

RETHINKING CALIBRATION FOR PROCESS SPECTROMETERS II

Will Warkentin

Short Range Refinery Planner
Chevron Products Co.
Richmond Refinery
841 Chevron Way
Richmond, CA 94801

Brian Rohrbach

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 favored technology to measure chemistry and is ubiquitous in the hydrocarbon processing industry. In a previous paper, we focused on a generic, machine-learning approach that addressed the primary bottlenecks of mustering data, automating analyzer calibration, and tracking data and model performance over time. The gain in efficiency has been considerable, and the fact that the approach does not disturb any of the legacy (*i.e.*, no changes or alterations to any analyzer or software in place) made deployment simple.

We also standardized a procedure for doing calibrations that adheres to best practices, archives all data and models, provides ease of access, and delivers the models in any format. What remains is to assess the speed of processing and the quality of the models. To that end, a series of calibration experts were tasked with model optimization, restricting the work to selecting the proper samples to include in the computation and setting the number of factors in PLS. The amount of time and the quality of the models were then compared. The automated system performed the work in minutes rather than hours and the quality of the predictions at least matched the best experts and performed significantly better than the average expert. The conclusion is that there is a large amount of recoverable giveaway that can be avoided through automation of this process and the consistency it brings to the PLS model construction.

INTRODUCTION

There is a lot of mundane work tied to the assembly of spectra and laboratory reference values to enable quality calibration work. There is also insufficient guidance when it comes to the model construction task. How much time should be spent on this task? How to best assess whether a spectrum-reference pair is an outlier or not? How many cycles of regression-sample elimination make sense? Where do we switch over from improving the model by adding PLS factors to overfitting and incorporating destabilizing noise?

Why is this discussion important? Optical spectroscopy is one of the few instrumental technologies that allow us to assess the chemistry inside our manufacturing processes; the other being chromatography [1]. Spectroscopy measures the concentration of chemical functional groups, allowing us to infer a great number of properties with both accuracy and precision. Use of spectroscopy shifts the quality assessment from daily to minute-by-minute. In a petroleum refinery, where blending shifts can occur hourly, lab results are simply too infrequent and not timely. So, we adopt optical spectroscopy (NIR, IR or Raman) to solve the timeliness problem, but the quality of the reported values is completely controlled by the quality of the calibration.

When we consider the costs of octane or RVP giveaway, the approach we take to optimize calibration models takes on a huge significance. There are automated optimization solutions that make sense to deploy as discussed in our earlier paper [2]. Are the costs of the solution justified? In this paper, we focus on two questions: one whether an automated calibration system can perform as well as true experts in the field, and the second to provide a mechanism for companies to objectively evaluate the quality of the models they produce in-house. To the second end, we propose a cost-free test to judge a company's ability to generate optimal calibration performance.

This paper builds on papers presented in the 2015, 2018, and 2019 ISA-AD Symposia [3, 1, 2]. The results reported in this paper supports our earlier contention that consistency in the approach to spectroscopic calibration is critical. For background on the multivariate calibration process, there are myriad articles and books. For the uninitiated or those interested in the organization of industrial projects, we recommend the text by Beebe et al. [4]; for more advanced practitioners interested in theory, Martens and Næs [5] is an excellent place to start.

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.

SPECTROMETERS

The spectrometer in place in the Chevron gasoline blending quality control station is a Raman system from Process Instruments (Salt Lake City, UT, USA). Additional data was processed from an FT NIR was processed from an ABB Bomem spectrometer (Quebec City, Quebec, Canada).

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. Because the Process Instruments Raman analyzer uses Solo models (Eigenvector Research, Manson, WA, USA), this software was incorporated into the Ai-Metrix workflow to generate the final models. Similarly, the Ai-Metrix system generated GRAMS models (Thermo Fisher Scientific, Madison, WI, USA) for the ABB analyzer. Validation samples were identified using the duplex algorithm (Infometrix, Inc.).

