Chroma: tutorials and descriptions

*Correspondence: Julian Traphagan (jtraph1@lsu.edu)*

**Table of Contents**

1. Overview of Chroma 2

1.1 Running base programs 2

1.2 Peak Integration 3

1.3 Data visualization 3

2. Vignette 4

3. Description of Indices 13

3.1 Carbon Preference Index (CPI) and Odd-Even Predominance (OEP) 13

3.2 Average Chain Length (ACL) 13

3.3 Paq and Terrigenous/Aquatic Ratio (TAR) 13

3.4 Pristane/Phytane (Pr/Ph) 14

3.5 Other Hydrocarbon Indices 14

4. References 15

*Description*: This document contains details on functions available in Chroma. Follow the **Vignette** section with the corresponding MATLAB file (vignette.m) to run basic example operations within the package. Function and input argument descriptions can be found here or by using the *help* command in MATLAB. Please contact the correspondence email with questions or suggestions for new features.

# Overview of Chroma

Hydrocarbon indices are frequently utilized in organic geochemistry as proxies for biological source, maturity, and degradation of organic compounds preserved in sediments. These molecular ratios provide information and constraints for paleoenvironmental and palaeoecological analyses and are thus powerful tools for interpretation and contextualization of the sedimentary record. Hydrocarbon indices are determined by identifying *n*-alkane chain lengths in chromatogram data and integrating them for compound abundance.

Chroma is a package of MATLAB functions for biomarker data processing and calculations in gas-chromatography analysis. The purpose of this package is to provide a group of efficient, accurate, and flexible tools for determining component-specific molecular abundances and hydrocarbon. The base calculations are performed in the title *chroma* function, which reads in a chromatogram of a sample analysis and automatically integrates targeted peaks which are referenced to a standard. Chroma was originally developed for the processing of chromatograms of *n*-alkane peaks produced by a flame ionization detector (FID) but can be adapted for identifying any time series of discrete peaks and will be updated periodically to include more post-processing calculations for other biomarker compounds.

List of available functions in Chroma:

1. *chroma*
2. *chromall*
3. *pkcomp*
4. *indall*
5. *prepfiles*
6. *prph*
7. *cpi*
8. *paq*
9. *acl*
10. *tarhc*
11. *oep*
12. *lh*
13. *lchsch*
14. *wi*
15. *c31c19*
16. *nar*
17. *ratio*
18. *stack*
19. *plotchrom*
20. *compcpi*
21. *compacl*

## Running base programs

The function *chroma* operates by reading in the time series information of a single sample and single standard of equal vector length. Target peaks (e.g., *n*-alkanes) are identified by locating the peaks in the standard and selecting the corresponding peaks in the sample. Non-target peaks (i.e., noise, contaminants, and other compounds in the sample) are automatically omitted from the sample peak identification if outside a specified time window of the peaks identified in the standard. The areas of the algorithm-selected peaks are determined by trapezoidal integration and stored in the output data table.

Input data must be read into functions in the Chroma package in a specific format. The function *prepfiles* automatically reads in a data file (e.g., file extensions .csv, .xslx, etc.) and restructures the data into a readable format. Data files must have the time information in the first column and the intensity information in the second column (any units are acceptable) to be properly formatted by *prepfiles*. If reading in multiple samples to be analyzed by *chromall*, data from each sample should be included in the data file following this column sequence (time, intensity; time, intensity; time, intensity…) with no space between columns.

The Chroma package includes operations for calculating several hydrocarbon indices either using individual functions for each or simultaneously using the *indall* function. The available indices to be calculated include the following: carbon preference index (CPI), average chain length (ACL), *P*­­aq, terrestrial/aquatic ratio (TAR), odd/even preference (OEP), low molecular weight/high molecular weight ratio (L/H), long chain/short chain ratio (LCH/SCH), weathering index (WI), and *n-*alkane ratio (C31/C19), among others. These calculations are all performed following the specifications of the original reporting literature.

## Peak Integration

The sample peak integration algorithm is initialized by setting the criteria for peak identification. These include the peak height threshold minimum, start time of the analysis, and the maximum time window around the reference peak to identify corresponding sample peak. If not included as arguments, the function will use default values for each parameter. These parameters should be adjusted according to peak density and sample rate for optimal peak detection and noise filtering. If the standard does not include peaks present in the sample, these peaks will not be detected in the sample. To include these peaks in the calculations, additional peaks can be manually added if the estimated retention time and component number is known.

## Data visualization

There are several functions included in the Chroma package for quick visualization and diagnosis of properties of a sample. Many of these are automatic outputs of the *chromall* function but can be individually performed as well. The *stack* function, for example, adds together and normalizes the chromatograms of multiple samples; the returned stacked chromatogram subdues the noise and emphasizes peaks which are present in all or most of the samples, thus visually identifying the dominant components present in the sample group. This calculation is included in the diagnostics output of the *chromall* algorithm, as well as a histogram of chain lengths, average integrated areas of each component, and the distribution of areas.

The *pkcomp* function is used for comparing the peak heights and areas of two samples or analyses. An example of a potential use for this function is determining the amount of material loss associated with urea adduction, a common procedure for noise reduction in gas chromatographic analysis. Comparing the peaks of a sample before and after urea adduction shows the percentage of material lost during the procedure.

The *plotchrom* function is a tool for easy data visualization, containing several options for plotting your data.

# Vignette

Use this section with the corresponding vignette.m file to run base scripts in Chroma. Use *help* in MATLAB for any function to see documentation.

1. To run functions in Chroma, input sample and standard files must be in a readable format. The function *prepfiles* reads most delimited or spreadsheet file extensions (e.g., .txt, .csv, .xls, .xlsx) and converts the data and header information into the Chroma format. If the file has sample ID information, the row in the file may be specified by the 'headerrow' argument. The base *chroma* function requires the input of a sample and standard chromatogram converted to this structure. The input data file should have the first column as the time and the second column as the intensity (any units are accepted). See the './Chroma/data/exampleDF.xlsx' for an example of a proper input file format.
2. Once the standard and sample data files are in the Chroma structure, you can run *chroma* to determine the sample peak areas. Required input arguments include the sample (DF), standard (RM), and the components present in the standard (refcomp). If the number of peaks in the standard chromatogram does not match the length of refcomp, *chroma* will return an error. In this case, we used the Indiana University at Bloomington standard B4 containing the *n*-alkane components C16-30. MA is the output table containing the peak times, areas, and heights.
   1. Change 'view' to 'yes' to open a diagnostic figure. Here, you can assess whether the peak detections were successful. If not, additional parameters may need adjusting.

