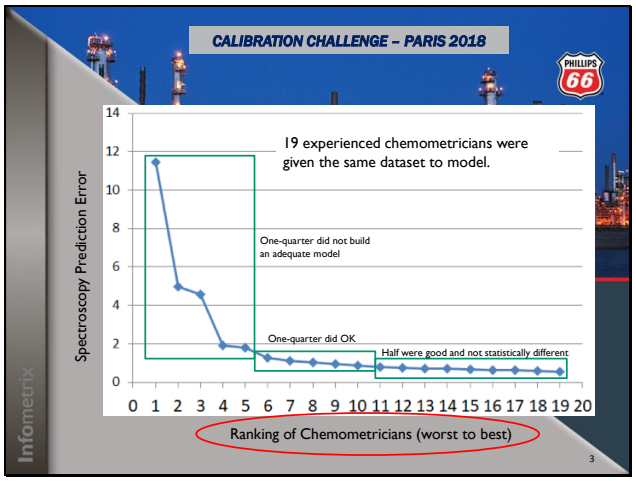
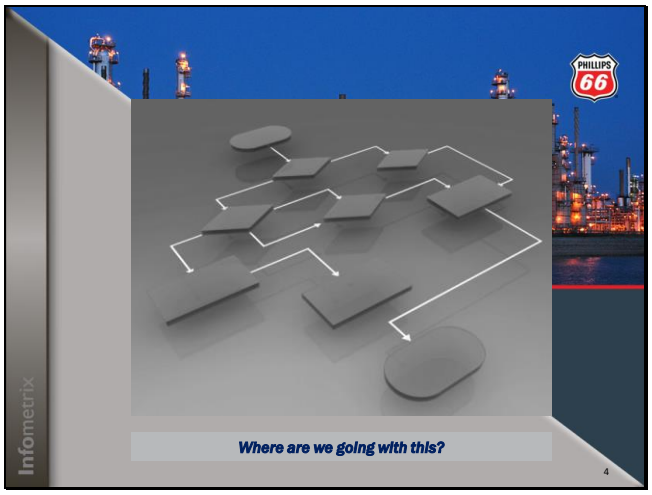
This slide deck was presented at the International Forum for Process Analytical Chemistry meeting in Washington DC, March 4th, 2019.

Infometrix, with feedback from Phillips 66 has undertaken a project to reduce the effort devoted to producing, maintaining, and stabilizing optical spectroscopy performance in routine quality assessment. We have examined an unprecedented historical collection of spectra from multiple spectrometers spanning 1-5 years from sixteen manufacturing settings, with the goal of developing and maintaining stable models for long-term deployment. The technologies utilized follow a pattern of best practices, including the use of Robust outlier diagnostics, local and hierarchical modeling, and model augmentation. The effort has resulted in significant progress towards automation of model creation, stability, and maintenance in an industrial process. The assembly of chemometrics technology with bits borrowed from current computer trends leads to a significant reduction in deployment and maintenance cost of these instruments.

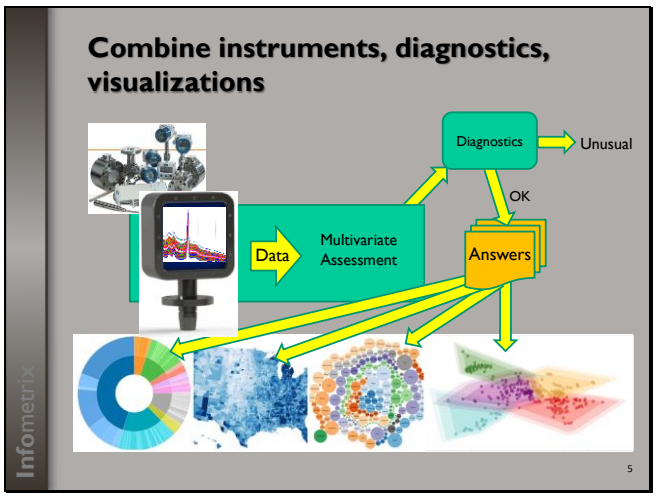


When you build a model, no one checks your work. How do you know how good the models are? We built a system to build models objectively to provide at least a reference to know that your models are good.

