VIAVI Solutions

Application Note

# Method Transfer across Multiple MicroNIR Spectrometers for Raw Material Identification

Raw material identification or verification (of the packaging label) is a common quality-control practice. In the pharmaceutical industry, the increasing global footprint of the supply chain and public health concerns[1] resulting from contaminated materials (or mislabeling) have driven many regulatory bodies to require inspection of every barrel in every shipment of materials used in the manufacture of pharmaceutical drugs.

Handheld analytical instruments are becoming more common for rapid and non-destructive testing of incoming raw materials directly in warehouses, reducing the cost and time needed for taking samples to the laboratory and quarantining incoming shipments until lab testing is complete.

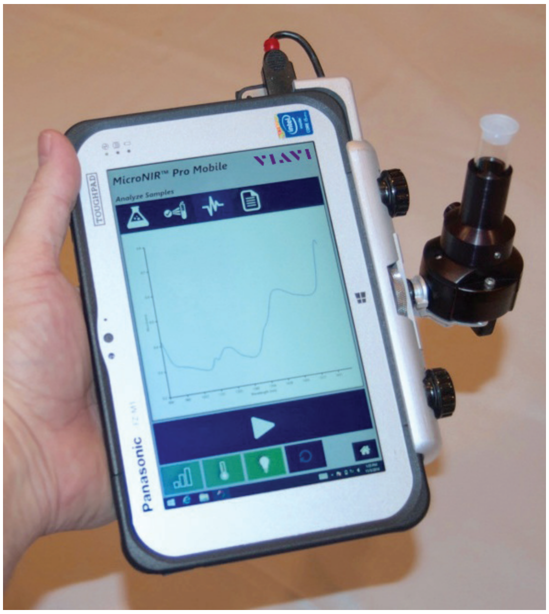


Figure 1. MicroNIR spectrometer tethered to a rugged 7" Windows 8.1 tablet; the MicroNIR is equipped with a vial holder in this image

## The MicroNIR spectrometer

The MicroNIR™ spectrometer is a near-infrared (NIR) handheld spectrometer tethered to a smart tablet and equipped with predictive modeling software[2] that identifies pharmaceutical raw materials. Advantages of the MicroNIR spectrometer relative to Raman and Fourier-transform infrared (FTIR) spectroscopy include:

- NIR spectroscopy does not suffer from the fluorescence problem encountered in Raman spectroscopy
- Measurement time with MicroNIR is very fast (<1 second)
- MicroNIR measures through plastic or glass containers
- Light sources are eye safe

This application note reports on a study conducted to demonstrate how the MicroNIR™ 1700 spectrometer[2-3] correctly classifies pharmaceutical raw materials and enables transferring the method from one or more master instruments to multiple target instruments. The study compared different classification algorithms to assess the number of master instruments needed to achieve 100% accurate prediction.

In addition to adopting the preferred classification algorithm, successfully transferring methods to target instruments depends highly on instruments that are consistent in performance spectrally, optically, and physically[3]. MicroNIR spectrometer technology, when coupled with common chemometric analysis techniques, is highly suitable for classification applications in high-volume applications.

## Methodology and Experimental Procedure

The study was conducted in two phases. Phase I consisted of the scanning of 19 of the most commonly used pharmaceutical excipients and active ingredients to build a classification library. Phase II aimed at providing a robust model that could be transferred to a new, "untrained" spectrometer. Spectra of all 19 compounds were collected from six instruments with cross-instrument validations using soft independent modeling of class analogies (SIMCA), partial least squares discriminant analysis (PLS-DA), and support vector machine (SVM) classification algorithms. Additionally, data pretreatment algorithms to find out the minimal number of spectrometers needed and the best combinations (and sequences) of pretreatment schemes for achieving optimal prediction performance were also studied.

