CharAnalysis 0.9: Diagnostic and analytical tools for sediment-charcoal analysis

User's Guide, updated January 2009

Philip Higuera

Montana State University, University of Illinois

philip.higuera@montana.edu

www.montana.edu/phiguera

http://CharAnalysis.googlepages.com

This program has been placed in the public domain with the understanding that it will not be used for profit and that the user has read and agrees to the disclaimer. The program has been tested, though not rigorously, and is correct to the best of my knowledge. If you find any errors or have suggestions for improvement, please contact me at the e-mail above.

Contents

Part I. Background

CharAnalysis is a set of diagnostic and analytical tools designed for analyzing sediment-charcoal records when the goal is peak detection to reconstruct "local" fire history. The analyses are based on the widely-applied approaches that decompose a charcoal record into low- and high-frequency components (e.g. Clark and Royall 1996; Long et al. 1998; Carcaillet et al. 2001; Gavin et al. 2006), and the program introduces a new technique of using a locally-defined threshold to separate signal from noise (Higuera et al. (2008); **?**). The program is setup to make explicit the range of choices an analyst has to make when implementing this approach. Diagnostic tools help determine if peak detection is warranted, and if so, what parameters are most reasonable. Sensitivity analyses illustrate the impacts of alternative analysis criteria on peak-based fire-history interpretations, and graphical displays and statistical analyses summarize peak-based fire history metrics.

The program is written and functions best in Matlab. A stand-alone version is also available for Windows, containing the same functionality. In both cases, the program depends on an .xls file containing the input charcoal dataset and analysis parameters. The Matlab code is organized in distinct components (.m files) described in this guide, and it is distributed in it entirety. Significant effort has been made to make the code well commented, and users are encouraged to "look under the hood", understand what's going on, and modify the program to suit individual needs. Simple modifications to this code can change pre-set analysis parameters, figure formats and text, and printing formats. Simple scripts can be written in Matlab to automate and standardize the analysis of multiple records.

Part II. Using *CharAnalysis*

1 Download and Installation

CharAnalysis for both Matlab and Windows (XP, Vista) can be downloaded from the web site `http://CharAnalysis.googlepages.com`. Updates to the program are documented on this web site, so check the website before undertaking significant analyses.

1.1 Windows XP and Vista

The Windows distribution contains four files:

1. CharAnalysis.exe: the main program the user interacts with.
2. templateChar.xls: template file where the user inputs the charcoal dataset and selects parameters for use by CharAnalysis.
3. MCRInstaller.exe: Matlab Component Runtime program, installed only once on a give computer.
4. CharAnalysis.ctf: Component Technology File archive, required by the computer but not accessed by the user.

To install CharAnalysis.exe:

1. Create a folder on your C drive or in your Programs folder named CharAnalysis, and put all the distributed files into this folder.
2. Before running the program for the first time, you must install the Matlab Component Runtime library by opening the file MCRInstaller.exe. This program installs a set of Matlab functions that will be accessed by the program CharAnalysis.exe. Installation takes a few minutes and requires that you have administrative privileges on your computer. Double-click on MCRInstaller.exe and it will open a command window and begin preparation for the installation. Again, you only need to do this once on a given computer.
3. You are now ready to use CharAnalysis.exe.

1.2 Matlab 7.0 or higher, on any platform

The Matlab distribution contains:

1. The 23 .m files needed to run the main function CharAnalysis.m.
2. templateChar.xls: template file where the user inputs the charcoal dataset and selects parameters for use by CharAnalysis.

To install CharAnalysis.m in Matlab:

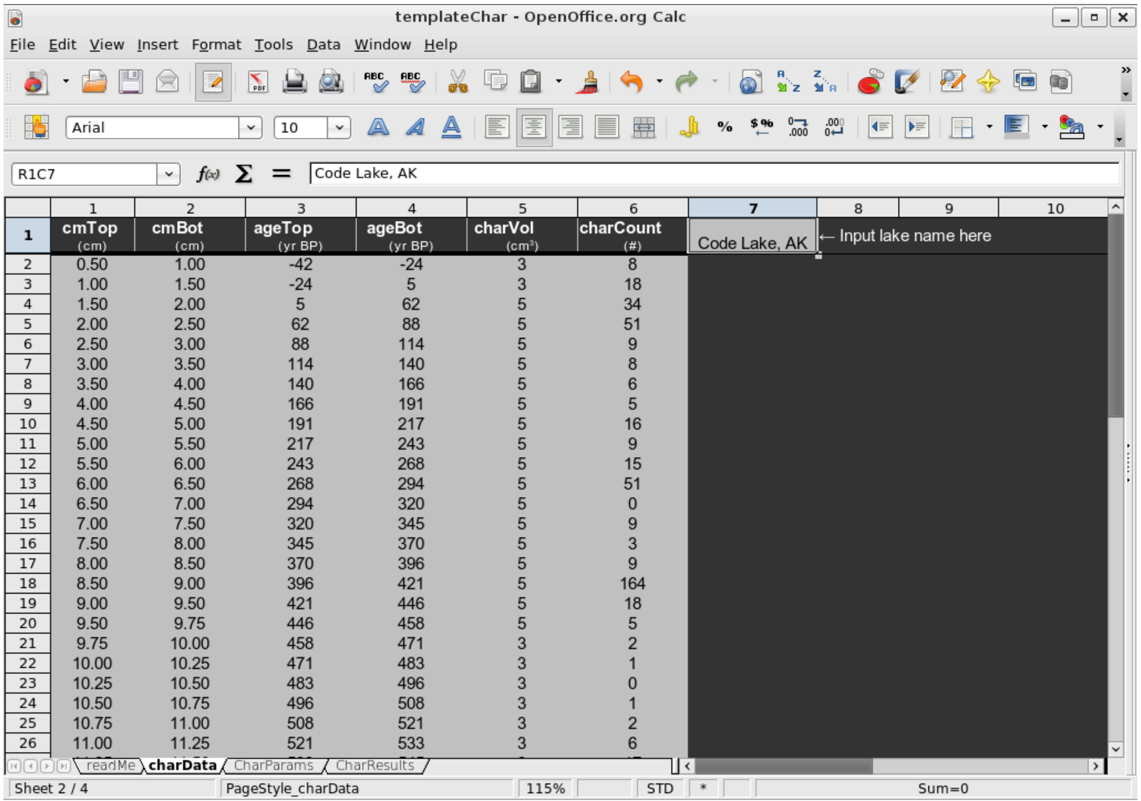
1. Create a folder, CharAnalysis, and put all the *.m files into this folder.
2. Add the CharAnalysis folder to the search path of Matlab. For example, enter addpath 'C:/Phiguera/CharAnalysis' into the Command Window.
3. Save the Matlab path by entering savepath into the Command Window.
4. You are now ready to use CharAnalysis.

2 Data Inpu and Parameter Selection

Open up the included .xls file named templateChar.xls and save it under a new name that identifies your site (e.g. COchar.xls). Place this file in the working directory of Matlab (or change the Matlab working directory), or in the directory that contains the program CharAnalysis.exe.

