CMSC/M/06/657

XLGENLINE Version 1.0 - Software Documentation

By
I M Smith and F O Onakunle

Issued 28 March 2007

Version 1.0

Unrestricted

Mathematics and Scientific Computing Group
National Physical Laboratory
Hampton Road
Teddington
Middlesex
TW11 0LW

Tel 020 8943 7071
Fax 020 8977 7091

Project name:	Software for generalised least-squares fitting
Project acronym:	SS*f*M-3 1.6.1 - XLGENLINE
Document reference:	CMSC/M/06/657
NPL File reference:	CMSC//09/03/M02
Approved by:	BJC

National Physical Laboratory

Hampton Road, Teddington, Middlesex, TW11 0LW

1 INTRODUCTION

This document describes how to use the Microsoft Excel spreadsheet XLGENLINEv1.XLS (which calls the DLL XLGENLINEv1.DLL) to carry out ordinary least-squares and generalised least-squares polynomial fitting and inverse polynomial evaluation.

The underlying software is written in Fortran 90 that has been compiled into a DLL using Compaq Visual Fortran (Version 6.6). The DLL is called from Microsoft Excel with the code used to implement this call written in Microsoft Visual Basic 6.0.

The software is downloaded as the ZIP file xlgenlinev1.zip (2.3 MB). The main files within this ZIP file are:

- XLGENLINEv1.dll,
- XLGENLINEv1.xls, and
- xlgenlinev1_sw_doc.pdf (this document).

The document is organised as follows. Section 2 lists the steps required to be able to call the DLL from Microsoft Excel on a PC running Windows 2000 or XP while Section 3 describes how to use the spreadsheet, detailing the format that input data is required to take and the outputs that are generated. Section 4 provides contact details for feedback and/or questions.

2 INSTALLING AND UNINSTALLING THE SOFTWARE

Prior to installing the software, ensure that any instances of Excel are closed.

To install the software, first extract the contents of the ZIP file to a local folder, double-click on setup.exe to commence the installation process, then:

1. Left-click on “Next”.
2. Read the licence agreement, left-click on the radio button corresponding to “I accept the terms in the licence agreement” and left-click on “Next”.
3. Installation of the Fortran run-time libraries will commence. Left-click on “Next”. Left-click on “Next”. (A message may be displayed indicating that the read only file fqwin.hlp was found when copying files to the destination location. If this message appears, left-click on “Yes” to overwrite the file.) Left-click on “Finish”.
4. Enter user name and organization and left-click the appropriate radio button to specify if XLGENLINEv1 is to be made available to all users of the PC or just to the currently logged-on user. Left-click on “Next”.
5. Select the destination folder for the XLGENLINEv1 files. (The default folder is C:Program Files\XLGENLINE but another folder can be chosen by left-clicking on “Change…” and navigating to that folder and left-clicking on “OK”.) Left-click on “Next”.

6. Left-click on “Install”.

7. Left-click on “Finish” to exit the wizard.

Note that it is possible to go back and change options chosen in steps 4 and 5 by left-clicking on “Back” and entering new choices.

The installation described above need only be carried out once, before first using the XLGENLINEv1 spreadsheet. The name of the folder containing the file XLGENLINEv1.DLL (chosen in step 5 above) is added to the PATH environment variable during the installation process.

While XLGENLINEv1.XLS contains example data, it can be considered to be a template spreadsheet. The user may rename the main worksheet, make and use a copy (or copies) of the main worksheet within the same spreadsheet, and make and use a copy (or copies) of XLGENLINEv1.XLS.

Provided that the name of the folder containing the file XLGENLINEv1.DLL is not removed from the PATH environment variable and that XLGENLINEv1.DLL is not removed from this folder, the user should be able to use the spreadsheet XLGENLINEv1.XLS, or any copies thereof, placed in this or any other folder.

To uninstall the software, double-click on setup.exe, then:

1. Left-click on “Next”.

2. Left-click on the radio button corresponding to “Remove” and left-click on “Next”.

3. Left-click on “Remove”.

4. Left-click on “Finish” to exit the wizard.

The uninstallation process deletes the folder where the XLGENLINE files were placed (chosen in step 5 of the installation process) and removes the name of this folder from the PATH environment variable.

3 USING XLGENLINEv1.XLS

Upon opening the spreadsheet (e.g., by double-clicking on it), a message may be displayed indicating that it contains macros and asking if the user wishes to disable or enable macros. The “Enable” option should be chosen.

The spreadsheet should appear as in figure 1.

