Manual

for the

Varve Interpolation Program

(VIP)

Manual version 1.0

for VIP 3.0.0

Gordon Schlolaut

contact: SchlolautG@gmail.com

Table of content

Note

This manual is in addition to the publication *Schlolaut, G. 2018. The Varve Interpolation Program 3.0.0 - A unique and easy to use tool for incompletely varved sediments. Quaternary Geochronology accepted*. Please cite this publication if you use(d) the VIP.

The VIP is open-source software; you are free to use, copy, distribute and modify it, but please do read this manual and accompanying paper before using the program. This software is distributed under the terms of the GNU General Public Licence. The VIP does not come with any warranty and the author does not assume any responsibility for the usefulness of any portion of this program.

Data requirements for interpolation

In addition to the requirements stated in Schlolaut (2018), such as that the data should have a log-normal VTF distribution etc., it should be pointed out that accurate measurements of the distances between counts are essential. The program itself works with an precision of 0.01 mm.

Tip: If data for interpolation have a very small mean SR relative to the precision of the program, e.g. 0.1 mm/a with a resolution of the measurements better than 0.01 mm, it could be beneficial to multiply the count distance measurements by 10. If the mean SR is relatively large, e.g. 10 mm/a, dividing the distance measurements by 10 will reduce calculation time greatly.

Installing the VIP

The VIP folder needs to be unpacked on the computer using the open source software 7zip (http://www.7-zip.org/). Matlab needs to be installed. To start the program open Matlab and change the current path to the VIP folder using either the address field or entering *cd('*pathname*')* in the command window. Then enter *VIP* in the command window and press enter.

Note that you can find example data from the Rehwiese record (Neugebauer et al., 2012, QSR 36, 91-102) for interpolation in the data folder.

Data Preparation

The program expects the count data in a four column matrix. The first column should contain the measured depths of single counts in mm (i.e. the cumulative count distances). These data will be used for all the essential calculations. In the second column the composite depth of the counts in cm can be entered. This is the depth scale the program will use for the interpolated age model and also is the depth scale on which event layer positions are to be provided. The differentiation is made because distance measurements made with, for example, a microscope are usually more accurate than composite depth definitions, which are often made on cores or core photographs. If this differentiation is not required, the first and second column can contain the same depth scale, though it should be kept in mind that the first column is in mm and the second in cm. The third column contains layer quality, i.e. information on the quality of the seasonal layers counted, for example from 1 (very good) to 4 (very poor). The forth column can contain the section quality, also in a numerical format. Note that if an interval is marked as quality 0 in the forth column the programme will not interpolate this interval. This can be useful if unusually thick varves occur, for example directly above a turbidite or such.

These layer and section quality information can be used to create quality selective datasets (cf. Schlolaut et al., 2012, Quat. Geochronol. 13, 52-69). If no quality information is available, these columns can simply be filled with ones.

The data need to be saved as an ascii-delimited file. For example, if the data are originally contained in an Excel spreadsheet, select all the data and copy it into a text file using NotePad or something similar. Save the file on the computer (using an extension like *.txt or *.dat). Consider saving it in the data folder within the VIP folder for quick access. The file should contain only numbers, i.e. for example no column headers.

If event layer information is available, i.e. the position of instantaneous deposits like tephras or turbidites in which no varve interpolation should be carried out, these should be saved in another ascii-delimited file which in the first column contains the top and in the second column the base of the event layers in composite depth, i.e. the depth scale contained in the second column of the count data matrix. The VIP will remove any counts within event layers if any should exist and will not interpolate within event layers.

The GUI of the VIP

Loading the data is simply done by pushing the buttons in the field load data and selecting the files created as described in the previous section. In the field 'Input Data Information' the position and age of an anchor point can be entered in the form depth age lower-error upper-error. The lower error should be a negative number, e.g. an anchor point at 72 cm depth with an age of 6735 years, a lower error of 34 years and an upper error of 47 years would be entered as 72 6735 -34 47. If nothing is entered, or the anchor point is outside the dataset, the topmost count will be the anchor-point with an age 0 years and an error of 0 years.

In the next field it is possible to define an interval for interpolation in case that not the whole dataset should be interpolated. It should be kept in mind that the boundaries must be given in the depth scale which was entered in the second column of the count dataset. The program will find the counts closest to the boundaries entered and use these as boundaries (since the program does not extrapolate).