2.1 Data Input

1. Paste charcoal data into the worksheet labeled charData in the .xls file (Figure 1). The format is the same as for the program *Charster*, written by Dan Gavin (`http://geography.uoregon.edu/gavin/charster/Introduction.html`). For each sample, enter: depth at top of sample, depth at base of sample, age at top of sample, age at bottom of sample, volume of sediment samples, charcoal count. *NOTE*: you must paste only values into this spreadsheet (i.e. no formulas). Missing values for charVol and charCount (but not cmTop, cmBot, ageTop, or ageBot) can be identified by any number < 0 (e.g. -999). CharAnalysis will interpolate across missing data to estimate values. If gaps in a record are not identified as missing values CharAnalysis will interpolate across samples with 0 values, potentially leading to 0 values for CHAR (if the gap is long relative to the interpolation interval).
2. Type in the site name in cell G1 (of column 7, row 1).



	cmTop (cm)	cmBot (cm)	ageTop (yr BP)	ageBot (yr BP)	charVol (cm³)	charCount (#)
2	0.50	1.00	-42	-24	3	8
3	1.00	1.50	-24	5	3	18
4	1.50	2.00	5	62	5	34
5	2.00	2.50	62	88	5	51
6	2.50	3.00	88	114	5	9
7	3.00	3.50	114	140	5	8
8	3.50	4.00	140	166	5	6
9	4.00	4.50	166	191	5	5
10	4.50	5.00	191	217	5	16
11	5.00	5.50	217	243	5	9
12	5.50	6.00	243	268	5	15
13	6.00	6.50	268	294	5	51
14	6.50	7.00	294	320	5	0
15	7.00	7.50	320	345	5	9
16	7.50	8.00	345	370	5	3
17	8.00	8.50	370	396	5	9
18	8.50	9.00	396	421	5	164
19	9.00	9.50	421	446	5	18
20	9.50	9.75	446	458	5	5
21	9.75	10.00	458	471	3	2
22	10.00	10.25	471	483	3	1
23	10.25	10.50	483	496	3	0
24	10.50	10.75	496	508	3	1
25	10.75	11.00	508	521	3	2
26	11.00	11.25	521	533	3	6

Fig. 1: The charData worksheet in templateChar.xls, where data input occurs.

2.2 Parameter Selection

Parameter choices are input into column C (3) of the CharParams worksheet in the templateChar.xls file (Figure 2). Parameter choices are divided into the four stages of program:

2.2.1 Pretreatment:

zoneDiv Years defining beginning and end of record and any zone divisions. Zone divisions are used for plotting and for analyzing fire return intervals. You must input at least two values, the beginning and end of your record, in ascending order (i.e. youngest age on top). You may leave the spaces you don't use blank.

yrInterp Years to interpolate record to. Charcoal counts, sample volume, and sample depths are all interpolated before calculating charcoal accumulation rates. Enter 0 here is you want to interpolate to the median sample resolution (yr sample-1) of the selected record.

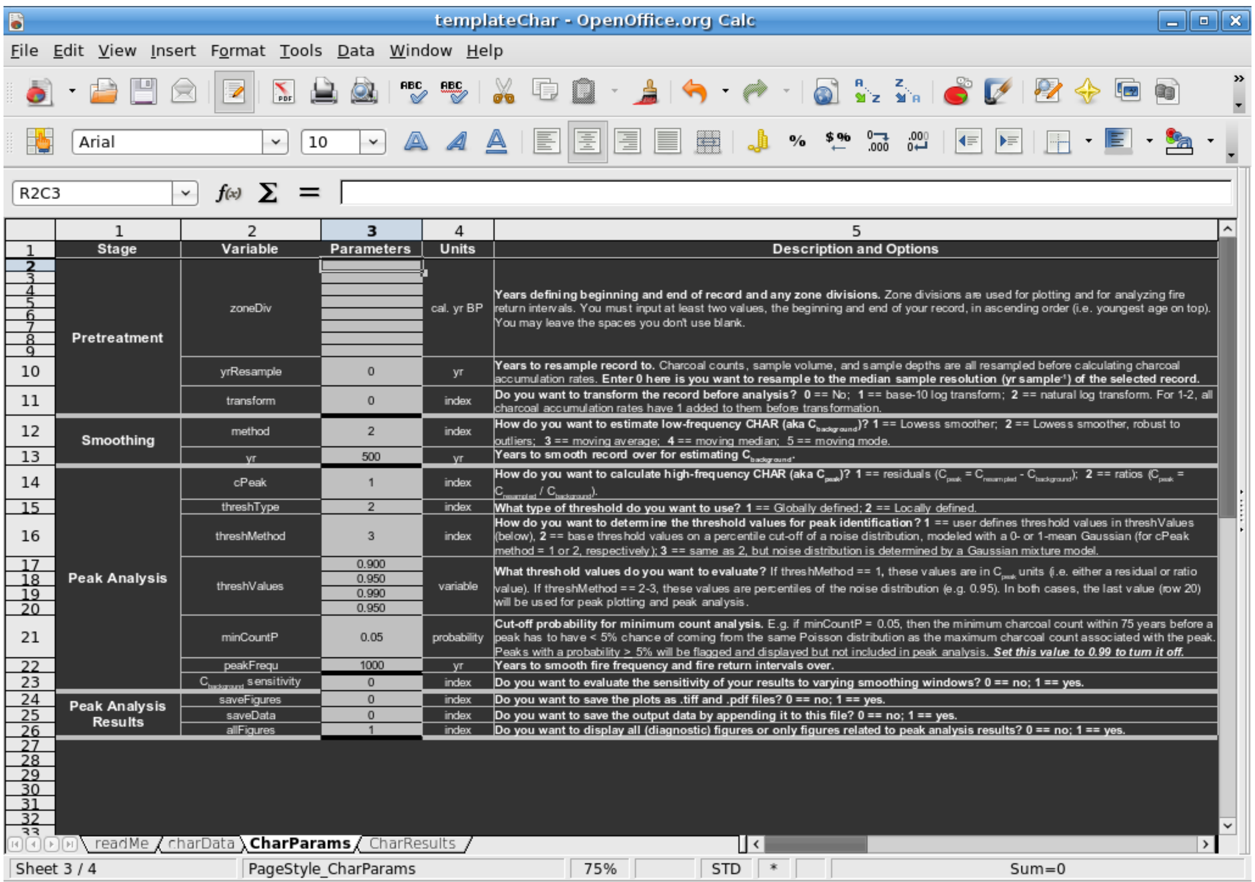


Fig. 2: The "charParams" worksheet in the template file "templateChar.xls", where input parameters are selected.

transform Do you want to transform the record before analysis? 0 == No; 1 == base-10 log transform; 2 == natural log transform. For 1-2, all charcoal accumulation rates have 1 added to them before transformation.

2.2.2 Smoothing:

method How do you want to estimate low-frequency CHAR (aka Cbackground)? 1 == Lowess smoother; 2 == Lowess smoother, robust to outliers; 3 == moving average; 4 == moving median; 5 == moving mode.

yr Years to smooth record over for estimating Cbackground.

2.2.3 Peak Analysis

cPeak – How do you want to calculate high-frequency CHAR (aka Cpeak)? 1 == residuals (Cpeak = Cinterpolated - Cbackground); 2 == ratios (Cpeak = Cinterpolated / Cbackground). threshType – What type of threshold do you want to use? 1 == Globally defined; 2 == Locally defined.

threshMethod How do you want to determine the threshold values for peak identification? 1 == user defines threshold values in threshValues (below), 2 == base threshold values on a percentile cut-off of a noise distribution, modeled with a 0- or 1-mean Gaussian (for cPeak = 1 or 2, respectively); 3 == same as 2, but noise distribution is determined by a Gaussian mixture model.

threshValues What threshold values do you want to evaluate? If threshMethod == 1, these values are in Cpeak units (i.e. either a residual or ratio value). If threshMethod == 2-3, these values are percentiles of the noise distribution (e.g. 0.95). In both cases, the last value (row 20) will be used for peak plotting and peak analysis. minCountP – Cut-off probability for minimum count analysis. E.g. if minCountP = 0.05, then the minimum charcoal count within 75 years before a peak has to have < 5% chance of coming from the same Poisson distribution as the maximum charcoal count associated with the peak. Peaks with a probability > 5% will be flagged and displayed but not included in peak analysis. Set this value to 0.99 to turn it off.

peakFrequ Years to smooth fire frequency and fire return intervals over.

Cbackground sensitivity Do you want to evaluate the sensitivity of your results to varying timescales used to define Cbackground? 0 == no; 1 == yes.

