CMSC/M/05/608

XGENLINE Version 8.1 Stand-Alone Executable - Software Documentation

By
I M Smith

Issued 20 May 2005

Version 1.0

Unrestricted

Division of Enabling Metrology
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:	XGENLINE
Document reference:	CMSC/M/05/608
NPL File reference:	CMSC//09/03/M02
Path and filename:	gasstnd\xgenline\software\docs\sw_doc_xgenlinev8_1_ext.doc
Approved by:	BJC

National Physical Laboratory

Hampton Road, Teddington, Middlesex, TW11 0LW

1 INTRODUCTION

This document describes how to install and run a stand-alone Windows (NT/2000/XP) executable to carry out ordinary least-squares and generalised distance regression polynomial fitting and evaluation.

The underlying software is written in Matlab (version 7.0.4) and the Matlab Compiler (version 4.2) has been used to create the stand-alone executable.

The software consists of the following four files (with file sizes in brackets):

- XGENLINEV8_1.EXE (11KB),
- XGENLINEV8_1.CTF (59KB),
- MCRINSTALLER.EXE (104,588KB),
- NPL_LOGO_BLACK_LARGE.JPG (7KB).

The document is organised as follows. Section 2 lists the steps required to install the software on a PC running Windows NT, 2000 or XP while Section 3 describes how to run XGENLINEV8_1, detailing the required format of input data files and showing some example results. Section 4 describes how to update the environment variable PATH for each of the Windows operating systems.

2 INSTALLATION

Installation steps are as follows:

1. Copy MCRINSTALLER.EXE to a local folder then double-click on the MCRINSTALLER.EXE icon to commence installation of the Matlab Component Runtime. You will be prompted to enter the directory (which will be denoted in this document by <MCR>) into which the libraries should be placed. [After installation you may delete the copy of MCRINSTALLER.EXE as well as the files InstMsiW.Exe, Setup.ini and MCRInstaller.msi that have been created in the local folder.]

2. Add the directory <MCR>\v72\runtime\win32 to the environment variable PATH (See Section 4 for further details). If you are installing the Matlab Component Runtime on a machine that already has an installation of Matlab, the PATH variable needs to be adjusted as follows:

   - To run the deployed component, the Matlab Component Runtime run-time directory must appear prior to the Matlab run-time directory on the path.
   - To run Matlab, the Matlab run-time directory must appear prior to the Matlab Component Runtime run-time directory.

   Note that <MCR>\v72\runtime\win32 may be added to the PATH variable automatically during the first installation step, but not necessarily before any Matlab directory already present.

Note that if the Matlab Component Runtime has previously been installed, then there is no need to reinstall it as long as the PATH variable is configured as above.

3 RUNNING XGENLINEV8_1

To run the graphical user-interface, either:

A. Double-click on the XGENLINEV8_1.EXE icon, or

B. Open a DOS window, change directory to that in which XGENLINEV8_1.EXE is located and type "XGENLINEV8_1".

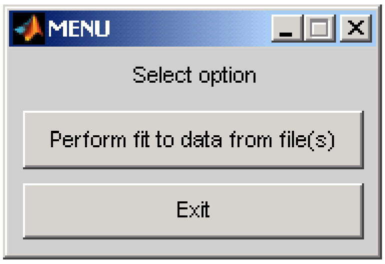
Note that option B should be used when you are fitting a straight line and want to be able to cut and paste some or all of the estimates of the straight line parameters and their associated uncertainties and covariance.

XGENLINEV8_1 performs the following operations:

1. Displays (in DOS window) header information identifying the software.

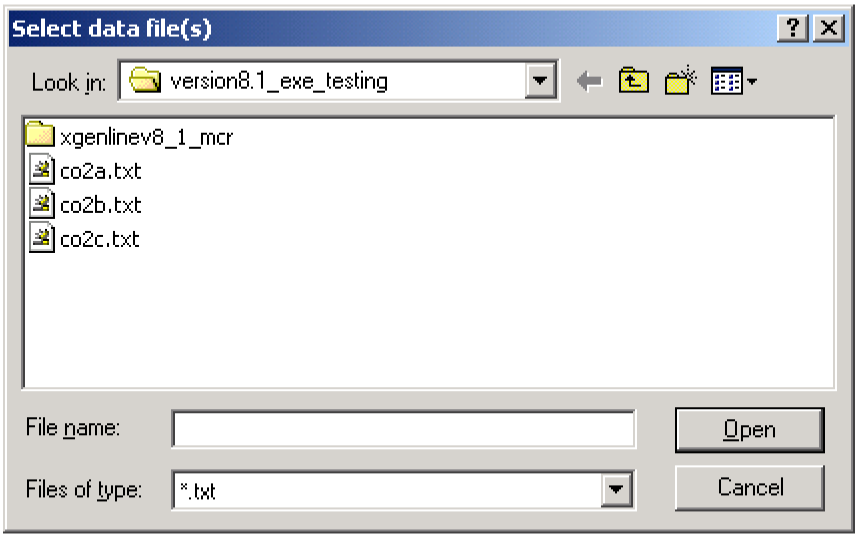
2. Displays (in DOS window) the current date and time.

3. Prompts the user to select "Perform fit to data from file" or "Exit".

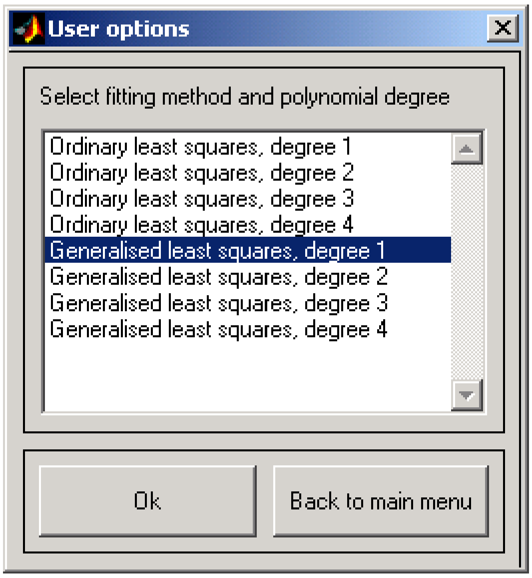


4. If "Exit" is selected, execution is terminated.

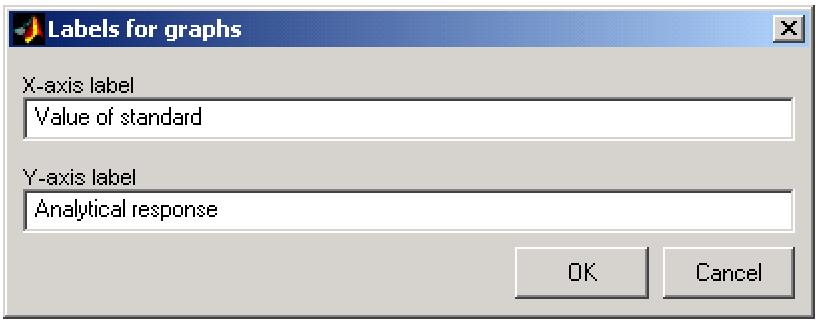
5. If "Perform fit to data from file(s)" is selected, prompts the user to select one or more input data files (see Section 3.1 for details of the required format of data files). A window opens listing all files with extension ".txt" in the current directory. The user may change, in the usual way, the type of files listed, and the directory. [Note that a folder (xgenlinev8_1_mcr) is created automatically upon first running of the executable.]



6. Prompts the user to select the fitting method (either ordinary least squares or generalised least squares) and polynomial degree *n* (1, 2, 3 or 4). The default option is "Generalised least squares, degree 1".



7. Prompts the user to select labels for the control (x) and response (y) variables. The default options are "Value of standard" for the x-axis label and "Analytical response" for the y-axis label.



