Computers & Geosciences 72 (2014) 255–261

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

paleofire: An R package to analyse sedimentary charcoal records from the Global Charcoal Database to reconstruct past biomass burning ☆

Olivier Blarquez [a,b,*], Boris Vannière [c], Jennifer R. Marlon [d], Anne-Laure Daniau [e], Mitchell J. Power [f], Simon Brewer [g], Patrick J. Bartlein [h]

[a] Centre d'étude de la Forêt, Université du Québec à Montréal, Montréal, Québec, Canada
[b] Natural Sciences and Engineering Research Council of Canada Industrial Chair in Sustainable Forest Management, Forest Research Institute, Université du Québec en Abitibi-Témiscamingue, Rouyn-Noranda, Québec, Canada
[c] Centre National de la Recherche Scientifique (CNRS), UMR Chrono-Environment, Besançon, France
[d] Yale School Forestry and Environmental Studies, Yale University, New Haven, CT, USA
[e] Centre National de la Recherche Scientifique (CNRS), Environnements et Paléoenvironnements Océaniques et Continentaux (EPOC), Unité Mixte de Recherche(UMR) 5805, Université de Bordeaux, Talence, France
[f] Natural History Museum of Utah and Department of Geography, University of Utah, Salt Lake City, UT, USA
[g] Department of Geography, University of Utah, Salt Lake City, UT, USA
[h] Department of Geography, University of Oregon, Eugene, OR, USA

ARTICLE INFO

Article history:
Received 7 January 2014
Received in revised form
29 July 2014
Accepted 31 July 2014
Available online 8 August 2014

Keywords:
Charcoal
Fire
Biomass burning
Databases
R statistical language

ABSTRACT

We describe a new R package, paleofire, for analysis and synthesis of charcoal time series, such as those contained in the Global Charcoal Database (GCD), that are used to reconstruct paleofire activity (past biomass burning). paleofire is an initiative of the Global Paleofire Working Group core team (www.gpwg.org), whose aim is to encourage the use of sedimentary charcoal series to develop regional-to-global syntheses of paleofire activity, and to enhance access to the GCD data by providing a common research framework. Currently, paleofire features are organized into three different parts related to (i) site selection and charcoal series extraction from the GCD; (ii) charcoal data transformation; and (iii) charcoal series compositing and synthesis. We provide a technical description of paleofire and describe some new implementations such as the circular block bootstrap procedure. We tested the software using GCDv3 data from eastern North America, and provide examples of interpreting results of regional and global syntheses.

1. Introduction

In the last decade, regional and global syntheses of sedimentary charcoal records have been used to examine broad-scale patterns in palaeofire activity (Carcaillet et al., 2002; Power et al., 2008; Daniau et al., 2012). The linkages among fire, climate, vegetation and humans at centennial–millennial timescales have likewise been examined using global and regional syntheses (Marlon et al., 2008; Ali et al., 2012). Syntheses of charcoal records can also aid the validation and calibration of fire simulations (Flannigan et al., 2001; Pechony and Shindell, 2009; Girardin et al., 2013; Brücher et al., 2014). Because fire influences ecosystems at all spatio-temporal scales (ranging from days to centuries and from microsites to biomes), a growing interest in paleofire research has emerged. Additionally, future wildfire regimes may have no analogues from recent decades, and so identifying reference conditions and baselines in the past has become crucial to projecting future wildfire activity (Girardin et al., 2013). Sedimentary charcoal series from individual sites are distributed worldwide, and are increasingly included in the Global Charcoal Database (Power et al., 2010, GCD), which provides the scientific community a global charcoal dataset for research and archiving for sedimentary records of fire (the GCD is available at http://gpwg.org/gpwgdb.html). Syntheses of spatio-temporal changes in fire at global (Marlon et al., 2008; Power et al., 2008; Daniau et al., 2012; Marlon et al., 2013) and regional

☆ paleofire package version 1.1.3

* Corresponding author at: Department of Geography, Université de Montréal, Montréal, Québec, Canada.
E-mail address: blarquez@gmail.com (O. Blarquez).

http://dx.doi.org/10.1016/j.cageo.2014.07.020
0098-3004/

(Marlon et al., 2009; Mooney et al., 2011; Vannière et al., 2011; Power et al., 2013) scales were obtained by applying several analytical steps implemented by a set of Fortran programs (Bartlein, unpublished). Because these statistical methods are not easily usable or modifiable, the Global Paleofire Working Group core team has developed a package using the open-source R statistical programming language. The new package should increase accessibility to paleofire data while providing the fire-science community with new analytical tools that include and extend the previously used functions for GCD data extraction and statistical analysis.