2.2.4 Results

saveFigures Do you want to save the plots as .tiff and .pdf files? 0 == no; 1 == yes.

saveData Do you want to save the output data by appending it to this file? 0 == no; 1 == yes.

allFigures Do you want to display all (diagnostic) figures or only figures related to peak analysis results? 0 == no; 1 == yes.

3 Running CharAnalysis

CharAnalysis.exe from Windows Once the MCRInstaller is installed, double-click on CharAnalysis.exe to open and run the program. The first time you run the program it will go through some set-up procedures that take a few seconds. This procedure will also create the folder CharAnalysis_mcr in the current directory which contains information needed to run the program but is not accessed by the user. NOTE: When first starting up, the .exe program may take up to several minutes to open. This varies with computer and on the number of background applications running. This should bring up the startup text displayed below. Input the datafile name into the DOS Window, with quotations (e.g. 'testChar.xls') and press ENTER to run the program.

3.1 CharAnalysis.m from within Matlab

Open Matlab and type CharAnalysis into the Command Window. This should bring up the startup text displayed below. Input the data file name into the Command Window, with quotations, and press Enter to run the program. (e.g.'testChar.xls') NOTE: In Matlab, you can also call CharAnalysis directly without having to enter the file name into the command window. This option is advantageous because it allows one to analyze multiple files with a few lines of code. For example, entering CharAnalysis ('testChar.xls') will analyze the file testChar.xls; entering CharAnalysis ('testChar.xls'); CharAnalysis ('test2Char.xls') will analyze both files. Finally, entering results = CharAnalysis ('testChar.xls') will return the array results to the workspace with all the results (and more) that are saved into the .xls file. CharAnalysis startup text:

Input file name, bounded with single quotations and including file suffix:

3.2 Citations, support, and updates for CharAnalysis

As of January 2009, the methods used for locally-defined thresholds are best described by Higuera et al. (2008). If you use these methods, please cite this paper. Higuera et al. (in press)contains more extensive description of locally-defined threholds. If you use CharAnalysis for other features, please note in the text (or a footnote) that the program is available for free at `http://CharAnalysis.googlepages.com`. Thank you.

Although I cannot and do not promise any support for users of CharAnalysis, I would like to know when the program is not working as designed. In the event of program crashes that you cannot troubleshoot, please send me (1) the file you are trying to run, and (2) a description of the problem (if using CharAnalysis.exe) or the error message returned by Matlab.

Updates to CharAnalysis will be made as needed and noted on the web page `http://CharAnalysis.googlepages.com`. Please be sure to run the latest version.

3.3 Causes of common errors

The following table contains a list of common problems, as of December, 2007, that cause CharAnalysis to quit:

Tab. 1: Causes of common errors in CharAnalysis, as of January 2009.

Error Description	**Cause**
Fire is loaded, but program quits while or after interpolating and smoothing record.	Input data is .xls file contain formulas (should be text only)
	Multiple depths have the same age assignment
	Smoothing window is longer than record

Part III. Understanding *CharAnalysis*

4 Terminology

Tab. 2: Terms used in CharAnalysis and this documnet.

Term	Description
C	charcoal accumulation rate (CHAR; pieces cm^{-2} yr^{-1})
C_{raw}	CHAR of raw record
C_{int}	CHAR of interpolated record
C_{back}	low-frequency trend in C_{int}, also termed "background CHAR" or "BCHAR" in the literature
C_{peak}	high-frequency trends in C_{int}, after $C_{back}$is removed
C_{noise}	one of the two addative components of C_{peak}
C_{fire}	the other addative components of C_{peak}
t	sample-specific threshold value used to seperate C_{noise} from C_{fire}
C_{thresh}	timeseries of t, which can be displayed in a variety of ways

5 General steps of the analyses

The structure of *CharAnalysis* reflects the main analytical components of most decomposition methods, with many details specific to this program. At each step in the analysis, the user must make one or more parameter decisions (Figure 3). Parameter choices are made before running the program, via an input file. The general steps are as follows:

1. Interpolate the components of the raw charcoal series (concentration [pieces cm^{-3}], sediment accumulation rate [cm yr^{-1}]) to equal intervals to define Cinterpolated (pieces cm^{-2} yr^{-1}).
2. Smooth C_{int} to model low-frequency trends and define C_{back}.
3. Remove $C_{back}$from C_{int} to create a series containing only high-frequency variations, C_{peak} .
4. Define threshold value, t, and apply to the peak series to separated fire-related samples from non-fire related samples
5. Screen peaks and remove any that fail to pass a minimum-count criterion

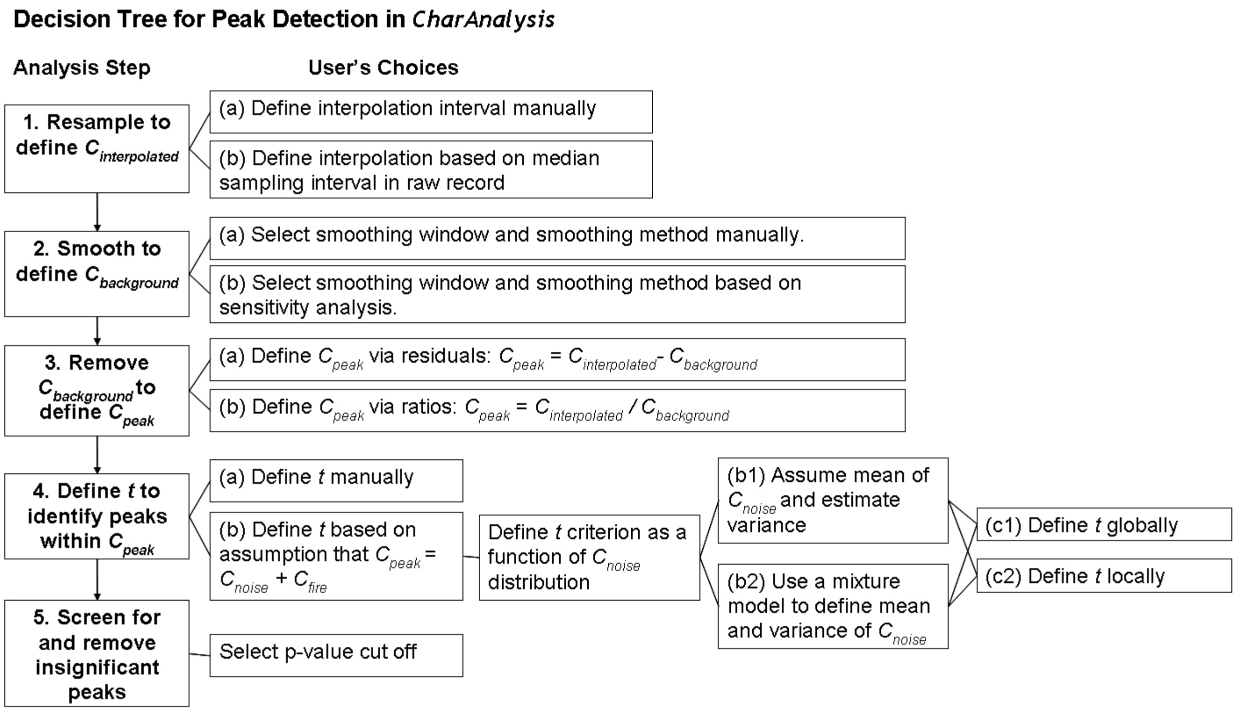


Fig. 3: Decision tree for peak detection in CharAnalysis

6 Analytical choices

6.1 Pretreatment

The analyst has sets of options within the pretreatment phase:

Stratifying Records Records can be divided into user-defined zones, identified by ages, which will be used for plotting and in several statistical comparisons.

