# Visualisation of sampling error effects in near infrared analysis—comparison between Petri dish, roll bottle and spiral sampler

Marie Sørensen,[a] Anders Larsen[a] and Kim H. Esbensen[b,c,d]
[a]Q-Interline A/S, Tølløse, Denmark
[b]Geological Survey of Denmark and Greenland (GEUS), Copenhagen, Denmark
[c]ACABS Research Group, Aalborg University (AAUE), Denmark
[d]Telemark University College (HIT), Porsgrunn, Norway

With spectroscopic methods, e.g. near infrared (NIR) analysis, using a constant beam aperture, the effective scanning footprint will be different for a spinning Petri dish, a rolling bottle and a new spiral sampler configuration. This will significantly influence the analytical accuracy and precision of a NIR analytical determination of heterogeneous materials, for example barley with differing protein contents. Here we present the results from a bench-top experiment that evaluates the total analytical bias and precision characteristics for three alternative sample presentation approaches using a mixture of two plastic polymer pellets as a test material with significant heterogeneity. After removal of all incorrect sampling errors (ICS), there are still significantly varying correct sampling error [Fundamental Sampling Error (*FSE*) and Grouping and Segregation Error (*GSE*)] uncertainties associated with these standard analytical approaches—but there is a clear winner.

## Introduction

Analytical spectroscopic scanning methods such as near infrared (NIR) spectroscopy are widely used to analyse a multitude of different types of materials, most of which are irregular (heterogeneous). Based on proper calibration and validation with respect to a suitable reference method, NIR analysis is a fast approach giving reliable results in a very short time (often <1 min), hence its enormous range of applications and number of successes. Much interest is traditionally given to the analytical performance of the instrument and this may vary depending on the instrument type, the acquisition conditions and the quality of the calibration. In many applications, however, this is of less importance if analysis is to be performed on materials with a significant degree of *heterogeneity*. In such a situation, it is of essential interest to focus both on how the analytical aliquot was produced (representativity of the full sampling process) as well as how it is introduced to the instrument. Most common laboratory instruments either use fixed volume vials or Petri-dishes for liquid and solid samples, respectively. While these are adequate methods for sample materials which are compositionally relatively uniform (e.g. one-phase liquid samples, finely ground, well-mixed materials and/or natural powders), significantly heterogeneous material like whole wheat, grass, hay, silage, meat, vegetables, fruits, berries etc. will present a severe challenge due either to particle size and/or degree of heterogeneity in relation to the absolute sample size possible (often fixed by the Petri dish diameter). A representative sample will often be difficult to define in such cases if not based on the full principles of the Theory of Sampling (TOS). In many analytical communities there has been, perhaps understandably, a tendency to the attitude of "too much of this sampling focus—let's get on with the analysis". It would, for example, be quite pleasant were such issues to be eliminated by a "smart, fit-for-purpose" sample presentation technique or if a universal accessory was available. At various times the spinning Petri dish, as well as the rolling bottle and other adaptions have both been hailed as more or less the final answer to these issues, and the newly developed "spiral sampler" is but the latest such candidate.

It was felt that it would be useful to both the NIR and sampling communities to compare these three widely-available sample presentation options based on a quantitative evaluation. We shall analyse the sample presentation and spectral acquisition situations from the principles of the TOS, but otherwise let the numbers speak…

Three alternative sample presentation approaches compared in the present study are shown in Figure 1. The effective sample mass (area) achieved in each is markedly different. The predominant effect and difference between the accessories are the observable surface, and hence sample