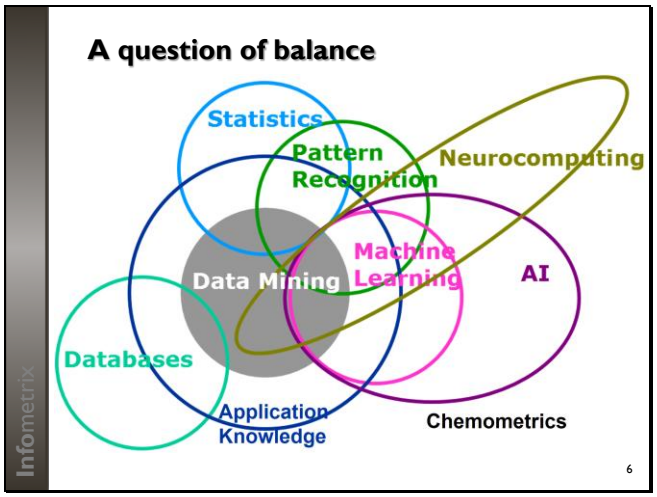
Here 19 chemometricians with varied experience levels were tasked with creating a model which, in turn, was tested against a validation set. One-quarter did not produce an acceptable model, one-quarter did not sufficiently optimize the model, half did fine and are not statistically different from one-another.



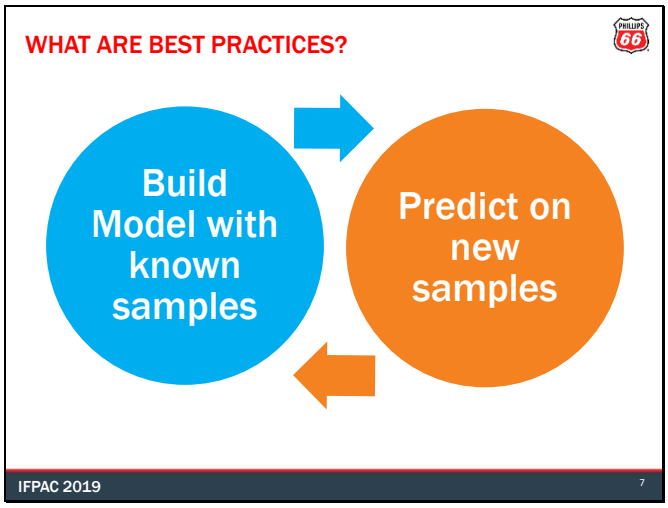
Our goal is to make the deployment of optical spectrometers easier and cheaper. What if we could install a NIR or Raman with the same level of effort that we put a temperature sensor in place. That goes for maintenance too. Let's define what we can do with the calibration task to support this endeavor.



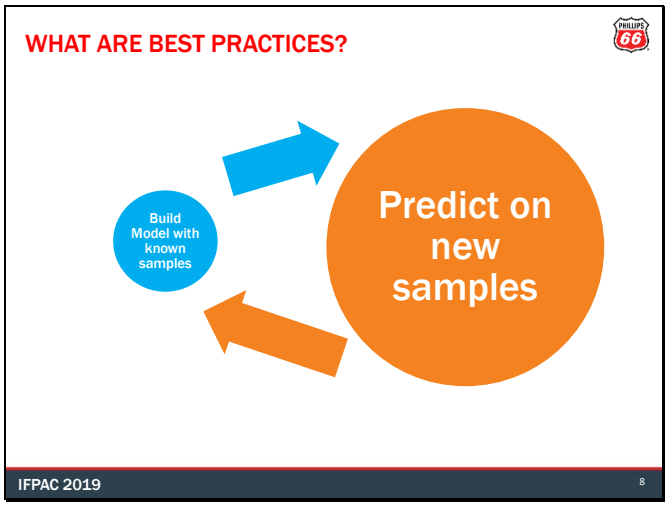
This is the informatics world that Infometrix plays in. Instruments and sensors are the data input to a multivariate processing world designed to convert that data stream into actionable information. Diagnostics are key in telling us the quality of this data-information transformation and we need to address how to push that information feed to the places where it can indeed be acted upon.



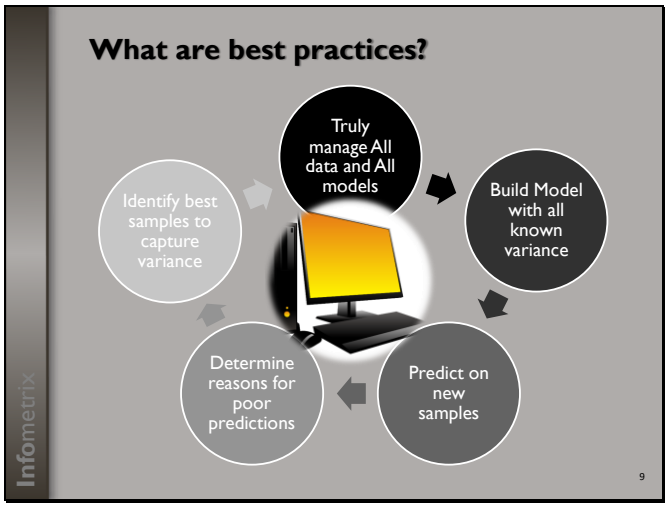
There is a huge amount of confusion and hype that surrounds the processing of data to form this information feed. Note there is significant overlap and the choice of using different terminology to cover the same algorithmic grounds is not helpful. A successful system needs to draw from all sources where there is a logical advantage of that field's approach. The system discussed here combines techniques from all areas displayed in a unique way.



