Novel mathematical and statistical approaches to uncertainty evaluation: Software for calibration problems developed and tested

Deliverable 1.3.5

of Work Package

WP 1 (Regression)

Authors:

Andrea Malengo
Francesca Pennecchi
Pier Giorgio Spazzini
INRIM

Peter Harris
Ian Smith
NPL

A report of the EMRP joint research project
NEW04 "Novel mathematical and statistical approaches to uncertainty evaluation"

1. Work package WP 1 (Regression)	**2. Deliverable number** D1.3.5	**3. Reporting date** May 2014
4. Title (and subtitle) Software for calibration problems developed and tested		
5. Author(s) Andrea Malengo (INRIM), Francesca Pennecchi (INRIM), Pier Giorgio Spazzini (INRIM), Peter Harris (NPL), Ian Smith (NPL)		**6. Lead author (e-mail)** f.pennecchi@inrim.it
7. Contributing researchers (institutes) INRIM, NPL		
8. Other contributing work packages None		**9. Lead researchers in other WPs** None
10. Supplementary notes None		

11. Abstract

This document reports the User manual of the "Calibration Curves Computing – CCC software", Release 1.2, which is a software for the evaluation of instrument calibration curves, specifically aimed at flow meter calibration curves. The software was developed by INRIM and tested and validated by NPL.

The software may be applied to pairs (x, y) of measurement values, where x is the independent/explanatory variable and y the dependent/explained variable. If uncertainties associated with the data are available, they can be provided as inputs to the software, otherwise the software is able to estimate them on the basis of repeated measurements.

The regression models which can be addressed by the software are (fractional) polynomial curves. In order to determine estimates of the parameters of the curve and evaluate associated uncertainty and covariance information, the software can perform the following kind of regression procedures, corresponding to statistical models 1a, 1b, 2a, 2b, 3a and 3b, respectively, described in deliverable D1.1.4 of the NEW04 project:

Ordinary least-squares regression (OLS) for models 1a and 1b;

Weighted least-squares regression (WLS) for models 2a and 2b;

Weighted total least-squares regression (WTLS) for models 3a and 3b.

12. Key words

Calibration software, ordinary least-squares regression, weighted least-squares regression, weighted total least-squares regression, flow meter.

Calibration Curves Computing – CCC Software User manual (for Release 1.2)

Abstract

This document is a User manual for the MATLAB-based software “Calibration Curves Computing – CCC Software” (Release 1.2), which is a software for the evaluation of instrument calibration curves, specifically aimed at flow meter calibration curves, and developed at Istituto Nazionale di Ricerca Metrologica (INRIM). The software constitutes the main component of deliverable D1.3.5 “Software for calibration problems developed and tested” of the EMRP project “Novel mathematical and statistical approaches to uncertainty evaluation” (EMRP NEW04).

The software may be applied to pairs (x, y) of measurement values, where x is the independent/explanatory variable and y the dependent/explained variable. If uncertainties associated with the data are available, they can be provided as inputs to the software, otherwise the software is able to estimate them on the basis of repeated measurements.

The regression models which can be addressed by the software are (fractional) polynomial curves. In order to determine estimates of the parameters of the curve and evaluate associated uncertainty and covariance information, the software can perform the following kind of regression procedures, corresponding to statistical models 1a, 1b, 2a, 2b, 3a and 3b, respectively, described in deliverable D1.1.4 of the NEW04 project:

Ordinary least-squares regression (OLS) for models 1a and 1b;

Weighted least-squares regression (WLS) for models 2a and 2b;

Weighted total least-squares regression (WTLS) for models 3a and 3b.

Table of contents

1. BACKGROUND

The "Calibration Curves Computing – CCC Software" constitutes the main component of deliverable D1.3.5 "Software for calibration problems developed and tested" of the EMRP project 'Novel mathematical and statistical approaches to uncertainty evaluation' (EMRP NEW04).
Deliverable D1.3.5 is part of the series of deliverables D1.1.4, D1.1.8, D1.1.12, D1.2.4, D1.2.8, D1.3.3 and D1.3.6, dedicated to the statistical issues related to flow meter measurement data and the determination of relevant calibration curves.
The software relies on the statistical models described in deliverable D1.1.4 [1], in particular on models 1a, 1b, 2a, 2b, 3a and 3b. Although the software was developed for flow meter calibration, it can easily be applied to similar calibration problems in other metrology areas.