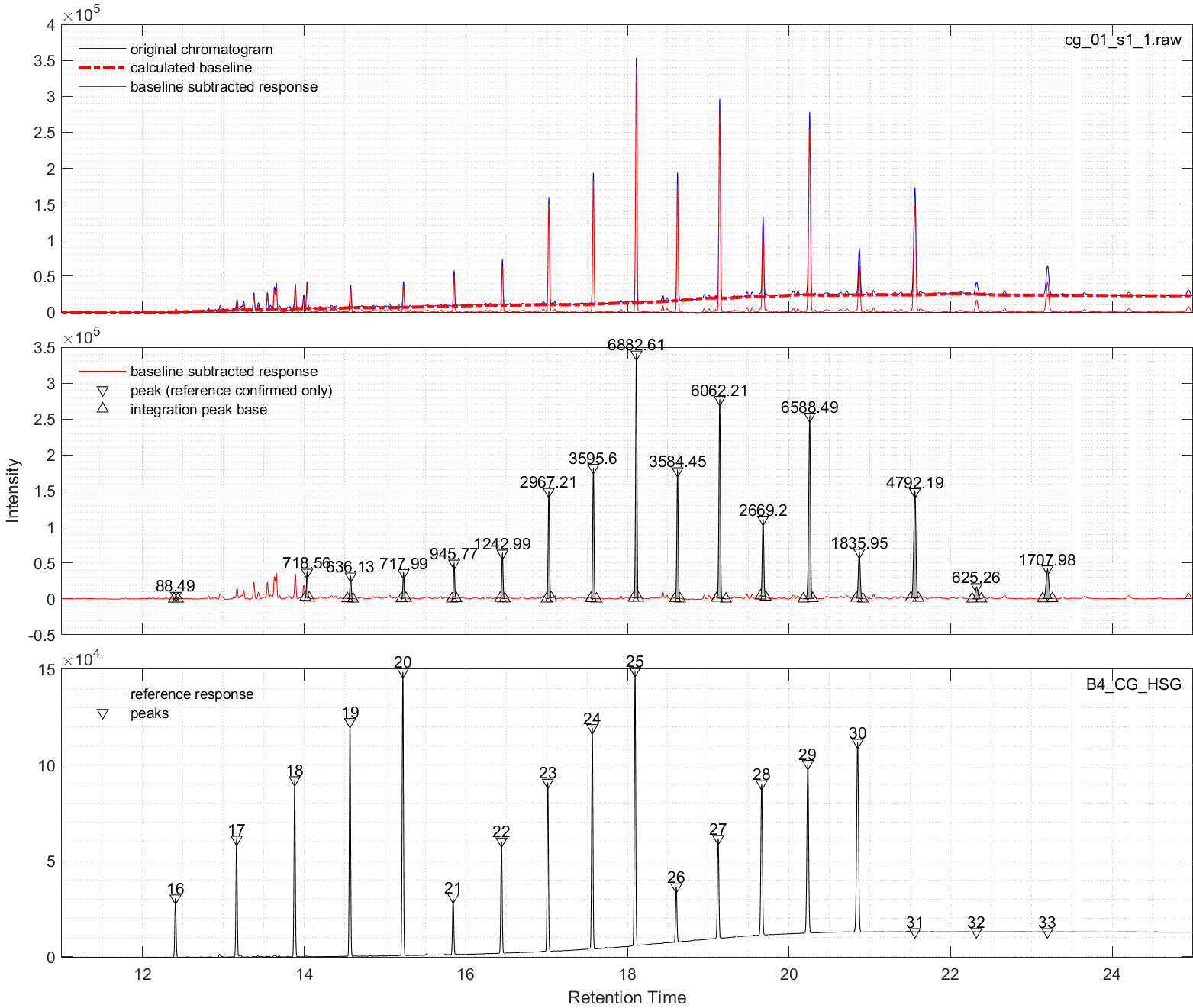


Fig. S1. Example of output diagnostic figure (CG 01) generated by *chroma* or *chromall* when the argument ‘view’ is set to ‘yes’. Components C31-33 were added by the user, based on the observed peak times in the sample, which demonstrate a characteristic odd-over-even distribution typical of intact terrestrial higher plants. The top panel shows the operation of subtracting the interpolated baseline in the sample. Integrated peak areas are indicated above each peak in the middle panel with up-facing triangles indicating the valley-to-valley integration boundaries and down-facing triangles indicating integrated peak identifications. Gray shaded areas show the integrated region of each peak, which are approximations for the relative abundances used in the index calculations. The bottom panel shows the reference standard used in the analysis with labeled peak component numbers. Syntax: chroma(DF, RM, refcomp, 'view','yes') or chromall(DF, RM, refcomp, 'view','yes').

1. The *chroma* function has several options for plotting and improved peak detections. These are all optional (with built-in default values), except for 'refcomp'
   1. 'refcomp' – a vector of numbers corresponding to components in the standard.
   2. 'pad' – add reference peaks not included in the standard. For example, B4 does not contain the C31-33 components, but they can be added to the analysis by specifying the target homologs in ‘pad’. To manually assign retention times to each target component, input for ‘pad’ should be a two-column with retention time in the first column (in the same units as the data) and the component number (e.g., 31, 32 33, etc.) in the second column. The user can also choose to automatically detect extra compounds in the trace by entering ‘pad’ as a single column matrix with the numbers of the target components only; this technique assumes a linear relationship between retention time and chain length.
   3. 'ds' – specify the reference detection window for peak correlations between the sample and standard. A smaller window provides a stricter criterion for peak correlations but may result in missed peaks if the standard and sample peak times do not align well. A larger window relaxes this criterion but may cause multiple sample peaks to be detected for a single sample, which will result in an error. In general, it is recommended to use the smallest value that is still able to capture all target peaks.
   4. 'view' – option for generating a figure ('yes' or 'no'). The figure is useful for diagnostic purposes. The top panel shows the spline interpolated baseline subtraction. The middle panel shows the integrated peak areas (gray). The bottom panel shows the standard and component numbers.
   5. 'smthreshold' – minimum peak threshold for sample chromatogram (units of intensity). Adjust this to filter out noisy peaks.
   6. 'rmthreshold' – minimum peak threshold for standard chromatogram (units of intensity). Make sure all standard peaks are above this threshold.
   7. 'xrange' – plotting range for x-axis (units of time) when 'view' is set to 'yes'.
   8. 'out' – output data format ('mat' or 'tab'). 'mat' will return a matrix and 'tab' will return a table.
   9. 'cutoff' – start time for the analysis. This tells *chroma* to ignore data before the specified time (useful for removing the solvent peak, for example).
2. If several samples need to be analyzed at once, the *chromall* function can read in multiple sample chromatograms and a single standard chromatogram to generate output hydrocarbon index data for all samples. The input sample data can be formatted from a single data file using *prepfiles*. Each sample should occupy two columns (for time and intensity) with no spaces between samples (see any of the Excel files in ./data for examples). For large datasets, formatting may take some time. The standard should still only contain one chromatogram. Input arguments for *chromall* are the same as *chroma*, with the addition of 'nfold' and 'prof':
   1. 'nfold' – directory for figure output when 'view' is 'yes'. If not specified, the figures will be put into a directory called 'chroma\_figs' in the current working directory. Output figures include diagnostic peak detection and pristane/phytane detection figures for each sample, a figure containing the calculated indices for the dataset, and a summary figure for distribution and uncertainty analysis.
   2. 'prof' – profile position of input sample chromatograms (i.e., depth, age, etc.). Values should be in the same order as the samples. If not specified, output figures will plot by the order number.

