^T \mathbf{J}^T \mathbf{c})^2. \quad (66)$$

If we choose, for example, the vector  $\mathbf{v}$  to hold only its first component,  $\mathbf{v}^T = (1, 0, \dots, 0)$ , the inequality (61) provides the lower bound for the prediction uncertainty at the first measurement point, due to systematic errors, by the expression

$$[\mathbf{U}_f]_{1,1} \geq \left( \sum_{i=1}^N c_i \frac{\partial f_1}{\partial a_i} \right)^2. \quad (67)$$

Evidently, this lower bound is combined of the joint effects of systematic errors at all data points via equation (65), not solely at the first measurement itself. Estimating those systematic errors from structures discernible in the scatter of regression residuals is the topic of the next section.

## 6. Estimating Systematic Errors from Regression Residuals

The Type B standard uncertainty “is evaluated by scientific judgement based on all of the available information on the possible variability” (GUM 2008, section 4.3.1 therein) of the input quantity. Given a set of measurement values and their Type A uncertainties, only a diagonal weight matrix, equation (43), is available. After the regression has been carried out by minimizing the related WLS sum, equation (1), by a suitable choice of regressor functions and subsequent determination of the regression coefficients,  $\mathbf{a}$ , each measured value,  $y_i$ , is then associated with a residual,  $\Delta y_i$ , of the fit,

$$\Delta y_i = y_i - f_i(\mathbf{a}, \mathbf{x}_i). \quad (68)$$

This residual may be regarded as the error of the measurement with respect to the fit. This error typically consists of a systematic part,  $s_i$ , and a random part,  $r_i$ ,

$$\Delta y_i = s_i + r_i. \quad (69)$$

For reasonably separating the given residuals  $\Delta y_i$  in two parts, information additional to the original mathematical problem is required.

The full input data set may be split into groups, each belonging to a certain data source (article, experiment, author, etc.). It may be assumed that some kind of systematic error is characteristic for all values that belong to the same group, and that this common kind of deviation is reflected in discernible structures in the fitting residuals of this group, such as an offset or trend relative to the overall fit covering all groups together. In such a case, the residuals of a selected group may again be fitted separately to an auxiliary empirical function,  $s(\mathbf{x})$ , that accounts for their systematic error. This fit leaves over a reduced but random residual that accounts for the random error of the group members. The offset,  $s_i = s(\mathbf{x}_i)$ , may then be identified with the systematic error,  $\mu_i$ , of the stochastic ensemble, equation (46), while the remaining random error,  $\sigma_i$ , may be derived from the combined uncertainty, equation (47), or alternatively from the scatter relative to the auxiliary function. With such estimates available for each measurement, an improved dispersion matrix, equation (49), can be constructed for the regression coefficients.

Structure in the residuals is evidence of commonality of error and hence correlation. This error could, at least in part, be corrected, leaving a random scatter of error characterised by the uncertainty in the correction. However, in many cases the residual error structure cannot be corrected, in which case we need to account for this error structure in the uncertainty budget. For example, a linear trend in the residuals of a particular set of data suggests a common error perhaps scaled by one of the measurement data quantities. The scatter of points around the linear trend represents the residual random error for that part of the fitted data curve, and can be quantified as a standard deviation. This approach is covered to some extent by Lovell-Smith *et al* (2017). It suggests that the systematic error terms in those data points which form a set sharing a trend or auxiliary empirical function, should be considered fully correlated, which is what the simple model (46) assumes.

As a cross-check, the full GLS regression may be repeated with the new weight matrix containing off-diagonal elements for inspecting the residual structure. Also, the uncertainty of the empirical equation derived from GLS may be compared with other uncertainty estimates possibly available from other sources. Iterative adjustments of the systematic-error estimates,  $\mu_i$ , may be carried out in order to achieve sufficient agreement. Because the dispersion matrix provides uncertainty estimates



also for quantities for which no such estimates were available before, these new estimates will this way become consistent with those available previously.

## 7. Error Blindness

It is possible that errors belonging to a class of data points with certain mathematical properties, regardless of how large those errors may be, do not propagate to related errors of the regression coefficients. In turn, those errors will not be reflected by the dispersion matrix of the regression coefficients, and will consequently neither affect the estimated uncertainties of the regression function nor of any properties derived thereof. This insensitivity will here be termed *error blindness* of the regression coefficients with respect to that error class. A simple example is the calculation of the arithmetic mean value with respect to systematic offsets of the data groups involved, see Appendix B.

The propagation of errors of the measurements to those of the regression coefficients is governed by equation (35). A regression function is termed *error-blind* with respect to a certain error  $\delta \mathbf{m} \neq 0$  if

$$\delta \mathbf{a} = \mathbf{K} \delta \mathbf{m} = 0. \quad (70)$$

Here,  $\mathbf{K}$  is the  $(N \times M)$  matrix defined by equation (20).

Error-blind regression functions may lead to significantly underestimated uncertainties of quantities computed thereof. If the dispersion matrix of the regression matrix is unable to propagate significant errors of the input data to the output, the reporting of that matrix along with the regression coefficients will represent an insufficient tool for the description of deviations of input data from predicted properties of the corresponding equation.

