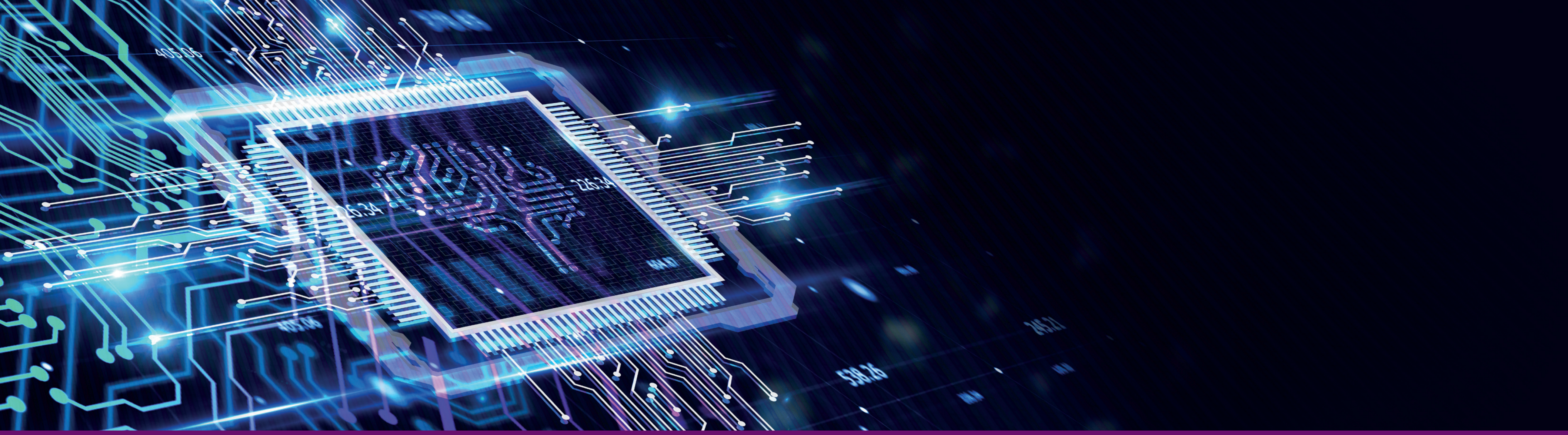


STREAMLINING THE USE OF AI/MACHINE LEARNING IN THE CHEMICAL INDUSTRY: CHEMOMETRICS

Abstract

Artificial intelligence and machine learning are inevitable results of the work driven by the consumer side of our economy. The question is not whether it will impact refining and chemical plant operation, but how soon and how long it will take for the benefits to outstrip the costs. The goal is to distinguish between vision and hallucination and to provide some practical guidance for making progress in this complicated set of fields. There are three categories of measurements that provide us with the data that will form the basis for any interpretation system: single-purpose sensors, chromatographs, and spectrometers. Chemometrics can be used in all three categories and, in fact, is critical to interpreting output from any type of spectrometer. We can easily demonstrate that the use of multivariate analysis for each of the three data sources, taken individually or assembled together, gives faster response, improved flow of information derived from these data, and a significant leg up for process understanding. This information is available and is nearly cost-free.

Introduction

As computers came on-line four decades ago, automated quality control became practical, but we still don't take advantage of that processing capability, instead sticking with engineering heuristics and rendering most of the available data impotent for control. This applies equally to the monitoring of process variables and to the analyzer population. In the process sensor world, ALL measurements could be combined to form a virtual instrument of a unit or even the entire plant, isolating the process-relevant signal from the noise. The advantage is that previously-unseen process upsets can be identified and managed in close-to-real time. Spectroscopic and chromatographic analyzers have enjoyed a more intimate relationship with the computer, but we have not yet tapped into the potential of processing raw data into actionable process information. The key to success is to NOT focus solely on extracting the information content of a stream of data using multivariate analytics. More critical is the blending of these results with application-specific product knowledge. Even that is not enough; all is lost if we do not focus on the effective delivery of that blended information content. Tools abound from multiple sources that can assist in achieving a proper blend, but the successful engineer needs to tackle the task with a systems engineering approach. Any integration must be both useful and used.

In the literature and in the media, there is terminology confusion that pervades any discussion of the use multivariate analysis to manage information gathering and performance prediction. AI, machine learning and chemometrics represent overlapping fields of study and do not in and of themselves dictate a path to follow in getting to understand what your data is telling you, let alone facilitate the building of useful process models. I choose the term chemometrics in that it blends the thought of chemistry understanding with the math we use, distilling our focus on industries like those tied to hydrocarbon processing.

