PRECONK

QUICK HELP FOR PRECON NOW CALLED PRECONK VERSION 5.11

(UNTIL I GET A MANUAL TOGETHER)

****

Harold C. Fritts

October 10, 1996

This manual assums that INSTALL.BAT has been used to install all files in directory C:\PRECON.

If you downloaded these files, to extract the files manually, enter the following command followed by a RETURN: PKUNZIP ARCFILE.ZIP *.* where "ARCFILE.ZIP" is the name of the archive file. Later, if a file becomes corrupted, you can retrieve it from the ARCFILE by substituting the name of the file for *.*. I recommend that you put all programs and related files in the same directory. However you can run PRECON from any directory if PRECON.BAT and PRECON32.BAT files are placed in the root directory or in a directory included in the PATH Statement. For those who purchased all of the U.S. climactic data, you can place the contents of each *.zip file on a separate floppy disk or directory. Remove earlier versions of PRECON.EXE or PRECONK.EXE before installing the new one. The remaining files are required to run version 5.11.

Climatic data are available for the entire U.S. and selected European stations. I include only the Arizona and European data in this packet. Data for the rest of the 48 contiguous United States can be obtained at cost.

To set up Graph-in-the-box Analytic (GBA) enter: GBASETUP and follow the menu to choose the monitor and printer. Do not change the color code in the GBSETUP program until you have thoroughly tested the results. It is difficult to predict the colors for different setups. If you are unhappy with the existing combinations, the color can be changed in the setup program, or the colors can be changed while running GBA by depressing the F6 and F7 keys. The latter change is not permanent.

Enter: GBA to put the graphics program into memory (The *.BAT files will do this automatically). ALT-G will call up the program when you are in DOS. It can be called from within PRECON also. PRECON will write files that GBA will recognize and you can use these to make plots. When finished, enter GBKILL to remove GBA from memory.

1) SELECTING AND READING DATA (INITIALIZATION)

Enter: "PRECON" or "PRECON32" which ARE the nameS of the BAT files that load GBA, runs PRECON16.EXE or PRECON32.EXE whichever version you have, saves the INI file if you requested it and reruns the program unless you exit by entering "X". If the BAT file does not run properly, then try upgrading to DOS 6.0 or higher.

PRECON prompts and offers default answers at every step in the Main Menus. It does not remember parameters within submenues such as in the Change Menu or while running Response etc. Enter your choice or simply use a "RETURN" to select the default.

The program pauses after the initial message. Enter a "RETURN" to continue. The program then offers four options. (1) is used to run the program with default prompts; (2) uses the programed defaults but also learns any changes you make; (3)

reads the defaults saved in options 2 and 3 and saves any changes made to them; (4) bypasses input parameters and starts with a previously saved file. You must have used the save option on an earlier run to use option 4. Answer "1" if you are new to the program. The default is "3" which will be convenient later when you are running a number of similar analyses. Use 2 if you want to run with the defaults and changes you make to them and save the commands in PRFILNEW.INI. Whem you exit PRECON within the BAT file, it will copy this new INI file to PRECFILE.INI. If you run the exe file outside the bat file, the program will create a file called NEW but not copy the PRFILNEW.INI file to the PRECFILE.INI. In this caes, if you want to use this file the next time you run, it is necessary to copy PRFILNEW.INI to PRECFILE.INI.

The program opens a file with a name derived from the chronology specifications. This file name ends with an "OUT" extent and must be present to run option (4) mentioned above. The OUT file contains a record of all choices, information for determining errors, the input and output data, and considerable detail about the analysis including basic information not shown on the screen. If you have any problem or forget what you did, look at this file with a word processor. Save this file on a separate disk for a permanent record of your work. If you have major problems, please send me a copy of this file with as complete description as you can make of your setup. I am providing the PRECON disk at less than cost because I need help in identifying all of the bugs. If I hear nothing, I will assume you have had no major difficulty using the disk and do not want updates.

The next prompt asks for the first and last year in the analysis. These dates are only the outermost boundaries of the 100-year interval in which you choose to work. It is not necessary to identify the actual years of the particular analysis. If a set of tree-ring and climate data spans fewer years than those defined by this block, the program finds the common period in the two data sets. However, all data outside this block of years will be excluded. If the period starts before your chronology, you will not have prior growth values available. If you want to include 3 years of prior growth for such short chronologies, then select the period to begin 3 years later than the first year of the chronology. In most cases the chronology is much longer and it is necessary only to select the 100 years spanning the climatic data. I recommend that you always select a 100 year period, except if you want to purposely exclude prior growth values from being available when you run PRECON.

Next, select one of four tree-ring chronology sets or indicate you will import files in the International Tree-Ring Data Bank format. For the latter option there is an opportunity to add a constant (add 1.0 if the indices have a mean of zero). Next, there is an opportunity to change the read format. In addition to the format, the program expects three lines of header information which will be displayed but ignored in the analysis. BE SURE THESE THREE LINES OCCUR IN ALL THE TREE-RING DATA SETS. In addition, place each chronology in its own file. Program ARSTAN produces three different chronologies. Separate these into different files and be sure to add three lines before the first data to be read.

For options 1-4 the computer scrolls through a summary listing of all chronologies in the selected set. Continue scrolling through this list by choosing the default (entering the RETURN key). You can scroll down through this list and return to the top as many times as you wish. When you identify the chronology to select, stop the scroll with an "N". The program prompts with a default chronology input. Indicate your choice by entering the 3 digit site code and 4 digit species code in UPPER CASE letters. This

is the only question that requires upper case input. The computer matches these letters with those of the table and there must be an exact match. If no match is found, the program returns to the previous prompt thus providing another chance. A map of chronologies 1-102 in Data Set 1 is included in file USTRMAP.PCX. This map can be displayed in the Windows Paint program or any other similar program capable of reading a PCX format.

