AVERAGE V2.22

Users manual

A program to merge single-crystal diffraction intensity data, with rejection of outliers.

RJ ANGEL

rossangelsoftware@gmail.com

http://www.rossangel.com

CONTENTS

INTRODUCTION

AVERAGE is a program to merge single-crystal X-ray intensity, with options to reject outliers from sets of symmetry-equivalent data. It can be used for data collected with either an area detector or a point detector.

History

AVERAGE was originally written while the author was at University College London in 1991 for Fortran running under VAX-VMS, and it used the 1984 release of the Cambridge Crystallographic Subroutine Library (CCSL) by Judy Matthewman to perform some of the symmetry manipulations. In the late 1990's the code was ported to run in a DOS box under Windows. Major modifications were then made in 2002-2003 at Virginia Tech Crystallography Laboratory (VTX), including:

- New format for control file, compatible with WinIntegrStp integration program.
- All controls set in the control file.
- Addition of option to reject negative outliers caused by DAC diffraction (see Angel et al. 2000 for a discussion of data reduction for DAC data).
- Full handling of SHELX *hkl* files in addition to RFINE format files.

The resulting version of AVERAGE was designated 2.0 and was distributed as a "beta-test" version. The addition of a Windows dialogue box to run the program with the same underlying code as v2.0-2.1 was designated v2.2. The current version is 2.22 as of October 2011.

The AVERAGE program is distributed on a non-commercial basis and the author would appreciate its use being acknowledged by reference to the source at www.rossangel.com in any publications.

If you would like to receive program updates (including bug fixes), please register with me as a user by e-mail (rossangelsoftware@gmail.com). If you discover apparent bugs in the program, please send me the *input* files, the *output* file and a full description of the problem by e-mail. Other suggestions for improvements and modifications will also be considered. Further information will be posted on the web site www.rossangel.com.

Acknowledgements

Thanks are due to the late Judy Matthewman for introducing me to CCSL and letting me as a young student take away a complete version of the code of the subroutine library. Many users have provided significant feedback and undertaken testing of the program, especially (victims can insert your names here!) Thankyou!

Disclaimer

While I try to ensure that the AVERAGE software is free of bugs and errors, people use it at their own risk. I cannot accept any responsibility whatsoever for either incorrect results or for any physical, mental or other damage arising from use of AVERAGE or from errors in this manual.

1. MANUAL AND SOFTWARE REVISIONS

October 1991:

- Original code written for VAX-VMS Fortran.

1991-1998:

- Minor modifications to fix bugs.

1998:

- Ported to Lahey Fortran for PC.

2001:

- Ported to Compaq/DEC Fortran for PC.

Nov 2002- Oct 2003:

- Switch from reading *ccsl* control files to *exp* control files to enable continuity with other data reduction programs WinIntegrStp and Absorb.
- Introduced reading of SHELX *hkl* files.
- Introduced more controls for rejecting reflections and controlling output.
- Introduced rejection for DAC "dips".
- Switched to working with F^2 instead of F.
- Restructured program to prepare for GUI-driven version.
- Introduced *cif* writer.
- Introduced GUI to act as file browser.

May 2006:

- Added View utility to main dialogue.
- Fixed bug with asymmetric unit arising from *hkl* files terminated with 000.
- Fixed bug in coverage calculation that caused program to crash with array bounds exceeded.

2007:

- Introduced new control card AVERAG FRIEDEL NOFLIP to explicitly prevent any reflection *hkl* being flipped to its Friedel mate.
- Changed test on the det(G*) to allow for unit cells with volumes up to $10^{4.5}$Å^3 (in line with Absorb program).
- Introduced new keyword to AVERAG LESSTHAN card: REJECT results in less-thans not being written to the *output* data file (useful for SHELX-L). Default is to keep them (keyword KEEP).

2010:

- Added 'NOSINGLET' as an optional argument to the AVERAGE OUTPUT keyword. When set, this stops singlet reflections being written to the *output* data file.

2011:

- Version 2.22 – minor revisions of manual and GUI.

2. INSTALLATION

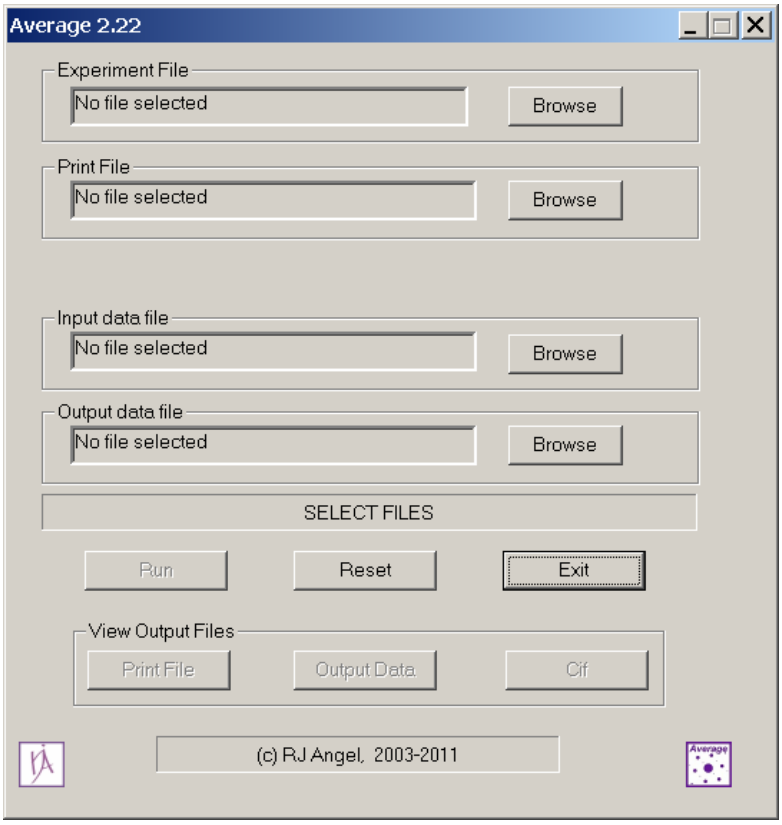
1. The average22.zip file contains the executable, a test datasets with *input* and *output* files, and this manual as a *pdf* file.
2. Unzip the average22.exe and the *pdf* file containing the manual into one folder. It is recommended that this folder is not used for data files.
3. Unzip the remainder of the files into a working directory. These contain example datasets and corresponding *experiment* files and *output* files.
4. Create a shortcut to average22.exe either in the working directory, or on the desktop, by right-clicking on the file, dragging to the working directory and then selecting "create shortcut".
5. Test the installation by running the program (see below) and using the example datasets and control files provided in the distribution.