Interpolating All records have to be interpolated to equal intervals to justify the techniques used for peak analysis. Resampling in CharAnalysis is based on the technique used in Charster, which (a) determines the relative proportion that each raw sample contributes to each interpolated interval, and then (b) weights the raw sample(s) within an interpolated interval based on this (these) proportion(s). This technique preserves the primary structure of the charcoal data better than binning pseudo-annual values derived by linear interpolation because it makes fewer assumptions about the pattern of charcoal deposition within sampling intervals. The interpolation interval can either be user-defined or determined based on the median sampling interval of all raw samples.

Data Transformation Before peak analysis, CHAR data can be transformed via log (base 10) or natural log transformations. In each case, CHARs have 1 added to them before transformation.

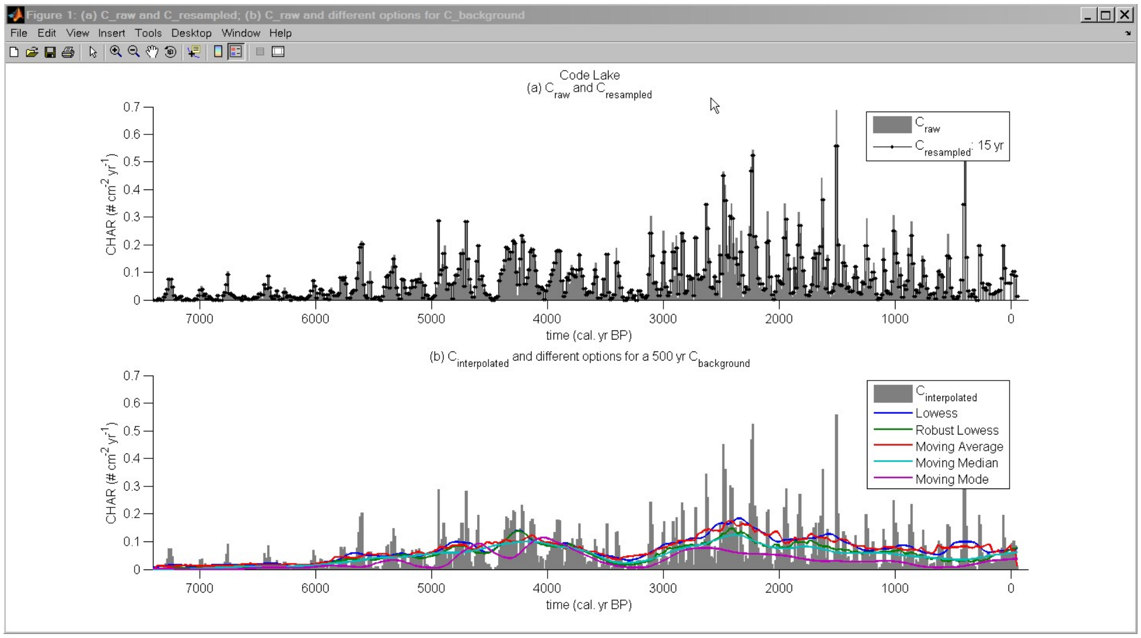


Fig. 4: Raw and interpolated CHAR series (a) and all fire smoothing options (b) are illustrated in Figure 1 of the CharAnalysis output.

6.2 Smoothing

Smoothing refers to the method used to model low-frequency trends in a charcoal record, Cbackground. The analyst must select a smoothing timescale that defines C_{back}. Tools to help make this choice are presented later in this guide. CharAnalysis includes five methods for smoothing the interpolated charcoal series, C_{int} (Figure 3):

1. Locally weighted scatter plot smooth using least squares linear polynomial fitting (i.e. Lowess) (smooth.m function in Matlab Curve Fitting Toolbox)
2. Lowess smoothing that is resistant to outliers (smooth.m function in Matlab Curve Fitting Toolbox)
3. Moving average filter (smooth.m function in Matlab Curve Fitting Toolbox)
4. Moving median. Each sample is assigned the median value of C_{int} within the smoothing window. This series is then smoothed with a Lowess filter.
5. Moving mode. Each sample is assigned the modal value from C_{int} within the smoothing window. Within each window, Cinterpolated values are divided into 100 equally-spaced bins. This series is then smoothed with a Lowess filter.

6.3 Peak Analysis

CharAnalysis includes several methods for peak analysis, reflecting both what has been used in the literature and new techniques for identifying charcoal peaks. The approach involves three steps:

6.3.1 Remove low-frequency trends in C_{int}, to obtain a peak CHAR series, C_{peak}.

There are two options for defining C_{peak}:

1. Define C_{peak} as a residual: $C_{peak} = C_{int} - C_{back}$. This approach assumes an additive relationship between C_{back} and charcoal introduced to a lake from "local" fires. i.e. charcoal introduced to a lake from a "local" fire will be X pieces greater than C_{back}.

2. Define C_{peak} as a ratio: $C_{peak} = C_{int}/C_{back}$. This approach assumes a multiplicative relationship between $C_{back}$and charcoal introduced to a lake from "local" fires. i.e. charcoal introduced to a lake from a "local" fire will be X times greater than C_{back}.

6.3.2 Determine and apply a threshold value, t, to each C_{peak} sample and flag the sample as a "peak" if $C_{peak}>t$.

There are two options here:

1. Define t manually: The user can input the threshold values, in C_{peak}units. e.g. if C_{peak} is defined by a ratio, then t is a ratio; if C_{peak} is defined as a residual, then t is a residual.

2. Define t based on the assumption that the C_{peak} series contains two components, C_{fire} and C_{noise}: Cnoise consists of the normally-distributed variation around C_{back} and C_{fire} consists of the high CHAR values that exceed C_{noise}. Mechanisms including sediment mixing and within-lake redistribution of charcoal create normally-distributed variability in charcoal records (see Clark et al. 1996; Gavin et al. 2006; Higuera et al. 2007). The user must choose where to cut off the C_{noise} distribution, and this is done by selecting a percentile value. e.g. setting t to 0.95 will place the threshold at the 95th percentile of the C_{noise} distribution. It is logical that the separation between C_{fire} and C_{noise} would fall at the upper end of the C_{noise} distribution. CharAnalysis contains two ways for estimating the C_{noise} distribution:

b1) Assume the mean of the C_{noise} distribution and estimate the variance. The mean of the C_{noise} distribution should be at or around C_{back} values, and this method thus assumes a mean of 0 (if C_{back} is defined by residuals), or 1 (if C_{back} is defined by ratios). The variance of C_{noise} is then estimated by calculating the variance of a population containing all C_{peak} values < 0 (or 1) and the absolute value of all C_{peak} values < 0 (or 1) - thus a symmetric population centered on 0 or 1 (Higuera 2006).

b2) Use a Gaussian mixture model to estimate the mean and variance of the C_{noise} distribution. In many records, the mean of the C_{noise} distribution is not exactly at 0 or 1, and in these cases the Gaussian mixture model typically identifies the mean more accurately. The mixture model assumes that each C_{peak} distribution is a mixture of two distributions (i.e. C_{noise} and C_{fire}), and it identifies the mean and variance of each. In practice, the assumptions of the mixture model better fit the mechanisms creating C_{noise}. The distribution of C_{fire} is likely not normal, but all we need to know are the parameters of C_{noise}. *CharAnalysis* uses the same mixture model used by Gavin et al. (2006) and is described therein. See Acknowledgments for the source of the mixture model.

3. Choose whether to define *t* globally (i.e. based on the entire C_{peak} distribution) or locally (i.e. based on the C_{peak} distribution in a defined region around each sample).

c1) Define t globally. This method assumes that variability around C_{back} does not vary through the record. In many cases this may be reasonable and this method has been widely used. For an example of this method applied with the Gaussian mixture model, see Gavin et al. (2006).

c2) Define t locally. The use of a locally-determined threshold is new (Higuera et al. in press, 2008; Briles et al. 2008), and it applies the concepts of a globally-applied threshold to every sample in a record. Specifically, the locally-defined threshold is based on the C_{peak} distribution within a given number of samples around each sample. The timescale used to define local thresholds is the same one used to define C_{back}; i.e. if a 500-yr window width is selected for defining C_{back}, then the threshold value for a given sample, t_{yr}, is defined by the C_{peak} values in the surrounding 500 years. In this example, threshold values for samples within 500 yr of the start (end) of the record are based on the first (last) 500 yr of the record. When using a locally-defined threshold, be sure to have a window-width large enough to include > 30 samples within each window. Defining the C_{noise} distribution with fewer than 30 samples is not advised.