Regardless of the label in use, the goal is to use mathematical tools to assess quality in products in a cost effective and reliable manner. In these examples, I use Infometrix software, but there are myriad options, particularly for process variables where there is significant overlap with the statistics community. Let's look at three areas where nearly instant financial benefits are available through a simple application of chemometrics.

Process variables

Multivariate data analysis, as powerful as it may be, can be daunting to the uninitiated. It has its own jargon and incorporates concepts which at first glance seem strange. Typically, a novice user becomes comfortable with the multivariate approach only after a period of confusion and frustration. Why bother? The rationale behind multivariate data analysis is simple: univariate methods, while well–understood and proven for many applications, can sometimes produce misleading results and overlook meaningful information in complex data sets.

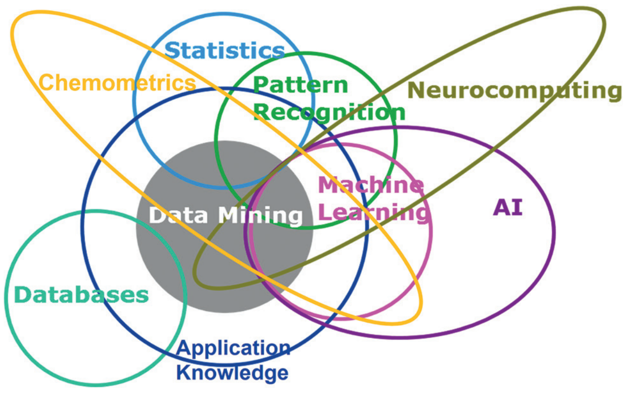


Figure 1: The terms used to label discussions in multivariate data processing are not distinct; overlap abounds.

Univariate methods were developed for univariate data. Applying univariate methods to process measurements may be useful but it is also tantamount to discarding all but one of the measurements. While some problems may yield to a thorough statistical analysis of a single variable, this approach has several drawbacks when applied to multivariate data. Basically, it is incomplete if multivariable relationships are important. If the measurements are numerous and/or correlated, processing them as a unit is advised. Chemometric techniques are natural garbage collectors as the data processing does not easily get hijacked by erratic sources of data. Applying first principals to guide the multivariate process models is critical to long-term success.

Most of the sensors in play in a process setting are pressure, temperature, flow, and level monitors that can be considered together, melding them into a custom process instrument. We can use one of two techniques to tease the information content from the mix: Principal Component Analysis helps us classify the state of the process (e.g., is there a process upset?); and Partial Least Squares allows us to predict performance measures (such as yield). These are the tools to use for tuning the results to suit the process at hand and for automating the routine evaluation of the data stream.

Chromatography

In the hydrocarbon processing industry, chromatography is the traditional, go-to technology for monitoring chemical composition. It is the most direct way to measure the concentration of individual molecules. With the current generation of fast, highly capable GCs and the advent of process ultra-high-performance liquid chromatographs, analyses can be completed at a rate commensurate with optical spectroscopy.

For a variety of reasons, processing chromatographic data using multivariate techniques has been largely neglected over the past few decades. There are two multivariate technologies necessary for automated processing of any-but-the-simplest chromatographic applications: a signal processing step to align the chromatographic traces and, as demonstrated for the process variables approach described in the previous section, a mining of the chromatographic trace (or tabular results) for full information content.

Spectroscopy

Chemometrics has seen much use in the management of optical spectroscopy data collected for monitoring a chemical process, and is routinely integrated into the analytical workflow both to improve the signal and to process the signature into useful quantitative and qualitative information. But, the typical employment in process systems is in need of an upgrade.

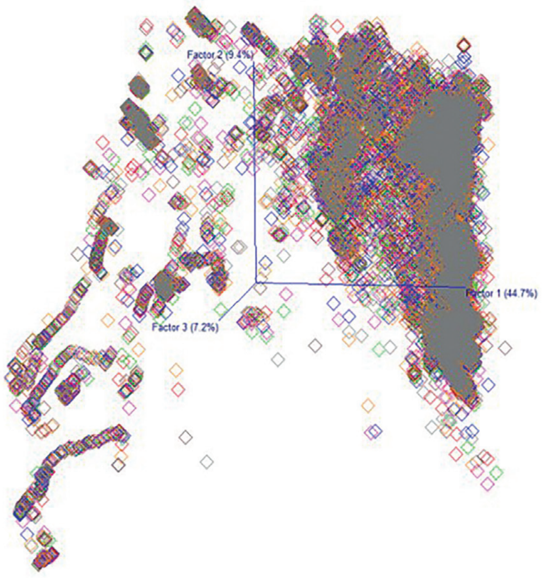


Figure 2: Process variables turn into inferentials and here clearly show process conditions that are normal (right) and show upset conditions (left). Principal Component Analysis distills all data into a single plot showing when the process is in control. Each datapoint represents the signature of the process for a specific time; the closer the points, the more similar the process conditions.