8. If ordinary least squares polynomial fitting is selected, displays (in DOS window):

8.1. The name of the file being processed.

8.2. The weighted residuals associated with the measured y-data and the fitted polynomial, i.e., the values

$$e_i = \frac{1}{v_i}\left(y_i - p(x_i, \mathbf{a})\right), \quad i = 1, \ldots, m.$$

8.3. The remaining sum of squares deviation associated with the fitted polynomial, i.e., the value

$$\sum_{i=1}^{m} e_i^2 .$$

8.4. A goodness of fit measure associated with the fitted polynomial, i.e., the value

$$\max_{i=1,\ldots,m} |e_i|.$$

8.5. The root-mean square residual associated with the fitted polynomial, i.e., the value

$$\sqrt{\frac{\sum_{i=1}^{m} e_i^2}{m-(n+1)}}.$$

8.6. If the polynomial degree chosen is 1, estimates of the gradient and intercept and their associated uncertainties and covariance.

8.7. A new figure window containing:

- A graph of the input data, the fitted polynomial and the position(s) on the fitted polynomial of the predicted measurement point(s). If the polynomial degree chosen is 1, then the following information is displayed:

- The gradient l_1 of the straight line,
- The uncertainty $u(l_1)$ associated with l_1,
- The intercept l_2 of the straight line with the *y*-axis,
- The uncertainty $u(l_2)$ associated with l_2,
- The covariance $u(l_1, l_2)$ associated with l_1 and l_2,
- If the interval $[l_1 - 3u(l_1), l_1 + 3u(l_1)]$ does not contain 0, the intercept l_3 of the straight line with the *x*-axis, and
- If the interval $[l_1 - 3u(l_1), l_1 + 3u(l_1)]$ does not contain 0, the uncertainty $u(l_3)$ associated with l_3.

- A graph of the *weighted* residuals associated with the measured *y*-data against *x*-values.
- Text listing the input data file, polynomial degree, fitting method, the goodness of fit measure (see 8.4), and the (up to a maximum of 30) predicted point(s) with their associated uncertainty(ies).

8.8. The name of the output file containing information relating to the predicted measurement point(s). This will take the form "*file*_o*n*.txt" in the case that the selected data file is "*file*.ext" and the selected degree of polynomial fit is *n*.

8.9. If the polynomial degree chosen is 1, the name of the output file containing weighted and unweighted residuals. This will take the form "*file*_o1_res.txt" in the case that the selected data file is "*file*.ext".

9. If generalised least squares polynomial fitting is selected, displays (in DOS window):

9.1. The name of the file being processed.

9.2. A message indicating whether the Gauss-Newton algorithm has converged to a local solution.

9.3. The weighted residuals associated with the measured *y*-data and measured *x*-data.

9.4. The remaining sum of squares deviation associated with the fitted polynomial, i.e., the sum of squares of the weighted residuals computed in 9.3.

9.5. A goodness of fit measure associated with the fitted polynomial, i.e., the absolute value of the weighted residual of largest absolute value.

9.6. The root-mean square residual associated with the fitted polynomial.

9.7. If the polynomial degree chosen is 1, estimates of the gradient and intercept and their associated uncertainties and covariance.

9.8. A new figure window containing:

- A graph of the input data, the fitted polynomial and the position(s) on the fitted polynomial of the predicted measurement point(s). If the polynomial degree chosen is 1, then the following information is displayed:
    - The gradient l_1 of the straight line,
    - The uncertainty $u(l_1)$ associated with l_1,
    - The intercept l_2 of the straight line with the *y*-axis,
    - The uncertainty $u(l_2)$ associated with l_2,
    - The covariance $u(l_1, l_2)$ associated with l_1 and l_2,
    - If the interval $[l_1 - 3u(l_1), l_1 + 3u(l_1)]$ does not contain 0, the intercept l_3 of the straight line with the *x*-axis, and
    - If the interval $[l_1 - 3u(l_1), l_1 + 3u(l_1)]$ does not contain 0, the uncertainty $u(l_3)$ associated with l_3.
- A graph of the *weighted* residuals associated with the measured *y*-data against *x*-values.
- A graph of the *weighted* residuals associated with the measured *x*-data against *y*-values.
- Text listing the input data file, polynomial degree, fitting method, the goodness of fit measure (see 9.5), and the (up to a maximum of 30) predicted point(s) with their associated uncertainty(ies).

9.9. The name of the output file containing information relating to the predicted measurement point(s). This will take the form "*file*_g*d*.txt" in the case that the selected data file is "*file*.ext" and the selected degree of polynomial fit is *d*.

