Natural Time-Series Analysis Software

Frederick W. Vollmer

July 15, 2020

This page intentionally left blank

Preliminaries

License

Antevs software and companying documentation are © 2011-2020 Frederick W. Vollmer. They come with no warrantees or guarantees of any kind. The software is free and may be downloaded and used without cost, however the author retains all rights to the source, binary code and accompanying files. It may not be redistributed or posted online. The only stipulation is that citation be given for any usage that leads to publications or presentations.

The software is provided "as is", without warranty of any kind, express or implied, including but not limited to the warranties of merchantability, fitness for a particular purpose and noninfringement. In no event shall the authors or copyright holders be liable for any claim, damages or other liability, whether in an action of contract, tort or otherwise, arising from, out of or in connection with the software or the use or other dealings in the software.

Citation

Any significant usage, such as a resulting presentation or publication, should include the following two attributions to the program and author:

Rayburn, J.A., and Vollmer, F. W., 2013. ANTEVS: A quantitative varve sequence cross-correlation technique with examples from the northeastern United States. GFF: Journal of the Geological Society of Sweden, v. 135, p 282-292.
Vollmer, F.W., 2020. Antevs: Natural Time Series Analysis Software. Retrieved from: www.frederickvollmer.com/antevs.

Any reference to the content of the User Manual (this document) may be cited as follows:

Vollmer, F.W., 2020. Antevs: Natural Time Series Analysis Software User Manual. Retrieved from: www.frederickvollmer.com/antevs.

Registration

Please consider registering the software, registration is free and helps determine the software usage and justify the time spent in it's upkeep. To register, simply send an email to me at vollmerf@gmail.com with your user name, affiliation, and usage. I will send you an email in reply with my thanks, and will not place you on a mailing list. For example, send me an email with something like:

    User: John Rayburn
    Affiliation: SUNY New Paltz, Geology Department
    Usage: Undergraduate geomorphology course and varve correlation research

The author is happy to take emails with questions and suggestions, either at the university (SUNY New Paltz) or at the gmail address used on his website. However he is not reliable about checking email, so please forgive him if he is slow in answering, he will try to respond in as timely a fashion as possible.

Installation

On Macintosh macOS, double click the disk image file (dmg), and drag the Antevs application to your Applications folder, or other desired location. Do not try to run the software from the disk image installation window, it must be moved to another location on your computer. The User Manual and Example Data folder can then also be copied to the same location.

Gatekeeper in macOS 10.7 and up must be set to allow applications other than from the Mac App Store to be opened. To do so, open the macOS *System Preferences* and the *Security & Privacy* option. Under *General* select *Allow apps downloaded from: Anywhere.*

On Windows, unzip the zip file using the *Extract All* option, and drag the Antevs application to any desired location. Make sure to entirely extract Antevs from the zip file, this is the most common installation problem. The User Manual and Example Data folder can also be copied to the same location.

On Linux unpack the tar file (tgz), and copy Antevs to any desired location. The User Manual and Example Data folder can also be copied to the same location.

The Example Data folder contains example data and images, these are referred to in this guide. After installing a new version, if any problems arise, try using the *Reset Preferences* command in the *Help* menu. Alternatively, start Antevs while holding down the *Shift* key. This will allow resetting the preferences to default values. The preferences are stored in the file Antevs.json, which is located in the folder Antevs in your operating system's application preferences folder. To deinstall simply delete the Antevs application, and optionally delete the preference folder. No other files are installed on your computer. No administrative permissions are required to install Antevs, and it is possible to keep a copy on a thumb drive to run on any computer.

SHA-256 hash files are provided for each download package. These can be used to verify the package integrity, i.e., that it has not been modified and is the same file that was uploaded by the author. Please contact your system administrator or computer technician if this is required.

Contents

1   Introduction

Time-series arise in situations where a value is measured through time, generally at discrete intervals. Annual intervals commonly occur in natural systems due to seasonal variations in temperature, as displayed in proglacial varves, tree-rings, and glacial ice. Antevs (*Automated Numerical Time-series Evaluation of Varying Sequences*) is designed primarily to work with such natural time-series, but can be applied to other discrete time-series as well. Antevs is designed to aid in correlation of time-series, where an undated times-series, the unknown, is compared to a previously dated series, a chronology, to determine if they overlap in time and to date the unknown series.

Antevs was developed for the correlation of proglacial varve sequences (Rayburn and Vollmer 2013), however it includes standard techniques for matching tree-ring sequences, and can convert files between varve and tree-ring formats. It includes routines for reading the Velmex UniSlide digital readout used for tree-ring measurements. A common problem with natural time-series is that of missing data, often due to broken cores, so Antevs includes interactive procedures to correlate sequences that contain missing values.

Antevs is designed to be user friendly and easy to use, but with powerful analytical tools. It undergoes extensive testing by faculty and undergraduates working on research projects in John Rayburn's varve and tree-ring laboratory at the State University of New York at New Paltz.

Antevs is developed on Macintosh macX and is compiled for Macintosh macOS, Windows, and Linux. Currently testing is done mostly on maxOS 10.13, Windows 7, and Linux Ubuntu 18.04. User input is appreciated to improve operation on various platforms.

1.1   Example Usage

To test the correlation of glacial varve series from the Hudson Valley at Newburgh, New York with the Connecticut Valley at Hartford, Connecticut, download the following North American Varve Chronology (NAVC) Normal Curves files from the North American Glacial Varve Project website (Antevs 1922 Ridge 2012 Ridge et al. 2012):

HUD29-32AM.txt
CON28-32AM.txt

or open the included tsv (tab-separated values) files:
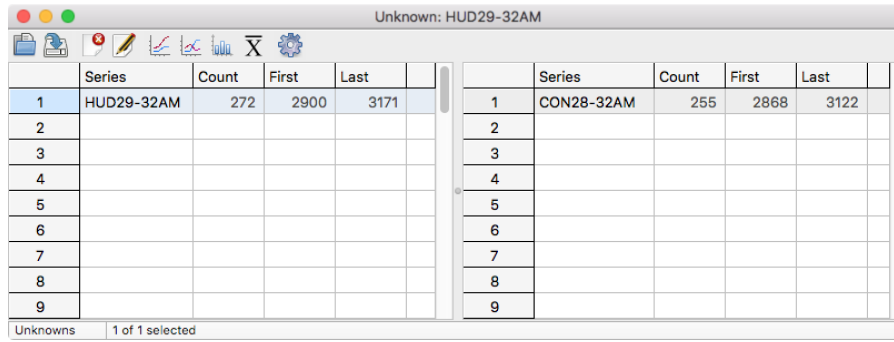
HUD29-32AM.tsv
CON28-32AM.tsv

Assume one is known and the other is unknown, so open HUD29-32AM as an unknown and CON28-32AM as a chronology (Figure 1.1). Select them both, and use the command *Analyze - Data Plot* to view the raw data (Figure 1.2). Note that the differences in scaling and long wavelength trends makes any correlation difficult to see.

To detrend and rescale the data select *Analyze - Detrend* to open the *Preferences* dialog, and set *Detrend* on, with a 16 term Fourier curve fit, using the *Standardized* normalization method (Figure 1.3a). In the *View* pane check *Detrended* data for both *Unknown* and *Chronology* (Figure 1.3b), and press *OK*.

The data plot will update to show the data detrended by removing long wavelength trends using a 16 term Fourier curve, and normalizing using standard deviations of the residuals, making the correlation more clear (Figure 1.4).

