



White Paper

Visible and near-infrared spectroscopy for the measurement of fresh-produce quality parameters

Non-destructive fresh-fruit-quality analysis is of wide interest to the stakeholders involved at various stages of the fresh-fruit chain. The need for quality assessment ranges from the decision-making during harvest to monitoring fruit ripeness during controlled storage at high-end fruit-storage warehouses [1]. Several non-destructive sensing techniques are available for fresh-fruit-quality analysis; however, most of them are still restricted to either research laboratories or are not cost-effective to be defined as a profitable solution for fresh-fruit analysis [2, 3]. A key technique being promoted from the past two decades is visible and near-infrared (Vis-NIR) spectroscopy (i.e., 400–2500 nm) [1, 4]. This is an optical technique that relies on the interaction of electromagnetic radiation (400–2500 nm) with the fruit. Basically, when the electromagnetic radiation interacts with the fruit, based on the physical and chemical properties of fruit, the electromagnetic radiation

(EMR) gets attenuated. Further, with advanced data analysis, the attenuation in the EMR radiation can be modelled as calibrations and later used for the prediction of real-time properties. Vis-NIR spectroscopy allows a non-destructive prediction of fruit pigments (such as chlorophyll, anthocyanin) in the visible part of the EMR, and the chemical properties such as moisture, sugar, fats and protein in the NIR part of the EMR [5, 6]. Further, the physical structure of the fruit appears as the scattering over the complete Vis-NIR signal [5, 6].

This report describes the state of the art of Vis-NIR spectroscopy for the use in the fresh-fruit chain. The principles of NIR spectroscopy are described to serve as a practical guide to successfully employ the power of Vis-NIR spectroscopy as an affordable and non-destructive method to measure fruit-quality parameters with the goal of making a more efficient fresh-fruit chain and delivering optimal quality fruits to the consumer.

CONTENTS

| | | | | | |
|---|---|---|---|---|----|
| 1 | Introduction | 1 | 6 | Applications in fruit-quality measurements | 7 |
| 2 | Basic concepts and principles of visible and NIR spectroscopy | 2 | 7 | Reference methods for calibrating NIR sensors | 9 |
| 3 | Measurement modes | 2 | 8 | Summary | 10 |
| 4 | Equipments | 3 | 9 | Further reading | 11 |
| 5 | Data modelling | 4 | | | |

2 Basic concepts and principles of visible and near-infrared spectroscopy

Vis-NIR spectroscopy covers a spectral range of 400–2500 nm. The visible part of the spectrum is from 400 nm to 700 nm, and the NIR part is from 700 nm to 2500 nm, the range between the visual spectrum and infrared (heat) radiation. A clear boundary between visible and near-infrared is not defined, and in many scientific literatures it has been debated. A main part of Vis-NIR spectroscopy is irradiation of samples with the light in the spectral range of 400–2500 nm. For visible regions, standard white light can be used, while for the NIR, halogen light sources are required. The light interacts, penetrates, scatters, absorbed and reflected from the sample or product tissue. Based on the physicochemical state of fruit, the information about chemical and physical properties of the sample or product is entangled in the reflected signal.

Chemically, the NIR irradiation interacts with molecules containing C–H, O–H and N–H molecular bonds. Therefore, it can provide content information about moisture content (MC) and biomolecules such as protein, starch, sugars, and oil. Differences in scattering (reflected light) of the NIR irradiation can provide information about physical properties such as surface texture, firmness, or particle size. These underlying properties are also determinants of fruit quality. Therefore, the NIR measurements indirectly determine quality aspects such as ripeness, firmness, and taste of fruits. Vis-NIR spectroscopy is popular for its ease of use with little or no sample preparation and the strong decline in the cost of spectrometers and sensors, enabling also in-field and in-line measurements next to laboratory-based measurements.

3 Measurement modes

All spectrometers consist of three major components: a light source (often a tungsten halogen bulb), a wavelength selector and a photodiode array sensor. The position of the light source in relation to the measured product and sensor determines the measurement mode. Three measurement modes can be distinguished: transmission, reflection and interaction (Figure 1). In reflection mode, the light source and sensor are positioned at an angle, often 45° relative to the product. In transmission mode, the light source and sensor are placed at opposite sides of the product (light is transmitted through the product). In interaction mode, the light source and sensor are positioned in parallel using a so-called bifurcated fibre optic cable near the product. The selection of the measurement mode depends on the application and product. As a rule of thumb: light intensity declines exponentially with the distance travelled through the product. In practice, the light will not penetrate more than 2–7 mm into fruits. For this reason, fruits are measured mostly in reflection or interaction mode due to the dense flesh or thick skin. The limited penetration depth of light into fruits when measured in diffuse reflection or interaction mode limits the possibilities for measuring internal fruit quality. The penetration depth of light in apples has been estimated at 2–7 mm depending on wavelength: 5–7 mm for light between 700 nm and 900 nm, and 2–4 mm for light between 900 nm and 1900 nm. In our own experience, this is similar for mango fruits. Avocado skin seems to block most of the light from

penetration into the fruit tissue. Otherwise, measuring dense fruits or with thick skin would require high halogen light intensities for sufficient penetration depth. However, the heat generated by such a light source may damage the fruits. An alternative approach may exist in using strong light pulses (short duration) for transmission measurement, enabling sufficient light penetration without applying excessive heat.

