RETHINKING CALIBRATION FOR PROCESS SPECTROMETERS

Will Warkentin
Quality Control Manager
Chevron Products Co.
Richmond Refinery
841 Chevron Way
Richmond, CA 94801

Brian Rohrback
President
Infometrix, Inc.
11807 North Creek Parkway S
Suite B-111
Bothell, WA 98011

KEYWORDS
Best Practices, Calibration, Cloud Computing, Database, Gasoline Blending, Optical Spectroscopy, PLS, Process Control

ABSTRACT
Optical spectroscopy is a great source of process chemistry knowledge. It has the advantage of speed, sensitivity, and simple safety requirements. As one of very few analyzer technologies that can measure chemistry, it has become a workhorse in the hydrocarbon processing industry. What if we could put a spectroscopy system in place and have it handle the application and communicate results as soon as it is turned on? Then, if predictions do not match legacy standards, the system dials itself in or calls for help. And, we are not constrained on either the hardware or the software front. In this paper, we address the primary bottleneck of mustering data, automating analyzer calibration, and tracking data and model performance over time. The approach is generic although the figures and tables are focused on gasoline blending.

INTRODUCTION
Our collective practical experience began in the late 1970s and early 1980s when the benefits of monitoring physical properties like the octane rating of gasoline by near infrared were becoming clear even though the computational tools were in their origin stage and the computers had not yet
made the leap required to manage the process. The computer environment improved in a compound fashion over the next two decades bringing practical use of a spectrometer in routine quality control to the forefront.

THE CURRENT SITUATION

As we fast-forward to present day, the approach we have taken to manage those spectrometers has not changed all that much. Computers have become more responsive, data storage has seen its cost drop to nearly nothing, spectrometers themselves have become more robust, and the gains in connectivity have gone through the roof. Despite these advances, the common process we go through to adapt a spectrometer to a specific application is largely stuck in time.

Several complications relate to this situation:

- As we have proven the value of optical spectrometers to do quality control at the speed of light, the number of analyzers based on this technology is increasing;
- Simultaneously, companies are reducing staff in support of these analyzers; and
- There are fewer students trained in the principles of multivariate analysis, which means that the fitness level of technicians is decreasing.

We recognize that the industry needs to adapt to an increased workload combined with changes in the experience and longevity of the technicians responsible for installation, calibration, and maintenance. In reaction, an effort began six years ago to re-evaluate how the calibration process for optical spectroscopy could be managed in support of the chemical, petrochemical, and petroleum industries.

This paper builds on papers presented in the 2015 and 2018 ISA-AD Symposia [1, 2].

THE CHEVRON CASE

Chevron owns a set of spectrometers that use bundled multivariate software both to process the data into information and to communicate results to an ultimate storage location. Disrupting this set-up is not an attractive option; best is to apply a system that is flexible enough to fit a variety of spectrometer configurations and still deliver value in terms of time-savings and improved calibrations. The preference is also to minimize and standardize the training of personnel so that we can achieve consistent, high-quality results across the entire organization. The focus of this paper is to outline the tasks and show how a cooperative relationship with a vendor delivered an approach that satisfied the need in a cost-effective manner.

To accomplish our goals, the component tasks were organized into six categories as follows:

1. Streamline the mustering of data to prepare for the calibration task;
2. Get as close as possible to full automation of the calibration process;
3. Monitor the quality of the data and models over time;
4. Optimize the calibrations and support any chemometric model format;
5. Handle the report writing tasks and model validations; and
6. Demonstrate the value proposition.

**METHODS**

The component pieces described here are assembled into a service called AI-Metrix, supplied by Infometrix (Infometrix, Inc., Bothell, WA). The individual components of this service are available from multiple vendors. The sources for these components are listed below by function.

**ALGORITHMS**

The algorithms are assembled from a set of MatLab routines (MathWorks, Natick, MA, USA) and Pirouette functions (Infometrix, Inc., Bothell, WA, USA). These were managed in a system driven by the database engine and were focused on three tasks: identification of outliers, calibrating the spectroscopy to fit the assessment, and identifying maintenance needs.

**DATABASE**

The centerpiece of this service is a SQL database (SQL Server, Microsoft, Redmond, WA, USA). This piece is the central organization and coordinator for all processing. The database assembles and matches spectra and reference values and tracks all metadata that allows us to associate the computational results back to the process. On request, it drives the modeling process, generated reports, and formats output for upload into the cloud for dashboard display.

**COMMUNICATION**

Given that there is information collected in separate silos and there are limitations to what types of computation are possible in each, we need to be able to assemble data into optimized bins for processing. For this work, the choice was to use a file sharing protocol, one of the many available from a dozen or more suppliers. The examples given here are shown using DropBox (DropBox, Inc., San Francisco, CA, USA). This allows a set of folders to be shared in a refinery with a remote processing portal.