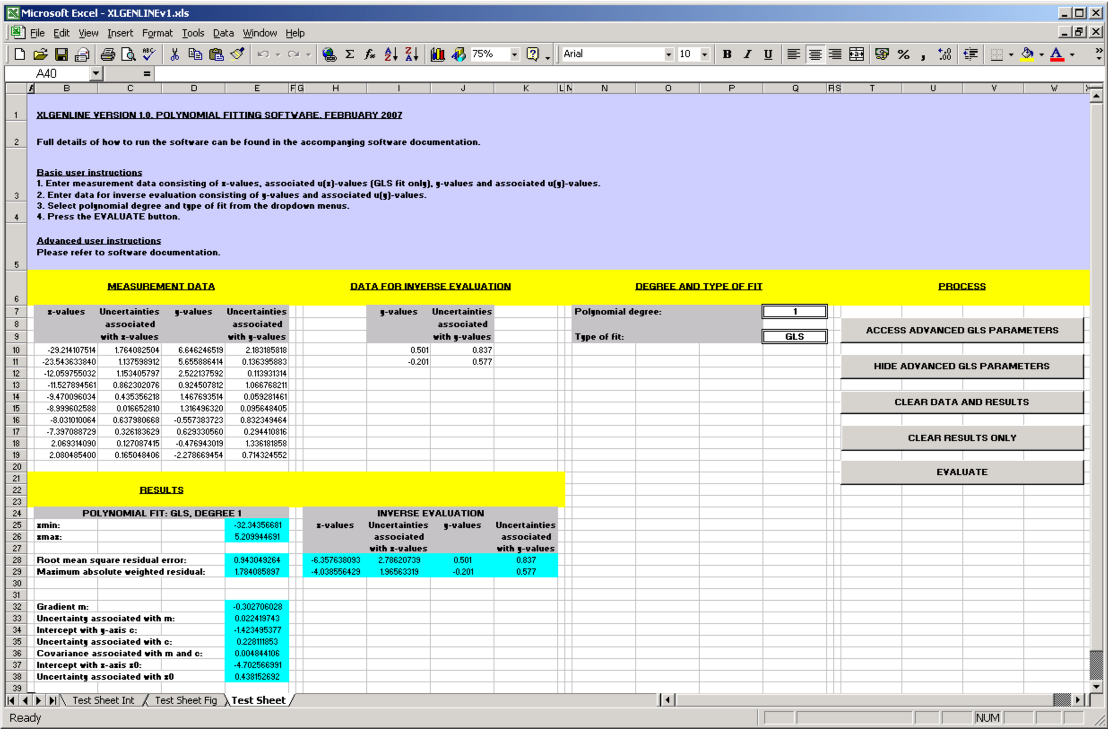


Figure 1: XLGENLINEv1 example spreadsheet.

3.1 Entering data and selecting fitting options

Figure 2 shows a close-up of the five buttons that may be pressed.

Any results on the current worksheet may be cleared by left-clicking on the “**CLEAR RESULTS ONLY”** button.

Results, measurement data and fitting options may be cleared by left-clicking on the “**CLEAR DATA AND RESULTS**” button.