To perform a cross-correlation and view the resulting correlogram, select *Analyze - Correlogram* (Figure 1.5). Use the mouse to examine the r correlation values. Note the strong spike at offset 0

Figure 1.1: *Data Sets* window with an unknown and a chronology file open and selected.
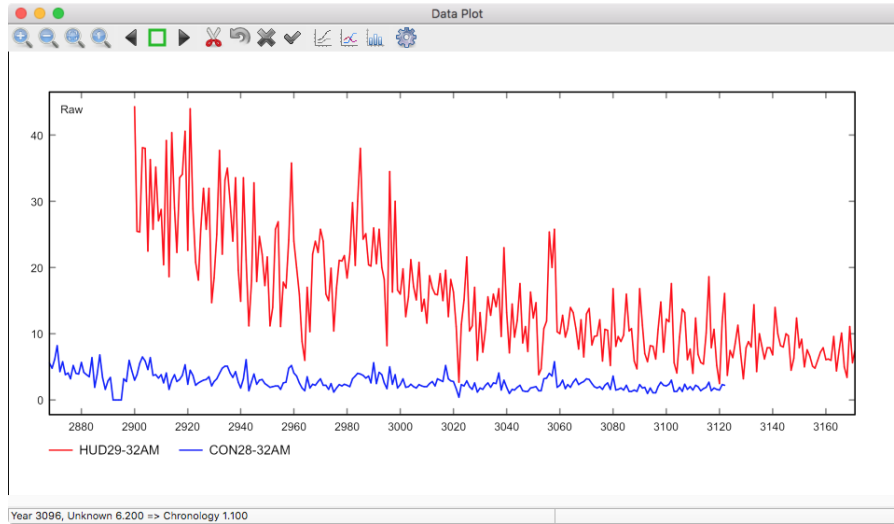
Figure 1.2: *Data Plot* window showing the two raw varve example data sets. The difference in scaling and the long term variations make the correlation difficult to see.
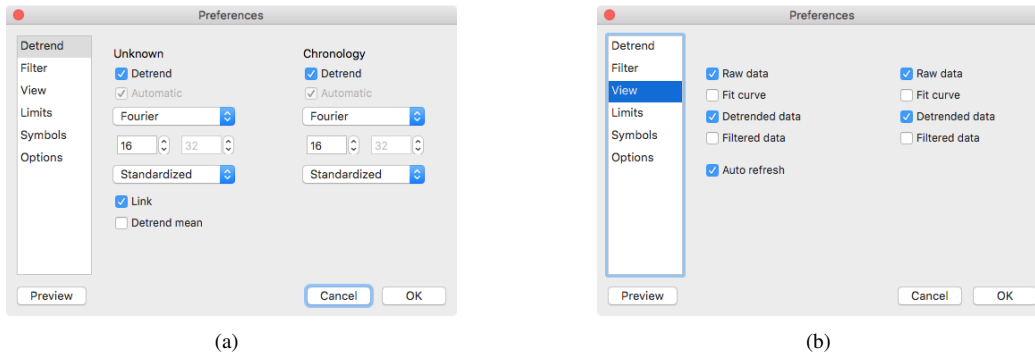
Figure 1.3: Antevs *Preferences* dialog. (a) The *Detrend* pane with the *Detrend* options on, and default *Fourier* and *Standardized* settings. (b) The *View* pane with *Raw* and *Detrended* options on.

and $r = 0.643$, the best correlation. This demonstrates that the data is correctly correlated, with zero offset between the two series.

Finally, select *Analyze - Bin Test* to see how subsets of the unknown match the known chronology. The default bin size is 50, with a minimum overlap of 50. Most of the subsets match with zero offset suggesting a robust match (Figure 1.6).

Tests on NAVC varve data sets known to be correlated suggest that a correlogram with a single correlation spike with $r \geq 0.6$ and $z \geq 6$ is a good potential correlation for varves (Rayburn and Vollmer 2013). However, definitive correlation acceptance criteria are difficult to quantify, rather, Antevs provides statistical and graphical tools for the evaluation of sequences, recognizing that the
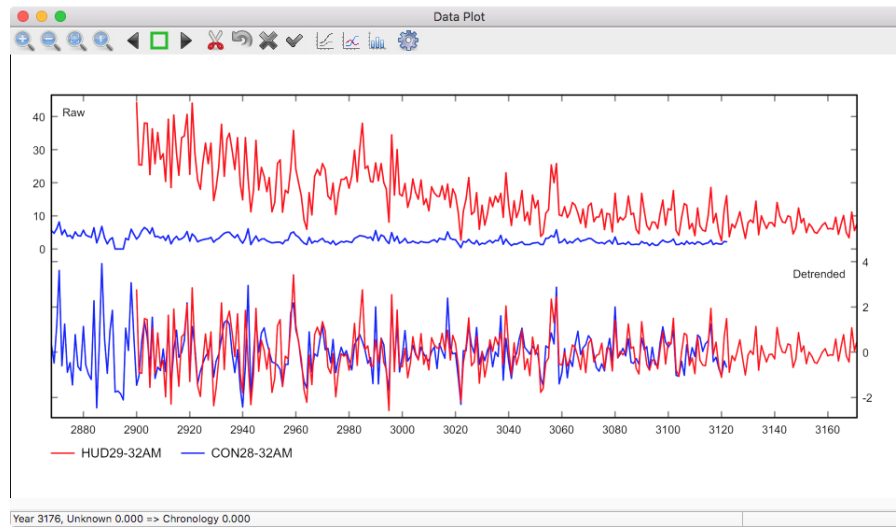
Figure 1.4: Data plots of raw and detrended data using a 16 term Fourier series.

ultimate decision for correlation, whether tree-rings, varves, or other time series, lies with the researcher (Grissino-Mayer 2001 Rayburn and Vollmer 2013).
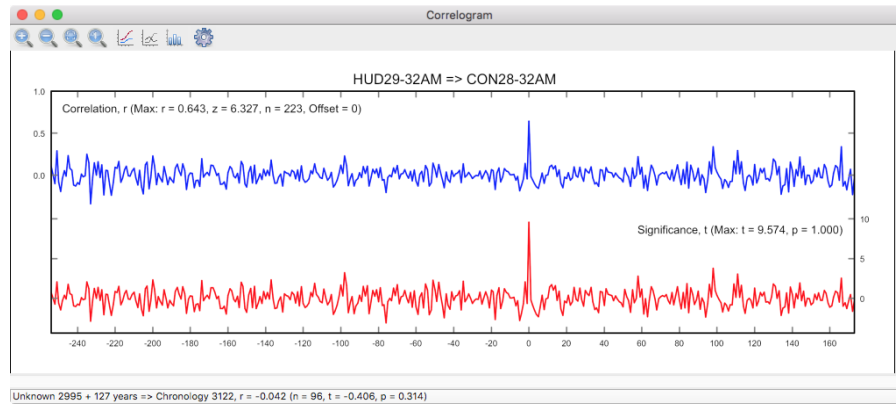
Figure 1.5: Correlogram for two varve sequences showing a strong correlation spike at the best matching year offset with $r = 0.643$, $z = 6.327$. In this case, the best matching year offset is 0, indicating that they are correctly correlated.
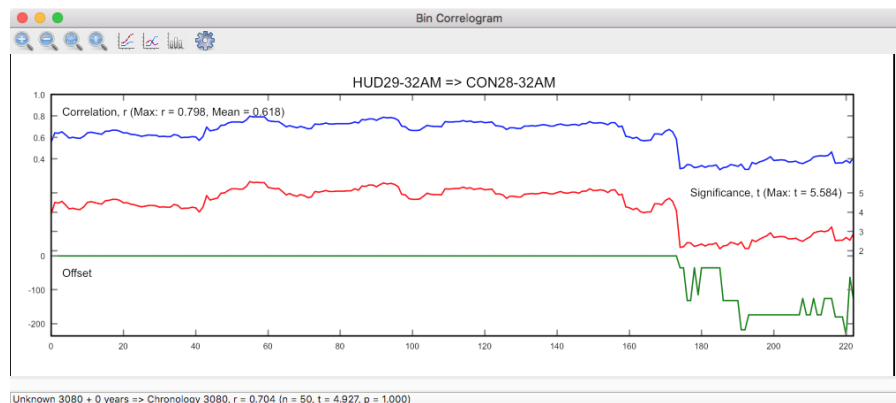
Figure 1.6: Bin correlogram for varve data suggesting a robust fit at 0 year offset. Bin correlation uses a sliding window, by default 50 years wide, to perform multiple correlations across the sequence

