Probabilistic approach for material verification and identification in pharmaceutical applications

Fundamentals and performance characterization of Thermo Scientific™ TruScan™ Handheld Raman Analyzers

Keywords

Raman spectroscopy, pharmaceutical inspections, material identification, Thermo Scientific TruScan Handheld Raman Analyzers, method development, statistical analysis

Thermo Scientific™ TruScan RM™ and TruScan G™ Handheld Raman Analyzer

Summary

Spectroscopic material verification is a critically important analytical application within the pharmaceutical industry because of the adoption of standards such as the Pharmaceutical Inspection Co-operation Scheme (PIC/S), Annex 8. PIC/S Annex 8 requires that individual samples be taken from all incoming containers and an identity test be performed on each sample. This is a major change from the traditional practice of allowing composite sampling of a statistical subset of the batch and identity testing of the single composited sample, in order to release the batch to manufacturing. Individual container identity testing puts drastically higher demands on expert analysts' time. The TruScan Handheld Raman Analyzer's revolutionary capabilities enable nonexperts, such as receiving personnel at the loading dock, to positively verify the identity of materials at the point of receipt, increasing an organization's throughput and cycle time. The TruScan Handheld Raman Analyzer is the first device of its kind to employ embedded uncertainty tracking and estimation. Tracking the uncertainty enables the onboard analytics to determine if the material is consistent with a particular reference -- without forcing the end-user to perform time-consuming calibrations or call on expert, manual interpretation.

Probabilistic approach

Verification of a material's chemical identity involves instrument hardware to measure the optical properties of the sample and, if expert oversight is to be avoided, embedded statistical

thermo scientific

analysis to determine if the material is consistent with a particular reference or references. Rapid advancements in optical components used for Raman spectroscopy have dramatically improved instrumentation hardware, with regard to ruggedness and portability. Thermo Scientific's TruScan platforms are expressly designed for setup and use by non-experts. In most direct terms, TruScan not only acquires the Raman spectrum of the material of interest but also – in real-time – determines the uncertainty of that measurement, given factors such as the sample characteristics, instrument telemetry, environment and testing environment. By "uncertainty" we mean how repeatable and reliable we expect that measured spectrum to be over similar or even different sampling conditions. In statistical terms, uncertainty refers to standard deviation.

TruScan poses and answers the following question:

- Is the measurement of the test material statistically consistent with the measurement of the reference material?
  - For the TruScan RM analyzer - The answer is "yes" (Pass) if no significant difference between test and reference measurements was observed (p-value >= 0.05), and "no" (Fail) otherwise. See Figure 1.

  - For the TruScan GP analyzer - The answer is "yes" (Positive Match) if no significant difference between test and any reference measurements was observed (p-value >= 0.05), and "no" (No Match Found) otherwise. See Figure 2.

Figure 1. TruScan RM authenification result output.

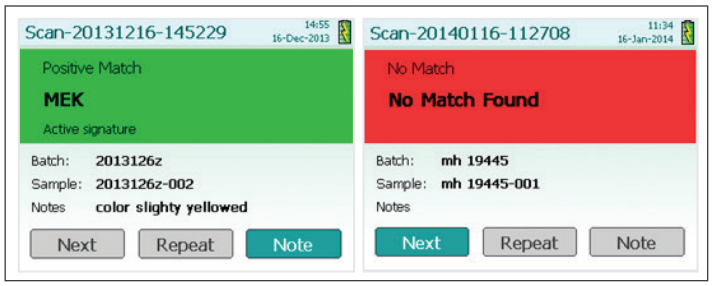

Figure 2. TruScan GP identification result output.

For illustration purposes, the two figures at the top of the next column graphically depict the pass/fail (Figure 3) and positive match/no match criteria (Figure 4) on a univariate t-distribution. The difference between the measured and reference spectra is the basis of the x-axis in the illustration.

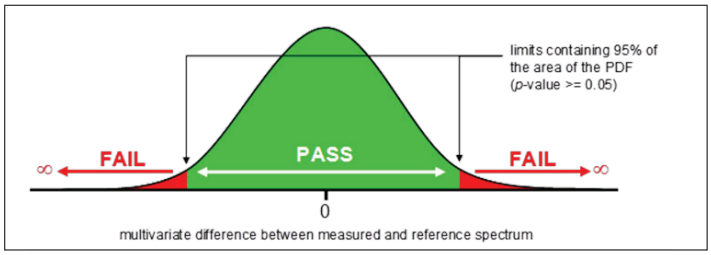

Figure 3. TruScan RM pass/fail univariate t-distribution.