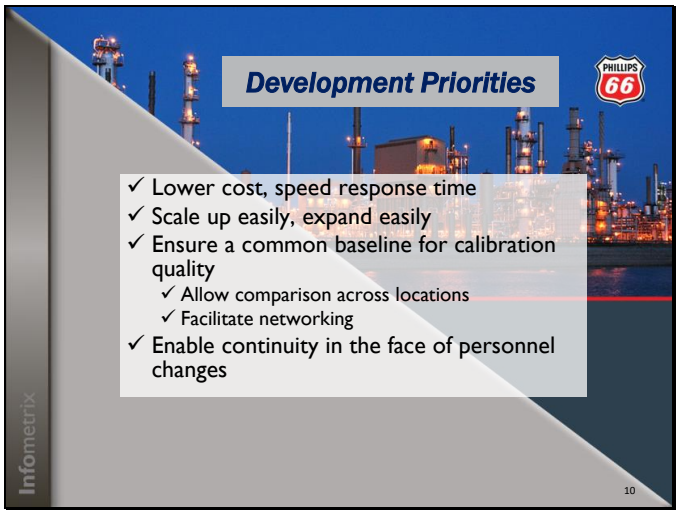
When a user considers the calibration of optical spectroscopy, the first view is of a traditional calibration, akin to tuning any instrument for its application assignment. We start with the known, build a method for handling the data, and use that calibration model to assess new samples. This is how analytical chemistry works.



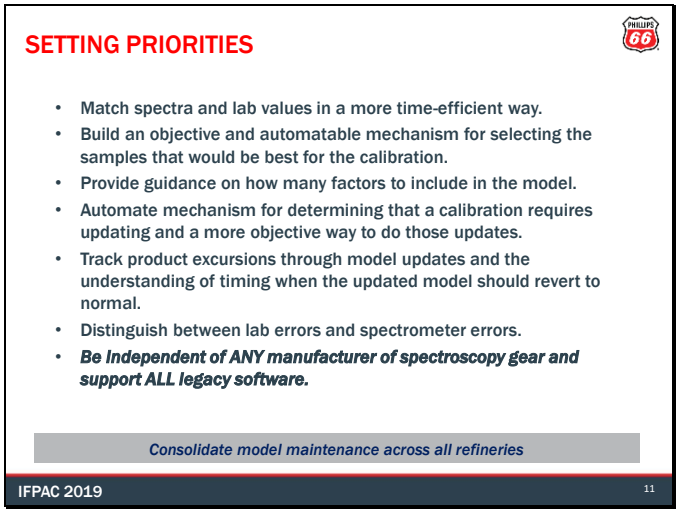
Customers have a lot more interest in extracting the information content of the new samples than in building and maintaining the calibration. This is as it should be, but then we must consider, particularly in applications where the calibration is subject to shifts, that the ultimate quality of our routine assessments is governed by the calibration's relevance to whichever sample happens to be in the spotlight.



Switching to the Infometrix perspective, we add a few twists to the two-bubble approach and find ourselves delving into details that we hope will improve model performance and robustness. We think of it as a model lifecycle and want to determine how the computer can help streamline and simplify the process.



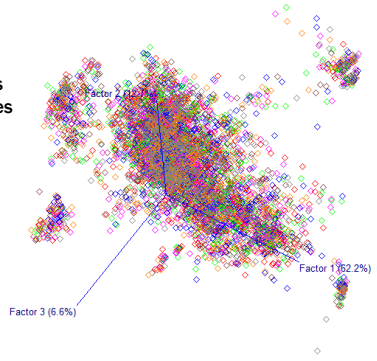
Both developer and end user really share the same goals. We want to develop and deploy systems that are both useful and used. We also need to be concerned with the outside factors that influence how the system should be structured, most notably the issues of training, how help is accessed, and the concern that accompanies personnel turnover.