3. RUNNING THE PROGRAM

The normal procedure is to first create an *experiment* file. This is an ASCII file which can be edited by any normal editor (e.g. Notepad, WordPad etc). It contains all of the information about the crystal symmetry and the controls for averaging. Details of the content of the *experiment* file are given later in this manual. For testing the program use the example files provided in the distribution.

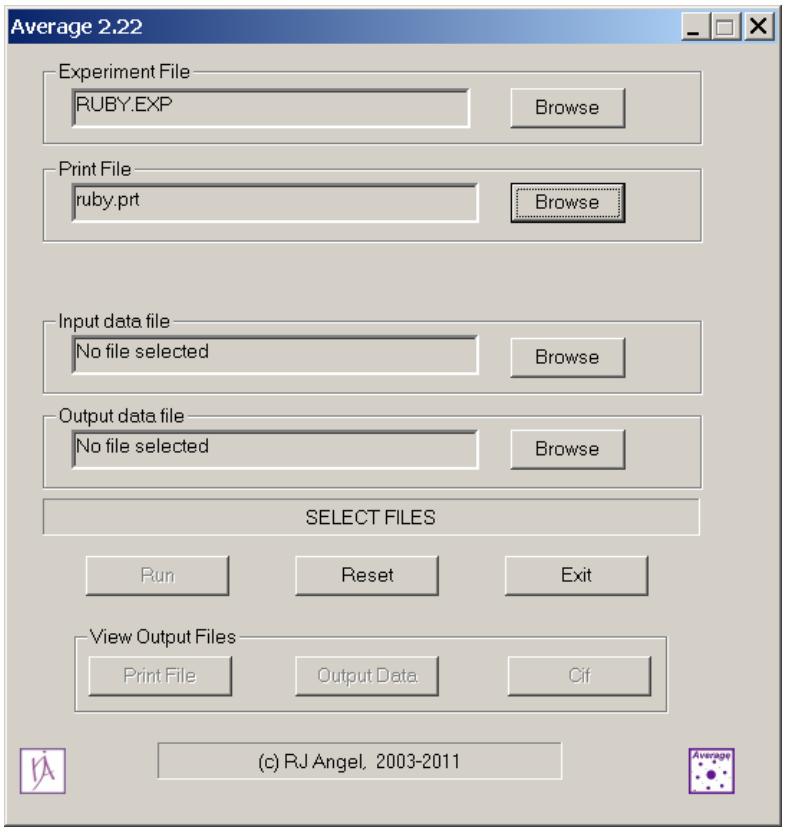
You also need the *input* data file containing the unmerged reflection data. For testing the program use the example files provided in the distribution (but read the next section *before* proceeding!).

Start the program by double-clicking on the shortcut. The main AVERAGE dialogue box will appear:



Some guidance as to what to do at each step is provided in the message window (that says "Select Files" in the picture).

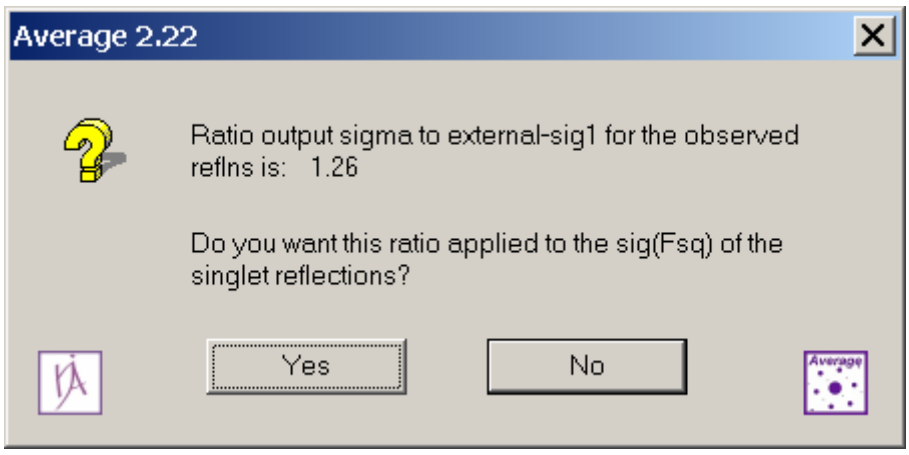
Use the *Browse* buttons to select an *experiment* file (must exist) and a *print* file. The *print* file will contain details of the calculations and results. It may exist – if it exists the new information will be appended to it. The file names will appear in the dialogue box:



Use the *Browse* button to select the *input* data file (must exist) containing the data that you wish to merge. The file extension *abs* is interpreted as a file produced by WinIntegrStp, the extension *hkl* is interpreted as a SHELX file. Other extensions can be used.

Use the *Browse* button to select the *output* data file to contain your merged data. The file extensions *avg* or *abs* are interpreted as a file to be produced in the RFINE format, the extension *hkl* is interpreted as a SHELX file. Other extensions can be used. If the file exists you will be asked if you want it to be over-written.

