

Hyperspectral imaging as process analysis technology for inline applications – Laboratory precision meets high sampling accuracy

Matthias Kerschhaggl

EVK DI Kerschhaggl GmbH
Josef Krainer Strasse 35, A-8074 Raaba

Abstract Hyperspectral Imaging (HSI) has found its way into modern production lines as yet another camera technology that combines the means of spectroscopy with computer vision. However, while its imaging capabilities with its increased amount of processed data are already subject to state of the art sorting machines used for product integrity checks and foreign body removal, the potential of doing real-time spectroscopy and quantitative chemical analysis of product streams has up to now not fully been harnessed for the production line. Respective industries still rely solely on highly precise but statistically limited, laboratory based, slow and invasive methods while HSI offers the data needed to perform real-time and statistically representative product sampling as add on. In this regard the potential of HSI as process analysis technology (PAT) will be outlined in this article presenting a first industrial field application based on the HELIOS near infrared hyperspectral imaging camera. Results from this up to now first commercialized application of this kind in the potato industry, i.e. an inline dry matter detection for potatoes, will be discussed.

1 Introduction

Modern factories and production lines consist of a variety of different process steps covered by fully automated machines comprising high-tech sensor units and electronic equipment. These artificial organs of an integrated factory body nowadays produce enormous amounts of data

most of which are only analyzed from a very local point of view with respect to the process location (i.e. data from one machine is used mainly for process control of this particular unit). On the other hand many inline process parameters with rapidly changing states on timescales of minutes at different locations in the product stream are still monitored by means of invasive, laboratory based analysis technologies suffering from low sampling rates (e.g. every 30 min) and limited (not representative) statistics (number of samples orders of magnitudes below the number of product items to be controlled). Recently more and more of these classical methods are combined with more rapid at- and inline process analysis technologies (PAT) with access to high data statistics and almost no latencies in the control and regulation cycle. Both of these developments are subject to a combined trend towards intelligent (in the machine learning sense), interconnected and self-regulating production sites where inline monitoring (and process control) machines (such as e.g. sorting machines) generate real time data at different points of the production flow. These are combined and analyzed (e.g. via multivariate data analysis) and ultimately used for process control at various process locations. Concerning inline PAT one has to keep in mind that it is not simply the goal to imitate and replace the high precision equipment of a well introduced laboratory method with its various advantages [1]. The actual objective is to add to a solid and high precision laboratory quality control the benefits of an inline technology that monitors the product stream all the time. Such it generates lots of representative data over the whole product stream thus harnessing the power of statistical analysis by which (expected) random statistical fluctuations and non-representative outliers can be averaged out. Such effects will subsequently not propagate into a meaningful mean measurement value. On the other hand real systematic deviations on shorter time scales can be observed and reacted at swiftly since statistical parameters such as target value variance or overall distributions are very sensitive to such changes which will not be marginalized. Thus it is crucial that a precise laboratory reference method is used to (re-)calibrate the inline device whose added value is the enhanced accuracy of the quality control via statistical representative expectation values being insensitive to Gaussian noise. With the help of PAT the different processes in a factory can be fine-tuned in order to arrive at the most cost effective production parameters such as energy, water and input of raw material.

In the following the inline application of quantitative dry matter detection in potatoes and French fries, respectively, by means of a HELIOS near infrared (NIR) hyperspectral imaging camera (EVK GmbH, Raaba/Austria [2]) is presented exemplary for the potential of HSI-PAT. Similar applications such as the quantitative analysis of active pharmaceutical ingredients (API), metal oxide content in ores or all kinds of constituents in a food matrix can be done along the same lines and are subject to current development activities [3]. The hyperspectral imaging (HSI) technology in the NIR regime is about to develop in the food processing industry from a sensor technology used mainly for classification tasks (e.g. removal of foreign bodies from an input stream) to a method for spatially resolved quantitative analysis of chemical food parameters (sugar, starch, proteins etc.). The inline inference of for instance dry matter values can help in order to save energy for the frying process and control/adjust the product quality for different products and customers on short time scales. The latter expect specific lower limits on the dry matter levels for their product.

2 Data Taking and Analysis

Hyperspectral imaging data in the NIR regime (1050-1690 nm) was acquired using a HELIOS NIR G1 HSI camera with a scanning rate of 330 Hz full frame translating into ≈ 81000 processed data points per second.¹ The inline inference of spatially resolved dry matter values in French fries and potatoes was investigated with this inspection technology. The outline of this work was first to perform a proof of principle test with a prototype setup in a French fry factory and then in a second step the transfer of the full functionality onto a commercializable monitoring machine.

