**Thermo Fisher**
**SCIENTIFIC**



Raman spectroscopy

# Probabilistic evaluation – first principles vs. empirical
## Fundamentals and performance characterization of the Thermo Scientific™ TruScan™ Handheld Raman Analyzers

**Keywords**
Raman spectroscopy, Thermo Scientific TruScan G3 Handheld Raman Analyzers, method development, statistical analysis

In most direct terms, the Thermo Scientific™ TruScan™ Handheld Raman Analyzers not only acquires the Raman spectrum of a material of interest, but also determines – in real time – the uncertainty of that measurement. Uncertainty here is defined as how consistent and reliable we expect measured spectrum to be in similar or dissimilar sampling conditions or in terms more common on a smaller-scale "standard deviation." The TruScan handheld analyzer is designed to account for all measurements of intrinsic variability factors, including sample characteristics, instrument telemetry, environment, and others.

The chief sources of uncertainty in Raman can be determined from first principles[1,2,3], and the optical and electrical performance of the system over a range of conditions. Thermo Scientific engineers and chemometricians spent a great deal of time on system design – specifying terms, accounting for how they affect measurements and recognizing when they dominate. What might sound rather complicated is simply propagation of error (e.g. adding up all the individual terms of uncertainty terms) into generalized estimations of uncertainty. With a Raman spectrum acquired and the multivariate estimate of its uncertainty determined, the TruScan handheld analyzer has the statistical measures necessary to perform an objective and statistically relevant comparison of the measured data to any reference spectrum of a material, resulting in a multivariate test of equivalence.



**The Thermo Scientific TruScan™ G3 Handheld Raman Analyzer**

thermo scientific

When the TruScan handheld analyzer begins to acquire a spectrum, it is tracking terms such as read and shot noise and calibrating fixed terms. In real time, it collects one spectrum and starts propagating all error terms for each measurement along that spectrum. By the first accumulation, the system has determined at the CCD pixel level an estimate of the uncertainty of each measurement.

Since TruScan models calculate uncertainty directly, there is no calibration or user-modeling involved with method development. A single reference spectrum typically suffices for method development with bulk materials, as physical properties of the sample have minimal influence on the Raman spectra. The remaining sources of variability are modeled directly by the embedded analysis. A high quality (high signal-to-noise ratio) reference spectrum is prerequisite and much higher in signal-to-noise ratio (SNR) than what is acquired for a sample analysis spectrum. Calculations implicitly assume that the uncertainty in the analysis measurement dominates the uncertainty in the reference spectrum.

All TruScan handheld analyzer scans -- both the reference and analysis measurement – are measured to fixed SNR values, not a predetermined fixed-time setting. In Raman spectroscopy, each chemical can differ quite substantially in the amount of shifted light (e.g., Raman scatter) returned to the detector independent of the scanning conditions (e.g., chemical not physical or environmental property). Figure 1 provides an example of two chemical scans both acquired for the same length of time (50ms).