2. INTRODUCTION

This User manual describes the "Calibration Curves Computing – CCC Software", software for the evaluation of instrument calibration curves, specifically aimed at flow meter calibration curves. The User provides measured data, uncertainty and covariance information associated with those data, and assigns the mathematical and statistical model from the options available. The software determines estimates of the calibration curve parameters and their associated covariance matrix, as well as estimates of values of the calibration curve and their associated covariance matrix. The software can perform the following kinds of regression:

- Ordinary least-squares regression (OLS);
- Weighted least-squares regression (WLS);
- Weighted total least-squares regression (WTLS).

The software relies on statistical models 1a, 1b, 2a, 2b, 3a and 3b described in deliverable D1.1.4.

Section 3 provides information about the software User licence agreement. Section 4 describes how to use the software: the software graphical user interface is described (Section 4.1), information on the input data file format is given (Section 4.2), guidance on the criteria for choosing the statistical model appropriate for the available measurement data is provided (Section 4.3), the list of possible regression exponents is given (Section 4.4), information on how processing the data (Section 4.5) and saving the results (Section 4.6) is also provided. Appendix A informs on the implementation of the different least-squares methods corresponding to the statistical models described in Section 4.3. Appendix B lists all the possible messages of warnings and errors.

The software is intended to run on a PC running Microsoft Windows. MATLAB [2] is required to be installed before running the software (MATLAB 2009 or later is recommended), together with the Optimization Toolbox™, which is needed for implementation of WTLS models (Release 1.2 was developed in MATLAB R2013a, with Optimization Toolbox™ 6.4, on a PC running Microsoft Windows 7).

3. SOFTWARE USER LICENCE AGREEMENT

Release 1.2 of the Calibration Curves Computing – CCC Software is provided as a collection of open source MATLAB codes intended for use by partners of the EMRP NEW04 project only, and to whom it will be made available through the JRP website (future versions of the software will be provided as an executable and will be made freely available for Users outside the EMRP project).
The software is provided without warranty of any kind. INRIM will try to solve any reported

problems found in running the software, but does not guarantee to fix any problems in the software or to provide any updates or solutions by a particular date.

INRIM is not liable for any damages arising from the use of the software. The entire risk arising out of the use or performance of the Software and documentation remains with the User.

4. USING THE SOFTWARE

4.1. General

After starting MATLAB, use the Current Folder window to access the directory containing the file `CCC_Software.m`. Then double-click on the file to open it and then run it from the Editor window, or type directly "`CCC_Software`" into the Command Window and press Enter.

The program opens a graphical user interface (GUI), shown in fig. (1), which allows the User to perform the following steps:

1. **SELECT INPUT EXCEL FILE**: the User selects a Microsoft Excel workbook. Data is loaded from the workbook and displayed on the GUI.
2. **SELECT STATISTICAL MODEL**: the User selects a statistical model from a list of possible options.
3. **SELECT REGRESSION EXPONENTS**: the User selects the exponents to be used in the regression model from a list of possible exponents.
4. **PROCESS DATA**: the User presses the `Elaboration` button.
5. **SELECT OUTPUT FILE**: the User saves the results to an output data file.
6. **EXIT**: the User quits the program by pressing the Exit button.

Each step is described in detail in the following subsections.

