RAL REPORT 89-129, THIS REVISION 16/ 11/ 05

# THE "FISH" REFERENCE MANUAL
# (DATA FITTING PROGRAM FOR SMALL-ANGLE DIFFRACTION etc.)

Dr. R.K.Heenan
ISIS Facility, Building R3,
Rutherford Appleton Laboratory
Chilton,
Didcot, OX11 0QA, OXON., England

e-mail: R.K.Heenan@RL.AC.UK
Tel.: 01235-446744
Fax: 01235-445720



**CONTENTS**

## INTRODUCTION

This report describes the data fitting program FISH, written by R.K.Heenan. The program is best suited to the fitting of a curve with a relatively large number of data points by a model with a few parameters. There are presently some eighty available models, mostly concerned with the fitting of small angle diffraction data, though others such as polynomial fitting and peak fitting are of more general use, as are the graphics routines.

The mathematical model used to fit the data and the values of the parameters in a model may be changed interactively by the user. This is in contrast to most "least squares packages" where a model routine has to be compiled and linked by the user with a standard library. Complete model descriptions or previously saved sets of parameter values may be recalled from a file at any time. Any set of results may be written out to file. All actions of the user are recorded in a monitor file, which may later be sent to a printer for closer examination. Values of parameters may be fixed, tied together or constrained as the iterative fit proceeds.

A simple graphics interface allows plots of the results. The FISH program is designed to be easily portable from one computer to another. The FISH2 program first released in Oct. 1999 runs under VMS, Windows and Unix, using PGPLOT graphics routines. (FISH3 with a graphical user interface in IDL is under development for VMS and Windows.)

A typical run of the program uses READ or READ3 open a data file, the user selecting a workspace ( data set or memory) numbered 1 to 24, to store the data. The command FIT then enters the set up routine for data fitting. A "model file" is read containing previously set up model descriptions, one of which is selected for use. Data set numbers are then chosen for the observed, calculated , background and other necessary data stores ( such as polydispersity, structure factor etc.). The interactive FIT routine itself may then be started. The RUN command causes a single cycle of least squares fitting to be done. PP for Print Parameters will then display the results. After STOP to leave the interactive routine and returning to the main program a LIST command may be used to save calculated data sets in a file. Detail information on each of these stages and on the format of the various files is given below. An example interactive session is included as Appendix D.


## ACKNOWLEDGEMENTS

A number of contributors are thanked for allowing inclusion of their own routines or for making modifications or additions to some parts of the code. These include C.Fagotti, J.Penfold, J.Hayter, A.Mackie, A.North, W.S.Howells and H.Stanley. Major contributions are acknowledged in the FORTRAN code and in the appropriate parts of this manual. References to formulae in the scientific literature are given where relevant. These should be consulted for the limitations and interpretation of the theories applied.


## FITTING METHODS

Fitting is by a standard iterative linear least squares method, involving computation of first derivatives of each calculated data point with respect to each parameter in the model. Derivatives are calculated analytically in easier cases, else from the result of a small numerical shift in one direction ( the user may adjust the size of these shifts). Convergence may often be improved by applying less than the computed least squares shifts. This may be done via the "partial shift" associated with each parameter, the value of which in the program also determines whether a parameter is refining or not (ON or OFF) or whether it is tied to another parameter ( > 0, = 0, or -1 respectively). Calculated shifts are often too large due to ignoring higher derivatives. If not "damped" down in some way dramatic oscillations in parameter values may occur and the fit then

only converges if just one or two parameters are adjusted at a time. The rate of convergence is often best when multiplying calculated shifts by a partial shift of about 0.4.

A variation of damping procedure is offered by the "Marquardt method" in which diagonal elements of the least squares matrix are multiplied by (1+l), where l starts large and is recomputed at each cycle to obtain optimum convergence. ( Consult standard texts on fitting procedures ). This has the effect of forcing the fit along the line of steepest descent when it is a long way from convergence, gradually moving back to the more tortuous least squares route, which is at right angles to this direction, as l decreases. In this method the goodness of fit is forced to always improve but there is no guarantee of finding even a local minimum. Runs with test data shows that the Marquardt method may not converge to the true least squares minimum where that minimum is not well defined (which is usually why one tries to use it). A parameter search and/ or different starting points should be used to give some idea of parameter correlation. Large values in the least squares correlation matrix itself (use command CC to see this ) will also indicate a poorly determined fit. Note that when using the Marquardt method the estimated standard deviations of the parameters become unreasonably small - run one cycle of normal least squares before recording their values.