DATABASE

Some form of database is critical for efficient operation. In the case of the Ai-Metrix example, the center 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 results. It also drives the modeling process, generated reports, and formats output for upload into the cloud for dashboard display.

RESULTS AND DISCUSSION

Two issues are the focus of this paper. First is to construct a test that measures the quality of the automated calibration. Second is to assess how responsive the system will be in routine use, both in the flexibility of the approach to handle custom requests and the speed of model delivery.

CALIBRATION QUALITY

The need for high-quality, consistent modeling practice is emphasized by a graph from our 2019 paper reproduced as Figure 1. In this case, a dataset was supplied to all scientists planning to attend a chemometrics conference in France. This challenge data set was processed by 19 chemometricians of varying experience. The results in the figure are ranked from highest error 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. One of the 19 was the Ai-Metrix approach, which yielded a result that fell into the “correct” category.

We want to be able to build calibration models in an automated manner, but if the quality of those models is not sufficient, the automation process is useless. The results below represent a reasonable initial test of the system, and provides confidence in moving toward on-line implementations.

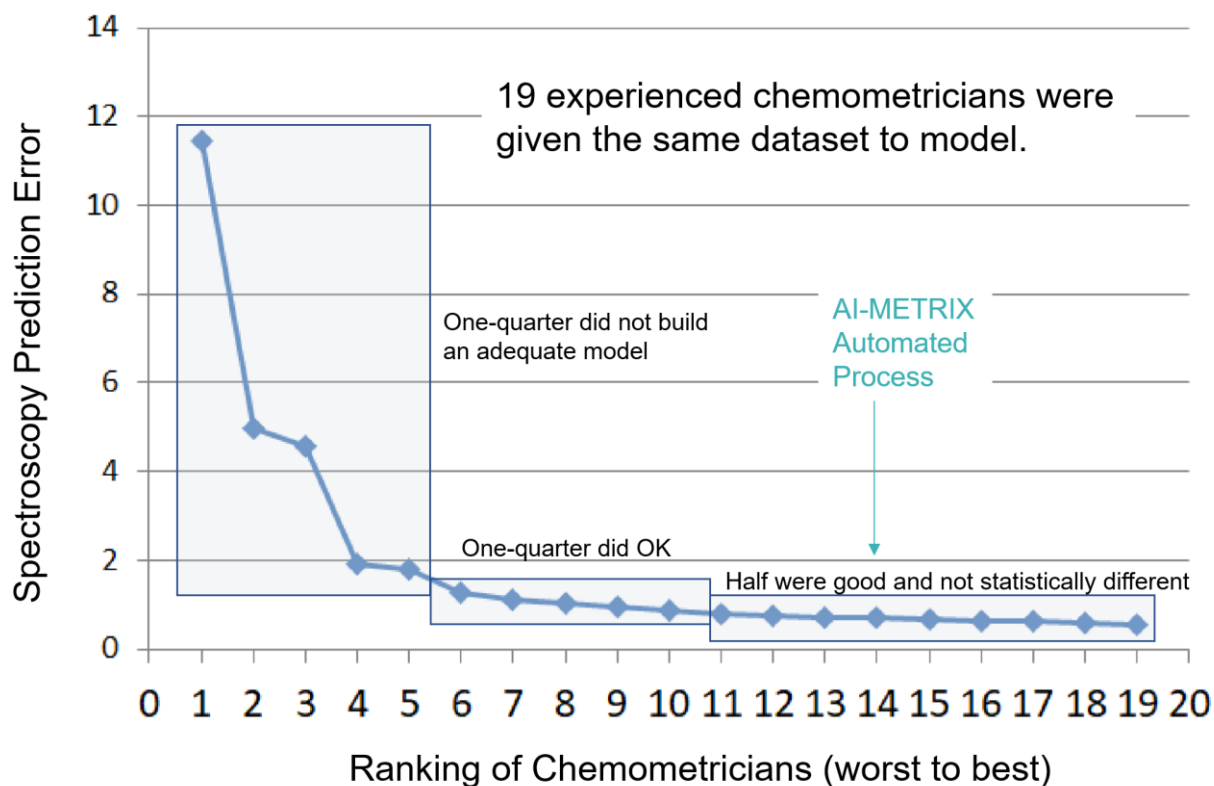


FIGURE 1: CALIBRATION QUALITY VARIATION AMONG PRACTITIONERS

The result above suggests a cause for concern when considering the procedure, the hydrocarbon processing industry follows in building calibrations that tie to product quality and ultimately corporate profitability. But, the example of Figure 1 represents a special case and one that does not necessarily correspond to our process environment. In the above test, the nature of the data was unknown, so no one had experience with similar data (it was an agricultural dataset). Table 1 shows the design of a hydrocarbon process dataset comparing the results of a PhD-level chemometrician to the automated system for a selection of properties important to gasoline blending. In this case, there is no statistically significant difference in the two sets of results.

TABLE 1: COMPARISON OF AUTOMATED TO MANUALLY-PREPARED MODELS

Spectroscopy	Component	Automated	Manual
Raman	Universal T-20	1.75	1.67
Raman	Universal Neat RON	0.35	0.37
Raman	Summer Neat RON	0.26	0.27