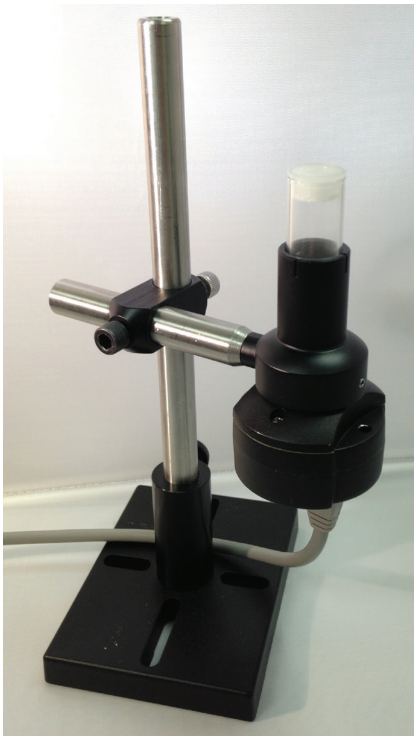


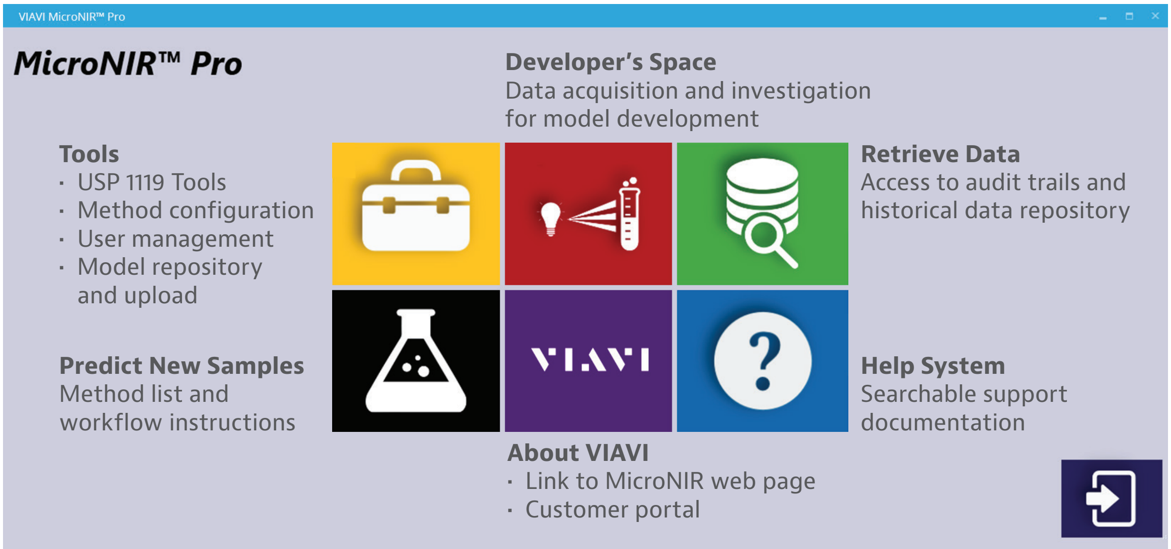
Figure 2. Experimental setup

The spectra of 19 pharmaceutical materials listed in Table 1 were collected in laboratory ambient conditions. All samples were presented to the spectrometer housed in 4 mm borosilicate glass vials with measurements performed through the bottom of the vials. A vial holder that slips over the end of the spectrometer was used by the operator to easily introduce and remove samples and also maintain an optimal 3 mm distance between the material and the face of the spectrometer. Figure 2 shows a picture of the vial holder in use. The vial was rotated approximately 10 to 15 degrees after every scan. Each scan had an integration time of 10 ms with spectrum averaged over 50 collections. Each material was scanned a minimum of five times.

Table 1. 19 pharmaceutical compounds used in the study

| Material ID | CAS Number | Short Names |
|---|---|---|
| Acetaminophen | 103-90-2 | Acetaminophen |
| Ascorbic acid, L-ascorbic acid | 50-81-7 | Ascorbic acid |
| Aspirin | 50-78-2 | Aspirin |
| Benzocaine (ethyl-4-aminobenzoate) | 94-09-7 | Benzocaine |
| Caffeine | 58-08-2 | Caffeine |
| Cellulose | 9004-34-6 | Cellulose |
| Corn Starch | 9005-25-8 | Corn starch |
| Fructose, D-(–)-fructose | 57-48-7 | Fructose |
| HPC, hydroxypropyl cellulose | 9004-64-2 | HPC |
| (Hydroxypropyl) methyl cellulose (HPMC) | 9004-65-3 | HPMC |
| Ibuprofen | 15687-27-1 | Ibuprofen |
| Lactose | 63-42-3 | Lactose |
| Magnesium stearate | 557-04-0 | Mg-stearate |
| Poly (ethylene oxide) (PEO) powder | 25322-68-3 | PEO |
| PVP, polyvidone, polyvinylpyrrolidone, povidone | 9003-39-8 | PVP |
| Polysorbate 80 | 9005-65-6 | Polysorbate80 |
| SSG, sodium starch glycolate, Explotab® | 9063-38-1 | SSG |
| Talc | 14807-96-6 | Talc |
| Titanium dioxide, titanium(IV) oxide | 13463-67-7 | TiO2 |

