Detection and Classification with Overlapping Signals

Neal B. Gallagher

Key words: multivariate analysis, classical least squares

As modern sensor systems are taken from the lab to less pristine field operations, more is being asked of them. Under these conditions, it is difficult to find sensors that are perfectly selective for an analyte of interest in the face of background signals they are expected to encounter. To be perfectly selective, a sensor must respond only to the analyte of interest and nothing else. However, this situation is often difficult or impossible to achieve. Instead it is more common to find sensors responding to multiple analytes as well as interferences, and it is the task of detection and classification algorithms to sort out this "cross talk" from different signal sources.

A simple example can be given by imagining a two-channel sensor. In the first case, the response on the two channels to Analyte A at unit concentration is [1 0] and to Analyte B is [0 1]. This states that Analyte A only has a response on Channel 1 and Analyte B only has a response on Channel 2 and this system is perfectly selective for both A and B. However, in the second case the response to A is [1 0] and to B it is [0.2 0.8]. This shows that the response to B now "bleeds" into the response on Channel 1. If this cross talk were not accounted for, high concentrations of B could easily result in a false alarm for A as well as an alarm for B. In spectroscopic measurements, this type of cross talk manifests as overlapping peaks and occurs even for measurements in well-controlled laboratory environments. There are several methods to account for cross talk and one highly successful method utilizes the linear mixture model.[1]

In spectroscopic applications, the linear mixture model was proposed because it is directly relatable to Beer's law for transmission measurements of multi-component mixtures. This model is also known as the classical least squares (CLS) model and it can be used with single or multiple target analytes. With advanced forms,[2] the CLS approach can be used to account for interferences as well. These models work by finding a response unique to each analyte in the system. For

example, in the second case discussed above the unique signal for A at unit concentration is N_A = [1 -0.25] and for B it is N_B = [0 1.25]. Now take the case where only pure B is measured in the system at unit concentration (i.e., the measured response is [0.2 0.8]). In the CLS model, the measured response is first multiplied by the unique A signal N_A yielding $[0.2\ 0.8]*[1\ -0.25]^T = 0.2*1 + 0.8*(-1.25) = 0.^†$ The second step multiplies the measured response times the unique B signal N_B yielding $[0.2\ 0.8]*[0\ 1.25]^T = 1*0 + 0.8*1.25 = 1$. Therefore, if pure B is present the CLS model will not indicate a response on A but will indicate a response on B. In this example, the CLS model utilized both channels of the instrument to obtain a unique response for both analytes; it used the multivariate signal to improve selectivity of the system over the univariate response. As a result, the CLS model did not false alarm for analyte A due to the presence of analyte B as it would have if only Channel 1 was considered.

The simple example considered above neglected measurement noise. However, including additional channels in the measurement system, which typically results in more redundancy in the signal, can also result in higher signal-to-noise.[1] Other advantages of adding more channels to the instrument include adding the ability to detect more analytes and to detect unusual signal.[1-3] This latter advantage is often overlooked but should be considered a distinct advantage over univariate systems (and many types of low number channel instruments) and is one of the principles behind anomaly detection and process monitoring.[4] This type of signal can be used to monitor the health of the analytical instrument as well as identify the presence of new uncharacterized interferences (i.e., tell the user when something other than A and B are present). This is important because uncharacterized signal can bias a model's output and easily result in false alarms if it is ignored. Because inverse least squares (ILS) methods are commonly encountered, it is good to note that ILS methods such as partial least squares (PLS) and principal components regression (PCR)

EIGENVECTOR
RESEARCH INCORPORATED
Research, Training, and Software

196 Hyacinth Road
Manson, WA 98831
www.Eigenvector.com

also find the signal unique to each analyte. The mathematical formulation for the ILS methods is a bit different but the concepts are similar to CLS models.

Because it is difficult or impossible to develop sensors perfectly selective for analytes of interest, modern sensor systems rely on multiple sensors (channels, frequencies or wavelengths) and multivariate analysis tools like classical least squares to extract relevant information. In fact, sensors are no longer viewed as just the measurement device alone. Instead, they are viewed as sensor *systems* that consider the measurement device, the signal due to analytes of interest and expected interferences (the sensing scenario), and the algorithms used to extract relevant signal. Coupled with good design of experiments and multivariate analysis tools, these systems are realizing the multivariate advantages of higher signal-to-noise, higher selectivity and enhanced measurement diagnostics.

[†] This multiplication is known as an inner product in linear algebra where the corresponding elements are multiplied and the result is summed.

1. Martens, H. and Næs, T., "Multivariate Calibration", John Wiley & Sons, New York, NY (1989).

2. Gallagher, N. B., Detection, Classification and Quantification in Hyperspectral Images using Classical Least Squares Models. In *Techniques and Applications of Hyperspectral Image Analysis*, Grahn, H. F.; Geladi, P., Eds. John Wiley & Sons: West Sussex, England, 2007; pp 181-201.

3. Jackson, J.E., "A User's Guide to Principal Components", John Wiley & Sons, New York, NY (1991).

4. Wise, B.M. and Gallagher, N.B., "The Process Chemometrics Approach to Chemical Process Monitoring and Fault Detection," *J. Proc. Cont.* **6**(6), 329–348 (1996).

5. "Chemical and Biological Sensor Standards Study," Defence Advanced Research Projects Agency, LTC John Caranno, Ph.D., Study Chair, Microsystems Technology Office.