( To use the Marquardt method enter K2=1 and initialise CON(1) for l by entering N1= 1; to return to least squares enter K2=0. )

Some more detailed notes on least squares and Marquardt fitting are given in Appendix E.

If these methods fail then "predicate observations" may provide an alternative to simply leaving some parameters at fixed values. An expected value of a parameter is included in the fit as a data item with a carefully assigned weight. This is explained further under "MODEL 5" below, and has proven particularly useful in the field of molecular structure determination where a bond length may be allowed to "float" about its chemically expected value.

Data points to be fitted may be given unit weights ( use K1=0 ) or weighted as 1/ $E^2$ (K1=1) or as 1/ data ( K1=2, for where E = sqrt(data) ), where errors E are read in with the data. If you have no errors, and the usual K1=1 is present in the most model file, then FISH will complain about “zero weights”, you will then have to set either K1=0 or K1=1. K1 is the first control parameter on the second line of the model file (more later or see LSQ model file format on p43).

Some experiments have been made with a maximum entropy weights scheme for particle size distributions, this requires further development.

**SOME PROGRAM PHILOSOPHY**

The main sections of the program are each controlled by a simple command language; if you are stuck simply type HELP or H for a list of relevant commands.

Data are stored in "sets" which are given a number (1-9), these are similar in concept to ( but considerably predate !) "workspaces" in the ISIS GENIE program. Arithmetic manipulation of these sets is catered for, however more complex operations are best done elsewhere such as in specific raw data reduction codes or in the GENIE program. For those not familiar with GENIE these workspaces are simply data storage arrays. The model description used to calculate each point in such an array is analogous to a list of sequential instructions given to a programmable calculator. However FISH does pre-search the model description for special cases, such as polydisperse particles, which require initialisation.

Since the initial use was to process small angle X-ray data from a linear position sensitve detector all data sets may be stored with "left" and "right" sides with Q or radius respectively descending and ascending in value. Raw PSD data may then be "centred" and adjusted before binning into Q. The data files allow for NCH data points, using points NC1 to NC2 on the "left" and NC3 to

NC4 on the "right". Thus poor data at the ends of the usual ascending Q range may simply be ignored, but kept in the file, by setting NC3 >1 and NC4 < NCH ( see the DATAFILE definition section).

More detail comments on programming styles are made in the section "Making changes to the program".

## HOW TO GET STARTED - the FISHPREF.TXT file

A file FISHPREF.TXT will be needed in your working directory, as illustrated below. A typical model file LSINP.DAT and an example of FISHPREF.TXT and distributed with the program download. (FISHPREF.TXT provides a simple way to keep the operation of FISH2 the same on any platform, by avoiding the use of system variables.)

```
! thsi FISHPREF.TXT file needs to be in your working directory
! lines with exclamation mark are comments and are ignored
! Need system type VAX (for vms) or WIN (for Win95 or WNT) or LNX (for any linux and unix)
WIN
! Need "source directory" for FISH's command definition files
! ( which are called FCOMMAIN.TXT, FCOMPLT.TXT and FHELPFIT.TXT )
C:\ FISH2\
!
! then the directory for logging file FISHLOG.LIS and
! graphics plot files ( scratch area on vax)
C:\ FISH2\ WORK\
!
! and finally the directory and name of the LSQFILE containing models
! file ( you can call it any name you choose.)
C:\ FISH2\ WORK\ lsinp.dat
! Inside FISH use SET to switch to another file.
```

Run the executable, then use READ *filename*, or READ3 *filename* command in the main routine to get the input data, then try FIT. Follow through the items 1 to 3 in the FIT control section menu, to decide which set to fit and which sets will store the calculated data, etc. Then proceed into the FIT interactive commands section. Study the example session at the end of this manual. Use of an existing LSQFILE model file (usually stored as LSINP.DAT) will be a great help as actual examples are easiest to follow.