All measurements were collected using MicroNIR Pro spectrometer software version 2.0 (Figure 3). A reference measurement was performed on the MicroNIR approximately 15 minutes after the lamps were turned on and every hour thereafter while performing scans. A 99% diffuse reflectance panel was used for the 100% reference value and the 0% reference value was taken by leaving the tungsten lamps on with an empty vial holder. This scenario was used to account for any scattered light from the sample vial holder.



**Developer's Space**
Data acquisition and investigation
for model development

**Tools**
- USP 1119 Tools
- Method configuration
- User management
- Model repository and upload

**Retrieve Data**
Access to audit trails and
historical data repository

**Predict New Samples**
Method list and
workflow instructions

**Help System**
Searchable support
documentation

**About VIAVI**
- Link to MicroNIR web page
- Customer portal

Figure 3. The MicroNIR Pro software home page

Most of the 19 pharmaceutical compounds were ordered from 2-3 different vendors to increase the coverage of samples' manufacture sources variations. Each compound from each manufacture source was scanned at three different dates and ambient temperatures. To study the models' transfer performance among different spectrometers, six spectrometers collected data.

After spectra collection, data was imported into an embedded version of Unscrambler® X software version 10.3 in the MicroNIR Pro software for spectral analysis and calibration model development. All spectra were pretreated using a Savitzky-Golay first derivative followed by standard normal variate (SNV), also performed within the MicroNIR Pro software. Figure 4 is a plot of the first derivate of the spectra of the 19 materials tested.

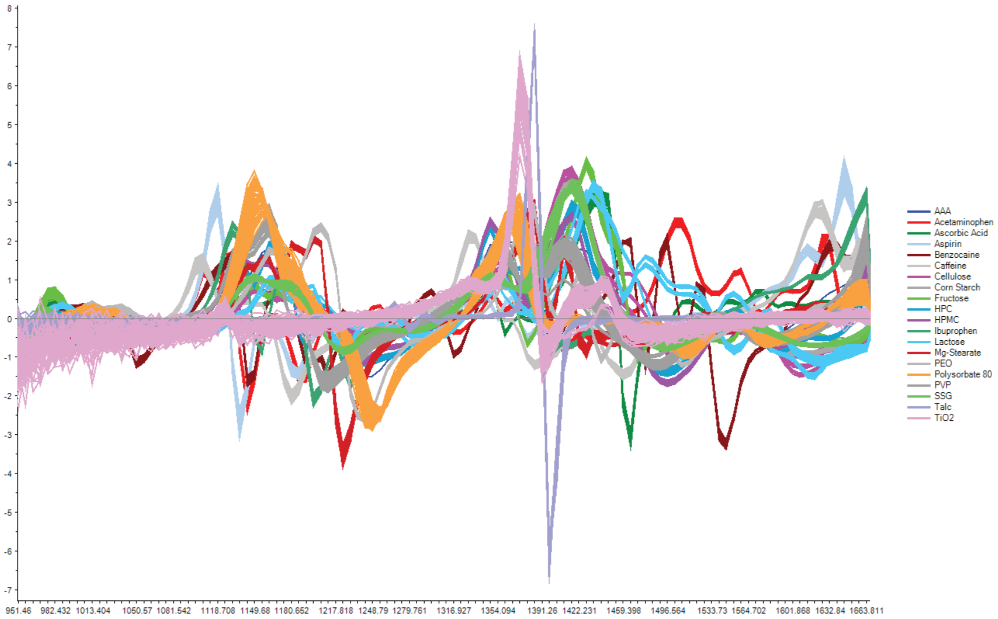


Figure 4. Savitzky-Golay first derivative and SNV pre-treated spectral data across 950 – 1650 nm

## Chemometric Library Development

Chemometric libraries are predictive and data-driven statistical models that use chemometric and machine-learning principles to extract key features and underlying relationships between spectra and their respective known identities (training sets). The libraries are used later to predict the same identities of unknown samples (test sets). In this study, we investigated the building of such models for pharmaceutical raw materials identification applications with spectra data from the MicroNIR 1700 instruments.