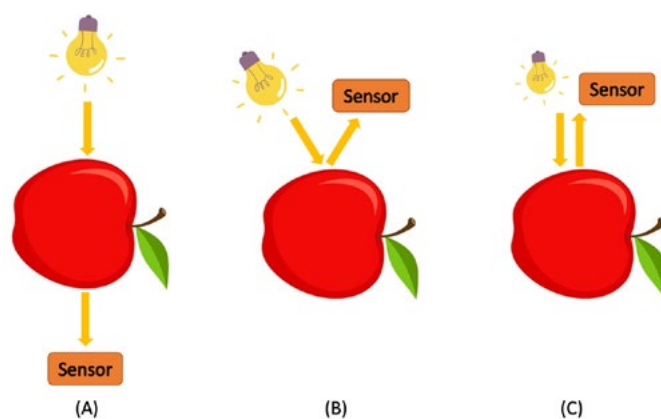


Figure 1 Measurement modes: (A) transmission, (B) reflection and (C) interaction.

4 Equipments

4.1 Laboratory and handheld equipments

Vis-NIR spectrometers are available from a large range of manufacturers, amongst others [Avantes](#), [Ocean Optics](#), [Felix Instruments](#), [Foss Analytics](#), [Zeiss](#) and [Malvern Panalytical](#). Spectrometers can be an integrated unit for the measurement of liquids and bulk materials.

Spectrometers can also be fibre-optics-based for the measurement of separate (non-bulk) products (e.g., fruits) either as a laboratory setup or integrated as in-line sorting technology. Handheld units are nowadays readily available for scanning of products on location or in the field. One company, [Consumer Physics](#), is even offering handheld spectrometers for the consumer, called SCIO, that are operated through a cloud-based app on a smartphone.

Spectrometers also come in different wavelength ranges, some also including the visual range of the spectrum (around 350–700 nm), resolutions and signal to noise ratios (SNRs; a poor SNR means that the signal is “overshadowed” by noise, rendering the measurement useless). Spectrometers with a large wavelength range, high resolution and excellent SNR properties will cost more compared with lower-end models with smaller wavelength ranges, lower resolution, and inferior SNR properties. For instance, the [Foss NIR spectroscopy DS2500](#) has a wavelength range of 350–2500 nm with 0.5 nm spectral sampling (i.e., a reflection measurement is performed each 0.5 nm between 350 nm and 2500 nm) and state-of-the-art SNR. By comparison, a [Felix F-750 handheld spectrometer](#) covers a wavelength range between 310 nm and 1100 nm with a spectral sampling of ~3.5 nm, with less optimal SNR performance (much noise is observed in the lower and upper part of the wavelength range specified). Good SNR performance is generally more important than resolution. A large wavelength range is desirable in two ways:

- Including the visual range can be beneficial, since colour information may support building robust models, as, in several fruits, the colour of the outer skin is highly correlated to the ripeness and maturity stage.
- Additional chemical and physical information may be embedded in the wavelength range between 1100 nm and 2500 nm.

Also, here a trade-off can be made: an NIR device with a range up to 1700 nm with good SNR performance is best chosen over an NIR device with a range up to 2500 nm with worse SNR performance. In literature, little is reported about the optimal wavelength range for fruit-quality assessment. A large wavelength range may not be needed in all applications. With increasing wavelength, the penetration depth of light into fruit tissue decreases. Several works have showed that the Vis-NIR wavelength range, i.e., 400–1000 nm, is sufficient for prediction of key chemical traits such as moisture, soluble solids and pigments in the fruit surface [5, 7–9]. The Vis-NIR spectroscopy can also be performed in imaging modality, which allows the spatially resolved optical properties of fresh fruit to be estimated. The imaging modality of spectroscopy is called hyperspectral imaging. Further, several hyperspectral cameras are available (e.g., Specim, Finland) depending on the spectral range desired. Further, cameras are available for both line-scan and snapshot modality. although practical application of snapshot cameras is currently limited to remote sensing applications. Further, for an in-line sorting facility where the fruit are measured on a continuous moving line, the line-scan modality of hyperspectral imaging is more practical. Like point spectroscopy, hyperspectral imaging can be performed in multiple models such as diffuse reflection, interaction and transmission. For example, in Figure 2, an example of performing hyperspectral imaging in three different modes is presented.