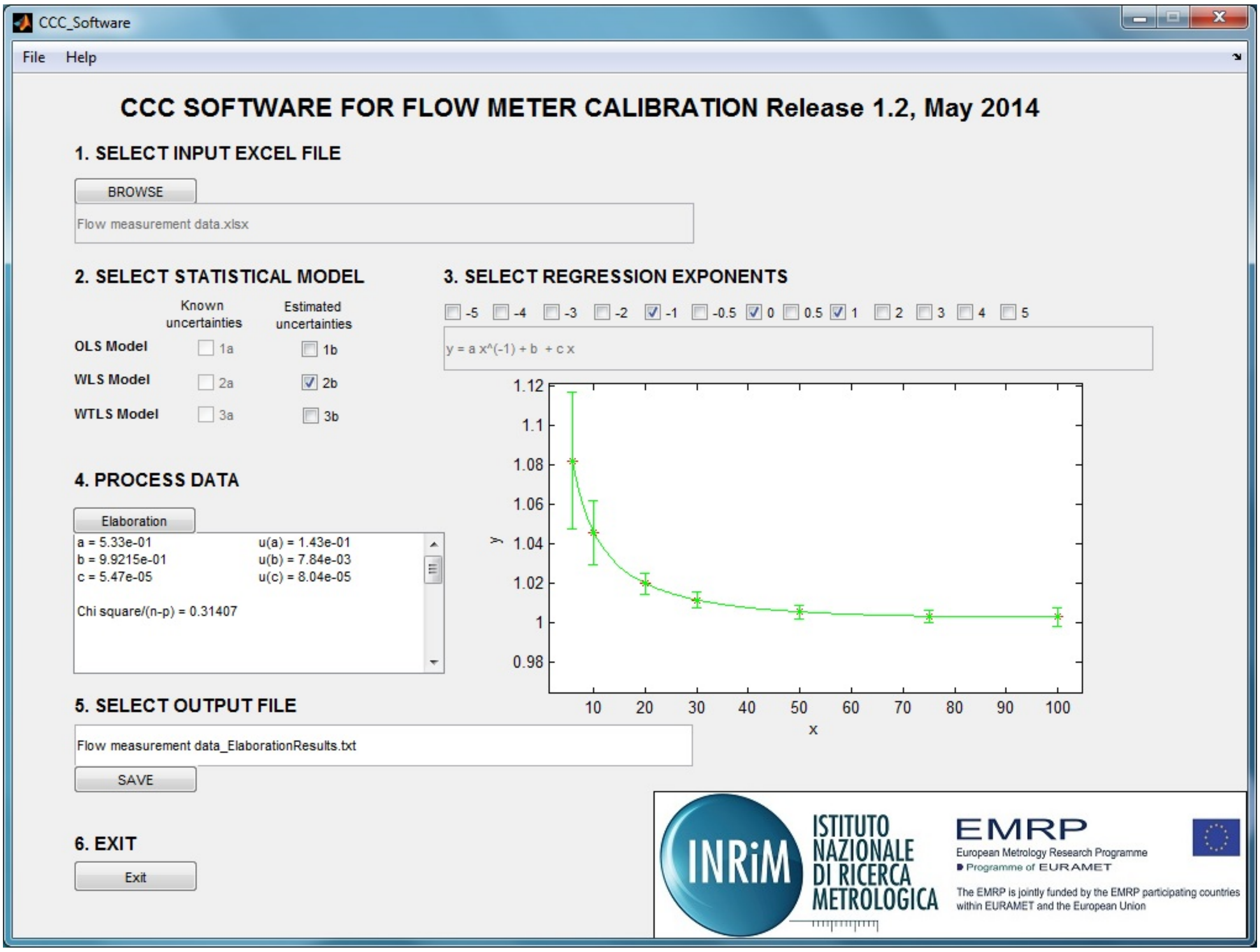

Figure 1: Main GUI of the CCC Software.

4.2. SELECT INPUT EXCEL FILE

Pressing the BROWSE button allows the User to select a Microsoft Excel workbook, i.e., a .xls or a .xlsx file. On pressing the BROWSE button, the software opens a browsing window. The default directory proposed by the software is the directory where the main MATLAB script is located, but the User can browse other directories by the usual commands of the browsing window.

The input data file should have the three worksheets titled and filled in as described in the following:

- Worksheet “Data”: the first column contains the *x* values and the second column contains the corresponding *y* values. The first line of both columns contains the corresponding headers, which will be displayed as axis legends of the data plot. If repeated measurements are made, they must have the same number of repetitions, each group of values being separated from the following by a blank cell.
- Worksheet “Var_x”: contains the covariance matrix associated with the *x* data. If worksheet “Data” contains *x* and *y* values divided into groups, worksheet “Var_x” must be empty.
- Worksheet “Var_y”: contains the covariance matrix associated with the *y* data (it can be provided in the form of a single value or a column vector or a positive definite square

matrix). If worksheet "Data" contains *x* and *y* values divided into groups, worksheet "Var_y" must be empty.

Once the data is loaded, the file name is displayed within the text box below the `BROWSE` button and the data couples (*x*, *y*) are shown as a scatter plot within a figure in the GUI.

Specific restrictions and requirements relevant to the input Excel file are addressed in Appendix B.

4.3. SELECT STATISTICAL MODEL

The following regression models are available, according to models described in [1]:

- OLS Model 1a: OLS model with known uncertainties;
- OLS Model 1b: OLS model with estimated uncertainties;
- WLS Model 2a: WLS model with known uncertainties;
- WLS Model 2b: WLS model with estimated uncertainties;
- WTLS Model 3a: WTLS model with known uncertainties;
- WTLS Model 3b: WTLS model with estimated uncertainties.