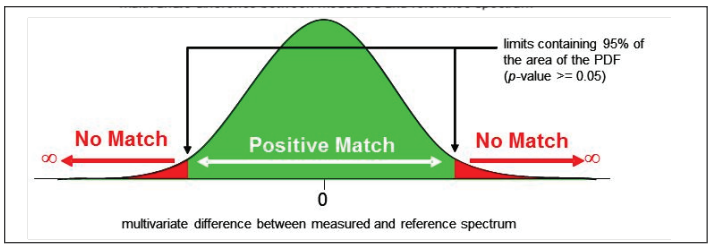

Figure 4. TruScan GP Positive Match/No Match univariate t-distribution.

If the difference falls within the limits corresponding to a p-value of 0.05 (the green shaded area) the sample measurement is considered consistent with the reference spectrum, within the uncertainty of the measurement, and the device will report "Pass" or "Positive Match". If the difference falls outside the limits (red shaded area), the device will report "Fail" or "No Match".

This device employs a 785 nm +/- 0.5 nm laser (excitation wavelength) with a 2 cm^{-1} line width and stability of < 0.1 cm^{-1}, resulting in a Raman shift range from 250 to 2875 cm^{-1} and spectral resolution of 8 to 10.5 cm^{-1} (FWHM). These specifications result in approximately 2000-dimensional spectral data, with the p-value test conducted in all dimensions simultaneously. Therefore, Figures 3 and 4 are extreme but representative simplifications.

The probability curve is developed by the software during method development and its shape is a function of easily modeled sources of uncertainty. Since the software models the 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 (e.g., particle size, packing density, humidity/ water) have minimal influence on the Raman spectra TruScan acquires, and the remaining sources of variability are modeled directly by the embedded analysis.

If more than one reference spectrum is used for a TruScan RM method, the question above is repeated for each reference spectrum, and if the outcome of any comparison is a Pass, then the method as a whole will pass. That is, the measurement of the material was consistent with at least one of the measurements of the reference materials. The analysis does not average the reference spectra together and perform a test of equivalence using the average spectrum.

In the TruScan GP, if the sample measurement is consistent (p-value >= 0.05) with more than a single reference spectrum in the library, all matches are reported in rank order of p-value.

Decisions and outcomes

End users of field-material identification systems are not usually spectroscopy experts and, therefore, rely on an algorithm to convert instrument data to a qualitative result. Like all data-driven judgments, this qualitative result is invariably right or wrong. For spectroscopic material identification, the outcome tree (Figure 5) is slightly more complex than simply right and wrong. Therefore, the critical question for the end user is "If my instrument fails to report a "Pass" or "Positive Match", how often am I correct in identifying the unknown, and how often am I mistaken?"