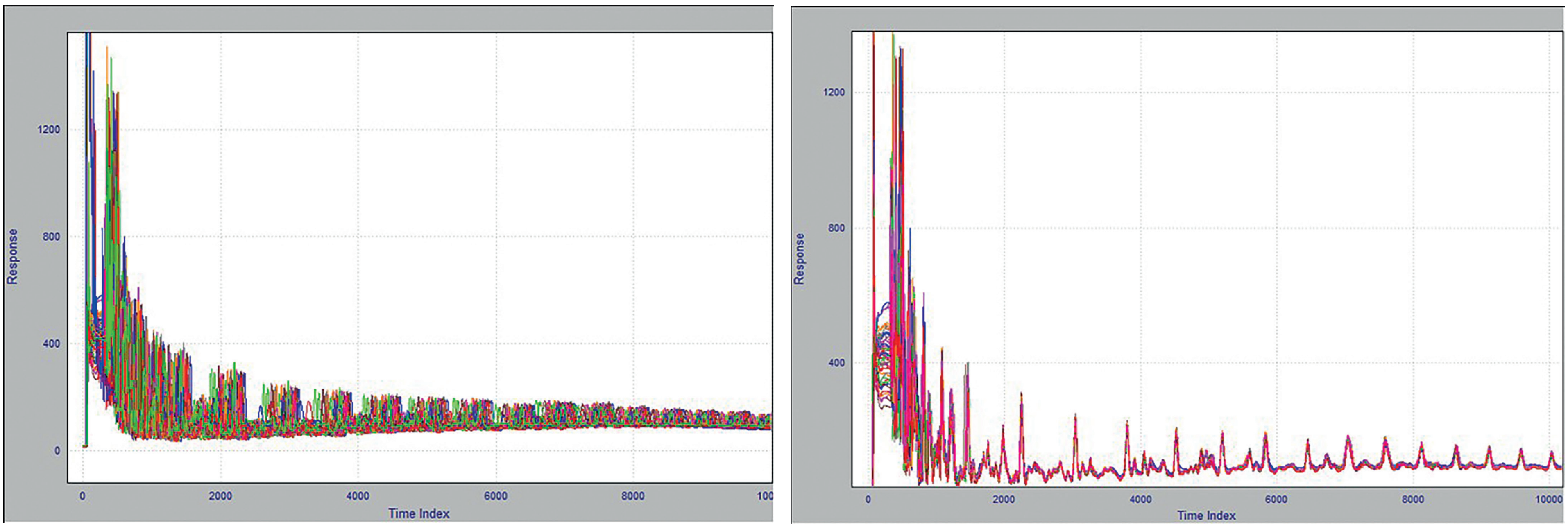


Figure 3: Repeat injections over 3 years for a reference oil standard: original files on the left and the same data chemometrics-corrected on the right. The use of correlation optimized warping to correct for retention time variability, here processed by a commercial package called LineUp from Infometrix.

Multivariate calibrations of instrument systems are performed when the instrument is first set up, but often suffer as either the instrument quality drifts, or the process conditions change.

We did this work better 20 years ago than we do today. Now, the computed result is monitored but there is little or no considerations of any of the quality diagnostics inherent in the processing. This is exacerbated by the lack of training: apparently an expense and a time commitment to be avoided.

There have been advances in chemometric processing and with some simple concepts integrated into the communication loop, the calibration procedures can be streamlined and automated. It is possible to integrate spectroscopic measurements with nearly the same simplicity as the effort with which we put simple temperature sensors in place.