The end user wants to perform his job in an efficient manner and avoid any mistakes that would lead to slower response or less-capable function. A priority list resulted from consideration of the tasks at hand, not just in consideration of one location's support needs, but how the system might flow across a company like Phillips 66. We experience several mind-numbing tasks that have a high potential for errors. We believe we model reasonably well, but we really have no way of knowing. We sometimes experience difficulties in preserving data and the models we generate and rarely track (never track?) the chemistry reason behind model updates. We have several different instrument types and different assessment software thoroughly plumbed into our systems; how is that legacy best preserved (or how easy is it to replace if we need to do so)?

GASOLINE QUALITY CONTROL IS COMPLEX

6,000 spectra spanning grades and blend recipes

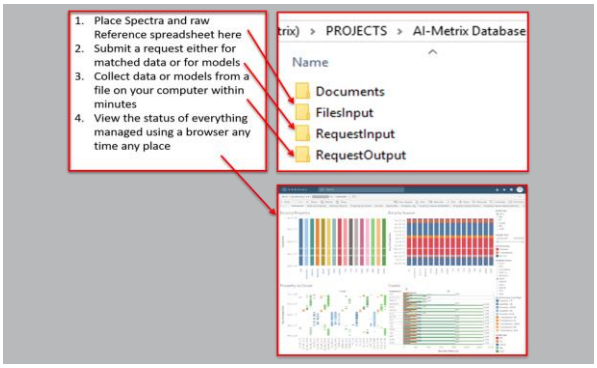


3 PCs capture >80% of the variance

IFPAC 2019 12

Gasoline blending is necessarily a dynamic process. This PCA plot, where each point represents the spectrum of a unique point in time, displays both the trends and the clustering of the data that is related to blend recipe and grade. The structure is there; we need to ensure we can make the most of the calibration opportunity.

PROCEDURE OVERVIEW

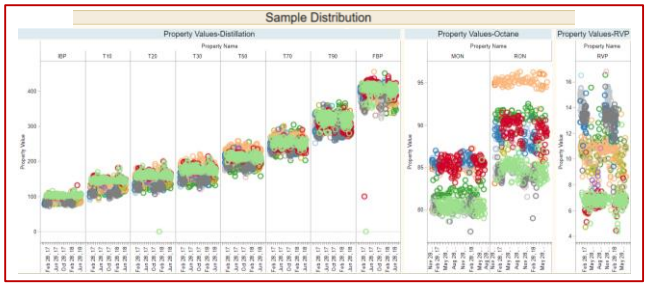


Streamlining and simplifying the user experience.

IFPAC 2019 13

It is up to the customer to muster the data, but in this system, the raw files are simply dropped into a folder structure, where a SQL Server database engine matches spectra to their corresponding lab values and forms a useful data repository. Visualization is updated within minutes in a web-based, customizable dashboard.

DATA ORGANIZATION

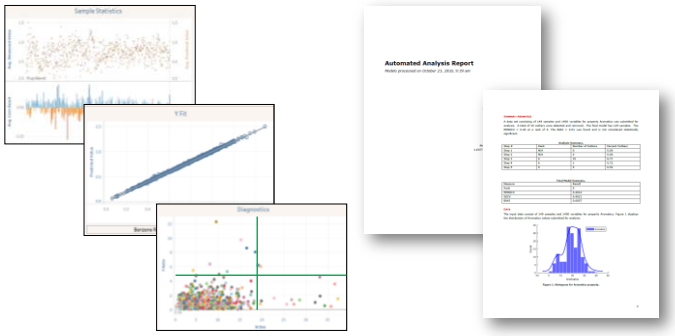


Real-time updating of all available data is critical.

IFPAC 2019 14

Custom views of the data can be color-coded by blend and anomalies are more apparent than they would be by scanning an Excel spreadsheet. These data views help the user decide how to construct a request for new models.

MODEL DIAGNOSTICS AND REPORTS

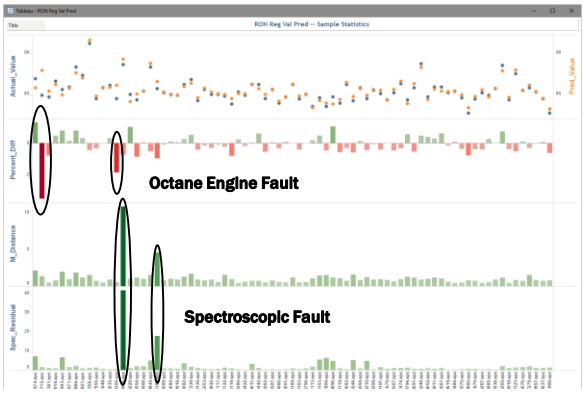


We want to be able to generate custom and industry-standard reports.

