DMTA Data Prep Vignette

Ryan Haupt <ryan.j.haupt@vanderbilt.edu>

December 16, 2011

Abstract

The purpose of the DMTA Data Prep Vignette is to show a user how to use the package to prepare a finalized results spreadsheet from raw Dental Microwear Texture Analysis (DMTA) results. This package is best suited for someone already familiar with DTMA methods, especially one with access to Scale Sensitive Fractal Analysis (SSFA) software which provides user with .csv files ideally suited for use with this package.

Contents

1. Background

Dental Microwear Texture Analysis (DMTA) is a method for determining diet by looking at the microscopic texture patterns on the enamel of teeth. The teeth are analyzed with a scanning white light confocal microscope and four adjacent scans are made per tooth. After scanning the tooth, Scale Sensitive Fractal Analysis (SSFA) software calculates values of complexity (Asfc), anisotropy (epLsar), scale of maximal complexity (Smc), Total fill volume (Tfv), and heterogeneity of complexity at various scales (HAsfc) for each portion of the scan. Further information about the exact definition of each of these characteristics can be found in Ungar et al. 2003. The results from the SSFA software are given as three separate .csv files. It is then up to the user to manually organize, process, and calculate median values for each characteristic using Ungar's DMTA Manual and Excel templates, neither of which are publicly available.

It is unlikely that anyone with access to SSFA software does not also have access to the DMTA Manual and Excel templates. Thus I argue that it would be more efficient and less labor intensive to use this implementation which carries out the task of formatting, calculating, and organizing data automatically.

2. Methods

The manual for how to process this data using Excel is not publicly available nor is it available for citation. My methods are modified from the original steps based on my own preferences for data compilation and organization.

This code begins with the .csv file containing Asfc, epLsar, and Smc data. It adds columns for taxa, specimen, facet, scan, and place. The filename column is broken into parts and used to fill in each new column except place which is done automatically as a repeating sequence of "A", "B", "C", and "D." Place and scan are double-checked against one another to make sure they match up, then the entire data set is sorted by place. The data set is broken down into four separate data sets, one for each scan (e.g. A, B, C, or D). Then a median value for each characteristic is calculated from the four scan data sets and the final value is added to a final data frame.

Next the .csv file with Tfv data is formatted with the addition of a place column. The place and scan columns are double-checked, as before, then the data set is sorted. As above, the data set is subdivided into separate data frames for each scan, and median values are calculated. The final results are added to the data frame containing the medians for Asfc, epLsar, and Smc.

The final .csv file contains heterogeneity data. Columns for taxa, specimen, facet, scan and place are added. Once again, filename is used to fill in appropriate information for the new columns except place which is a sequence of repeating letters A through D. Place is compared to scan to make sure the columns match, then the entire data set is sorted by place. Splitting the data into separate data frames based on scan (A, B, C, or D), the code then calculates median values for HAsfc 2x2 through 11x11 and outputs the results as a new data frame.

The Master_prep function then compiles both new data frames and organizes them with user-provided date. The finalized data frame can then be exported by the user or analyzed within the R environment.

3. Example

This example is based upon real and as yet unpublished data. However, the data has been modified to a degree where it does provide publishable results upon completion of this implementation.

4. Future Work

The most pressing improvement to the code that could be made is the ability to test not only against specimen id but also against facet. Sometimes the same specimen is scanned multiple

times for certain types of analyses, but as of this version the code cannot distinguish between these variations.

Future versions of this code may provide more robust statistical implementations for the user. There are tests which are common the most DMTA studies which could be run automatically upon completion of the master file. Furthermore, code could be improved to include methods to aggregate implementation into a true "Master spreadsheet." Functions could then be run which recognize different taxa and run the appropriate statistical comparisons of DMTA characteristics for the user.

5. References