One will observe in Figure 1, the quality (SNR) between the two chemicals is quite significant. Therefore why would one want to use the same sampling conditions (e.g., total number of scans and/or length of time) for all samples of various chemistries? The answer is one should not, as the results of any analysis based on those
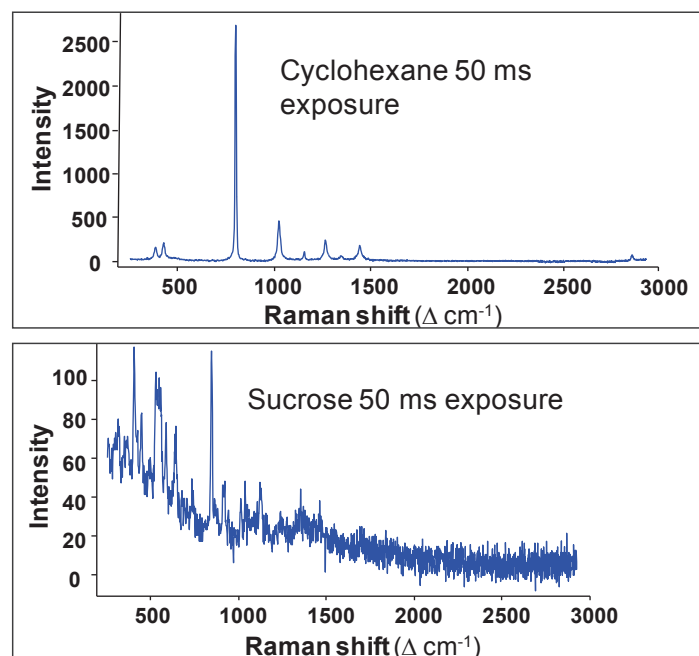
scans would suffer significantly. Any system set to a fixed time will either scan too long for materials of high Raman activity and not long enough for those of low activity. Those not scanned long enough increase the rate of both false-positive or false-negative results. Fixed-scan length also never adjusts to environmental influences affecting the scan quality in real-time (Figure 2). We designed the TruScan handheld analyzer to address this concern. The TruScan handheld analyzer system monitors and adjusts for ambient light, such as sunlight and room light, because handheld units are meant to be deployed in continuously changing environments.
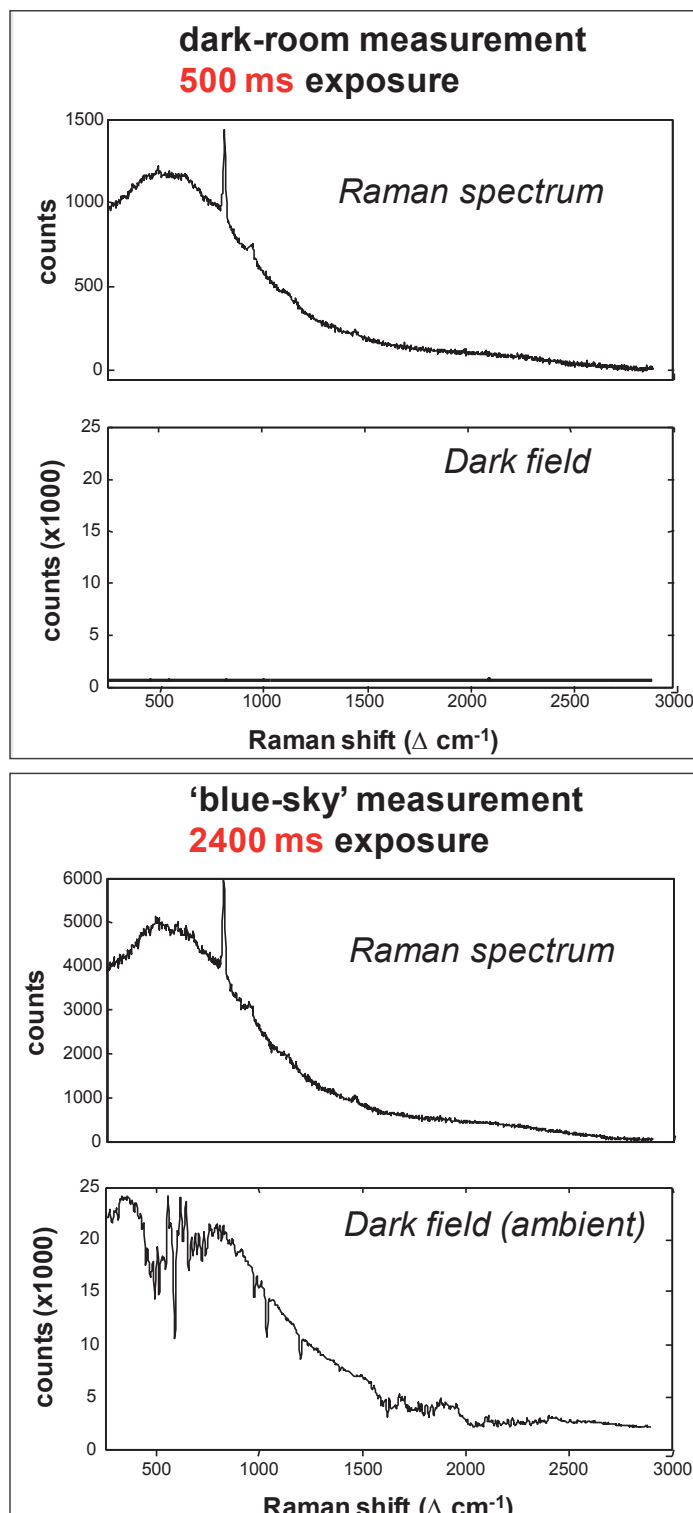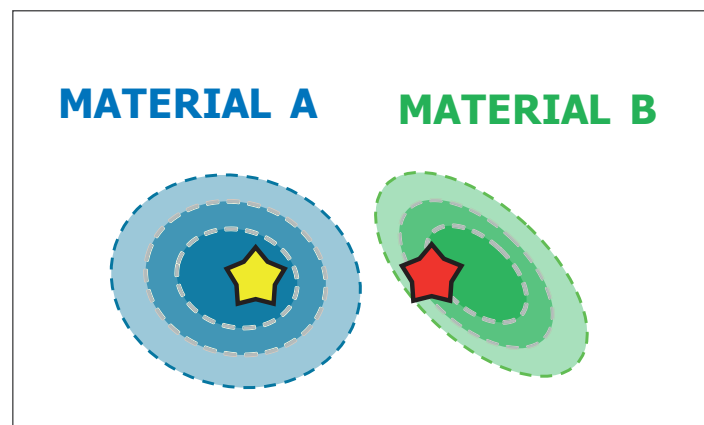