For data sets 1-2 the next prompt asks whether you want residual or standard chronologies. If a residual chronology is not available, a message will indicate that fact and the standard chronology will be selected. (The residual chronology has been ARMA modeled to remove all auto-correlated relationships.) For data set 4, there are five chronology types to select from: RING WIDTH, EARLYWOOD, LATEWOOD, MAXIMUM DENSITY and MINIMUM DENSITY.

Next, select the number (1 or 2) of climatic variables to be read. Temperature, Precipitation or Palmer Drought Severity Indices (PDSI) are calculated using a Thornthwaite water balance to estimate the presence or absence of drought. PDSI values are available only for the North American data sets. PDSI for at least March through October must be selected to run the ZAHNER model offered at a later prompt.

For sets 1 and 2, choose the state code number used by NOAA. These are listed on NOAA data sheets and the codes for western U.S. are displayed on a map, which is provided in file CLDIVMAP.PCX or a hard copy can be provided if you wish. For example Arizona is the second state in the alphabet so its number is "2". For New Mexico it is "29". You then must decide which climatic division within the state is appropriate for the chronology that was selected. Locate the chronology and check it against the map of the divisions. If I can get some programing help I will try to produce these maps in the program. Until then, use these PCX files or the hard copy maps.

The next question asks for the drive which contains the climatic data. For chronology sets 1 and 2, it is necessary to enter the 2-character Postal abbreviation of the state (AZ for Arizona, CA for California). This is required to construct the file name of the correct climatic data set.

The next two prompts describe the interval of months to be selected for analysis of the growth response. First, indicate the number (within the year) of the first month to be used. For example to start with the month of June, select "6" or for August select "8". Second, indicate the number of months to be selected from each data set. For a period starting with June preceding the year of growth and continuing through July of the following year (year of growth), enter 14. If two variables are to be used, the same period will be applied to both. In the 14 month case, variable numbers 1-14 will be assigned to the first variable (either Temperature or Precipitation) to represent the months June-August while numbers 15-28 will be assigned to the second variable (either Precipitation or Palmer Drought Severity Indices (PDSI). Keep this in mind when you are selecting these variables later in the analysis. Choose only one continuous interval for any one analysis. For periods within one calender year the prompt will ask whether the data are for the year before the tree-ring date or for the same year. In most cases the latter is appropriate.

For chronology set 4 from Europe, select SWISS or SCANDINAVIAN climatic data depending upon the location of the chronology that was selected. A list of temperature stations are displayed. Stop the scroll by entering 'N'. The programmed default is the nearest station. Enter a RETURN to select that or enter the 5-digit Identification number for the climatic station you want.

The program then displays the data for the selected station and asks for confirmation as to whether it is correct. With a 'Y' answer or a RETURN the program searches for the desired climatic data set. If a second variable is indicated, the program asks: 'Do you want to use Matching Climatic Data' (data from the same climatic station). A 'N' will allow you to select a different station in case there is one closer to the data set. I have tried to build the maximum flexibility in choosing climatic data sets.

2) THE CORRELATION MENU

Now the data are read and the action begins!!!! Simple correlations and a stepwise multiple regression analysis are calculated. The first menu deals with the correlation analysis.

Options in this menu can be selected in any order except for 6. Option 4 will exit from this menu for more advanced analysis. Each selection is numbered, the number written in the *.OUT file as a permanent record and past selections are displayed in the Menu to remind you of past choices. The names of the graphics files will include the sequence number of the step in which they were produced.

Options 1 and 2 write output files that are read by the graphics program to display plots. The file names are based on the Site, Species, Type of chronology selected and Menu step. Information on the data source is included in the headings of the plots. Option 2 makes a scatter-gram plot of one variable on the x axis against two variables on the y axis. Look at the variables using option 3 and then choose those you want to plot in 2. The independent variables can be plotted against one another or against the Actual, Estimate or Residual of the dependent variable.

Option 3 displays the values of the simple correlations along with the means, maxima and minima of each monthly climatic variable used in the analysis.

Option 5 will save the important information in a file in case it is necessary to exit at this time. This file can be read at a later time and you can continue with the analysis. Be sure to keep the .OUT file for this run. When you read this file in it will open the .OUT file and begin writing on it.

Option 6 terminates the program without saving a file.

Option 7 temporarily returns to DOS so that the Graph-in-the-box Analytic program (GBA) can display the plots (This Option is not needed for PRECON32.EXE. To see the plots enter "ALT G" at any time.) First enter "ALT G"(to activate the graphics program) followed by "9" ("ALT G" can be used after entering 9 if you forget to enter it earlier, then enter a DOS command such as "PAU" to "wake up" GBA and display the results. If you prefer, wait until you have finished all analyses, save the file, exit the PRECON program and call up the program with an ALT G. If your computer memory does not have room for both PRECON and GRAPH-IN-THE-BOX, use PRECOWIN.BAT at the beginning instead of PRECON.BAT or edit PRECON.BAT not to put GBA into memory until you exit PRECON. Then enter GBA followed by ALT G to wake it up. After finishing with GBA, or on exit from PRECON.BAT with a Contorl C, ALWAYS enter GBKILL to remove GBA from memory. If you have saved the file before leaving PRECON you can return to where you saved the file using Option 4 on the first prompt after you start PRECON again. This option reads the file you had saved and resumes the analysis at that step. You must not delete the file containing the record of your analysis which has an extension *.OUT if you want to use Option 4. This OUT file can be examined and

edited with a word processor to retrieve data from the analysis in ASCII format.

3) RUNNING PROGRAM GRAPH-IN-THE-BOX ANALYTIC

The first screen appearing in GBA is in "Capture Mode". This mode is never used a PRECON analysis. Exit this mode by entering an "ESCAPE" to activate the menu. (If you become lost in GBA just enter an "ESCAPE" and it will return to this menu.) Choose Files by entering "F" or moving the cursor to the Files position and then use the space bar to toggle through the available options. Select the "Load" option. (The program remembers whatever option is selected as long as it is in memory.) Then move the cursor with the right arrow to "Name". To display the list of names, enter "F1", and then select the desired file for plotting by moving the cursor to the file name with the right and down arrow. Load that file by pressing "Enter". Then press "S" to show the plot. After you have examined the plot, enter an "Escape" to return to menu and "P" to print.