**DASHBOARD TECHNOLOGY**

The dashboards created for this paper were done using software and utilities specifically designed for business intelligence. Tableau Creator was employed to create dashboard formats and facilitate processing of Excel and SQL output. Tableau Explorer was the interface to the data in the Tableau Online repository. These are products of Tableau Software, Inc., Seattle, WA, USA.
RESULTS AND DISCUSSION

The need for high-quality, consistent modeling practice is emphasized by the graph in Figure 1. A conference was held in 2018 where a challenge data set was provided to 19 chemometricians of varying experience. The results are ranked from worst at the left to lowest error on the right (based on an independent validation set). Roughly one-quarter of the chemometricians failed to model the system adequately and about half were both good and statistically indistinguishable from one another. The automated approach yielded a result that fell into the “correct” category.

![Graph showing chemometricians' performance](image)

**FIGURE 1: CHEMOMETRICS CHALLENGE RESULTS**

We do validate our models before they go live, but no one can be absolutely certain of the quality; there is some concern that models will not generate optimal results when produced by technicians with less experience than the individuals in Figure 1’s example above. A good, objective calibration mechanism is useful as a baseline quality control tool.

Given that the target is to be able to process chemometric models in an automated manner, there is a need to consider all parts involved in the calibration process. Here, we revisit the component tasks indicted in the Introduction.
SET-UP

Data from spectrometers and from the reference sources must be harvested and matched to provide a viable source of data. In addition, storing and being able to recall any or all of these data is critical to the instrument support. This is performed without requiring either a specific brand of instrumentation or a pre-set spectral file format. For pairing of spectra and lab values, there are tools within the SQL database environment that allow developers to match based on the naming convention or the time stamps either on the files or incorporated as part of the header information.

The system can be thought of as a simple set of 4 tasks to cover the functions of the database and calibration system. For the example implemented, the user supplies files into two folders and output appears in two return folders. Because these folders are shared both on the customer side and on the cloud side, any data or results added will be duplicated in near-real-time on the receiving computer. For convenience, a cloud-based dashboard is updated with any change in data under management or evaluation of any models created. The user experience is summarized in Figure 2. The user either submits data (spectra or lab reference values or both) into one folder or submits a request for organized data or model creation in a second folder. These folders are watched by the centralized SQL Server.

FIGURE 2: A STREAMLINED USER EXPERIENCE ENABLES MODEL REQUESTS TO BE FULFILLED AUTOMATICALLY

When the Server has completed the request, any status report generated will show up in the Documents folder and the models and matched datasets are placed in the RequestOutput folder.

Any data placed in the FilesInput folder will be checked for conformity and entered into the database. If needed, an automated, three level matching algorithm insures that the spectral data and references are matched appropriately. In the data assembly, we put the effort into the construction of a custom file read to parse the information from the LIMS output of the raw data file.
With the data matching process done, the SQL Server automatically uploads the status of the database to an on-line viewer, giving response within minutes of the user’s original file deposits. An interactive dashboard allows modification of the displays to identify missing and mislabeled data plus evaluate the quality and consistency of the data in place. Figure 3 gives an example of a dashboard one showing refinery data color-coded by blend code.

**FIGURE 3: DASHBOARD DISPLAY OF DATABASE CONTENTS**

The number and layout of these data displays is limited only by the data and the metadata dropped into the database. In addition, any new data introduced into the database is predicted using the current model to determine if a model update is required.

**CALIBRATION**

The calibration approach follows the best practices guidelines in several texts [3, 4]. The addition of Robust statistics, specifically Robust PLS, provides an automatable approach to identifying the spectra that should be included in the calibration [5]. Our experience has shown that a careful application of Robust principles results in a calibration set that closely mimics what an experienced chemometrician would choose and can be fully automated. This avoids one of the most common mistakes by modelers with lower levels of experience.
Calibration requests in the current system are tailored for each spectrometer’s application; an example is shown in Figure 4. The request is made through an Excel interface that has been populated with customer-specific metadata and date ranges.

If the user has specified that models be created in a specific format, that model will be constructed as part of the automated process; there is no requirement to use any particular model format.