Figure 2. Environmental Factors influencing on measurements.



Figure 1. Fixed measurement time signal quality greatly influenced by material.

Thermo Scientific TruScan Raman systems continue to collect data and measure sources of uncertainty until they arrive at an SNR threshold sufficient for chemical identification. The null hypothesis is claiming that a measurement spectrum belongs to the population of the reference library spectrum, given the measurement uncertainty. The alternative hypothesis is claiming that measurement spectrum does not belong to the population of the reference library spectrum.
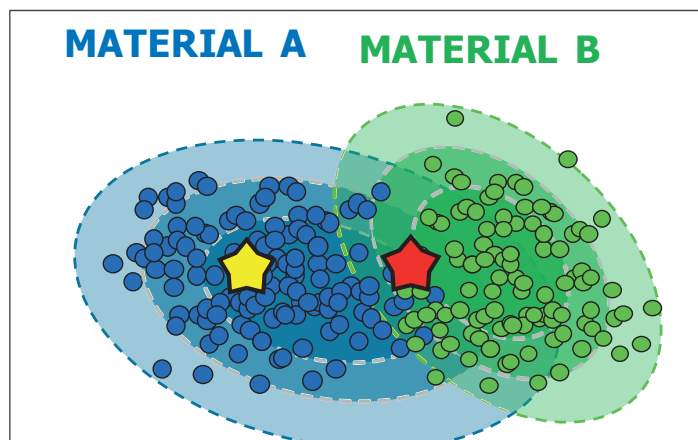
Thus, p-value is the probability of observing a spectrum more extreme (worse) than the sample spectrum if the sample spectrum belongs to the population of library spectrum (i.e., when null hypothesis is true). A significant level of 0.05 is used here. Thus, if p-value is not less than 0.05, the measurement is considered consistent with the reference spectrum and the device will report PASS or Positive Match. Otherwise, the device will report FAIL or No Match, depending on the unit configuration.

One can think of this in terms of a massively multivariate version of a statistical t-test. This is also similar in spirit to the Hotelling statistics calculated on the F-value of the residuals for empirical chemometric models[4] such as Soft Independent Modelling of Class Analogy (SIMCA). The exceptions are that no low-rank latent variable modeling is performed, and, of course, that no empirical calibration is involved.