2   Data Entry

Antevs can read numerous data file formats, including most common spreadsheet files (e.g., Excel, LibreOffice). A *header* row in these files, as described in Section 2.1, is used to indicate the type of data in the subsequent rows. Additionally, data can be entered directly using the *Data Edit* window. A third option is to use automated data collection with the Velmex Unislide digital readout system, commonly used for tree-ring measurements.

2.1   Data Files

Numerous data file formats are supported, the simplest are spreadsheet (e.g., Excel, LibreOffice) compatible lists with a header. The default is a simple tab delimited, or tab-separated value (tsv) format, consisting of optional comment lines, a header identifying the data columns, and rows of data. For example:

```
// Hudson Valley; Newburgh; NY
// This is a comment
//
Year<tab>Raw<tab>Notes
2900<tab>44.300<tab>
2901<tab>25.500<tab>approximate value
```

Headers can include: *ID, Year, Raw, Fit, Detrend, Count,* and *Mean. Count* and *Mean* are special output headers used when creating a mean chronology. *Summer, Winter, Notes* headers are also read and written, but are not otherwise used. It is best to avoid commas in comment lines. The tsv file format is recommended over the space-delimited txt format, as it are can be read and written by most spreadsheet programs, and the comment and header lines provide self-documentation of the content.

A comma-separated value (csv) format is identical to the tsv format, but uses commas as data delimiters. Note that commas in data fields can potentially cause problems. This format can be also read and written by most spreadsheet programs, for example:

```
// Hudson Valley; Newburgh; NY
// This is a comment
//
// Hudson Valley; Newburgh; NY
// This is a comment
//
Year,Raw,Notes
2900,44.300,
2901,25.500,approximate value
```

A space-delimited format (txt), which is for a single sample and has year or sequence number (first column), and total thickness as decimal values in centimeters in the second column. If four columns are detected they are assumed to be year, summer thickness, winter thickness, total thickness. A tsv format is recommended over this format. Note that all text file formats use decimal point separators ("."").

Three additional formats used by spreadsheet programs such as LibreOffice Calc and Microsoft Excel are supported: OpenDocument (ods), Office Open XML (xlsx), and Excel (xls) spreadsheet formats. The spreadsheet may have comments at the beginning starting with"//", and must have a header line as described above.

2.2 Tree-Ring Files

A number of specialized formats have been developed for tree-ring files, some of which were based on Fortran punch card formats. Antevs can read and write many of these, and convert them into other formats.

A Tucson decadal tree-ring raw measurement format (rwl) contains fixed column records with core id, decade or start year, and thicknesses. Multiple cores per file are supported. Thicknesses are stored as six digit integers in units of 0.01 mm or 0.001 mm. When opening a rwl file the precision is determined by the data stop values stored in the file. The desired output format can be set with the *RWL units* setting in the *Preferences* dialog, the default is 0.001 mm.

A Tucson tree-ring chronology format (crn) file contains a single composite mean chronology which is detrended and has a mean value of 1. To compare a raw unknown measurement to a crn chronology, the *Ratio* method should be used for the unknown, as the *Standardized* method gives a mean of 0. Note that a crn chronology, as well as a *Mean* master chronology (Section 3.4), are not detrended or filtered as this is done when creating the mean chronology. The *Detrend* and *Apply normalize* filters settings are therefore ignored by default for mean chronologies. This may be overridden, however, using the *Detrend mean* and *Normalize mean* options in the *Preferences* dialog.

Some Tucson tree-ring chronology files contain more than one series, each of which has been processed differently (e.g., RAW, STD, RES, ARS). If multiple series are found in a file, a dialog box will be displayed, from which one of the series can be selected. If multiple series from the same file must be opened, duplicate the file with a new name, open it, and select the required series.

A NOAA (National Oceanographic and Atmospheric Administration) chronology format (noaa.crn) file contains the same data as a crn file (year, raw thickness, and count data), stored in a simple tab delimited format. A noaa.crn file typically contains extensive metadata in the file header. Note that, unlike most file formats, the name of the file must include "noaa" as the last four characters of the file name, as well as the extension "crn".

2.3 Manual Data Entry

While it is often convenient enter data into a spreadsheet such as Excel or LibreOffice, entering a new time series directly into Antevs is straightforward. In the main *Data Sets* window select the *Unknowns* grid, and then the *File - New* command. A new time series entitled "untitled" will be created (Figure 2.1). Change the title as required, and use the *Edit Data* command (pencil and paper icon) to open the *Data Edit* window (Figure 2.2).
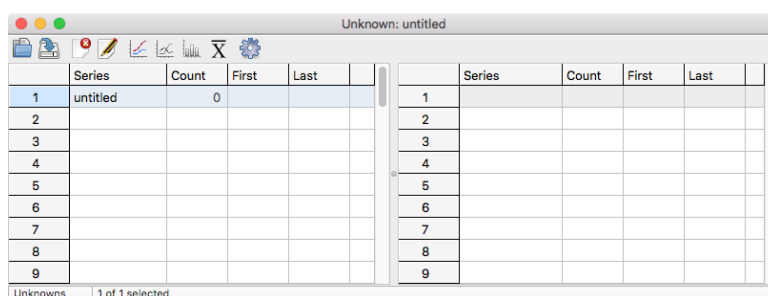
Figure 2.1: The *Data Series* window after using the *File - New* command to create a new series. The series can be renamed as required.

By default the first year is 1900, edit the starting year as required (this can also be changed later), and then begin entering data in the *Raw* column. Each entry advances the year by one and moves the cursor down one row (Figure 2.3). Notes can be entered in the *Notes* column to indicate anything distinctive, such as damage to the varve or tree-ring.

After data has been entered, the *Cancel Edits* (red X icon) and *Accept Edits* (green checkmark icon) commands become available. Use *Accept Edits* to accept the data, this will update any currently open plots. *Cancel Edits* will revert the data to the previous state last accepted. The *Revert* command
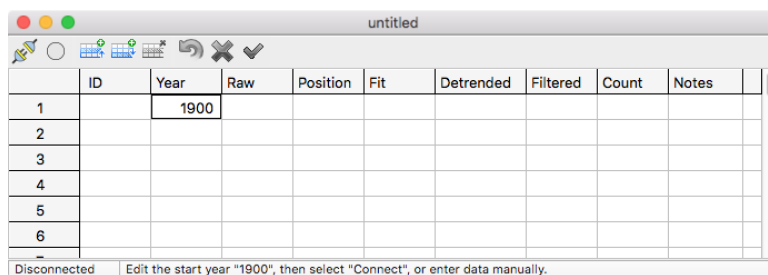
Figure 2.2: The *Data Edit* window after creating a new series. The start year is given a default value of 1900, edit this as required.