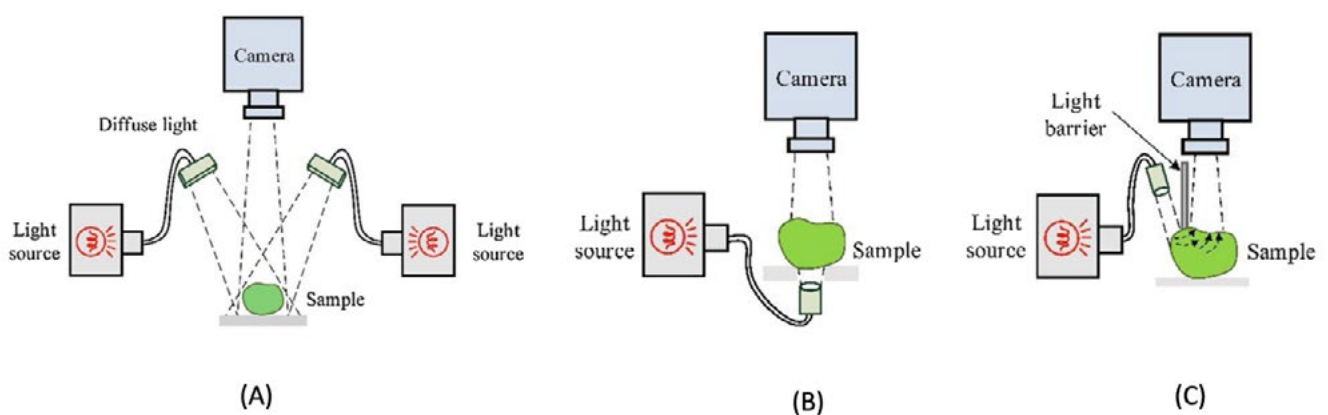


Figure 2 Different modalities of hyperspectral image acquisition [10]. (A) Diffuse reflection, (B) transmission and (C) interaction.

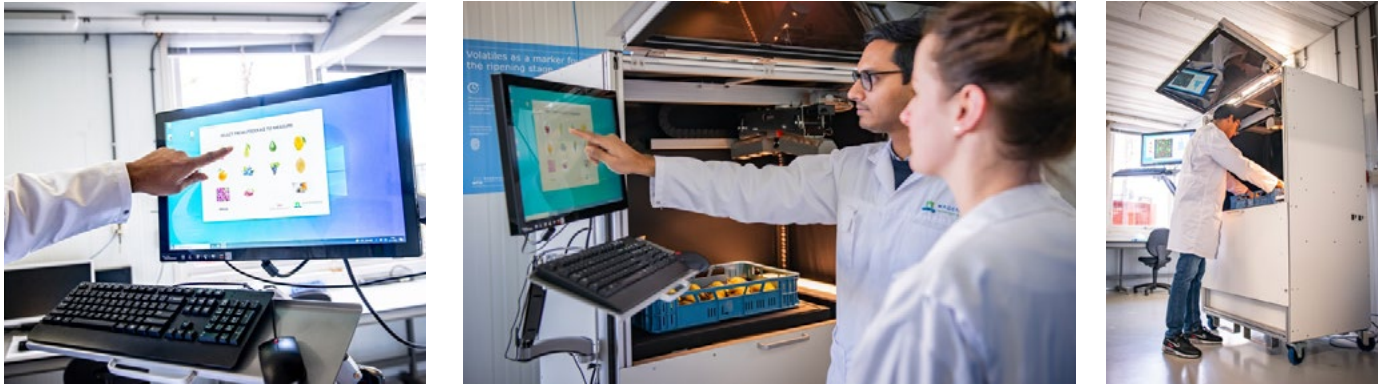


Figure 3 A new spectral imaging setup at Wageningen Food and Biobased Research to analyse fresh produce.

The diagram is taken from reference [10] and modified for this report. In the easiest approach, the hyperspectral imaging can be performed in diffuse reflection mode (Figure 2A), where the incident light and the sensor are above the sample. Such an integration is also easy to deploy in real-world in-line sorting facilities, as it requires minimal adjustment to the existing sorting lines. However, a main limitation of the diffuse reflection mode is that it mainly captures the surface properties of the fruit that are sufficient to predict chemical properties such as MC and soluble-solids content (SSC). However, measurement in diffuse reflection mode may not capture the internal changes in fruit, as can be captured with transmission mode (Figure 2B). The main challenge with the transmission mode is related to its integration in the in-line sorting facilities, as it required the illumination and camera to be at opposite sides of the fruit. To deal with the drawback of diffuse reflection and combine the advantages of both modes, hyperspectral imaging can also be performed in the interaction mode (Figure 2C), where the light and camera are placed on the same side of the fruit, as in diffuse reflection implementation. However, the camera view is set at a slightly spatial offset such that instead of measuring diffuse reflection, the light that eventuates after interacting with the internal structure of the fruit is measured.