To determine uncertainty empirically via user calibration, the calibration phase is very resource intensive in terms of time and expertise to determine accurately, requiring replicates over multiple lots, preparations and wide-ranging sample and instrumental considerations. If the calibration set is too small, the mapped space of the method when challenged only with a single or few samples, will give the initial impression of selectivity between two materials. However, if the calibration set included the true expected variance, it would have demonstrated that the appearance of selectivity is false. Illustrated examples are provided in Figures 3 and 4. Figures 5 and 6 are based on an empirical SIMCA data model calculated on a population of Raman spectra (2 factor, principal component).
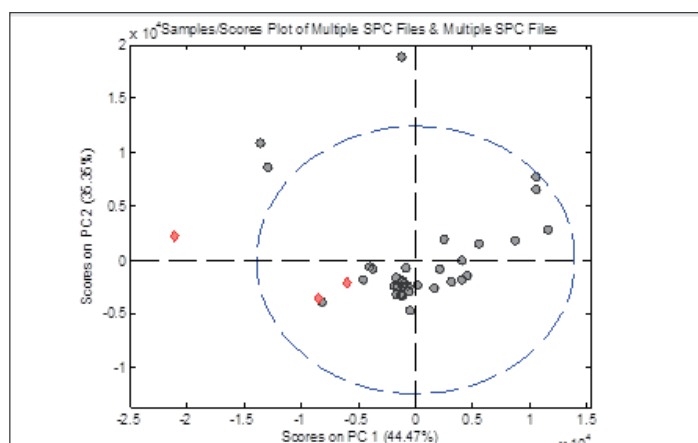


Figure 4. Multiple sample reference, multiple sample challenge sample test.
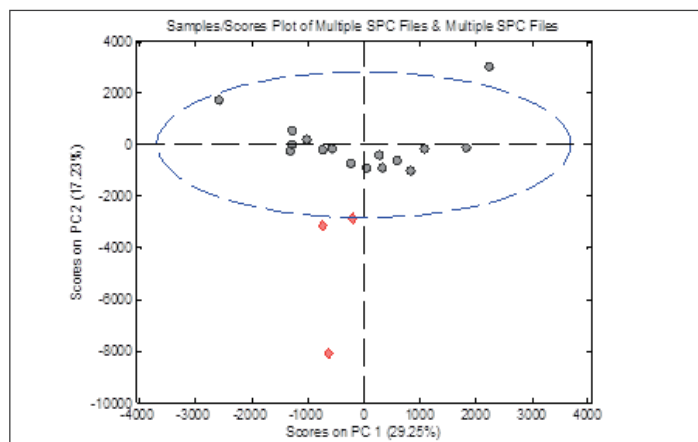


Figure 5. 1 sample reference, replicate scan calibration set (black), 3 sample challenge test set (red) with 95% variance ellipse (blue line).



Figure 6. Multiple sample, multiple scan reference calibration set (black), multiple sample challenge sample test set (red) with 95% variance ellipse (blue line).

No two instruments can be made with absolutely identical performance attributes. An empirical-based model that other Raman vendor instruments employ, if only scanned with one instrument, may not have successful calibration transfer when moved to a second instrument (Figure 7) or a third, or more.