The *Run* button should now be enabled. Select it and AVERAGE will start processing your data, as indicated in the message window at the bottom. Once the data have been merged the program calculates the average ratio of $\sigma\left(\overline{F}^2\right)$ to $\sigma(external2) = \left[n/\sum 1/\sigma^2\left(F^2\right)\right]^{1/2}$. The former value is based upon the agreement of the symmetry-equivalent reflections and the latter is based upon the reported uncertainties of the individual reflections. Therefore the ratio represents a measure of how well the individual uncertainties, $\sigma(F^2)$, in the *input* data file represent the true reproducibility of the F^2. If the ratio is larger than 1, then it may be appropriate to apply this ratio to the $\sigma(F^2)$ of the singly-determined reflections. If you have requested that averaged reflections are written to the *output* data file, then you are given the choice in the following dialog that appears:

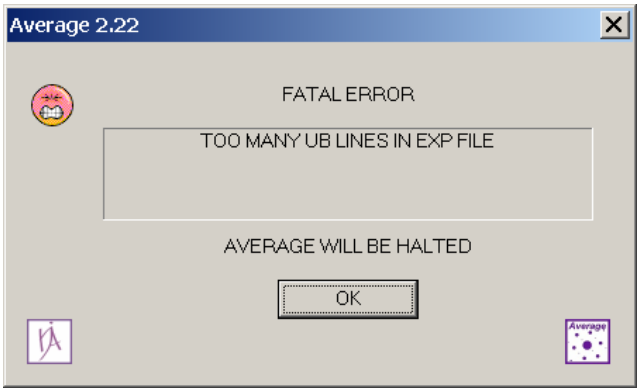


The issues involved in making a choice are:

- If the *Rint* and *Rsigma* values indicate a high-quality dataset, and the sigma-ratio is reasonably small (usually between 1 and 2) then it is recommended to choose *Yes*. This will then increase the $\sigma(F^2)$ of the singlet reflections so as to avoid them being over-weighted in the subsequent structure refinement.
- If the ratio is less than 1, then it is inappropriate to apply this ratio. Select *No.*
- If the ratio is larger than 2, then *Rint* is probably large and the data quality is probably low, so you are on your own!

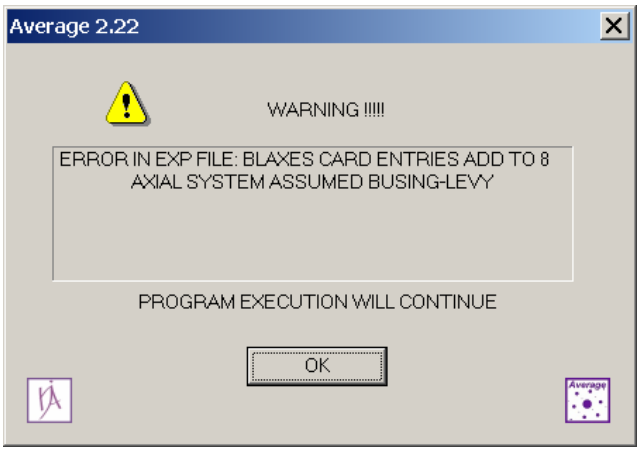
Once the data are processed, the program writes the information about the corrections to the *print* file and also to a *cif*, given a name based upon the name of the *output* data file, together with "*_average.cif*". A message that AVERAGE has run successfully will appear on the main dialog box.

If, however, there is an error in the program, one of two types of message box will appear. If the problem is sufficiently serious to prevent the program from executing, you will see the following error box, with an explanation of the problem:



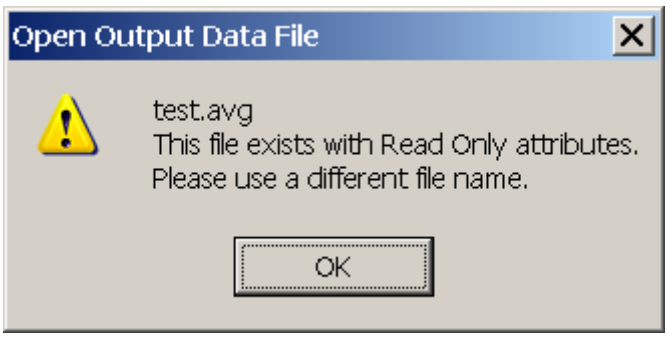
After you select *OK* you will be forced to exit the program because the error will prevent the program from processing your data. You will need to correct the error in your *experiment* file.

You may also see a warning box:



Warnings alert you to *potential* problems, usually with the *experiment* file, that may or may not be fatal to the program. In this case the BLAXES card in the *exp* card could not be interpreted, but since this is not critical for the merging process, this does not prevent program execution. After you select *OK* on the warning box you will be returned to the main dialogue box.

You will also receive a warning if the program cannot open the file that you selected. for example:



Common causes for this error are that the file is either already open and locked by another program, or that the file has been set to "read-only". When you are returned to the open file dialog, just select another file.

For both warnings and errors, the same information is printed to the *print* file.

At the end of execution you can either view the *output* files directly via the buttons on the main dialogue window, or you can open them separately with an editor such as Notepad.

4. TEST DATASETS

Run the program with the test datasets, creating *new* names for the *print* file and the *output* data file, e.g. *test1.prt* and *test1.abs*. Do not use *example1.prt* as this will overwrite the file provided in the distribution!

Carefully compare the output from your test run with the *.prt*, *.cif* and *.abs* or *.hkl* files provided in the distribution. Report any discrepancies to the author at rossangelsoftware@gmail.com. Please send copies of the files, a full description of the problem especially noting the *first discrepancy* from the example output, and details of your operating system.