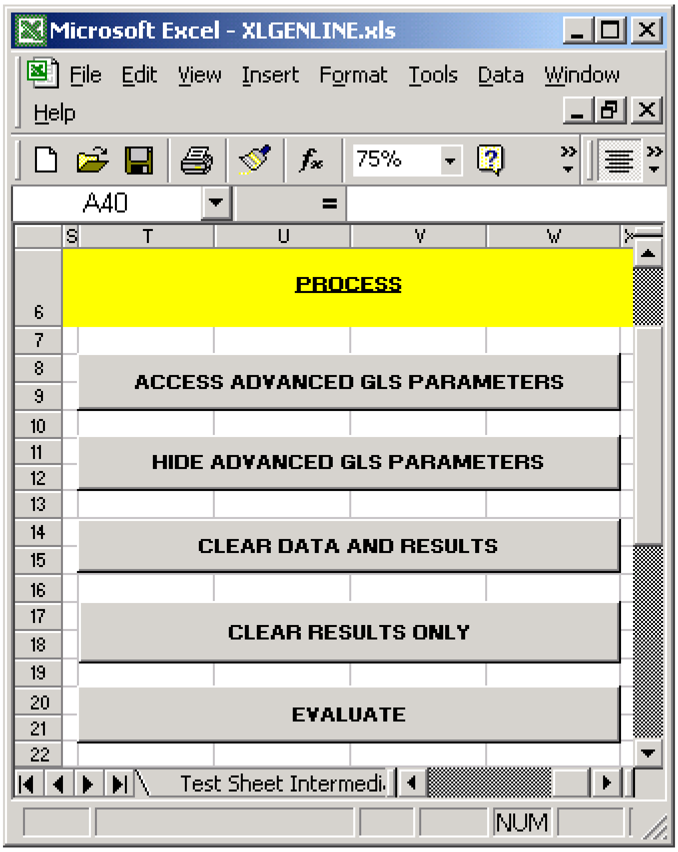


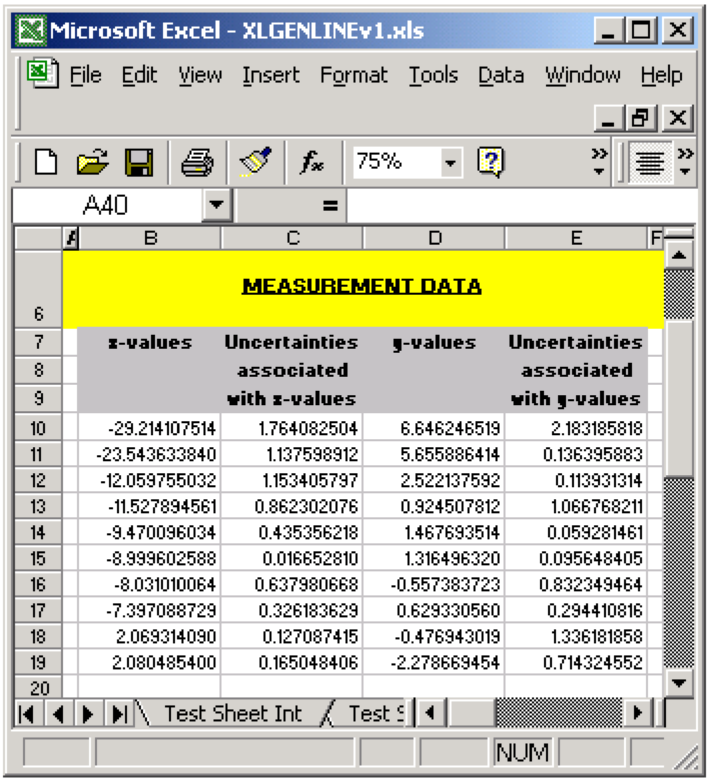
Figure 2: XLGENLINEv1 processing buttons.

3.1.1 Measurement data

For ordinary least squares (OLS) fitting, measurement data consisting of *x*-values, *y*-values and uncertainties associated with *y*-values should be entered into columns B, D and E, respectively, starting at row 10.

For generalised least squares (GLS) fitting, measurement data consisting of *x*-values, uncertainties associated with *x*-values, *y*-values and uncertainties associated with *y*-values should be entered into columns B, C, D and E, respectively, starting at row 10.

Figure 3 illustrates ten measurement data points for OLS or GLS fitting.



MEASUREMENT DATA

Row	x-values	Uncertainties associated with x-values	y-values	Uncertainties associated with y-values
10	-29.214107514	1.764082504	6.646246519	2.183185818
11	-23.543633840	1.137598912	5.655886414	0.136395883
12	-12.059755032	1.153405797	2.522137592	0.113931314
13	-11.527894561	0.862302076	0.924507812	1.066768211
14	-9.470096034	0.435356218	1.467693514	0.059281461
15	-8.999602588	0.016652810	1.316496320	0.095648405
16	-8.031010064	0.637980668	-0.557383723	0.832349464
17	-7.397088729	0.326183629	0.629330560	0.294410816
18	2.069314090	0.127087415	-0.476943019	1.336181858
19	2.080485400	0.165048406	-2.278669454	0.714324552

Figure 3: Example measurement data.

3.1.2 Data for inverse polynomial evaluation (OPTIONAL)

Data consisting of *y*-values and uncertainties associated with *y*-values should be entered into columns I and J, starting at cell 10. If no inverse evaluation is to be carried out, these columns should be left blank.

Figure 4 illustrates two data points for inverse evaluation.

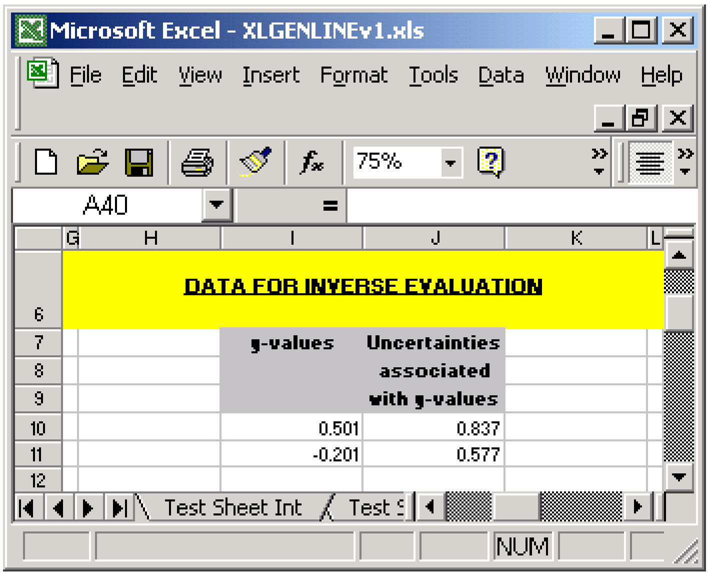


Figure 4: Example data for inverse evaluation.