If an error is indicated, there may be a problem with the grid size that PRECON entered for the plot. To correct the error, toggle F9 to display the data in the file. Toggle F9 again to show the options and headings. You can toggle back and forth using F9. Use arrow keys to move around these screens. Find the scale for the LEFT Y and RIGHT Y axis. Enter a larger value for the Grid line step or toggle from "Manual" to "Automatic" and enter F10 to display the plot. If the error is still present, go back to F9 and experiment with larger Grid line steps. The original file will not change unless you enter "Save" in the File Menu without changing the file name. The original file can be read again if things become mixed up. If you don't want a particular graph to appear in the plot, you can select the "COLOR" option for that particular variable and toggle with the space bar until "NOCOLOR" appears.

Use F7 and F8 to change the color combinations of the plot or change the number for color of individual variables in the F9 options as mentioned above. Line styles and the information appearing on the plot can be altered to change it's appearance.

Enter ESCAPE to return to the Main Menu; select "P" from the Main Menu to print the plot. Within the print menu use the space bar to toggle through the options and the arrows to move the cursor to different items. One, two or three plots can be made on the same page depending upon the size selected. Toggle to send a line feed if it is needed for your setup.

To save changes, toggle "SAVE" in the files menu. Return to the Files menu to choose a different plot, etc.

RECOMMENDATIONS FOR USING HP LASER JET SERIES III
FOR PRINTOUT GRAPHS WITH PRECON AND GBA
by Jaroslav Dobry
Forest Sciences Dept. UBC
270-2357 Main Mall
Vancouver, B.B.
Canada V6T 1Z4

(1) Select printer in GBSETUP: HP LaserJet II.

(2) Switch the printer OFF (direct computer-printer connection is recommended)

(3) Remove (pull out) the PostScript Cartridge

(4) Switch the printer ON

(5) "On Line" button should be ON

(6) To print use the command PRINTOUT (button menu).
Select size = "SMALL SIZE", "MEDIUM SIZE", OR "LARGE SIZE" and unit = "PRINTER", then enter print.

Note: For more than one picture per page, SIZE must be adjusted to SMALL or MEDIUM (Step 6). In case of "LARGE SIZE" the printout is automatically done in Landscape Mode.

(7) (A) To print one picture per page: After the printed picture disappears from the screen, switch "ON LINE" button on printer to OFF, then the "Form Feed" button on printer to ON. The printer should start to print. After the printout is finished switch the "ON LINE" button again to ON. The system is ready to print another picture (Return to step 6).

(B) To print more than one picture on one sheet of paper: Leave the "ON LINE" button ON (Return to step 6).

END OF RECOMMENDATION

4) THE REGRESSION MENU

When you are finished looking at simple correlations and have made some plots, exit from the Correlation Menu using option 4 (Other Models). This brings up the second main menu, which is similar to the first.

Option 1 in this menu makes a plot of the partial regression coefficients.

Option 2 makes a time-series plot of the of the (ACT)ual index measurements, (EST)imates from the regression and the (RES)iduals, which are the differences between the actual values and estimates. Variable CLC is the climatic change selected in the Change menu described below and TRC is the growth estimated by adding the climatic change to the actual climatic data. These variables have values of zero unless the Climatic Change option has been activated.

Option 3 is the same as option3 in the CORRELATION MENU.

Option 4 includes a sub menu that can be used to examine some of the results and to make a variety of changes.

"S" shows the correlation coefficients again.

"R" shows the Regression Coefficients.

"F" allows changing of the F-level.

Choose "Q" to quit and run the Stepwise Multiple Regression (MLR) over again.

"B" is used to identify a sub period for analysis, which is less than the length of the existing record. To recalibrate the regression relationship for a particular sub period, select the sub period, then exit with a "Q" but answer "Y" to the next question (Do you want to run Stepwise Multiple Regression?). Then return to Option 4 again and change the sub period back to the entire period, but exit by answering "N" to running the multiple regression again. Whenever the regression is run, new regression coefficients are obtained using the sub period you have chosen. Instead of calculating the coefficients all over again, select option 5 to calculate the estimates and residuals using

the existing coefficients. Then choose option 2 to display the new time series. This combination can be used to assess the potential effects of Air Pollution and similar factors.

"C" is used to add (or subtract) a change in climate (CLC) from the existing climatic data and to calculate the estimated growth due to the change (TRC). The time-series plot for this option shows the change in climate (CLC) that you selected and the growth calculated by adding that change to the actual climatic data (TRC) and applying the coefficients obtained by multiple regression or by the response function selected at a previous step in the analysis. These values are zero and are plotted along the 0 axis when the climate change option is not selected. Try this option. It should be fairly easy to understand. It is not necessary to recalculate the regression or use option "2" to replot, that is done automatically. You can delete the change by entering option "D".

There is space for only 40 transformations. Option M can be used to make your own transformations. The relevant portions of the manual to the Multiple Regression Program (Appendix 1) are included to help if you want to use this option.

Option "A" assists you in making the transformations necessary to test for interactions. It takes all variables that are significant in the regression, calculates averages, then calculates the squares and then all cross-products. The first entry in this option asks how many groups of variables are to be averaged. (Try "1" first.) Then the program takes each group and asks for the number of items to be averaged. Select the variables by entering their numbers after the prompt. Select up to 10 variables to average. For example, if the significant variables in regression are 16,18,19 22, 23 and you want to make an average of the interval including them, enter "1" to indicate one average is to be produced, "8" on the second prompt for eight climatic variables and "16 17 18 19 20 21 22 23" for the following prompt. This will take 8 of the 40 lines available for transformations, so you must be careful not to be too ambitious. In this example, no variables would be left for automatic transformation.

After making the desired averages, transformation codes are generated that will be used by the MLR program to 1) move the dependent variable into the last position, column 40, in the array, 2) sum the variables selected for averaging, 3) divide the sums by the number in the sum, 4) move all remaining significant variables into positions 1, 2 to n in the array, 4) calculate the squares of all significant variables and move them into the next n + 1 positions, 5) calculate the cross-products of all combinations of significant variables and move them to the next array positions, 6) move the averages into the next n + 1 to n + n' array positions and, 7) move the dependent variable from array position 40 into the last variable used in the regression.