4.2 In-line sorting equipments

Companies such as Aweta, Ellips and Greefa develop, build, and sell in-line sorting equipment for fruit grading. Third-party Vis-NIR spectroscopy instruments (both point and hyperspectral cameras) are usually integrated in these sorting lines. Fruits are measured in diffuse reflection or transmission mode to predict chemical properties and detect defects and internal quality aspects. Nevertheless, no data have been provided about the accuracy of the sorting applications for the various fruit species. For sorting lines, line-scan imaging modality of Vis-NIR spectroscopy (hyperspectral imaging) is highly interesting [5, 6]. Apart from estimating the fruit properties, hyperspectral imaging allows the spatial distribution of fruit properties to be explained. Such spatial distribution allows any localised anomalies on the fruit to be detected. Several applications of hyperspectral imaging related to fresh-fruit analysis can be found in a recent application review [10]. Recent at WUR, a new spectral imaging setup for fruit quality analyse was developed which allows analysis of wide range of fresh fruit for traits such as MC, SSC, size and colour. The setup can be understood in Figure 3.

5 Data modelling

Vis-NIR sensing technology alone is of no use. After successful implementation of NIR sensors (either point sensor or a camera), the main step is to calibrate the NIR sensor that allows the property of interest to be predicted. Hence, before the application of NIR spectroscopy, a pre-experiment is required where several fruits are measured with spectroscopy and with reference properties such as the measurement of MC and SSC. Later, such data

generated are used to develop linear or non-linear regression models that can be used in future application. A calibration of spectral sensors is required, as each sensor is different or has small differences that limit the generalisability of one model to be used on all future instruments. However, once a sensor is calibrated, it can be easily used to predict properties of interest.

The data modelling for spectral data requires some extra efforts compared with traditional machine learning approaches. A reason is because Vis-NIR data capture the chemistry information of the samples, and it is always useful to understand the background chemistry of the models. For example, when a model related to MC prediction is to be made, then the aim of the data models is to give high regression weights to the wavelengths related to the OH bonds. Several of the steps involved in making useful Vis-NIR models are highlighted as follows.

5.1 Interpretation of Vis-NIR spectra

The first main step of the Vis-NIR data modelling is to check the spectra to determine whether they make any sense. This is required to see whether the sensor measured the signal correctly or there are any artefacts in the data. A typical Vis-NIR reflection signal from fresh agri-food material such as fruit and leaf is presented in Figure 4. The signal shown in Figure 4 is taken from reference [11]. In the signal, at first in the spectral range of 400–700 nm, all information related to colour can be found. For example, if the outer skin of the fruit is green, then peaks related to green colours can be found in the visible range. After the visible range, a high reflection is noted in the spectra range of 700–1300 nm. In this range, the main signal is masked by the high reflection due to water, but overtones of chemical-bond vibrations are noted. After 1300 nm until 2500 nm, overtones of all major chemical bonds, such as OH, CH and NH, that are directly related to several fruit properties are presented.

Water is the main constituent of plant and fruit tissues (~80–90%). Hence, the information embedded in the NIR spectra is dominated by water peaks. In addition, these

spectra contain information about the chemical composition and physical properties. NIR spectra are a collection of broad shoulders and peaks. The position of the peaks is influenced by various effects such as temperature, instrument and scattering effects, all wavelength dependent. In general, no specific absorption bands can be directly assigned to the studied properties. Hence, chemometric and machine learning approaches are employed to correlate (or calibrate) the studied reference values (response variable), for instance, water content, ripeness, or degree Brix, to the complex information embedded in the NIR spectra (predictor variable). However, the visible part of spectra has distinctive peaks for different pigments that can be related to pigments such as chlorophyll, anthocyanin and carotenoid.

5.2 Pre-processing of spectral data

After the visual inspection of the spectral signals and confirmation that the signals look fine, the next step in data modelling is to select or identify a proper pre-processing approach to spectra. The pre-processing of spectral data is required, as these data consist of a mix of light scatter and absorption. The light scattering is the result of the complex interaction of light with the internal structure of fruit, while the light absorption is caused by the chemical constituents in the fruit. The scattering and absorption information interfere with each other while modelling. Hence, it is highly recommended in NIR data processing to pre-filter out the scattering information, especially when the aim is to make models for the prediction of chemical composition. Such prediction based solely on absorption information led to robust models that did not vary according to differences in the physical structure of fruit. To reduce the effect of scattering, several techniques are available, such as multiplicative scatter correction, standard normal variation, variable sorting for normalisation, and second derivatives. Recently, a review work summarised all major approaches regarding spectral data pre-processing [12]. The optimal combination of pre-processing techniques is partly a matter of experience or can be determined by comparing different combinations of methods concerning the calibration-model performance. However, there are so many different pre-processing approaches available in the domain of NIR spectroscopy that the user is often trapped in the loop of trial and error. To solve this problem, recently the concept of pre-processing approaches was proposed in the domain of chemometrics [12]. The pre-processing fusion approaches allowed an ensemble learning to be performed with differently pre-processed data such that the model can learn the complementary information from each pre-processing stage to boost the NIR models. Two main approaches developed in this direction are the sequential and parallel orthogonalized pre-processing fusion approaches, called SPORT and