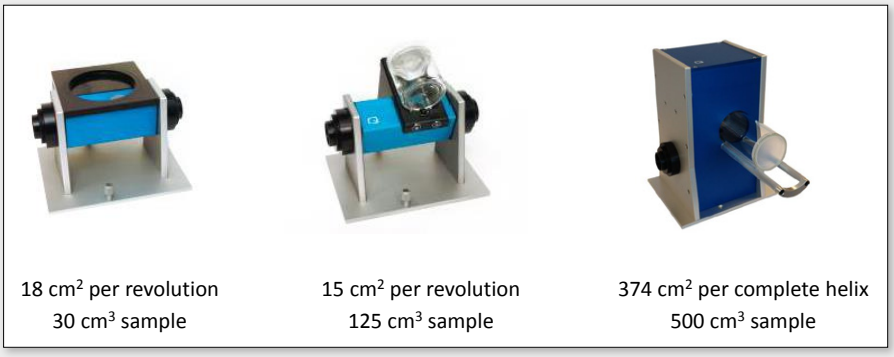


Figure 1. Illustration of the different NIR accessories [Petri sampler (left), bottle sampler (centre) and spiral sampler (right)] which present the sample to the spectrometer in three different ways. The effective scanned areas are also indicated (dependent on method/number of scans) and the maximum sample volume for the respective sample containers.

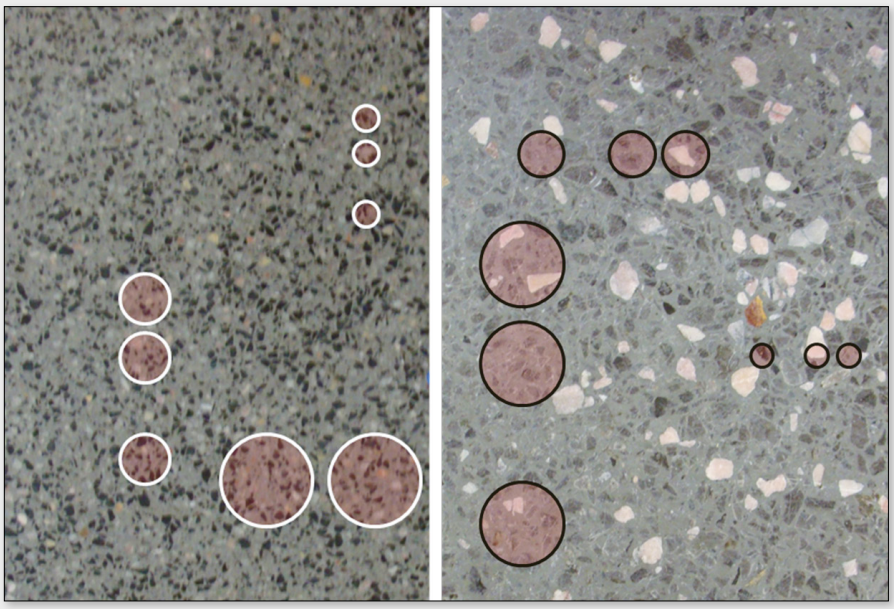

Figure 2. Illustration of the effect of increment size (circles) for two materials with markedly different heterogeneity. In NIR analysis the increment size can be increased by acquiring a larger sample area (a larger circle area in this illustration). Using "too small" a sample area (small circles) can clearly lead to significantly different analytical results when performing repeated analysis, due to significant *GSE* and *FSE*.

sampling will never result in exactly identical concentration values. Repeated sampling, even with exactly the same procedure, will unavoidably extract *different* increments of the heterogeneous lot material. This is the effect of what is termed the Fundamental Sampling Error (*FSE*) in TOS. In practice, the effect is often augmented by uncertainty stemming from the Grouping and Segregation Error (*GSE*) which originates from meso-scale spatial heterogeneity. The primary way to reduce *FSE* and *GSE* is by reducing the degree of heterogeneity, which can be done effectively by comminution and subsequent mixing or just mixing (to reduce the existing spatial heterogeneity). Alternatively a larger sample size has to be used—or best, a combination of all of the above.