In a sense, error-blindness is an inherent feature of least-squares regression. It is not difficult to imagine that various different point clouds may be represented by exactly the same regression function. Least-squares fitting is even designed as a method for suppressing noise in data, rendering details of the noise irrelevant for the result. The  $M - N$  dimensional vector space of  $M$  differential errors  $d\mathbf{m}$  that belong to the same regression coefficients  $\mathbf{a}$  obeys a homogeneous system of  $N$  linear equations (70) in the form

$$\mathbf{K} d\mathbf{m} = 0. \quad (71)$$

Thus, in the sense that there are much more error components,  $M$ , in equation (71) than there are restricting conditions,  $N$ , one may even argue that the regression coefficients are “blind to most of the errors”, in particular also to any multiples of the residual vector, equation (28). Systematic errors,  $\boldsymbol{\mu}$ , that obey equation (71),

$$\mathbf{K} \boldsymbol{\mu} = 0, \quad (72)$$

do not raise the final uncertainty, equation (66), despite arbitrarily large residuals.

In the context of this paper, the most relevant question is what kind of systematic errors contained in the GLS weight matrix (48) may not propagate into the dispersion matrix (53). The governing uncertainty propagation equation is

$$\mathbf{U}_a^{-1} = \mathbf{J}_f \mathbf{U}_y^{-1} \mathbf{J}_f^T. \quad (73)$$

Writing the weight matrix (49) in compact form,

$$\mathbf{U}_y^{-1} = \text{diag}(\boldsymbol{\sigma}^{-2}) - \frac{[\text{diag}(\boldsymbol{\sigma}^{-2}) \boldsymbol{\mu}][\text{diag}(\boldsymbol{\sigma}^{-2}) \boldsymbol{\mu}]^T}{1 + \boldsymbol{\mu}^T \text{diag}(\boldsymbol{\sigma}^{-2}) \boldsymbol{\mu}}, \quad (74)$$

its multiplication (73) with the Jacobians  $\mathbf{J}$  gives

$$\mathbf{U}_a^{-1} = \mathbf{J}_f \text{diag}(\boldsymbol{\sigma}^{-2}) \mathbf{J}_f^T - \frac{[\mathbf{J}_f \text{diag}(\boldsymbol{\sigma}^{-2}) \boldsymbol{\mu}][\mathbf{J}_f \text{diag}(\boldsymbol{\sigma}^{-2}) \boldsymbol{\mu}]^T}{1 + \boldsymbol{\mu}^T \text{diag}(\boldsymbol{\sigma}^{-2}) \boldsymbol{\mu}}. \quad (75)$$

Evidently, systematic errors  $\boldsymbol{\mu}$  do not propagate to the dispersion matrix  $\mathbf{U}_a$  of the regression coefficients if they fall under the condition

$$\mathbf{J}_f \text{diag}(\boldsymbol{\sigma}^{-2}) \boldsymbol{\mu} = 0. \quad (76)$$

Uncertainty estimates derived from the dispersion matrix for the regression function will be “blind” to any systematic errors with this property, even if those errors may be substantial and significant.

Calculating from equation (74) the product

$$J_f W \mu = J_f U_y^{-1} \mu = J_f \text{diag}(\sigma^{-2}) \mu - \frac{[J_f \text{diag}(\sigma^{-2}) \mu][\text{diag}(\sigma^{-2}) \mu]^T}{1 + \mu^T \text{diag}(\sigma^{-2}) \mu} \mu \quad (77)$$

demonstrates that the condition (76) implies error blindness, equation (72), for  $K$  defined by equation (20) in the special case of simulation,  $J = J_f$ . Evidently, systematic errors that do not propagate to the dispersion matrix do neither propagate, too, to the regression coefficients.

The decomposition of the regression residuals into systematic and random errors,

$$y - \hat{f} = r + s, \quad (78)$$

may be done as described by equation (69) or in any other suitable way. Independently of that particular separation method applied, by considering the property (28), the multiplication of (78) with the matrix  $K$  always results in

$$Ks = -Kr. \quad (79)$$

However, if the systematic error  $\mu$  is identified with its estimate  $s$ ,  $Ks = K\mu$  is the systematic contribution to the uncertainty, equation (65). It follows that the effect of error blindness does not fully suppress the non-statistical effect of systematic error, but reduces its magnitude to that of the residual random error.

## 8. Summary

When reference-quality empirical equations of state are constructed from large, heterogeneous sets of data for various quantities that had been published by various authors in the course of decades of years, such as the fundamental geophysical equations for air (Lemmon *et al* 2000), fluid water (Wagner and Pruß 2002), ice (Feistel and Wagner 2005, 2006) or seawater (Feistel 2003, 2008, 2010), the extended requisite set of regression functions involved is typically selected by numerous laborious trial-and-error iterations. A subjectively “best” version is obtained as a suitable compromise between a minimum value of the goodness of fit, equation (8), a minimum number of regression coefficients, minimum mathematical complexity of the optional regression functions, a smooth and reasonably extrapolating overall regression function without spurious oscillations of its first three or four partial derivatives, and well-balanced scatter of uncertainty-weighted residuals in the various regions of the state space and of the various derived quantities. One may hardly imagine that this demanding, complicated process may be satisfactorily accomplished by any simple, fully automatic algorithm.