1. Dataset collected under ambient conditions:

    a. There is one example input dataset from a standard ruby crystal provided in two formats, an RFINE *abs* file and a SHELX *hkl* file.

    b. There are two *exp* files that can be used with either *input* data file. *Laue.exp* demonstrates the symmetry as specified by Laue group, while *ruby.exp* explicitly specifies the symmetry. Other controls are also set differently to illustrate various options within the program.

    c. There are four sets of *output* files corresponding to the four possible combinations of *input* data files and *experiment* files:

        i. Example1: from *Laue.exp* and *Ruby.abs*

        ii. Example2: from *Laue.exp* and *Ruby.hkl*

        iii. Example3: from *Ruby.exp* and *Ruby.abs*

        iv. Example4: from *Ruby.exp* and *Ruby.hkl*

    d. The output from each example consists of a *print* file, an *output* data file in either *avg* or *hkl* format, and a *cif*.

2. Dataset from a diamond-anvil cell, *dacdemo.abs*, and associated *exp* and *output* files to illustrate the operation of the rejection of DAC dips.

5. PROGRAM DESCRIPTION

Execution

The overall sequence of program operations is as follows:

1. The contents of the *experiment* file are read:
    a. The entries on the cell and UB cards are checked for consistency.
    b. The metric and reciprocal metric tensors, and the B matrix are calculated.
    c. The symmetry information is checked. If the symmetry was specified by symmetry operations the full space group is built from them.
    d. The unit-cell and symmetry operations are written to the *print* file.
    e. A summary of the control parameters and instructions requested by the user in the *exp* file are written to the *print* file.

2. The data are read in from the *input* data file:
    a. If the *input* data file is an *abs* file, the F and $\sigma(F)$ are converted to F^2 and $\sigma(F^2)$.
    b. If the *input* data file is an *hkl* file, the 2θ value is calculated:
        i. from the UB matrix and direction cosines if present, or
        ii. from the unit-cell and the direction cosines if present, or
        iii. from the unit-cell and indices if the unit-cell or UB is present but the direction cosines are not.
    c. The indices of the reflection are transformed by each of the symmetry operators in turn until the transformed indices lie inside the specified asymmetric unit.
    d. The indices of the reflection and the values of F^2 and $\sigma(F^2)$ are stored in groups of symmetry equivalent reflections.

3. The symmetry-equivalent groups of reflections are processed and outliers are flagged in the following steps, applied in turn to each set of symmetry equivalent reflections:
    a. If DAC rejection has been selected all of the individual reflections in the set are averaged, and the statistical tests of Blessing (1987) are applied only to those reflections with $F^2 < \overline{F}^2$, so as to reject those reflections affected by multiple diffraction events in the DAC.
    b. The remaining reflections (or all, if no DAC rejection applied) in the set are averaged, and the statistical tests of Blessing (1987) are applied to identify all outliers.
    c. The surviving individuals in the set are re-averaged to obtain $\overline{F}^2$.
    d. $\sigma\left(\overline{F}^2\right)$ is set to the larger of σ(internal2) and σ(external2), except when only one reflection remains after rejection of outliers, in which case the larger of σ(internal1) and σ(external1) is used.

4. Groups of symmetry-equivalent reflections that are space group violations and those that fall outside the specified 2θ limits are flagged.

5. The numbers of reflections passed and rejected in certain categories are calculated (from the flags set previously), and the results are written to the *print* file.

6. The *Rint* values for merging and *Rsigma* values are calculated for reflections in various categories and the results are written to the *print* file.

7. The $\sigma(F^2)$ of the singlets is modified, if requested by the user (only possible if you are writing averaged reflections to the *output* data file).

8. The reflection data is written to the *print* file grouped into symmetry-equivalent sets, along with the internal and external sigma values.

9. The data are written to the *output* data file under the controls specified by the user in the *exp* file.

10. A *cif* is created with information about the dataset and the averaging statistics.

Conventions

If the symmetry is specified by symmetry operators in the *exp* file then the user is free to choose the orientation of the unit-cell and space group. It should, of course, match that of the indexing of the reflections in the *input* data file.

If the symmetry is specified by Laue group in the *exp* file, then the following conventions are used:

- For the monoclinic Laue class, *b* is the unique axis.
- For the uniaxial Laue classes, *c* is the unique axis.
- For all Laue classes except 2/m, the averaged data have $l \geq 0$.
- For Laue class 2/m, the averaged data have $h \geq 0$ and $k \geq 0$.
- For all Laue classes except -1 and 2/m, the asymmetric unit is defined to be all or part of the octant $h \geq 0$, $k \geq 0$, $l \geq 0$.

The diffractometer angles on the input *abs* data files are not used, except for the absolute value of 2θ which will be used to reject data outside of the specified 2θ limits. Therefore diffractometer circle parities are not relevant.

The UB matrix, if supplied, is only used for calculating the 2θ of reflections from the direction cosines on the input *hkl* data files. This calculation is performed by extracting the B matrix from the UB matrix, and using the B matrix to calculate the scalar product between the direction cosines of the incident and diffracted beams. Therefore the orientation of the ϕ-axis coordinate system on which the UB is defined is not relevant.

The units used for the unit-cell parameters and the wavelength in the *exp* file must be in Ångstrom.