A graph of a graph

Description automatically generated

Fig. S2 Example of index calculation figure output from *chromall*. Syntax: chromall(DF, RM, refcomp, 'view','yes')

A group of graphs showing different sizes and shapes

Description automatically generated

Fig. S3. Example of summary figure output from *chromall*. Syntax: chromall(DF, RM, refcomp, 'view','yes')

A graph with numbers and symbols

Description automatically generated

Fig. S4. Example of plot made by *chromall* for examining the difference between retention times of homologs in the sample and external standard.

1. The function *pkcomp* compares the peak heights and areas of two samples. This is useful for identifying material lost during procedures like urea adduction. Input arguments require the sample and standard chromatograms for both samples, as well as the component numbers for each standard. Optional input argument used in *chroma* may be used in *pkcomp*. The plotting option is given by the argument 'plt'. The files example1.xlsx and example2.xlsx contain the chromatograms of HGZ sample 12 before and after urea adduction, respectively. The output table T shows the relative amount of material lost during the urea adduction procedure. Standard intensities were adjusted to share an appropriate 'rmthreshold' value. S1 and S2 correspond to the samples in the first and second arguments of the command.

A graph of a chemical reaction

Description automatically generated

Fig. S5. Example of analysis-to-analysis peak comparison in Chroma. Percentages indicate the area (abundance) of the original analysis (black) recovered after urea adduction (red). Peak height differences are indicated by the bracketed regions as intensity lost from the original.

1. The function *indall* can be used to generate a table of multiple index values from a single sample chromatogram. Use *chroma* to get the peak areas and component numbers, setting the 'out' argument to 'mat'. Then run *indall* from these areas. Individual indices can be calculated using the following functions: *acl*, *c31c19*, *cpi*, *lchsch*, *lh*, *oep*, *prph*, *paq*, *tarhc*, and *wi*.
2. The function *plotchrom* plots the chromatograms for easy visualization. Plotting options (after 'plt') include 'layer', 'stack', 'layout', and '3d'. The 'layer' option shows all available chromatograms in one panel window. The 'stack' option plots the sum of all chromatograms (normalized). The 'layout' option plots each chromatogram in individual panels. The '3d1' option plots all chromatograms in a rotatable (click and drag) 3D figure colored by sample number. The '3d2' option plots the same figure but colored according to intensity. The 'single' option plots a single chromatogram from the dataset, specified by the additional argument 'sn'.

A graph of different colored lines

Description automatically generated

Fig. S6. Example of output from *plotchrom* with syntax plotchrom(DF,'plt','layer','xrange',[11 25])

A graph of a number of objects

Description automatically generated

Fig. S7. Example of output from *plotchrom* with syntax plotchrom(DF,'plt','stack,'xrange',[11 25])