As a result of such a “best fit”, certain data groups are represented very well, scattering randomly about the regression curve. Other data groups may deviate more systematically with clear trends or offsets. Both situations should properly enter the uncertainty estimates associated with the final regression function. However, only in rare cases such empirical equations of state were published together with a covariance or dispersion matrix of the regression coefficients, such as suggested by the GUM (Feistel *et al* 2016, Lovell-Smith *et al* 2017). Rather, uncertainty estimates are reported for only a few selected quantities, in particular for those where measurements actually entered the fit. Those estimates are often obtained from unsystematic and subjective, expert-based assessments. One of the main reasons for this unsatisfactory situation is the fact that, starting from available estimates for systematic and random uncertainties of the input data, there is no formal procedure available for the construction of a reasonable dispersion matrix of the regression coefficients from those input uncertainties. The aim of this paper is to reduce this profound deficiency.

The mathematical formalism developed in this article allows including systematic errors in the weight matrix of a fit, and accordingly in the dispersion matrix of the regression coefficients. The regression procedure suggested here consists of the following steps:

- (i) Accepting uncertainties,  $\omega$ , such as those reported by authors along with their measurements,  $y$ , a WLS fit with a diagonal weight matrix, equation (44),

$$U_y^{-1} = W' = \text{diag}(\omega^{-2}). \quad (80)$$

is performed. It is common practice that this step may be repeated for optimizing the set of regressor functions and their adjustable coefficients. Finally, from the structure of the residuals, systematic errors,  $\mu$ , and the range of random scatter,  $\sigma$ , are estimated, as described in section 6. The regressor function values  $f'$  simulated by the fit, and their scatter  $(y - f')$ , respectively, are obtained from the projectors  $P'$  and  $(I - P')$ , equations (27), (29), applied to the vector of measurements,  $y$ . Here, the prime ' added at the variables is used to indicate that these least-squares quantities are tentative ones, belonging to the preparatory first step of the method.

- (ii) With the same data points and regressor functions as before, the fit is repeated using a GLS weight matrix, equations (48), (49),

$$W^{-1} = \mu \mu^T + \text{diag}(\sigma^2). \quad (81)$$

The result of this fit is considered the final empirical equation. Consistent with its regression coefficients, their dispersion matrix, equation (53),

$$U_a = (J_f W J_f^T)^{-1} \quad (82)$$

is calculated for being reported along with the regression coefficients.

- (iii) External subsequent users of the published equation may estimate uncertainties (and possibly also dispersion matrices associated with sets of output data) of values  $h$  calculated from the equation (within its range of validity) by equation (9),

$$U_h = J_h^T U_a J_h. \quad (83)$$

The main diagonal elements of the dispersion matrix  $U_h$  represent the squared uncertainties of the predicted values,  $h$ .

Uncertainties of empirical equation estimated this way, including random and systematic errors of individual data sets, represent the reliability and robustness of the given equation with respect to variation of the input data within their uncertainties. The output data uncertainty, equation (83), should not be confused with, as the GUM is calling it, a *safety limit to some quantity*. As well, it should not be confused with the likely range of scatter of possible future measurements under reproducibility, or similar, conditions.

**Acknowledgement:** The authors are grateful to the referee for critical comments and helpful suggestions. This paper contributes to the tasks of the Joint SCOR/IAPWS/IAPSO Committee on the Properties of Seawater (JCS). Organisations financially supporting this work include the Scientific Committee on Oceanic Research (SCOR), with funding from national SCOR committees, and the International Association for the Properties of Water and Steam (IAPWS), with funding from IAPWS National Committees. Part of this work was funded by the New Zealand Government as part of a contract for the provision of national measurement standards.

## Appendix A: Additional Properties Related to the Simulation Projector $P$

The  $(M \times M)$  simulation projector matrix  $P$  is defined by equation (14),

$$P = U_f U_f^{-1} = U_f W = J^T A^{-1} J W = J^T (J W J^T)^{-1} J W, \quad (A.1)$$

A quadratic matrix is regular only if its rank equals its order. The least-squares problem, equation (1), has a unique solution only if the  $(N \times N)$  normal matrix  $A$  is regular, which implies that we consider only cases for which

$$\text{rank}(A) = N. \quad (A.2)$$

According to the Binet-Cauchy theorem, the rank of a product of matrices cannot exceed the rank of any of its factors (Gantmacher 1959, §I.3 therein). Therefore, if  $A = J W J^T$  is regular, also

$$\text{rank}(J) = N \quad (A.3)$$

holds, because the rank of the  $(N \times M)$  matrix  $J$  cannot exceed  $N$  if  $N < M$ . The  $(M \times M)$  input covariance matrix,  $U_f = W^{-1}$ , is assumed to exist, so that

$$\text{rank}(W) = M. \quad (A.4)$$

It follows from equation (12) that the rank of the simulation covariance matrix is,

$$\text{rank}(U_f) = N < M, \quad (A.5)$$

so that  $\mathbf{U}_f$  is singular.

Several mathematical properties related to the simulation projector matrix  $\mathbf{P}$  are summarized below, in addition to those given in Section 3.