OLS models assume that only the *y* values are subject to uncertainty (or, equivalently, the uncertainties associated with the *x* values are negligible) and that all the *y* values have the same uncertainty.
WLS models assume that only the *y* values are subject to uncertainty and are uncorrelated[1], but without necessarily being homoscedastic.
WTLS models assume that the *x* values are also subject to uncertainty. The *x* values may be correlated among themselves, and the *y* values may be correlated among themselves, but there may be no correlation between *x* and *y* values.

Models "a" assume that uncertainty or covariance information associated with the measurement data are known and supplied by the User as described below:

- OLS Model 1a: a single value is provided as the common squared standard uncertainty associated with the *y* values in any of the cells within the "Var_y" worksheet;
- WLS Model 2a: a column vector of (positive) values having the same size as the data is provided, in any place within the "Var_y" worksheet, as the squared standard uncertainties associated with the *y* values (a corresponding diagonal covariance matrix is then constructed having that vector on its diagonal);
- WTLS Model 3a: a square $n \times n$ positive definite matrix (not necessarily diagonal), with n equal to data size, is provided, in any place within the "Var_x" worksheet, as the covariance matrix associated with the *x* values and a square $n \times n$ positive definite matrix (not necessarily diagonal), is provided, in any place within the "Var_y" worksheet, as the covariance matrix associated with the *y* values.

[1] Release 1.2 follows model 2a as defined in [1], in which the *y* values are uncorrelated, but it could be easily generalized to the case of a full covariance matrix associated with *y* values.

Models "b" assume that no uncertainty or covariance information associated with the measurement data is available. In this case, the measurement data within the "Data" worksheet must be organized into subgroups of repeated measurement values separated by blank spaces, each subgroup having the same number of repetitions, and the "Var_x" and "Var_y" worksheets must be empty. From the repeated measurement values, the software evaluates the uncertainty or the covariance matrices as follows:

- OLS Model 1b: the common standard deviation of the *y* data is estimated by means of the standard error of the regression, i.e., the square root of the sum of the squared residuals divided by the number of degrees of freedom (that is, the number *n* of data minus the number *p* of model parameters);
- WLS Model 2b: the covariance matrix associated with the *y* data is constructed as a diagonal matrix having on its diagonal the sample variance of each subgroup of data, repeated a number of times equal to the number of the measurement repetitions. This construction means that all the data pertaining to the same subgroup are considered as uncorrelated and having the same uncertainty, and that data pertaining to different subgroups are uncorrelated;
- WTLS Model 3b: both the covariance matrices associated with the *x* and *y* values are constructed as described for WLS Model 2b above (it is also assumed that there is no correlation between the *x* and *y* values).

Depending on the information available about the uncertainty or the covariance matrices associated with the data, only those models which make sense with the input data are selectable; this fact is graphically evidenced by disabling the check boxes corresponding to not appropriate models.

4.4. SELECT REGRESSION EXPONENTS

A subset of fractional polynomials [3] constitutes the class of curves among which the User can choose the regression model, i.e., positive, negative and fractional exponents are allowed (logarithmic transformation of the independent variable and repeated powers are not). The choice of the curve to be fitted to the data is performed by selecting the desired exponents among those in the following list:

```
-5, -4, -3, -2, -1, -0.5, 0, 0.5, 1, 2, 3, 4, 5
```

This list derives from the experience developed by personnel performing flow meter calibrations over the years; it was observed, as reported in [1], that in general the calibration curve of a flow meter can be adequately fitted using a polynomial of degree less than or equal to 5; often, negative exponents are required to fit high-slope regions in the vicinity of the instrument threshold (usually, negative exponents -1 and -2 are sufficient, but for symmetry it was decided to set -5 as the minimum exponent value). Moreover, fractional exponents -0.5 and 0.5 were also included.

Only those exponents that are appropriate for the input data may be selected; this fact is graphically evidenced by the software disabling the check boxes corresponding to invalid exponents (see Appendix B for restrictions on the exponents).

As the exponents are ticked, the text box below displays the updated expression for the resulting calibration curve.

4.5. PROCESS DATA

When the statistical model has been selected and at least one exponent has been chosen, the `Elaboration` button becomes active. The number of selected exponents must be smaller than the number of data.

On pressing the `Elaboration` button, the optimization problem is solved according to the chosen statistical model and the selected regression exponents (see Appendix A for more details on the optimization being implemented).

Following solution of the optimization problem, the text box below the `Elaboration` button displays the estimates of the calibration curve parameters together with their associated standard uncertainties, and the normalized chi-squared value, i.e., the sum of the squared normalized residuals divided by the number of degrees of freedom. Moreover, the resulting regression curve is added to the scatter plot of the data, and the expanded uncertainty bars (corresponding to a coverage factor $k = 2$) associated with the fitted[2] y values are visualized.