Files

AVERAGE uses five files:

- Information about the crystal symmetry, the experiment, and the controls for the averaging are provided by the user in the *experiment* file. Details of the format of this file are provided in a separate section below.
- Output from the program including a listing of input parameters, all of the data grouped by symmetry-equivalent sets, and the averaging statistics for the whole dataset are provided in a *print* file. Further explanation of the output is provided in a separate section below.
- Intensity data for merging is supplied in the *input* data file. The formats recognised by the program are:
    - RFINE *abs* format, as produced by the WinIntegrStp program (Angel 2003) and ABSORB programs. The file contains one line per reflection with: *hkl*, setting angles, F, $\sigma(F)$, a sequence number and some flags. The setting angles must be in the order 2θ, ω, χ, ϕ.
    - SHELX *hkl* format, with *hkl*, F^2, $\sigma(F^2)$, and direction cosines with respect to the crystal axes (*not the Busing-Levy coordinate system)*. The direction cosines are only used to calculate the 2θ value of each reflection, but in order for this calculation to be performed the *experiment* file must contain the UB matrix or the unit-cell parameters. Conversion of the direction cosines to an orthonormal coordinate system follows the methods of Busing and Levy (1967) and Allan et al. (2000).
- Processed data are written to the *output* data file. The two supported file formats are RFINE *abs* (also called *avg*) and SHELX *hkl,* with the same formats as given above. Whether the *output* data file contains averaged or individual reflections is controlled by the user.

For both *output* file formats the values of F and $\sigma(F)$, or F^2 and $\sigma(F^2)$, are dynamically formatted to fill the output field. The values of $\sigma(F)$ or $\sigma(F^2)$ depend on the output option selected by the OUTPUT card (see description in the section on the *experiment* file). Values of $\sigma(F)$ or $\sigma(F^2)$ that overflow the format are set equal to the maximum value allowed by the format.

SHELX-format *output* files do not contain the direction cosines. RFINE-format *output* files will contain the 2θ values of reflections if they were either read from an input *abs* file, or could be calculated from the direction cosines in an input *hkl* file.

- Information about the dataset and the *Rint* values are written to a *cif*, given a name based upon the name of the *output* data file, together with "*_average.cif*". Thus an *output* data file with the name *example1.hkl* will have an associated *cif* named *example1_average.cif*.

Code Validation

The operation of the *CCSL*-based version of AVERAGE was examination of the results for "reasonableness" ....i.e. no-one complained about the results for about two decades. Version 2.22 produces the same results as the original code, allowing for the changes from processing F to processing F^2.

6. THE EXPERIMENT FILE

The *experiment* file is a text file that can be edited by any text editor, such as Notepad. The general *format* of the instrument parameter file is that the first six characters of each line are read as a label. The label defines the content of the rest of the line. If the first six characters of a line are blank, then the remainder of the line is ignored; blank labels can therefore be used to space out the information or to add comments (see the example files). The lines can appear in any order within the file.

The information is read by Fortran read statements, so floating-point values should include a decimal point, and integer values must not contain a decimal point. Otherwise the format is free. Individual values can be separated by commas or spaces.

If you used WinIntegrStp (Angel 2003) to integrate step-scanned intensity data then you probably used that program to create an *experiment* file, to which you have subsequently added information about the absorption correction. Some of the entries resulting from these two programs may be used again by AVERAGE. You will also have to add the information and controls for the AVERAGE program to the *exp* file. The recognised labels and information that are read by AVERAGE listed below. If items are not present, the program uses the default values as described.

MISCELLANEOUS INFORMATION

TITLE example title
Default: none

Title written onto *print* file

WAVEL 2, 0.709316, 0.713606, 0.50

Default	1,0.7093	
Variables:	nwave	number of wavelengths, 1 or 2
	wave(1)	wavelength in Ångstrom of α_1 component
	wave(2)	wavelength in Ångstrom of α_2 component (if nwave=2)
	wratio	intensity ratio of α_2/α_1 (if nwave=2)

Only the first wavelength value is used by AVERAGE for the following calculations:

- Conversion of components of UB supplied on UBL cards to true UB.
- Calculation of 2θ values from *hkl* if no direction cosines are present in the *input* data file.
- Calculation of 2θ values from *hkl* in the calculation of coverage.

CELL 7.5, 5.2 ,6.7 ,90., 113.7, 90.,
Default: none

Unit-cell parameters, a, b, c, α, β, γ, and Volume. If the volume is not given, then it is calculated from the cell parameters. Not needed if the UB matrix is input.

BLAXES -2, 1, 3
Default: 1,2,3

Defines the sense of the axes of the orthogonal coordinate system used by the diffractometer control software in calculating the UB matrix, relative to the positive axial system defined by Busing and Levy (1967). The three digits refer to the *x,y,z* axes of the diffractometer coordinate system. In the example:

- the –2 in the first position indicates that *+x(diffractometer) = -y(Busing-Levy)*
- the 1 in the second position indicates that *+y(diffractometer) = +x(Busing-Levy)*
- the 3 in the third position indicates that *+z(diffractometer) = +z(Busing-Levy)*

Note that when the diffractometer circles are all at zero, *+x(Busing-Levy)* is along the diffraction vector (towards $2\theta = +90^o$), *+y(Busing-Levy)* is along the X-ray beam, *+z(Busing-Levy)* makes a right-handed set.

This information is only used to rotate the UB matrix provided in the *experiment* file to the Busing-Levy axial system. The converted UB matrix is only used by AVERAGE to convert direction cosines associated with SHELX datafiles. See also description of the UB cards.

For data collected with the Crysalis software and Xcalibur diffractometers of Oxford Diffraction, use -2,1,3.

UB 0.019490 -0.021849 -0.125444
UB 0.129750 0.123922 -0.001238
UB 0.125763 -0.124432 0.020638
Default: none

UB matrix. If the entries of the UB matrix contain the wavelength (i.e. UB(input) = UB(Busing-Levy).λ) then the UBL cards should be used instead.

On input, the UB matrix is rotated to the Busing-Levy coordinate system by use of the information provided on the BLAXES card. The rotated UB matrix is only used to:

- determine the unit-cell parameters for use in the coverage calculation.
- convert the direction cosines in SHELX data files to the Busing-Levy ϕ-axis coordinate system.