3.1.3 Selecting fitting parameters and select type of fit

To select the polynomial degree, first left-click on cell Q7 then left-click on the arrow that appears to the right of the cell. A dropdown menu appears from which the polynomial degree (between 1 and 4 inclusive) can be chosen by left-clicking on the appropriate number.

To select the fitting type, first left-click on cell Q9 then left-click on the arrow that appears to the right of the cell. A dropdown menu appears from which the fitting type can be chosen by left-clicking on “OLS” (for ordinary least squares fitting) or “GLS” (for generalised least-squares fitting).

Figure 5 shows a close-up of the degree and type of fit choices.

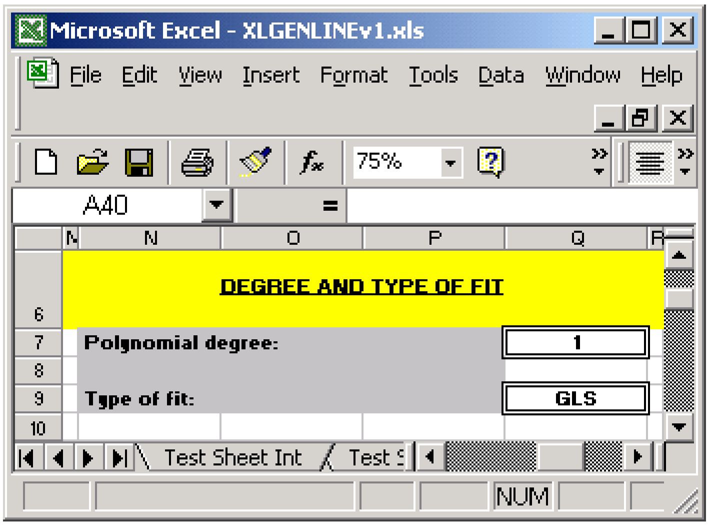


Figure 5: Example fitting parameters.

For GLS fitting, the user may also access additional advanced parameters and enter values for those parameters (see section 3.4).

3.2 EVALUATING

Having entered measurement data, (optionally) data for inverse evaluation and chosen the polynomial degree and type of fit, fitting (and, if required, inverse evaluation) is carried out by left-clicking on the "**EVALUATE**" button.

If the fitting is carried out successfully, then the following information will be displayed (see figure 6):

- The values taken by *xmin* and *xmax* (see section 3.4).
- The root-mean-square (weighted) residual error and the maximum absolute value of weighted residual.
- *If the polynomial degree is 1*, estimates of the gradient and intercept of the best-fit straight line and their associated standard uncertainties and covariance, and the intercept of the best-fit straight line with the *x*-axis and its associated standard uncertainty.

If inverse evaluation has been undertaken, the *y*-values and uncertainties associated with the *y*-values entered by the user are displayed along with the corresponding estimates of the *x*-values[1] and uncertainties associated with the estimates.[2]

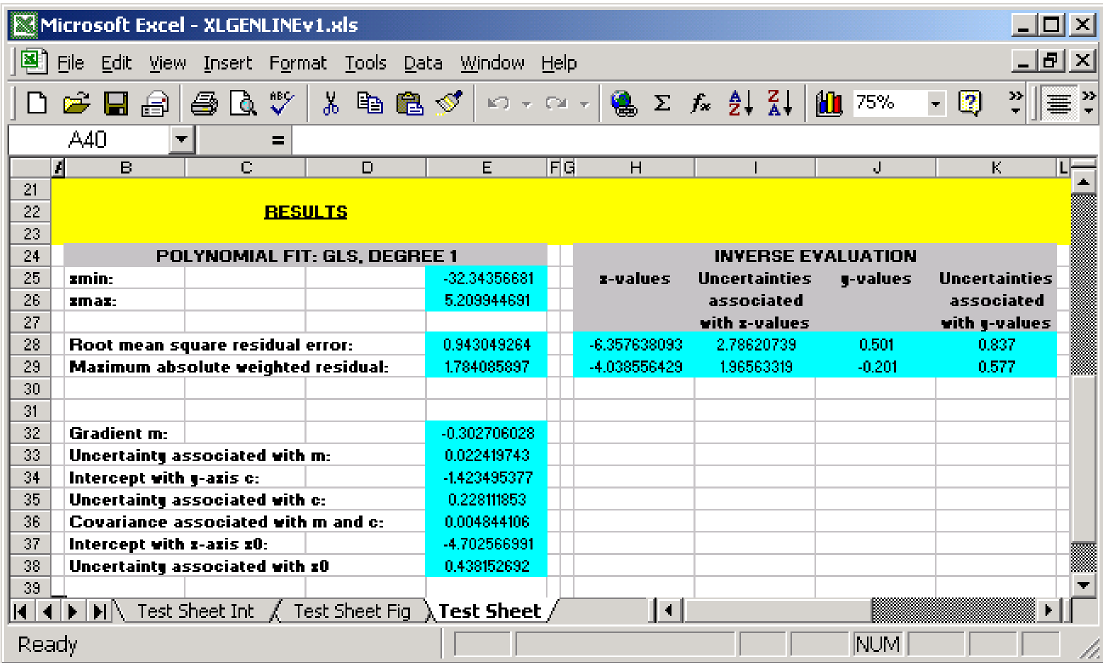


Figure 6: Example results.