What is needed in a specific situation will be determined by the decision of the acceptable level of the *TSE* (Total Sampling Error)—to which must always be added the Total Analytical Error (*TAE*). The main issue is that *TSE* is nearly always significantly larger than *TAE* (factors of 5–20 are normal and may be even higher in particularly adverse situations). *FSE can* be estimated mathematically following Pierre Gy's formula[1–3] which, with *some* approximation, can be used to estimate the minimum sample mass, or conversely, the maximum grain size that corresponds to an *a priori* given (*TSE*+*TAE*) level. In general, *FSE* is inversely correlated to the effective aliquot mass being analysed and is also dependent upon the analyte concentration. However, the influence of *GSE* is **not** included in any of this type of *FSE* estimation, for which reason many practitioners prefer to establish the *effective* sample mass vs acceptable *TSE*+*TAE* levels by some appropriate *empirical* approach; this forms the basis for the present evaluation. For a more in-depth introduction to the specific TOS issues, see, for example DS 3077[4] and the many references found therein.

## Materials and methods

White polyethylene (PE) pellets and white polystyrene (PS) pellets were purchased from industrial plastic manufacturers. The density of each type of pellet was determined experimentally. Master samples of 2%, 10% and 20% PS concentration levels in PE (vol/vol) were prepared based on appropriate masses and pellet densities. The accuracy (absolute level) of these concentration levels 2%, 10% and 20% is not important with respect to the conclusions but all master samples were still prepared with the outmost care, allowing us to assume the error contribution from preparation can be neglected in this experiment when compared to the errors arising from sampling/presentation. Sampling from master sample lots was carried out using a spoon with a size chosen such that the volume in each type of sample container was achieved by combining 10–12 composite increments. The filling degree was 100% for all containers in the present study.

Figure 3. Illustration of the different sample masses typically achieved with the Petri dish (left), rolling bottle (centre) and spiral sampler (right). Typical proportions between the effective analytical masses are 35 g, 50 g and 600 g, respectively. N.B. the analytical NIR sensor system does not interact with the entire mass in the containers.

NIR scanning was performed on a Quant FT-NIR instrument (Q-Interline, Tølløse, Denmark), equipped with an InAs detector. The spectra were acquired in the range from 4000 $cm^{-1}$ to 12,500 $cm^{-1}$ (800–2500 nm) with a data point spacing of 8 $cm^{-1}$; 180 scans were recorded for each analysis. For the needs of the present comparison experiment, all analysis was carried out under the same ambient conditions.

## Experimental design

The experimental design (Figure 4) imitates sampling from a realistic primary lot followed by packing in container and analytical measurement. Each of 10 sampling replicates from the master lot was measured with the standard NIR approach (experimental level S1). One random replicate was re-packed and measured 10 times (experimental level S2). Finally, one randomly selected replicate was measured 10 times to estimate repeatability (experimental level S3). This design was applied identically to all three sample presentation methods.

## Spectral data prediction model

To predict all concentrations in this binary experiment, a PLS1 calibration model was constructed. Samples used for calibration comprised a single spectrum from each sample presentation method and each concentration level in this study, plus spectra of four additional concentration levels. The spectral range used was 5870–9025 $cm^{-1}$ and pre-processing comprised Savitzky–Golay 1st derivate, 13 points. As could be expected, the model needed only two PLS components to explain close to 100% of all variation. This model was used to predict all the PS concentrations in the study.

## Results and discussion

For an apparently simple and straight-forward analytical method, such as NIR, there are still multiple error sources that contribute to total uncertainty variance. The primary sampling error effects are associated with how to get a representative sample with respect to the whole lot without contributions from incorrect sampling errors. Here, the total sample mass must be considered relative to the inherent heterogeneity of the lot. Very heterogeneous materials would of course benefit from a larger sample size as compared to more uniform materials (but **only** if based on proper composite sampling). This is needed in order to get an acceptable, reduced contribution from *FSE* and *GSE*. If sub-sampling must be employed, each such stage forms a completely new "primary sampling" scenario at a reduced scale. Petersen *et al.*[3] present a complete survey of all available techniques for this purpose, including empirical evidence for selecting optimal approaches only (splitting).