IFPAC 2019 15

When a set of models are requested, a five-step model construction process is triggered using a bundle of technologies that Infometrix has optimized over the past 6 years (and based on 41 years of total experience in this field). The resulting model is characterized both in an on-line dashboard and in a report delivered to the desktop. The speed of model construction is a function of the number of samples in the portion of the database used and the number of models requested. Typically, this process is measured in minutes.

USING THE DIAGNOSTICS



IFPAC 2019 16

We can mine these data and their resulting models for myriad bits of knowledge. Here we are looking to distinguish between lab errors at the top and "spectroscopic errors" (stemming from the model not encompassing the spectrum that was flagged or an issue with the spectrometer itself).

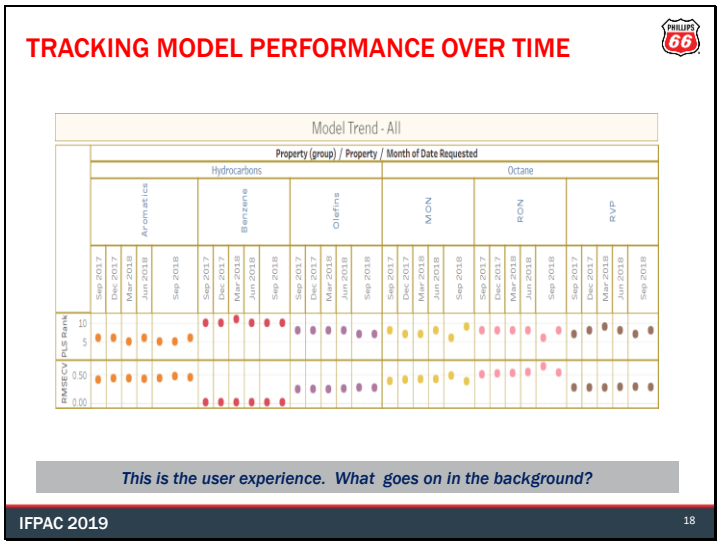
HOW GOOD ARE THE MODELS?

	Automated	Manual
AllBOB_T20	1.752	1.666
Universal_RONN	0.349	0.366
SUBSummer_RONN	0.262	0.266
RBOBWinter_RONN	0.374	0.364
RBOBSummer_T50N	1.454	1.574

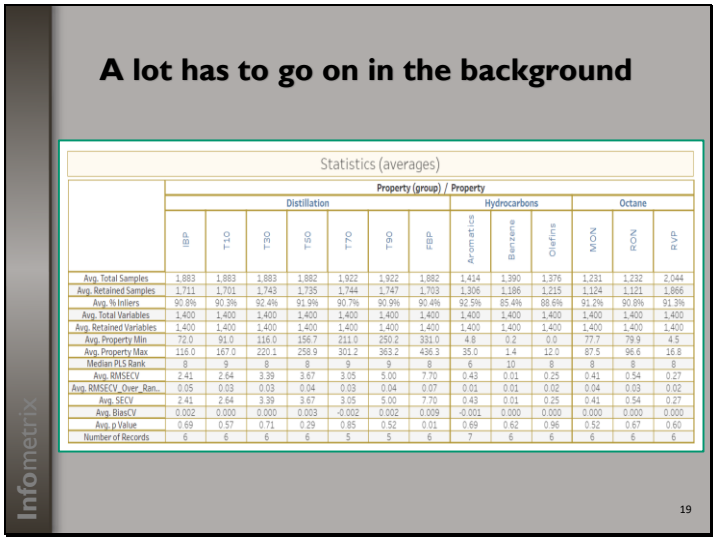
There is no statistically-significant difference between the automatically-generated models and those built by an experienced chemometrician.

