AUTOMATING ASTM SPECTROSCOPY VALIDATION

Scott Ramos, Randy Pell, Brian Rohrback
Infometrix, Inc.
19119 North Creek Parkway, Suite 100
Bothell, WA 98011

KEYWORDS

Chemometrics, Best Practices, Model Tracking, Optimization, Calibration, Fuel Blending, Optical Spectroscopy, Partial Least Squares, Process Control, Automation

ABSTRACT

If a company decides to follow ASTM-compliant reporting to meet industry standards for spectroscopy model calibration and process monitoring, one finds that current systems struggle to deliver this compliance information efficiently. The manual validation process defined in D6122 is tedious and fraught with steps that are conducive to error. A mechanism has been developed that automates this evaluation, presents the results in a dashboard, and sends an email to the user when a fault is detected. The automated approach is simple to integrate into any system regardless of the source of the analyzer and independent of any software in place. This approach eliminates hands-on work and provides near real-time monitoring to flag non-compliant data or processes, ensuring continuous adherence to ASTM standards.

INTRODUCTION

One cannot control what is not being measured, and optical spectroscopy provides those measurements in near real-time. The spectrum generated is not a primary measurement, however, and needs to be correlated to the properties of interest via a calibration model. The quality of the correlation is a function of the analyzer’s base technology (IR, NIR, Raman, NMR typically), the quality of the hardware itself, and the quality of the calibration model. An ASTM committee was formed to update a series of standards governing how to manage the deployment of optical spectrometers for quality control measurements in the hydrocarbon processing industry. This involved updating a series of standards that detail best practices for all steps, from setting up the spectrometer to managing the calibration process. These standards recognize, particularly true in the hydrocarbon processing industry, that calibrations can go stale; spectrometers age, the input hydrocarbon streams can change in composition, units operating on various streams can shift compositions of a given stream based on temperature, flow, and catalyst. The ASTM committee wrote D6122 to provide an objective means of continuously evaluating model quality [1]. This paper was written to encourage the application of D6122 validation, to explain the steps necessary to follow the standard, and to introduce an automated mechanism that implements D6122 to save operator time.

- The rewriting of standards under the ASTM D8340 banner enables quality control based on the optical spectroscopy results alone.
- This requires more frequent validation of the calibration model in use.
- There is a specified series of steps required to manage the process.
- These steps can be managed by software and benefit greatly from automation.

This paper continues a three-decade long pattern of talks presented at the Analyzer Technology Conference and its predecessor, the International Society of Automation – Analyzer Division (ISA-AD) [2-7]. A series of detailed discussions form the basis of best practices for spectroscopy calibration and optimal strategies for automation. In parallel, extending the automation tactics to this ASTM model validation simplifies adherence to the standard and forms the basis for reducing the number of reference samples that need to be processed in the laboratory.

OUTLINE

The goals here are first to enumerate each step necessary to follow the D6122 Standard and second to define the procedural tasks to follow for complete automation. The first steps, listed below, are to create a calibration model for routine use:

1. Data from the spectrometer and the laboratory need to be assembled and matched.
2. The regions of the spectrum to be used in the model need to be defined.
3. The preprocessing of the spectra needs to be defined.
4. A robust analysis needs to be performed to filter out bad spectra and misreported reference values.
5. A calibration model needs to be generated, optimized, tested, and installed for use.
6. The samples included in the above calculation are then used to calculate the sample-specific guidelines that will be employed to assess ongoing model applicability.

The only additional task for current spectroscopy users is step #6 above. With a functioning model in place, additional reference values will need to be collected in the laboratory and routed with the associated spectra through the D6122 process, listed below, to ensure that the model is operating properly:

1. As in the calibration step, data assembly, matching reference values to spectra, and filtering of egregious outliers must be done.
2. Examine the difference between the predicted value from the spectroscopy to the lab reference and check it against the allowable deviation for that given concentration and flag that sample if it fails the test.
3. If more than a given number of failures occur within the last 20 qualified samples, the validation is judged to be invalid and a bias adjustment, model augmentation, or a recalibration is required.
4. After recalibration and an evaluation period, the system can return to assessing model validity.