We used three types of classification algorithms characterized by specific principles:

**SIMCA**[4] — a principal component model represents each class in the data set

**PLS-DA**[5] — X- and Y-scores are chosen so that it will seek directions in the factor space that are associated with high variations in responses but biasing them toward directions that are accurately predicted

**SVM**[6] — Class boundaries based on maximizing the separation margin between classes: samples at the borders of each class play a major role

Table 2. Pros and cons for each classifier

| Classifier | Pros | Cons |
|---|---|---|
| SIMCA | Easy to understand, implementation is simple | Difficult to optimize and maintain parameters for each class |
| | | Cannot generalize well for heterogeneous structured datasets |
| PLS-DA | Can calculate each variable's contribution | Can run into overfit issues for complicated datasets |
| SVM | Robust and good generalization capabilities | Implementation is more complicated but commercial software packages such as CAMO's The Unscrambler offers this capability |
| | | Cannot calculate each variable's contribution |

## Results and Discussion

Figure 5 is the principle component analysis (PCA) score plot showing PC-1 plotted against PC-2. The PCA scores plot provides an initial indication of spectral differentiation in the calibration set. Chemically-similar materials tend to cluster closer to each other versus chemically different materials.

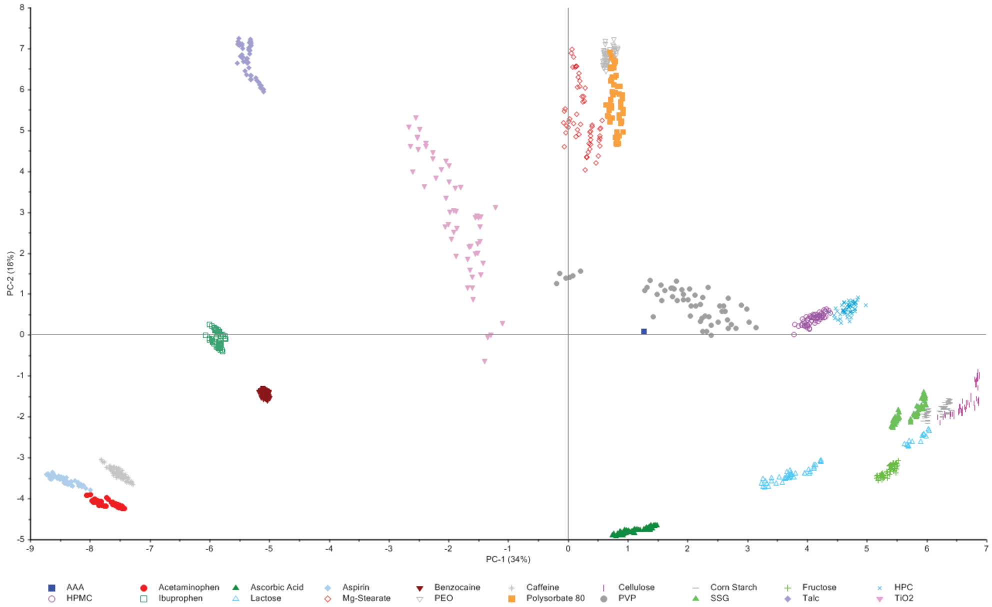
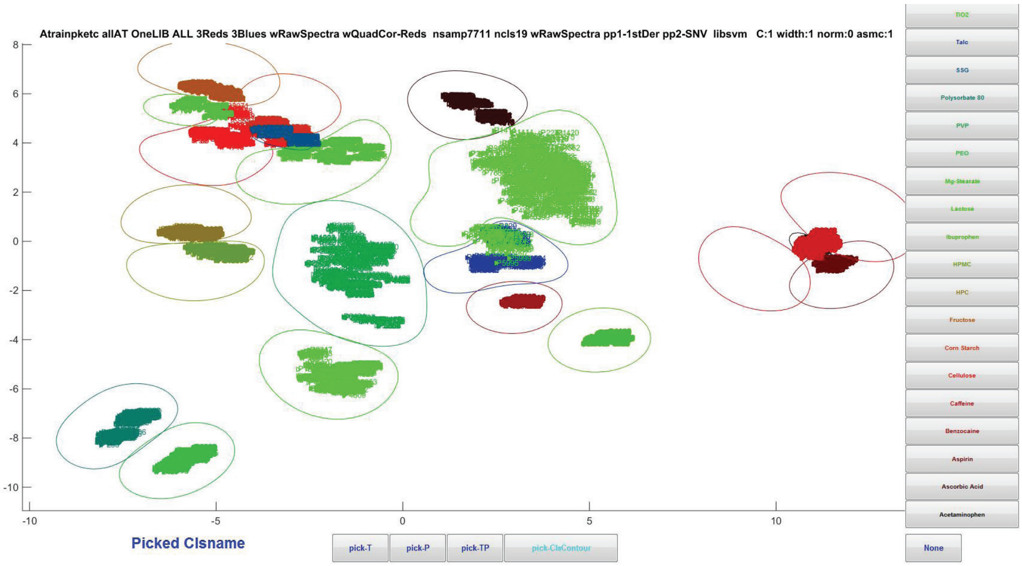