2.1 Inline Determination of Dry Matter

As was already shown in [4] & [3] the quantitative inference of dry matter levels in potato tuber slices is feasible using NIR hyperspectral imaging in conjunction with a factor analytical ansatz such as partial least

¹ A more thorough description of the HELIOS camera and hyperspectral imaging in general can be found in [4] & [3]

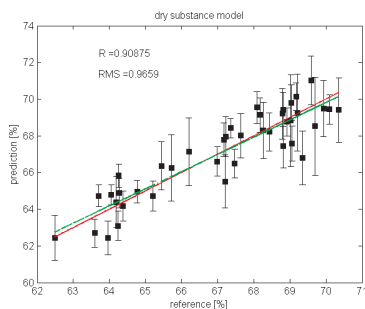
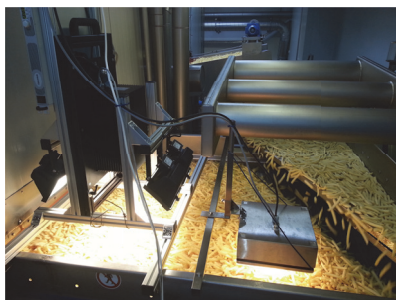


Figure 22.1: (left) Inline test setup for dry matter measurements of French fries using HELIOS NIR G1. Figure taken from [3]. (right) Correlation between reference analysis and model prediction. The red line indicates perfect correlation whereas the green line depicts a fit to the data.

squares regression (PLSR) [5]. While this could be shown for a laboratory based test setup the next project step included the validation of this method during an inline test run in a French fry factory and the ultimate transfer to a working machine fit for commercialization [6]. The inline method validation was carried out by mounting a HELIOS NIR G1 camera right after the freezer² unit on top of the product stream (see fig.22.1). The product was illuminated by 4 halogen lights mounted at an inclination of roughly 60° with respect to the camera optical axis. Such the camera was positioned head on to the inspection line allowing for an optimal trade-off between diffuse reflection, i.e. reduced specular reflection and overall image intensity. Overall roughly 3 min of data, corresponding to 15 million spectra, were taken over a full day in single shots of 2 seconds duration each. The target value range was spanned by deliberately varying the product dry matter, i.e. manipulating the frying time of the frying unit. For each data acquisition shot a reference analysis was performed on a sample of about 5 kg of French fries drawn from the product stream at the moment of data taking. Each sample was analyzed for its average dry matter content by means of a *SMART Turbo Solids Analyzer* [7]. The dataset was pre-processed using the spec-

² At this point of the production line the French fries feature a stable temperature of about 5° .

tral first derivative and split into two subsets one for implementing a chemometric calibration based on PLSR with 5 latent variables and one (by selecting every second data) for internal validation. The correlation plot between reference values and model prediction is shown in the right plot of fig. 22.1. As can be seen from the Pearson correlation $R \approx 0.9$ and the root mean square error of calibration $RMSEC < 1\%$ the suggested model is within statistical margins fully compatible with the reference analysis method. The figures for the validation set showed a somewhat reduced but stable prediction behaviour at levels of $R = 0.86$ and $SEP = 1.3\%$ for original spectra and $R = 0.74$ and $SEP = 1.9\%$ for the normalized spectra.

While R and $RMSEC$ are in particular sensitive to the calibration precision, i.e. repeatability of the measurement, it is more important for an inline calibration to focus on the accuracy of a model, i.e. the degree of closeness of the measurement to the true value, especially if the dataset comprises high statistical, Gaussian scatter that will be easily marginalized under the impression of the enormous inline data statistics. One has to note that the almost perfect alignment between the (green) linear fit to the correlation plot fig. 22.1 and the (red) line denoting a correlation coefficient of one is in fact a better measure for accuracy of the HSI method than any figure of merit prone to noise (see also [8]). While the model precision is based on only 2 seconds of data taking per measurement point, the accuracy of the method will surpass any laboratory based method by averaging out fluctuations that otherwise propagate into the low sampling rate results of classical at line methods. The method's accuracy is thus to first order only limited by the intrinsic sampling error of the reference method.

2.2 Dry Matter Monitoring Machine

With the inline test meeting the accuracy requirements for a field application the ultimate goal of this work was the transfer of the described HSI dry matter measurement to the technically more challenging environment of a real monitoring machine. Such a machine in contrast to the described test setup suffers e.g. from suboptimal local resolution (due to commercially triggered maximization of the working width) and other optical limitations (reduced focal length of the fore lens and thus enhanced lens aberrations etc.). Thus a proof of principle as described

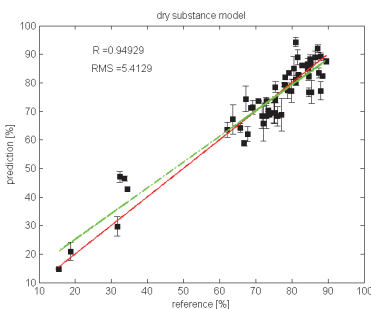


Figure 22.2: (left) HSI monitoring machine *Sherlock Observer* from INSORT. (right) Correlation plot dry matter prediction vs. reference values for *Sherlock Observer*. The red line indicates perfect correlation whereas the green line depicts a fit to the data.