Next, a quality selective interpolation can be defined (Schlolaut et al., 2012, Quat. Geochronol. 13, 52-69). Often incompletely varved intervals contain very low quality layers, which can make measuring of distances difficult or make it difficult to distinguish seasonal layers from non-seasonal layers. By assigning a quality to each count, it is possible to interpolate the

dataset using all counts and using only high quality (i.e. certain) counts. Under ideal conditions the interpolation result should be the same in both cases. However, if the results differ, this may point towards issues in the dataset. Similarly, an interval quality rather than a layer quality can be used. It should be kept in mind that zero as an interval quality is reserved for intervals in which the program should not add interpolated counts.

Next, breakpoints can be entered, separated by spaces, in composite depth. If no breakpoints should be used this field should contain NaN (stands for not a number), i.e. it should not be left empty. Breakpoints outside the dataset are ignored, i.e. the breakpoints don't need to be adjusted when interpolating only an interval of the dataset. Note, that the program will define the count closest to a breakpoint position as the actual breakpoint, since the program interpolates and does not extrapolate, thus interval boundaries must be defined by a count. (There is a work-around to make the program extrapolate by adding dummy-counts at appropriate positions.)

The next field is 'Edit Settings' on the right. All values there but one have been described in Schlolaut et al., 2017. The variable not mentioned is resolution. This field controls the resolution of ranges being run in the fitting process. For example, the standard sigma range is 0.2 to 0.6. With a standard resolution of 1/10 the sigma values used will be 0.2, 0.24, 0.28 … 0.52, 0.56, 0.6. The resolution will also be used for other cases. For instance, the mode position is determined with an uncertainty, i.e. range, for which the resolution applies, too. While a resolution as high as possible is beneficial, a very high resolution may result in too long a calculation time, thus 1/10 is the standard setting to compromise between the two.

The settings as they are, are the standard settings, and using these is the automated version of the program. The effect of changes in the settings (for advanced users) will be explained later separately. If changes to the settings, and/or the field 'Input Data Information' are made, these settings can be saved by pushing the button 'save settings' and retrieved anytime by pushing 'load settings'.

This leaves the middle of the console to explain. If the checkbox 'plot result after calculation' is checked, a figure will be opened in a new window showing the age-depth model (ADM) with error estimates, either as line graphs or as errorbars. Via the button 'load reference ADM' another ADM (for example from ^{14}C dating or the original count dataset) can be loaded, which will be plotted together with the interpolation result. 'Start Calculation' obviously starts the calculation and the resulting ADM will automatically be saved as ADM.dat. Note that if this file

is not renamed manually after creation it will be overwritten with the next calculation. After a calculation the 'analyse sub-section' button will open a new window showing plots and other information relating to the finished calculation (see next section). The data used for 'analyse sub-section' are stored in a file called LastResult.mat, which will be overwritten with each new calculation and each new interval separated by breakpoints. To avoid overwriting 'Date and Time' in the lower left 'save options' field should be checked, in which case the filename will be Results followed by data and time. From the analyse sub-section window result files can be loaded, though LastResult.mat is always loaded by default.

At the very bottom of the GUI is a progress bar, which has yellow background during calculation, which turns green when the calculation is finished. A red bar shows the approximate progress.

Sub-section analysis GUI

The sub-section analysis GUI allows the visualisation of the data for each sub-section. It can be started from the main VIP GUI or by entering f_analyse in the Matlab command window. In the upper left corner a results file can be loaded (by default LastResults.mat is loaded). Below that the approximate starting depth of a sub-section can be specified. The program will find the sub-section with the closest starting point to this value. Below that the sub-section length control parameter and the maximal allowed difference between SR_{est} and SR_{int} can be specified. Of course only those values will be available which were used in the calculation. Pressing the 'show' button will produce the plots relating to the sub-section and settings specified. The sub-section analysis GUI produces five plots. The upper centre figure shows the count distances versus depth. The red line denotes the mode position of the CDF distribution. The dashed red line shows the position of a secondary mode position, which is based on the data smaller than the main mode position. In this range the program will test which mode position gives the best fit in the Monte Carlo approach. The upper right plot shows the CDF distribution. The bin size is 1/5 of the model value of the CDF distribution. The red lines have the same meaning as in the upper centre plot. The lower centre plot shows the cumulative fit between the actual CDF distribution and the fitted CDF distribution to the upper limit of 2.5 times the estimated mean SR. The lower right plot shows the approximate fit in a frequency plot. The fitted CDF distribution is adjusted so that the measured and fitted CDF distribution have the same mode value. Blue bins mean that the distance data are smaller than 2.5 times

the estimated mean SR, light blue bins means that the data are larger. The bin size is 1/3 of the estimated mean sedimentation rate. The lower left plot shows the final age depth model (before removal of event layers and including interpolation in quality 0 intervals) as well as the age depth model which would derive from the settings selected in the upper left corner. For example, if a calculation has been run with sub-section lengths of 13, 15 and 17 it can be checked what the ADM would look like if only 13 or 15 or 17 were used compared to the final product (before event layer removal). Pushing the 'clear ADM figure' will remove all plotted age models except for the final age model plot.