---

[1] For a given *y*-value, the number of corresponding *x*-value estimates may be between zero and the polynomial degree inclusive

[2] Uncertainty evaluation is undertaken according to the GUM Uncertainty Framework [1].

In addition, a figure is produced (see figure 7), plotting the measurement data with "1 sigma error bars" (i.e., measured value plus and minus one standard uncertainty), the fitted polynomial and (optionally) the inverse evaluation data with "1 sigma error bars" (i.e., value plus and minus one standard uncertainty). The name of the worksheet containing the figure is determined by the name of the worksheet containing the data[3] – if the data worksheet is "Worksheet_name", the figure worksheet is "Worksheet_name Fig". An additional worksheet named "Worksheet_name Int" is also generated and is used to store data for the generation of the figure.

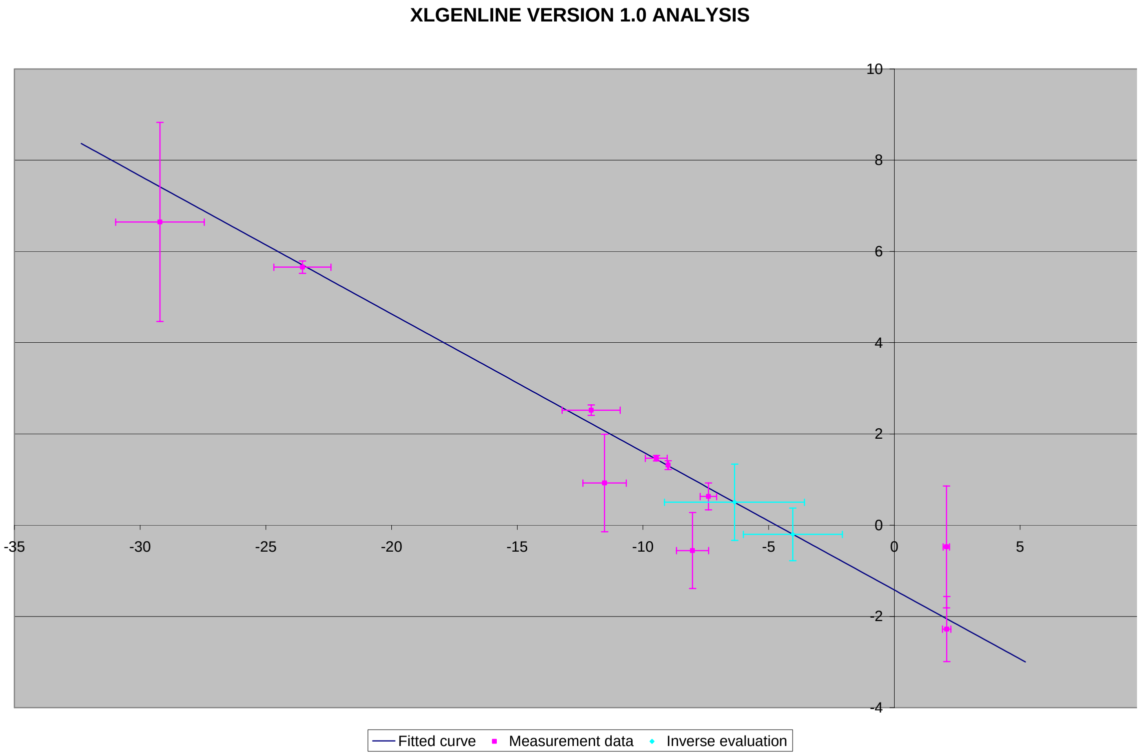


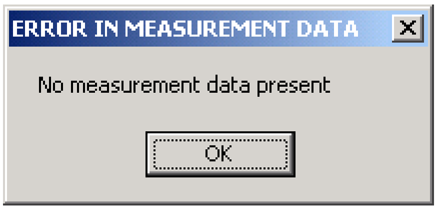
Figure 7: Example figure.

[3] The maximum number of characters in a worksheet name is 31. Since the names of the worksheets containing the figure and the data for the generation of the figure will contain 4 characters more than that of the main worksheet, the name of the main worksheet should consist of 27 or fewer characters.