If no averages are made, up to 7 significant variables can be transformed with the automatic option. If there are more than 7, some interaction terms will not be computed unless you average some of the variables. If more than 7 variables are significant, then reduce the number by combining several into one or more averages, or increase the F level to reduce the number of significant variables and running the regression over again. Keep increasing the F value until the number of significant variables is reduced appropriately.

The averaging of variables uses some of the 40 lines available for the transformations so check the transformations that are generated. The transformation codes are shown on the screen and written in the *.OUT file and hopefully the computer will warn you if it is likely that the automatic transformation was cut

short. If all transformations are generated, the last items in the code will be the cross-product of the last two linear variables, any averages that were generated and the final move of the chronology values to the last variable of the transformed data set. If you are not confused by now, you have not been reading carefully! At least try the A (Automatic) transformation option and check the code written in the OUT file with the information in Appendix 1. You might be able to understand what is happening. I have not had time to make the Manual transformation code user friendly, but I hope to do so.

After producing the transformation code, an opportunity is provided to edit them. I haven't worked with this part of the program very much, as I have spent most of the time making the other selections as automatic as possible. Select the defaults then enter "Y" to calculate the Multiple Regression. This uses the code you generated, to move and transform the original data into a new data set which is used to recalculate the multiple regression. The program also tries to label the variables that were transformed. Only one transformation can be made in one run. The plots from Option 1 in the main menu are no longer intelligible but the other options can be run again.

When choosing "Q" to quit the Change Module you will always have the opportunity to recalculate the regression coefficients and change the F level.

Once the automatic transformations have been made, you can use option 10 to display the form of the interactions.

Option 10 lists the names of the significant variables after transformations have generated new regressions with interaction terms as follows:

SEQ NO	VAR NO	ORIGINAL NO	VARIABLE NAME	NO	VARIABLE NAME	TYPE	COEFFICIENT
1	3	18	PREAV1	3	NEWV	Linear	1.08223271
2	6	18	PREAV1	18	PREAV1	Product	-0.15638722
3	7	4	TEMOCT	12	TEMJUN	Product	-0.00747843

It will also list the pairs of variables and show the coefficient terms for them. Select the number of pair that you wish to plot. If there are interactions, this option will plot them in a sort of 3-D plot on a 2-D surface.

The program allows for rotating the plot, that is to switch the x and y axis. If the first plot is not clear, use this option again with the same variables, but answer "Y" to the Rotation prompt.

Option 5 in the REGRESSION MENU recalculates the estimates and residuals over the period selected using the last coefficients that were calculated by MLR or by the Response Function.

Option 6 calculates growth using Palmer Drought Severity Indices (PDSI)[available only for the USA] and the algorithms described by Zahner, R. and C.E. Grier. 1990. Concept for a model to assess impact of climate on the growth of the southern pines. In: Dixon, R.K., R.S. Meldahl, G.A. Ruark and W.G. Warren (eds). Forest Growth: Process Modeling of Responses to Environmental Stress. Timber Press, Oregon. pp. 383-392. They assume that the tree responds only to PDSI values and weight the monthly PDSI values by the relative amount of growth expected to occur during that monthly period. I have extended their program to cover trees at higher elevations in Western USA using some of my tree-growth measurements (see Fritts et al. 1991. Climatic variation and tree-ring structure in conifers: empirical and mechanistic models of tree-ring width, number of cells, cell size, cell-

wall thickness and wood density. Climatic Research 1:97-116. for more details regarding this and other PRECON options).

Option 7 saves current information in a file.

Option 8 exits the program.

Option 9 is used with ALT-G will temporarily exit the program so that GBA can be run. Follow directions and enter "PAUSE" after entering ALT-G and 9. This wakes up GBA.

Option 11 is not operational in this version.

Option 12 calculates a Response Function. Up to three years prior growth can be analyzed (rotated) along with the climatic variables. THIS OPTION IS THE ONLY WAY TO CREATE PRIOR GROWTH VARIABLES. These new variables will be retained for all subsequent analyses.

I use the RESBO subroutine written by Joel Guiot (See Appendix 1) with bootstrap methods to obtain reliable error estimates. It calculates k different response functions by sampling the data n times using random sampling. Then for each case the data points that were not selected were sampled n times and applied to the coefficients of the response function to provide independent verification statistics. Intermediate response function output data are generated in the *.OUT file. You can look at this file for such details.

The response function solution is very sensitive to the number of variables and the period used in the analysis. Be cautious about interpreting differences between response functions. They may simply represent statistical variability. In the older versions of Response the numbers of variables that were significant were often very large especially when only one or two principal components were entered into regression. The errors were underestimated. This version gets around some of these difficulties. However, beware of interpreting results that explain small amounts of climatic variance and for which the independent statistics are low (below 30%). If a sub period is chosen before entering the Response option, the response will be calculated from the sub period. If the sub period is too short, the computer will tell you of this fact. A response function plot is automatically produced, but option 2 must be selected to view the new estimates from the response. The response function coefficients can be used for the options available in the Change Sub menu.

Option 13 shows statistics that compare the 6 series entered in the time-series array. It also allows exchanging data sets in this array. When reading input for version 2 of PRECONK, a prompt asked whether a second chronology was to be read. Option 13 is used to compare these 2 data sets with each other and with other data in the array. The first chronology is placed in column 1 and the second is placed in column 6 of the time-series data set. Column 2 contains the estimates from the actual data in column 1, column 3 contains the residuals ,column 4 the climatic change and column 5 the calculated growth from the change in climate. The second chronology data set in column 6 is not used in any analysis unless this array is exchanged with another data set included in columns 1-5. The EXCHANGE option allows this to be done. In addition, the EXCHANGE option calculates some comparative statistics of the data in each column of the time-series data set.