- (i)  $\mathbf{P}$  has  $N$  eigenvalues equal to 1 and  $M - N$  eigenvalues equal to 0. (A.6)

*Proof:* Any idempotent matrix has only 0 or 1 as its eigenvalues. The rank of an idempotent matrix equals the number of its non-zero eigenvalues. Because of (15), the eigenvalue 1 of  $\mathbf{P}$  has the multiplicity  $N$ , the eigenvalue 0 has the multiplicity  $M - N$ .

*Note:* The additivity of rank of  $\mathbf{P}$  and  $(\mathbf{I} - \mathbf{P})$  is subject to Cochran's theorem (Tian and Styan 2001).

- (ii)  $\mathbf{K}$  is of rank  $N$ . (A.7)

*Proof:* Because  $\mathbf{I}$  in (23) is of rank  $N$ , the rank of the factor  $\mathbf{K}$  cannot be less than  $N$ . Because  $\mathbf{K}$  is of the order  $(N \times M)$  with  $N < M$ , the rank of  $\mathbf{K}$  cannot be greater than  $N$ . So,

$$\text{rank}(\mathbf{K}) = N. \quad (\text{A.8})$$

- (iii) The biorthogonal Hilbert-Schmidt expansion (or *spectral decomposition*) of  $\mathbf{P}$  into pairwise conjugate eigenvectors takes the form  $\mathbf{P} = \mathbf{J}^T \mathbf{K}$ . (A.9)

*Proof:* By the definition of  $\mathbf{J}$  and  $\mathbf{K}$ ,

$$\mathbf{J}^T \mathbf{K} = \mathbf{J}^T (\mathbf{J} \mathbf{W} \mathbf{J}^T)^{-1} \mathbf{J} \mathbf{W} = \mathbf{P}. \quad (\text{A.10})$$

- (iv) The rows of the  $(N \times M)$  matrix  $\mathbf{k}$ ,

$$\mathbf{k} = \mathbf{J} \mathbf{W}, \quad (\text{A.11})$$

are left eigenvectors of  $\mathbf{P}$  to the eigenvalue 1.

*Proof:* By the definition (14) of  $\mathbf{P}$ ,

$$\mathbf{k} \mathbf{P} = \mathbf{J} \mathbf{W} \mathbf{J}^T (\mathbf{J} \mathbf{W} \mathbf{J}^T)^{-1} \mathbf{J} \mathbf{W} = \mathbf{J} \mathbf{W} = \mathbf{k}. \quad (\text{A.12})$$

*Note:*  $\mathbf{k}$  is biorthogonal with  $\mathbf{J}^T$  if and only if the normal matrix  $\mathbf{A}$  is diagonal. Evidently,  $\mathbf{k} \mathbf{J}^T = \mathbf{J} \mathbf{W} \mathbf{J}^T = \mathbf{A}$ . (A.13)

- (v) The columns of the matrix  $\mathbf{j}^T = \mathbf{J}^T (\mathbf{J} \mathbf{W} \mathbf{J}^T)^{-1}$  are right eigenvectors of  $\mathbf{P}$  to the eigenvalue 1. (A.14)

*Proof:* By the definition (14) of  $\mathbf{P}$ ,

$$\mathbf{P} \mathbf{j}^T = \mathbf{J}^T (\mathbf{J} \mathbf{W} \mathbf{J}^T)^{-1} \mathbf{J} \mathbf{W} \mathbf{J}^T (\mathbf{J} \mathbf{W} \mathbf{J}^T)^{-1} = \mathbf{J}^T (\mathbf{J} \mathbf{W} \mathbf{J}^T)^{-1} = \mathbf{j}^T. \quad (\text{A.15})$$

*Note:*  $\mathbf{j}^T$  is biorthogonal with  $\mathbf{K}$  if and only if the normal matrix  $\mathbf{A}$  is diagonal. Evidently,  $\mathbf{j} \mathbf{K}^T = (\mathbf{J} \mathbf{W} \mathbf{J}^T)^{-1} \mathbf{J} \mathbf{W} \mathbf{J}^T (\mathbf{J} \mathbf{W} \mathbf{J}^T)^{-1} = \mathbf{A}^{-1} = (\mathbf{K} \mathbf{J}^T)^{-1}$ . (A.16)

- (vi) The eigenvectors given by  $\mathbf{j}$  and  $\mathbf{k}$  are pairwise mutually orthonormal,  $\mathbf{k} \mathbf{j}^T = \mathbf{j} \mathbf{k}^T = \mathbf{I}$ . (A.17)

*Proof:* By the definition of  $\mathbf{j}$  and  $\mathbf{k}$ ,

$$\mathbf{k} \mathbf{j}^T = \mathbf{J} \mathbf{W} \mathbf{J}^T (\mathbf{J} \mathbf{W} \mathbf{J}^T)^{-1} = \mathbf{I}, \quad (\text{A.18})$$

And

$$\mathbf{j} \mathbf{k}^T = (\mathbf{J} \mathbf{W} \mathbf{J}^T)^{-1} \mathbf{J} \mathbf{W} \mathbf{J}^T = \mathbf{I}. \quad (\text{A.19})$$