Raman	Winter Neat RON	0.37	0.36
Raman	Summer Neat T-50	1.45	1.57

The results reported in Table 2 are again of value to show that the automated modeling can perform as well as an expert, but the test represents a single instance, an anecdote. We set out to design a test that is of more statistical relevance. To that end, we prepared six sets of data drawn from two different spectroscopies and different refineries. We also wanted to test data sets of different sizes that would span typical calibration problems seen in the hydrocarbon processing industry. Table 2 lists the number of spectra/lab value pairs and the component under test.

TABLE 2: OVERVIEW OF THE SIX DATASETS USED IN THE COMPARISON

Spectroscopy	Component	# Cal	# Val
NIR	Benzene	2118	200
NIR	Octane	2136	181
Raman	Octane	586	60
Raman	Octane Special Blend	273	27
Raman	RVP	675	60
Raman	RVP Special Blend	277	23

Because octane and RVP represent the most interesting characteristics of gasoline as far as giveaway is concerned, the focus of these data was on those parameters. In addition, as the data sets were assembled, a subset of the samples was set aside as a validation set. This collection of samples was designed to be representative of the entire set, but without compromising the samples left in the calibration suite. The validation sets give us a means of grading models in an objective manner.

With these data sets in hand, we contacted experienced chemometricians, asking them to process the data sets. Eight responded with models and it should be noted that the average chemometric modeling experience level for these 8 was more than 25 years. In this experiment, the only task was to identify the outliers to exclude and specify the rank, or number of PLS factors to use. The identical data sets were sent through the automated Ai-Metrix process.

Table 3 reports the results of this test. Here, the six data sets are shown with the model with the lowest prediction error (RMSEP) at the top of the column grading down to the highest prediction error. The automated approach we are using for calibration consistently scored near the top, with low prediction errors. This demonstrates the fact that human factors, even among the most experienced practitioners, will have an impact on calibration quality; even if an expert produces an optimal model for a property, it does not guarantee that they can repeat the model quality every time. We are simply rushed to get all our tasks done.

These data sets are available for anyone who wants to participate in the comparison.