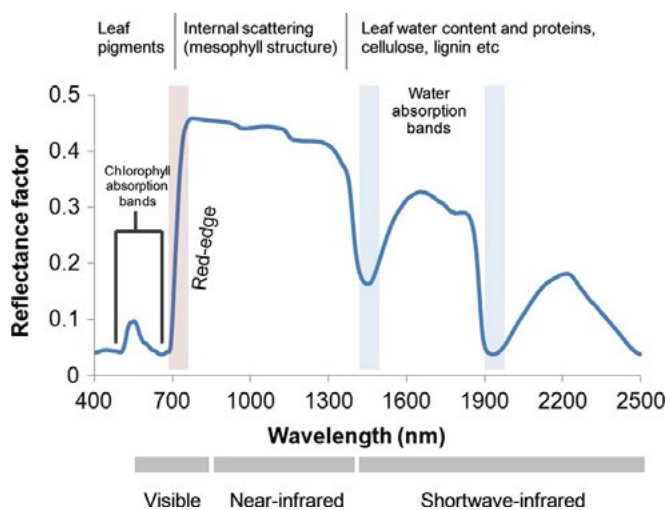


Figure 4 An exemplary visible and near-infrared spectrum of fresh agro-food material [11].

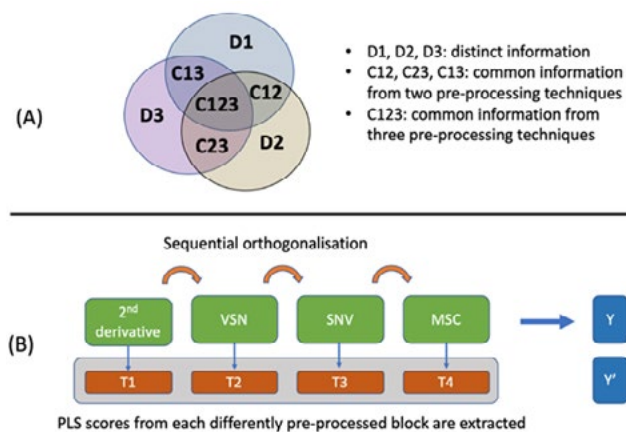


Figure 5 Schematic of parallel (A) and sequential (B) pre-processing fusion approaches. T1–T4 are the scores from sequential partial least squares (PLS) regression [17].

PORTO, respectively, for short [13, 14]. A schematic of the SPORT and PORTO modelling approaches is shown in Figure 5. Both these approaches have outperformed all previous chemometrics approaches for several challenges such as fruit, fuel, tablets, meat, and seeds [13–19].

5.3 Predictive modelling for spectral data

Vis-NIR spectroscopy can be used for performing several tasks; for example, it can be used for predicting continuous variables such as chemical component, i.e., MC and SSC. On the other hand, the Vis-NIR data can be used to perform classification of samples. Hence, both regression and classification modelling approaches are used widely for Vis-NIR data modelling. A basic underlying challenge with the Vis-NIR data is the highly correlated variables that require special attention to develop robust models. Typically, latent-space modelling approaches are used to decorrelate the variables and capture the underlying variability most related to the property of interest. A common approach to process Vis-NIR data is with partial least squares (PLS) regression and discriminant analysis. The PLS approach performs a dimensionality reduction of the Vis-NIR data before the development of multi-linear regression models on the transformed space. The main advantage of the PLS approach is related to its capability to decorrelate the data, thus reducing the collinearity in the data [20–22]. PLS-based methods are the gold standard for analysis of spectral data in chemometrics; nevertheless, recently, deep learning methods have appeared to be more powerful than the PLS models [23, 24]. However, the main challenge with such deep learning models is the requirement for large data sets on which complex and non-linear patterns can be attained. When there are small data sets, the PLS-based approaches can perform equally well.

5.4 Calibration vs cross-validation vs independent test

A main step in training the models is to tune the hyperparameters and validate the model on an independent test set to check its robustness. Hence, before NIR modelling, usually the data are partitioned into calibration and independent test sets. Unlike machine learning, random data split is not highly recommended in NIR data modelling. However, two main techniques, i.e., Kennard–Stone and DUPLEX algorithms, are used widely to split NIR data. The calibration of NIR models, for example, requires tuning the total number of latent variables (LVs) before making the final calibration. The total number of LVs are usually explored with the cross-validation procedure, which led to an error plot as a function of LVs. The error plot usually carries a shoulder pattern that allows optimal LVs to be selected. Once the optimal LVs are identified, the model can be calibrated and tested on the independent test set. Testing a model using an independent test set is highly recommended, as the main criterion in judging the accuracy of a model is whether it can be used on unseen data.

5.5 Model performance

The performance of the NIR models is expressed as the root mean squared error of cross-validation and prediction as measured by cross-validation and testing it on an independent test set. For fruit traits such as MC and SCC, a root mean squared error of less than 1% is usually acceptable. Model performance depends on not only precision but also robustness. Model robustness can be defined in terms of performance of a model across a larger set of samples, which include more biological variation, such as growth location, season and cultivars. In other words, how well does a model perform for different batches of the fruit of different origin and cultivar? Model robustness is of key importance for successful application. A natural solution to attain robust models is to measure as wide a variation of samples as possible. In scientific literature, a lot of attention is paid on developing multi-season, multi-cultivar models to increase the robustness of NIR models [25, 26].

