Application note

Classification of herbs by FT-NIR spectroscopy

Authors

Martin Hollein, Nicolet CZ s.r.o., Prague, Czech Republic, Todd Strother, Thermo Fisher Scientific, Madison, WI, USA

Introduction

Vibrational techniques like Fourier transform near-infrared (FT-NIR) spectroscopy are well-suited for raw material identification. This is because FT-NIR is sensitive to the characteristic molecular vibrations that occur for specific compounds. Because of this ability to detect molecular vibrations, NIR spectrometers are widely used in the pharmaceutical industry for classifying pure chemicals or their mixtures. Compounds usually contain highly absorbing functional groups (alkyls, phenyls, amines, thiols, hydroxyls, acids, esters, etc.), which allow for unambiguous identification within seconds. The Thermo Scientific™ Antaris™ II FT-NIR Analyzer line (Figure 1) is ideally suited for raw material identification of pure chemicals as well as complex biological samples used in pharmaceutical, nutraceutical, and related industries.

While herbs and their extracts are often used in cosmetics and nutraceuticals, they also find use as raw materials in classic pharmaceutical industries. Often, these herbal active ingredients, or their intermediates, are isolated from raw materials using extraction or chromatographic techniques. Herbs used in pharmaceuticals must be treated like any other raw material in the manufacturing facility and must fulfill all of the same regulatory demands.

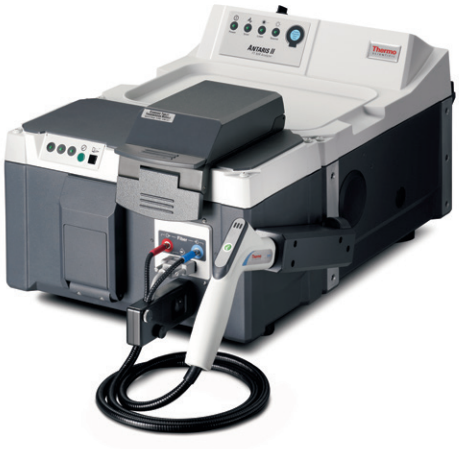

Figure 1. The Antaris II FT-NIR Analyzer showing different available detection options.

thermo scientific

Researchers at IREL, spol. s.r.o. (Brno, Czech Republic) explored FT-NIR spectroscopy as a means to rapidly identify the herbs used to produce extracts found in nutraceuticals, cosmetics, and pharmaceuticals. Their biological origin means that many herbs contain similar components, like cellulose, proteins and sugars, which tend to result in similar spectra. The purpose of this application note is to demonstrate that complex herbal materials can be correctly classified using FT-NIR spectroscopy much more rapidly than conventional techniques.

Experimental

The standard herbs used to develop the analysis method were supplied by IREL, spol s.r.o. and used directly to produce the fluid extracts—no additional grinding or sieving was performed. Particle size varied substantially among the samples. Leaves were typically thin and as large as 6x6 mm. Roots, galls, and orange peels were crushed to maximum particle sizes of 10x10x5 mm. The list and description of the 14 herbs used in the data collection are summarized in Table 1.

Herbal name	Scientific name
Agrimony—top	Agrimonia eupatoria L.
Buckbean—leaf	Menyanthes trifoliata L.
Calamus—root	Acorus calamus L.
Chamomile—flowers	Matricaria recutita
Gentian—root	Gentiana lutea L.
Hot pepper—powdered fruit	Capsicum sp.
Myrrh—powdered gum resin	Commiphora molmol E.
Oak apple—crushed gall	Diplolepis quercus-folii
Orange—peel	Citrus Aurantium L.
Sage—top	Salvia officinalis L.
Valerian—root	Valeriana officinalis L.
Walnut—leaf	Juglans regia L.
Witch hazel—leaf	Hamamelis virginiana L.
Wormwood—top	Artemisia absinthium L.

Table 1. List of herbal samples used in analysis.

Three or four samples of each herb were scanned with an Antaris II FT-NIR Analyzer in the range between 10,000 and 4,000 cm^{-1}. The materials were placed in a closed rotating sample cup with a 30 mm window over the integrating sphere (Figure 2). Spectra and other calibration data were collected and archived using validated Thermo Scientific RESULT Software. Each sample spectra was the result of 50 scans at 4 cm^{-1} resolution. These parameters allowed the material to be scanned through two full rotations of the sample cup in less than one minute.