TABLE 3: DETAILED RESULTS LISTED FROM BEST TO WORST

NIR Octane	Prediction Error		Raman Octane	Prediction Error		Raman Octane2	Prediction Error
Expert #5	0.274		Expert #1	0.397		Expert #7	0.348
Expert #1	0.276		Automated	0.397		Expert #3	0.362
Automated	0.276		Expert #3	0.413		Expert #1	0.383
Expert #7	0.277		Expert #4	0.428		Automated	0.385
Expert #3	0.278		Expert #7	0.430		Expert #5	0.403
Expert #6	0.279		Expert #6	0.447		Expert #4	0.412
Expert #4	0.280		Expert #5	0.454		Expert #6	0.472
Expert #8	0.304		Expert #8	0.496		Expert #2	0.568
Expert #2	0.304		Expert #2	0.585		Expert #8	0.596
Raman RVP	Prediction Error		Raman RVP2	Prediction Error		NIR Benzene	Prediction Error
Expert #3	0.221		Expert #5	0.140		Expert #7	0.0211
Automated	0.226		Automated	0.147		Automated	0.0212
Expert #6	0.227		Expert #1	0.155		Expert #1	0.0214
Expert #8	0.227		Expert #6	0.176		Expert #5	0.0216
Expert #5	0.234		Expert #7	0.179		Expert #3	0.0227
Expert #1	0.235		Expert #4	0.180		Expert #2	0.0229
Expert #4	0.238		Expert #3	0.189		Expert #4	0.0246
Expert #7	0.240		Expert #8	0.214		Expert #6	0.0247
Expert #2	0.349		Expert #2	0.284		Expert #8	0.0248

Given that there are errors in the measurement of any property, it is not possible to know when we have reduced the spectroscopic assessment error beyond some limit. For a more detailed discussion of this issue, please refer to Trygstad et al. [3]. For this discussion, the authors have defined the best result in Table 3 above to represent the point at which anything higher is recoverable by a better calibration model. Thus, this *recoverable giveaway* represents the potential windfall for a plant if they were able to get the best possible model.

So, by subtracting the difference between any individual's error result compared to the best individual's result for that parameter, we have an estimate of the benefit that each participant's calibration models would compare to all others. Here, we calculate the percent difference to place all values on the same scale, then take the average across all six models to yield the chart shown in Figure 2.

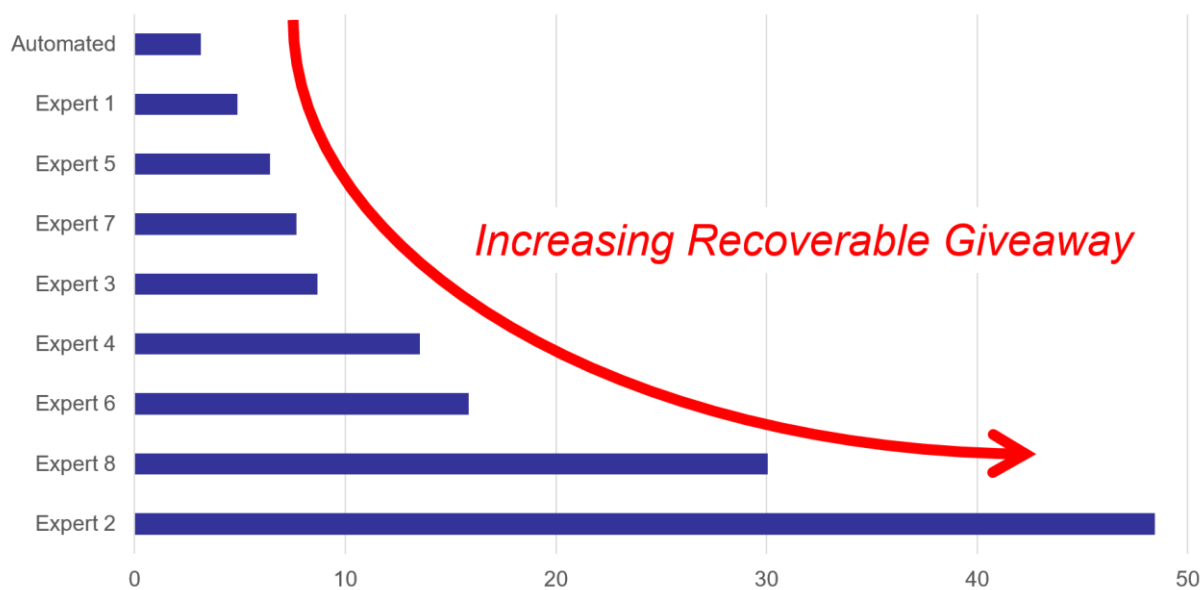


FIGURE 2: AVERAGE PERCENT HIGHER THAN THE BEST TEST SCORE

Even though one of the experienced chemometricians always scored the best and established our border for recoverable giveaway, no one chemometrician was consistently the best. Considering all six results, the best overall score was achieved by the automated system, in part due to its consistency in applying the principles of best practices described in our earlier paper [2].