4. Screen "peaks" passing the threshold criterion and eliminating peaks that result from statistically insignificant variations in charcoal counts. Each sample exceeding the threshold for a given year, t_{yr}, is subjected to "minimum count" screening before it is classified as a charcoal peak. This screening is adopted from the program Charster (see acknowledgments) and calculates the probability that two charcoal counts could arise from the same Poisson distribution. First, the sample volume, V1, and charcoal count, X1, from the sample with the lowest charcoal count from all samples within the 75 years preceding the potential peak are identified. This sample must also occur after the preceding charcoal peak. This is compared to the charcoal count, X2, and sample volume, V2, from the sample with the largest charcoal count occurring within 75 years after the potential peak using a two-sample Poisson test, modified for unequal “frames” (or sample volume in this case; Shiue and Bain 1982). The test statistic, d, is compared to the Student’s t distribution and is calculated as follows:

$$d = \frac{\left|X_1-(X_1+X_2)(\frac{V_1}{V_1+V_2})\right|-0.5}{\sqrt{(X_1+X_2)(\frac{V_1}{V_1+V_2})(\frac{V_2}{V_1+V_2})}}$$

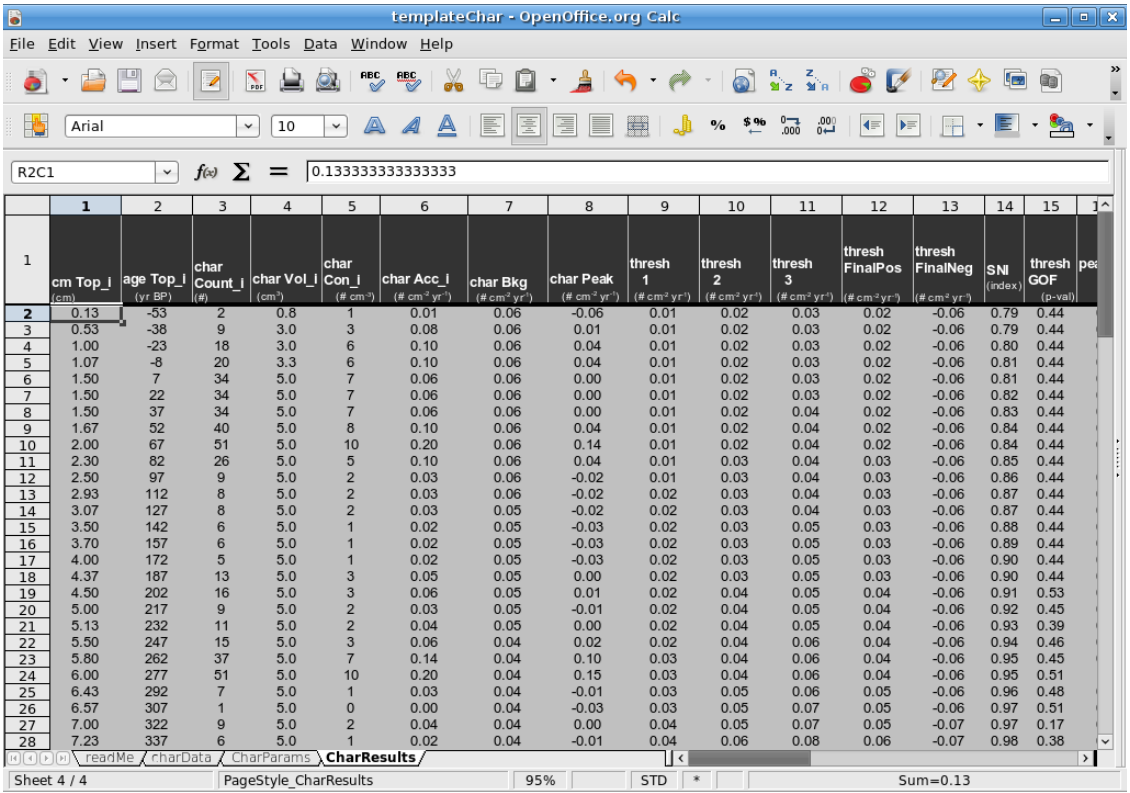
If the probability of obtaining a given d is > 0.05, i.e. there is a greater than 5% chance that the two charcoal counts came from the same Poisson distribution, then the potential peak is eliminated. Conversely, only peaks only peaks that have a 5% of less probability of coming form the same Poisson distribution as the minimum charcoal count proceeding the peak are identified.

6.3.3 Results

Choices in the results section on the input file include:

peakFrequ. Peak analysis summaries include a smoothed fire-frequency curve and a smoothed fire return intervals (FRIs) curve. The smoothing window for these analyses is entered in the parameter peakFrequ (in years). In the case of fire frequencies, the total number of fires within a 1000-yr period are summed, and then this series is smoothed with a Lowess smoother. For the FRI curve, the raw FRIs are interpolated to annual values, and then this series is smoothed with a Lowess smoother.

Cbackground sensitivity. is a binary variable that turns on (off) this option by entering 1 (0). When turned on, peak analysis is repeats 5-10 times (depending on previous peak analysis parameters used) using a range of value for the smoothing window defining C_{back}. If using a global threshold, a three-variable plot is produced illustrating the number of peak identified (z) as a function of threshold value (x) and smoothing window (y). If a local threshold is used, three plots are produced: (a) box plots of the KS goodness-of-fit test for the C_{noise}distribution (y) for different smoothing windows (x), (b) box plots of the signal-to-noise distribution for (y) for different smoothing windows (x), and (c) the sum of (a) and (b), which helps select the optimal smoothing window when (a) and (b) change contain opposite trends.