At some point, the proper sample size (mass) has been achieved, however, and is now to be presented to the NIR instrument. This also has to be carried out in a representative manner, i.e. all parts of the scanned aliquot (sample) should have the same probability of contributing to the analytical spectrum. This is often not the case with the three options being investigated. This is often a direct effect of the type of sample preparation, forced by the design of the sample presentation method and accompanying sample container (Petri dish, bottle, spiral sampler tube). Additionally, the spatial filling of these sample containers can contribute to *GSE* as there may be a tendency towards different packing as a result of differences in density, surface properties and shape of particles in the lot material.

The final focus for the present study is the sample presentation methodology, which will influence the validity of the analytical results with regard to effective scanning area relative to sample size (a *FSE* issue) and the physical sample presentation that should seek to minimise effects from *GSE*. The *effective* scanning area is the area of the sample surface that is actually scanned in depth and which contributes to the acquired NIR spectrum. Typical proportions between the effective scanned areas are 18 $cm^2$, 30 $cm^2$ and 300 $cm^2$ for the Petri dish, the bottle spinner and the spiral sampler, respectively.

### Petri sampler

For the Petri sampler, the effective scanning area is an annular area measured on the bottom side of the Petri dish which in the present case corresponds to ~18 $cm^2$. Increasing the number of scans above the acquisition time corresponding to a full annular revolution will not reveal any new sample surface area but merely results in repeated scanning of the *same* sample surface as has already been fully covered. As a result, multiple measurements (S3) of the same Petri dish give excellent repeatability (Figure 5). However, if the sample is re-packed or re-sampled in the Petri dish, a completely different result is revealed. This can be seen in Figure 5 as a significantly larger standard deviation for re-packed and re-sampled Petri dishes (S1 + S2) compared to repeated measurements (S3).

Relatively large bias values are also characteristic for the Petri dish presentation (Figure 6); this does appear also to be the result of the small scanning area. Combining the reproducibility (here repeatability) and bias into representativity, a central

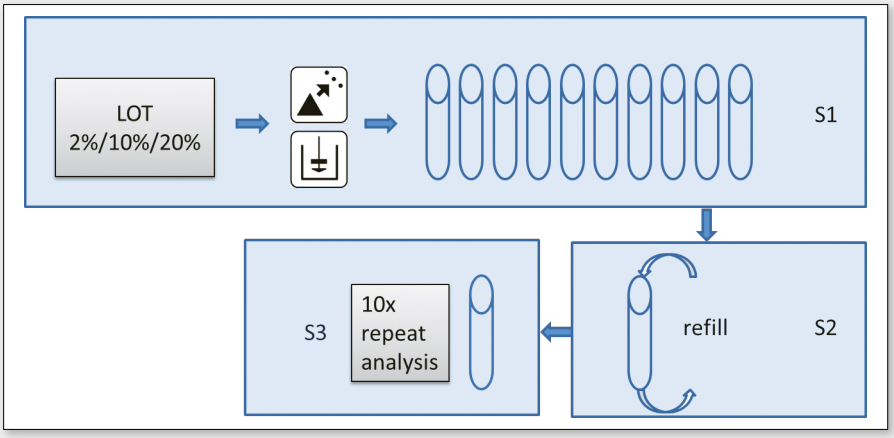


Figure 4. Flow-path diagram of the experimental design. Experimental level S1: the three master sample lots at 2%, 10% and 20% PE were mixed thoroughly and then, using composite sampling, 10 sample containers (Petri dish/rolling bottle/spiral spinner tube) were filled and analysed by NIR with each of the respective sample presentation methods. Experimental level S2: after NIR analysis, one of these sample containers was emptied and re-packed into the same container 10 times, each analysed by NIR. Experimental level S3: one container was finally re-analysed 10 times. Each cylinder in this illustration represents either a Petri dish, rolling bottle or spiral sampler tube.