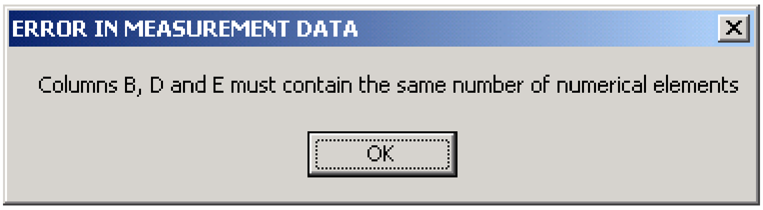
3.3 MESSAGES

The following message boxes may appear after pressing the "Evaluate" button:

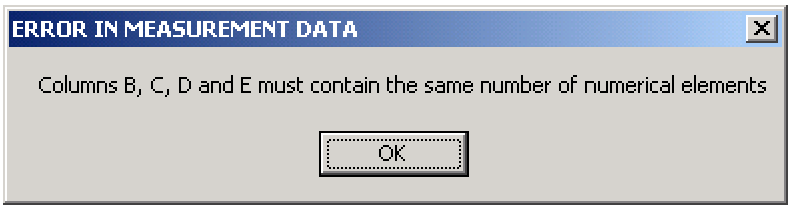
1. There is no measurement data in columns B to E.



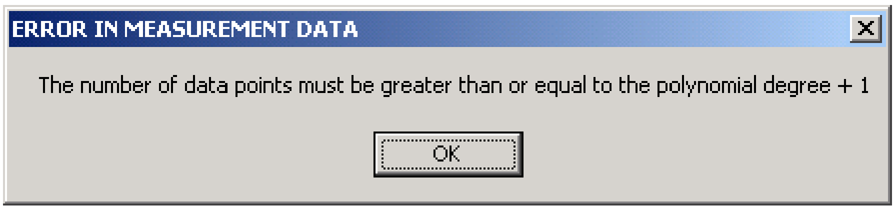
2. In the measurement data for OLS fitting, the numbers of *x*-values, *y*-values and uncertainties associated with *y*-values are not all the same.



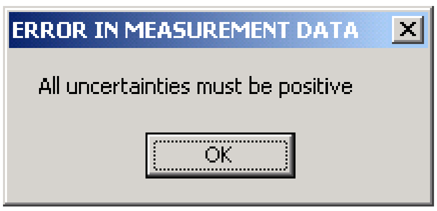
3. In the measurement data for GLS fitting, the numbers of *x*-values, uncertainties associated with *x*-values, *y*-values and uncertainties associated with *y*-values are not all the same.



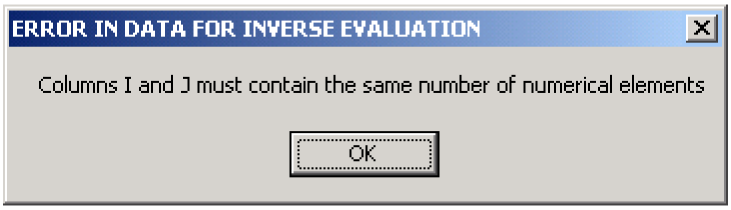
4. There are insufficient measurement data points to fit a polynomial of the chosen degree.



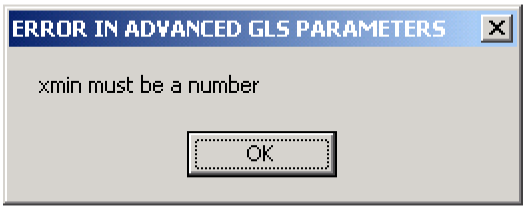
5. At least one uncertainty is less than or equal to zero.



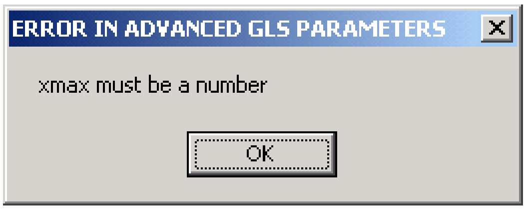
6. In the data for inverse evaluation, the numbers of *y*-values and uncertainties associated with *y*-values are not the same.



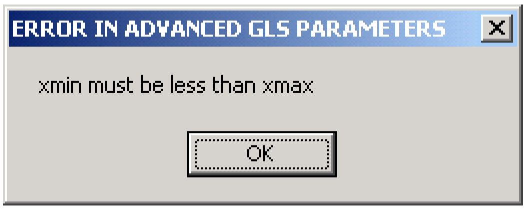
7. A non-numerical value for advanced GLS parameter *xmin* has been entered.



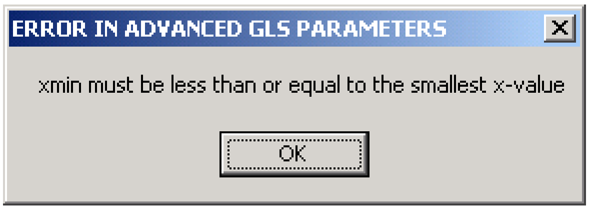
8. A non-numerical value for advanced GLS parameter *xmax* has been entered.



