New Technology for Finding Optimal Spectral Matches in Reference Databases

Ty Abshear[1] and Karl Nedwed[2]
[1] Bio-Rad Laboratories, Inc., Informatics Division, 2754 Compass Dr. Suite 360, Grand Junction Colorado, 81506, USA
[2] Bio-Rad Laboratories, Inc., Informatics Division, Radegunderstr. 108c, 8045 Graz, Austria

Spectroscopy 210404

Abstract

Optimized Corrections is a new technology to optimize spectral search results. It iteratively optimizes the query and/or reference for each comparison by automatically applying multiple corrections to compensate for differences between spectra. This technology solves many of the issues surrounding spectral search that cannot be adequately addressed by traditional algorithms and manual methods.

Introduction

Spectral searching is one of the most important tools researchers use to classify or identify materials, yet it continues to be plagued by errors and imperfections. During a spectral search, a sample spectrum is compared to a database of reference spectra. To ensure that an optimal match is found in the database, spectra can be adjusted to compensate for differences between spectra caused by variability in instruments, accessories, environmental conditions, and other factors.

According to ASTM's guide on spectral searching[1], various algorithms and manual methods exist to adjust spectra to get reasonable match scores when two compared spectra of the same compound differ for various reasons. While these methods may work in select cases, subtle discrepancies such as a shift of the X-axis are very hard to identify and correct manually. The inflexible mathematical algorithms typically employed do not compensate for these types of errors in spectra that are flawed. Manual corrections can be made by expert spectroscopists, but those less experienced in spectroscopy are often unaware of how to perform the necessary corrections on their sample spectrum to achieve the best search result.

To address this growing concern, Bio-Rad has introduced Optimized Corrections, a breakthrough patent-pending technology that performs a computationally complex set of multiple corrections on all query and reference spectra in a search to find the optimal match between the query and each individual reference spectrum. A series of examples will demonstrate how the Optimized Corrections technology yields better matches between query and reference spectra than can be attained using rigid search algorithms alone or with manual methods to optimize spectra for searching.

Methods

The Optimized Corrections technology in Bio-Rad's KnowItAll® spectral search software application allows a query spectrum (defined as a series of X/Y data points) to be compared to each reference spectrum in a database, in an iterative process that applies one or more mathematical corrections to the X/Y data points of one or both spectra. One or more correction parameters is adjusted in each iteration to find the best match between the spectra as defined by a scoring algorithm that assigns a relative numerical value to the match of the two spectra in each iteration. Once all of the correction parameters that give the best match have been determined, the Optimized Corrections are applied to the two spectra being compared, which are then displayed visually in the software in an overlapped, stacked, or offset mode.

The following types of corrections are applied automatically by the algorithm: baseline correction, vertical clipping, horizontal shift, vertical shift, and intensity distortion. If an ATR correction needs to be applied to compare an ATR IR spectrum to a

transmission IR spectrum, Optimized Corrections for ATR correction attenuation or non-polarization effect compensation are also applied. Finally, for Raman spectra, a Raman-specific intensity distortion can be applied in an attempt to compensate for intensity differences in Raman spectra measured with different excitation laser wavelengths.

The Optimized Corrections technology works with all industry-standard scoring methods, such as Euclidean Distance, Correlation, or First Derivative Euclidean Distance.

Results

Optimized Curve Matching - Example One

The following example demonstrates the Optimized Correction technology being applied to find the optimal clipping value to compensate for truncated peaks as encountered often when the concentration of the sample is too high during an FTIR run. A search of 2-Furaldehyde compared with the reference spectrum match of the same compound using the Correlation Algorithm and industry standard normalization, yields a match score of 87.6% (Figure 1). Using only the Optimized Clipping Correction, an improved match score of 96.2% is obtained. Using all of the Optimized Corrections, a match score of 97.6% is obtained (Figure 2).