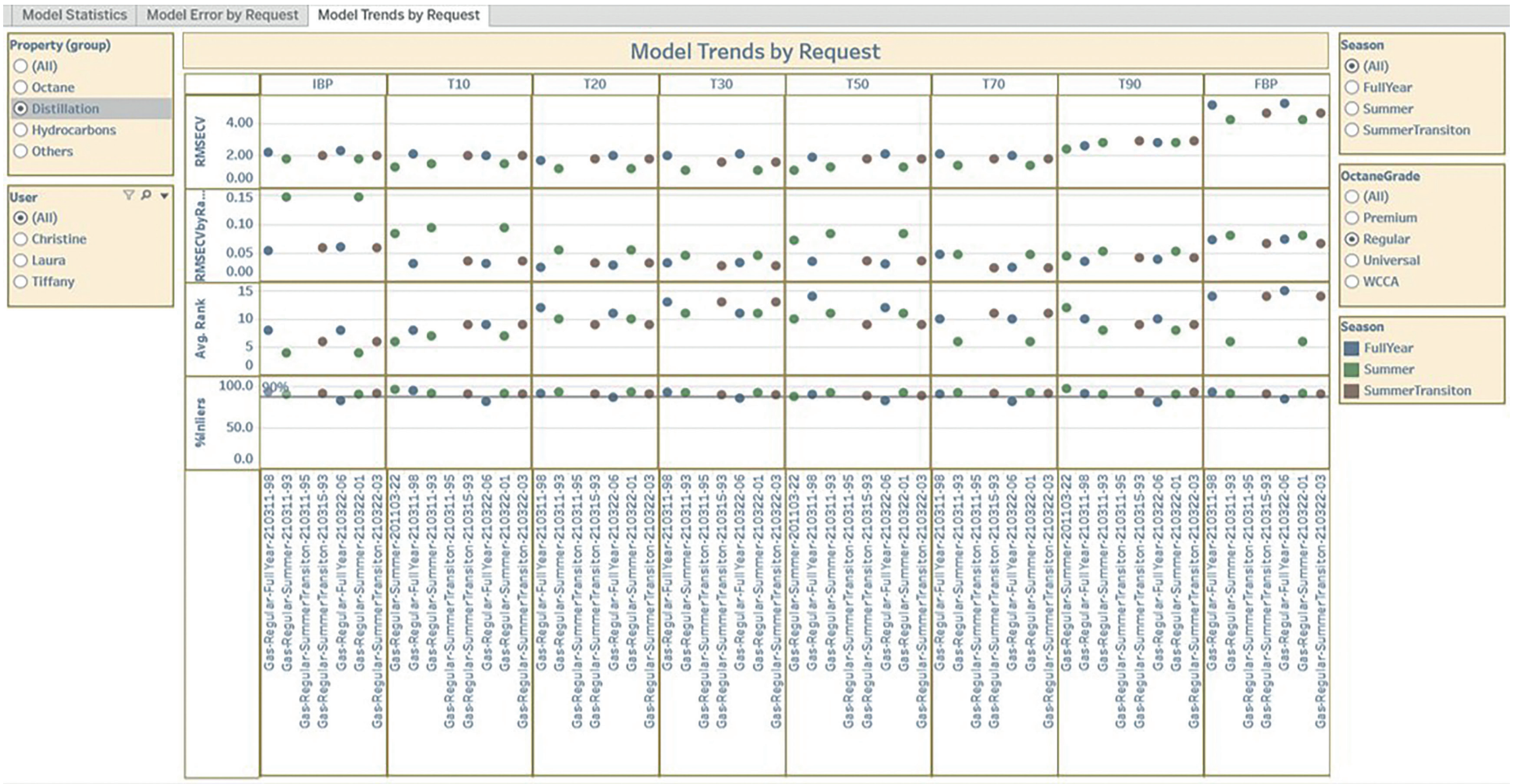
Conclusions

The oft-quoted statement "you can't control what you don't measure" leads easily to "the quality of your measurement dictates the quality of your control". We can further break down measurement quality into two sub-parts: 1) the precision and specificity of the measurement; and 2) the system we employ to extract the information content from these data. In the hydrocarbon processing industry, we have three primary sources of data to utilize, with the collection of univariate sensors (typically temperature, pressure, flow, and level), spectrometers (mostly optical), and chromatographs (mostly GCs).

Each of these data source categories require a calibration step to enter the world of multivariate control.

- For a collection of process variables, we would expect the models to last a long time; most of the upset warnings will be tied to failure in individual sensors.
- In chromatography, much of the variability is solved using correlation-based chemometric alignment, giving these analyzers vastly-lower calibration requirements and giving us the opportunity for instrument interchangeability and a common interpretive base instrument to instrument or even plant to plant.
- Even though spectral analysis has long been tied to multivariate analysis, few companies are taking advantage of tools that can simplify calibration, ease the burdens of maintenance, and improve model quality to be less dependent on different operator experience levels.