Figure 5. PCA scores plot, PC-1 vs. PC-2

Figure 6 shows the partitions of classes in SVM models by the PCA-SVM plot, where the x-axis and y-axis represent PC-1 and PC-2 scores and the class decision boundaries were generated by SVM calculations. This gives an overall picture of the relative positions of each class and their class boundaries.



Atrainpketc allAT OneLIB ALL 3Reds 3Blues wRawSpectra wQuadCor-Reds nsamp7711 ncls19 wRawSpectra pp1-1stDer pp2-SNV libsvm C:1 width:1 norm:0 asmc:1

Figure 6. PCA-SVM plot for 19 pharmaceutical compounds based on SVM models

## Model Building Times

Different classification algorithms take different amounts of time to build training models. Model building time in offline learning cases may not be crucial but if it takes too long to build, it will be difficult to study and optimize modeling processes. One key factor in deciding model building time is the size of the training set (number of samples in the training set). Figure 7 compares a model's building time vs. the size of the training set for the three types of classification algorithm we studied here.

Both PLS-DA and SIMCA require optimization of their model parameters (number of PLS factors for PLS-DA and number of principal components for each class in SIMCA) through a training set cross-validation process while for SVM there is no need for cross validation. As shown in the figure, it will take about 10,000 times longer to train PLS-DA vs. SVM and roughly 100 times longer to train SIMCA vs. SVM. It was also estimated that PLS-DA model building times increased with a power of about 1.9 vs. sample size and SIMCA increased with a power of about 1.5. With SVM, model building times increased almost linearly with training set sample sizes.

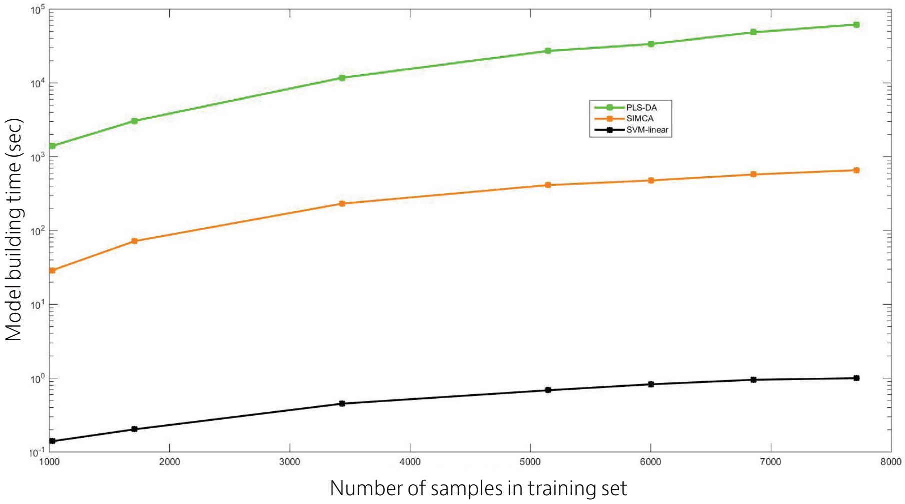


Figure 7. Model building times for three kinds of classification algorithms vs. sizes of training sets

In this study, and for reference, with two instruments used for the training set, the model building time was about 1 hour, 100 sec, and 1 sec for PLS-DA, SIMCA, and SVM respectively.

## Building Robust Models

With point-of-use NIR applications, it is possible that hundreds of spectrometers will be deployed in the field. It is almost impossible to train each spectrometer with the specific dataset of materials it will analyze. The practical approach is to collect a few complete sets of spectra covering all variances from chemicals, environmental factors (temperatures, humidity), and user variations (path length, for example) from a couple of master spectrometers. These models are then transferred to all other target or production spectrometers.