## FORTRAN FILES TO BE ASSIGNED

There are several files required by or generated by FISH, these can be changed from inside the program using SETUP in the main program. They are summarised in the table below. Files may be called by any appropriate name, but will be referred to by the names in the table below throughout this manual.

| Fortran stream number | File name in this manual | Purpose | Present default name |
|---|---|---|---|
| INPUT | | | |
| | FISHPREF.TXT | Defines important directories and initial name of LSQFILE | FISHPREF.TXT |

| | | | |
|---|---|---|---|
| 1 | DATAFILE | Actual data e.g. $I_{obs}(Q)$ | nnnn.Q |
| 3 | LSQFILE | Least squares model description(s) | LSINP.DAT |
| OUTPUT | | | |
| 2 | MONITOR | To record what you did | MON.LIS |
| 4 | NEWDATA | Modified or calculated data. FISH will prompt for a file name. | FOR004.DAT |
| 8 | NEWLSQ | Modified or new model descriptions e.g. to save latest parameter values. FISH will prompt for a file name. | FOR008.DAT |
| graphics | | graphical output, default postscript | FISHPLOT.PS |

In addition there are three text files which define the command names and minimum abbreviations for the main program, the PLOT routine and the interactive FIT routine. NOTE that the MON.LIS file can become quite large, you may need to purge or delete it regularly.

## COMMANDS IN THE MAIN CONTROL ROUTINE

These are the commands available on first entering FISH. Most people will only need to use READ, FIT, LIST, PLOT, INDEX, STOP and possibly RANGE. The minimum numbers of characters for an acceptable abbreviation are underlined. Input strings are converted internally to uppercase, up to the first space (so that case sensitive file names in Unix are preserved). The Nov. 2000 version allows spaces in filenames (for Windows).

HELP

Lists the entire detailed help file.

H

Lists command names only.

READ-DATA *[filename]*

for ASCII (normal character) files in standard DATAFILE format. Uses DATIN routine and channel 1. Filename is optional, if missing the file previously set up will be used, as files may contain many sets of data.

READ3 or R3 *filename*

for ASCII (normal character) 3 column Q, Intensity, Error file with a single data set in Fortran free format.

READ2 or R2 *filename*

for ASCII (normal character) 2 column Q, Intensity (without Errors) file with a single data set in Fortran free format. If fitting you may need to use switch K1=2 or K1=1 to set leass squares weights.

RNILS *filename*

for standard ILL data file with Q, Intensity, Error.

OTOKO *filename*

attempts to read OTOKO binary format Xray file, NOT quite working ?, only works on WINDOWS, please consult RKH if you need to import such data.

GETWIR

for BINARY linear data input, old Daresbury SAXS format in file [RKH.DL]Snnnn.FV4, where nnnn is input run number.

LIST

to list data set to monitor file ( for printer) or to write out an ASCII data set to the NEWDATA file, in which case new title records will be asked for. ( The ouput ASCII file, default FOR004.DAT, may be read into a GENIE workspace by COLETTE command OLD.)

INDEX

show details of the sets you have stored.

ARITHMETIC

add, subtract, divide, scale, normalise data sets.

QBIN - for processing raw linear detector data.

First enters the RANGE routine, calculates Q values for set I and will rebin into set J, if J non-zero. Assumes that raw data channels are equi-spaced, will ask for camera distance, detector element spacing and incident X-ray or neutron wavelength. (Enter a negative Q BIN DELTA Q value to jump back to main routine if things go wrong.)

CENTRE - helps to find middle of Daresbury SAXS data set,

asks for pixel range L1,L2 to left of beam stop and a guess for mid-point. The mid-point is entered multiplied by 10, to give the nearest 0.1 division. Then it sums (N(R)-N(L) )**2 over R corresponding to L =L1,L2 where N(i) is the count per channel. A small range of mid-points is chosen, the"best" may be at the minimum sum. Use with care and plot results to test !!!

L-R

Put Left-Right differences for a Daresbury SAXS set into another data set, e.g. ready to plot. Results are placed on the left side for the full range over which left and right sides overlap. If data is in Q then a linear interpolation is done on the right side to force Q values to match the lfet side.

RANGE n

Allows removal of points at the ends of the range of data set n. e.g. to use points 10 to 70 in a set with normal ascending Q enter 0 0 0 10 70

PLOT

Enters longwinded, complicated, but very flexible plot routine. First you will need to set up a list of sets to plot, several can be placed on the same axes. If in doubt about optional parameters enter zeroes to get default action. PLOT has its own command language and help facility as detailed below.

FIT

Enters least squares model-fitting program via the FIT menu. This has its own interactive language and HELP command. A large number of models are available. See separate sections below. It is possible to fit more than one linear data set simultaneously with overlapping parameters.