4.6. SELECT OUTPUT FILE

Once the elaboration is completed, the `Save` button under the “SELECT OUTPUT FILE” box becomes active and, in the box itself, a name is automatically suggested for the output file: “Filename_ElaborationResults.txt”, where “Filename” is the name of the Excel input data file. On pressing the `Save` button, the software displays the browsing window “Select the Output data file”, which automatically opens the same directory where the main MATLAB script is located. However, the file name and its location can be modified, placing the file in the desired directory and saving it by pressing the `Save` button of the browsing window. If such a file already exists in the selected directory, it is possible to overwrite it or to save it with a different name. The chosen file name is then displayed within the text box below the `Save` button of the GUI, and the `Save` button is disabled.

The output file reports the following information:

- Main information on the software;
- Name of the input data Excel workbook;
- Two (column) vectors of the experimental x and y data;
- Number of experimental data couples;

[2] For OLS and WLS models, fitted y values are calculated for the data x values, whereas for WTLS models they are calculated for the estimated x values (see also Appendix A).

- Covariance matrices (or single variance value) associated with data *y* and, possibly, with data *x* (the covariance matrices are provided by the User, for models “a”, or are estimated by the software, for models “b”);
- Statistical model;
- Regression curve;
- Parameter estimates and associated standard uncertainty;
- Normalized chi-squared value of the regression;
- Covariance matrix associated with the parameter estimates;
- Fitted *y* values and the associated covariance matrix.

The plot of the regression curve displayed in the GUI is automatically saved in a .tif format as “Filename_ElaborationResults_plot.tif”, in the same directory where the output file is saved.

In the following, an example output file is reported:

```
EMRP NEW04 CCC Calibration Software: Release 1.2
  (programming language: MATLAB R2013a)

 Program identifier:

 EMRP NEW04 CCC Calibration Software (Release 1.2)
 Authors: A Malengo, F Pennecchi, P Spazzini
 Istituto Nazionale di Ricerca Metrologica - INRIM, Italy
 Release Date: May 30, 2014
 Developed in the framework of WP1, EMRP Project NEW04
 Project financed by EURAMET

  Data identifier:

 Ordinary, weighted and weighted total least-squares method
 for fitting (fractional) polynomial curves to given data (x, y, Ux and Uy)
 according to models 1a, 1b, 2a, 2b, 3a and 3b defined in Deliverable 1.1.4

  Input data from File: C:\Few data.xlsx

  Experimental data x and y

                  1                                10
                1.5                                 9
                  4                                38
                3.5                                42

  Total experimental points no = 4

  Known covariance matrix associated with x:

         0.4166666667                          0                         0                         0

                    0               0.4166666667                         0                         0

                    0                          0              0.4166666667                         0

                    0                          0                         0              0.4166666667

  Known covariance matrix associated with y:

          2.916666667                          0                         0                         0

                    0                2.916666667                         0                         0

                    0                          0               2.916666667                         0

                    0                          0                         0               2.916666667
```

```
 Regression Model: Model 3a

 Fitted polynomial curve: y = a  + b x

  REGRESSION RESULTS:

a = -6.4281692              u(a) = 9.370249411
b = 12.47127207             u(b) = 3.367400858

Chi square/(n-p) = 0.58342

 Covariance matrix associated with parameter estimates:

        +87.80157402             -28.34845408

        -28.34845408             +11.33938854

 Fitted values y

         9.8295830373
         9.1412094040
        38.2350197603
        41.7941885397

 Covariance matrix associated with fitted values y:

         +2.82789527            -0.9809788842            -0.7409801278            -0.6600553358

       -0.9809788842            +39.03355671             +2.341229847             -1.773198798

       -0.7409801278            +2.341229847             +135.8204274              +118.775383

       -0.6600553358            -1.773198798              +118.775383             +153.1013684
```

The results in the output file are reported up to ten significant digits[3], in order to provide the User with all the available information, in case it should be used for subsequent evaluations. However, for readability reasons, the results shown in the GUI are rounded in the following way: uncertainties associated with the parameter estimates are reported to the third significant digit and relevant estimates are reported to the corresponding decimal digit.

4.7. EXIT

Pressing the EXIT button of the GUI terminates the program and closes the GUI. Alternatively, the User may click on ‘File’ in the main menu of the GUI and then click on ‘Exit’ (or click on the upper-right cross of the GUI window).