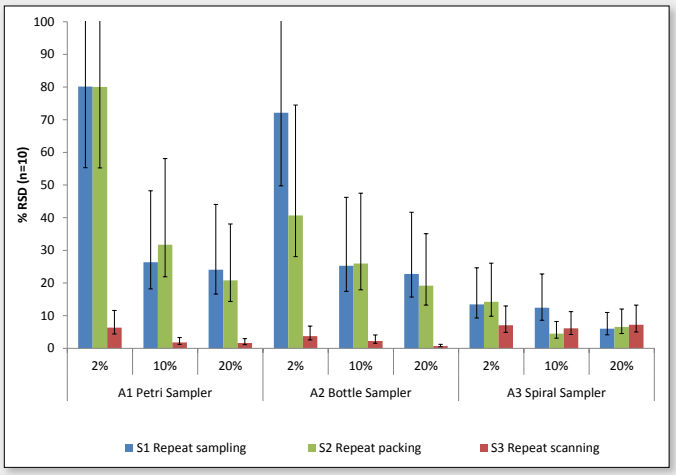


Figure 5. Replication error contributions shown expressed as relative standard deviations obtained for 10× repeated sampling, 10× repeated packing of sample containers and 10× repeated measurements for each sample presentation method and at each of the three concentration levels. Error bars show 95% confidence limits.

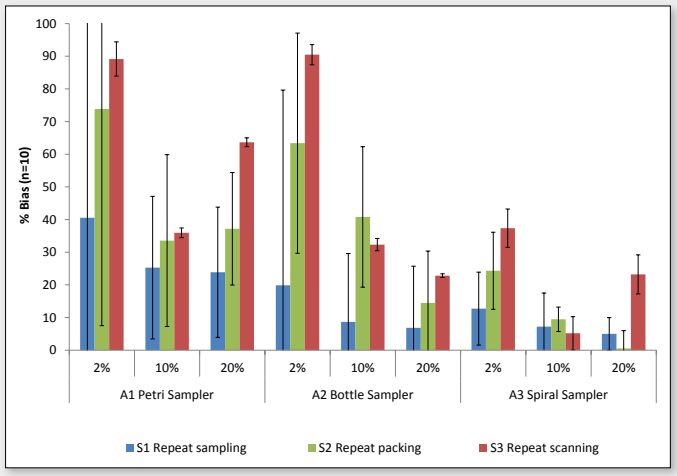


Figure 6. Bias contribution shown as relative bias [%] as obtained for 10× repeated sampling, 10× repeated packing of sample containers and 10× repeated measurements for each sample presentation method and at each of the three concentration levels. Error bars show 95% confidence limits.

tenet of TOS, one reaches the conclusion that the Petri spinner has the worst representation of "the truth" which is known in this controlled experiment (Figure 7). The representativity quickly becomes devastatingly worse at low concentrations because of the progressive influence from the irregular distribution of the analyte (increasing *GSE*). This feature disqualifies the Petri dish as a valid element in any *representative* measurement system.

This is exactly what should be expected in the light of TOS. Since the Petri dish is measured on the base, what is measured is dominated by material that settles at the bottom. For sample material prone to segregation (very many types of foods, feeds, powders) this results in significant *GSE* and may cause a large bias as well. Because of this risk of segregation and the limited size of the dish, the Petri sampler is only suitable for fine powders and even here thorough mixing is a requirement; for more heterogeneous material, the contribution of *GSE* will reduce analytical performance rapidly.

## Bottle sampler

The bottle sampler accessory is designed to rotate a semi-filled 125 mL bottle around its length axis while positioned at an angle of 22° to the horizontal; this offset is in order to stimulate tumbling/mixing (however, in this study the bottle spinner is used with 100% filling for comparative reasons). The scanned area is the circular belt area around the cylindrical sides of the bottle, which corresponds to a mere 15 $cm^2$. But, for bottle contents that mix during spinning, increasing the number of scans will, to *some degree*, increase the effective scanning area as long as scanning continues. For a full bottle, however, the effective total scanning area will not increase above 15 $cm^2$ and in this regard the bottle sampler should then resemble the Petri dish and therefore show similar trends for standard deviation and bias (Figures 5 and 6) which is indeed the case. Compared to a Petri sampler, there is a tendency for a lower bias (not significant) for the bottle sampler. Since the glass quality is similar and all other parameters are kept identical, this difference could relate to the target material. The plastic pellets are perhaps packing better in the bottle as compared to the Petri dish bottom. The end result (Figure 7) is that bottle sampler representativity is slightly better than the Petri dish, especially at higher concentrations which is in support of the thesis that the pellets are not perfectly identical and hence pack differently. These minor differences will of course be larger with increasing contrast between different particle sizes and/or densities.

