Assessing *Serenoa repens* (Arecaceae) Quality at the Retail Level Using Spectroscopic and Chemometric Methods

Bryan A. Hanson*, Tao Ye[+] & M. Daniel Raftery [+]

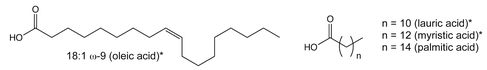
**Department of Chemistry & Biochemistry, DePauw University, Greencastle Indiana 46135 USA*

[+] Department of Chemistry, Purdue University, West Lafayette, Indiana 47907 USA

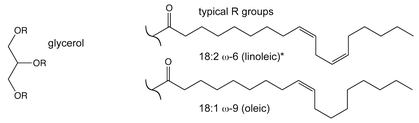
Background

Quality control and authenticity of medicinal plant products is critical. The lipido-sterolic extract of *Serenoa repens* (Arecaceae) is an important herbal treatment for benign prostatic hyperplasia (BPH) in men. The free fatty acids present in the extract act by inhibiting the formation of dihydrotestosterone from testosterone via the prostatic specific enzyme 5-α-reductase. Typically the extract is obtained via extraction with hexane or supercritical carbon dioxide. Our goal was to develop and apply appropriate statistical methods for the spectroscopic analysis of retail samples of *Serenoa repens*. NMR and IR spectra of 16 retail samples of *Serenoa repens* were recorded, along with spectra of evening primrose oil and olive oil as outliers. The spectra were analyzed using multivariate chemometric methods appropriate to high dimensional data sets, using functions written in the open source computing environment R.

Serenoa repens (saw palmetto) contains mostly free fatty acids

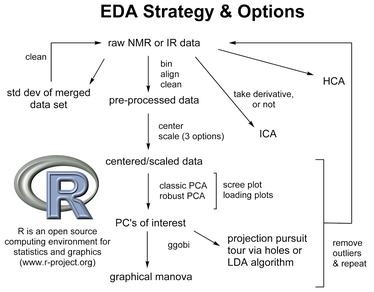


Olive Oil & Evening Primrose Oil (outgroups in this study) contain triglycerides



* these fatty acids are the strongest inhibitors of 5-α-reductase (depends upon isoform)

EDA Strategy & Options



Always return to the spectra & interpret them in light of preliminary results!

Summary of Methods & Terms Used in Exploration of High Dimensional Data Sets

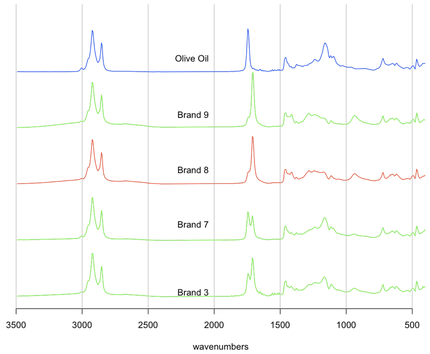
High Dimensional Data Sets	Data sets composed of relatively few samples but many observations on those samples (e.g. 32K NMR data points). Many classical algorithms fail under these conditions.
Chemometrics	The use of appropriate statistical methods to look at large chemical or spectroscopic data sets and extract information from them, make predictions and so forth.
EDA (exploratory data analysis)	A general term referring to a heuristically guided data mining process. *Let the data speak!*
HCA (hierarchical cluster analysis)	Computes "distances" of each sample from the others, producing a cladogram showing similarities.
PCA (principal component analysis)	Computes a set of PC's which represent the minimum number of factors needed to fully describe the data, discarding noise in the process.
ICA (independent component analysis)	Computes independent components, which in principle correspond to the spectra of the separate compounds, or perhaps compound classes.
Projection Pursuit Tour	A method of finding informative views of high dimensional data sets which fall into various categories.
Outliers	Data points which may represent a measurement error, or an incorrectly classified sample.
Robust Methods	Statistical methods which deal with outliers in a more sophisticated way than classical methods.

Analysis of IR Spectra

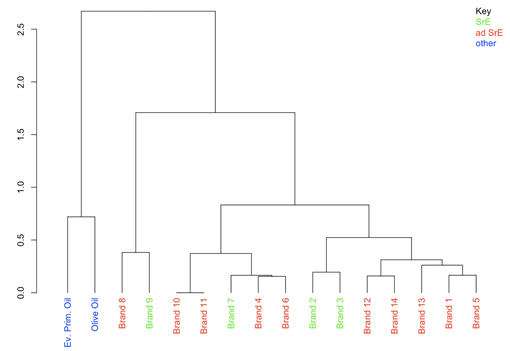
Data Abbreviations

SrE	*Serenoa repens* extract
ad SrE	adulterated extract (typically added olive oil)
other	outgroups: olive oil or evening primrose oil