4.8. HELP

Clicking on ‘Help’ in the main menu of the GUI and then clicking on ‘About CCC Software’ causes a message box displaying information about the software to be displayed. Click on ‘OK’ to close this message box.

Clicking on ‘Help’ in the main menu of the GUI and then clicking on ‘Help Documentation’ causes the User manual (this document) to be opened (using the default program for viewing PDF files).

[3] The validation process on OLS and WLS models pointed out an agreement of 8 to 10 digits between the results produced by the CCC software and those produced by an independent software working in higher-than-normal precision arithmetic.

Clicking on 'Help' in the main menu of the GUI and then clicking on 'Licence Agreement' causes the licence agreement to be opened (using the default program for viewing PDF files).

5. ACKNOWLEDGMENT

The authors are grateful to NPL colleagues for assisting with validating and testing Release 1.1 of the software, and for fruitful discussions on a preliminary version thereof.

The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union.

6. REFERENCES

[1] Gertjan Kok, Adriaan van der Veen, Peter Harris, Ian Smith, "Statistical model and prior knowledge for the determination of calibration curves of flow", Deliverable D1.1.4 of WP 1, report of the EMRP joint research project NEW04 "Novel mathematical and statistical approaches to uncertainty evaluation".

[2] The MathWorks, Inc. MATLAB, http://www.mathworks.com

[3] Royston, P., and D. G. Altman. 1994. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. Applied Statistics 43: 429–467.

[4] A. Malengo and F. Pennecchi, A weighted total least-squares algorithm for any fitting model with correlated variables, *Metrologia* (2013), **50**, 654.

A. STATISTICAL MODELS

In this Appendix, information is provided on the implementation of the different least-squares methods corresponding to the statistical models described in Section 4.3.

A.1 Notation

$\boldsymbol{x}$ vector of x values (n x 1)

$\boldsymbol{y}$ vector of y values (n x 1)

$\boldsymbol{X}$ design matrix (n x p)

$\boldsymbol{\beta}$ vector of regression parameters (p x 1)

$\boldsymbol{\varepsilon}$ vector of errors in variable y (n x 1)

$\boldsymbol{\delta}$ vector of errors in variable x (n x 1) (for WTLS models only)

u_y common standard uncertainty associated with y values (for OLS models)

$\boldsymbol{V}_y$ covariance matrix associated with y values (n x n) (for WLS and WTLS models)

$\boldsymbol{V}_x$ covariance matrix associated with x values (n x n) (for WTLS models only)

$\widehat{\boldsymbol{\beta}}$ vector of parameter estimates (p x 1)

$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ covariance matrix associated with $\widehat{\boldsymbol{\beta}}$ (p x p)

$\widehat{\boldsymbol{y}}$ vector of fitted $\hat{y}$ values (n x 1)

$\widehat{\boldsymbol{x}}$ vector of fitted $\hat{x}$ values (n x 1) (for WTLS models)

χ^2_{N} normalized chi-squared value of the regression

The mathematical symbols mentioned above for uncertainties and covariance matrices associated with data are used in this section irrespective of whether they have been provided by the User, for models "a", or they have been estimated by the software, for models "b".

B.1 Models

For OLS and WLS fitting, the underlying statistical model is given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \tag{1}$$

where

$$\boldsymbol{X}=\begin{bmatrix} {x_1}^{\beta_1} & \cdots & {x_1}^{\beta_p} \\ \vdots & \ddots & \vdots \\ {x_n}^{\beta_1} & \cdots & {x_n}^{\beta_p} \end{bmatrix}. \tag{2}$$

For WTLS fitting, in which also the *x* values are subject to uncertainty, the model is given by

$$\boldsymbol{y}=\boldsymbol{X}^{*}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\,, \tag{3}$$

$$\boldsymbol{x}=\boldsymbol{x}^{*}+\boldsymbol{\delta}\,. \tag{4}$$

The choice of the fitting procedure depends on the assumptions made about the nature of errors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$, according to models 1 to 3 described in [1].