UBL 0.019490 -0.021849 -0.125444
UBL 0.129750 0.123922 -0.001238
UBL 0.125763 -0.124432 0.020638
Default: none

UB matrix input if the entries of the UB matrix contain the wavelength (i.e. UB(input) = UB(Busing-Levy).λ). This is the case for UB matrices used by the Crysalis software system of Oxford Diffraction.

On input, the UB matrix is rescaled by λ and then used as described by the UB cards.

CONTROLS FOR AVERAGE PROGRAM

There are a number of entries starting with AVERAG that provide controls of the averaging process.

AVERAG 2THETA MIN 5. MAX 60.
Or
AVERAG 2THETA MIN 5.
Or
AVERAG 2THETA MAX 60.

Default: 0.,180.

Restricts the data to the specified 2θ range (in degrees). Reflections lying outside the specified range will not be output by the program. It is not necessary to specify both MIN and MAX, one only can be specified if desired. But if both limits are required then they must both appear on one line as in the first line above, not on two separate lines.

Reflections which exceed these limits are flagged with a "T" in the *print* file and are not written to the *output* data file. The upper 2θ limit is also used in the coverage calculation reported in the *cif* as _diffrn_measured_fraction_theta_full.

AVERAG SCALE 1.65
Default: 1.0

The values of F^2 and $\sigma(F^2)$ are multiplied by this scale factor. If the card is not present, no rescaling is applied.

AVERAG LESSTHAN 2.0 *keyword*
Default: 0.0 KEEP

Reflections with $F^2/\sigma(F^2)$ less than this value are considered "unobserved" , and are marked as such in the output *avg* data file and on the *print* file listing.

The keyword specifies whether unobserved reflections are written to the *output* data file:

- if *keyword* is REJECT, these reflections are not written to the *output* data file.
- if *keyword* is KEEP, these reflections are written to the *output* data file (default).

AVERAG ABSENT *keyword*
Default: REJECT

If space group symmetry operations have been specified in the *exp* file, then reflections in the *input* data file that should be absent can be identified. This card specifies how these absent reflections appear in the *output* data file:

- if *keyword* is REJECT, these reflections are not written to the *output* data file.
- if *keyword* is LESSTHAN, these reflections are written to the *output* data file. In the *avg* format file they are marked as "less-thans".
- if *keyword* is INCLUDE, these reflections are written to the *output* data file as normal reflections.

AVERAG OUTPUT *keyword*
Default: AVERAGE

- if *keyword* is AVERAGE the *output* data file will contain those averaged reflections passed by the program. The $\sigma(F)$ or $\sigma(F^2)$ values on the *output* data file will be set as follows:
    - for averaged reflections it will be the larger of σ(internal2) or σ(external2) calculated in the second pass of the averaging process.
    - for averaged reflections for which only one individual is not rejected, it will be the larger of σ(internal1) or σ(external1) calculated in the first pass of the averaging process.
    - for singlet reflections it will be the input $\sigma(F^2)$ multiplied by the average of the ratio σ(output)/ σ(external2) calculated for "observed" reflections that were averaged.
- if *keyword* is INDIVIDUAL the *output* data file will contain those individual unaveraged reflections not rejected in the averaging process. In this case the AVERAGE program is being used to reject outliers and to calculate *Rint* values only. The $\sigma(F)$ or $\sigma(F^2)$ values on the *output* data will be copied from the *input* data file.
- if *keyword* is NOSINGLET the singlet reflections will not be written to the *output* data file. This may be used in combination with either AVERAGE or INDIVIDUAL.

AVERAG OUTLIER *keywords*
Default: SYMMETRY

- if SYMMETRY is the only *keyword* outliers in symmetry-equivalent groups of reflections will be identified by the Blessing criteria and will be excluded from averaging and output.
- if *keyword* is NONE no rejection of outliers will be performed.
- if DAC is the only *keyword* then negative outliers in symmetry-equivalent groups of reflections will be identified by the Blessing criteria and will be excluded from averaging and output. This is intended to eliminate "diamond dips", reflections whose intensities are lowered by simultaneous diffraction events in the diamond anvils of a pressure cell.
- if *both* SYMMETRY and DAC appear, then the program applies the rejection criteria as described by the entry for the DAC keyword, and then applies the normal Blessing criteria to the remaining reflections.

AVERAG BLESSING 0.05 3.3 2.60 1.00
Default: 0.05 3.3 2.60 1.00

The four values c_1, c_2, c_3, c_4 for the four statistical criteria specified in section 6.9.1 of Blessing (1987) for the rejection of outliers.

SYMMETRY INPUT

The information required to identify which reflections are symmetrically equivalent can be specified to AVERAGE by either specifying the Laue group by name, or by explicitly listing the symmetry operations of the space group, along with the asymmetric unit. The specification by Laue group is simpler. However, you may wish to use the alternative method for one or more of the following reasons:

- Your data include reflections forbidden by space group symmetry and you wish to exclude these from either the calculation of the averaging statistics such as *Rint*, and/or exclude them from the output dataset.
- Your data have a non-conventional orientation of a space group, such as *c*-unique in the monoclinic system. The Laue option assumes that *b* is unique in monoclinic and *c* is unique in uniaxial crystal systems.
- You wish to merge the data into a different asymmetric unit from the one used when the symmetry is specified by the Laue option.
- Your crystal is acentric and has significant anomalous scattering, so that Friedel mates have significantly different intensities.
- You want AVERAGE to perform a correct calculation of the data coverage of the symmetrically unique reflections in a dataset that does not include the space group forbidden reflections.

AVERAG LAUE 2/m
Default: no symmetry

The Laue symmetry as specified by the conventional Laue group symbol. The recognised symbols are -1, 2/m , mmm, 4/m,4/mmm, -3, -3m1, -31m, 6/m, 6/mmm, m3, and m3m.
Note that *no spaces* should appear between the symmetry symbols.