This paper first discusses the step of preparing the sample specific limits that will be used in routine evaluations.

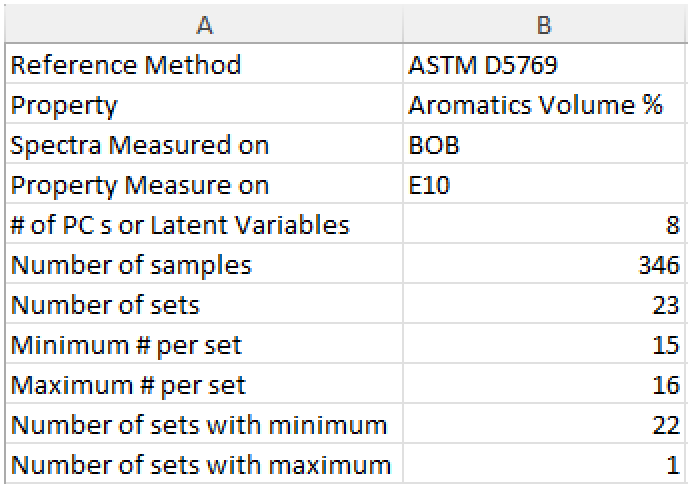
SAMPLE-SPECIFIC GUIDELINES

The most tedious task is in assessing heteroscedasticity of the predicted property residuals (*i.e.,* the difference between the spectrometer reported value and the reference lab reported value) and using it to set the sample-specific guidelines to use in assessing the applicability of the calibration model. This will then be used to assess quality for the continuing stream of analyses. Because this evaluation is specific for every reference property under management, each property will need to be processed separately.

ORGANIZE THE CALIBRATION DATA

There is an accompanying Excel® spreadsheet available through ASTM that works several examples to illustrate the procedure. Captures of portions of the spreadsheet are reproduced here with a description of the tasks required. First is to pull together a summary of the model parameters as shown in Table 1.

TABLE 1. CALIBRATION MODEL DETAIL



A	B
Reference Method	ASTM D5769
Property	Aromatics Volume %
Spectra Measured on	BOB
Property Measure on	E10
# of PC s or Latent Variables	8
Number of samples	346
Number of sets	23
Minimum # per set	15
Maximum # per set	16
Number of sets with minimum	22
Number of sets with maximum	1

Based on the calibration model, it is necessary to extract the rank (number of factors/principal components) for the model. With the assembled calibration data of the laboratory and the spectrometer values, compute the average of the reference measurement and the predicted value from the calibration model. Then the samples need to be ordered by this average value from low to high.

The next step is to divide the list of average values into groups of 15. If the last group does not contain 15 samples, some groups will be given 16 samples until there are no orphan samples remaining. By convention, the addition is done for the higher reference value groups. The minimum number of spectrum/reference pairs necessary is 60 (*i.e.,* 4 groups of 15).

CALCULATE GUIDELINES FOR ROUTINE ASSESSMENT

To compute sample specific guidelines, the data assembled above will be used to determine if the dataset displays an increase or a decrease in the error of prediction as a function of concentration. If the error is not constant across the range of values in the calibration set, it is necessary to fit a function that accounts for that variation. There are two types of modifications in that case. A linear correction is correct when the calibration error consistently increases or decreases as the property value increases. A power function can be fitted if the rate of increase is not constant. Figure 1 illustrates the three error behaviors from which to select; increasing concentration left to right is on the x-axis and error is on the y-axis.