A.3 OLS fitting

When model errors ε_i in (1) are assumed to be i.i.d. as $\varepsilon_i \sim \mathrm{N}(0, u_y^2)$ (normal distribution with zero mean and standard deviation equal to u_y), for $i = 1, \ldots, n$, then the following formulae are implemented:

$$\widehat{\boldsymbol{\beta}}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}\,, \tag{5}$$

$$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}=u_y^2(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}, \tag{6}$$

$$\widehat{\boldsymbol{y}}=\boldsymbol{X}\widehat{\boldsymbol{\beta}}\,, \tag{7}$$

$$\boldsymbol{V}_{\widehat{\boldsymbol{y}}}=\ \boldsymbol{X}\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{X}^{\mathrm{T}}\,, \tag{8}$$

$$\chi_{\mathrm{N}}^2=\frac{1}{n-p}\sum_{i=1}^{n}\left(\frac{y_i-\hat{y}_i}{u_y}\right)^2. \tag{9}$$

A.4WLS fitting

When model errors $\boldsymbol{\varepsilon}$ in (1) are assumed to be distributed as $\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{V}_y)$ (multivariate normal distribution with zero mean vector and covariance matrix $\boldsymbol{V}_y$), then the following formulae are implemented:

$$\widehat{\boldsymbol{\beta}}=\left(\boldsymbol{X}^{\mathrm{T}}{\boldsymbol{V}_y}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}{\boldsymbol{V}_y}^{-1}\boldsymbol{y}\,, \tag{10}$$

$$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}=\left(\boldsymbol{X}^{\mathrm{T}}{\boldsymbol{V}_y}^{-1}\boldsymbol{X}\right)^{-1}, \tag{11}$$

$$\widehat{\boldsymbol{y}}=\boldsymbol{X}\widehat{\boldsymbol{\beta}}\,, \tag{12}$$

$$\boldsymbol{V}_{\widehat{\boldsymbol{y}}}=\ \boldsymbol{X}\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{X}^{\mathrm{T}}\,, \tag{13}$$

$$\chi_{\mathrm{N}}^2=\frac{1}{n-p}\,(\widehat{\boldsymbol{y}}-\boldsymbol{y})^{\mathrm{T}}\,{\boldsymbol{V}_y}^{-1}\,(\widehat{\boldsymbol{y}}-\boldsymbol{y})\,. \tag{14}$$

A.5 WTLS fitting

When model errors $\boldsymbol{\delta}$ in (4) and $\boldsymbol{\varepsilon}$ in (3) are assumed to be distributed as $\boldsymbol{\delta}\sim \mathrm{N}(\mathbf{0}, V_x)$ and $\boldsymbol{\varepsilon}\sim \mathrm{N}(\mathbf{0}, \boldsymbol{V}_y)$, respectively, then the following formulae are implemented:

$$[\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\beta}}] = argmin_{\{\boldsymbol{x}^*,\boldsymbol{\beta}\}} \{ (\boldsymbol{x}^* - \boldsymbol{x})^{\mathrm{T}} \boldsymbol{V}_x{}^{-1} (\boldsymbol{x}^* - \boldsymbol{x}) + (\boldsymbol{X}^*\boldsymbol{\beta} - \boldsymbol{y})^{\mathrm{T}} \boldsymbol{V}_y{}^{-1} (\boldsymbol{X}^*\boldsymbol{\beta} - \boldsymbol{y})\}, \quad (15)$$ [4]

$$\boldsymbol{V}_{[\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\beta}}]} = \boldsymbol{H}^{-1}\boldsymbol{D}\, \boldsymbol{V}\, (\boldsymbol{H}^{-1}\boldsymbol{D}\,)^{\mathrm{T}}, \quad (16)$$

where $\boldsymbol{H}$ is the Hessian matrix of the cost function (i.e., the expression to be minimized) in (15); $\boldsymbol{D}$ is the matrix of the mixed second-order derivatives of the cost function with respect to parameters $\boldsymbol{x}^*$ and $\boldsymbol{\beta}$, and with respect to input data $\boldsymbol{x}$ and $\boldsymbol{y}$; and

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_x & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_y \end{bmatrix}.$$

Covariance matrix $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ is the lower right sub matrix in $\boldsymbol{V}_{\{\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\beta}}\}}$ of dimension (p x p).
Moreover,

$$\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}\,, \quad (17)$$

$$\boldsymbol{V}_{\widehat{\boldsymbol{y}}} = \boldsymbol{J}\, \boldsymbol{V}_{[\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\beta}}]}\, \boldsymbol{J}^{\mathrm{T}}, \quad (18)$$

where $\boldsymbol{J}$ is the matrix of the first-order derivatives of the regression model $\boldsymbol{X\beta}$, with respect to both $\boldsymbol{x}$ and $\boldsymbol{\beta}$, calculated at estimates $[\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\beta}}]$, and

$$\chi^2_{\mathrm{N}} = \frac{1}{n-p}\{ (\widehat{\boldsymbol{x}} - \boldsymbol{x})^{\mathrm{T}} \boldsymbol{V}_x{}^{-1} (\widehat{\boldsymbol{x}} - \boldsymbol{x}) + (\widehat{\boldsymbol{y}} - \boldsymbol{y})^{\mathrm{T}} \boldsymbol{V}_y{}^{-1} (\widehat{\boldsymbol{y}} - \boldsymbol{y})\}. \quad (19)$$

For more details on the implemented WTLS fitting procedure, see [4][5].