For example, switch the second chronology read with the first by exchanging columns 1 and 6 or exchange the estimates in column 2 with the actual data in column 1. This allows one to look at the

estimates, residuals and to perform additional analysis. I am not through with this option, so the labels of any plots will not correspond with the new variables made with exchange. Experiment with this option if you wish, but for most analyses, it may be easier and less prone to error if you simply run PRECON a second time to analyze the second data set. This option is useful if you suspect that the residuals may still contain useful information. Simply exchange the residuals (column 3) with the original data (column 1) and perform further analysis. I use this option to read in the results from our mechanistic model, compare them to the results from the PRECON model, and then to run them through Response.

Option 14 (Kalman Filter) prepares a file that can be read by program KALMAN. I have modified Ed's Kalman Filter program to run on a PC and to take an input file prepared by program PRECONK. To generate the input file for KALMAN, simply choose the Kalman option in PRECONK and indicate which data are to be selected for input to KALMAN. The program uses the transformation option in the stepwise multiple regression program so data can be combined into up to 10 averages of up to 10 monthly variables. WHEN YOU EXIT YOU MUST RUN THE MULTIPLE REGRESSION TO MAKE THE TRANSFORMATIONS. I suggest selecting all variables so that the program will write all monthly climate variables and prior growth variables that were not combined into averages. A subset of these can be selected while running KALMAN. The transformation moves the dependent variable to position 40, makes any averages, then puts monthly variables into the first positions 1-n, adds the averages, and then places the dependent variable at the end.

Once you make these transformations, there is a limited number of things you can do in PRECON. For example, you can not make any more transformations, and the response plot will not be labeled correctly after transforming the variables. Climatic Change can be run, but the effects of different seasons may not be calculated correctly. The sub period can be changed and the regression using the transformed variables can be run again. For more analyses after generating the input for Kalman, save the information before choosing the Kalman Filter option. Quit after generating the Kalman input and start PRECON again selecting option 4 to read the file and start again.

5) RUNNING KALMAN

Enter "KALMAN" to begin the Kalman Filter. Select the default option to read the PRECON file. The program allows some manipulation of the predictand data such as Differencing or fitting a Polynomial. Lags and leads can be added as predictors. Take the defaults unless you know what you are doing. The next prompt asks "Are there any Exogenous Predictors?" Answer "Y" to get the next prompt: "ENTER "I" TO RUN 1 OR MORE PREDICTORS Individually"; "ENTER "M" TO RUN More THAN 1 PREDICTOR". Select "I" to run each variable individually through Kalman with the option to take all those with a decrease in the AIC statistic as multiple Kalman Filter predictors. Select "M" to run up to 10 predictors in the Kalman Filter selected from the PRECONK output. Option 1 is essentially the recommendation of Visser. See the reference cited. After selecting the options and indicating what data are to be included in the analysis, take the default options unless you understand the Kalman Filter. After calculating the filter, if the change in AIC statistic is negative, a plot will be made automatically. If the change is 0 or positive, the program asks whether you want to plot it. If the variance reduced is over 10% make plots even though the AIC statistic did not decline. You can always delete them. Each time plots are made, the number in the file is incremented. The file

name ending with "A" includes the time series of the actual values, estimates, residual and the 95% confidence limits. The file name(s) ending with "K" include time series of 3 Kalman Filter weights. If there are more than 3 weights additional files ending in "K" are produced. More information on the Kalman Filter is provided in Appendix 1.

The Kalman Filter analysis as Ed originally designed it can be run by answering 'N' to the first question. With this option two files must be present, one with the dependent data and one with the independent data. Transformations can be made only with his version of the program. To run the PRECON generated data with Ed's version, strip the header information from the KALMANPR.INP file, copy it to a different name, read the dependent data in the last column of one file and the independent data from the other file. The format is (4X,1X,n(7.3)) where "n" is the number of columns to be read. If the dependent data is in the 18 column, the format to read it is (124X,F7.3).

6) USING PRECON FOR ANALYZING DATA SETS

I recommend that you begin analyses by running the Simple Correlation, Multiple Regression (which is done automatically), and the Response Function. Make plots where necessary. The Multiple Regression can be run again after Response depending upon which method of calibration is wanted in the change option. It is easy to try both methods and see the differences. Either will work. After experimenting with these 1) run the Kalman Filter; 2) run Automatic Transformation and look at the interactions; or 3) if you are real brave and have memorized the instructions in the Appendix, run Transformations Manually. The latter is very tricky. It is probably easier to make transformations before running PRECON. To run 2 or more of the above, read in the data a second or third time to make the different transformations. This is easy to do if the learning mode (2) is selected at the beginning of the PRECON. The answers to the prompts will be recorded. On exit from PRECON, answer "Y" to the question about saving the file and it copies the "Learned" FILENEW.INI TO FILE.INI which can be read the next time PRECON is run by choosing option 3 at the beginning. Any changes from these input will be written in FILENEW.INI.

If you start at the beginning of a chronology, there will be no lags. To create lags for a file beginning in 1920, enter 1923 for the first year of the analysis (Prompt 2). Then to generate lags as prior growth variables, run Response on the entire period, not on a subperiod. If you use a subperiod, the lags will not be generated for the entire period. After that, one can choose a subperiod, return to Response, choose lags of 3 before calibrating the subperiod. To project that model over the entire period, return to Change, extend the subperiod, and recalculate using option

More information on PRECON can be found in:

Fritts et al. 1991. Climatic variation and tree-ring structure in conifers: empirical and mechanistic models of tree-ring width, number of cells, cell size, cell-wall thickness and wood density. Climatic Research 1:97-116.

Fritts and Shashkin. In Press. Modeling Tree-Ring Structure as related to Temperature, Precipitation, and Day Length. In: Lewis, T.E., ed. "Dendrochronology and Dendrochemistry." CRC Press, Cleveland. USA.

(Reprints are available at a nominal charge if you cannot find these publications.)