(blue curved arrow icon) is not currently available because the initial data series was empty. When editing a data set opened from a file, *Revert* will revert to the state when the file was opened. When finished, accept the edits, close the *Edit Data* window, and save the series from the *Data Sets* window using the *Save As* command.

Figure 2.3: Manual entry of data in the *Data Edit* window, ready to enter data for the year 1883. Grayed entries can not be edited. The *Accept Edits* icon is now enabled.

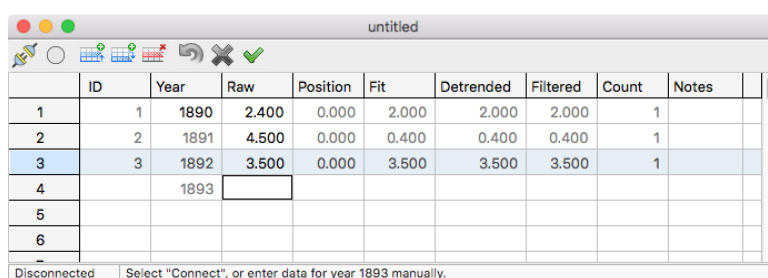Antevs can be used for editing existing time series, and has additional commands such as for splitting and joining rows. If the time series start year must be changed, enter the new value in row 1, and use the *Edit - Renumber* command. Missing values are entered as "NaN", or "nan" (Section 4.5.1).

2.4   Automated Data Entry

Antevs allows automated data collection using the *Velmex Unislide* digital readout system, commonly used for tree-ring measurements. This has been tested on Windows and Macintosh systems. To collect data measurements, click on the *Unknown* grid list and use the *File - New* command to create a new file (Figure 2.1). Select the file, rename it as desired, and select *Edit - Edit Data*. In the *Data Edit* window (Figure 2.2), select *Edit - Connect* (blue plug and socket icon) and enter the required serial port parameters.

2.4.1   Serial Port Connection

The Velmex Unislide digital readout system communicates via a serial port (RS-232 DB9), which requires a USB to serial (RS232) converter on most computers. Numerous converters are available, and their setups, specifications, and quality differ. Converters contain different internal components (processor chip, serial driver chip, output power, surge protection, etc.) and must be compatible with your operating system to ensure a stable connection. Consult your computer technical support to aid in selecting a suitable converter.

Antevs is set for the default parameters that work in our laboratory on Windows (COM3, 9600 baud, 8 data bits, 2 stop bits, no party, no flow control). If the default settings do not work, check the information that comes with your USB to serial (RS232) converter, and the Windows hardware

information (*Control Panel - Device Manager - Ports* on Windows 7) to determine the correct settings for your system.

On a Macintosh open the *Terminal* from the *Applications Utilities* folder, and enter the following command:

ls /dev/tty.*

This will list available serial ports. For the Macintosh using a Prolific PL2303 chip set serial USB converter the required port is:

/dev/tty.usbserial

This port works in our laboratory (Figure 2.4). An open source driver is available for the PL2303 chip set that works on Macintosh, and is included on the Linux Ubuntu distribution. A free download-loadable version from Prolific includes an installation manual (find it using a web search on *Prolific PL2303*).
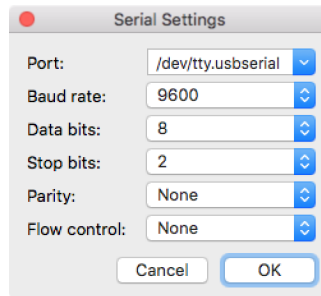

Figure 2.4: Serial port dialog showing settings for a Macintosh computer, these may differ on your configuration.

Future versions of Antevs will attempt to connect with other data collection devices. The author will work with researchers who are attempting to automate their data collection, currently this must be through a serial port connection.

2.4.2   Data Coordinate Entry

After connecting to the recording device in the *Data Edit* window using the *Edit - Connect* (plug and socket icon) command, the red *Record* icon indicates a connection has been made, the status has changed from *Disconnected* to *Connected*. Now move the device to the *beginning* of the first year in the sequence, and press the red *Record* icon to enter *Initializing* mode (yellow icon). Now Antevs is ready to accept the coordinate of the *start* of the data sequence (Figure 2.5). Antevs has four modes during automated data entry: *Disconnected*, *Connected*, *Initializing*, and *Recording*. These are indicated in the status bar, as well as by the color of the *Record* icon (gray, red, yellow, and green respectively). Each mode change is indicated by an audio cue, a click. User instructions are also given in the status bar.

Edit the initial year as required (this can be changed later), position the recording device at the *beginning* of the data sequence, in this case at the start of the year 1900. The device should be reset to zero at this point, although this is not critical. When ready, select the red *Record* icon to go into *Initializing* mode (yellow icon). Now use the device to send the coordinate, and the mode changes to *Recording* (Figure 2.6).

Once in recording mode, advance the recording device to the *end* of the first year, in this case the end of 1900, and send the coordinates from the device. The measurement for the first year is now displayed, and Antevs is ready to accept the coordinate for the end of the next year, in this case 1901 (Figure 2.7). Notes can be entered in the *Notes* column to indicate anything distinctive, such as damage to the tree-ring.
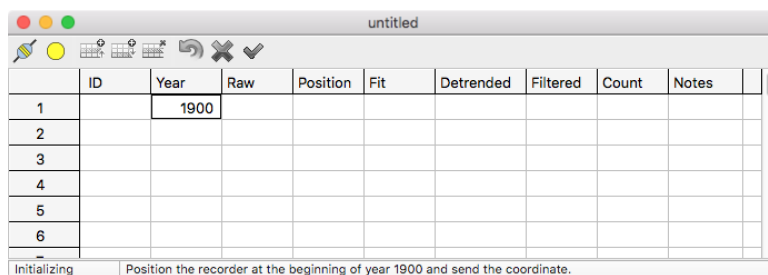
Figure 2.5: The *Data Edit* window in *Initializing* mode (yellow icon), after connection with the recording device, and pressing the red *Record* icon. Antevs is now ready to accept the coordinate of the start of the data sequence. It is a good idea, but not critical, to zero the device before proceeding. Edit the start year if desired, and move the recording device to the *start* of the sequence, and send the coordinate.
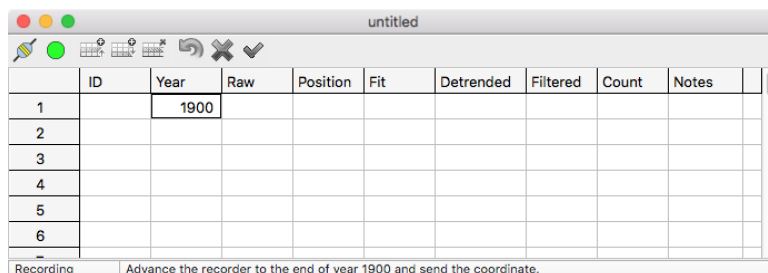

Figure 2.6: The *Data Edit* window in *Recording* mode, after sending the coordinates of the start of the data sequence, and ready to accept the coordinate of the *end* of the first year.

Figure 2.7: The *Data Edit* window in *Recording* mode, after sending the coordinate of the end of the first year, 1900, and ready to accept the coordinate of the *end* of the second year, 1901.

    The *Cancel Edits* (red X icon) and *Accept Edits* (green checkmark icon) commands are now available. Use *Accept Edits* to accept the data, this will update any currently open plots. *Cancel Edits* will revert the data to the previous state last accepted. The *Revert* command (blue curved arrow icon) is not currently available because the initial data series was empty. When editing a data set opened from a file, *Revert* will revert to the state when the file was opened.