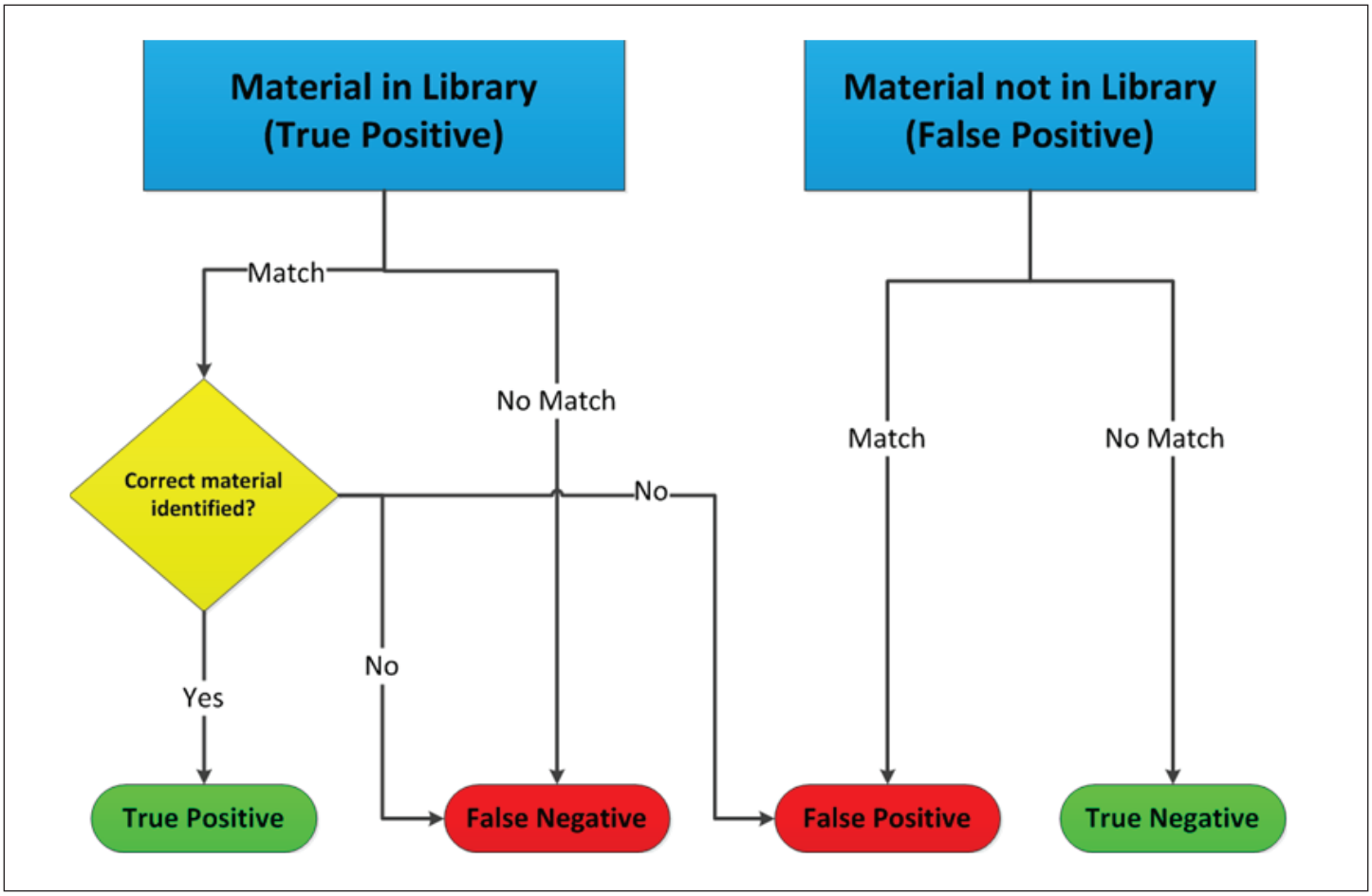

Figure 5. Outcome tree for decision making based on a qualitative result.

Performance characterization

Receiver Operator Characteristic (ROC) curves[1-3] have long been used to depict the tradeoff between sensitivity of detection and false-positive rates for qualitative tests. While they are easily understood graphics of test capability, ROC curves are insufficient for the material identification task in at least one sense: spectroscopic material identification systems could return multiple material records in response to a single measurement query. Qualitative precision is, therefore, an additional performance characteristic that is important from an end-use perspective.

The effectiveness of information retrieval systems are more typically characterized by precisionrecall (PR) curves.[4] McLafferty et al. employed PR curves[5] in their evaluation of mass spectral library search software. However, their objective was substructure identification rather than full molecular identification. The PR curve also does not reflect false-positive rates because in document retrieval it is assumed that there are always relevant records in the database. This is an inappropriate assumption for spectroscopic material identification because the systems will invariably encounter materials that are not in the device's library of materials.

Therefore, it is necessary to construct new parameters for a spectroscopic material identification system, which can be characterized by the following:

- true-positive rate @ t : if materials in the system library are tested under field conditions, how likely is the system to declare a match of an unknown to the correct library materials?
- false-positive rate @ t: if materials that are not in the system library are tested under field conditions, how likely is the system to declare false matches of an unknown to other irrelevant materials in the system library?

To examine the system's operating characteristics, we also have to assume that the spectroscopic material identification device has a tunable parameter *(denoted "t" in the definitions above and is the p-value threshold value in the ROC study)*, which would allow it to be more liberal or conservative in suggesting a material identification.

Experiment

We examined the performance of Thermo Scientific's handheld Raman verification system, as performance verification of the system is critical. A multi-device, multi-user experiment was designed and executed over 8 weeks, with materials chosen at random by serial number. This involved:

6 devices	335 measurements in vial-holder
6 operators	455 measurements in free space
261 unique chemical materials	376 liquid measurements
790 total measurements	414 solid measurements

Five of the users were characterized as having light experience with the device (equivalent to two days of training) and one user was a novice, having approximately five to ten minutes of operational training at the start of the study.

Of the 790 measurements, 664 were made on materials represented in the system library. For a more rigorous assessment of false-positive rates, the results of these tests were also re-analyzed by removing the relevant library record and re-examining test results for any false-positives that would have occurred. The devices executed the embedded version of Thermo Scientific's probabilistic material verification software in real time, and were operating in "auto-measure" mode, in which the device governs all tunable software parameters on a measurement-by-measurement basis.

Results of the experiment

The aggregate ROC curve[6] for all six devices is shown in Figure 6. The experimental uncertainty (95th percentile) in the point estimates is indicated by the blue shaded area while the measured curve itself is in red. The area under this ROC curve, a bulk measure of qualitative accuracy, is 96.9% with an uncertainty of 1.0% (95th percentile). The results were very consistent across the 6 devices/users participating in the study.