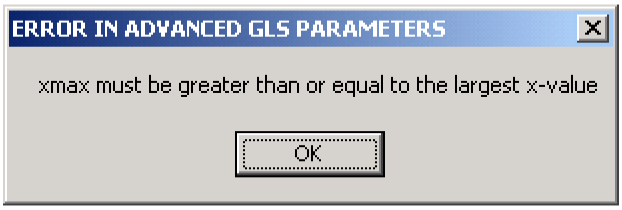
9. Values of *xmin* and *xmax* have been entered with $xmin \geq xmax$.



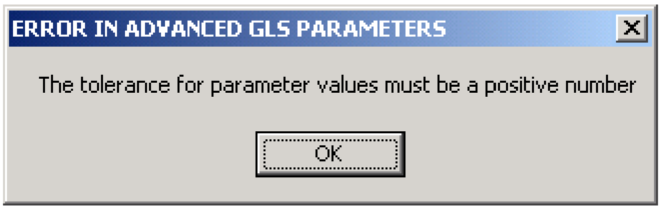
10. A value of *xmin* has been entered that is greater than the smallest *x*-value.



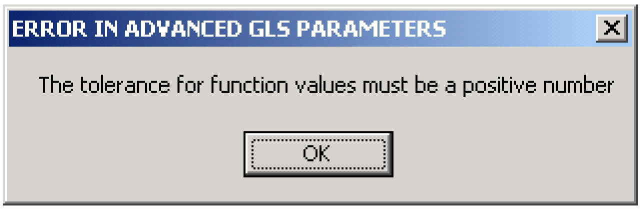
11. A value of *xmax* has been entered that is less than the largest *x*-value.



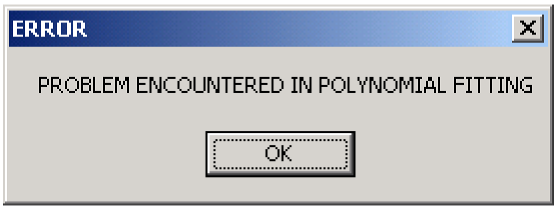
12. A non-positive value of *tola* has been entered.



13. A non-positive value of *tolf* has been entered.



14. The software has encountered a problem when trying to solve the polynomial fitting problem.



The user should be able to take steps to address messages 1 to 13.

Message 14 is more serious. In the first instance, it is recommended that the user check that:

- The measurement data has been entered correctly into the cells,
- The measurement data is appropriate for the choice of polynomial degree.
- The measurement data is not very small or very large and should be rescaled.

In the case of GLS fitting, the user may change the values of one or more of the advanced GLS parameters (see Section 3.4).

If there is no obvious reason as to why any error is encountered, the user may contact NPL as described in Section 4.

3.4 ADVANCED GLS PARAMETERS

The user may access advanced parameters for polynomial fitting as follows:

- Press the "**ACCESS ADVANCED GLS PARAMETERS**" button to access the advanced parameters for GLS (see figure 8).

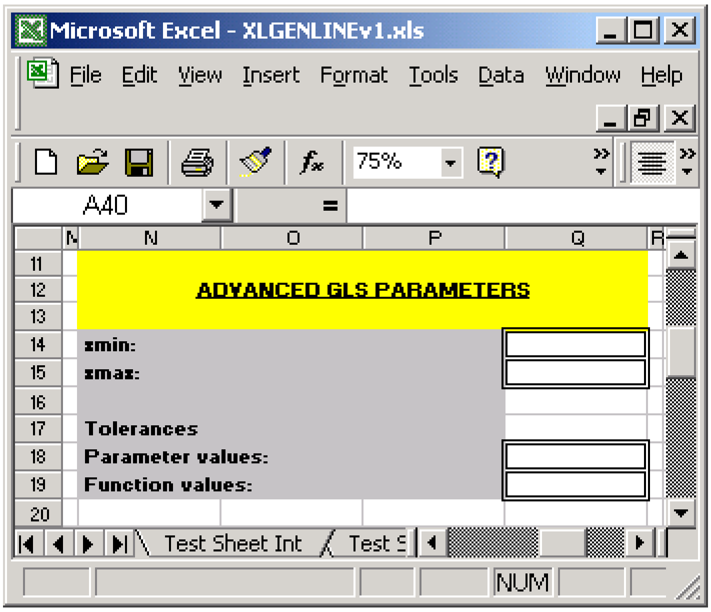


Figure 8: Example advanced GLS parameters (note that this figure may change depending on inclusion or non-inclusion of tolerances.

- Enter values for *xmin* (cell Q14), *xmax* (cell Q15), the tolerance *tolp* on parameter values and/or the tolerance *tolf* on function values (see sections 3.4.1 and 3.4.2 for descriptions of these parameters). Leaving any cell (or indeed all four cells) empty means that the default value(s) for the corresponding parameter is used.

- Press the "**EVALUATE**" button to process the data using the choices of advanced GLS parameters.

The advanced parameters for GLS fitting may be cleared and hidden by pressing the "**CLEAR ADVANCED GLS PARAMETERS**" button.