    Advance the device to the end of the next year, here 1901, and send the coordinates. Continue with subsequent years (Figure 2.8) until data collection is complete. Antevs gives an audio cue, a beep, when a decade entry, divisible by 10, is made. Data entry may be paused by selecting *Record*, and continued by reinitializing the beginning of the next year to be entered (the end of the last year entered). When finished, use *Accept Edits*, then *Save As* in the *Data Series* window to save to disk.
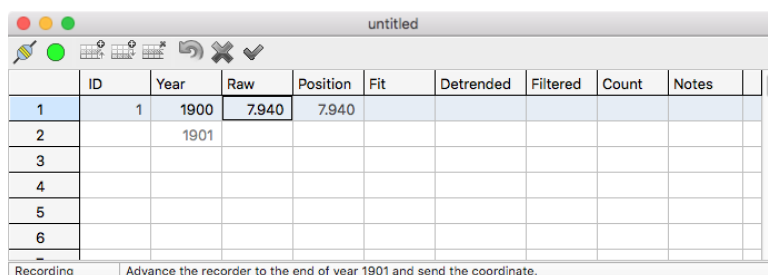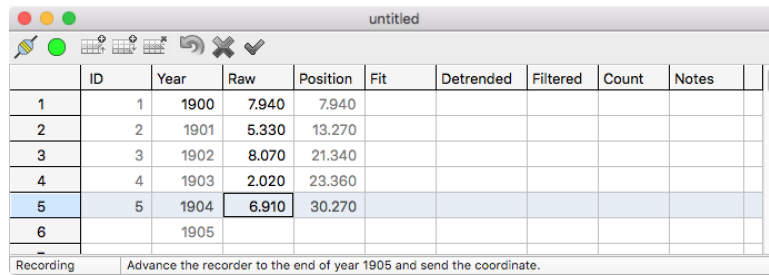
Figure 2.8: The *Data Edit* window in *Recording* mode, after sending the coordinates years 1900 to 1904, and ready to accept the coordinate of the *end* of the next year, 1905.

3   Time-Series Analysis

Time-series arise in situations where a value is measured through time, generally at discrete intervals. In natural systems annual intervals commonly occur due to seasonal variations in temperature, as seen in proglacial varves, tree-rings, and glacial ice. Antevs is designed primarily to work with such natural time-series, but can be applied to other discrete time-series as well. It is designed to aid in correlation of time-series, where an undated times-series, the unknown, is compared to a previously dated series, a chronology, to determine if they overlap in time and to date the unknown series.

The analysis is broken down into three main parts. *Detrending* is done to remove longer term trends or wavelengths, and to rescale the time-series so they can be more easily compared. A curve is fit to the time-series, and the fitted curve is removed from the data values. *Filtering* may optionally be done following detrending, and can potentially enhance correlations by using neighboring values to smooth local values. Finally, the two time-series are *cross-correlated* to determine if a match can be found between them.

Data plots can be used to visually compare the raw or detrended series, and a correlogram allows visual evaluation of the cross-correlation. Additionally, a *bin test* can be done to test the robustness of the match. Finally, a group of correlated series can be used to create a mean series, to serve as a *master chronology*.

3.1   Detrend

Detrending rescales the raw data and removes longer wavelength trends by fitting a curve to the data set, and adjusting the raw data values, y_i, using the fitted values, $\hat{y}_i$. This allows sequences with different environmental inputs to be more easily correlated, such as distal and proximal varves, or young and old tree-ring growth. The *residuals* are the data values minus the fitted (estimated) values

$$r_i = y_i - \hat{y}_i \tag{3.1}$$

where $\hat{y}_i$ is the estimate calculated by fitting a curve or function to the data.

3.1.1   Mean

Mean fits to the mean of y

$$\hat{y}_i = \frac{1}{n} \sum y_i \tag{3.2}$$

3.1.2   Linear

Linear fits a linear curve

$$\hat{y}_i = a + bx \tag{3.3}$$

by minimizing the sum of the squared residuals in y.

3.1.3   Quadratic

Quadratic fits to the quadratic curve

$$\hat{y}_i = a + bx + cx^2 \tag{3.4}$$

by minimizing the sum of the squared residuals in y.

3.1.4   Cubic

Cubic fits to the cubic curve

$$\hat{y}_i = a + bx + cx^2 + dx^3 \tag{3.5}$$

by minimizing the sum of the squared residuals in y.

3.1.5 Exponential

Exponential fits an exponential curve

$$\hat{y}_i = ae^{bx} + c \tag{3.6}$$

to the data by minimizing the squared residuals in *y* using an iterative method. This can be useful for tree-rings that have a regular decrease in width with age, but can fail in many cases.

3.1.6 Automatic

The *Automatic* method selects a curve fitting method using several criteria. For cubic, quadratic and linear curve fitting a fit is rejected if a negative thickness is predicted, and a lower order curve is tried, until the mean is used. For an exponential fit, the curve is rejected with $b > 0$ or $c < 0$, and a linear curve fit is tried. If a linear fit is rejected with $b > 0$, or a negative thickness is predicted, the data is fits to the mean value of *y*. Automatic exponential is a common choice for some tree-ring data sets (Cook and Holmes 1986).

3.1.7 Spline

The *Spline* method fits a cubic spline to the raw data. A cubic spline is constructed of piecewise third-order polynomials which pass through a set of control points, maintaining slope at the connection points (Press et al. 2007). The *Stiffness* parameter controls the tightness of the curve. This method is commonly used for tree-ring correlation (Cook and Peters 1981).

3.1.8 Fourier

The *Fourier* method fits a Fourier series to the raw data. A Fourier series is a periodic function composed of harmonically related sinusoids (Press et al. 2007). The *Low pass frequencies* parameter controls the number of components in the series. This is the default method used for varve analysis, with a default setting of 16 (Rayburn and Vollmer 2013).

3.2 Detrend Method

The *Detrend Method* sets the method by which the fit curve is used to detrend the raw data. Common settings are *Standardized* and *Ratio*, both of which give unitless values. The *Standardized* is the default setting, and is recommended unless *Ratio* is required for compatibility with other files or workers.

3.2.1 Fit

Fit displays the fit curve, $\hat{y}_i$.

3.2.2 Residual

The *Residual* method uses the residuals, subtracting the predicted value, $\hat{y}_i$, from the data value, y_i

$$r_i = y_i - \hat{y}_i \tag{3.7}$$

These are not normalized, so units are in centimeters or millimeters. The residuals will contain positive and negative values with a mean of zero.

3.2.3 Standardized

The *Standardized* method standardizes, or normalizes, the residuals by dividing by the standard deviation of residuals. The data is transformed to standard deviations from the fit curve.

$$y_i' = (y_i - \hat{y}_i)/\sigma \tag{3.8}$$

Where σ is the standard deviation. This allows comparison of differing series, as it gives unitless numbers with a mean of zero. This is the default method for varve analysis.

3.2.4 Ratio

The *Ratio* method uses the ratio

$$y_i' = y_i/\hat{y}_i \tag{3.9}$$

The ratio method is commonly used for tree-ring data. It can be biased however, and using residuals (standardized method) is more robust. Values of $\hat{y}_i$ near zero are discarded to avoid divide by zero errors. The mean of the detrended values is 1.

3.2.5 Ratio log

Ratio log uses the *Ratio* method (Section 3.2.4) and applies a log transform for variance stabilization after adding a constant of one-sixth of the series mean to each value.