Detailed inspection of Figures 5 and 6 reveals minor differences in the relationships between the performance of the rolling bottle and Petri dish with respect to both bias and replication variability. These differences are not statistically significant but are stochastic reflections of the interplay between heterogeneous materials being repeatedly sub-sampled, re-packed and re-analysed. No general conclusions can be drawn on this basis; there is always such a random effect in the sampling plus analysis system.

## Spiral sampler

For the spiral sampler, the scanned area is a belt wrapped around the glass tube in a helical fashion. This helical belt ensures that the beam footprint continues to cover new sample material along the entire cylinder length and effective scanned area is therefore limited only by measurement time and tube length. For the size of glass tubes used, the maximum possible area is 375 $cm^2$. Experiments reported here were actually limited by measurement time (here 180 scans) corresponding to 275 $cm^2$, which is still many times larger than for the other methods (>five times). Due to the larger scan area, the repeatability of scanning is not as low as for the other sample presentation methods (Figure 5, S3). While for both the Petri spinner and bottle sampler there were significantly higher standard deviations for repeated sampling (S1) and re-packing (S2) compared to repeat scanning (3), this is not the case for the spiral sampler. This means that one properly taken sample fully represents the heterogeneous lot under study and that repeated, repacked sampling reveals no new information. This

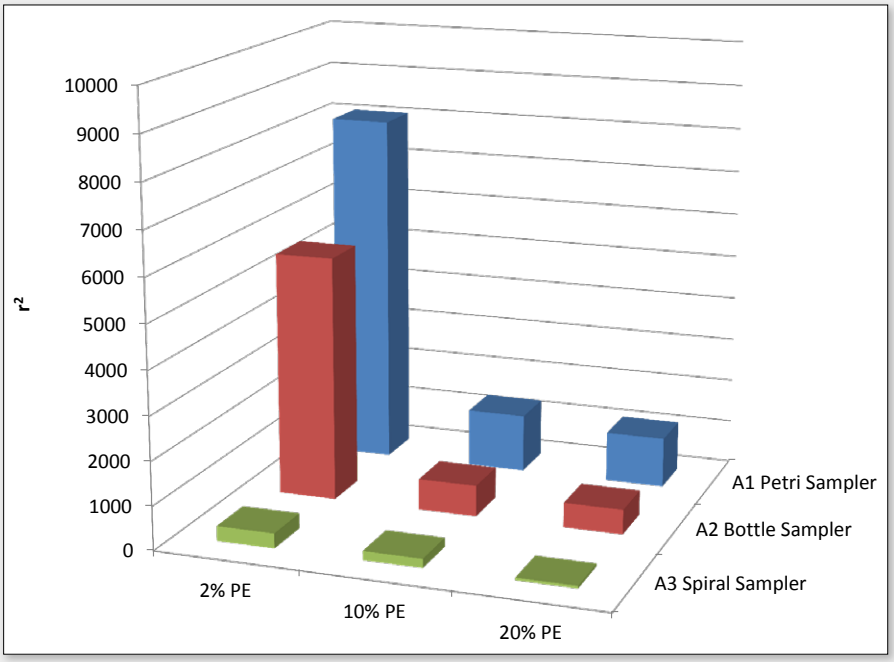


Figure 7. Total representativity characterisation for 10-fold repeated sampling, sample presentation and analysis using A1: Petri sampler, A2: bottle sampler and A3: spiral sampler.

becomes very clear when representativity calculations are done for the spiral (Figure 7) in which case the results are magnitudes better than for the other two methods.