A graph of a graph

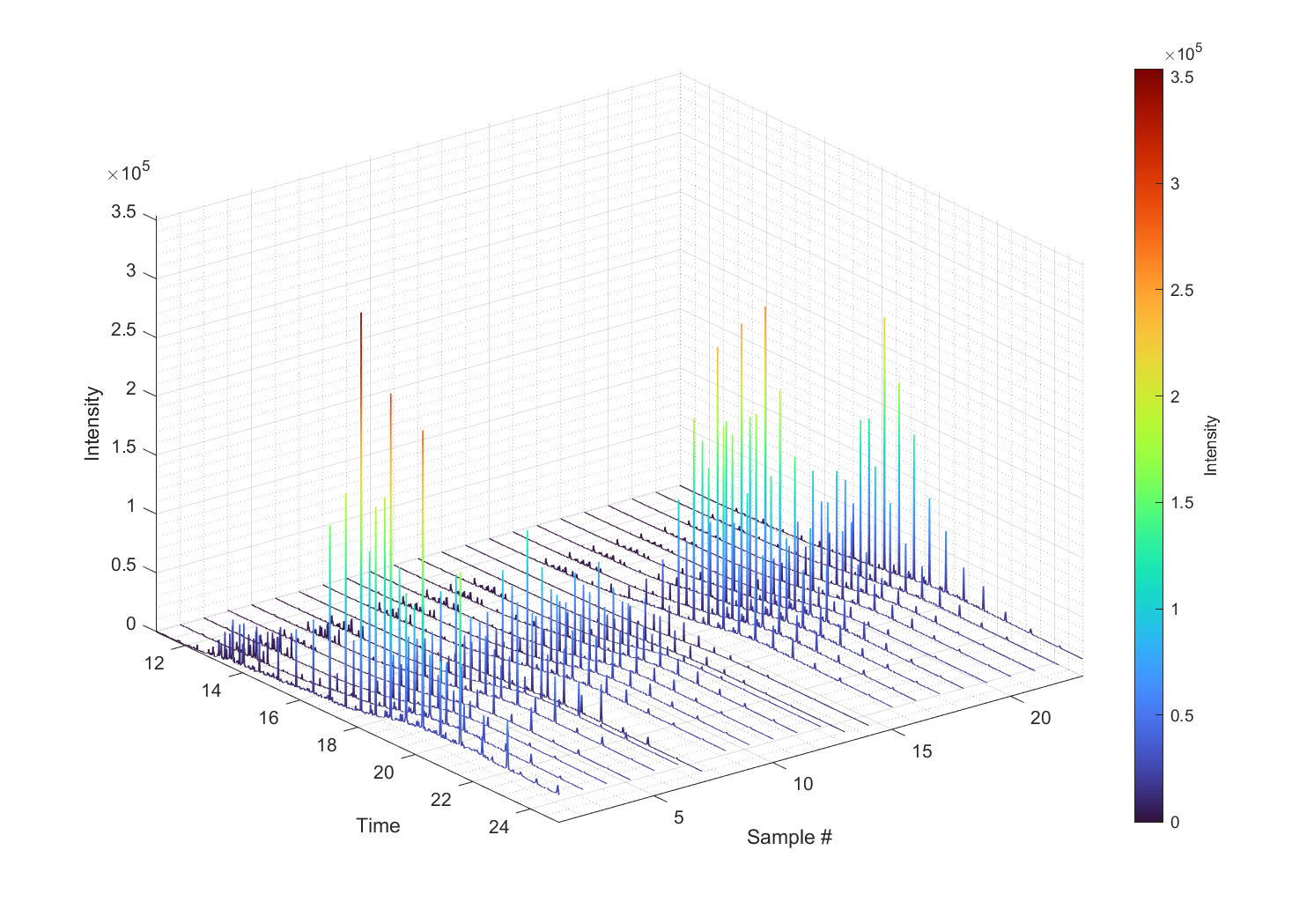
Description automatically generated

Fig. S8. Example of output from *plotchrom* with syntax plotchrom(DF,'plt','3d1','xrange',[11 25])

Fig. S9. Example of output from *plotchrom* with syntax plotchrom(DF,'plt','3d2','xrange',[11 25])

A graph with lines in the middle

Description automatically generated

Fig. S10. Example of output from *plotchrom* with syntax plotchrom(DF,'plt','single','sn',7,'xrange',[11 25]). This plots the 7th chromatogram in DF. If 'sn' is not specified, the first chromatogram in DF will be plotted.

1. Use *prph* to determine the abundance of pristane and phytane. When 'yes', the 'view' argument generates an output figure for assessing the automatic detection.

A graph of a graph of a person

Description automatically generated

Fig. S11. Example of output from *prph*. Syntax [PRPH,PR,PH] = prph(DF,RM,refcomp,'view','yes','ds',30', 'rmthreshold', 2.5e4,'xrange',[12 16])

1. Use *compcpi* and *compacl* to compare the output of different methods for calculating CPI and ACL, respectively. Optional detection input arguments are the same as *chroma* and *chromall*. The 'plt' argument options are '1:1', 'profile', and 'none'.

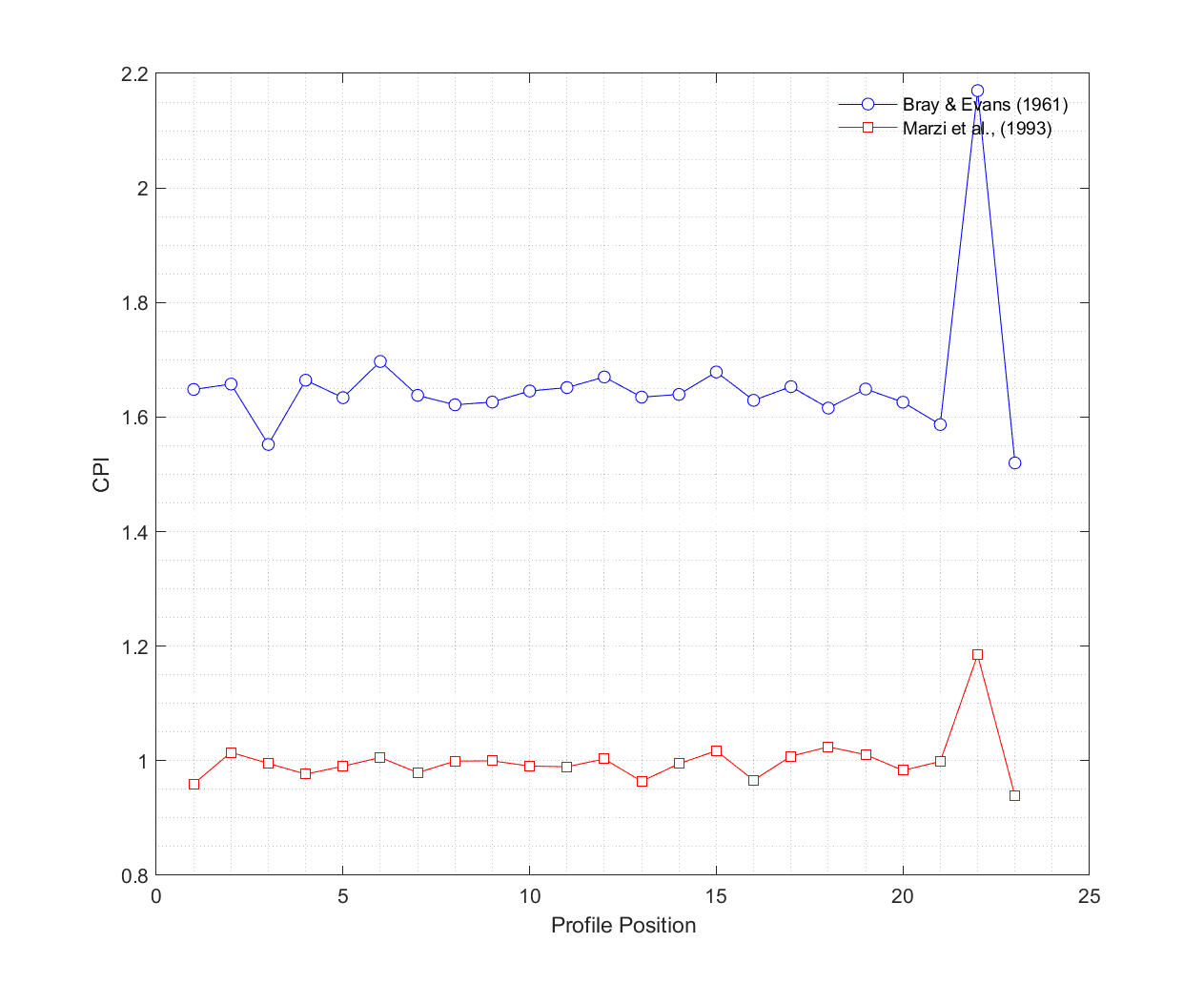


Fig. S12. Example of *compcpi* output figure with syntax [CPI2,CPIBE] = compcpi(DF,RM,refcomp,'profile');