3.4.1 *xmin* and *xmax*

For reasons of numerical stability, polynomials are expressed in terms of a Chebyshev basis of a normalised variable, *viz.*, the polynomial of order n (degree $n - 1$) is written as

$$p(\boldsymbol{a}; x) = \sum_{j=1}^{n} a_j T_j(X),$$

where

$$X = \frac{(x - xmin) - (xmax - x)}{xmax - xmin},$$

and $T_j(X)$ is the jth order Chebyshev polynomial of the first kind [2, 3].

For OLS fitting, *xmin* and *xmax* are chosen to be the minimum and maximum x-values, respectively.

For GLS fitting, given initial estimates of coefficients a_j and abscissae x_i, an iterative algorithm (the Gauss-Newton algorithm) is used to find updated estimates. It is possible for one or more of the updated abscissae estimates to be assigned values that are less than the minimum x-values and/or greater than the maximum x-value, and values of *xmin* and *xmax* that allow for this possibility should therefore be assigned.

In the XLGENLINEv1 spreadsheet, *xmin* and *xmax* are assigned the default values

$$xmin = \min\{x_i\} - 0.1 \times [\max\{x_i\} - \min\{x_i\}],$$

$$xmax = \max\{x_i\} + 0.1 \times [\max\{x_i\} - \min\{x_i\}].$$

While for the majority of data sets, the default values of *xmin* and *xmax* will allow fitting to be undertaken successfully, there may be some data sets for which the iterative process generates abscissae estimates that lie outside the range [*xmin*, *xmax*]. For such data sets, a message box will be displayed (see message number 12 in Section 3.3).[4]

3.4.2 Tolerances on parameter and function values

As described in section 3.4.1, the GLS parameter values are determined by an iterative scheme. Tolerances are required to test whether or not the scheme has converged. The convergence criteria are related to the change in parameter values a_j and the change in function values from iteration to iteration. The parameter values a_j are on the same scale as the y-data. Thus, if the y-values range from –1000 to 1000, the a_j will be in a similar range.

The tolerance for the parameters has a default value of

$$tolp = 10^{-6} \times \max\{1, \max\{|y_i|\}\}.$$

If the range of y-values is much smaller than 1, a smaller tolerance than 10^{-6} may be appropriate. If the data is very noisy, then a larger tolerance may be appropriate.

The second tolerance *tolf* is compared with the change in function values. If the uncertainties associated with the x- and y-values entered in the spreadsheet are a reasonable reflection of the true uncertainties associated with the x- and y-values, then the function values should lie within a range comparable with [–1, 1].

The tolerance for the function values has a default value of

[4] Note that problems with the value(s) of *xmin* and/or *xmax* are not the only reason this message box may appear.

$$tolf = 10^{-8}.$$

If the uncertainties associated with the *x*- and *y*-values are an overestimate, then a smaller value for *tolf* may be appropriate. Conversely, if they are an underestimate, a larger value for *tolf* may be appropriate.

The default values of the tolerances are designed so that the solution determined by the solver will differ from the true mathematical solution by an amount that is much smaller than the uncertainties associated with the fitted parameter values. If the tolerances are too small, the GLS solver will usually return the best solution that can be expected from the data set. If the tolerances are too large, the GLS solver may return with solution parameters that could differ from the true solution by amounts greater than that consistent with the noise in the data.

When using the default values of the tolerances, the risk of the software failing is reduced when the user's measurement data (*x*-values and/or *y*-values) lie (approximately) within the range $[10^{-3}, 10^{3}]$. Scaling of the measurement data may therefore be required. For straight-line fitting, the estimates of the parameters of the straight-line and their associated uncertainties returned by the software must be rescaled appropriately. For inverse evaluation, the *x*- and/or *y*-values and their associated uncertainties must be rescaled.

4 ENQUIRIES AND FEEDBACK

Questions about and general comments/feedback on XLGENLINE are welcome and should be sent by e-mail to eurometros@npl.co.uk. (On the EUROMETROS homepage http://www.eurometros.org/, clicking on "Feedback" automatically generates a blank e-mail with the above e-mail address filled in.)

5 ACKNOWLEDGEMENTS

The National Measurement System Directorate of the UK's Department of Trade and Industry supported the development of this software as part of its Software Support for Metrology (SS*f*M) programme. The contribution of Frank Guenther (NIST) is also appreciated.

6 REFERENCES

[1] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML. *Guide to the measurement of uncertainty in measurement*, 2nd edition (Geneva, Switzerland: International Organisation for Standardisation) ISBN 92-67-10188-9, 1995.

[2] L Fox and I B Parker. *Chebyshev polynomials in numerical analysis*. Oxford University Press, 1968.

[3] R M Barker, M G Cox, A B Forbes and P M Harris. *SSfM Best Practice Guide No. 4: Discrete Modelling and Experimental Data Analysis*, version 2.0, 2004. http://www.npl.co.uk/ssfm/download/bpg.html#ssfmbpg4.