3.3 Filter

The *Filter* pane in the *Preferences* dialog displays options for filtering the data. Filters are applied after detrending the data to *normalize* the data and potentially emphasize correlations. Normalizing involves smoothing the data by using neighboring values.

3.3.1 Delta bar

Delta bar smooths using the average of the nearest neighboring data points, for 3 terms

$$y_i' = \frac{1}{3}(y_{i-1} + y_i + y_{i+1}) \tag{3.10}$$

Rayburn and Vollmer (2013) suggest this filter may help emphasize some varve correlations. Figure 3.1 shows the Hudson Valley - Connecticut Valley correlogram (Figure 1.5) with the delta bar filter applied.
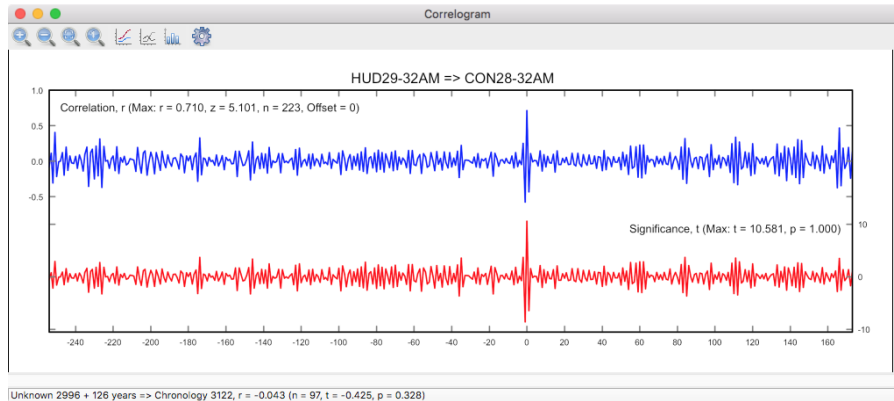

Figure 3.1: Correlogram for the Hudson Valley - Connecticut Valley varve sequences (two varve sequences (Figure 1.5) after applying the delta bar filter. In this case the correlation spike at the best matching year offset gives $r = 0.710$, $z = 5.101$.

3.3.2   Cubic derivative

The *Cubic derivative* method applies an orthogonal cubic first derivative filter to smooth using the cubic derivative of the nearest neighboring data points. This filters the data to look for changes in slope of the series using the first derivative of a local cubic polynomial (Davis 1986). The number of terms can be 5, 7, or 9. The 5 term cubic derivative filter can be particularly effective for correlation. The 5 term cubic derivative filter is

$$y_i' = y_{i-2} - 8y_{i-1} + 8y_{i+1} - y_{i+2} \tag{3.11}$$

The 7 term cubic derivative filter is

$$y_i' = 22y_{i-3} - 67y_{i-2} - 58y_{i-1} + 58y_{i+1} + 67y_{i+2} - 22y_{i+3} \tag{3.12}$$

The 9 term cubic derivative filter is

$$y_i' = 86y_{i-4} - 142y_{i-3} - 193y_{i-2} - 126y_{i-1} + 126y_{i+1} + 193y_{i+2} + 142y_{i+3} - 86y_{i+4} \tag{3.13}$$

3.3.3   Ratio

The *Ratio* method smooths using the ratio of current and previous years

$$y_i' = y_i/y_{i-1} \tag{3.14}$$

3.3.4   Natural

Natural uses the logarithmic, or natural, change from previous year to current year (Hollstein)

$$y_i' = \ln(y_i/y_{i-1}) \tag{3.15}$$

3.3.5   Relative

Relative uses the relative change from previous to current year

$$y_i' = (y_i - y_{i-1})/y_{i-1} \tag{3.16}$$

3.3.6   Log five year

Log five year applies a five year natural logarithmic smoothing filter (Baillie and Pilcher 1973)

$$y_i' = \ln(5y_i^2)/(y_{i-2} + y_{i-1} + y_i + y_{i+1} + y_{i-2}) \tag{3.17}$$

3.4   Mean Master

The *Mean* command calculates a mean master series of all open data series, either chronologies or unknowns. Each series is detrended before calculating the mean. Before creating a mean master series, the *Detrend* option should be checked in the *Preferences - Detrend* panel, and either *Standardized* (recommended) or *Ratio* methods selected. Detrending rescales the series so each is weighted equally. The data series are not normalized when calculating the mean, so the normalize filter setting does not matter.

   Note that detrending using the standardized method gives values with a mean of zero, and therefore will create negative values. The *Treat negative as missing* and *Display missing as -9* options (Section 4.5.1) must be off. The ratio method gives values with a mean of 1, it is less robust, and only recommend if it is required for compatibility, such as when using a Tucson CRN tree-ring mean chronology.

   When using either a mean master chronology, or a Tucson CRN tree-ring mean chronology to correlate an unknown raw time-series, the mean chronologies are not detrended or filtered as detrending was already done when creating the mean chronology. The *Detrend* and *Apply normalize filters* settings are therefore ignored for mean chronologies. This behavior can be overridden with the *Detrend mean* and *Normalize mean* options in the *Preferences* dialog.

4   Time Series Plots

Various plots allow the graphical, and analytical, evaluation of the time series data. The simplest is a plot of the data sets, both unknowns and chronologies, by year. These plots can be of the raw or detrended data, including any applied filters, and can the display fitted curves used for detrending. A *cross-correlation* plot compares one time series (an unknown, or undated, series) to another (a dated chronology) at successive time lags, allowing the identification of the match with the highest significance. Finally a *bin correlation* plot uses a sliding window to test how well shorter data segments of the series correlate. Plots can be zoomed in to examine the details of a correation.

4.1   Data Plot

The *Data Plot* command plots unknown and chronology thickness, or normalized thicknesses, versus year. It optionally plots best fit curves and raw data for unknowns and chronologies. This plot can be zoomed for a closer view of the data (Section 4.4), and can be used to segment the time-series if there are gaps or missing data (Section 4.5.1).

As described in previous sections, data plots allow two basic transformations to be applied to assist in correlation. First the data can be *detrended*, to remove trends that may differ between two locations, for example proximal versus distal varves, and then rescaled so the two series are more easily compared. An effective technique is to use a 16 term Fourier function as a low pass filter, with a standardized rescaling using the residuals from that curve (Rayburn and Vollmer 2013). However, numerous other methods have been proposed for tree-ring data, many of which are implemented in Antevs, e.g. exponential, spline, and ratio (Section 3.1).

Secondly, it may be useful to *filter* the original data using a function based on the neighboring data points (Section 3.3). This has similarities to, for example, blur and sharpen filters applied in image software such as Adobe Photoshop and GNU Gimp. The utility of these filters is best determined by trial and error, although the delta bar filter has been found useful for varves. Other common filters, including those proposed particularly for tree-ring studies, are also provided.

Section 4.5 describes how the unknown data can be segmented to assist in reconstructing cores with missing segments. This can be a problem when analyzing tree-ring cores with rotted sections, or varve cores with destroyed sections. By sliding individual sections of the core relative to one another, it may be possible to reconstruct and correlate the series.

4.2   Cross-Correlation

Cross-correlation is a technique for comparing one time series (an unknown, or undated, series) to another (a dated chronology) at successive time lags, thus allowing the identification of the match with the highest significance (Davis 1986). This allows the objective time correlation of two cores, and is the foundation for automated varve and tree-ring correlation techniques. For each of the offsets, Antevs calculates a correlation coefficient, r, which varies from -1 to +1, where +1 indicates a perfect correlation. A t value is also reported, which gives an approximate test of the significance of the correlation, the t value also takes into account the number of data points used in the calculation (Davis 1986). The offset is reported, which is the offset of the unknown series. The offset for the highest t value, is the most likely correction that should be applied to the unknown years.