5.6 Transferring calibration models

A key drawback of NIR models is that they are highly specific to the sensor, and often a model made on one NIR instrument cannot be directly used on a new NIR instrument. Different NIR instruments can have differences in sensitivity, resolution, and wavelength calibration. Of course, when the NIR instruments are similar, the main difference between NIR spectrometers is the varying wavelength calibration. In other words, a certain peak may appear at a slightly shifted wavelength in the spectrum when comparing two spectrometers. Further, each type of spectrometer comes with different

optics, sensor, or light source. Also, changes in the environmental conditions, especially temperature during measurement, may cause a drift in the wavelength response [27, 28]. For this reason, calibration models cannot be used with spectra generated with a second spectrometer from the same type and specifications (without proper precautions), let alone a spectrometer of a different type. This inherent obstacle of NIR spectroscopy can be overcome in two ways: (1) instrument-alignment correction, and (2) calculating spectral difference transfer functions.

For instrument-alignment correction, spectrometers are calibrated for their wavelength responses with the help of precise (master) calibration standards. This ensures that spectrometers of a certain model or type perform within tight tolerances, and calibration models can be transferred. Here the hardware is fine-tuned.

Calibration transfer may also be implemented at the level of the spectral data and corresponding calibration models through calculating transfer functions between spectra. This involves the measurement of a calibration standard or set of samples with a so-called parent or primary spectrometer and a child or secondary spectrometer. Using the same calibration standard on both spectrometers can thus enable the calculation of a

wavelength transform function. With this transform-function, data from the parent spectrometer can be converted into data as if measured with the child spectrometer. The wavelength transform-functions compensate for inconsistencies between spectrometers regarding the position of the different wavelengths. Several mathematical methods for developing transform-functions have been described in the literature [29]. A main challenge to the currently existing methods for the calibration transfer is that they rely on the measurement of standard samples. Such measurement of standard samples involves appraising a set of the same fruit samples on two different instruments. Thus, since the fruit is same, any difference in the spectra of the fruit measured on two instruments is related to the differences in the two instruments. Measurement of new fruit requires extra time and efforts; hence, this is not a welcome approach in portable spectroscopy. To eliminate the requirement for new measurements, recently new standard-free approaches to calibration transfer have emerged in the domain of chemometrics [30]. The standard-free approach allows the re-use of models on a new instrument without any new reference measurements. Several applications of such standard-free approaches have emerged in the domain of fresh-fruit analysis [9, 31].

6 Applications in fruit-quality measurements

6.1 Academic literature

A very large body of academic literature, encompassing hundreds of studies, can be found on the use of NIR spectroscopy to study fruit-quality aspects [1, 4, 32, 33]. These reports show reasonable to good calibrations for fruit-quality attributes. Good correlations can be found for both chemical as well as physical properties. The key chemical properties that are widely predicted with NIR spectroscopy are MC and the SSC. Both these properties can be predicted with an error of <1% [7, 34]. Works have also presented the usefulness of NIR spectroscopy for predicting firmness in fruit; however, predicting firmness is a slight challenge compared with investigating the chemical components, as the firmness models are based on secondary correlations captured by NIR spectroscopy [8]. Like any other scientific field, there are far more applications existing in academic scenarios than in real-life practical applications. Several reasons can be identified to explain this:

- The experiments underlying academic studies have been carried out under well-controlled laboratory

conditions, constant temperature and ambient MC.

These occur often in a limited time/budget project that ends before such factors can be finalised or made into a practical real-life application.

- All measurements have been performed with a single spectrometer with ideal measurement setups. Usually, to perform a real-life application, a new spectrometer needs to be integrated in the system; however, because of the limitations of academic studies in terms of time/budget, this is usually not possible.
- In some academic studies, the collections of fruits may be rather homogeneous with respect to variety or cultivar, developmental stage, growth location and season. Calibrating any of these fruit-quality parameters for a batch of fruits from the same variety, location, season, and developmental stages leads to a model with limited scope of application. Building a robust model that can be applied to fruits at different developmental stages, from different cultivars, seasons, or locations, is a much greater challenge in not only modelling but also organizing such experiments. Several publications can be found on such

homogeneous fruit collection. However, recently, academic studies focussing on measurement of samples from diverse cultivars, seasons and maturity stages are emerging [7, 25, 26]. Such studies have shown that modelling a wider biological variation increased the robustness of predictive models [25, 26].

6.2 Towards real-life applications

Recent advancement in sensor technology and data analysis methods has shown a tremendous increase in academic literature. However, applications of NIR spectroscopy in the real-life working environment such as fruit-sorting lines are still in progress. This very much depends on not only the type of fruit but also the type of application, e.g., SSC, DM or maturity. Some further explanation is needed, as follows.