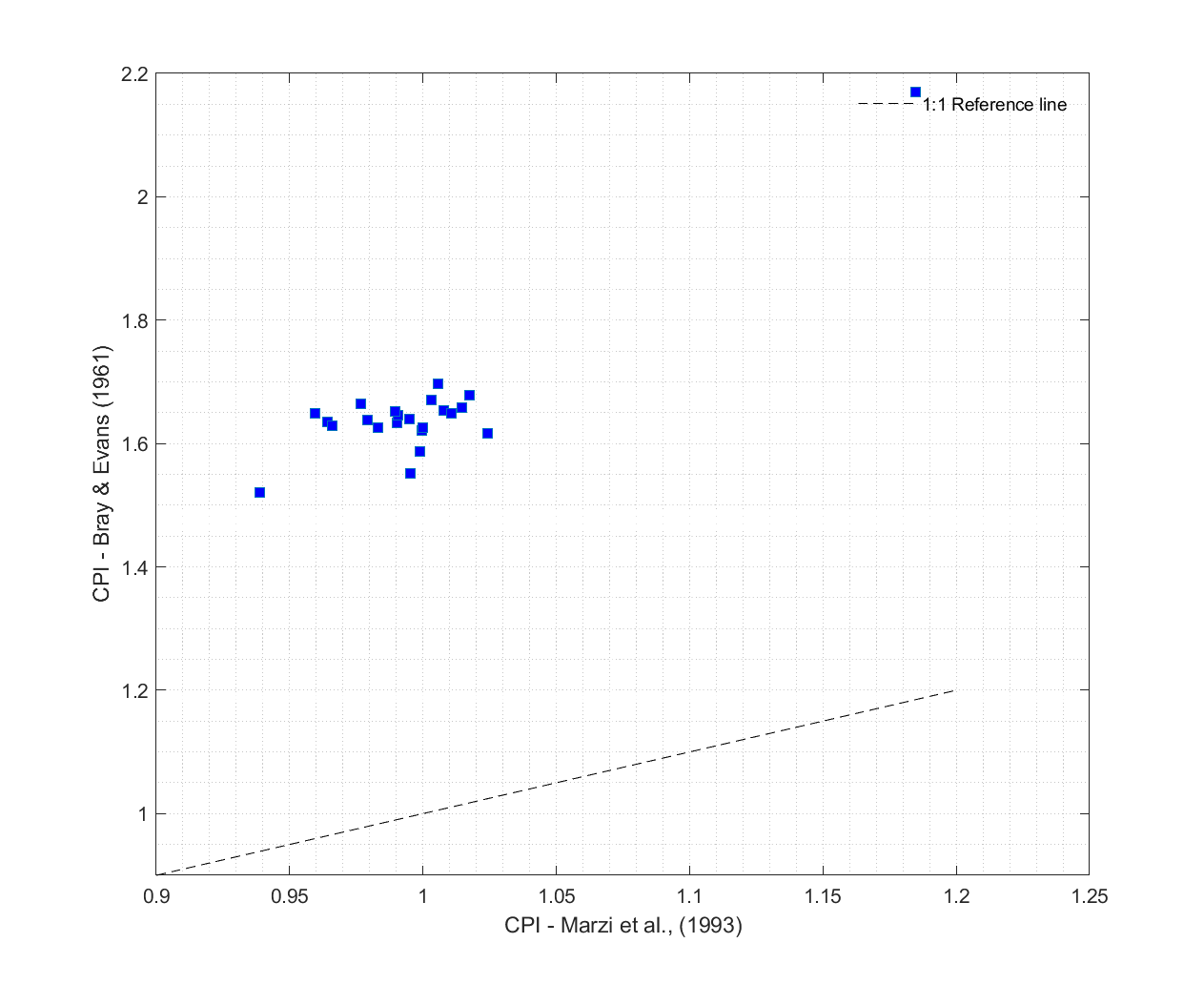


Fig. S13. Example of *compcpi* output figure with syntax [CPI2,CPIBE] = compcpi(DF,RM,refcomp,'plt','1:1');

A graph of different colored lines

Description automatically generated

Fig. S14. Example of *compacl* output figure with syntax [A1,A2,A3] = compacl(DF,RM,refcomp,'plt','profile');

