De novo Models for Raman: Comparison to and Advantages Over Traditional Chemometrics Models

Written in summary, of the white paper "De novo Approaches for Bioprocess Parameter Estimation" please view the full version here for the detailed explanations, figures, and references.

Raman in Bioprocess

Raman spectroscopy has enjoyed substantial growth in pharmaceutical applications over the last 20 years initially as high-efficiency material identity testing and more recently in bioprocess applications. Raman is a light scattering technique that involves molecules scattering light from a high-intensity laser light source. Raman inherently provides high analyte specificity and flexibility in measurement geometries as the measurements can be made through plastic windows, walls, glasses, flow cells, or with direct immersion. However, without a MAVERICK device, the interpretation of the Raman raw data requires extensive computational assistance using traditional chemometric analysis. Modeling of critical process parameters (CPPs) and product critical quality attributes (CQAs) using Raman in bioprocess is generally done using the partial least squares (PLS) approach, which has several challenges - as listed below. An advantage of the PLS approach is that it requires little to no detailed understanding of the bioprocess and instrument physics.

7 reasons why Raman spectroscopy with PLS models is challenging for bioprocess applications:

1. The PLS approach implies that every bioprocess must run the same, with consistent correlations between different chemical species. It also implies that the instrument measurement variance is always the same, which in bioprocess is never an accurate assumption: The variation in biomass and, thus, fluorescence alone would invalidate this assumption.

2. Co-variance: chemical species often correlate, which is widely utilized in Raman bioprocess PLS models. However, this is dangerous, as the signals are thus unspecific and do not actually measure the molecule of interest but rather measure the unknown analyte(s) which correlate with it.

3. Exchangeability and cross-validation – difficult for the reasons mentioned in 1 and 2.

4. Modeling pre-processing, normalization, correction, etc. are handled with the attitude "*Try and see what seems to work*" as there is little theoretical justification to choose one approach over another.

5. Empirical modeling approaches don't provide direct estimates of the central figures of merit and, therefore cannot provide estimates of selectivity, linearity, precision, limit of detection, and sensitivity, as other analytical tools do.

6. When spectrometers are exchanged or sources/detectors replaced, the multivariate model typically needs to be corrected for relevance to the new spectrometer properties. This "calibration transfer" can be very time-consuming and costly.

7. The black-box (only data-driven) nature of empirical calibration methods requires extensive empirical validation efforts to demonstrate sensitivity, selectivity, linearity, and robustness.

In the literature, several groups have reported success building 'generic' models using PLS with various mathematical pre-treatment methods and have reported reasonable success for defined platform processes. However, often upwards of 25-30 process runs were involved in these efforts at very considerable expense—several years of process time—and that excludes deployment and maintenance activities that follow.

Given these challenges, it is little wonder that robust Raman method development and deployment has been a particularly vexing challenge in bioprocess applications.

De novo Model – MAVERICK

Conventional Raman modeling is challenging due to points 1-7 noted above. Conversely, MAVERICK *de novo* model does not use any empirical data in its first-principles predictions. The MAVERICK approach thus eliminates the issues of deciding what experiments should be done, on what hardware, across which conditions, how the raw data should be processed/manipulated prior to calculations, and how the resultant model performs in the next bioprocess run. The MAVERICK *de novo* model uses the key features of:

- Specific Raman reference information for glucose, lactate, and a formulary of many other chemical components that may be present in cell culture media. Glucose and lactate are effectively just two items in the formulary that we use our two-point calibration to convert from "relative weight" to g/L.
- The dynamic characteristics of each unique MAVERICK system from an environmental, optical, electronic, and measurement error model perspective

The *de novo* model is **dynamic**, and the measurements are adjusted in real-time – actively during the bioprocess run. One might wonder how it is possible to cover all possibilities in the formulary, but there are a few helpful boundaries. While it is very likely that the number of chemical/biochemical species in an active bioprocess may number in the thousands, the limited sensitivity of Raman spectroscopy implies that one really needs to consider only the major components above approximately 0.01 g/L. At these limits in mammalian cultures, we have found that a few hundred chemical species are relevant, along with non-biologic additives (e.g. surfactants, anti-foaming agents).

When you eliminate non-Raman active components (like salts) the list is even further narrowed.

Many of the largest sources of error in Raman systems are fundamental to the system's optical design and electronics, and MAVERICK's internal system model allows it to estimate the measurement error covariance in real time. Related, the system model also allows for adaptation to, for example, changing lighting, temperature, and turbidity conditions, including fluorescence.

It is quite rare to find publications demonstrating Raman application in bioprocesses citing standard analytical specifications such as sensitivity, selectivity, LOD for the resulting models. This is because these specifications are complicated for multivariate models. In contrast, sensitivity and selectivity factors consistent with the IUPAC definitions can be directly estimated from the *de novo* model for MAVERICK.

Using MAVERICK for Bioprocess

MAVERICK is designed to be plug-and-play across measurement modules, probe adapters, and probes. To enable this, a quick setup is required to confirm quantitative system suitability before bioprocess analysis can begin. This is a 3-step process, guided on screen by the

MAVERICK's user interface:

- Immerse probe into "LOW" calibration standard, press go (wait approximately 4 minutes)
- Immerse probe into "HIGH" calibration standard, press go (wait approximately 4 minutes)
- Autoclave the probe for immersion in the actual bioprocess

Steps 1 and 2 check that several properties of the MAVERICK and probe are conforming with the *de novo* model, and a minor correction is made to the *de novo* model outputs for the particular combination of MAVERICK measurement model, probe adapter and probe. This information also allows for audited performance qualification and tracking with the serialized/microchipped probe. MAVERICK also supports a single point 'live' reference which can be helpful in eliminating any small observed biases that may be consistent between a particular offline reference analyzer and MAVERICK's *de novo* outputs.