- (vii) The column vectors of  $\mathbf{U}_f = \mathbf{P} \mathbf{U}_y$  are right eigenvectors of  $\mathbf{P}$  to the eigenvalue 1. (A.20)

*Note:* Each of the  $M$  columns of  $\mathbf{U}_f$  is a linear combination of the  $N$  columns of  $\mathbf{J}^T$ . (A.21)

*Proof:* Because  $\mathbf{P}$  is idempotent,  $\mathbf{P}^2 = \mathbf{P}$ , we find  $\mathbf{P} \mathbf{U}_f = \mathbf{P}^2 \mathbf{U}_y = \mathbf{P} \mathbf{U}_y = \mathbf{U}_f$ . (A.22)

- (viii) The row vectors of  $\mathbf{U}_f = \mathbf{P} \mathbf{U}_y$  are left eigenvectors of  $\mathbf{P}^T$  to the eigenvalue 1. (A.23)

*Proof:* The transpose of equation (A.22) reads  $\mathbf{U}_f \mathbf{P}^T = \mathbf{U}_f$ . (A.24)

- (ix) The mean quadratic uncertainty of the simulated values,  $\text{tr}(\mathbf{U}_f)/M$ , is restricted by the inequality, 
$$\frac{1}{\text{tr}(\mathbf{U}_y^{-1})} \leq \frac{\text{tr}(\mathbf{U}_f)}{N} \leq \text{tr}(\mathbf{U}_y). \quad (\text{A.25})$$

Here,  $\text{tr}(\mathbf{Q})$  is the trace of a quadratic matrix  $\mathbf{Q}$  which is defined by the sum of its main diagonal elements.

*Proof:* The trace of an idempotent matrix equals its rank (Penrose 1955, Baksalary *et al* 2010),  
 $\text{tr}(\mathbf{P}) = \text{rank}(\mathbf{P}) = N.$  (A.26)

$\mathbf{P}$  is positive semidefinite.  $\mathbf{U}_y$  is positive definite. The trace of the product of positive semidefinite matrices of equal order is sub-multiplicative (Yang 2000, with reference to Coope 1994). From  $\mathbf{U}_f = \mathbf{P}\mathbf{U}_y$ , we get

$$\text{tr}(\mathbf{U}_f) = \text{tr}(\mathbf{P}\mathbf{U}_y) \leq \text{tr}(\mathbf{P})\text{tr}(\mathbf{U}_y) = N \text{tr}(\mathbf{U}_y), \quad (\text{A.27})$$

and from  $\mathbf{U}_f \mathbf{U}_y^{-1} = \mathbf{P}$ , we get

$$\text{tr}(\mathbf{P}) = N = \text{tr}(\mathbf{U}_f \mathbf{U}_y^{-1}) \leq \text{tr}(\mathbf{U}_f) \text{tr}(\mathbf{U}_y^{-1}). \quad (\text{A.28})$$

Combination of the two inequalities (A.27), (A.28) results in the interval (A.25).

*Example:* In the special case of weighted least squares with uniform measurement uncertainty,  $\mathbf{U}_y = \omega^2 \mathbf{I}$ , the interval (A.25) takes the form

$$\frac{1}{M} \leq \frac{\text{tr}(\mathbf{U}_f)}{N\omega^2} \leq M. \quad (\text{A.29})$$

For large  $M$ , the restriction expressed by this interval is only weak.

(x) The matrix norm of  $\mathbf{U}_f$  is restricted by the inequality

$$\frac{1}{\|\mathbf{U}_y^{-1}\|} \leq \frac{\|\mathbf{U}_f\|}{\|\mathbf{P}\|} \leq \|\mathbf{U}_y\|, \text{ where } \|\mathbf{P}\| \geq 1. \quad (\text{A.30})$$

*Proof:* There exist various matrix norms  $\|\cdot\|$ , each of them is sub-multiplicative,

$$\|\mathbf{QR}\| \leq \|\mathbf{Q}\| \|\mathbf{R}\|, \quad (\text{A.31})$$

for any matrices  $\mathbf{Q}, \mathbf{R}$ , and each of the norms obeys

$$\|\mathbf{P}\| \geq 1, \quad (\text{A.32})$$

for any idempotent matrix  $\mathbf{P}$  (Horn and Johnson 1984, §5.6 therein).

Application of (A.31) to  $\mathbf{U}_f = \mathbf{P}\mathbf{U}_y$  gives  $\|\mathbf{U}_f\| \leq \|\mathbf{P}\| \|\mathbf{U}_y\|$ , and to  $\mathbf{U}_f \mathbf{U}_y^{-1} = \mathbf{P}$  gives  $\|\mathbf{P}\| \leq \|\mathbf{U}_f\| \|\mathbf{U}_y^{-1}\|$ . In combination, the two inequalities result in the interval (A.30).

(xi) The simulation covariance,  $\mathbf{U}_f$ , does not “exceed” the measurement covariance,  $\mathbf{U}_y$ , in the sense that  $\mathbf{y}^T \mathbf{W} \mathbf{U}_f \mathbf{W} \mathbf{y} \leq \mathbf{y}^T \mathbf{W} \mathbf{U}_y \mathbf{W} \mathbf{y}$ . (A.33)

*Proof:* This inequality follows directly from equation (30) and  $p \geq 0$ , by definition.