above does not at all guarantee feasibility for field applications. Therefore the dry matter measurement method using a HELIOS NIR camera was implemented on a monitoring machine of the type *Sherlock Observer* from INSORT GmbH (see fig. 22.2 (left) and [6]). The implementation on the machine was performed using raw potato slices to arrive at similar optical conditions, i.e. surface reflectance etc., than for e.g. French fries. The various slices were measured on the *Sherlock Observer* featuring a working width of 300 mm. Reference analysis was carried out along the same lines as for the factory test with a *SMART Turbo Solids Analyzer* using only 10x10x10 mm cubes from the central parts of the corresponding slices. Accordingly, only spatially resolved spectra from the same center region from the HSI image were used for the PLSR multivariate analysis. The sample preparation included the generation of two different populations. Population I spanned an extended value range from below 20% to over 90% moisture in the tubers. In order to generate very dry potato probes a desiccation chamber for dehydrating the samples was used. Population II comprised only natural sampling, thus featuring a confined range in moisture levels. Figure 22.3 shows the validated correlation of the model prediction based on splitting the data into a training and a test set for the population I model (left) as well as the model performance for the confined population II (right). Both models

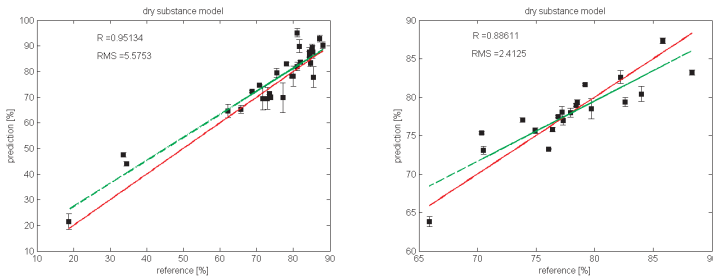


Figure 22.3: (left) Validated correlation between model prediction and reference analysis of potato moisture levels over the extended dry matter level range (population I). (right) Model correlation for confined dry matter levels (population II). The red lines indicate perfect correlation whereas the green lines depict a fit to the data, respectively.

show a fairly high correlation with the true values from the reference analysis, i.e. $R = 0.95$ and $R = 0.88$ for population I and II, respectively. However, the RMS model error is enhanced for the population I, i.e. $RMSEC = 5.4\%$, in comparison to acceptable $RMSEC = 2.4\%$ for population II. In general the RMS is higher for both models as compared to the inline tests results (fig. 22.1). Presumably, this is related to an enhanced sampling error using only a limited number of spectra from spatially confined cubic probes as compared to the much broader inline data acquisition with hundreds of thousands of spectra per measurement which allow averaging out intrinsic sample scatter. It has to be noted that in this study the natural scatter of dry matter values across a 1×1 cm spatial region of interest on a potato slice was at best of the order of 2% and could be as high as $\sim 10\%$ if moisture inhomogeneities such as water channels were part of the inspection region. This can also be seen in fig. 22.4 showing the spatially resolved moisture levels of a potato slice as measured with the *Sherlock Observer*. Peculiar moisture inhomogeneities in the fruit cross section as well as the fact that potato tubers show a dry matter gradient from the centre to the outer rims are obviously traced by the NIR HSI online measurement.

Validation runs were done by measuring several 1 ccm raw potato cubes with the machine in comparison to the SMART device. Table 22.1 shows the outcome of these tests in terms of several statistical figures

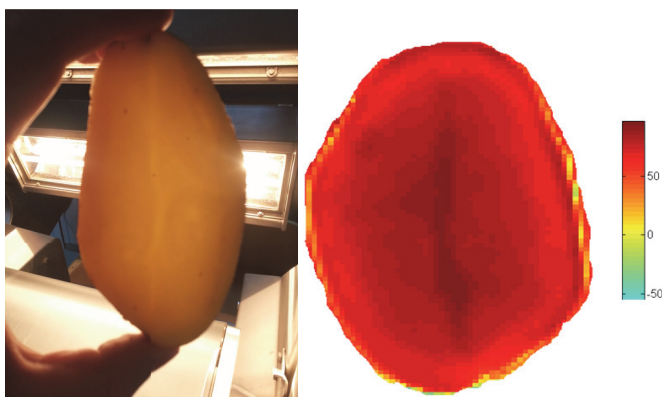


Figure 22.4: (left) Potato slice including water channels as seen with the naked eye. (right) Same slice viewed as color coded moisture map (in % water content) as seen with *Sherlock Observer* (crf. [6]).