****

APPENDIX 1

A Generalized Stepwise Multiple Regression Analysis
University Computer Center
University of Arizona
Tucson, Arizona

AUTHOR: Charles K. Hussar

INTRODUCTION

The mathematical method for a stepwise multiple regression analysis (including flow chart) has been discussed by M. A. Efroymson of ESSO Research and Engineering Company in Mathematical Methods for Digital Computers, edited by A. Ralston and H. S. Wile, and published by John Wiley and Sons, Inc., 1960.

A maximum of 40 variables and 20 constants can be read and (after transformations) 40 "output" variables can be used for the regression analysis. There is no limit on the number of observations or sets of "input" variables. (The only limit is the overflow condition on the uncorrected sums of squares i.e.,

$$\sum_{i=1}^{n} d \quad X2<1050$$

Transformations can be applied to the input variables to obtain the desired output variables.

STEPWISE PROCEDURE

At any stage of the stepwise procedure only one variable is entered or removed from the regression equation, and a variable to be removed takes priority over the one to be added. The contribution of a variable, in reducing the variance, is compared to the contribution of all other variable in the regression equation. If the contribution of a particular variable is insignificant (at a specified F-level), this variable is removed, otherwise the variance reduction for all variable not in the regress equation is considered. The variable that reduces the most variance (if the calculated F-level exceeds a specified F-level to include variables) is then added to the regression equation. If there are no additional variables to be examined or both of the above significance tests fail, the algorithm is terminated.

A check is made for the possibility of degeneracy resulting from an independent variable being approximately a linear combinations of the other independent variables. If such a case were encountered, the variable in question would not be entered into the regression equation.

A test for serial (or auto) correlation of the residuals is run, where R is the Van Neumann's ratio and U is its normal approximate for large samples. A significance table for R can be found on page 341 of Ezekiel & Fox, Methods of Correlation and Regression Analysis, Third Edition, Wiley 1959. For example if |U| is less than 1.96 then there is no evidence of serial correlation at the 5% significance level. If U <-1.96, it indicates a positive serial correlation. If U> 1.96, it indicates

a negative serial correlation, (of the residuals). If there is evidence of serial correlation, the usual significance test may not be valid.

Data Area

The data area consists of 99 fields, and these fields are indexed from one to 99. The first 40 fields are used both for the input and the output variables. The next 20 fields, from 41 to 60, are used when constants are read. Fields from 61 to 70 contain floating point numbers, 1.0 to 10.0 respectively. Fields 71 through 80 contain useful constants and conversion factors as follow:

Index	Value	Remarks
71	3.1415926	Pi
72	6.2831853	2Pi
73	2.7182818	e
74	1.7453290x10-2	radians in 1 degree
75	2.9088820x10-4	radians in 1 minute
76	4.8481368x10-6	radians in 1 second
77	0.3937	inches in 1 centimeter
78	2.54	centimeters in 1 inch
79	28.349527	grams in 1 ounce
80	0.45359243	kilograms in 1 pound

Fields 81 through 99 may be used to store the intermediate results of the transformations in addition to other fields which are not in use, (for example, if 4 input variables are read, then fields 5 through 60 are available; but if 2 additional constants are read, then the fields indexed 41 and 42 are not available). The constants in field 61 through 80 are generated at the start of the program, and if they are destroyed by storing the intermediate results of the transformation in these fields, the constants will not be regenerated for the next regression problem. The constants in fields 41 through 60 are read before the data, and are destroyed either by transformations or by reading a new set of constants. (It is not necessary to read the same set of constants for every problem, only for the first problem).

Reading Cycle

During one cycle only, one set (year) of input variables is read into the fields indexed 1 through 40. If there are transformations to be made, they are calculated using the 1-30 column data area. The transformed data are placed are then placed in the order they are to be read by the regression, fields 1 through n. The n-th field contains the dependent variable. These n variables are used to calculate the sums, the raw sums of squares and the cross products; then the next set of input variables are read.

Transformations

The purpose of the transformations is to process the original input data. Transformations produce output variables from input variables). New variables can be generated (e.g., sin (x), log (2x), (xty)z, etc), and variables can be scaled and rearranged (eg., 10x, A=B, etc.).

Each transformation is defined by an 8 digit field and consists of 4 two digit numbers (4I2) which are identified as C, I. J, and K. C is the transformation code defining a FORTRAN operation or

function. I, J and K are index numbers defining the fields in

the data area. I and J are used for identifying the fields (variables or constants) which are to be used in the transformation, while K, if not zero, designates the field in which the result of the transformation is to be placed. If K is equal to zero, the result will be placed in field I, (and thus destroys its previous content). Note: J is not used in connection with the transformation codes 1 through 10.

Let X(I), X(J) and X(K) be the contents of the fields indexed I, J and K respectively (Z is a temporary storage). The transformation codes are defined below.

Codes	FORTRAN Operation or Function
00	Continue (i.e., No transformation)
01	Z = SQRTF (X(I))
02	Z = SINF (X(I))
03	Z = COSF (X(I))
04	Z = ATANF (X(I))
05	Z = ASINF (X(I))
06	Z = EXPEF (X(I))
07	Z = EXPF (X(I))
08	Z = LOGEF (X(I))
09	Z = LOGF (X(I))
10	Z = ABSF (X(I))
11	Z = X(J) X(I) = X(J) or better X(K)= X(J)
12	Z = X(I) + X(J)
13	Z = X(I) - X(J)
14	Z = X(I) * X(J)
15	Z = X(I) / X(J)
16	Z = X(I) ** X(J)

Decision routine to place the result of the transformation (Z)

```
      IF(K) 1,1,2
    1 X(I) = Z
      GO TO 3
      2 X(K) = Z
    3 CONTINUE
```