(xii) The residual vector,  $(\mathbf{y} - \mathbf{f})$ , is a right eigenvector of  $\mathbf{P}$  to the eigenvalue 0,

$$\mathbf{P}(\mathbf{y} - \mathbf{f}) = 0. \quad (\text{A.34})$$

*Proof:* Using equation (27),  $\mathbf{P}(\mathbf{y} - \mathbf{f}) = \mathbf{P}\mathbf{y} - \mathbf{P}\mathbf{f} = \mathbf{P}\mathbf{y} - \mathbf{P}^2\mathbf{y} = 0$ .

(xiii) The residual vector,  $(\mathbf{y} - \mathbf{f})$ , is a right eigenvector of  $(\mathbf{I} - \mathbf{P})$  to the eigenvalue 1,

$$(\mathbf{I} - \mathbf{P})(\mathbf{y} - \mathbf{f}) = (\mathbf{y} - \mathbf{f}). \quad (\text{A.35})$$

*Proof:* Using equation (A.34),  $(\mathbf{I} - \mathbf{P})(\mathbf{y} - \mathbf{f}) = (\mathbf{y} - \mathbf{f}) - \mathbf{P}(\mathbf{y} - \mathbf{f}) = (\mathbf{y} - \mathbf{f})$ .

## Appendix B: Tutorial Example for GLS and a Dispersion Matrix

Let  $M$  repeated measurement results  $\mathbf{y}$  consist of two independent groups, the first group with  $M_1$  data and  $M_2$  data of the second group. The repeatability  $\sigma$  in both groups may be the same. Let the mean of group 1 be

$$c_1 = \frac{1}{M_1} \sum_{i=1}^{M_1} y_i, \quad (\text{B.1})$$

and similarly  $c_2$  of group 2. The WLS weight matrix is, with  $\mathbf{I}$  denoting the unity matrix,

$$\mathbf{W} = \frac{1}{\sigma^2} \mathbf{I}. \quad (\text{B.2})$$

The regression function is simply  $f = a$ , with  $a$  being the regression coefficient. The WLS solution for  $a$  is well known to be the arithmetic mean of all input values,

$$a = \frac{M_1}{M} c_1 + \frac{M_2}{M} c_2. \quad (\text{B.3})$$



If the scatter within each group is much smaller than the offset between the groups,  $\sigma \ll |c_1 - c_2|$ , the structure of the residuals is clearly visible. Fitting constant offsets separately to the residuals of the two groups, the estimated systematic errors are

$$\begin{aligned}\mu_1 &= c_1 - a = \frac{M_2}{M}(c_1 - c_2), \\ \mu_2 &= c_2 - a = \frac{M_1}{M}(c_2 - c_1).\end{aligned}\quad (\text{B.4})$$

Note the fixed “barycenter” of the systematic errors,

$$\mu_1 M_1 + \mu_2 M_2 = 0. \quad (\text{B.5})$$

The related GLS weight matrix is a block matrix for the two groups,

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}, \quad (\text{B.6})$$

with the blocks,

$$\begin{aligned}\mathbf{W}_{11} &= \frac{1}{\sigma^2} \left\{ \mathbf{I} - \frac{\mu_1^2}{\sigma^2 + M_1 \mu_1^2 + M_2 \mu_2^2} \mathbf{1} \right\}, \\ \mathbf{W}_{12} &= -\frac{1}{\sigma^2} \left\{ \frac{\mu_1 \mu_2}{\sigma^2 + M_1 \mu_1^2 + M_2 \mu_2^2} \right\} \mathbf{1},\end{aligned}\quad (\text{B.7})$$

and similarly for  $\mathbf{W}_{21}$  and  $\mathbf{W}_{22}$ . The matrices  $\mathbf{1}$  consist of elements equal 1 each.

The Jacobian  $\mathbf{J}$ , eq (11), is a matrix with a single row consisting of two blocks of 1's,

$$\mathbf{J} = (\mathbf{J}_1, \mathbf{J}_2) = \mathbf{1}. \quad (\text{B.8})$$

This leads to

$$\mathbf{J}_1 \mathbf{W}_{11} \mathbf{J}_1^\top = \frac{\mathbf{J}_1 \mathbf{J}_1^\top}{\sigma^2} - \frac{1}{\sigma^2} \frac{\mu_1^2}{\sigma^2 + M_1 \mu_1^2 + M_2 \mu_2^2} \mathbf{J}_1 \mathbf{1} \mathbf{J}_1^\top = \frac{M_1}{\sigma^2} - \frac{1}{\sigma^2} \frac{\mu_1^2 M_1}{\sigma^2 + M_1 \mu_1^2 + M_2 \mu_2^2}$$

and

$$\mathbf{J}_1 \mathbf{W}_{12} \mathbf{J}_2^\top = -\frac{1}{\sigma^2} \left\{ \frac{\mu_1 \mu_2}{\sigma^2 + M_1 \mu_1^2 + M_2 \mu_2^2} \right\} \mathbf{J}_1 \mathbf{1} \mathbf{J}_2^\top = -\frac{1}{\sigma^2} \frac{\mu_1 \mu_2 M_1 M_2}{\sigma^2 + M_1 \mu_1^2 + M_2 \mu_2^2}, \text{ etc.}$$