	cm Top_i (cm)	age Top_i (yr BP)	char Count_i (#)	char Vol_i (cm³)	char Con_i (# cm⁻³)	char Acc_i (# cm⁻² yr⁻¹)	char Bkg (# cm⁻² yr⁻¹)	char Peak (# cm⁻² yr⁻¹)	thresh 1 (# cm⁻² yr⁻¹)	thresh 2 (# cm⁻² yr⁻¹)	thresh 3 (# cm⁻² yr⁻¹)	thresh FinalPos (# cm⁻² yr⁻¹)	thresh FinalNeg (# cm⁻² yr⁻¹)	SNI (index)	thresh GOF (p-val)
2	0.13	-53	2	0.8	1	0.01	0.06	-0.06	0.01	0.02	0.03	0.02	-0.06	0.79	0.44
3	0.53	-38	9	3.0	3	0.08	0.06	0.01	0.01	0.02	0.03	0.02	-0.06	0.79	0.44
4	1.00	-23	18	3.0	6	0.10	0.06	0.04	0.01	0.02	0.03	0.02	-0.06	0.80	0.44
5	1.07	-8	20	3.3	6	0.10	0.06	0.04	0.01	0.02	0.03	0.02	-0.06	0.81	0.44
6	1.50	7	34	5.0	7	0.06	0.06	0.00	0.01	0.02	0.03	0.02	-0.06	0.81	0.44
7	1.50	22	34	5.0	7	0.06	0.06	0.00	0.01	0.02	0.03	0.02	-0.06	0.82	0.44
8	1.50	37	34	5.0	7	0.06	0.06	0.00	0.01	0.02	0.04	0.02	-0.06	0.83	0.44
9	1.67	52	40	5.0	8	0.10	0.06	0.04	0.01	0.02	0.04	0.02	-0.06	0.84	0.44
10	2.00	67	51	5.0	10	0.20	0.06	0.14	0.01	0.02	0.04	0.02	-0.06	0.84	0.44
11	2.30	82	26	5.0	5	0.10	0.06	0.04	0.01	0.03	0.04	0.03	-0.06	0.85	0.44
12	2.50	97	9	5.0	2	0.03	0.06	-0.02	0.01	0.03	0.04	0.03	-0.06	0.86	0.44
13	2.93	112	8	5.0	2	0.03	0.06	-0.02	0.02	0.03	0.04	0.03	-0.06	0.87	0.44
14	3.07	127	8	5.0	2	0.03	0.05	-0.02	0.02	0.03	0.04	0.03	-0.06	0.87	0.44
15	3.50	142	6	5.0	1	0.02	0.05	-0.03	0.02	0.03	0.05	0.03	-0.06	0.88	0.44
16	3.70	157	6	5.0	1	0.02	0.05	-0.03	0.02	0.03	0.05	0.03	-0.06	0.89	0.44
17	4.00	172	5	5.0	1	0.02	0.05	-0.03	0.02	0.03	0.05	0.03	-0.06	0.90	0.44
18	4.37	187	13	5.0	3	0.05	0.05	0.00	0.02	0.03	0.05	0.03	-0.06	0.90	0.44
19	4.50	202	16	5.0	3	0.06	0.05	0.01	0.02	0.04	0.05	0.04	-0.06	0.91	0.53
20	5.00	217	9	5.0	2	0.03	0.05	-0.01	0.02	0.04	0.05	0.04	-0.06	0.92	0.45
21	5.13	232	11	5.0	2	0.04	0.05	0.00	0.02	0.04	0.05	0.04	-0.06	0.93	0.39
22	5.50	247	15	5.0	3	0.06	0.04	0.02	0.02	0.04	0.06	0.04	-0.06	0.94	0.46
23	5.80	262	37	5.0	7	0.14	0.04	0.10	0.03	0.04	0.06	0.04	-0.06	0.95	0.45
24	6.00	277	51	5.0	10	0.20	0.04	0.15	0.03	0.04	0.06	0.04	-0.06	0.95	0.51
25	6.43	292	7	5.0	1	0.03	0.04	-0.01	0.03	0.05	0.06	0.05	-0.06	0.96	0.48
26	6.57	307	1	5.0	0	0.00	0.04	-0.03	0.03	0.05	0.07	0.05	-0.06	0.97	0.51
27	7.00	322	9	5.0	2	0.04	0.04	0.00	0.04	0.05	0.07	0.05	-0.07	0.97	0.17
28	7.23	337	6	5.0	1	0.02	0.04	-0.01	0.04	0.06	0.08	0.06	-0.07	0.98	0.38

Fig. 5: Example of CharResults worksheet after running CharAnalysis and saving data.

saveFigures and saveData. The user can choose to (not) save the output figures and/or output data by entering (0) 1 for the saveFigures or saveData parameters. When using the program in Matlab, figures are saved as both .tiff and .pdf files. When using the stand-alone version, CharAnalysis.exe, only .tiff files are saved. Output data are appended to the input .xls file. NOTE: this file must be closed for the program to save output data. 4.Finally, the allFigures parameter allows one to suppress diagnostic figures (1, 2, and 9) by entering 0. Entering 1 will display all figures.

allFigures Enter 1 to display all figures or 0 to not display diagnostic figures (1, 2, and 9).

7 CharAnalysis output

7.1 Data

Nearly all the data needed to create the figures output by CharAnalysis are provided to the user so you can (1) understand the figures, and (2) create your own figures, should you prefer. The columns of the worksheet CharResults contain the following output data (Figure 5):

A) cmTop_i interpolated top depth (cm)

B) ageTop_i interpolated top age (input units, yr BP) C)charCount_i – interpolated charcoal counts (pieces)

D) charVol_i interpolated sample volume (cm3)

E) charCon_i interpolated charcoal concentration (pieces cm-3)

F) charAcc_i interpolated charcoal accumulation rate (pieces cm-2 yr-1)

G) charBkg low-frequency trend in CHAR, Cbackground (pieces cm-2 yr-1)

H) charPeak detrended CHAR, Cpeak (pieces cm-2 yr-1)

I) thresh1 1st CHAR threshold value selected (pieces cm-2 yr-1)

J) thresh2 2nd CHAR threshold value selected (pieces cm-2 yr-1)

K) thresh3 3rd CHAR threshold value selected (pieces cm-2 yr-1)

L) threshFinalPos Final positive CHAR threshold value (pieces cm-2 yr-1)

M) threshFinalNeg Final negative CHAR threshold value (pieces cm-2 yr-1)

N) SNI Signal-to-noise index. For each threshold value (i.e. one, if using a global threshold, or for each interpolated sample, if using a local threshold):

$$SNI = \frac{var(C_{fire})}{var(C_{fire})+var(C_{noise})}$$

where C_{fire} is the population of samples that exceeds the threshold value, and C_{noise} is the population of samples that is less-than or equal to the threshold value. If using a global threshold, *SNI* will be the same for all samples; if using a local threshold, SNI will differ for each sample.

O) threshGOF Threshold goodness-of-fit measure. For each distribution of peak CHAR, C_{peak}, threshGOF is the p-value from a Kolmogorov-Smirnov test comparing the fitted noise distribution and C_{noise} samples (C_{noise} = CHAR samples less-than or equal to the threshold value). In this case, a "large" p-value suggests that the two populations come from the same distribution (although technically, a large p-value fails to suggest that the two populations do not come from the same distribution). A "small" p-value suggests the opposite. It is up to the analyst to decide on a "large" p-value, but 0.10 would be a reasonable choice according to the author.

P) peaks1 binary variable indicating the start of an identified CHAR peak, using thresh1

Q) peaks2 same as peaks1, but for thresh2

R) peaks3 same as peaks1, but for thresh3

S) peaksFinal same as peaks1, but for threshFinalPos

T) peaksInsig. Identifies peaks that failed to pass the Poisson minimum-count criterion, set in the input file.

U) peakMag Peak magnitude (pieces cm-2 peak-1) is the sum of all samples exceeding threshFinalPos for a given peak. The units are derived as follows: [pieces cm-2 yr-1] * [yr peak-1] = [pieces cm-2 peak-1].

V) smPeakFrequ For every interpolated sample i, the binary series peakFinal is summed over peakFrequ (user-set value) years, centered on year i. This series is then smoothed to peakFrequ years using a Lowess filter (as described in the section Smoothing).

W) smFRIs The time series of fire return intervals (years per fire; FRI) is interpolated to annual resolution, and then smoothed with a Lowess filter (as described in the section Smoothing).

X) nFRIs Total number of FRIs in each zone selected in the input file.

Y) mFRI Mean FRI within each zone selected in the input file. Each zone is treated independently, so the first fire in each zone does is not used to calculate a FRI (which would be based on the last fire in the previous zone).

Z) mFRI_uCI Upper 95% confidence interval for mFRI, based on 1000 bootstrapped samples..

AA) mFRI_lCI Lower 95% confidence interval for mFRI, calculated as for mFRI_uCI.

BB) WBLb Maximum likelihood estimate of the Weibull b parameter (yr) for the distribution of FRIs within each zone selected in the input file, determined using the Matlab function wblfit. The goodness-of-fit of each Weibull model is tested using a one-sample Kolmogorov-Smirnov test, and Weibull models are only reported if the p-value from this test is > 0.10 (if nFRI < 30), or > 0.05 (if nFRI > 30).

CC) WBLb_uCI Upper 95% confidence interval for WBLb, based on 1000 bootstrapped samples.

DD) WBLb_lCI Lower 95% confidence interval for WBLb, based on 1000 bootstrapped samples.

EE) WBLc Maximum likelihood estimates of the Weibull c parameter (unitless), based on the same method used to estimate WBLb.