SETUP

Allows new input or output files, so there is no need to exit the progam to switch to another data file. Enter zero to just list currently opened files.

STOP

Use this for a graceful exit from the program, don't forget to PRINT MON.LIS to the printer if the monitor might contain useful information. It should provide a record of what you did on the screen. Tidy up any output files, note that on a VAX their names default to FOR00n.DAT, where n is a channel number, if they were not assigned elsewhere.

QUIT

Same as STOP

## COMMANDS FOR THE PLOTTING ROUTINE

These commands are available after giving PLOT in the main program, note that a completely new command language is then in use. The input here is extremely tedious, however complicated overlaid plots may easily be built up using a variety of graph markers. To get into this routine use PLOT in the main routine. To generate a plot a recipe has to be set up, by READ, with a list of which work spaces are to be plotted and with which symbols. (i.e. you have to know in advance exactly which sets you want to plot as later overlays are not possible at present.) This recipe is preserved upon subsequent calls to PLOT (use LIST to see it), so if you wish to change any of the workspaces a repeat plot can be made without having to re-do a READ. A SCREEN or FILE command will then produce the plot to the appropriate device.

NOTE the PLOT command in the interactive FIT routine becomes PICT and PLOT then enters a high speed routine to show OBS, CALC and OBS-CALC for the present cycle of refinement. This does however redefine the plot recipe ! Skip the rest of this section if you are only interested in doing simple fits. You may however need to use this routine to plot say a polydispersity function or to change axis types (e.g. PICT, LOGLOG, STOP from FIT interactive commands to start Log(Y) vs. Log(x) plots).

HELP

Full help instructions.

H

Quick list, or H Command for information on any Command.

SYMBOL

lists standard line and symbol types

READ

Asks for a list of curves (sets) to be drawn, all on same axes with different symbols or line types for each. ( do SYMBOL for more information) You will be asked for:

(a) (I1) Number of curves (up to 9)

(b) for each curve:

(i) (5I1) LTYPE - see below, line type e.g. solid or symbols
LSYMBOL - see below, marker type e.g. cross or asterix
IFOLD =1 to fold about centre
IEB =n to add error bars of +-n*E(i)
I_NUM_NOT_Q=1 to plot against channel rather than Q
e.g. for + signs with error bars enter 5301

(ii) if data has left and right sides, (2I1) LTYPE, LSYM for the left side data
(iii) YSHIFT - added to Y values at plot time

line types LTYP are: 1 - straight line segments,
2 - smooth curve(local polynomials) 3 - dashed, 4 - dotted
5 - use symbols of type LSYM as specified by :
LSYM= 1 - up triangle, 2 - down triangle, 3 - + , 4 - X , 5 - square,
6 - diamond, 7 - circle, 8 - *

NOTE: Some symbol or line type options may not be implemented on particular computers.
For symbols on a solid curve, enter the set twice, once as solid, again with symbols. If in doubt about any values leave zero to get defaults.

LIST

Gives details on data sets chosen by READ

SCREEN

Draw on screen, the LOCAXIS routine will ask about the axis ranges - dependant upon local installation, then you will get the plot !

FILE

Plot to a file. (The default postscript type can at present only be changed from PLOT in the FIT routine).

INDEX

Usual master index of all sets in memory.

STOP

Returns to calling routine.

LINEAR

use Y against X values just as supplied.

LOG

use LOG(Y) against X, or use before other plot types to return to base 10 logs.

LOGLOG

use LOG(Y) against LOG(X) (or LN(Y) against LN(X) if first issue a LN command ).

LN

choose LN(Y) against X, or use before other plot types to switch to natural (base e) logarithms.

GUINIER

LOG(Y) against X**2, uses log to base 10 unless you have previously issued a LN command.

ZIMM

1/Y against X**2.

RODS

LOG(Y*X) against X**2 for thin rods.

SHEETS

LOG(Y*(X**2)) against X**2.

USER

choose your own transformations. These are of the form (X**i)(Y**j)LOG( (X**k)(Y**l) ) where i,j,k,l may be defined differently for both X and Y axes.

FIT n m

straight lines to be fit to sets n to m, will ask for X range for each set, where the units of X may be the originals, as transformed or as channel number. Will turn "off" the fit if one is already "on". Use LIST to see the gradient etc. but after first doing a SCREEN or FILE command to see the plot.