Note also, that if your dataset comes from a crystal whose space group results in symmetry-forbidden reflections, and those symmetry-forbidden reflections are not present in the *input* data file, then the coverage calculation (reported in the *cif*) will be incorrect when the symmetry is specified by Laue group.

To specify the symmetry by symmetry operators you must specify two sets of cards, as follows:

AVERAG SYMOP x,1/2-y,1/2+z
Default: no symmetry

One SYMOP card has to be specified for each symmetry operator entered. The program combines all of the specified operators to try and form a group, so only the *generators* of the group have to be entered. If more operators are provided in the *exp* file this is not a problem.

The *rotational* part of the symmetry operator is used to transform the indices of the input reflections into the asymmetric unit. Therefore, if rejection of space group absences is not required you can choose to specify only the rotational components of the operators.

AVERAG UNITIN 0 0 1 0
Default: no asymmetric unit

One UNITIN card has to be specified for each boundary of the asymmetric unit in reciprocal space. If the four input numbers on a card are *m,n,o,p* they are interpreted as restricting the asymmetric unit in reciprocal space by: $mh + nk + ol \geq p$. The example given above therefore specifies $l \geq 0$.

The card AVERAG UNITIN 1 -1 0 0 specifies $h \geq k$.

The program uses the specified symmetry operators to rotate the indices of the input reflections until they meet all of the requirements specified by all of the UNITIN cards. Reflections with identical rotated indices are then considered to be symmetry equivalent reflections. Therefore it is critical to specify a consistent and complete set of SYMOP and UNITIN cards in order to obtain a correct averaged dataset. The following example shows the cards required for space group *R-3c* and will produce an averaged dataset in which all reflections have $h > 0$, $k > 0$, $l > 0$.

```
AVERAG SYMOP  2/3+x,1/3+y,1/3+z
AVERAG SYMOP  -y,x-y,z
AVERAG SYMOP   y,x,1/2-z
AVERAG SYMOP  -x,-y,-z
AVERAG UNITIN   0   0   1   0
AVERAG UNITIN   0   1   0   0
AVERAG UNITIN   1   0   0   0
```

AVERAG FRIEDEL *keywords*
Default: FLIP

The default operation of the program is to allow the *hkl* of reflections to be flipped to their Friedel mates, even if an inversion centre has not been specified by SYMOP cards. To prevent reflections being converted to their Friedel mates, specify FRIEDEL NOFLIP.

7. THE PRINT FILE

The exact contents of the *print* file depend on the control information that you provided in the *exp* file. But in general:

- The first part of the *print* file contains information about the unit-cell and symmetry derived from the *experiment* file. If the symmetry was specified as a LAUE class then the printed symmetry elements correspond to a primitive space group with the same rotational symmetry elements as the Laue group. For example, when the Laue class is specified as 4/m the space group symmetry is set to P4/m.
- The next part, headed *General Controls* consists of a listing of instructions in the *experiment* file. Read this carefully and compare it with the *experiment* file to ensure that the program is performing the merging in the way you want it to!
- The *Summary of Averaging* section first provides a listing of the numbers of individual reflections as read from the *input* data file, and then the numbers of those excluded, and those passed, by the averaging process. The next sub-section provides similar statistics on the sets of symmetry-equivalent reflections. The last part of this section provides merging statistics and signal-to-noise measures for various categories of reflections, each further sub-divided to reflections that were classified as observed and unobserved.

  The two *R*-values for merging the reflections listed are defined as:

  - $R\,\mathrm{int} = \dfrac{\sum\left|F^2 - \overline{F}^2\right|}{\sum F^2}$: the merging *R* value used by SHELX and reported in the *cif*.
  - $R\,\mathrm{int}\,2 = \dfrac{1}{Nset}\sum_{1}^{Nset}\left(\dfrac{\sum^{set}\left|F^2 - \overline{F}^2\right|}{\sum^{set} F^2}\right)$

  In both cases, the summations are only performed over reflections which belong to symmetry-equivalent sets; singlets are excluded. Note that *Rint* is defined by the summation over all reflections of the deviation of each reflection ($\left|F^2 - \overline{F}^2\right|$) from the mean value ($\overline{F}^2$) of the set of symmetry-equivalent reflections to which it belongs, divided by the sum of F^2 over all reflections. As such, it will be biased to low values by good agreement on a few well-determined strong reflections. By contrast, *Rint2* provides a possibly statistically-meaningful average of *Rint* evaluated over each symmetry-equivalent set of reflections.

  The *Rsig* values provide a measure of the signal-to-noise of the data:

- $Rsig = \sum \sigma(F^2) \Big/ \sum F^2$: the *Rsigma* from SHELX
- $Rsig2 = \frac{1}{Nref} \sum \left(\frac{\sigma(F^2)}{F^2} \right)$

  The difference between these values is that the first will provide a smaller number because the summations of the sigmas and the F^2 are performed separately, and both the sums in numerator and denominator will be dominated by the contributions of the strong reflections. By contrast, *Rsig2* provides the mean value of $\sigma(F^2)/F^2$, or information content, of the entire dataset. Note however, that reflections with $F^2 = 0$ have to be excluded from the summation of *Rsig2*.