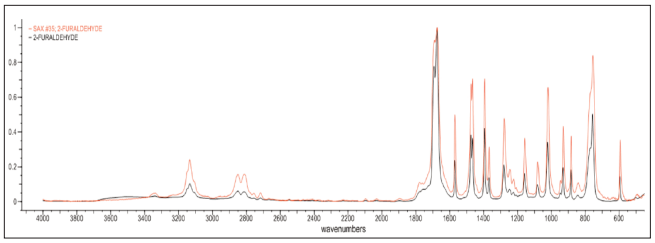

Fig. 1. Results of a search of 2-Furaldehyde (black) compared with the reference spectrum match (red) of the same compound. The Optimized Corrections technology is not enabled during the search. Using the Correlation Algorithm and industry standard normalization, a match score of 87.6% is obtained.

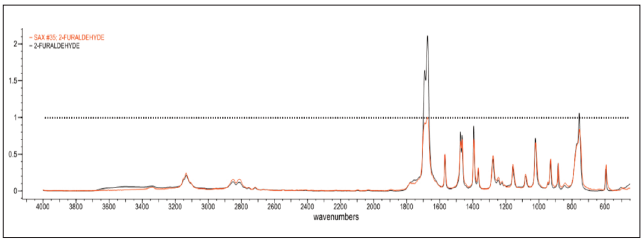

Fig. 2. Results of a search of 2-Furaldehyde (black) compared with the reference spectrum match (red) of the same compound. This time the Optimized Corrections technology is enabled during the search. Using all of the Optimized Corrections, a match score of 97.6% is obtained. The dotted line shows where the vertical spectral clipping was performed.

Optimized Corrections - Example Two

The following example shows that a search of Polypropylene, compared with the reference spectrum match of the very same compound using the Correlation Algorithm and industry standard normalization, yields a match score of only 28.9% (Figure 3). Using all of the Optimized Corrections in the KnowItAll software, multiple corrections are applied, and a match score of 97.6% is obtained (Figure 4). The combination of Optimized Corrections that gave the best match are correcting the baseline of the query spectrum, performing a Raman-specific intensity distortion adjustment factor of 84.6% to all regions above 2471.8 cm^{-1}, clipping the top 17.2% of the query spectrum, and offsetting the query spectrum horizontally

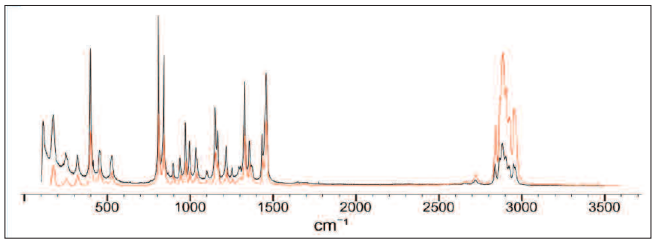

Fig. 3. Results of a search of Polypropylene (red) compared with a reference spectrum match (black) of the same compound. The Optimized Corrections technology is not enabled during the search. Using the Correlation Algorithm and industry standard normalization, a match score of only 28.9% is obtained.

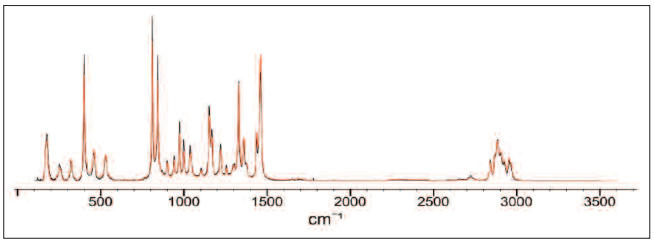

Fig. 4. Results of a search of Polypropylene (red) compared with the reference spectrum match (black) of the same compound. This time the Optimized Corrections technology is enabled during the search. A match score of 97.6% is obtained.

Comparison of Sadtler Standard Spectral Library to the *NIST Chemistry WebBook*

To assess the prevalence of horizontal shifting among infrared spectra, 1,365 IR spectra from the Bio-Rad Sadtler Standards (Comprehensive) database were compared to corresponding 1,365 IR replicate spectra from the *NIST Chemistry WebBook*[2] using only the Horizontal Shift Optimized Correction to improve the match quality of each comparison. The average shift of the Bio-Rad Sadtler spectra relative to the *NIST Chemistry WebBook* spectra was found to be -2.47 cm^{-1}, and 50.8% of the comparisons have X-axis shifts of more than + or -4 cm^{-1}. The overall frequency of shifting was significant and skewed slightly in the negative direction (Figure 5).