FIGURE 4: EXCEL SPREADSHEET FOR CALIBRATION REQUESTS

QUALITY CONTROL

When a calibration is performed, an evaluation of the model is made relative to a validation set and is compared to previous models on earlier data drawn from the same spectrometer. Results of the calibration are pushed into a cloud dashboard similar to the data review example shown in Figure 3. Tabular information is assembled for all properties requested and are displayed as in Table 1.
TABLE 1: DESCRIPTIVE METRICS FROM AUTOMATED MODEL GENERATION

<table>
<thead>
<tr>
<th>Property (group) / Property</th>
<th>Property</th>
<th>Other Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Octane</td>
<td>Hydrocarbons</td>
<td>Distillation</td>
</tr>
<tr>
<td>MON</td>
<td>RON</td>
<td>RVP</td>
</tr>
<tr>
<td>Total Samples</td>
<td>576</td>
<td>576</td>
</tr>
<tr>
<td>Retained Samples</td>
<td>467</td>
<td>518</td>
</tr>
<tr>
<td>% Inliers</td>
<td>81%</td>
<td>90%</td>
</tr>
<tr>
<td>Total Variables</td>
<td>1,400</td>
<td>1,400</td>
</tr>
<tr>
<td>Retained Variables</td>
<td>1,400</td>
<td>1,400</td>
</tr>
<tr>
<td>Property Min</td>
<td>32.70</td>
<td></td>
</tr>
<tr>
<td>Property Max</td>
<td>1,400</td>
<td></td>
</tr>
<tr>
<td>Median PLS Rank</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>RMSECV Over Range</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>SECV</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>Bias CV</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>p.Value</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

Other views are useful to track the model changes and their performance over time. Figure 5 shows the tracking for successive models charting PLS rank (number of factors in the model), the standard error of cross validation, and the error as a fraction of the data range.

FIGURE 5: TRACKING MODELS THROUGH TIME

OPTIMIZING CALIBRATIONS

We also must be careful not to overfit the models, which can leave them fragile and in need of more frequent updates. As models are created, the system must purge the data of outliers, either tied to a spectroscopy problem or from a mistake by the reference technique. In this implementation, we blended several techniques for determining the best number of factors to include in the final model. The decision is made as a function of the spectral distribution.
The true test of an automated system is to examine how it performs relative to an experienced chemometrician. Because the R&D effort tied to this project extended over 6 years, there were many petroleum refinery data sets to test performance. These data spanned FTNIR, NIR and Raman spectroscopy sources. An example comparison is summarized in brief in Table 2.

TABLE 2: COMPARISON OF AUTOMATED TO MANUALLY-PREPARED MODELS

Table 2 presents data from manual and automated calibrations on the same five prediction sets. In each case, either the human or the computer made decisions on outliers and number of factors for the models. The results shown here are statistically indistinguishable.

HANDLING THE PERIPHERY

The system is designed to allow flexibility in the dashboard displays and in other forms of reports. A model report in PDF format is automatically generated with each request. This document outlines the processing steps and describes the model in enough detail to be able to build it easily in any software environment. In addition, standard reporting (ASTM, CARB) is available along with custom reports. In Chevron’s case, we also build output that can be read directly into Spotfire.
RETHINKING SPECTROSCOPIC CALIBRATIONS

JUSTIFYING THE CHANGE

The fact that industry is moving to more analyzers and fewer technicians means that the calibration paradigm must change. The industry really has no choice. We present here one option, but one that uses generic principles. This is an approach that works.

CONCLUSIONS

The deployment of ever more specialized analytical technology results in fragmented data that can be tough to wrangle and keep track of for modeling. A database solution can ease and automate the routine aspects of this process, allowing for better calibration modeling. As we have steadily added spectroscopy-based analyzers to our process monitoring approach, we need to apply some effort to technology management in order to optimize the use of personnel. The approach described here not only streamlines the calibration process, it is an equalizer across multiple locations providing a common baseline for this activity.

Assembling the disparate functions into a single, coordinated system generates documentable benefits:
1. It lowers the cost of calibration and speeds response time;
2. The process scales easily allowing the incorporation of additional analyzers under a common computational and visualization umbrella;
3. It ensures a common baseline for calibration quality across locations and across analyzer technicians;
4. Communication among locations is facilitated and experiences in one location are transferable to all other similar installations; and
5. It enables continuity in the face of personnel changes.

ACKNOWLEDGEMENTS

As always, work that goes into a paper like this is not the isolated and unique thoughts of the listed authors. Many people assisted with insights, ideas, refinements, data and model management ideas, and description of use cases specific to their organizations. The author’s thanks go to the following exceptional people:

- Scott Ramos, Randy Pell, and Stan Christie, Infometrix
- Roger Bledsoe and Randy Ridge, Marathon
- Mary Beth Seasholtz, Dow Chemical
- Ron Skelton and Paul Weider, Shell
- Andrea Rauch, Honeywell/UOP
- Cory Phillips, Phillips 66
REFERENCES