## FIT ROUTINE - MAIN MENU

This menu appears after issuing FIT in the main program, it guides the set up needed before entering the interactive fit routine. Normal usage involves first READing some experimental data in the main routine, then enter FIT, and work through options 1, 2 and 3 in this menu.

1 - Read model file.

This reads the LSQFILE of model descriptions, enter 1 when you reach the appropriate model.

2 - Choose OBS,CALC etc.

First asks how many data set you want to fit - normally just 1, for multi-data sets fits the model must be specifically designed with this in mind.

Follow the instructions given to choose workspace numbers to be used for the OBServed data, the CALCulated data, BKG for experimental background (model 3), POLY polydispersity function (model 6) and WRK to store scaled background (models 3 and/or 4) etc.

If in doubt enter 1 2 3 4 5 6 7 8 9 assuming your data is in workspace 1, but in theory leave a zero for a set not required e.g. 1 2 0 0 3 0 4 5. Missing a space or entering zero for a workspace that is actually required may result in the program stopping with a subscript out of range error. You may have to re-enter option 2 after replacing one model by another using option 1 if extra workspaces are required or if new data with a different number of points has been READ in.

For multiple data set models you may skip missing sets by entering -1 for OBS. e.g. if the model has core, shell & drop contrasts but you only want to input and fit core and drop, then say you have 3 sets to fit, but enter say 1, -1 and 2 in the OBS column.

3 - Enter fit routine

Goes into the interactive fitting routine, see the next section.

4 - Calculate only or set Q.

Used instead of option 2 if there is no experimental data and you wish to do a calculation only.

5 - Index

As usual, helps if you have forgotten which sets are in use.

6 - Return

Goes back to the main control routine.

31- Enters derivative test routine TESTER.

32- Enters derivative shift and integration scheme set up routine DELSET

## FIT ROUTINE COMMANDS - INTERACTIVE LEAST SQUARES FITTING

This routine is reached from the FIT routine main menu, option 3, above. A highly abbreviated command language is used here due to the large number of times each command is used. To start with try out the commands P, PP, RUN, n=r (e.g. 1=123.45 ), n=ON (e.g. 1=ON=0.5 ) , n=OFF.

Parameters are in array V(i), commands below may have ranges n,m where n and m are integers. If m is not present then m=n by default. Real number r does not need a decimal point. , but no E+0n allowed (yet).

| | |
|---|---|
| R or RUN | do one cycle of calculation. |
| ST or STOP or RE or RETURN | return to calling routine. |
| n m or n,m or n.m | prints values of V(n) to V(m) on screen. |
| n=r | sets value of V(n) to real number r |
| n,m=ON | sets status of V(n) to V(m) to ON by setting partial shifts to 1.0 |
| n,m=ON=r | additionally resets partial shift to value r (e.g. use 0.4) which multiplies calculated shifts, reduces oscillations in many cases. |
| n,m=OFF | sets status of V(n) to V(m) to OFF by setting partial shifts to 0.0 |
| ON or OFF | changes status of whole fit regardless of partial shifts (NOTE: OFF followed by ON zeroes all the calculated shifts, before they are applied, useful if something has gone wrong.) |
| n,m=TIE | turns back on constraints turned off by setting partial shift to -1.0. |
| P | Prints all V(i) parameter values on screen ( same as 1,99) |
| PP | Prints all Parameter information to screen, including titles, constraints etc. |
| PF | Print to File, lists all data and correlation matrix to MONITOR file. |
| FF | Fill File writes current model and parameter values to file NEWLSQ (default FOR008.DAT). |
| TT | Terminal Table, lists OBS,CALC data etc. to screen. |
| CC | prints least squares Correlation Coefficient matrix on screen. |
| PS or S | " data Sets in use |
| PC or C | " Constraint relations |
| PN or N | " Numerical constants, CON(1) is Marquardt lambda |
| PT or T or PK or K | prints control records and titles |
| Sn,m or Cn,m or Nn,m | prints a range of the above |
| Cn= or Tn= | will invite a replacement for a constraint or title |
| Pn,m= | " " " " " parameter records to change or construct a model, ( regret no insert instruction). |
| Nn=r | resets constant n, e.g. N1=1.0 for Marquardt l, N4 for Dr, N5 for $R_{max}$. |
| Kn=i | " control flag n, e.g. K2=1 for Marquardt method, K2=0 least squares, K5=1 to re-initialise numerical integration schemes. |
| IN or INDEX | usual index of data sets |
| PICT | enters normal long-winded plot routine e.g. PICT, LOGLOG, STOP to switch PLOT to log-log axes. |
| PLOT | special plot for least squares data -see note below. |
| H or HELP | prints a shortened version of this section on screen. |

Most sensible combinations of command and range are allowed.