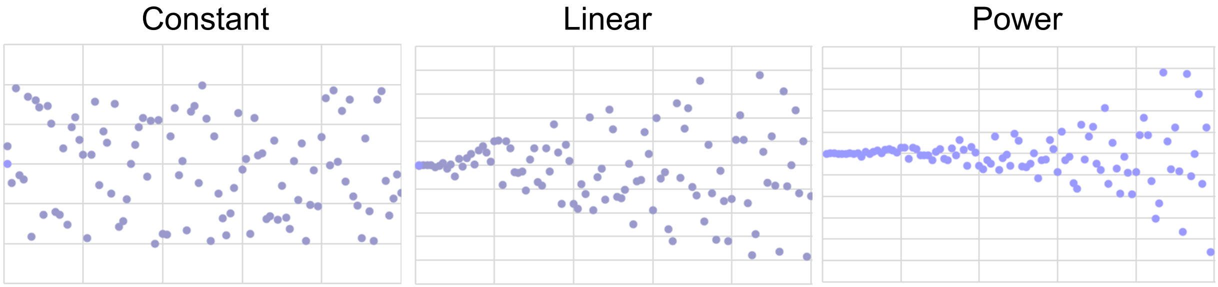


FIGURE 1. CONCENTRATION VERSUS ERROR FOR DATA REQUIRING CONSTANT, LINEAR, OR POWER CORRECTIONS

To decide which guideline to use, a calculation needs to be made to see which equation best models the calibration data for a given property. To start, pull the Standard Error of Calibration (SEC) from the model file and calculate the degrees of freedom by subtracting the number of model factors (latent variables) in that model from the total number of calibration samples. Take the list of residuals collected above, square each value, and calculate the sum. This sum of squared residuals (SSR) will be used for comparison.

Next, compute the slope and intercept of the line fitted through the grouped data values and calculate the sum of squared residuals for the Linear model. Take the ratio of the Constant SSR to the Linear SSR and compare it to the F-ratio limit; if the SSR ratio is larger than the F-ratio limit, the Linear sample specific guidelines will be used. A similar approach is used to see if a Power curve fit outperforms the Linear. Note that the Power curve fit requires a non-linear fitting of the data. An example from the ASTM spreadsheet is reproduced in Table 2.

TABLE 2. COMPONENTS USED IN THE CALCULATION OF SIGNIFICANCE

Constant		**Linear**	**a*m+b**	**Power**	**a*(m+b)^c**
SEC	0.2571	a	0.02511	a	0.0232
dof	338	b	-2.0629	b	-83.12
Sum Squared Residuals	0.09173	dof	337	c	1.039
		Sum Squared Residuals	0.08569	dof	336
		F-ratio	1.0704	Sum Squared Residuals	0.08569
		F-limit	1.2383	F-ratio	1.0000
		Significance	FALSE	F-limit	1.239
				Significance	FALSE

Model to Use	Constant

This computation outlined in Table 2 is tedious, even with the assistance of a spreadsheet. One should note that the calculations need to be performed every time a new model is built and repeated for every quality parameter the company wishes to track. ASTM D6122 generates a significant amount of work that can lead to errors if performed manually. Automating this process is critical if a company wants to efficiently follow D6122.

Take the example of a medium to large US refinery. Across all products, all seasons, all grades, and all component streams, it measures a total of 238 properties. This results in 2,205 models built over the last few years. Of this total, 1,253 models did not have enough samples (60) to qualify for the ASTM significance test. Of the rest, almost two-thirds fell into the Constant realm signifying that there was no concentration dependance in computing the sample-specific errors. Most of the remaining showed a linear increase in error diagnostics as a function of concentration. Very few fell into the Power category. Graphics showing the automated evaluation of all models are shown in Figures 2 and 3.

238 reference values have been modeled over time.

43% of the models had enough calibration spectra to evaluate.