Figure 2. Antaris II Analyzer with the closed sample cup and spinner.

The resulting spectra were evaluated with Thermo Scientific TQ Analyst Software using a discriminant analysis algorithm in order to classify the samples. Multiplicative signal correction (MSC) pathlength type was chosen because it is designed for analysis of samples, such as solids, where it is difficult to obtain an independent pathlength measurement. Because of light scattering, pathlength is then treated as a multiplicative contribution to the spectral signal. The software applies the same function to the standards and the unknown samples. The raw spectra in the range between 9,900 and 4,100 cm^{-1} were used for the calibration algorithm. A linear removed baseline correction was used but no other mathematical smoothing or derivative functions were applied.

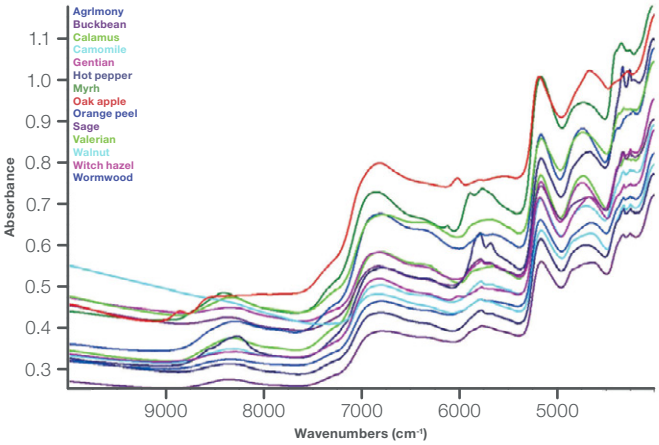

Figure 3. Representative spectra of the 14 samples used in the analysis.

Results

Figure 3 shows the average diffuse reflection spectra of the 14 herbs. The samples show substantial spectral variation that can be exploited for building the classification model. Using discriminant analysis as the chemometric method, 99.5% of the spectral variability was described with five principal components. Principal components are orthogonal vectors that describe spectral variation in a set of standards. The first principal component will account for most of the spectral variation and subsequent components account for the remaining variation. Generally, the fewer components required to describe the total variation, the more robust the method.

This method correctly classified all of the standards as well as randomly chosen validation spectra from each class. One means of determining the quality of a chemometric method is through Mahalanobis distance analysis. The Mahalanobis distance can be described as the spectral distance a particular sample is from the "center of mass" of the group. For samples that are spectrally similar to a particular class or group, the Mahalanobis distances will be relatively small. Samples that are spectrally very different than the class will have high Mahalanobis distances. For the samples under consideration, the Mahalanobis distances of the nearest incorrect class were at least twice the distance to the correct class; indicating good spectral separation and accurate classification. The distances of the nearest and next nearest classes are shown in Table 2.

Class ID	Distance	Next distance	Next class
Agrimony	0.5	4.0	Chamomile
Agrimony	0.7	3.6	Chamomile
Agrimony	1.0	3.5	Buckbean
Buckbean	0.2	3.3	Chamomile
Buckbean	0.2	3.3	Chamomile
Buckbean	0.4	3.1	Chamomile
Calamus	0.7	2.8	Orange peel
Calamus	0.7	3.7	Orange peel
Calamus	0.4	3.0	Orange peel
Chamomile	1.0	2.1	Wormwood
Chamomile	0.9	2.1	Wormwood
Chamomile	0.6	2.2	Wormwood
Gentian	0.2	5.0	Calamus
Gentian	0.4	5.0	Calamus
Gentian	0.4	4.8	Calamus
Hot pepper	0.1	13.2	Wormwood
Hot pepper	0.1	13.3	Wormwood
Hot pepper	0.1	13.2	Wormwood
Myrrh	0.3	7.4	Sage
Myrrh	0.4	7.3	Sage
Myrrh	0.5	7.3	Wormwood
Oak apple	0.5	5.5	Witch hazel
Oak apple	0.9	5.0	Witch hazel
Oak apple	0.8	6.2	Witch hazel
Oak apple	0.9	6.4	Witch hazel
Orange peel	1.3	3.1	Calamus
Orange peel	0.8	2.8	Calamus
Orange peel	0.7	3.6	Calamus
Orange peel	1.2	3.6	Calamus
Sage	1.0	4.6	Wormwood
Sage	1.0	6.2	Wormwood
Sage	0.8	5.0	Wormwood
Valerian	1.1	7.0	Gentian
Valerian	0.9	6.6	Gentian
Valerian	0.7	6.6	Gentian
Walnut	1.6	10.6	Agrimony
Walnut	2.0	13.0	Agrimony
Walnut	1.1	10.2	Agrimony
Walnut	1.7	11.3	Agrimony
Witch hazel	0.8	5.9	Agrimony
Witch hazel	0.9	5.7	Oak apple
Witch hazel	0.6	5.5	Oak apple
Wormwood	0.6	2.2	Chamomile
Wormwood	0.7	1.8	Chamomile
Wormwood	0.2	2.1	Chamomile