IFPAC 2019 17

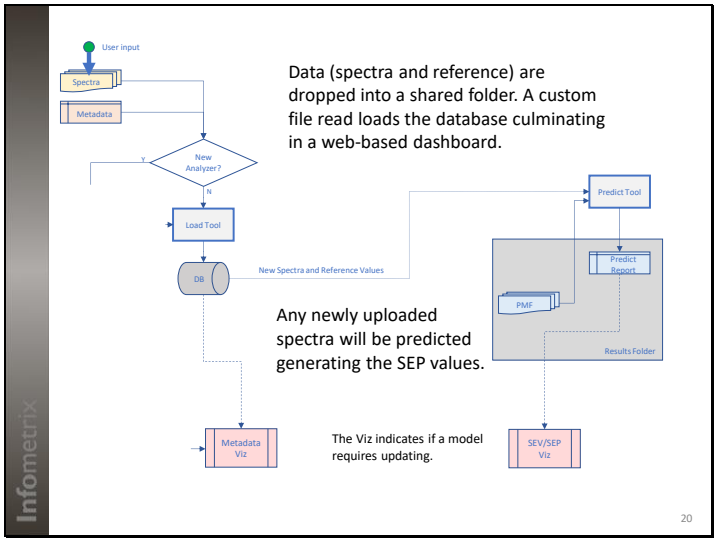
Here is a comparison of a few models built by an experienced PhD-level chemometrician and the automated process that was built. There is not statistically-significant difference in these data or any of the 88 models so tested.



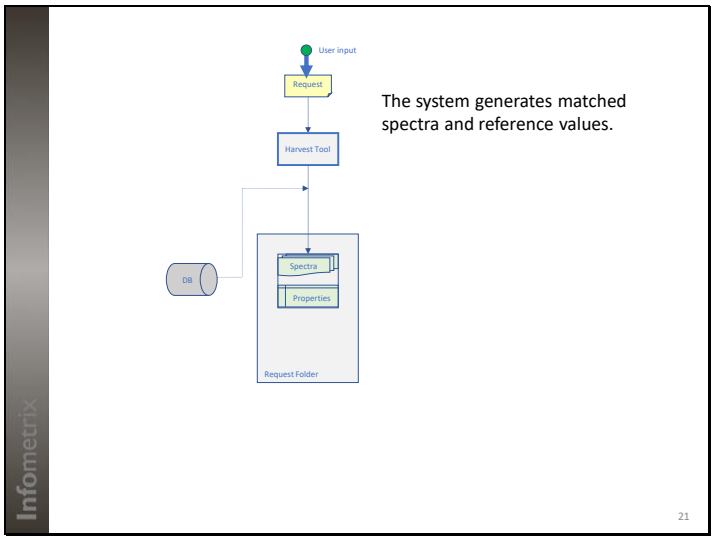
A database shows its potential for better management in that everything is contained in one place and comparisons can be made against history instantly.



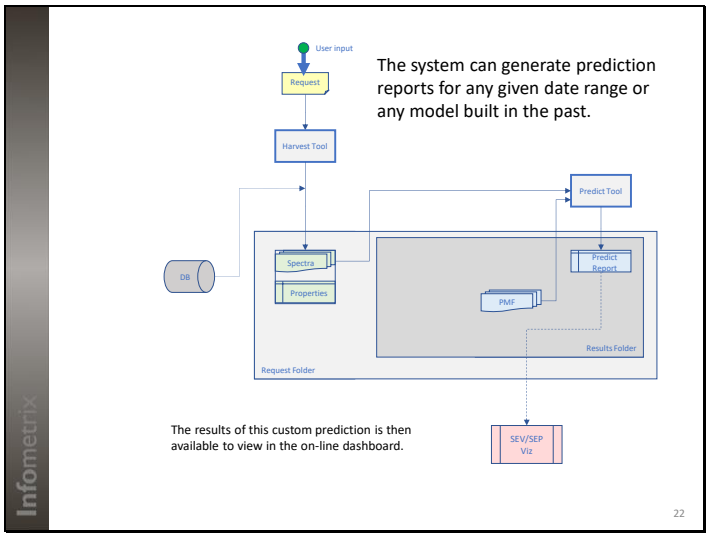
Of course, a lot of diagnostics are recorded when a model is built; a few are shown here. These tables are viewable in the dashboard, are immediately available, and can be customized to suit.



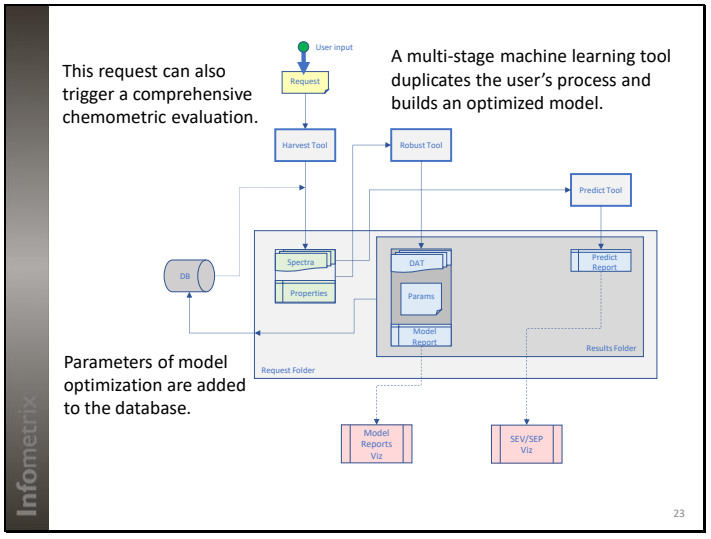
This is a breakdown of the processing done to set up the SQL database and the visualizations seen during the loading of data. Once built, any new data is shipped to the on-line viz but also predicted using the most recently-generated model. This tells us when a model requires updating.