9.10. If the polynomial degree chosen is 1, the name of the output file containing weighted and unweighted residuals. This will take the form "*file*_g1_res.txt" in the case that the selected data file is "*file*.ext".

10. Return to Step 3.

A figure may be printed and/or saved in any of a number of different formats by clicking on "File" and selecting the required options.

3.1 Sample data

A data file is normally laid out as follows:

x_1	u_1	y_1	v_1	c_1	
x_2	u_2	y_2	v_2	c_2	
…	…	…	…	…	
x_m	u_m	y_m	v_m	c_m	
Unknown		Unknown		$y_{0,1}$	$v_{0,1}$
Unknown		Unknown		$y_{0,2}$	$v_{0,2}$
Unknown		Unknown		$y_{0,3}$	$v_{0,3}$

Here, x_i is a measurement of a control variable and u_i is its associated standard uncertainty, and y_i is a measurement of the corresponding response variable and v_i is its associated standard uncertainty. The value c_i is the covariance between x_i and y_i, and the values c_i, $i = 1, \ldots, m$, may be omitted from the data file in the case where there is no correlation between x_i and y_i. The values $y_{0,j}$ are measured response variables with $v_{0,j}$ their associated standard uncertainties. The values $x_{0,j}$ and $u_{0,j}$, i.e., corresponding control variables and associated standard uncertainties, are to be determined and are represented in the data file by the keyword "Unknown". Any number of new samples can be considered but note that a maximum of 30 predicted points and their associated uncertainties can be displayed on the figure.

Note that *for straight line fitting only* ($n = 1$), the measurements $y_{0,j}$ are permitted to lie *anywhere* outside the range of the measurements of the response variable.

The following shows an example data file, called CO2A.TXT:

```
0.046423949	1.39275E-05	2666.761189	48.1197939
0.341410326	8.85762E-05	22250.65976	98.12421813
1.077653296	0.000265142	71348.97126	307.686787
1.932620989	0.000436213	128738.1555	224.4903343
3.490923473	0.000696884	232914.0801	146.3498099
7.027159247	0.001206868	470579.6974	829.9047798
4.92503803	0.000718402	330205.7472	787.17142
Unknown	Unknown	109627.37963	121.25630
```

In the above file, the items of data are separated by "tabs". Spaces may be inserted to align the columns and make the presentation of the data more user-friendly. Thus, the following data file, called CO2B.TXT, is also acceptable:

```
0.046423949    1.39275E-05    2666.761189    48.1197939
0.341410326    8.85762E-05    22250.65976    98.12421813
1.077653296    0.000265142    71348.97126    307.686787
1.932620989    0.000436213    128738.1555    224.4903343
3.490923473    0.000696884    232914.0801    146.3498099
7.027159247    0.001206868    470579.6974    829.9047798
4.92503803     0.000718402    330205.7472    787.17142
Unknown        Unknown        109627.37963   121.25630
```

Finally, data will be read until an "end-of-file" character or an empty line is encountered. Consequently, text may be added at the end of a data file provided it is separated from the data by (any number of) empty lines. Thus, the following data file, called CO2C.TXT, is also acceptable:

0.046423949	1.39275E-05	2666.761189	48.1197939
0.341410326	8.85762E-05	22250.65976	98.12421813
1.077653296	0.000265142	71348.97126	307.686787
1.932620989	0.000436213	128738.1555	224.4903343
3.490923473	0.000696884	232914.0801	146.3498099
7.027159247	0.001206868	470579.6974	829.9047798
4.92503803	0.000718402	330205.7472	787.17142
Unknown	Unknown	109627.37963	121.25630

Created 04-01-2001

If data is held within an Excel workbook, a data file may be created in the following way:

1. Copy the data to its own worksheet.

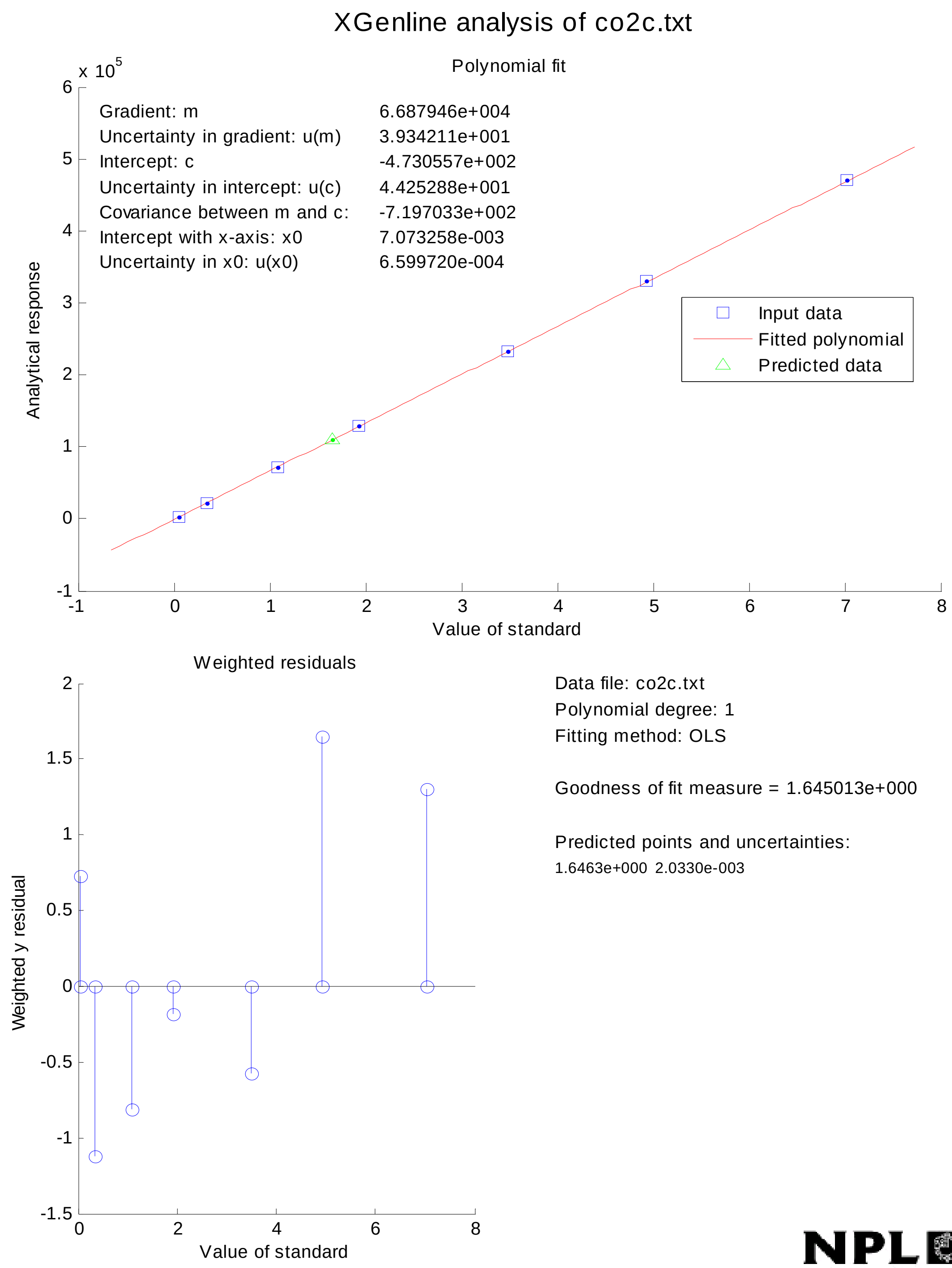
2. Save the worksheet as a “Text (tab delimited)” file.