6.2.1 Direct vs indirect measurements

Fundamentally, the NIR irradiation interacts with molecules containing C–H, O–H and N–H molecular bonds. Therefore, it can provide content information about moisture and biomolecules such as protein, starch, sugars and oil. Differences in scattering (reflected light) of the NIR irradiation can provide information about physical properties such as texture, firmness, or particle size. These chemical and physical attributes can also be indirectly related to quality-related measurement, such as firmness or maturity, within the context of fruit quality [1, 4]. Measuring moisture in fruit is a rather direct method with strong definable water bands in NIR spectra. Measurement of SSC is rather less direct. The SSC in fruits is a mix of dissolved sugars, proteins, pigments, vitamins, minerals and phenolics. SSC is indicated as one of the most important parameters for describing sweetness. Determining SSC is performed mostly by a refractometer. Although a large part of the SSC value is accounted for by fruit sugars, other compounds, such as anthocyanin, may influence the refraction index strongly. This may be the case for grapes and strawberries. Even if the fruit sugars account for the biggest refraction index, the relative ratios between sucrose, glucose and fructose are influenced strongly through biological variation (age of the fruit, growing locations, etc.). This composition is also influenced by many factors such as age, growth and climate conditions in which the fruit is cultivated. Fruit firmness or maturity is the most indirect measurement within the applications mentioned here. Fruit maturity or firmness is a complex arrangement of MC, degree Brix, DM and physical properties, e.g., texture, colour or gloss of the fruit skin, again influenced by biological and climatic factors.

6.2.2 Biological variation

Biological variation can be defined as any variation in the to-be-measured fruit-quality parameters influenced by

factors such as variety, cultivar, fruit-maturity stage, storage conditions, growing location, positional effects within a tree or plant and season. This list does likely not fully account for all imaginable biological variation. Biological variation strongly influences the robustness and precision of the calibration models, often in a negative manner. As a rule of thumb, it can be stated that the influence of biological variation on the robustness of the calibration model increases with the indirectness of the measurement.

To deal with biological variation, calibration models must be based on large data sets comprising NIR spectroscopy measurements from different varieties, seasons, orchards, or any other external factor that may underlie the biological variation. Many times a calibration model based on fruits of location “A” do not perform well when applied for the same types of fruits and cultivars from location “B”. The same is true for different seasons. When datasets of different locations and seasons are combined, the overall calibration model performs more robustly. In other words, the prediction error is bigger when compared with a calibration model specific for a location or season (trade-off). This also highlights the importance of having a variety of specific models for reliable predictions. However, recent advancement in new chemometric methods such as domain adaption and the removal of standard-free external effects has given new hopes to NIR modelling. Basically, these new methods allow any non-relevant differences to be removed from the data of different seasons or cultivars, which improves the performance of NIR models [9, 31]. However, the application of these new approaches is still emerging in the domain of fruit analysis [9, 31]. On the other hand, utilising advanced non-linear methods such as artificial neural networks (ANN) can also help in dealing with such problems by capturing more detailed features invariant to subtle biological variations. The biological variations do not bring a completely new peak in the NIR spectroscopy data, they will show just some differences in intensities at designated wavebands or some shifts in peaks along the wavebands. However, modelling the completely raw spectra with highly non-linear ANN models can help to solve this problem to some extent by modelling also different scattering and diffuse reflection properties that are discarded in multivariate statistical analysis with the help of different pre-processing and data normalisation. Further, a future technique – deep learning – could be a solution to this.

Collecting a large dataset is good to incorporate robustness to biological variation. The “size” of the needed dataset may vary per situation. Much depends on the fruit-quality parameter to be measured, the number of varieties and/or production locations that can be expected or the required precision of the calibration model.

Complex or indirect quality parameters, such as fruit maturity, need a dataset based on a larger biological variation. Here, it would help if the company wishing to apply a non-destructive NIR spectroscopy-based measurement for fruit quality has tight control over its “ecosystem”. For instance, if the company in question deals with a limited number of varieties per fruit sort from a limited number of growing locations, it may be enough to collect data from these varieties x locations for 2–3 seasons. The robustness of these calibration models may be challenged over time and likely need maintenance. This can be performed by routinely (for instance every season) incorporating new information into the calibration models by testing several fruits with the reference method for that specific fruit-quality parameter. This will also prove whether a calibration model needs an update when factors relating to a new set of fruits are not precisely predicted.

6.2.3 Academic research vs real-life applications

Academic research into NIR spectroscopy aims at the highest possible correlation between spectra and the attributes of the studied fruits. This is done mostly

through chemometrics and machine learning techniques. In real-life applications, this may not be necessary. For instance, classification in distinct quality categories may be more useful than calibration. The best approach depends on the way NIR spectroscopy is used. Optimisation of the fresh-fruit supply chain by estimating shelf life or ripeness, hence reducing waste, is a different application to prediction of SSC. Classification of fruits in different ripeness classes with a lower precision, compared with the academically determined R^2 of 0.95 for regression, may still be economically viable if this leads to the reduction in the number of discarded fruits. Thus, the application of NIR spectroscopy in the fresh-fruit supply chain is more about probability and classification than exact prediction through regression. Some examples that can be found in literature were based on NIR spectroscopy, whereby the biological age of fruit is estimated, expressed as a probability for a certain developmental stage/biological age. This is used to harvest fruit at an (estimated) optimal stage for fruit development and subsequent shipping.