Sample Serenoa repens Extract IR Data

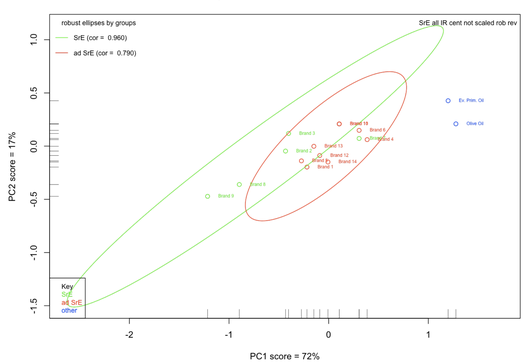


Serenoa repens Extract IR Data: HCA Analysis



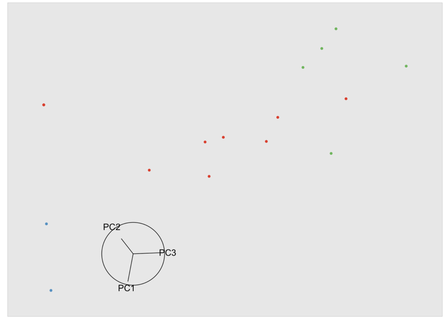
Based upon initial examination of PCA results, & reinspection of the IR spectra, it was determined that Brand 8 was adulterated with safflower oil, which contains mostly free fatty acids. Therefore it was reclassified as a "pure" SrE for subsequent analyses.

Serenoa repens Extract IR Data: PCA Score Plot

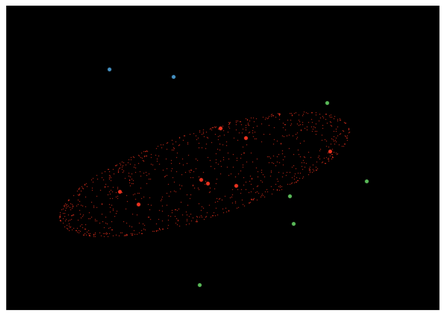


Analysis of IR Spectra, con't

IR Results: Projection of 1st 3 PCs Using ggobi 2D Tour



IR Results: Graphic Manova of 1st 3 PCs Using ggobi

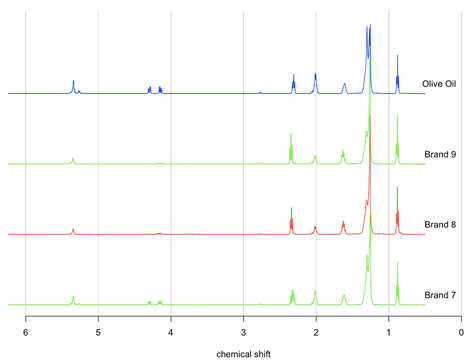


(ellipse drawn at 90% confidence level)

****

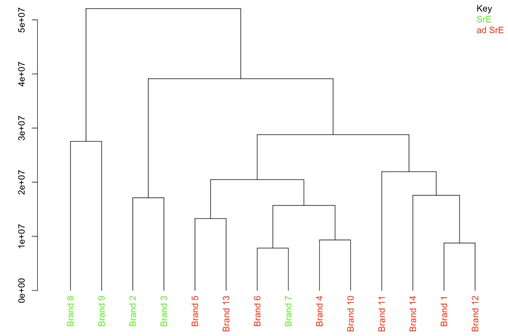
Analysis of NMR Spectra

Sample Serenoa repens Extract NMR Data



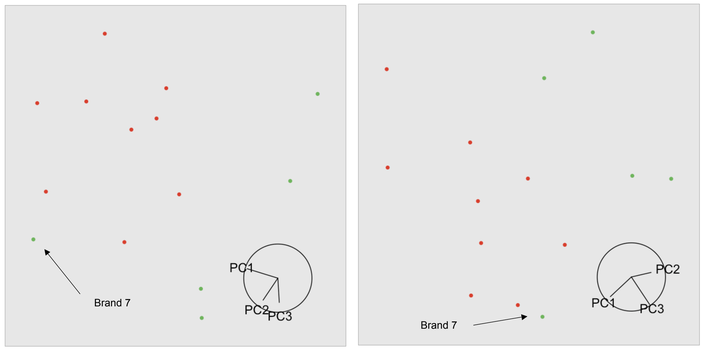
Analysis of NMR Spectra, con't

Serenoa repens Extract NMR Data: HCA Analysis



NMR Results: Projection of 1st 3 PCs Using ggobi 2D Tour

Left: Robust PCA. Right: Classic PCA



Conclusions

PCA on either the IR or NMR data readily distinguishes non-SrE samples (which are esters) from SrE samples (which are fatty acids). This might not be surprising in principle, as the spectra of these compound classes are rather different, but keep in mind that most SrE samples have added olive oil. Apparently PCA is sensitive to subtle features in the data, features which a spectroscopist might miss.

Among the SrE samples, preliminary explorations indicated that brand 8, an adulterated SrE, behaved like a pure SrE. As it turned out, it was adulterated with safflower oil, which is composed of free fatty acids (and not olive oil, an ester), so it was reclassified as a pure SrE for further analysis. Further work to distinguish between adulterated SrE and pure SrE was only partially successful, giving clusters that overlapped. One reason for this is that the adulterated SrE samples have added olive oil, but the amount added is not known (though from the spectra it appears that the amount of olive oil is 40-60% of the total). For both NMR and IR data, and by all chemometric methods, brand 7 appears to be an adulterated SrE, though the evidence is not strong. The label does not indicate any added ingredients.

IR and NMR spectra were about equally useful in discriminating among the samples. HCA provided a good initial view of the data. In the PCA experiments, unscaled data (in which large peaks exert a greater influence on the results) were more informative than autoscaled data. Robust and classical PCA methods were both effective. Which one was superior depended upon the particular data set.

Presented at the Society for Economic Botany meeting, June 2008