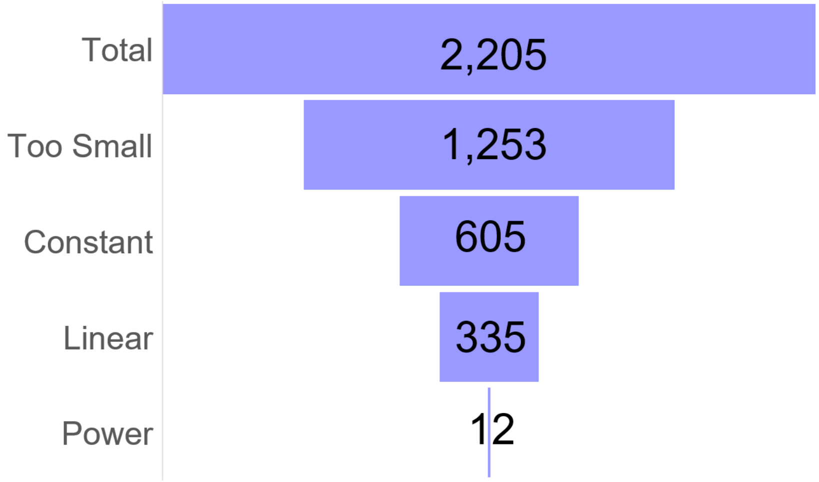


FIGURE 2. CASE STUDY: MODELS FOR AN EXAMPLE REFINERY

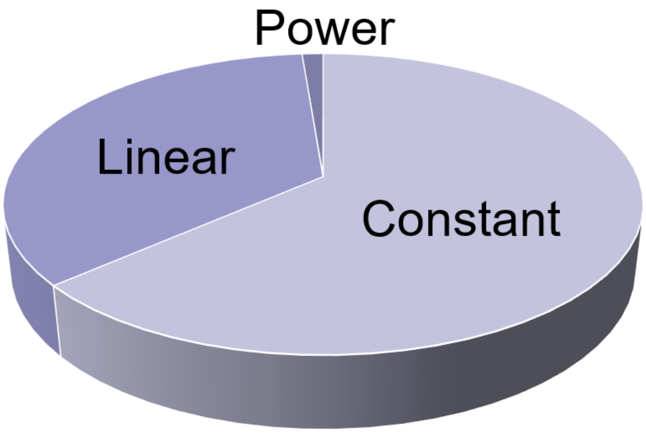


FIGURE 3. BREAKDOWN OF MODEL CHARACTERISTICS

ROUTINE ANALYSIS

The above handling of the calibration data provides the basis for the routine analysis during prediction of properties of new samples. With this enhanced understanding of error, the sample-specific error as reported by ASTM D6122 [1] can be calculated based on Equations 1, 2, or 3, as shown below.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n} \delta_i^2}{n}} \qquad \text{Constant} \tag{1}$$

$$RMSEC_L = \sqrt{\frac{\sum_{i=1}^{n} (a\ m_i\ +\ b)^2 + SSE_L}{n}} \qquad \text{Linear} \tag{2}$$

$$RMSEC_P = \sqrt{\frac{\sum_{i=1}^{n} (a\ (m_i\ +\ b)^c)^2 + SSE_P}{n}} \qquad \text{Power} \tag{3}$$

Whereδ = the difference between the measured and the predicted value
n = number of calibration samples
m = average of the measured and predicted values
RMSEC = Root Mean Square Error of Calibration
SSE = Standard Error of the Estimate
a, b, and c = fitted values

Look at the details for this computation in the ASTM document. There is a lot of ongoing work involved in following the standard. Fortunately, everything involved can be completely automated.

QUALIFYING THE MODEL FOR USE

With the calibration model in place, samples will continue to be evaluated using the reference method. The ASTM standard states that 20 comparisons of spectrometer prediction versus laboratory value must fall within the sample specific limits before the spectrometer values are deemed to be reliable. In this initial model validation, only samples that are both within the guidelines for sample residual and are also within the limits for in-model and out-of-model diagnostics qualify to be counted.

CONTINUOUS MODEL QUALITY EVALUATION

Once this test phase is done, the sample residual and the outlier diagnostics are used to flag model deviations. This is done on a moving window of results until there is a statistically significant deviation, which necessitates a recalibration, a model augmentation or a bias correction to continue. An example of this tracking is shown in Figure 4 below.