- The main part of the *print* file consists of a listing of the sets of symmetry-equivalent reflections, one set per line. A typical line looks like this, with 6 equivalent reflections read in, 2 rejected as outliers, and 4 used to determine $\overline{F}^2$:

```
H K L    F2(AVE) SIGF2  INT   EXT   INT2 EXT2 NEQ NAVE       INDIVIDUAL SQUARED STRUCTURE
3 3 6 G  16.51   0.43   1.52  0.46  0.36  0.43 6  4   19.58B  16.40    15.05B  16.05   16.69   16.88
```

  - *hkl*: the indices of this set of equivalent reflections within the asymmetric unit.
  - *Flag*: indicates the status of the averaged reflection:
    - *G* reflection passed by rejection criteria, with $F^2 > n\sigma(F^2)$ where *n* is set on the AVERAG REJECT card.
    - *L* reflection passed by rejection criteria, with $F^2 < n\sigma(F^2)$.
    - *T* reflection rejected for being outside the specified 2θ limits.
    - *S* reflection that is a space-group absence.
  - *F2(ave)*: the average value of the structure factor squared, $\overline{F}^2$, for this set, after rejection of outliers.
  - *SigF2*: the esd applied to $\overline{F}^2$. It is the larger of σ(internal2) and σ(external2), except when only one reflection remains after rejection of outliers, in which case the larger of σ(internal1) and σ(external1) is used.
  - *Int*: Population standard deviation of the values of all the F^2 in the set of reflections, excluding those rejected as DAC outliers:

$$\sigma(internal) = \left[\frac{n}{n-1} \sum \left(F^2 - \overline{F}^2 \right)^2 \right]^{1/2}$$

  - *Ext*: The rms average experimental uncertainty of all of the individual values of F^2 in the set of reflections, excluding those rejected as DAC outliers:

$$\sigma(external) = \left[\frac{1}{n} \sum \sigma^2 \left(F^2 \right) \right]^{1/2}$$

  - *Int2*: Population standard deviation of the values of all the F^2 in the set of reflections that were passed by the program, and contribute to F^2*(ave)*.
  - *Ext2*: The rms average experimental uncertainty of all of the individual values of F^2 in the set of reflections that were passed by the program, and contribute to F^2*(ave)*:

$$\sigma(external2) = \left[n / \sum 1/\sigma^2 \left(F^2 \right) \right]^{1/2}$$

  - *Neq*: The number of symmetry-equivalent reflections in the *input* data file.
  - *Nave*: The number of reflections remaining after rejection of outliers and contributing to $\overline{F}^2$.

- The rest of the line consists of the individual values of F^2 as read from the *input* data file. Those reflections rejected as DAC dips are flagged by the letter *D*, the reflections rejected by the normal Blessing criteria with the letter *B*. This listing will extend to subsequent lines if *Neq* > 8.

After the symmetry-equivalent sets have been listed, the singlet reflections are listed.

8. CIF

The *cif* written by AVERAGE contains the following items.

```
_diffrn_reflns_limit_h_max          7
_diffrn_reflns_limit_h_min         -7
_diffrn_reflns_limit_k_max          7
_diffrn_reflns_limit_k_min         -7
_diffrn_reflns_limit_l_max         22
_diffrn_reflns_limit_l_min          0
```

The minimum and maximum indices found in the *input* data file, irrespective of whether they are symmetry-allowed, within the 2θ limits, or outliers, or not.

```
_diffrn_reflns_number              871
```

The number of reflections found in the *input* data file, irrespective of whether they are symmetry-allowed, within the 2θ limits, or outliers, or not.

```
_diffrn_reflns_theta_full          30.00
```

The θ limit to which the coverage reported as _diffrn_measured_fraction_theta_full is calculated. The value corresponds to that specified as the 2THETA MAX in the *exp* file.

```
_diffrn_reflns_theta_max           39.97
```

The θ limit to which the coverage reported as _diffrn_measured_fraction_theta_max is calculated. The value is the maximum θ found in the *input* data file.

```
_diffrn_reflns_theta_min           5.86
```

The minimum θ value for a reflection found in the *input* data file.

```
_diffrn_measured_fraction_theta_max  0.985
```

The number of sets of symmetry-equivalent reflections in the dataset divided by the number of symmetry-allowed reflections in the asymmetric unit out to the maximum θ value found in the *input* data file.

_diffrn_measured_fraction_theta_full 1.000

The number of sets of symmetry-equivalent reflections in the dataset divided by the number of symmetry-allowed reflections in the asymmetric unit out to the value of θ full, set as one-half of 2THETA MAX in the *exp* file.

_diffrn_reflns_av_R_equivalents 0.020

Rint calculated for all sets of reflections in the *input* dataset that have more than one symmetry-equivalent, including all space group absences, reflections lying outside the specified 2θ limits, and reflections subsequently rejected as outliers.

_diffrn_reflns_unetI/netI 0.024

Rsigma calculated for all sets of reflections in the *input* dataset (includes all reflections subsequently rejected).

_reflns_number_total 179

Total number of symmetry-equivalent sets of reflections plus singlets in the *input* data file.

_reflns_number_gt 85

Total number of symmetry-equivalent sets of reflections (plus singlets) written to the *output* data file.

_reflns_threshold_expression 'Fsqd > 2.0u(Fsqd)'

Level in $\sigma(F^2)$ at which reflections are flagged as observed.

REFERENCES

Allan DR, Clark SJ, Parsons S, Ruf M (2000) A high-pressure structural study of propionic acid and the application of CCD detectors in high-pressure single-crystal x-ray diffraction. Journal of Physics C: Condensed Matter 12:L613-L618.

Angel RJ (2003) Automated profile analysis for single-crystal diffraction data. J Appl Crystallogr 36:295-300.

Angel RJ, Downs RT, Finger LW (2000) High-pressure, high-temperature diffractometry. In: R T Downs (Ed) High-pressure and high-temperature crystal chemistry, MSA. pp 559-596.

Blessing RH (1987) Data reduction and error analysis for accurate single crystal diffraction intensities. Crystallography Reviews 1:3-58.

Busing W, Levy H (1967) Angle calculations for 3- and 4- circle X-ray and neutron diffractometers. J Appl Crystallogr 22:457-464.