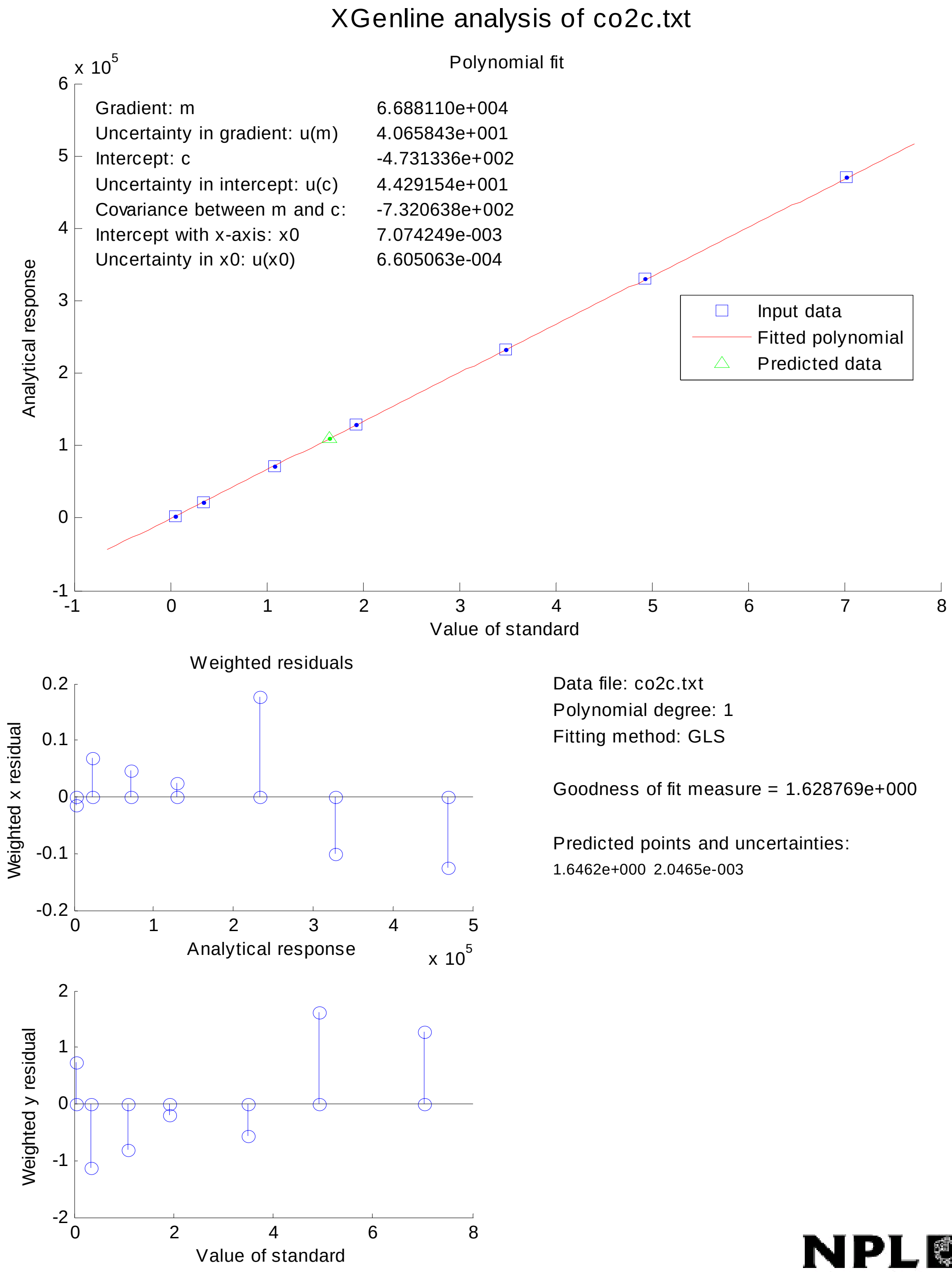
3.2 Sample results

Using XGENLINEV8_1 to obtain first degree, i.e., straight-line, polynomial fits to the sample data file CO2C.TXT listed above, the following screen and graphical output is obtained:

```
----------------------------------------------------
XGENLINEV8_1.M

Software for ordinary and generalised least-squares
polynomial fitting to data.

Author       I M Smith and P M Harris, MSCG
Version      8.1
Last amended 2005-05-20
----------------------------------------------------
Run on  2005: 5:20   at       11.30



Processing file ... co2c.txt

Ordinary Least-Squares Polynomial Fit

Weighted residuals (y):
    0.7275
   -1.1172
   -0.8153
   -0.1846
   -0.5736
    1.3015
    1.6450


Remaining sum-of-squares deviations = 7.20515

Goodness of fit measure = 1.64501

Root-mean-square residual, r0 = 1.20043

Gradient: m                     6.687946e+004
```

```
Uncertainty in gradient: u(m)    3.934211e+001
Intercept: c                     -4.730557e+002
Uncertainty in intercept: u(c)   4.425288e+001
Covariance between m and c:      -7.197033e+002

Saving fitted data to ... co2c_o1.txt

Saving residuals to ... co2c_o1_res.txt



Processing file ... co2c.txt

Generalised Least-Squares Polynomial Fit

Gauss-Newton algorithm has converged

Weighted residuals (y, x):
    0.7273   -0.0141
   -1.1180    0.0675
   -0.8180    0.0471
   -0.1951    0.0254
   -0.5559    0.1770
    1.2756   -0.1241
    1.6288   -0.0994


Remaining sum-of-squares deviations = 7.13955

Goodness of fit measure = 1.62877

Root-mean-square residual, rg = 1.19495

Gradient: m                      6.688110e+004
Uncertainty in gradient: u(m)    4.065843e+001
Intercept: c                     -4.731336e+002
Uncertainty in intercept: u(c)   4.429154e+001
Covariance between m and c:      -7.320638e+002

Saving fitted data to ... co2c_g1.txt

Saving residuals to ... c02c_g1_res.txt
```


XGenline analysis of co2c.txt
Polynomial fit
x 10^5
Gradient: m 6.687946e+004
Uncertainty in gradient: u(m) 3.934211e+001
Intercept: c -4.730557e+002
Uncertainty in intercept: u(c) 4.425288e+001
Covariance between m and c: -7.197033e+002
Intercept with x-axis: x0 7.073258e-003
Uncertainty in x0: u(x0) 6.599720e-004
Input data
Fitted polynomial