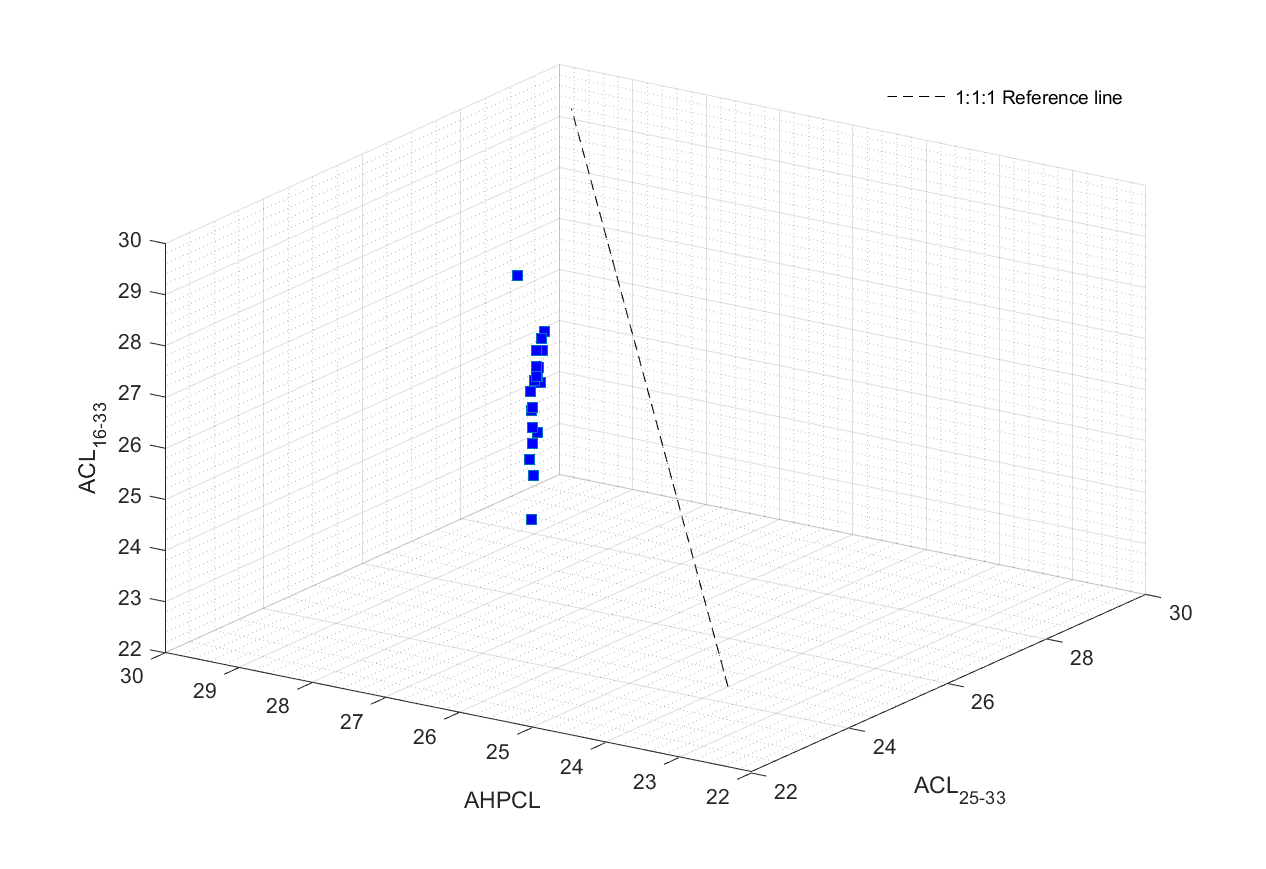


Fig. S15. Example of *compacl* output figure with syntax [A1,A2,A3] = compacl(DF,RM,refcomp,'plt','1:1');

# Description of Indices

Hydrocarbon indices of *n*-alkanes are unitless ratios of the molecular abundance of different carbon chain lengths and are used in various applications as proxies for degree of thermal alteration (Ishiwatari et al., 1977; Rabinowitz et al., 2017), paleoclimate change (Li et al., 2020; Rao et al., 2009), and maturity of organic matter (Rielley et al., 1991; Xie et al., 2004). The hydrocarbon indices available for calculation in Chroma include the following: carbon preference index (CPI), average chain length (ACL), *P*aq, terrigenous/aquatic ratio (TAR), odd-even predominance (OEP), low molecular weight/high molecular weight ratio (L/H), long chain/short chain ratio (LCH/SCH), and weathering index (WI). Chroma also determines the ratio of pristane and phytane (Pr/Ph), a common proxy for redox conditions (Large and Gize, 1996; Pancost et al., 1998).

## Carbon Preference Index (CPI) and Odd-Even Predominance (OEP)

CPI is a measure of the relative predominance of odd-numbered and even-numbered chains of carbon atoms of *n*-alkanes in a sample (Bray and Evans, 1961; Marzi et al., 1993):

where the abundance (or concentration) of each *n*-alkane component is denoted as C*n* hereafter, with *n* being the component number. The ratio is commonly used in petroleum geochemical analysis as an indicator for organic maturity and source, where higher values (CPI > 1) of the index reflect a preference toward odd-numbered carbon chains and therefore indicate terrestrially sourced organic matter and greater thermal immaturity (Eglinton and Hamilton, 1967). Terrigenous vascular (higher) plants typically have CPI values over 3, while lower CPI values (CPI < 1) may indicate a higher contribution of organic matter from petroleum and aquatic bacteria. Sediment and organic matter input from marine microorganisms and recycled material are represented in the CPI ≈ 1 range (Herrera-Herrera et al., 2020). We report CPI values using a component range of C23-33.

OEP is an alternative ratio for the relative proportion of odd-numbered and even-numbered chains (Scalan and Smith, 1970); the index is typically given in the form OEP = (C27 + C29 + C31 + C33)/(C26 + C28 + C30 + C32). Like CPI, OEP can be used as an indicator for the presence of kerogens. OEP values of 4-8 typically indicate terrestrial leaf wax *n*-alkanes, while lower values may indicate crude oil or aquatic lipids (Hoefs et al., 2002; Struck et al., 2020; Zech et al., 2009).

## Average Chain Length (ACL)

ACL is the weighted average of available chain lengths of *n*-alkane molecules (Jeng, 2006; Poynter and Eglinton, 1990):

Leaf lipids of vascular plants typically produce a greater number of *n*-alkanes with chain lengths in the C23-33 range and a strong predominance of odd-numbered components; *n*-alkanes produced by marine organisms like bacteria are typically shorter in chain length. Thus, ACL is commonly used as a proxy for vegetation type and associated climates. For example, grassland-derived leaf lipids produced in more arid climates may have higher ACL values than lipids produced by plants in relatively humid forests (Cranwell, 1973). The production of longer chain *n*-alkanes is climatically associated with warmer or more arid climates (Leider et al., 2013; Simoneit et al., 1991).

## Paq and Terrigenous/Aquatic Ratio (TAR)

*P*aq is a proxy for the relative input of submerged floating aquatic macrophytes and emergent or terrestrial plants in lacustrine environments (Ficken et al., 2000):

Typical values of *P*aq in modern plants generally follow these ranges: *P*aq = 0-0.1 for terrestrial plants, *P*aq = 0.1-0.4 for emergent macrophytes or a mixture of terrestrial and aquatic plants, and *P*aq = 0.4-1 for submerged and floating species (Ficken et al., 2000).

TAR is a similarly defined proxy index for algal productivity and terrestrial organic matter and is given by the relationship TAR = (C27 + C29 + C31)/(C15 + C17 + C19). TAR values of 1 and higher suggest plant lipids derived from terrestrial watershed sources, while lower values indicate a higher lipid matter contribution from aquatic sources (Bourbonniere and Meyers, 1996).

## Pristane/Phytane (Pr/Ph)

The Pr/Ph ratio is an index for interpreting the redox conditions of the depositional environment. The isoprenoid alkanes pristane (Pr) and phytane (Ph) are produced preferentially by the oxidation or reduction pathways of the chlorophyll constituent phytol, respectively (Didyk et al., 1978; Powell and McKirdy, 1973b). The index is sensitive to input from saline and evaporitic environments (Pr/Ph < 0.8) and oxic or suboxic terrigenous sources often associated with crude oil (Pr/Ph > 3). Generally, Pr/Ph < 1 indicates oil formed from sapropelic marine organisms, while Pr/Ph > 1 indicates oil formed along continental margins and coastal depositional environments (Bylinkin, 1987). The compounds are identified in gas-chromatography as detector responses directly following those of the *n­*-alkanes C17 (for pristane) and C18 (for phytane). The ratios of Pr/C17 and Ph/C18 are also used in source locations of crude oil, where open-ocean conditions produce Pr/C17 < 0.5 and inland peat swamps generate Pr/C17 > 1 (Brooks et al., 1969; Powell and McKirdy, 1973a). The alternative relationship (Pr + C17)/(Ph + C18), which was shown to be less influenced by thermal alteration, was proposed to account for the high aerobic biodegradation sensitivity in *n*-alkanes relative to pristane and phytane (Alexander et al., 1981).