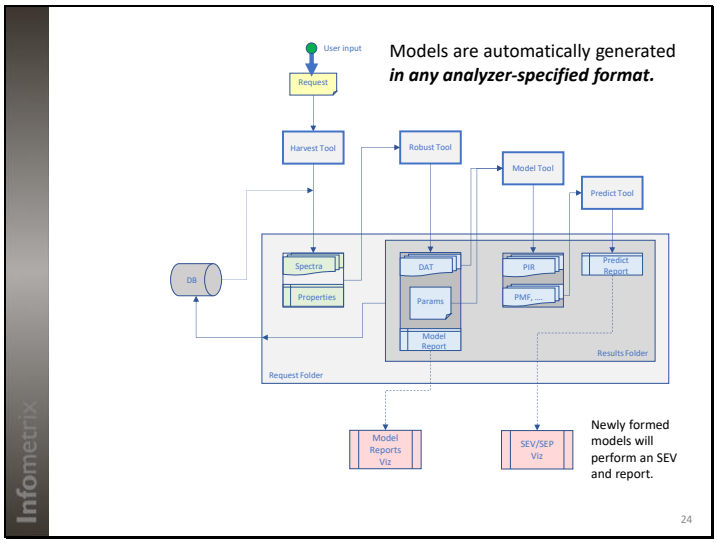
The second user process is to make a request on the database. These requests can be to simply extract the matched lab and spectra from the database given the constraints set by the user.



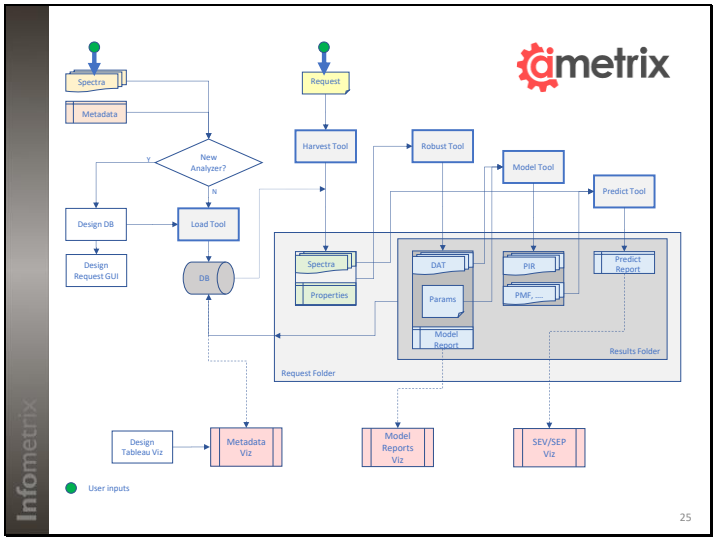
Because data and models coexist in a managed way, the user can specify a data range and apply any current or historical set of models to run custom predictions. These are shipped to the on-line viz.



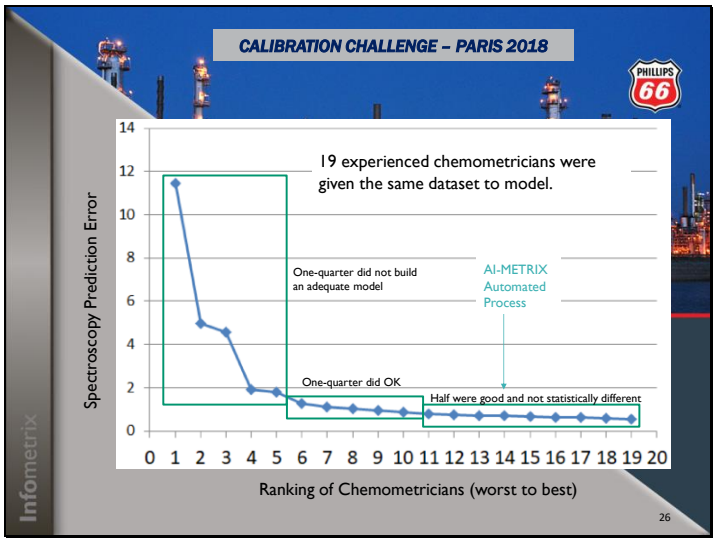
Typically, this route will be used for generating new models. The data to be used is specified as before, but here we ship data to a Robust model optimization tool that purges outliers that degrade the model performance and determine the model complexity or rank. The results are shown on-line and encapsulated into a report. These parameters are also shipped back to the database for safekeeping.