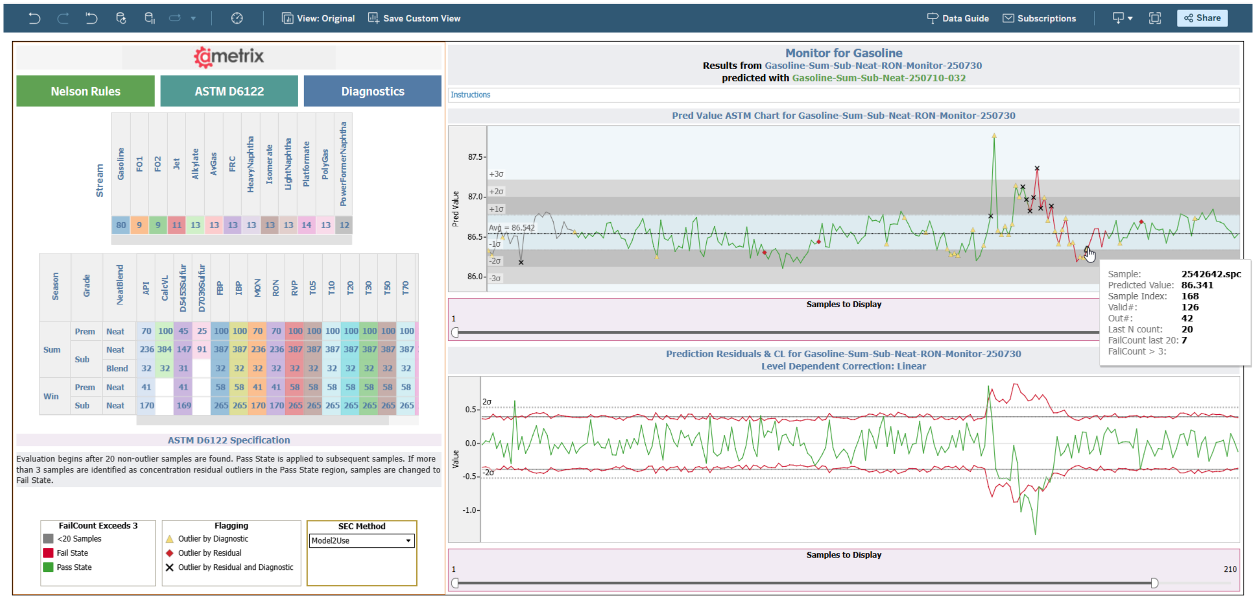
FIGURE 4. EXAMPLE TRACKING OF MODEL APPLICABILITY

The top part of Figure 4 shows the initial training range for the model quality as a gray line; the line turns green when the model qualifies for routine analysis. Red squares show where the difference between the spectral measurement and the lab are outside of the sample-specific guidelines. Yellow triangles identify analyses where one or both outlier diagnostics fail. The black X shows points where both the sample residual and one or both outlier diagnostics show the sample to be both a bad fit to the model and a too high residual. The math for assessing outliers is listed in the ASTM standard [1].

The lower chart shows the prediction residuals as a green line and the sample-specific prediction limits in red. This MON example has an increase in acceptable error bounds as a function of concentration and employs a linear correction factor.

AUTOMATION

Clearly, there is a lot to track and a lot to calculate. Doing this work by hand is subject to error making it the ideal case for automation. Infometrix fully automated D6122 in the Ai-Metrix® calibration service and found a series of steps were necessary to follow as discussed below. Figure 5 shows a complete view of the automated tracking mechanism which summarizes the calibration features that should be tracked in an automated implementation. The observations summarize the learning done as the development process was completed.



Stream	Gasoline	FO1	FO2	Jet	Alkylate	AvGas	FRC	HeavyNaphtha	Isomerate	LightNaphtha	Platformate	PolyGas	PowerFormerNaphtha
	80	9	9	11	13	13	13	13	13	13	14	13	12

Season	Grade	NeatBlend	API	CalcVL	D5453Sulfur	D7039Sulfur	FBP	IBP	MON	RON	RVP	T05	T10	T20	T30	T50	T70
Sum	Prem	Neat	70	100	45	25	100	100	70	70	100	100	100	100	100	100	100
	Sub	Neat	236	384	147	91	387	387	236	236	387	387	387	387	387	387	387
		Blend	32	32	31		32	32	32	32	32	32	32	32	32	32	32
Win	Prem	Neat	41		41		58	58	41	41	58	58	58	58	58	58	58
	Sub	Neat	170		169		265	265	170	170	265	265	265	265	265	265	265

ASTM D6122 Specification

Evaluation begins after 20 non-outlier samples are found. Pass State is applied to subsequent samples. If more than 3 samples are identified as concentration residual outliers in the Pass State region, samples are changed to Fail State.