To convert bragging rights into dollars, we need to consider the cost of this recoverable giveaway. Figure 3 shows one of the published charts of the cost of octane and RVP giveaway [6].

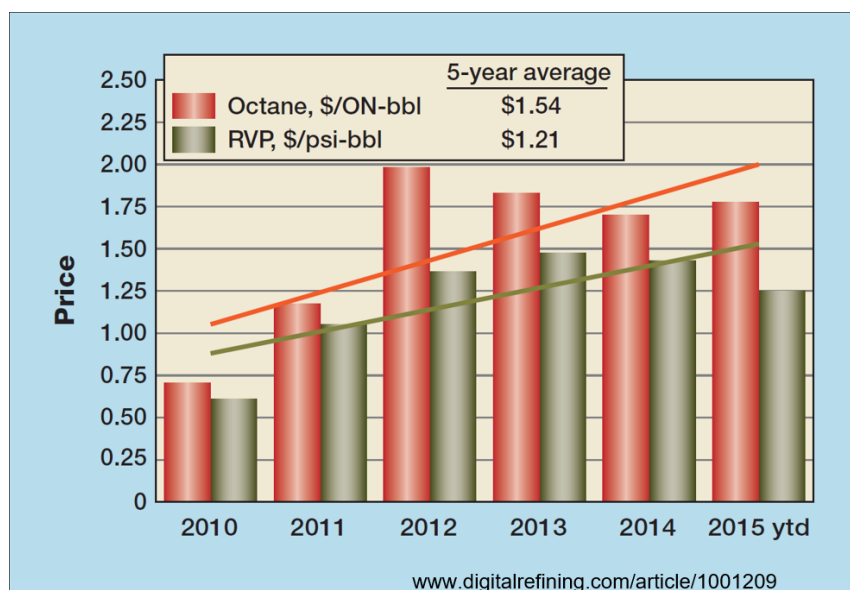


FIGURE 3: ESTIMATING THE COST OF GIVEAWAY IN DOLLARS PER BARREL

RETHINKING SPECTROSCOPIC CALIBRATIONS II

Estimates of the cost of giveaway for octane and RVP typically range from \$1.00 to \$2.00 per barrel of product. If we assume \$1.25/bbl as the cost of RVP giveaway and \$1.50/bbl as the analogous cost for octane, we have the basis for moving this evaluation into the accounting realm. If we further assume that we are operating a small refinery, with a capacity of 100,000 barrels of gasoline per day, the numbers work out as shown below.

- Expert #1 had the best score of the experts. In this case, recoverable giveaway would be
 - **\$240,000 per year.**
- The average of the 8 experts compared to the automated system yields a loss of
 - **\$3.7 million per year.**
- If we take the worst case of the experts, the number rises to an astonishing
 - **\$13 million per year.**

Given that the above calculations are for work performed by top experts in the field, the consideration of optimizing the calibration models is likely more valuable in many settings. Can you use this higher-precision assessment to manage giveaway better or possibly squeeze out additional blends?

CALIBRATION EFFICIENCY

So, high quality models are a hallmark of the automated process tested for this paper. The other issue is model cycle time. Here we consider several issues.

- Identification of when a model needs to be updated
- Capturing the time required to do (or have done) the modeling given other time demands
- Speed of the modeling process itself.

The Ai-Metrix system tracks the model quality whenever new results are added to the system, so the user is always aware of the need for a replacement model. Capture of time to match spectra with laboratory reference values and build a quality model is easier if the time required is small, but capturing the time for a model redo for 10 or 20 quality parameters often triggers a delay of days or weeks, even months. The result is we do not update our models as often as we should.