Summing up over the 4 blocks,

$$\mathbf{J}_f \mathbf{W} \mathbf{J}_f^\top = \mathbf{J}_1 \mathbf{W}_{11} \mathbf{J}_1^\top + \mathbf{J}_1 \mathbf{W}_{12} \mathbf{J}_2^\top + \mathbf{J}_2 \mathbf{W}_{21} \mathbf{J}_1^\top + \mathbf{J}_2 \mathbf{W}_{22} \mathbf{J}_2^\top$$

we find

$$\mathbf{J}_f \mathbf{W} \mathbf{J}_f^\top = \frac{M}{\sigma^2} - \frac{1}{\sigma^2} \frac{(\mu_1 M_1 + \mu_2 M_2)^2}{\sigma^2 + M_1 \mu_1^2 + M_2 \mu_2^2}. \quad (\text{B.9})$$

The normal matrix, (4), results to have just the one element,

$$\mathbf{A} = \mathbf{J}_f \mathbf{W} \mathbf{J}_f^\top = \frac{M}{\sigma^2}. \quad (\text{B.10})$$

Here, because of (B.5), all contributions from the systematic errors have mutually cancelled in (B.9). Correspondingly, the uncertainty of the regression coefficient is,

$$\mathbf{U}_a = \mathbf{A}^{-1} = \frac{\sigma^2}{M}, \quad (\text{B.11})$$

and is the same as that of the usual mean value, unaffected by the systematic errors. Similarly, also because of (B.5), the GLS result for the regression coefficient, eq (a6),

$$a = \mathbf{A}^{-1} \mathbf{J}_f \mathbf{W} \mathbf{y} = \frac{M_1}{M} c_1 + \frac{M_2}{M} c_2 \quad (\text{B.12})$$

is identical with the WLS result, (B.3).

If the scatter within each group is much smaller than the offset between the groups,  $\sigma \ll |c_1 - c_2|$ , as assume above, **all data points will be located far outside the uncertainty range** (B.11) of the regression result (B.12).

The goodness of the WLS fit is

$$\begin{aligned}
 p(a) &= \frac{1}{\sigma^2} \sum_{i=1}^M (y_i - a)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{M_1} (y_i - c_1 + \mu_1)^2 + \frac{1}{\sigma^2} \sum_{i=M_1+1}^M (y_i - c_2 + \mu_2)^2 \\
 &= \frac{1}{\sigma^2} \sum_{i=1}^{M_1} (y_i - c_1)^2 + \frac{M_1}{\sigma^2} \mu_1^2 + \frac{1}{\sigma^2} \sum_{i=M_1+1}^M (y_i - c_2)^2 + \frac{M_2}{\sigma^2} \mu_2^2
 \end{aligned} \tag{B.13}$$

Assuming the repeatability to match exactly,

$$\sigma^2 = \frac{1}{M_1} \sum_{i=1}^{M_1} (y_i - c_1)^2 = \frac{1}{M_2} \sum_{i=M_1+1}^M (y_i - c_2)^2, \tag{B.14}$$

we get

$$p(a) = M + \frac{M_1}{\sigma^2} \mu_1^2 + \frac{M_2}{\sigma^2} \mu_2^2 = M + \frac{M_1 M_2}{M \sigma^2} (c_1 - c_2)^2 \tag{B.15}$$

This scatter is dominated by the systematic errors, quite in contrast to the uncertainty, (B.11).