The aim of this paper is to describe the paleofire R package that facilitates the analysis of charcoal records contained in the GCD. The paleofire package functions are organized into three parts: (i) GCD site selection and data extraction using a variety of criteria (geographic, sedimentary, etc.), (ii) charcoal data transformation, including re-scaling, single-record variance homogenization and nonparametric trend estimation and (iii) data synthesis, including confidence limit estimation using resampling procedures.

2. Global Charcoal Database to paleofire data synthesis

The paleofire package works in conjunction with the GCD R package that contains a simplified version of the charcoal dataset in order to accommodate the different update frequency between the paleofire package (frequent updates) and the Global Charcoal Database (infrequent updates). The checkGCDversion() function can be used to determine whether the GCD data package is current. If it is not, the function asks whether the user wants to update the data. The minimal package version numbers required for running the examples presented in this study are 1.1.3 and 3.0.3 for the paleofire and GCD packages respectively. Backward compatibility will be ensured from these versions.

As of August 8, 2014, the GCD v3.0.3 contains a total of 736 charcoal records and is provided as a Microsoft Access database available at http://gpwg.org. The GCD data package is a simplified and reduced version of the GCD. The package consists of two data frames containing site metadata and charcoal data. The two data frames combine several tables from the GCD in order to simplify analysis in R, and exclude chronology development information, such as radiocarbon dates; these will likely be added however in future releases. Although most analyses in paleofire may use data directly from the dataframes in the GCD package, it is also possible to analyze user-defined database extracts or other charcoal series not in the database using the pfAddData function. There is currently no mechanism for permanently adding data to the GCD package automatically; interested contributors should contact the GPWG instead. The site metadata is accessible by typing the data(paleofiresites) command at the R prompt. This data frame provides a unique identifier for each site in the id_site column and associated metadata such as chronological, sedimentary or geographical information.

The raw charcoal data are accessible with the data(paleofiredata) command. The data consist of a seven column data frame containing: (i) site unique identifier, (ii) sample depth, (iii) sample age, (iv) sample charcoal value, (v) sample charcoal unit, (vi) extraction method (sieved charcoal, pollen slide charcoals, etc.) and (vii) sample type unit (influx or concentration). The default setting for the R package is to select the preferred units (e.g., concentration or influx values are typically preferred over charcoal-to-pollen ratios if both are available) used in previous analysis (see Daniau et al., 2012 or Power et al., 2008), however paleofire allows one to analyse charcoal records with user-defined units or methods (e.g. sieved *vs* pollen slide charcoals, see pfSiteSel and pfTransform functions help for details).

3. Technical description of paleofire

Here we provide a technical description and several illustrative examples of paleofire. The paleofire package is written in the R scientific computing language R Core Team (2011), and was developed under the R 3.0.3 version but remains compatible with R >= 2.10.0. The functions in paleofire are arranged into three groups associated with data selection, charcoal series transformation, and synthesis. We used the S3 method scheme to implement generic plotting and summary functions.

To present some of the paleofire capabilities, the examples below use charcoal series from Eastern North America. For additional examples and a detailed overview of individual functions, the reader is referred to the online help available at http://cran.r-project.org/web/packages/paleofire/paleofire.pdf.

3.1. Site selection

Two functions are dedicated to site selection. The first one, pfInteractive requires users to interactively draw a polygon on a map to select sites with respect to their geographic location. The function returns a list object containing site names and identifiers that is further used in the following analysis steps and is called using pfInteractive().

```
> install.packages("paleofire", repo = " http://cran.r-project.org")
> library(paleofire)
```

The pfSiteSel function is more versatile than pfInteractive and has arguments for a variety of user-defined criteria. In the example below we select charcoal series between 30° and 90° latitude and −100° and −50° longitude, and include only those with at least one geochronological (^{14}C or ^{210}Pb dating, tephra layer, etc.) control point each 2500 year.

```
> ID <- pfSiteSel(lat > 30 & lat < 90, long > -100 & long < (-50), date_int <= 2500)
> length(ID$id_site)
[1] 71
```

Seventy-one sites are selected and stored in the ID object of the class pfSiteSel. The summary function associated with the pfSiteSel object returns a table (Table S1) containing site information, including geographic (latitude, longitude and elevation) and chronological descriptors (number of chronological control points, number of samples, minimum and maximum estimated ages). In the