In the model quality testing for the six data sets described above, we asked that the experts report the time spent on all six model optimizations. The times reported ranged between 2 hours and 3 hours for the 6-model task.

Triggering the automated process was done by opening Excel to a custom spreadsheet and setting the process to run [2]. The operator time was less than 2 minutes. The time required to optimize - eliminating outliers and establishing the complexity (number of factors: rank) - for each of the six models is shown in Figure 4.

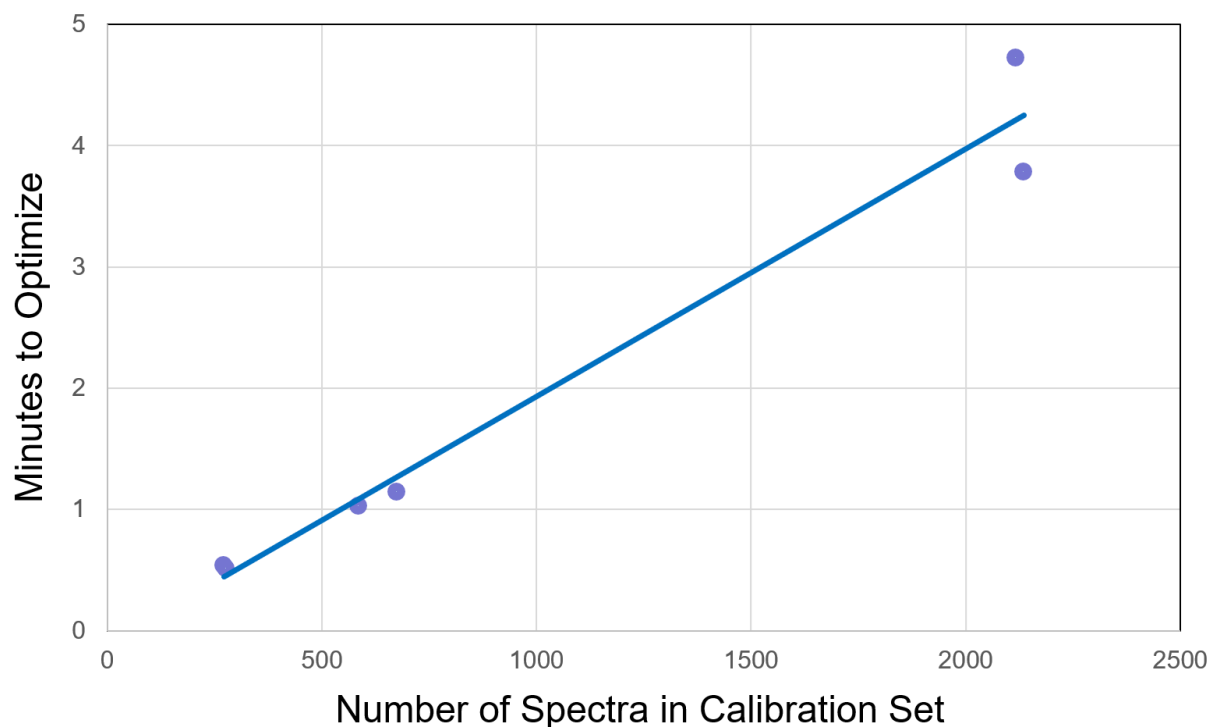


FIGURE 4: TIME TO OPTIMIZE MODELS AS A FUNCTION OF DATASET SIZE

The time to optimally understand exactly which samples should be retained for the calibration (eliminating outliers) and set the number of PLS factors ranged from around 30 seconds for a 275-spectrum model to 4 - 5 minutes for datasets of 2,000 spectra. Because of the machine learning approach to the optimization step, the times are not a strict function of number of spectra, but processing times are always close to the trend line above. So, automation did all this calibration work in just over 11 minutes, 10-20 times faster than the experts.