---

[4] The expression to be minimized is non-linear in its parameters $\boldsymbol{\beta}$ and $\boldsymbol{x}^*$, hence a numerical solution is found by implementing an algorithm based on the MATLAB function `fminunc.m`, available within the MATLAB Optimization Toolbox™.

[5] Slightly different results from those reported in the paper may be obtained since in Release 1.2 of the CCC software the minimization option "GradObj" of the `fminunc.m` function has not been implemented.

B. MESSAGES

In this section, a list of the messages that can be visualized when running the software is reported:

Errors and warnings when inputting data:

- “`An input data file must be selected`”: the User has pressed the `Cancel` button and therefore not selected a file.
- “`The input data file must be an Excel workbook`”: the extension of the input data file is not .xls or .xlsx.
- “`The input data file must contain the worksheets "Data", "Var_x" and "Var_y"`”: at least one of the prescribed worksheets is not present in the input data file.
- “`The worksheet "Data" must contain x and y data`”: measurement data are not present in the worksheet "Data".
- “`The worksheet "Data" must contain headers for x and y data`”: headers for *x* and *y* data are not present in the worksheet "Data".
- “`The number of data x must be equal to that of data y`”: *x* and *y* data have different size.
- “`The number of groups within data x must be equal to that within data y`”: *x* and *y* data are divided into a different number of groups.
- “`The number of data within each group must be the same`”: groups of data have different dimensions.
- “`When x and y data are divided into groups, the worksheets "Var_x" and "Var_y" must be empty`”: covariance matrices are provided although data are grouped.
- “`The groups within x and y data must be separated by a single empty row`”: groups of data are separated by more than one blank cell.
- “`Some groups of x data contain the same repeated values. When such repetitions occur, only Models 1b and 2b are allowed`”: one or more groups of data *x* contain the same repeated values.
- “`Some groups of x or y data contain the same repeated values. When such repetitions occur, only Model 1b is allowed`”: one or more groups of data *y* and/or data *x* contain the same repeated values.
- “`When x and y data are not divided into groups, at least the covariance matrix Vy must be given in the worksheet "Var_y"`”: no covariance matrix is provided although data are not divided into groups.
- “`The dimensions of the covariance matrix Vy must agree with the size of y data`”: matrix dimension do not agree with data.
- “`The dimensions of the covariance matrix Vx must agree with the size of x data`” : matrix dimension do not agree with data.
- “`The squared standard uncertainty associated with y data must be positive`”: a non-positive value has been provided as the squared standard uncertainty of *y* data.
- “`Covariance matrix Vy must be positive definite`”: the covariance matrix associated with *y* values are not positive definite.

- “`Both covariance matrices must be positive definite`”: either or both of the covariance matrices are not positive definite.
- “`Covariance matrices Vx and Vy must contain only numeric values`”: some not-a-number (NaN) value is present within the covariance matrices.
- “`All input data x should be non-negative. When some are negative, only integer exponents must be used`”: some *x* data are negative[6].
- “`When some x data are null, only positive exponents must be used`”: some *x* data are zero.

<u>Errors and warnings when selecting the regression exponents</u>:

- “`The number of selected exponents must be smaller than the number of data`”: when selecting the regression exponents, the number of the ticked exponents has reached the number of the experimental data.

<u>Errors and warnings when processing the data</u>:

- “`Minimization problems, the result may be unreliable`”: the software has encountered a problem when trying to solve the fitting problem (only for WTLS models).
- “`Warning: Matrix is close to singular or badly scaled`”: this warning can occasionally be displayed in the MATLAB Command Window when using OLS and WLS models involving ill-conditioned or nearly ill-conditioned linear systems[7] (due to roundoff error, the linear system can be seen as nearly singular).

<u>Errors and warnings when saving the results</u>:

- “`An output file must be selected`”: the User has pressed the `Cancel` button and therefore not selected a file.

[6] All *x* data should be positive, since they are flow values; should some *x* data be negative, the calibration is still possible, but without selecting fractional exponents.

[7] Scaling the measurement data may help fixing the problem.