Both the relatively large sample size and the large scanned surface area influence the *FSE* and *GSE* in a positive direction.

In this study, the spiral sampler was tested completely filled but can also be used with a fill level that enables mixing, as can the bottle sampler. Such a strategy will, in general, not change the present findings significantly.

## Conclusion

The three alternative sample preparation methods have very different characteristics in terms of precision and accuracy. It is evident, not only from the present results, that analytical precision alone is **not** an adequate measure of the performance of a NIR method as has otherwise been considered good practice for a while. Only after careful evaluation of the potential offset of the results, stemming from **both** *GSE* and *FSE*, may a particular NIR measurement *system* be successfully applied without significant risks of faulty and potentially expensive, wrong conclusions.

No doubt the well-known and easy-to-use Petri dish is the winner in the battle for best precision under repeatability conditions, but the characterisation as best precision *per se* is a complete *mirage*. It is abundantly clear that this only reflects the ability to predict the same wrong result several times in a row (it is in fact simply analysing "precisely wrong"). The bottle spinner is, in general, slightly better although not hitting any highs compared to the new spiral sampler, which combines good precision with low bias and thus very clearly comes out on top with respect to a full definition of representativity. Its ability to lower *FSE* and *GSE* significantly is due to the much larger composite sample mass and a much larger effective scan area.

What happens to a sample received in the analytical laboratory is not a trivial matter; significant sample preparation and presentation errors can arise. Still, much will depend on the validity of the full sampling plus analysis process—i.e. the first sampling stage is of critical importance concerning the accuracy with respect to the original lot (potentially creating a significant, inconstant bias). The entire "lot-to-aliquot" pathway is analysed rigorously from the standpoint of TOS in a new international sampling standard, DS 3077.[4] The task of being able to produce correct predictions, i.e. accurate and precise predictions closely resembling the real world, may often necessitate that the full chain of actions from primary sampling to NIR acquisition must be rewritten, away from what is the most "convenient" to what is most accurate and follows closely the principles of TOS.

The experimental binary "product" used here mimics many types of real-world counterparts and materials, for example freshly harvested sugarcane, silage, corn or the likes with stems, seeds and fragments displaying areas close to, or even larger than, one singular NIR beam footprint. For all such material types, as well as for all materials with *similar* heterogeneity characteristics, the clear winner is the spiral sampler.

Further, there is a significant potential for transfer of the present results to other applications within the area of process quality control employing PAT, often also using fixed-beam NIR sensor technologies, whether in-line, on-line or at-line. The same TOS principles invoked here can also be applied there, see, for example, Esbensen and Mortensen.[5]

## References

1. P. Gy, *Sampling for Analytical Purposes*. Wiley (1998). ISBN 0-471-97956-2

2. F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practise: Heterogeneity, Sampling Correctness and Statistical Process Control, 2nd Edn*. CRC Press (1993). ISBN 0-8493-8917-8

3. L. Petersen, C. Dahl and K.H. Esbensen, "Representative mass reduction in sampling—a critical survey of techniques and hardware", in *Special Issue: 50 years of Pierre Gy's Theory of Sampling. Proceedings: First World Conference on Sampling and Blending (WCSB1)*, Ed by K.H. Esbensen and Minkkinen. *Chemometr. Intell. Lab. Sys.* **74(1),** 95–114 (2004).

4. K.H. Esbensen (chairman taskforce F-205 2008-2013), *DS 3077. Representative Sampling—Horizontal Standard*. Danish Standards (2013). www.ds.dk

5. K.H. Esbensen and P. Mortensen, "Process sampling (Theory of Sampling, TOS)—the missing link in process analytical technology (PAT)", in *Process Analytical Technology,* 2nd Edn, Ed by K.A. Bakeev. Wiley, pp. 37–80 (2010). ISBN 978-0-470-72207-7