Predicted data
Analytical response
Value of standard
Weighted residuals
Weighted y residual
Value of standard
Data file: co2c.txt
Polynomial degree: 1
Fitting method: OLS
Goodness of fit measure = 1.645013e+000
Predicted points and uncertainties:
1.6463e+000 2.0330e-003


XGenline analysis of co2c.txt
Polynomial fit
x 10^5
Gradient: m 6.688110e+004
Uncertainty in gradient: u(m) 4.065843e+001
Intercept: c -4.731336e+002
Uncertainty in intercept: u(c) 4.429154e+001
Covariance between m and c: -7.320638e+002
Intercept with x-axis: x0 7.074249e-003
Uncertainty in x0: u(x0) 6.605063e-004
Input data
Fitted polynomial
Predicted data
Analytical response
Value of standard
Weighted residuals
Weighted x residual
Analytical response
x 10^5
Weighted y residual
Value of standard
Data file: co2c.txt
Polynomial degree: 1
Fitting method: GLS
Goodness of fit measure = 1.628769e+000
Predicted points and uncertainties:
1.6462e+000 2.0465e-003

4 UPDATING THE ENVIRONMENT VARIABLE "PATH"

4.1 Windows NT

1. Right-click on "My Computer" and select "Properties".

2. Click on the "Environment" tab.

3. Under "System Variables", select "Path". The current list of directories will be displayed under "Value" and will take the form

   DIR1;DIR2;DIR3;DIR4

   and the new directory (DIR5, say) can be added by appending ";DIR5" to the above list. If DIR5 needs to be placed before, say, DIR3, then insert ";DIR5" immediately before ";DIR3;DIR4". (Note that there should be no spaces either within the directory names or before or after the semi-colon.)

4. Click "Set" followed by "OK".

5. Restart the computer.

4.2 Windows 2000 and Windows XP

1. Right-click on "My Computer" and select "Properties".

2. Click on the "Advanced" tab then select "Environment Variables…".

3. Under "System variables", select "Path" and click "Edit…". The current list of directories will be displayed under "Variable value" and will take the form

   DIR1;DIR2;DIR3;DIR4

   and the new directory (DIR5, say) can be added by appending ";DIR5" to the above list. If DIR5 needs to be placed before, say, DIR3, then insert ";DIR5" immediately before ";DIR3;DIR4". (Note that there should be no spaces either within the directory names or before or after the semi-colon.)

4. Click "OK". Click "OK". Click "OK".