## References

1. Aitken A C 1936 IV. — On Least Squares and Linear Combination of Observations *Proc. Roy. Soc. Edinburgh* **55** 42–8, <https://doi.org/10.1017/S0370164600014346>
2. Baksalary O M, Bernstein D S and Trenkler G 2010 On the equality between rank and trace of an idempotent matrix *Appl. Math. Comp.* **217** 4076–80
3. Coope I D 1994 On matrix trace inequalities and related topics for products of Hermitian matrix *J. Math. Anal. Appl.* **188** 999–1001
4. de Courtenay N and Grégis F 2017 The evaluation of measurement uncertainties and its epistemological ramifications *Stud. Hist. Phil. Sci.*, <http://dx.doi.org/10.1016/j.shpsa.2017.05.003>, in press
5. Cox M G and Harris P M 2006 Uncertainty evaluation. Software support for metrology, best practice guide No. 6 *NPL Report DEM-ES-011* National Physical Laboratory, Teddington, UK ([http://publications.npl.co.uk/npl\\_web/pdf/dem\\_es11.pdf](http://publications.npl.co.uk/npl_web/pdf/dem_es11.pdf))
6. Davidson R and MacKinnon J G 1993 *Estimation and Inference in Econometrics* (New York: Oxford University Press)
7. Davidson R and MacKinnon J G 2004 *Econometric Theory and Methods* (New York: Oxford University Press)
8. Deletic A, Dotto C B S, McCarthy D T, Kleidorfer M, Freni G, Mannina G, Uhl M, Henrichs M, Fletcher T D, Rauch W, Bertrand-Krajewski J L and Tait S 2012 Assessing uncertainties in urban drainage models *Phys. Chem. Earth* **42–44** 3–10
9. Draper N R and Smith H 1998 *Applied Regression Analysis* (New York: Wiley Interscience)
10. Feistel R 2003 A new extended Gibbs thermodynamic potential of seawater *Progr. Oceanogr.* **58/1** 43–115
11. Feistel R 2008 A Gibbs function for seawater thermodynamics for -6 °C to 80 °C and salinity up to 120 g/kg *Deep-Sea Res. I* **55** 1639–71
12. Feistel R 2011 Extended equation of state for seawater at elevated temperature and salinity *Desalination* **250** 14–8
13. Feistel R 2011 Stochastic ensembles of thermodynamic potentials *Accred. Qual. Assur.* **16** 225–35
14. Feistel R and Lovell-Smith J W 2017 Defining relative humidity in terms of water activity. Part 1: definition *Metrologia* **54** 566–76
15. Feistel R, Lovell-Smith J and Hellmuth O 2015 Virial Approximation of the TEOS-10 Equation for the Fugacity of Water in Humid Air *Int. J. Thermophys.* **36** 44–68
16. Feistel R, Lovell-Smith J W, Saunders P and Seitz S 2016 Uncertainty of empirical correlation equations *Metrologia* **53** 1079–90
17. Feistel R and Wagner W 2005 High-pressure thermodynamic Gibbs functions of ice and sea ice *J. Marine Res.* **63** 95–139
18. Feistel R and Wagner W 2006 A new equation of state for H<sub>2</sub>O ice Ih *J. Phys. Chem. Ref. Data* **35** 1021–47
19. Gaiser C, Fellmuth B, Haft N, Kuhn A, Thiele-Krivoi B, Zandt T, Fischer J, Jusko O and Sabuga W 2017 Final determination of the Boltzmann constant by dielectric-constant gas thermometry *Metrologia* **54**, 280–9
20. Gantmacher F R 1959 *The Theory of Matrices Vol. 1* (Providence, Rhode Island: AMS Chelsea Publishing)
21. GUM 2008 Evaluation of measurement data - guide to the expression of uncertainty in measurement. JCGM 100:2008 [www.bipm.org/en/publications/guides/gum.html](http://www.bipm.org/en/publications/guides/gum.html)

22. GUM 2011 Guide to the expression of uncertainty in measurement, Supplement 2, JCGM 102:2011 [www.bipm.org/en/publications/guides/gum.html](http://www.bipm.org/en/publications/guides/gum.html)
23. Horn R A and Johnson C R 1984 *Matrix analysis* (Cambridge: Cambridge University Press)
24. IAPWS AN6-16 2016 Advisory Note No. 6: Relationship between Various IAPWS Documents and the International Thermodynamic Equation of Seawater - 2010 (TEOS-10) (Dresden, Germany: The International Association for the Properties of Water and Steam) <http://www.iapws.org>
25. IOC, SCOR and IAPSO 2010 *The international thermodynamic equation of seawater—2010: Calculation and use of thermodynamic properties* Intergovernmental Oceanographic Commission, Manuals and Guides No. 56, UNESCO (English), Paris. <http://www.TEOS-10.org>
26. Lemmon E W, Jacobsen R T, Penoncello S G and Friend D G 2000 Thermodynamic properties of air and mixtures of nitrogen, argon and oxygen from 60 to 2000 K at pressures to 2000 MPa *J. Phys. Chem. Ref. Data* **29** 331–62
27. Lovell-Smith J W, Feistel R, Harvey A H, Hellmuth O, Bell S A, Heinonen M and Cooper J R 2016 Metrological challenges for measurements of key climatological observables. Part 4: atmospheric relative humidity *Metrologia* **53** R40–R59
28. Lovell-Smith J W, Saunders P and Feistel R 2017 Unleashing empirical equations with “Nonlinear Fitting” and “GUM Tree Calculator” *Int. J. Thermophys.*, DOI 10.1007/s10765-017-2282-y, in press
29. Marsaglia G and Styan G P H 1974 Equalities and Inequalities for Ranks of Matrices *Linear Multilinear Algebra* **2** 269–92
30. Penrose R 1955 A generalized inverse for matrices *Math. Proc. Cambridge Phil. Soc.* **51** 406–13, <https://doi.org/10.1017/S0305004100030401>
31. Saunders P 2014 Nonlinear fitting software instructions version 5.0 *Callaghan Innovation Report* 18 (5.0) Measurement Standards Laboratory of New Zealand, Lower Hutt, New Zealand
32. Seber G A F and Lee A J 2003 *Linear Regression Analysis, 2nd edn* (Hoboken, New Jersey: John Wiley)
33. Strutz T 2011 *Data Fitting and Uncertainty* (Wiesbaden: Vieweg + Teubner)
34. Tian Y and Styan G P H 2001 Rank equalities for idempotent and involutory matrices *Linear Algebra Appl.* **335** 101–17
35. Wagner W and Pruß A 2002 The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use *J. Phys. Chem. Ref. Data* **31** 387–535
36. Willink R 2013 *Measurement uncertainty and probability* (Cambridge: Cambridge University Press)
37. Yang X 2000 A Matrix Trace Inequality *J. Math. Anal. Appl.* **250** 372–4

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.