## Other Hydrocarbon Indices

Some less frequently reported hydrocarbon indices of *n*-alkanes are also useful indicators for certain applications. The additional indices included in Chroma include L/H, LCH/SCH, WI, and C31/C19. L/H, given by L/H = ΣC15-20 / ΣC21-34, is a proxy for the relative input marine macrophytes to vascular plants, where values of L/H < 1 indicate contributions from terrestrial higher plants and L/H ≈ 1 is representative of degraded oil or planktonic input (Adeniji, 2017; Fagbote and Olanipekun, 2012; Gearing et al., 1976). The *n*-alkane ratio C31/C19 is a similar index for the relative abundance of terrigenous to aquatic biogenic sources, with C31/C19 > 0.4 indicates greater terrestrial input (Bush and McInerney, 2013; Yang et al., 2011). Another index for the type of biogenic source is LCH/SCH, a proxy for the predominance of phytoplankton-derived organic input (Fagbote and Olanipekun, 2012). LCH/SCH values, determined by LCH/SCH = (C27 + C29 + C31)/(C15 + C17 + C19), below 0.80 indicate high phytoplankton input and values above 4.0 indicate terrigenous origin. LCH/SCH values between 2.03 to 4.33 indicate a mixture of the two sources (Beard, 2007; Bolaji et al., 2023). Often, these indices are used when CPI values are close to 1 or the range of available sample *n*-alkanes is more appropriate. A combination of these proxies is ideal for confirming interpretations of CPI or other indices.

WI (or U/R) is specifically utilized for describing the weathering behavior of crude oil (Wang and Fingas, 1995); this index is calculated by WI = (C8 + C10 + C12 + C14)/(C22 + C24 + C26 + C28) and is negatively correlated with chemical weathering percentages, where WI values approach zero with increased weathering due to the preferential depletion of lower molecular weight *n*-alkanes relative to higher molecular weight *n*-alkanes (Volkman et al., 1984; Wang and Fingas, 1995; Wenger et al., 2002). WI is related to the ratio of the unresolved hydrocarbon components to the resolved components and can therefore be related to the prominence of an unresolved complex mixture (UCM) and petrogenic contamination (Gogou et al., 2000; Gough and Rowland, 1990; Killops and Al-Juboori, 1990); values of WI > 4 typically indicate an abundance of degraded petroleum residues and WI < 1 indicates lower contamination from crude oils (Adeniji, 2017).

# GC Conditions and Preparation of Samples Analyzed in Vignette

For reproducibility, relevant GC-FID settings and sample preparation steps are detailed here. All *n*-alkane detector responses were generated using a Thermo Scientific Trace 1310 Gas-Chromatography analyzer with a Flame Ionization Detector (GC-FID), fitted with a programmable-temperature vaporization (PTV) injector and TG-1MS column (60 m length, 0.25 mm inner diameter, 0.25 μm film thickness), and digitized in Xcalibur. The GC oven sequence is initialized at a starting temperature of 60°C (held for 1 min) and ramps to a final temperature of 320°C (held for 20 min) at rate of 15°C/min. Solutes were eluted to the FID with capillary column gas flows set at the following flow rates: hydrogen (35 ml/min), air (350 ml/min), and nitrogen make-up (35 ml/min). The carrier gas (helium) was set with a splitless constant column flow pressure of 240 psi at 2 ml/min. The sample solutes were prepared for GC-FID analysis by Soxhlet total lipid extraction using a solvent mixture of dichloromethane and methanol (DCM:MeOH; 2:1 *v*/v), then separated by column chromatography into aliphatic, aromatic, and polar fractions by eluting a series of organic solvents (hexane, DCM, then methanol, respectively) through an activated silica gel packed Pasteur pipet and collecting 4 ml of each eluate. Digitized chromatograms (time and intensity) were exported directly from Xcalibur and copied into an Excel spreadsheet (a delimited file may also be used) to be later read into Chroma. Once the data are saved to a file, the contents can be reformatted by Chroma for easy use in all package functions.

# References

Adeniji, A., 2017. Petroleum Hydrocarbon Profiles of Water and Sediment of Algoa Bay, Eastern Cape, South Africa. International Journal of Environmental Research and Public Health, 14: 1263.

Alexander, R., Kagi, R.I., Woodhouse, G.W., 1981. Geochemical Correlation of Windalia Oil and Extracts of Winning Group (Cretaceous) Potential Source Rocks, Barrow Subbasin, Western Australia1. AAPG Bulletin, 65(2): 235-250.

Beard, C.L., 2007. Environmental Chemistry of Waters and Recent (3420 ± 36 Years BP) Sediments in Tonle Sap, Cambodia. Ph.D. Thesis, The University of Manchester (United Kingdom), England, 367 pp.

Bolaji, T.A., Victor-Oji, C.O., Adedeji, O.A., 2023. Geochemical characterisation of the Turonian-Maastrichtian bituminous sands in Dahomey Basin, Southwestern Nigeria. Energy Exploration & Exploitation, 41(3): 922-940.

Bourbonniere, R.A., Meyers, P.A., 1996. Sedimentary geolipid records of historical changes in the watersheds and productivities of Lakes Ontario and Erie. Limnology and Oceanography, 41(2): 352-359.

Bray, E.E., Evans, E.D., 1961. Distribution of n-paraffins as a clue to recognition of source beds. Geochimica et Cosmochimica Acta, 22(1): 2-15.

Brooks, J.D., Gould, K., Smith, J.W., 1969. Isoprenoid Hydrocarbons in Coal and Petroleum. Nature, 222(5190): 257-259.

Bush, R.T., McInerney, F.A., 2013. Leaf wax n-alkane distributions in and across modern plants: Implications for paleoecology and chemotaxonomy. Geochimica et Cosmochimica Acta, 117: 161-179.

Bylinkin, G.P., 1987. The informativeness of the pristane/phytane ratio as a genetic index International Geology Review, 29(9): 1117-1120.

Cranwell, P.A., 1973. Chain-length distribution of n-alkanes from lake sediments in relation to post-glacial environmental change. Freshwater Biology, 3(3): 259-265.

Didyk, B.M., Simoneit, B.R.T., Brassell, S.C., Eglinton, G., 1978. Organic geochemical indicators of palaeoenvironmental conditions of sedimentation. Nature, 272(5650): 216-222.

Eglinton, G., Hamilton, R.J., 1967. Leaf Epicuticular Waxes. Science, 156(3780): 1322-1335.

Fagbote, E.O., Olanipekun, E., 2012. Characterization, Distribution, Sources and Origins of Aliphatic Hydrocarbons of the Soils of Agbabu Bitumen Deposit Area, Western Nigeria. African Journal of Scientific Research, 10: 1-24.