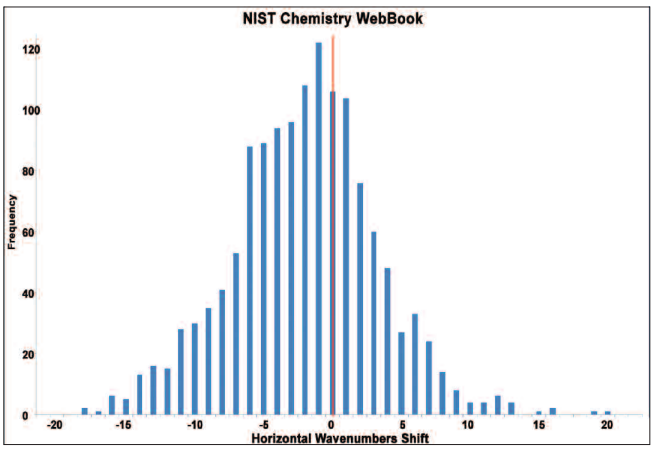

Fig. 5. Horizontal Shift Optimized Correction of 1,365 IR spectra from the Bio-Rad Standards (Comprehensive) database compared to the corresponding IR spectra from the NIST Chemistry WebBook.

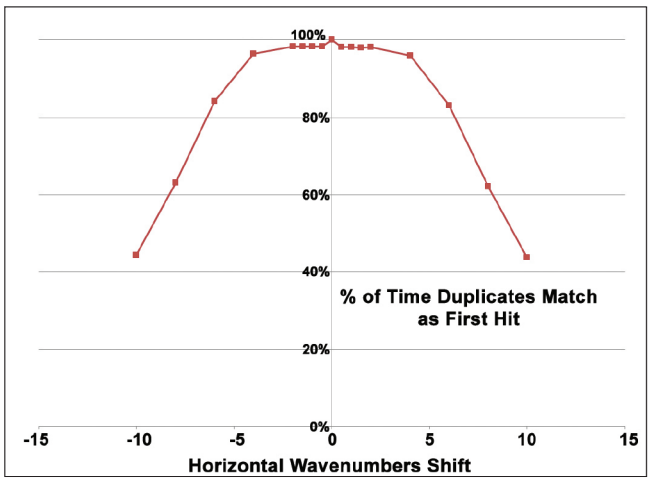

Fig. 6. Impact of horizontal shift on hit list ranking for a search of 1,000 query IR spectra against a database of 75,549 IR spectra that includes the 1,000 query IR spectra as duplicates.

Impact of Horizontal Shift on Hit List Order

The hit list order for 1,000 IR spectra from the Bio-Rad Sadtler Standards (Comprehensive) database used as queries and searched against the entire Bio-Rad Sadtler Standards (Comprehensive) database of 75,549 spectra was calculated as a function of horizontal X-axis shift from -10 to 10 cm^{-1} (Figure 6). When searching for a spectrum against itself with no horizontal shift, a perfect match is obtained, and the correct match appears as the first hit in the hit list. Horizontally shifting the query spectrum up to + or -2 cm^{-1} causes the correct hit to drop below the first hit only occasionally. Horizontally shifting the query spectrum above + or -4 cm-^{1}, however, has a significant impact on the correct hit appearing as the first hit in the hit list: with a shift of + or -10 cm^{-1}, the correct hit is no longer the first hit over 55% of the time.

Conclusion

Optimized Corrections provides a new methodology for spectral searching and visualization. It iteratively optimizes the query and/or reference for each comparison by automatically applying multiple corrections to compensate for differences between instruments, environmental conditions, sample concentration, etc. Through the examples shown, one can see how this new technology can produce better matches than traditional algorithms alone and manual methods that are currently used to optimize spectra for searching.

Reference

[1] E2310-04 - Standard Guide for Use of Spectral Searching by Curve Matching Algorithms with Data Recorded Using Mid-Infrared Spectroscopy, 2009. ASTM International Web Site. http://www.astm.org/Standards/E2310.htm (accessed March 4, 2015).

[2] NIST Chemistry WebBook. http://http://webbook.nist.gov/ (accessed March 4, 2015).