Lastly, at the right above each plot are little radio-buttons. If the button 'figures in new windows' is pressed all checked figures will be opened in normal Matlab figure windows, enabling zooming, saving, colour changes etc.

Changing settings in the VIP GUI

In the 'Edit Settings' panel on the right the parameters for calculation can be adjusted. The parameter at the top, sub-section length control, is well covered in Schlolaut (2017), thus only one point shall be added here. Using different sub-section lengths, rather than just one, reduces the danger of a misfit in a sub-section making it into the age model. A misfit can be coincidental, for example due to some random data distribution resulting in too large or too small mode position in a specific sub-section. Excluding data or including more data, i.e. using different sub-section lengths, may avoid the misfit in other sub-sections and thus avoiding its inclusion in the final age model.

The next parameter is overlap, or differently phrased: the frequency with which new sub-sections start. A large overlap is helpful in identifying changes in SR and reduces the chance of a misfit in a certain sub-section influencing the age depth model for reasons explained above. However, a large overlap can also be problematic, for instance if there is a trend in the mean SR or high frequency-low magnitude changes in SR. While in such situations the mean SR within single sub-sections may be reconstructed correctly, given that slight changes in SR do not pose a problem for reasons described in Schlolaut (2017), the overlap of the sub-sections can produce a large spread since every sub-section may have a different SR. Minimum sub-section length essentially just makes sure that there is enough data to be statistical meaningful. Otherwise the statements about sub-section length apply here, too.

The standard sigma range is relatively wide to cover ideally all likely cases. Narrowing the range down, for example by analysing well varved intervals, can be helpful since a shorter sigma range means that sigma will be higher resolved in the Monte Carlo fitting process and misfits become less likely, given that very large and very small values are excluded. Importantly, just the starting range is defined here. Through the optimisation loop the program can derive sigma values outside this range.

The last setting is tolerances. A very high tolerance means that the best fit in the fitting process is used. A small tolerance means the fitted and measured distribution need to have a similar or the same SR. As with the sub-section length control it is possible to use multiple values, which will add results to the Interpolation Matrix. In an ideal situation the tolerance should have no effect, since a perfectly correct result will produce an excellent fit which has the same SR as the interpolated result. Thus, a small tolerance will be particularly beneficial if the plot is noisy or high percentages of counts are indistinguishable; a high tolerance will be beneficial if the CDF distribution is well defined and in particular if the annual sediment accumulation variability is rather high at the same time. Ultimately it is suggested to compare interpolation results from a run with a low and a run with a high tolerance; pronounced differences between the runs can reveal issues with the dataset, which can help selected the best setting.

Using the Matlab command window

Users familiar with Matlab may prefer to use a command rather than the GUI. The command is

```
[ADM,IntMat]=mfVIP(Data,Events,Anchor,Breakpoints,LQltSel,SQltSel,SSL,
MinLength,OL,Sigma,Tolerance,Resolution,D3);
```

The output is the age depth model and the Interpolation Matrix. Data and Events should be matrices as described in 'Data preparation'. Events can be replaced by NaN if event layers are not present. Anchor is a 1x4 vector (or NaN), Breakpoints is 1xn vector (or NaN), LQltSel/SQltSel stands for Layer/Section Quality Selection (1xn vectors or NaN), SSL are the sub-section length control values (1xn vector), MinLength the minimum length of sub-sections, OL is the overlap, Sigma the sigma range (1x2 vector), Tolerance is the tolerance between SRint and SRest (1xn vector), Resolution is the same as resolution in the GUI and D3 is to be 0 if the fitting should exclude a distribution relating to counts three years apart and 1 if it should be included.

Note that the program will display the 'Fit Matrix' as it progresses, to show the progress. The first column of the Fit Matrix is depth from and the second column is depth to. For each value in 'Tolerance' one line per sub-section will be displayed. An example command list is provided in the main VIP folder.

Questions and answers

If you have any questions or suggestions feel free to send a mail to SchlolautG@gmail.com.