of merit. According to this the two models show a stable behaviour in the validation in comparison to the calibration with respect to the magnitude of model deviations (RMSEC and SEP, respectively). However, both calibrations lead to a prediction bias of the order of 3%. While such a bias if purely Gaussian and statistical in nature can always be corrected for by e.g. a linear correction of the multivariate parameters and has thus no impact for a working machine, it hints to the fact that chemometric calibrations in quantitative applications can be sensitive to changes in process parameters other than the interesting target variables. Presumably the bias encountered here goes back to alterations of the optical setup and feeder mechanics of the *Sherlock Observer* in-between calibration and validation runs. It has to be noted that a comparison of the RPD values between the two models with respect to the poor $RPD = 0.7$ for model I has to be taken with care since the calibration set I had a very sparse coverage and the validation set was limited in sample numbers. As shown in [9] dataset distributions deviating from normality can introduce a bias to the RPD value. With a $SEP = 2\%$ calibration II seems suitable for an inline application in the production line. The fact that model I features a somewhat poorer $SEP = 5\%$ with a systematic bias towards drier samples, the sparse midrange dry matter

Table 22.1: Statistics of validation runs for population I and II calibration. Shown are the model bias, the standard error of prediction (SEP), standard deviation of prediction (SDP), standard deviation of reference (SDR) and the ratio between standard error of prediction and standard deviation of the reference (RPD), respectively.

	Model I	Model II
BIAS [%]	-3.5	-3.0
SEP [%]	5.3	2.1
SDP [%]	4.5	3.0
SDR [%]	3.7	4.2
RPD	0.7	2.0

value data coverage (right plot in fig. 22.2) and the observation that samples pre-treated in the desiccation chamber show deviations from the usual spectral potato signature presumably due to changes in the fruit matrix (e.g. gelatinization effects) would suggest that it is favourable to only use natural potato probes as done for model II and stick to a more confined value range as typically present in production lines (e.g. 10 – 35% dry matter). In general any calibration model will need thorough recalibration and maintenance over the whole life cycle of the production process due to changes in related parameters ultimately also impacting the HSI NIR spectra. [10].

3 Conclusions

NIR-HSI is by now an established technology in different industries such as food, mining, recycling and pharmaceuticals. It has been developing from a mere inspection technology to a fast, reliable and spatially resolved quantitative analysis method suitable as PAT tool. Results in applying these innovative sensor systems for the inline measurement of dry matter levels in potato processing suggest its suitability for industrial use with a precision in the %-regime being presumably only limited by the sampling statistics of the training set. With the technology being fit for commercialization the first machines operating with according functionalities (dry matter detection, inference of residual peel amounts) are presently being deployed in the field [6].

Acknowledgments

I would like to thank the staff of INSORT GmbH for their great assistance and support with the measurement setup. Valuable discussions with E. Leitner (TU-Graz) and W. Märzinger (i-RED Infrarot Systeme GmbH) were highly appreciated.

This work was in parts carried out within the project *Forschung-Technologie-Innovation (FTI) - Inline Food Analyser* and was supported by the Austrian research council FFG under the project number 834298.

References

1. K. H. Esbensen, P. Geladib, and A. Larsenc, "The goal for my on-line nir is to be an automated copy of the lab technician," *NIR news*, vol. 25, no. 8, pp. 30–32, 2014.
2. <http://www.evkbiz/en/products/helios-hyperspectral-imaging-camera-system.html>.
3. M. Kerschhaggl, "Hyperspectral imagery for real-time quantitative inline analysis," *Sensor-Based Sorting 2014*, GDMB Verlag GmbH, vol. 135, 2014.
4. M. Kerschhaggl, W. Märzinger, E. Leitner, N. Haar, M. Zangl, M. Jeindl, and P. Kerschhaggl, *Inline HSI food inspection and concentration measurements of pharmaceuticals – a report from an industrial environment*. Karlsruhe: KIT Scientific Publishing, Karlsruhe, 2013.
5. S. Wold, M. Sjöström, and L. Eriksson, "Pls-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.
6. <http://potatopro.com/news/2014/insort-sherlock-analyst-sets-new-standards-process-monitoring-potato-processing>.
7. <http://www.cem.com/smart-turbo.htmlg>.
8. P. P. Mortensen, "Myth: A partial least squares calibration model can never be more precise than the reference method," *NIR news*, vol. 25, no. 3, pp. 20–22, 2014.
9. K. H. Esbensen, P. Geladi, and A. Larsen, "The rpd myth..." *NIR news*, vol. 25, no. 5, pp. 24–28, 2014.
10. M. B. Mercader and A. R. Puigdomènech, "Near infrared multivariate model maintenance: the cornerstone of success," *NIR news*, vol. 25, no. 7, pp. 7–9, 2014.