So, what is the value of more frequent updates? Given that we can optimize the calibration settings in a small number of minutes, it addresses situations where our traditional hand-held modeling approach suffers. More timely access to calibration models would be extremely valuable during a major unit shutdown, where the refinery posture and gasoline blend chemistry can dramatically change. The shift is often sudden and, unless the chemometrician is refreshing models daily, there is a long lag where the refinery must endure poor performance and use greater recipe cushions while “enough” data is collected to warrant an update and the time allotted to the calibration processing step. Major shutdowns typically occur on a five-year cycle and unless the data contains sample pairs from the last shutdown, there’s a good chance the models haven’t been trained to understand octane without olefins or RVP without C5/C6 isomerate.

Another timeliness example is during the transition from summer to winter gasoline specs. When we fire up our first winter blend of the year, we will be making predictions using data collected the previous winter. A lot can change in a refinery over eight months and it is interesting to see how models perform through the first weeks until we again collect “enough” data to update them.

CONCLUSIONS

We have been designing and improving a simplified approach to assembling lab values and spectra into a common database and using it to provide objective, high-quality calibration models on demand automatically. With technologies made commonplace in the information processing industries, there are new, better mechanisms to track data, track process quality, and do so inexpensively. The fact that the approach will work with any brand of spectrometer and any chemometrics software package makes it easy to deploy within existing infrastructure.

Given that a PLS calibration is not trivial and the pool of available calibration personnel is not keeping pace with analyzer installations, there must be a shift in how we manage these assets. Automated calibration adds a layer of both accuracy and consistency that in turn fills a need by optimizing the use of personnel, simplifying and speeding the delivery of high-quality calibration models. The approach also provides a corporate-wide common basis for calibration across instrument types and experience levels.

A single, coordinated system:

1. Speeds response time;
2. Lowers the cost of calibration;
3. Scales easily under a common computational and visualization umbrella;
4. Ensures a common baseline for calibration quality across the company; and
5. Enables continuity in the face of personnel changes.

Extending this thought brings up an intriguing possibility in this world of Industry 4.0. As we are now successful in automating the calibration process with competent, optimized models on demand, it raises the possibility of closing the loop of calibrating and maintaining spectroscopic instruments completely and adapts to the *chemistry du jour*. We can detect the degradation of a calibration model, prepare its replacement, and install the update all through the circulation of a pile of electrons. Implementing this will require additional high-level corporate buy-in, but the possibilities for increased profit by decreasing giveaway are enormous.

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 authors' thanks go to the following exceptional people:

- Scott Ramos, Randy Pell, and Stan Christie, Infometrix
- Mary Beth Seasholtz, Dow Chemical
- Brian Goodlander and Jon Stockert, Procter & Gamble
- Andrea Rauch, Honeywell/UOP
- Yangdong Pan, Josh Heltsley, and Cory Phillips, Phillips 66

REFERENCES

1. Rohrbach, B. “Prioritizing Multivariate Control”, Proceedings of the Instrument Society of America Meeting, Analysis Division, April 24, 2018, Galveston, Texas.
2. Warkentin, W., and Rohrbach, B. “Rethinking Spectroscopic Calibrations”, Proceedings of the Instrument Society of America Meeting, Analysis Division, May 7, 2019, Galveston, Texas.
3. Trygstad, M., Pell, R., and Roberto, M. “Motor Fuel Property Prediction by Inferential Spectrometry 2. Overcoming Limitations”, Proceedings of the Instrument Society of America Meeting, Analysis Division, April 28, 2015, Galveston, Texas.
4. Beebe, K., Pell, R. and Seasholtz, M.B. *Chemometrics: A Practical Guide*, (John Wiley & Sons: New York, 1998).
5. Martens, H. and Næs, T. *Multivariate Calibration* (Chichester: John Wiley & Sons, 1989).
6. Seiver, D. “Minimizing Gasoline Specification Give-Away”, Digital Refining, October 2015, www.digitalrefining.com/article/1001209