EXAMPLES

CIJK	FORTRAN EQUIVALENT	REMARKS
00010103	CONTINUE	C-00
07050000	X(5)=EXPF (X(5))	
07050004	X(4)=EXPF (X(5))	
07050204	X(4)=EXPF (X(5))	j is not used
09010000	X(1)=LOGF (X(1))	
10010002	X(2)=ABSF (X(1))	
11010500	X(1)=X(5)	K=0, I used
11010502	X(2)=X(5)	K=0, I not used
11000502	X(2)=X(5)	K=0, I not used
12010200	X(1)=X(1) + X(2)	
13010203	X(3)=X(1)-X(2)	
14017097	X(97)=X(1)*X(70)	Scaling by 10
15610106	X(6)=X(61)/X(I)	reciprocal
16016600	X(1)=X(1)8(74)	degrees to radians
03010000	X(1)=COSF(X(1))	then cosine
14010102	X(2)=x(1)*(1)	if X(1)=R,then X(2)=R2
14627203	X(3)=X(62)*X(72)	X(3)=2(2)=4
14020304	X(4)=X(2)*X(3)	X(4)=4 R2=surface
15016305	X(5)=X(1)/X(3)	X(5)= R/3
14040506	X(6)=X(4)*X(5)	X(6)=4/3R3 =volume

Note: (1) For example, when the cosine of an angle is to be used, and the angle is read in degrees.

(2) Given the radius of a ball, when its surface and its volume is to be used in the regression equation, 5 transformations will do the trick.

Error Messages

There are four error exists which terminate the algorithm.

Error No. 1: There is a negative or a zero variance at the

beginning of the stepwise procedure.

Error No. 2: There are no degrees of freedom for the error term.

Error No. 3: A negative diagonal element is below the tolerance limit (0.0001) in the inverse matrix.

Error No. 4: The standard deviation of a variable is equal to zero.

THE BOOTSTRAPPED RESPONSE FUNCTION

Guiot, J., Laboratoire de Botanique Historique & Palynologie,
UA CNRS
1152, F-13397 Marseille cedex 13

(See: Guiot, J. 1993. The Bootstrapped response Function. Tree Ring Bulletin 51:39-41.)

The principle of the response function is largely described in Fritts (1978) and discussed in Hughes, et al. (1982). To avoid problems when there are a large number of regressors and their intercorrelations, Fritts (1971) has introduced the regression on principal components. For all regression methods, the main problem is that the significance test of the coefficients is biased, and the way to test the stability of the estimates is incorrect. In addition, the response function obtained for a dendrochronological sample should be tested by applying it to growth values over years independent of those used for calibration. The most straightforward way to assess the stability is to divide climatic and tree-ring data into a dependent calibration set and an independent verification set (Fritts, 1976). If the estimated tree-ring indices of the independent data using the regression coefficients calibrated on the dependent data set are close to the observations, then the response function is judged to be reliable.

Gordon et al. (1982) set out the problem of verifying the predictive ability of a model calibrated on one data set when applied to another data set. Because regression coefficients are fitted only to the dependent data, they result in overconfidence in the predictive power of the model. This can be shown by simulating tree-ring indices by random numbers and by calculating response functions with real climatic data (Guiot, 1981; Cropper 1985). These authors showed that response functions of tree-ring series obtained by simulation had fewer significant regression coefficient than those judged significant by standard Student tests. This is due mainly to an inadequate number of degrees of freedom. To test regression coefficients, the Student test involves n-k-1 degrees of freedom where n is the number of observations and k the number of regressors. If k is set to the number of principal components actually introduced in the regression on the basis of their correlation with the predictand

(stepwise regression), the significance of the coefficients is overestimated, because the number k should have been chosen by a priori considerations, i.e. by considerations independent from the predictand. A good practice is to select a relatively large number of principal components taking into account 90 or 95% of the variance of the climatic data or to use the PVP criterion of Guiot (1981; 1985). The number k is then the number of principal components selected by an a priori criterion, not the number of variables entered into the stepwise regression.

The bootstrap procedure of Efron (1979) provides an alternative approach to test the significance and stability of the regression coefficients within a specific time period. Such an approach has been applied to response functions of tree-ring data by Guiot (1990) and Till and Guiot (1990). The lack of information on the statistical properties of the data is replaced by a large number of estimates each based on different sub-samples of data. A comparison of these sub-sample estimates is used to assess the variability of the estimates. The sub-sampling is done by random extraction with replacement from the initial data set. The size of each sub-sample is the same as the number of observations in the initial data set (n), thus avoiding bias (Efron, 1983). Each sub sample forms a bootstrap test useful for cross-validation. Guiot (1990) has shown that after 50 sub-samples, the results do not change significantly.

For each sub sample, the regression coefficients and the multiple correlation are computed on the observations randomly selected (some observations of the initial data set are selected several times in the random selection process, others are absent). An independent verification is done on the unselected observations in each sub-sample. If it is done 50 times, we obtain 50 sets of regression coefficients, 50 multiple correlations and 50 independent verification correlations. The means of the regression coefficients, the multiple correlations, and the independent correlations with their respective standard deviations are computed using these 50 estimates. The bootstrapped regression coefficients are judged significant at the 95% level, if they are twice, in absolute value, their standard deviation (see Fritts and Dean, submitted). A more precise method is to compute the interval between the 2.5th and the 97.5th percentile, which is a 95% confidence level.

In addition, we can predict a tree-ring index from climate using these the mean coefficients from these 50 response functions accompanied with a confidence interval. This feature can be used not only to evaluate the reliability of the results but also to assess simulations. For example, in the spirit of the paper of Fritts and Dean (submitted), the confidence interval can be used to calculate the amount of precipitation that must subtracted or added to the real precipitation series to produce a significant change in growth.

A subroutine RESBO which computes a bootstrapped response function has been added to the program PRECON of H. Fritts. In this interactive program, the user can observe in real time the evolution of the multiple correlations step by step and stop the bootstrap procedure when no significant change in statistics is detected.

A complete package of statistical programs, including bootstrapped regression (but also time series analyses, multivariate analyses, transfer functions...), with a 250 page user's guide (Guiot, 1990b is available by the author for 25 US $.

LITERATURE

Cook, E. R. and L. A. Kairiukstis. 1990. Methods of

Dendrochronology: Application in the Environmental Sciences, Kluwer Academic Press and IIASA, Dordrecht.