As the two series are compared by applying a lag, or offset, the number of overlapping data points goes to zero. As a small number of data points is unlikely to give a significant correlation, a minimum overlap is set to avoid this. The default minimum overlap required for a test is 50, this can be changed in the *Preference* dialog *Limits* panel. The *Cross-correlation* command applies the current filters, first detrending then normalizing, and calculates the cross-correlations and t statistics for each possible match.

4.3   Bin Tests

Two bin or windowing correlation tests can be used to test the robustness of the cross-correlation, and to see if the correlation changes through time. A bin correlation test sequentially groups the unknown data into bins, and runs each bin against the selected chronology. A sliding correlation test uses a window of a set radius that slides along overlapping chronology and unknown segments to look for changes in the correlation through time. The settings for these can be changed in the *Preference* dialog *Limits* panel.

4.3.1   Bin Correlation

A bin correlation test sequentially groups the unknown data into bins, or subsamples, and runs each against the selected chronology displaying the best correlations based on the *r* and *t* statistics. This is used to test the robustness of the match and to potentially locate missing varves or tree-rings in an unknown. For example, a marked change in the offset at a given point in the time series may indicate a missing varve or tree-ring at that point. The bin size can be changed in the *Preferences* dialog *Limits* panel. The bin size must be larger or equal to than the minimum overlap.

4.3.2   Sliding Correlation

A sliding correlation test may be done after an unknown data sequence has been matched with a chronology. A sliding bin, or window, of a set size, given by its half-width, or radius, about an unknown sequence data point, is correlated with the same time window in the chronology sequence for each year. The *r* and *t* correlation statistics are calculated from the beginning of the overlapping time sequence plus the window radius, to the end of the overlapping time sequence minus the radius.

This test can show changes in the correlation through time, potentially reflecting local changes in input or environmental conditions between the two sample localities. The test is selected by checking the *Sliding* checkbox, and setting the value of the *Sliding radius* in the *Preferences* dialog *Limits* panel.

4.4   Data Windowing

The *Data Plot* can be zoomed in to a portion of the time-series allowing more detailed examination. The *Zoom Window* command (green square icon) zooms in on a segment of the time-series (Figure 4.1). The length of the segment is set in the *Preferences* dialog *Limits* pane using the *Data window width* setting. Use the *Pan Left*, *Pan Right*, *Pan First*, and *Pan Last* commands to move the window. The pan step size can be set in the *Preferences* dialog *Limits* pane using the *Data window step* setting. The *Zoom Window* command toggles between the zoomed in window and the full time-series window.

The *Pan Left* and *Pan Right* icons (green arrow icons) also pan the data window, and holding down the *Shift* key changes these to *Pan First* and *Pan Last*. Additionally, the bracket keys ("[" and "]") are provided as shortcut keys for *Pan Left* and *Pan Right*.

4.5   Missing Values and Segments

A common problem in natural data sets such as varves and tree-rings is missing data points, as cores or observable varve sequences can be broken or discontinuous. Antevs treats missing data points as NaN (Not a Number) values, which have no numeric value. There can be any number of NaN data points in a sequence. In order to deal with this internally and avoid numerical errors, Antevs uses linear interpolation over the NaN values. This is, for example, the default used by MATLAB to deal with NaN values in numerical analysis.

Breaks can be inserted into the unknown time-series, allowing individual segments of an unknown time-series to be interactively shifted in time by incrementing or decrementing them. The segments
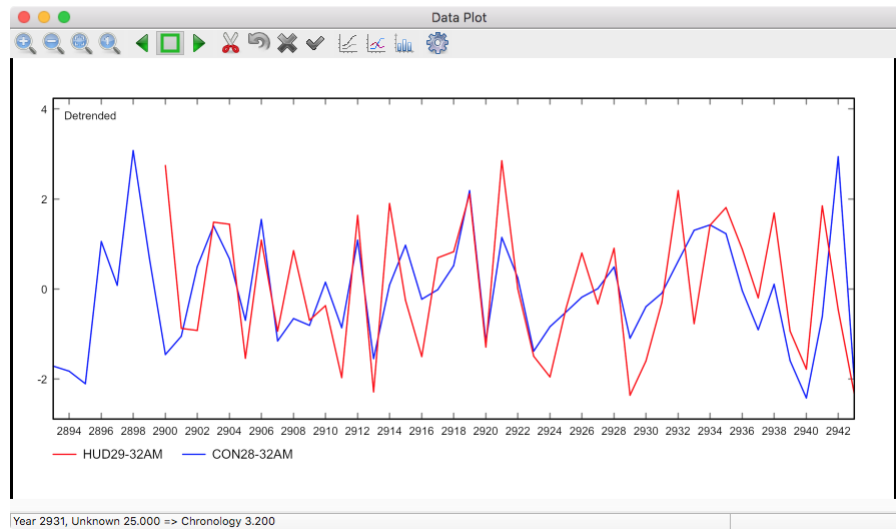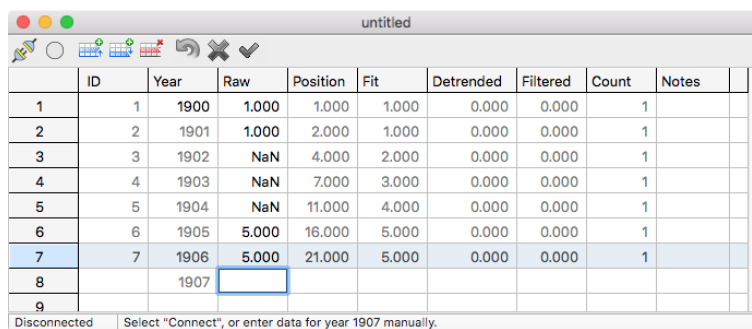
Figure 4.1: Zoomed in *Data Plot* window showing details of the detrended Hudson Valley - Connecticut Valley varve data (Figure 1.4), using the *Zoom Window* command.

can be individually compared to the chronology, and slid forward or backwards to match. Additionally, the cross-correlation can be automatically calculated at each step to test the strength of the match.

4.5.1   Missing Values

Antevs treats missing data points as NaN (Not a Number) values, which have no numeric value, and there can be any number of NaN data points in a sequence. In order to deal with this internally and avoid numerical errors, linear interpolation is used over the NaN values. This is, for example, the default used by MATLAB to deal with NaN values in numerical analysis.

To maintain operability with time series that may contain negative values (such as temperature), there are two settings to control how NaN values are treated. These are set in the *Preferences* dialog *Options* panel. If *Treat negative as missing* is checked (by default it is not), any negative number is treated as a NaN. The second option, *Display missing as -9*, controls how the NaN is displayed in the *Data Edit* window and output files. This is also off by default, however, a common convention in tree-ring data is to use "-9" to signify a missing value. If that is important, turn this option on, and "-9" will be displayed instead of "NaN" in the *Data Edit* window, and in saved files. Note, however, that treating negative values as NaN will create problems when creating a mean series using standardized detrending. Standardized detrending produces a series with a mean of zero, and will contain negative values.

Figure 4.2 shows a simple example of a time series with NaN values in the *Data Edit* window, and Figure 4.4 shows the *Data Plot* of this series. The three NaN values between 1901 and 1905 are interpolated. NaN values van be entered in the *Data Edit* window, or in input files, as "NaN", or "nan", or, if *Treat negative as missing* is checked, any negative number.