Table 2. Chart showing the Mahalonobis distance of each sample to its identified class. The distance to the next nearest class is included as a quantitative indication of how close the samples are to an incorrect class. Lower numbers indicate spectral similarity to a class, higher numbers indicate spectral difference. All of the samples were correctly identified.

Principal component score plots show spectra variation for samples in two dimensions. Figure 4 is a principal component score plot that shows clusters of different classes described by the first two principal components. Most of the classes are tightly clustered with standards that are near each other. Extremely homogeneous samples were most tightly clustered (i.e., powdered hot pepper). Those that have dispersed clusters are classes with irregularly shaped particles or heterogeneous samples (i.e., walnut leaf, crushed oak apple gall). The principal component score plot shows no overlap between the different classes when viewed in multidimensional space indicating that the samples can be successfully and easily classified.

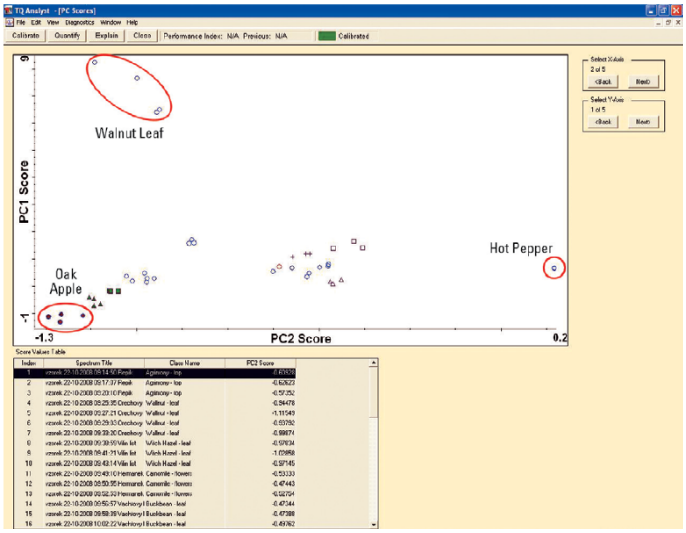

Figure 4. Principal component score plots showing class clustering as a function of the first and second principal components. Heterogeneous samples tended to have widely scattered clusters (walnut leaf, oak apple), while homogeneous samples were tightly clustered (hot pepper). Clusters were well defined and there were no overlapping clusters for the classes when viewed in 5-dimensional principal component space.

Conclusion

The Antaris II FT-NIR Analyzer offers an excellent alternative to traditional identification and classification methods (i.e., descriptive morphology, TLC and HPLC chromatography) of herbal extracts used in nutraceuticals, cosmetics, and pharmaceuticals. There is no need for additional grinding, cutting, sieving or other preparation of the samples. This greatly reduces the time and labor required to properly classify incoming raw materials used in these industries. Additionally, the sample cup spinner accessory with the closed sample cups allows for the reproducible collection of spectra from heterogeneous samples. The discriminant analysis method developed here for analysis of the 14 herbs is shown to be robust and reliable while providing answers within seconds.

Learn more at **thermofisher.com/nirfood**