Ficken, K.J., Li, B., Swain, D.L., Eglinton, G., 2000. An n-alkane proxy for the sedimentary input of submerged/floating freshwater aquatic macrophytes. Organic Geochemistry, 31(7): 745-749.

Gearing, P., Gearing, J.N., Lytle, T.F., Lytle, J.S., 1976. Hydrocarbons in 60 northeast Gulf of Mexico shelf sediments: a preliminary survey. Geochimica et Cosmochimica Acta, 40(9): 1005-1017.

Gogou, A., Bouloubassi, I., Stephanou, E.G., 2000. Marine organic geochemistry of the Eastern Mediterranean: 1. Aliphatic and polyaromatic hydrocarbons in Cretan Sea surficial sediments. Marine Chemistry, 68(4): 265-282.

Gough, M.A., Rowland, S.J., 1990. Characterization of unresolved complex mixtures of hydrocarbons in petroleum. Nature, 344(6267): 648-650.

Herrera-Herrera, A.V., Leierer, L., Jambrina-Enríquez, M., Connolly, R., Mallol, C., 2020. Evaluating different methods for calculating the Carbon Preference Index (CPI): Implications for palaeoecological and archaeological research. Organic Geochemistry, 146: 104056.

Hoefs, M.J.L., Rijpstra, W.I.C., Sinninghe Damsté, J.S., 2002. The influence of oxic degradation on the sedimentary biomarker record I: evidence from Madeira Abyssal Plain turbidites. Geochimica et Cosmochimica Acta, 66(15): 2719-2735.

Ishiwatari, R., Ishiwatari, M., Rohrback, B.G., Kaplan, I.R., 1977. Thermal alteration experiments on organic matter from recent marine sediments in relation to petroleum genesis. Geochimica et Cosmochimica Acta, 41(6): 815-828.

Jeng, W.-L., 2006. Higher plant n-alkane average chain length as an indicator of petrogenic hydrocarbon contamination in marine sediments. Marine Chemistry, 102(3): 242-251.

Killops, S.D., Al-Juboori, M.A.H.A., 1990. Characterisation of the unresolved complex mixture (UCM) in the gas chromatograms of biodegraded petroleums. Organic Geochemistry, 15(2): 147-160.

Large, D.J., Gize, A.P., 1996. Pristane/phytane ratios in the mineralized Kupferschiefer of the Fore-Sudetic Monocline, southwest Poland. Ore Geology Reviews, 11(1): 89-103.

Leider, A., Hinrichs, K.-U., Schefuß, E., Versteegh, G.J.M., 2013. Distribution and stable isotopes of plant wax derived n-alkanes in lacustrine, fluvial and marine surface sediments along an Eastern Italian transect and their potential to reconstruct the hydrological cycle. Geochimica et Cosmochimica Acta, 117: 16-32.

Li, C. et al., 2020. Assessment of the relationship between ACL/CPI values of long chain n-alkanes and climate for the application of paleoclimate over the Tibetan Plateau. Quaternary International, 544: 76-87.

Marzi, R., Torkelson, B.E., Olson, R.K., 1993. A revised carbon preference index. Organic Geochemistry, 20(8): 1303-1306.

Pancost, R.D., Freeman, K.H., Patzkowsky, M.E., Wavrek, D.A., Collister, J.W., 1998. Molecular indicators of redox and marine photoautotroph composition in the late Middle Ordovician of Iowa, U.S.A. Organic Geochemistry, 29(5): 1649-1662.

Powell, T., McKirdy, D., 1973a. Relationship between Ratio of Pristane to Phytane, Crude Oil Composition and Geological Environment in Australia. Nature, 243: 37-39.

Powell, T.G., McKirdy, D.M., 1973b. Relationship between Ratio of Pristane to Phytane, Crude Oil Composition and Geological Environment in Australia. Nature Physical Science, 243(124): 37-39.

Poynter, J., Eglinton, G., 1990. 14. Molecular composition of three sediments from hole 717c: The Bengal fan, Proceedings of the Ocean Drilling Program: Scientific results, pp. 155-161.

Rabinowitz, H.S., Polissar, P.J., Savage, H.M., 2017. Reaction kinetics of alkenone and n-alkane thermal alteration at seismic timescales. Geochemistry, Geophysics, Geosystems, 18(1): 204-219.

Rao, Z. et al., 2009. CPI values of terrestrial higher plant-derived long-chain n-alkanes: a potential paleoclimatic proxy. Frontiers of Earth Science in China, 3(3): 266-272.

Rielley, G., Collier, R.J., Jones, D.M., Eglinton, G., 1991. The biogeochemistry of Ellesmere Lake, U.K.—I: source correlation of leaf wax inputs to the sedimentary lipid record. Organic Geochemistry, 17(6): 901-912.

Scalan, E.S., Smith, J.E., 1970. An improved measure of the odd-even predominance in the normal alkanes of sediment extracts and petroleum. Geochimica et Cosmochimica Acta, 34(5): 611-620.

Simoneit, B.R.T. et al., 1991. Molecular marker study of extractable organic matter in aerosols from urban areas of China. Atmospheric Environment. Part A. General Topics, 25(10): 2111-2129.

Struck, J. et al., 2020. Leaf wax n-alkane patterns and compound-specific δ13C of plants and topsoils from semi-arid and arid Mongolia. Biogeosciences, 17: 567-580.

Volkman, J.K., Alexander, R., Kagi, R.I., Rowland, S.J., Sheppard, P.N., 1984. Biodegradation of aromatic hydrocarbons in crude oils from the Barrow Sub-basin of Western Australia. Organic Geochemistry, 6: 619-632.

Wang, Z., Fingas, M., 1995. Study of the effects of weathering on the chemical composition of a light crude oil using GC/MS GC/FID. Journal of Microcolumn Separations, 7(6): 617-639.

Wenger, L.M., Davis, C.L., Isaksen, G.H., 2002. Multiple Controls on Petroleum Biodegradation and Impact on Oil Quality. SPE Reservoir Evaluation & Engineering, 5(05): 375-383.

Xie, S. et al., 2004. Molecular and isotopic stratigraphy in an ombrotrophic mire for paleoclimate reconstruction2 2Associate editor: R. Summons. Geochimica et Cosmochimica Acta, 68(13): 2849-2862.

Yang, C. et al., 2011. Chemical Fingerprints of Alberta Oil Sands and Related Petroleum Products. Environmental Forensics, 12(2): 173-188.

Zech, M. et al., 2009. Reconstructing Quaternary vegetation history in the Carpathian Basin, SE Europe, using n-alkane biomarkers as molecular fossils. Quaternary Science Journal, 58: 148-155.