FF) WBLc_uCI Upper 95% confidence interval for WBLc, based on 1000 bootstrapped samples.

GG) WBLc_lCI Upper 95% confidence interval for WBLc, based on 1000 bootstrapped samples.

7.2 Figures

Output figures provide a detailed look into what the program is doing numerically, and they provide publication-quality displays of charcoal series and summaries of peak-analysis results (Figure 6). Nearly all figures are based on data output by *CharAnalysis* and described above. All time series plots except Figure 2 are scaled horizontally based on the amount of time analyzed; thus, longer (shorter) records will take up more (less) of the display. If you use a figure from *CharAnalysis* in a presentation or publication, please cite this User's Guide and the website hosting the program.

Figure 1: (a) Craw and Cinterpolated; (b) Craw and different options for Cbackground. (a) Raw CHAR, Craw, is displayed as bars, so sampling intervals are visible, and charAcc_i (labeled Cinterpolated) is displayed as a stair-step plot. (b) Cinterpolated with the five different options available for estimating low-frequency trends, Cbackground. Areas of the record with missing values are identified by a gray box.

Figure 2: Global or Local distribution of Cpeak values. This figure varies, depending on the type of threshold selected, i.e. "local" or "global". In both cases, the distribution of peak CHAR values, Cpeak, is displayed in a histogram. If using a local threshold, then this is done for multiple, non-overlapping, time periods that span most/all of the record. Also in this case, each subplot contains (1) the modeled noise distribution, (2) the local threshold value for each year, prior to smoothing (labeled tyr) (3) SNI for each year (labeled SNI), and (4) threshGOF for each year (labeled KS p-val).

Figure 3: Interpolated charcoal (C_int), low-frequency trends (C_back), and detrended series (C_peak). (a) charAcc_i (labeled C_{int}) with charBkg (labeled C_{back}). (b) charPeak (labeled C_{peak}) with charThreshPos and charThreshNeg defining C_{noise}, and peaksFinal plotted as "+" symbols. Peaks that fail to pass the Poisson minimum-count criterion are dis paled as gray dots.

Figure 4: Sensitivity to alternative thresholds and quality of record. (a) charAcc_i (labeled C_{int}), charBkg (labeled C_{back}) and threshFinalPos (in red), with peaksFinal plotted as "+" symbols, and peaks1 and peaks2 plotted at gray dots. (b) Mean fire return interval and 95% confidence limits (y-axis) for each zone (x-axis), based on (from left to right) peaks1, peaks2, and peaks3. This illustrates the sensitivity of FRI interpretations to the three alternate thresholds entered in the input file. (c) The SNI for each sample, illustrating how the variability above threshFinalPos varies throughout a record. (d) Boxplot of all SNI values displayed in (c). Horizontal lines represent the 10th, 25th, 50th, 75th, and 90th percentiles. NOTE: Y-axes in (b) and (c) are log scales, and the minimum value is set to 0.5 when the median SNI > 0.5, or 0.01 if the median SNI < 0.5.

Figure 5: Cumulative peaks through time. The cumulative sum of peaksFinal (y-axis) as a function of time (x-axis). The slope of this plot at any point in time is the instantaneous fire frequency (fires year-1). Areas of the record with missing values are identified by a gray box.

Figure 6: Fire return intervals by zone, with Weibull models if GOF test is passed. The proportion or scaled density (y-axis) of FRIs within each zone displayed in a histogram with 20-yr bins. If the fitted Weibull model passes the goodness-of-fit test described for the variable WBLb, then all model parameters and 95% confidence estimates are listed. Also listed are mFRI and 95% confidence estimates and nFRI.

Figure 7: Continuous fire history: peak magnitude, FRIs through time, and smoothed fire frequency. Top panel displays a time series of peakMag as bars and peaksFinal as "+" symbols. Peaks that fail to pass the Poisson minimum-count criterion are displayed as gray dots. Middle panel displays a time series of fire return intervals (i.e. the time between peaks) and smFRIs. Bottom panel displays a time series of smPeakFrequ. In all panels, areas of the record with missing values are identified by a gray box, and in the bottom panel, this box is extended to cover all data affected by missing values.

Figure 8: Between-zone comparisons of raw CHAR distributions. This plot provides a "bare-bones" analysis of difference in a charcoal record between two or more zones. Left panel displays the cumulative distribution (CDF) of raw CHAR values within each zone selected in the input file. If two or more zones were input, then a two-sample Kolmogorov-Smirnov (K-S) tests the null hypothesis that the CHAR distribution in zone 1 does not differ from that in zone 2. Multiple comparisons are performed if more than two zones are input. A table of p-values is output within this plot, with the zone ID as the first row and first column; all other values are p-values resulting from the comparison of the column zone to the row zone. Right panel displays a box plot of the raw CHAR values from within each zone. Horizontal lines represent the 10th, 25th, 50th, 75th, and 90th percentiles, and dots represent outliers. NOTE: Data from this plot are not output into the CharResults worksheet.

Figure 9: Alternative displays of threshold value(s). This figure displays (a) the C_{int} series with C_{back} and C_{thresh}, and peaks identified, as in previous figures. Panels (b) and (c) illustrate the C_{peak} and C_{thresh} series in two different "domains": the ratio domain and the residual domain. The purpose of this figure is to illustrate how the selected threshold would look if C_{peak} were calculated either as a ratio or residual. Panel (b) displays both series as a ratio to background CHAR. i.e. $C_{peak} = C_{int}/C_{back}$, and $C_{thresh} = C_{back} * t$. Panel (c) displays both series as residuals to background CHAR. i.e. $C_{peak} = C_{int} - C_{back}$, and $C_{thresh} = C_{back} + t$.

Figure 10: Sensitivity to different background windows. Choosing a time scale to define Cbackground can be arbitrary, but the quality-of-record measures and sensitivity analyses in CharAnalysis provide one way to either quantify the sensitivity of peak-analysis results to different background windows, or better yet, choose a background window based the quality-of-record measures. If you are analyzing a record for the first time, consider running this sensitivity analysis to help decide what window width to This figure displays the results from the KS goodness-of-fit test results (p-values) and the SNI distributions from multiple analyses, all using the given input parameters, but varying the timescale for defining Cbackground.

8 Details and structure of the Matlab scripts

CharAnalysis.m is the main function running the program. It calls upon the other functions included in the distributions, including those listed below. Matlab *.m files not listed below are

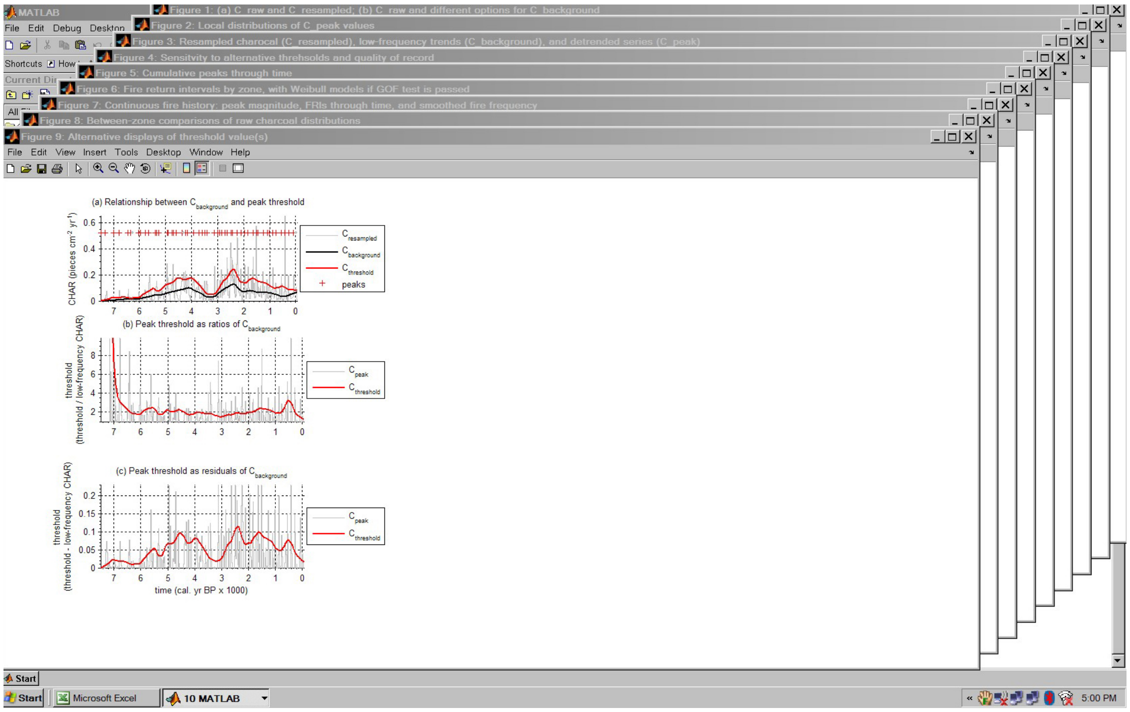


Fig. 6: Figure output array after running CharAnalysis.