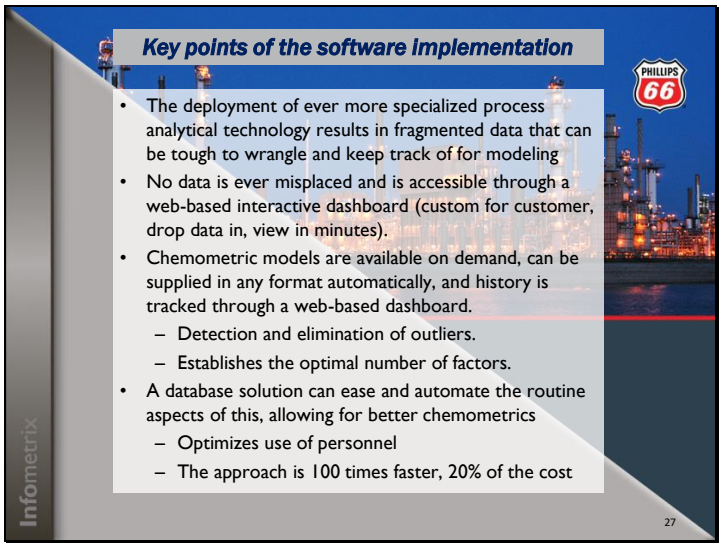
Once the model parameters are known, we capture the data and build the model in Pirouette, then run the new data to determine the quality both as an SECV (self-diagnosis) and, if a validation set has been specified, as a SEP. If a non-Pirouette model is required, it is built in this step.



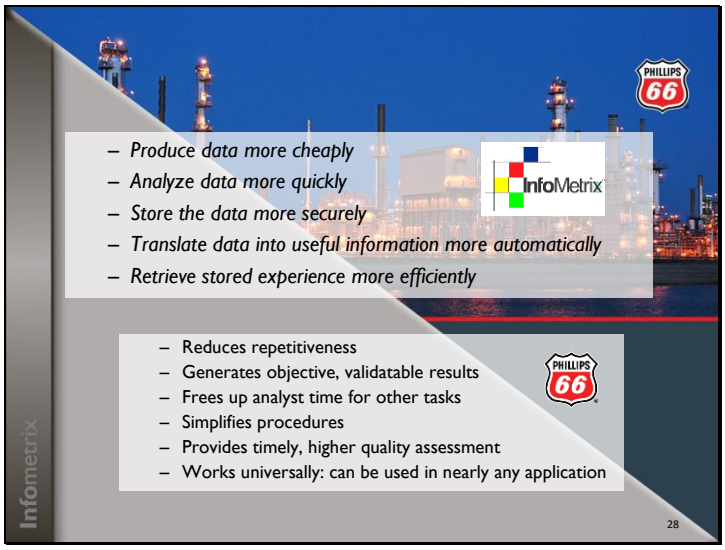
Here is the whole AI-Metrix process. The five Tools are what comprise the backbone of the assembly and processing engine. It uses commercially-available (and thus maintained) tools to step through the process. The calibration engine uses the best practices we have gleaned over the past years in optimizing 16 refinery case studies.



Recalling the issue that started this discussion. Actually, 18 chemometricians were in on this study, the 19th captured the AI-Metrix automated system approach.



We are really forced to reconsider the historical approach of manually generating calibration models. We have more spectrometers in place and more places where we project an advantage of seeing more real-time aspects of the chemistry in process. The paradigm must change, and this approach is objective, accurate, faster, and more cost-effective.



Infometrix has a mantra that outlines an overarching philosophy about handling data and the information content they have hidden away. From the Phillips 66 perspective, we recognize that approaching a calibration procedure in a consistent and transferable manner will reap substantial benefits, particularly as we factor in the variability in experience for those we task with model maintenance. If you purchase a car, you want to maintain that asset in the best possible way. We have an enormous investment in spectrometers, installation, and people; we should do no less. Interestingly, in this case, the best is faster and cheaper.