FailCount Exceeds 3

- ▨ <20 Samples
- 🟥 Fail State
- 🟩 Pass State

Flagging

- ▲ Outlier by Diagnostic
- ◆ Outlier by Residual
- ✕ Outlier by Residual and Diagnostic

SEC Method: Model2Use

Monitor for Gasoline
Results from Gasoline-Sum-Sub-Neat-RON-Monitor-250730
predicted with Gasoline-Sum-Sub-Neat-250710-032

Pred Value ASTM Chart for Gasoline-Sum-Sub-Neat-RON-Monitor-250730

Sample:	2542642.spc
Predicted Value:	86.341
Sample Index:	168
Valid#:	126
Out#:	42
Last N count:	20
FailCount last 20:	7
FaliCount > 3:	

Prediction Residuals & CL for Gasoline-Sum-Sub-Neat-RON-Monitor-250730
Level Dependent Correction: Linear

FIGURE 5. AUTOMATED TRACKING OF 238 MODELS

The left side of the screen lists the streams monitored on the top with the number of times a model has been requested for that stream in the colored squares. Below the streams layout are the latest models built arranged by season and grade with the number of calibration samples for that measure. The right side is an example of tracking the model quality for RON as a function of time. The top portion of the right side displays the RON values and, as in Figure 6, the out-of-spec samples are marked. If the number of evaluation failures reaches the critical point, the tracking line turns red. The lower portion of the right side of Figure 5 shows the sample residual (the difference between the spectrometer value and the lab value) along with the sample-specific quality guidelines in red.

The design outlined below reflects a poll of end user comments and Infometrix company experience. Experience suggests that all the automation parameters should be discovered using the following steps.

- Select only the most recent calibration request for each permutation of contexts.
- Use only samples dated after the last calibration sample.
- Fix the date range: start after the last calibration sample or 365 days before the current date, whichever is later.
- Schedule the frequency to process updates for the visualizations depending on client needs.
- Archive everything.

CONCLUSIONS

The establishment of ASTM D6122 is a significant advance in managing optical spectrometer results in an objective, scientific manner. Proper application of the guidelines provides an unbiased assessment of calibration model quality and its fit for routine assessment of chemical composition and physical properties. Implemented properly, the approach enables product release decisions based on the spectroscopy alone. It can reduce the need for confirming laboratory analysis and make the process of maintaining calibration models vastly more efficient and objective.

ACKNOWLEDGEMENTS

As always, work that goes into a paper like this is not the isolated and unique thoughts of the listed authors. Thanks go to Roger Bledsoe of Marathon and Bryan Bowie of ExxonMobil who did the herculean task of corralling a series of procedures, modernizing them, and generating an integrated set of standards to guide the calibration of optical spectroscopy analyzers and laboratory instruments. Their guidance and instructive discussion were critical to understanding ASTM goals. Thanks also to the users of the Ai-Metrix® calibration service for critical comments, useful feedback, and enduring a succession of updates to the D6122 implementation as it progressed to being a product.

REFERENCES

1. “Standard Practice for Validation of the Performance of Multivariate Online, At-Line, Field and Laboratory Infrared Spectrophotometer, and Raman Spectrometer Based Analyzer Systems”, ASTM D6122-23 Approved July 1, 2023.

2. Workman, J., Kowalski, B., and Mobley, P., “The Design of an Expert Calibration System for Spectroscopic Based Process Analytical Chemistry”, Proceedings of the Instrument Society of America Meeting, Analysis Division, 1995.
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, 2015.
4. Rohrback, B., “Prioritizing Multivariate Control”, Proceedings of the Instrument Society of America Meeting, Analysis Division, 2018.
5. Warkentin, W., and Rohrback, B., “Rethinking Spectroscopic Calibrations”, Proceedings of the Instrument Society of America Meeting, Analysis Division, 2019.
6. Warkentin, W. and Rohrback, B., “Rethinking Spectroscopic Calibrations II”, Proceedings of the Instrument Society of America Meeting, Analysis Division, 2020.
7. Rohrback, B., “Fundamentals of Quality Control Employing Optical Spectroscopy”, Proceedings of the Analyzer Technology Conference, 2024.
8. Rohrback, B., “Minimizing Error in Calibrating Spectrometers”, Proceedings of the Analyzer Technology Conference, 2025.