**Notes on PLOT command inside FIT which rather cryptically asks:**

PLOT CONTROLS IDEV= 0-EXIT, 1-plot, 2-save file, 3-new screen, 4-new file type

error bars are IEB*sigma

IPW=1 adds scaled wts

IDEL spreads graphs apart (IDEL=2 is default)',

ITXT = 0 table in file, 1 no table, 2-both, 3-screen only',

IDEV,IEB,IPW,IDEL,ITXT = (5I1)

Note this expects a string up to 5 characters long, but giving just 11 will give IDEV=1, IEB=1 to get a plot with error bars on the screen.

Answering just 2 will give IDEV=2 to save the plot previously drawn to screen in a file, default postscript.

Answering 3 may be used to change the default screen type, or 4 to change the default hard copy file type (e.g. for / CGML to get a file that may be incorporated into WORD, or / GIF ). The latter two options also prompt for new values of the default line-width, character size and marker size so that plots may be further customised. The hard copy file also contains a table of parameters. Parameter ITXT can add this table to the screen version (e.g. if you need to make a screen dump) or remove it from the hardcopy by using say 11002 or 11001 respectively.

The upper trace in the plot is the observed data ( + markers with added error bars if IEB=1), the lower trace is the difference OBS-CALC, with error bars (if IEB=1).

If you have asked to store P(Q), S(Q) or β(Q) you will be prompted to over plot these, enter 1 for solid line, 3 for dashed, 5n where n=1 to 6 for different marker types. S(Q) and β(Q) will be rescaled to suit the plot axes.

## ABSOLUTE INTENSITIES

The "scale" parameters for each model have a physical significance when the input data is in absolute units. The formulae supplied should enable them to be understood. Unfortunately is was not convenient to use an entirely consistent basis for all of them. Some scale factors may not be in their most familiar form e.g. by having a particle volume included or removed. This may avoiding a divide by zero possibility at R=0 or reduce parameter correlation by making the scale constant independent of particle size. The polydisperse spheres distribution are all normalised to total volume, allowing change from one distribution to another with minimal parameter shifts and also the possibility of fixing the total volume fraction at a known value.

In general for neutron small angle scattering the probability of scattering per unit solid angle per unit thickness of sample, in $cm^{-1}$, is

$$d\Sigma(Q)/ d\Omega = N(\Delta\rho)^2V^2P(Q)$$

where there are N particles $cm^{-3}$ ( typically $10^{16}$) of volume V and particle form factor P(Q) is normalised such that P(Q=0)=1.0. $\Delta\rho$ is the scattering length density difference between two phases for which

$$\rho = \Sigma n_i b_i \,.\, \rho_{bulk} \,.\, N_A / MW \quad \text{where } b_i \text{ is a scattering length in cm.}$$

e.g. for $D_2O$ $\rho = ( 2*0.667 + 0.580)x10^{-12}cm \times 1.1g.cm^{-3} \times 6.02x10^{23}mol^{-1} / 18g.mol^{-1}$

$= 6.4 \times 10^{10}cm^{-2}$

If $\Delta\rho$ is entered in units of $10^{10}cm^{-2}$ then the scale factors given below for each model are multiplied by $10^{20}$ and they then generally all work without underflow or overflow occuring in the FORTRAN programs. Molecular dimensions R are assumed in Å (= $10^{-10}$m = 0.1nm ) and scattering vector Q in $Å^{-1}$ throughout.

For X-ray scattering $\rho$ may be most easily expressed in electrons per unit volume, calculated from atomic numbers and molar volumes. The probability of scattering by one electron then needs to be included,

$I/ Io = e^4( 2 -2sin^2(2\theta) )/ (m^2c^4) \sim 7.94x10^{-26}cm^2$ at small angles. The actual "cross section" units for X-rays should be carefully checked for a given instrument, one may for example have to divide by the sample-detector distance squared in order to get counts per unit area detector per unit volume of sample or possibly have to allow for sample thickness or incident beam monitor somewhere.