either associated with the Gaussian mixture model referenced in ACKNOWLEDGMENTS, included to free the user from having to have the Matlab Curve Fitting or Statistics toolboxes, or they play a minor role in the program.

CharPrameters.m Reads the parameters and data in the input variable fileName, referencing the input file in .xls format, and returns parameters and charcoal dataset for use in CharAnalysis.

CharPretreatment.m Interpolated charcoal data to resolution defined by yrInterp, derives raw and interpolated charcoal accumulation rates, and log transforms charcoal accumulations rates, if desired.

CharSmooth.m Smooths interpolated charcoal series to estimate C_{back} using 1 one of 5 methods described perilously.

CharThreshGlobal.m Determines a threshold value for each interpolated sample, based on the distribution of C_{peak} values within the entire record and either a Gaussian mixture model or the assumption that the noise component of C_{peak} is normally distributed around 0 (if C_{peak} is defined by residuals) or 1 (if C_{peak} is defined by ratios). This function calls upon the Gaussian mixture model CLUSTER, referred to in the Acknowledgments section.

CharThreshLocal.m Determines a threshold value for each C_{int} sample, based on the distribution of C_{peak} values within the selected window (yr) and either a Gaussian mixture model or the assumption that the noise component of the peak charcoal record (C_{peak}) is normally distributed around 0 (if C_{peak} is defined by residuals) or 1 (if C_{peak} is defined by ratios). This function calls upon the Gaussian mixture model CLUSTER, referred to in the Acknowledgments section.

CharPeakID.m Identifies charcoal samples exceeding threshold value(s) determined in CharThrshLocal.m or CharThreshGlobal.m, and screens these values according to the minimum-count criterion selected.

CharPeakAnalysisResults.m Plots 6, 9 or 10 figures, depending on parameter choices, illustrating results of CharAnalysis. Data are saved to the .xls input file and figures are saved as .tiff files and .pdf files, if desired.

bkgCharSensitivity.m Analyzes a charcoal record using multiple timescale for defining C_{back} (but all with the same smoothing method) and plots a series of graphs illustrating the sensitivity of results to varying Cbackground definitions. These graphs vary depending on parameter choices.

Part IV. Acknowledgments

Many features in CharAnalysis are based on analytical techniques contained within the programs CHAPS, by Patrick Bartlein, and Charster, by Daniel Gavin (`http://geography.uoregon.edu/gavin/charster/Introduction.html`). In particular, the resampling algorithm and the minimum-count screening used in CharAnalysis were developed directly from these features in Charster. The peak magnitude and smoothed fire frequency results and display were developed based on CHAPS. CharAnalysis, like Charster, uses the Gaussian mixture model CLUSTER, which was created and is distributed as Matlab code by Charles Bowman (`http://cobweb.ecn.purdue.edu/~bouman/software/cluster/`). Development of the program has benefited greatly from discussions with and testing by members of the Whitlock Paleoecology Lab at Montana State University, Dan Gavin and Patrick Bartlein. CharAnalysis was written in Matlab (`http://mathworks.com`) with resources from the University of Washington, Montana State University, and the University of Illinois.

Part V. Disclaimer

THIS SOFTWARE PROGRAM AND DOCUMENTATION ARE PROVIDED "AS IS" AND WITHOUT WARRANTIES AS TO PERFORMANCE. THE PROGRAM CharAnalysis IS PROVIDED WITHOUT ANY EXPRESSED OR IMPLIED WARRANTIES WHATSOEVER. BECAUSE OF THE DIVERSITY OF CONDITIONS AND HARDWARE UNDER WHICH THE PROGRAM MAY BE USED, NO WARRANTY OF FITNESS FOR A PARTICULAR PURPOSE IS OFFERED. THE USER IS ADVISED TO TEST THE PROGRAM THOROUGHLY BEFORE RELYING ON IT. THE USER MUST ASSUME THE ENTIRE RISK AND RESPONSIBILITY OF USING THIS PROGRAM.

THE USE OF THE SOFTWARE DOWNLOADED FROM THE MONTANA STATE UNIVERSITY WEBSITE IS DONE AT YOUR OWN RISK AND WITH AGREEMENT THAT YOU ARE SOLELY RESPONSIBLE FOR ANY DAMAGE TO YOUR COMPUTER SYSTEM OR LOSS OF DATA THAT RESULTS FROM SUCH ACTIVITIES.

Part VI. References

References

Briles, C. E., C. Whitlock, P. J. Bartlein, and P. E. Higuera. 2008. Regional and local controls on postglacial vegetation and fire in the Siskiyou Mountains, northern California, USA. Palaeogeography Palaeoclimatology Palaeoecology **265**:159–169.

Carcaillet, C., Y. Bergeron, P. Richard, B. Frechette, S. Gauthier, and Y. Prairie. 2001. Change of fire frequency in the eastern Canadian boreal forests during the Holocene: does vegetation composition or climate trigger the fire regime? Journal of Ecology **89**:930–946.

Clark, J. S., and P. D. Royall. 1996. Local and regional sediment charcoal evidence for fire regimes in presettlement north-eastern North America. Journal of Ecology **84**:365–382.

Clark, J. S., P. D. Royall, and C. Chumbley. 1996. The role of fire during climate change in an eastern deciduous forest at Devil's Bathtub, New York. Ecology **77**:2148–2166.

Gavin, D. G., F. S. Hu, K. Lertzman, and P. Corbett. 2006. Weak climatic control of stand-scale fire history during the late Holocene. Ecology **87**:1722–1732.

Higuera, P. E., 2006. Late Glacial and Holocene Fire History in the Southcentral Brooks Range, Alaska: Direct and Indirect Impacts of Climatic Change on Fire Regimes. Ph.d. dissertation, University of Washington.

Higuera, P. E., L. B. Brubaker, P. M. Anderson, T. A. Brown, A. T. Kennedy, and F. S. Hu. 2008. Frequent Fires in Ancient Shrub Tundra: Implications of Paleorecords for Arctic Environmental Change. PLoS ONE **3**:e0001744.

Higuera, P. E., L. B. Brubaker, P. M. Anderson, F. S. Hu, and T. A. Brown. in press. Vegetation mediated the impacts of postglacial climate change on fire regimes in the southcentral Brooks Range, Alaska. Ecological Monographs .

Higuera, P. E., M. E. Peters, L. B. Brubaker, and D. G. Gavin. 2007. Understanding the origin and analysis of sediment-charcoal records with a simulation model. Quaternary Science Reviews **26**:1790–1809.

Long, C. J., C. Whitlock, P. J. Bartlein, and S. H. Millspaugh. 1998. A 9000 year fire history from the Oregon Coast Range based on a high-resolution charcoal study. Canadian Journal of Forest Research **28**:774–787.