The discussion in this paper addressed the information extraction process and how leveraging chemometric analysis can evaluate the data more quickly and automatically. The output of this analysis is objective, can be validated, and is a universal approach. The benefit is a simplified procedure, freeing up analyst time for other tasks. We need to invest in new technologies if we hope to make this all work; we cannot do things like we used to. Think of how many analyzers we deploy that have identical function; we need to manage them in a global way. Chemometric tools are off-the-shelf that facilitate this management task and will work with most computer-sentient legacy equipment. That really does make the transition to multivariate essentially free and takes us a major step beyond the current analyzer maintenance "IIoT" focus on simply digitizing paper control charts. It also enables a streamlining and simplification of the calibration process and provides an objective, automatable evaluation system.


Model Statistics
Model Error by Request
Model Trends by Request
Model Trends by Request
IBP
T10
T20
T30
T50
T70
T90
FBP
RMSECV
4.00
2.00
0.00
RMSECVbyRa...
0.15
0.10
0.05
0.00
Avg. Rank
15
10
5
0
%Inliers
100.0
90%
50.0
0.0
Gas-Regular-FullYear-210311-98
Gas-Regular-Summer-210311-93
Gas-Regular-SummerTransiton-210311-95
Gas-Regular-SummerTransiton-210315-93
Gas-Regular-FullYear-210322-06
Gas-Regular-Summer-210322-01
Gas-Regular-SummerTransiton-210322-03
Gas-Regular-Summer-201103-22
Gas-Regular-FullYear-210311-98
Gas-Regular-Summer-210311-93
Gas-Regular-SummerTransiton-210311-95
Gas-Regular-SummerTransiton-210315-93
Gas-Regular-FullYear-210322-06
Gas-Regular-Summer-210322-01
Gas-Regular-SummerTransiton-210322-03
Gas-Regular-FullYear-210311-98
Gas-Regular-Summer-210311-93
Gas-Regular-SummerTransiton-210311-95
Gas-Regular-SummerTransiton-210315-93
Gas-Regular-FullYear-210322-06
Gas-Regular-Summer-210322-01
Gas-Regular-SummerTransiton-210322-03
Gas-Regular-FullYear-210311-98
Gas-Regular-Summer-210311-93
Gas-Regular-SummerTransiton-210311-95
Gas-Regular-SummerTransiton-210315-93
Gas-Regular-FullYear-210322-06
Gas-Regular-Summer-210322-01
Gas-Regular-SummerTransiton-210322-03
Gas-Regular-Summer-201103-22
Gas-Regular-FullYear-210311-98
Gas-Regular-Summer-210311-93
Gas-Regular-SummerTransiton-210311-95
Gas-Regular-SummerTransiton-210315-93
Gas-Regular-FullYear-210322-06
Gas-Regular-Summer-210322-01
Gas-Regular-SummerTransiton-210322-03
Gas-Regular-FullYear-210311-98
Gas-Regular-Summer-210311-93
Gas-Regular-SummerTransiton-210311-95
Gas-Regular-SummerTransiton-210315-93
Gas-Regular-FullYear-210322-06
Gas-Regular-Summer-210322-01
Gas-Regular-SummerTransiton-210322-03
Gas-Regular-Summer-201103-22
Gas-Regular-FullYear-210311-98
Gas-Regular-Summer-210311-93
Gas-Regular-SummerTransiton-210311-95
Gas-Regular-SummerTransiton-210315-93
Gas-Regular-FullYear-210322-06
Gas-Regular-Summer-210322-01
Gas-Regular-SummerTransiton-210322-03
Gas-Regular-FullYear-210311-98
Gas-Regular-Summer-210311-93
Gas-Regular-SummerTransiton-210311-95
Gas-Regular-SummerTransiton-210315-93
Gas-Regular-FullYear-210322-06
Gas-Regular-Summer-210322-01
Gas-Regular-SummerTransiton-210322-03

Property (group)
○ (All)
○ Octane
● Distillation
○ Hydrocarbons
○ Others

User
● (All)
○ Christine
○ Laura
○ Tiffany

Season
● (All)
○ FullYear
○ Summer
○ SummerTransiton

OctaneGrade
○ (All)
○ Premium
● Regular
○ Universal
○ WCCA

Season
🟦 FullYear
🟩 Summer
🟫 SummerTransiton

Figure 4: Dashboards (in this case using Tableau On-Line) can provide visualization techniques to track all data and all process models across the organization in one place. Combined with automated and optimized creation of calibration models, here using Infometrix' Ai-Metrix system, true quality control of the quality control process can be achieved.

Author Contact Details

Brian Rohrback - President, Infometrix, Inc. • 11807 North Creek Parkway S, Suite B-11, Bothell, WA 98011, USA • Tel +1 425 402-1450 x118

• Email: brian_rohrback@infometrix.com • **Web: www.infometrix.com**