Recall that the ROC curve does not convey precision information. Tabulation of the precision characteristics in the experiment is shown below.

# of Possible materials identified	% of cases	Interval estimate
0	92.7%	[89.7% 94.9%]
1	5.2%	[3.4% 7.8%]
2	1.1%	[0.5% 2.7]
3	0.6%	[0.2% 2.0%]
≥4	0.4%	[0% 1.6%]

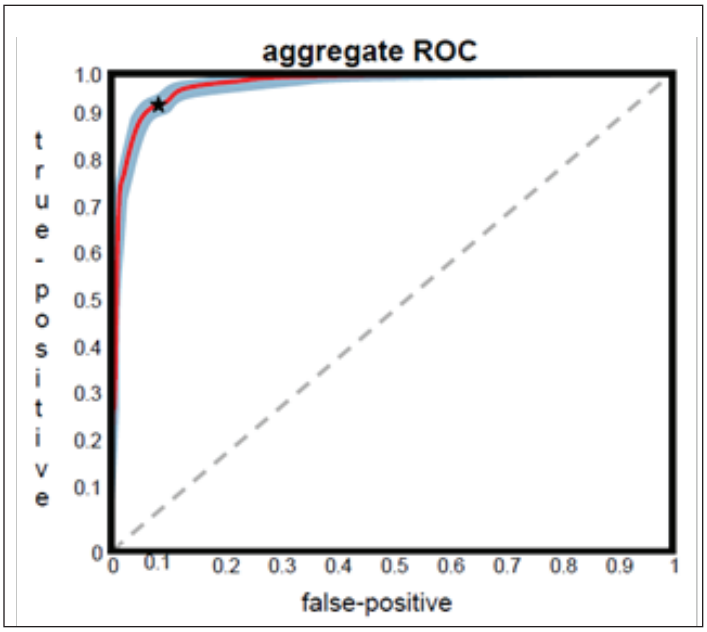

Figure 6. Aggregate ROC curve for all devices in the experiment.

In **93 percent** of all test cases, the system reported only the correct material as a plausible match. In **99.4 percent** of all test cases, the system reported the correct material as the first (e.g. most likely) or only choice.

Summary

Analytical devices employed in the pharmaceutical environment require performance characterization. In applications for material verification and identification, strict rules must be defined for robust, repeatable operation of these devices by non-expert users. Traditional analytical methods of measuring qualitative performance (SNR, SEL, SEN, LOD) are unsuitable for characterizing identification of unknowns. ROC curves and precision characteristics represent a promising means to measure spectroscopic discovery library performance.

When rigorously tested in an end-use situation, the Thermo Scientific handheld Raman devices accurately and precisely identify the correct or most likely material being analyzed. These spectroscopic material identification tools have been employed by non-expert field users and are found acceptable for pharmaceutical validation requirements.

In all, TruScan uses very advanced processing during both data collection and analysis to ensure that sound statistical judgment supports every material-verification decision. It is the first device of its kind to employ embedded uncertainty tracking and estimation, rather than forcing the end-user to calibrate. This translates to very rapid method development, and extremely high objective selectivity in a device in the hands of the non-expert.

References

1. JA Swets et al., “Assessment of diagnostic technologies” Science 205:753-759 (1979)
2. TD Wickens, Elementary Signal Detection Theory, Oxford University Press 2001
3. CD Brown, and HT Davis, “Receiver operating characteristic curves and related decision theory measures: a tutorial”, Chemometrics and Intelligent Laboratory Systems 80:24-38 (2006)
4. D Lewis, “Representation quality in text classification: an introduction and experiment”, Proceedings of the Workshop on Speech and Natural Language, Morgan Kaufmann 1990; pp. 312-318
5. GM Pesyna et al. “Probability based matching system using a large collection of reference mass spectra” Analytical Chemistry 48:1362-1368 (1976)
6. Proc. of SPIE, spie.org, Vol. 6378 637809,1-11

i CD Brown and RL Green, “Considerations for embedded authentication using intelligent portable Raman systems,” FACSS 2006, Orlando, FL

ii CD Brown, “Revisiting the Assumptions of Spectral Library Search in Light of New Frontiers,” FACSS 2005, Quebec, PQ

iii J Li, DB Hibbert, S Fuller, G Vaughn, “A comparative study of point-to-point algorithms for matching spectra,” Chemom.Intell. Lab. Syst. 82:50 (2006)

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

v BT Bowie, DB Chase, PR Griffiths, “Factors affecting the performance of bench-top Raman spectrometers. Part II: SampleEffects” Appl. Spectrosc., 54:200A (2000)

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

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

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