Figure 3. 1 sample reference, 1 sample challenge sample test.
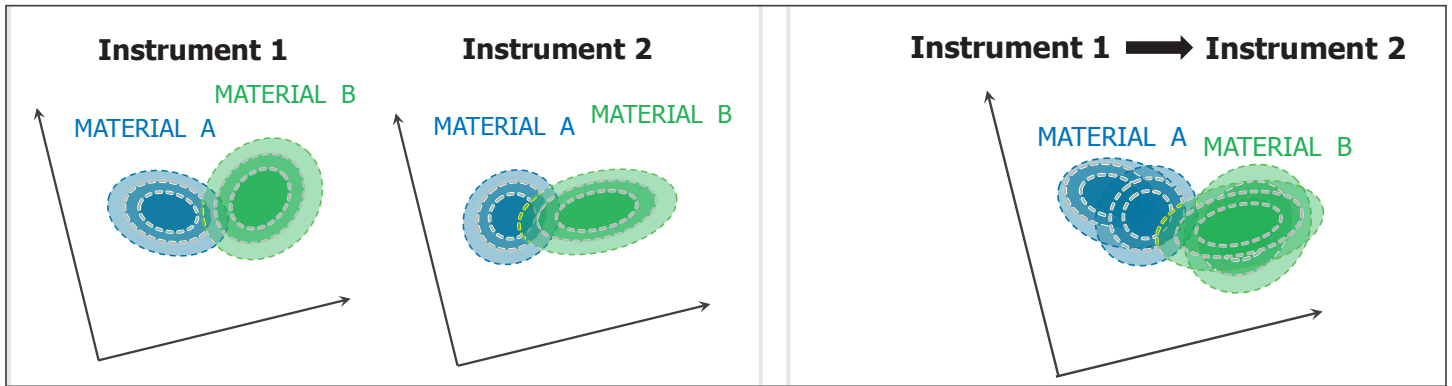
**Figure 7. Instrument to Instrument Variation Calibration Transfer Impact.**

The optical and electrical performances of the Thermo Scientific TruScan systems were rigorously tested on numerous devices over a range of conditions during device development. This performance was also assessed relative to the TruScan system set p-value threshold of 0.05. Another Raman systems' empirical approach places a huge burden on their users to build proper variance into the model to capture the design space, cover all legitimate variances and the p-value relative to those user determined variances, if the p-value is adjustable in their systems.

A common question that arises about the calculated p-values is, whether the reported values should in general be close to 1 for positive matches. Any other Raman system that reports p-values of 1 violates a basic law of statistics because all systems contain some probability of error and can only approach 0 or 1. This is why, if someone were using a statistical test where the p-values were always near 1, one should doubt the validity of the method.