Cropper, J. P., 1985. Tree-ring response functions: an evaluation by means of simulations. Ph.D. thesis, University of Arizona, Tucson, 132 pp.

Efron, B. 1979. Bootstrap methods; another look at the jackknife. The Annals of Statistics, 7, 1-26.

Fritts, H. C., T. J. Blasing, B. P. Hayden, and J. E. Kutzbach. 1971. Multivariate techniques for specifying tree-growth and

climate relationships and for reconstructing anomalies in paleoclimate. Journal of Applied Meteorology, 10, 845-864.

Fritts, H. C. 1976. Tree Rings and Climate. Academic Press, London, 567 pp.

Fritts, H. C. and J. S. Dean. submitted. Dendrochronological modeling of the effects of climatic change on tree-ring width chronologies from Chaco Canyon and Environs. Tree-Ring Bulletin.

Gordon, G. A., B. M. Gray, and J. R. Pilcher. 1982. Verification of dendroclimatic reconstructions. In Hughes, M. et al. (eds): Climate from Tree-Rings. Cambridge University Press, Cambridge, 58-62.

Guiot, J. 1981. Analyse mathematique de donnees geophysiques, application a la dendroclimatologie. Ph.D. thesis, Institut d'Astronomie et Geophysique, Louvain-la-Neuve.

Guiot, J. 1985. The extrapolation of recent climatological series with spectral canonical regression. Journal of Climatology, 5, 325-335.

Guiot, J. 1990a. Methods of calibration. In Cook E. and Kairiukstis, L. (eds): Methods of Dendrochronology: Application to Environmental Sciences, Kluwer Academic Press and IIASA, Dordrecht, 165-178.

Guiot, J. 1990b. Methods and programs of statistics for paleoclimatology and paleoecology. In "Quantification des Changements Climatiques: Methodes et programmes", edited by Guiot and Labeyrie, monographie 1, INSU, Paris, 253p.

Hughes, M. 1982. Climate from Tree-Rings. Cambridge University Press, Cambridge.

Till, C., and Guiot, J. 1989. Reconstruction of precipitation in Morocco since AD 1100 based on Cedrus Atlantica tree-ring widths. Quaternary Research, in press.

SOME KALMAN FILTER NOTES

by

Edward Cook

edited for Preconk by

Harold C. Fritts

The Kalman Filter program was developed by Edward Cook and revised for PRECON (now referred to as PRECONK.EXE) by Harold C. Fritts. The program is totally interactive and the original "menu driven" version of Ed's can be run by answering the first question with a 'N'. The prompts and queries are (hopefully) fairly understandable and intuitive. For now anyway, always invoke the data normalization option until Ed sorts out some odd behavior that he discovered in running unnormalized data. For

references to the Kalman Filter as implemented in this program, Ed suggests reading):

Harvey, A.C. 1984. A unified view of statistical forecasting procedures. Journal of Forecasting 3:245-275.

Visser, H. and J. Molenaar. 1988. Kalman Filter analysis in dendroclimatology. Biometrics 44:929-940.

The input data file(s) for Ed's version may have any name(s). If you choose the PRECON version with a 'Y' answer at the beginning, the program automatically reads a file named KALMANPR.INP which can be generated by the PRECONK PROGRAM by choosing the option Kalman Filter. Both versions prompt for the desired information and, in most cases, Hal Fritts has added default answers that can be chosen using "ENTER". Only Ed's version assumes that any exogenous predictors are in a different file from that holding the predictand.

The predictand and exogenous predictors need not cover the same time span. The program automatically adjusts either data set to the time period common to the two data sets. In the case of the PRECONK output, both data sets have already been reduced to the same length. However, all of the exogenous predictors are assumed to have the same number of observations. If only lagged predictands and/or a polynomial time trend are used as predictors, no exogenous predictor file needs to be opened. In Ed's version, the data may be read in two generic forms: casewise or sequential, with the actual data formats being user defined.

The program presently creates 2 output files: Kalman.out and Kalman.dmp. Logical unit numbers 6 (IU6) and 7 (IU7) are assigned, respectively, to these files in the program. The latter has a lot of difference stuff, including the unnormalized design matrix and the detailed grid search results. Be advised that the higher powers (i.e. >1) and cross-products created by using the polynomial detrending and/or variable interaction options are created after the data are corrected by subtracting the mean. This reduces the multicollinearity problems associated with creating such variables. In addition, all math is done in double precision. The version for PRECON will not generate transformations. You will eventually be able to do this in PRECONK before you prepare the file for Kalman.

The program is set up to handle a maximum of 10 total predictors, which includes any lagged predictands, polynomial time trend variables, exogenous variables, and interaction variables. The grid search routine used to maximize the log likelihood function is tolerably fast and quite robust, but its speed will depend on the shape and dimension of the surface. If you think that the solution may have gotten stuck in some local minimum, run the problem again after changing the starting q-values (a runtime option). You may also constrain some variables to have constant coefficients or coefficients with specified q's. By all means experiment and call Ed at if you have questions or problems.

Ed has tried to make the program interface fairly robust to operator screw-ups, but it is not bomb-proof. It has a fairly large amount of error recovery code in it and a limited amount of "help" information. Ed will probably enhance the interface in the future, but it is an enormous time-sink. Good luck on running it and, again, call if you have any problems. Also, lots of luck in interpreting any time-dependent relationships that you discover. It's an ugly world out there.

Additional notes by Harold C. Fritts

The Kalman terminology is a little confusing to me as it may be to you. For more explanation read the references and then perhaps you should contact Ed or Hans Visser. I am not used to this jargon, either. I reduce the length of the arrays to 100 years in this version so the program would fit into conventional memory of a DOS system.

I have checked the program against Ed's original Kalman program running on the Univ. of Ariz. Vax and the answers are the same.

For more information on the Kalman Filter or Ed's source code, please contact him. I will take responsibility for any problems generated by the PRECON adaptation, but have used Ed's calculations throughout. I hope this combination of the two programs is helpful and does not hamper the good work of Ed trying to make the Kalman Filter available to dendrochronologists.