4.5.2   Segmented Series

A common problem in natural time-series is missing data, caused, for example, by broken or incomplete cores. As described in Section 4.5.1, NaN values can be inserted to break a series into individual segments, however inserting NaN values manually can be cumbersome. Antevs therefore provides an interactive segment editing mode. The series can be broken up into any number of segments, and the segments can be shifted in time to locate a better match.

To begin segment analysis, open a time-series as an unknown. Segment analysis can only be done on one unknown time-series at a time, this will be the first unknown series selected if multiple unknowns are selected. Next select the *Edit - Edit Segments* command in the *Data Plot* window to

Figure 4.2: *Data Edit* window with simple example of a time series with NaN values. The NaN values between 1901 and 1905 are interpolated linearly.
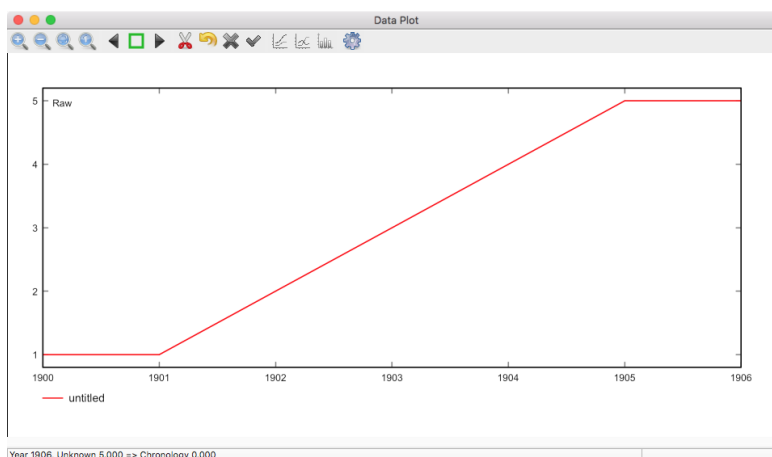

Figure 4.3: Data plot of the time series in Figure 4.2. The NaN values between 1901 and 1905 are interpolated linearly.

go into segment edit mode. All data points will be marked with a white circle. To begin editing, click on the symbol of the data point at the start or end of a segment that you wish to shift. The segment start point will then be marked with a red circle. Shifting the point to the right (to decrease its age) or left (to increase its age) will create a break by inserting a NaN data point, which is marked by a small red circle. Once a segment is created, the segment endpoint will be marked with a yellow circle. The symbols are displayed in the *Raw*, *Detrended*, and *Filtered* curves, any of these can be used for editing. The default symbols can be changed in the *Preferences* dialog.

To shift a segment, use the *Increment Segment* and *Decrement Segment* commands. The *Right arrow* and *Left arrow* keys are shortcuts, and holding down the *Shift* key will increment or decrement by 10. Moving one segment will insert or remove NaN data points and shift the segment. For example, if there are three segments, 1, 2 and, 3 from left to right, shifting segment 2 to the right will add NaNs to its left, and remove them to its right. Once all NaNs to the right are removed, segments 2 is joined to segment 3. Selecting the first or last point in the series will shift the entire series.

Mistakenly reversed segments may occur during data collection, for example during the collection of broken cores. The *Reverse Segment* command may be used to reverse an individual segment to look for a match.

If open, the *Correlogram* window will update and show the correlation values. Use the *Accept Edits* command to accept the edited values. The *Cancel Edits* command will revert to the state when *Accept Edits* was last called. The *Revert* command reverts to the state on opening the file.

Note that the *ID* field is used to give each entry a unique identification number. When shifting segments new data points are added as NaN values, and these will be given the next available highest *ID*. The *ID* can be any integer value, such as the initially estimated year sequence, or a series starting with 0 or 1. When reading in a data file, an *ID* value will be assigned if one is missing. This allows the *Year* field to be modified, as the *ID* is not modified when renumbering or inserting missing values. When deleting, splitting, or combining rows, or editing segments, the next available unique
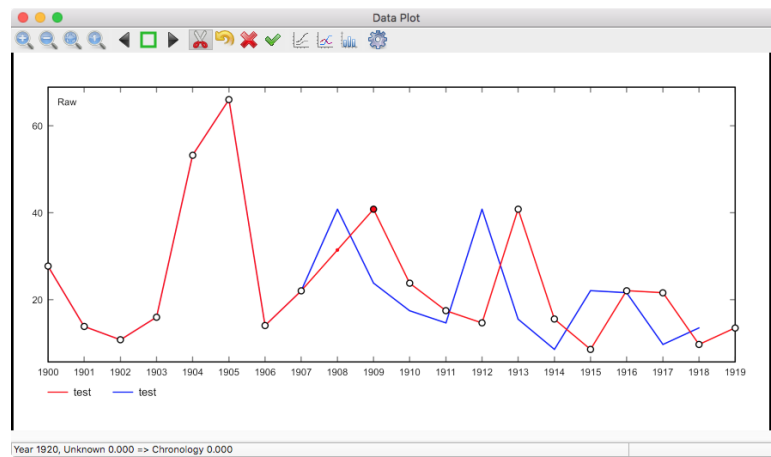
Figure 4.4: *Data Plot* window with a segmented unknown (red "test") compared to an identical chronology (blue "test"). The unknown has two segments, the second, starting at 1909, as indicated by the red circle , has a one year offset from its original start position at 1908. A NaN value has been inserted at 1909, as indicated by the small red dot.

ID number is assigned. The *Set ID from Year* command in the *Data Edit* window will set the *ID* field to the *Year* if desired.

Acknowledgements

References

Antevs E (1922) The recession of the last ice sheet in new england. American Geographical Society Research Series 11:1–120

Baillie M, Pilcher J (1973) A simple crossdating program for tree-ring research

Cook E R, Holmes R L (1986) Guide for computer program ARSTAN. The international tree-ring data bank program library version 2(16):75–87

Cook E R, Peters K (1981) The smoothing spline: a new approach to standardizing forest interior tree-ring width series for dendroclimatic studies. Tree-Ring Bulletin 41:45–53

Davis J C (1986) Statistics and Data Analysis in Geology. New York: Wiley, 2 edition

Grissino-Mayer H D (2001) Evaluation crossdating accuracy: a manual and tutorial for the computer program cofecha. Tree-Ring Research 57(2):205–221.

Press W H, Teukolsky S A, Vetterling W T, Flannery B P (2007) Numerical Recipes: The Art of Scientific Computing, 3rd edition. Cambridge University Press, Cambridge, 1235 p. Press, W.H, Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P., 2007. Numerical Recipes: The Art of Scientific Computing. Cambridge: Cambridge University Press, 3 edition

Rayburn J A, Vollmer F W (2013) ANTEVS: a quantitative varve sequence cross-correlation technique with examples from the Northeastern USA. GFF 135(3-4):282–292, ISSN 11035897

Ridge J (2012) The north american glacial varve project. Technical Report http://eos.tufts.edu/varves, The National Science Foundation and The Department of Earth and Ocean Sciences at Tufts University, Medford, Massachusetts

Ridge J C, Balco G, Bayless R L, Becks C C, Carter L B, Dean J L, Voytek E B, Wei J H (2012) The new north american varve chronology: a precise record of southeastern laurentide ice sheet deglaciation and climate, 18.2–12.5 kyr bp, and correlations with greenland ice core records. American Journal of Science 312:685–722

This page intentionally left blank