Figure 8 below is a statistical distribution plot. On the top is the probability density function vs. distance and on the bottom is the p-value vs. distance. The distance is what is computed for each new scan to the center of the reference scan in multivariate space. One can see that to get a p-value close to 1, the distance must be exceptionally close to 0. This is difficult to achieve in practice. Random noise associated with a scan should make the distance away from zero and therefore lower the p-value. Random noise associated with a scan should move the distance away from zero and therefore, lower the p-value. Obtaining p-values in the 0.2 to 0.6 range are more likely and intuitive than obtaining values in the 0.8 to 1 range, if a valid method and statistical test is used.
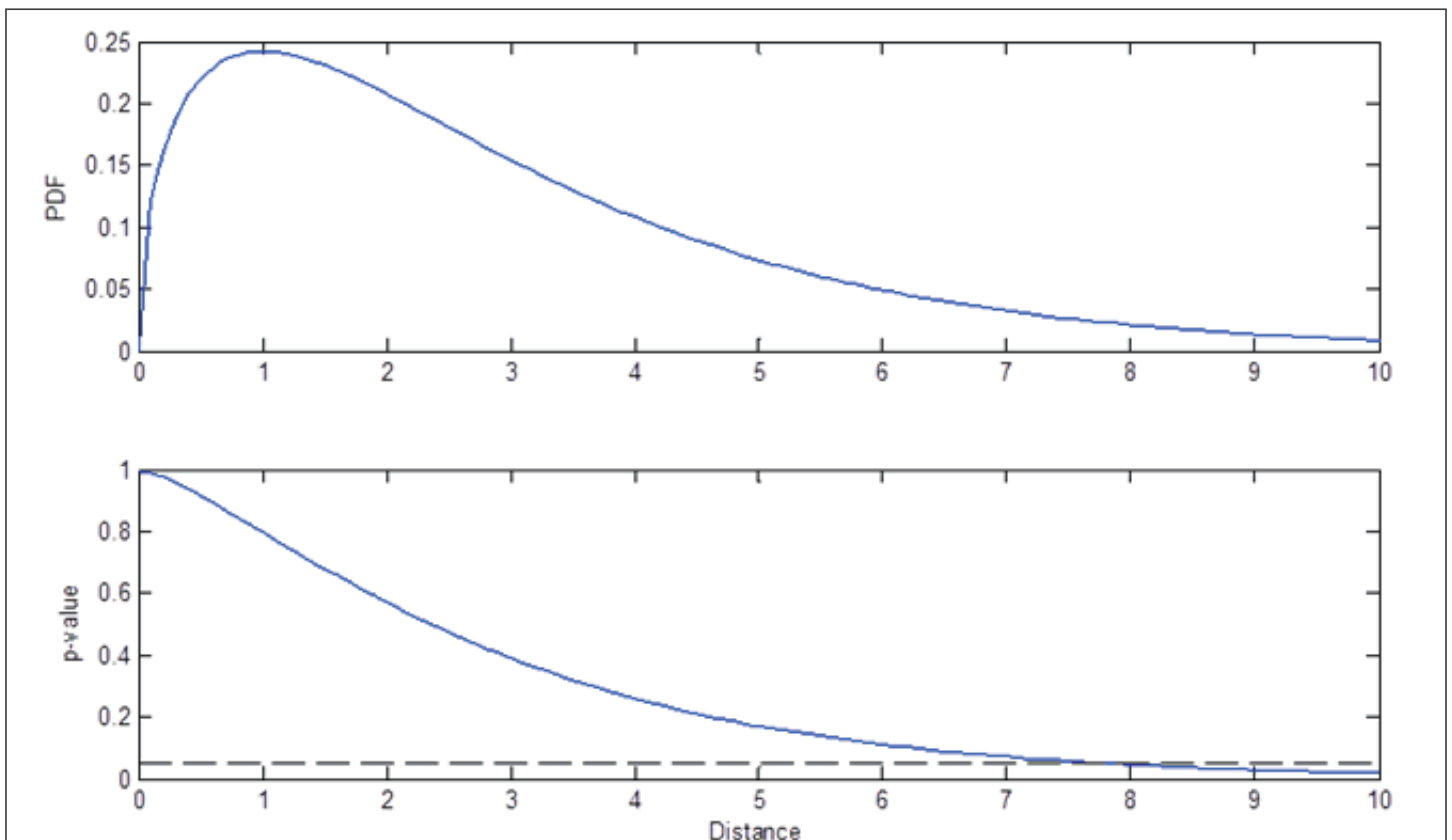


**Figure 8. Statistical distribution plot.**

## Summary

TruScan not only acquires the Raman spectrum of a material of interest, it determines – in real time – the uncertainty of that measurement. Since the software models uncertainty directly, there is no calibration or user-modeling involved with method development. A single reference spectrum typically suffices for method development with bulk materials, because the physical properties of the sample have minimal influence on the Raman spectra acquired and the remaining sources of variability are modeled directly by the embedded analysis.

Performance in terms of both robustness and selectivity for methods based on empirical SIMCA or other similar modeling techniques, which other Raman systems may use, can vary quite dramatically because they depend heavily upon the user's expertise to adequately account for all expected variations in a sample, as well as, environmental and instrumental conditions relative to the p-value.

## References

1. BT Bowie, DB Chase, PR Griffiths, "Factors affecting the performance of bench-top Raman spectrometers. Part I:Instrumental Effects", Appl. Spectrosc., 54:164A (2000)

2. BT Bowie, DB Chase, PR Griffiths, "Factors affecting the performance of bench-top Raman spectrometers. Part II: Sample Effects" Appl. Spectrosc., 54:200A (2000)

3. RL McCreery, Raman Spectroscopy for Chemical Analysis, Wiley (2000)

4. DL Massart et al., Handbook of Chemometrics and Qualimetrics: Part A, Elsevier (1997)

Learn more at **thermofisher.com/truscan**