In this study, we collected six libraries of spectral data from six different MicroNIR spectrometers and determined how many spectrometers would be needed for building a robust model that will be successful when applied to spectrometers that were not part of the training set. When constructing these kinds of consolidated libraries (based on spectra collected from more than one spectrometer), a sample selection scheme can drastically reduce the size of these training sets. This means less storage space and quicker model building times.

The study used the Kennard-Stone[7] sample selection scheme to select the most representative samples. Table 3 shows results where nTUx (x=1 to 6) denotes how many spectrometers were used to build training models and the remaining spectrometers' data was used as test set. For example, nTU2 shows average prediction results when any two spectrometers were used to build models and the remaining four spectrometers' data served as a test set. In the case of nTU6, since there are only six spectrometers, the training set consisted of samples chosen by Kennard-Stone algorithm and the remaining samples served as a test set.

Also shown in the table is the number of spectrometers needed to achieve 100% classification prediction. Based on this analysis, for SVM it would only need one spectrometer's data to achieve perfect prediction while for PLS-DA and SIMCA, it would take all six spectrometers' data to build models to reach 100%. However, even with PLS-DA and SIMCA, accurate prediction levels of >98.8% are achieved with only two instruments for training.

Table 3. Number of spectrometers needed to build robust models

| Classification | nTU1 | nTU2 | nTU3 | nTU4 | nTU5 | nTU6 | # Units for 100% |
|---|---|---|---|---|---|---|---|
| SIMCA | 97.395 | 98.851 | 99.496 | 99.808 | 99.961 | 100 | 6 |
| PLS-DA | 99.664 | 99.68 | 99.904 | 99.961 | 99.99 | 100 | 6 |
| SVM | 100 | 100 | 100 | 100 | 100 | 100 | 1 |

SVM is the preferred model for classifying raw materials. The model has been tested on over 20 different MicroNIR spectrometers manufactured over different lots. 100% successful prediction has been achieved every time.

## Conclusion

In this study, we showed that the MicroNIR spectrometer is effective for use in pharmaceutical raw materials identification. Three machine learning pattern classification model-building algorithms were tested: SIMCA, PLS-DA, and SVM. All delivered near-perfect identification success rates, with increasingly better performance from SIMCA to PLS-DA to SVM and correspondingly lower model building times. Model building and transfer with MicroNIR Pro software requires minimal time and performs with high reliability. Furthermore, model transfer from a master instrument to a number of other instruments is easily achievable due to the high reproducibility of the MicroNIR instruments in production.

## References

[1] http://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2012/05/16/heparin-a-wakeup-call-on-risks-to-the-us-drug-supply

[2] O'Brien, N., Hulse, C., Friedrich, D., Van Milligen, F., von Gunten, M., Pfeifer, F., Siesler, H., "Miniature Near-Infrared (NIR) Spectrometer Engine For Handheld Applications." Proc. SPIE, Ed. M. Druy, and R. Crocombe, 8374, p 837404-1-8 (2012).

[3] Friedrich, D., Hulse, C., von Gunten, M., Williamson, E., Pederson, C., O'Brien, N., "Miniature near-infrared spectrometer for point-of-use chemical analysis." Proc. SPIE, Ed. Y. Soskind and G. Olson, 8992 (2014).

[4] Wold, Svante, and Sjostrom, Michael, 1977, SIMCA: A method for analyzing chemical data in terms of similarity and analogy, in Kowalski, B.R., ed., Chemometrics Theory and Application, American Chemical Society Symposium Series 52, Wash., D.C., American Chemical Society, p. 243-282.

[5] Wold, S; Sjöström, M.; Eriksson, L. (2001). "PLS-regression: a basic tool of chemometrics". *Chemometrics and Intelligent Laboratory Systems* **58** (2): 109–130.

[6] Cortes, C.; Vapnik, V. (1995). "Support-vector networks". Machine Learning **20** (3): 273.

[7] R. W. Kennard and L. A. Stone (1969): Computer Aided Design of Experiments, Technometrics, 11:1, 137-148.

VIAVI Solutions

Contact Us **+1 844 GO VIAVI**
(+1 844 468 4284)

To reach the VIAVI office nearest you,
visit viavisolutions.com/contact