7 Reference methods for calibrating NIR sensors

NIR spectroscopy calibration models can only be as precise or robust as the quality of the reference measurement allows. To determine reference values for quantitative parameters such as soluble sugar content, these measurements may be destructive but precise. For quality-based attributes such as fruit ripeness, these measurements are likely to be more complex. Only if we can measure ripeness reliably and reproducibly can it be successfully correlated to NIR spectroscopy.

7.1 Measurement of moisture and soluble-solids content

MC and SSC are the fruit traits predicted widely with NIR spectroscopy. However, for calibration purposes, initially, destructive MC and SSC measurements are required. MC is measured by a combination of weighing and oven drying. Usually, after NIR measurement, a sample is extracted from the same location for moisture measurements. The fresh weight of the samples is weighed, the samples are dried in a hot air oven for 4–7 days, depending on the type of fruit, and the dried weight of the samples is noted. After noting the fresh and dried weight of the samples, the total moisture in percent can be calculated. Unlike moisture measurement, SSC measurement is fast and can be performed with refractometers. For SSC measurement, some juice from the fruit is extracted and, with the help of a refractor, the

total degree Brix in the fruit juice is estimated. A key point to consider is that, for both MC and SSC measurements, the fruit sampling should be performed from the same location of NIR measurement. Measurements at the same locations are required as fruit can be very homogeneous and measuring properties at different places may lead to poor modelling.

7.2 Mechanical firmness measurement techniques

Several mechanical techniques to measure firmness and ripeness attributes of fruits are at our disposal. Roughly, these mechanical measurements can be divided into two types: impact analysis and vibration (acoustical) analysis. With impact analysis, an external force is applied to the fruit, and the responding deformation is measured. Penetrometers and the Firmtech FT7 apparatus are examples of this. With vibration analysis, a fruit is tapped with a small impactor, and the corresponding acoustic response is measured with a microphone. The Aweta firmness tester is a well-known example. Several aspects must be taken into consideration when using either type of measurement. Both types of methods are intended to measure fruit firmness, but they measure different physical properties and are prone to different measurement errors. When both types of methods would be applied to the same fruit, a good correlation is not

always apparent. Impact analysis measures a combination of compression and shear strength, whereas the vibrational measurements characterise the acoustical inertness/elastic properties of fruit tissue. For instance, it has been reported that impact analysis can continue to measure fruit softening to more progressed stages compared with acoustical measurements. Further, impact analysis is insensitive to fruit shape, although the influences of fruit size have been described. Acoustical methods, on the other hand, are best suited for spherical and more solid types of fruits as opposed to irregularly shaped fruits and/or soft fruit. Finally, differential influences of the MC of ageing fruits on both types of methods have been reported. For precise fruit-firmness measurements, these factors should be taken into consideration. Reproducibility of both types of methods should also be considered and, if possible, confirmed. It goes without saying that, in an experimental setup, this includes measuring the same fruit several times not only in a single position but also in different positions relative to the measurement apparatus, especially in the case of irregularly shaped fruits as opposed to spherically shaped fruits.

7.3 Combination of techniques and mathematical models

There may also be alternative approaches for measuring and describing fruit ripening with the aim of increased precision. This can be performed by combining the results

of the two types of measurement; however, also examples can be found that describe the use of mathematical models to express a ripeness index. These models calculate the ripeness index based on, for instance, a combination of SSC, MC, and total titratable acidity also with the help of sensory evaluation panels.

7.4 Systems biology

Nowadays, we can study a plethora of molecules with the help of methods such as liquid chromatography mass spectrometry (LC-MS) and gas chromatography mass spectrometry (GC-MS) that can determine the complete transcriptome (mRNA), proteome (protein) and metabolome (metabolites) of fruit at a given developmental stage. These methods have been shown to describe the physiological state of certain fruit and provide an objective measure of fruit-quality parameters, such as the fruit-ripening stage. Also, here, building robust models including a large biological variation is a challenge. Nonetheless, these methods give a more detailed picture than the industry-standard mechanical methods. With the aid of these technologies, we may be able to maximise the use of NIR spectroscopy as a non-destructive method for determining fruit quality.

8 Summary

NIR spectroscopy is a promising technology for the non-destructive measurement and prediction of fruit-quality parameters. However, employing NIR spectroscopy successfully in, for instance, in-line sorting of fruits requires dealing with different challenges. Potential applications can be tested in a laboratory setup and later to build robust and accurate models to be employed in the fresh-fruit chain; however, data collection of samples covering a wide variation needs to be executed on the actual setup or spectrometer that will be used for the intended application. For instance, if in-line sorting for fruit ripeness is to be performed, data may also need to be collected with the in-line sorting equipment or the model might need to be adapted with advanced statistical methods. A large variation in the model is needed to account for biological variation: however, at post-harvest facilities at [Phenomea](#), a wide variation and chemical and physical properties can be induced artificially. Such artificially induced variation allows development of robust models. To maximize the potential of NIR spectroscopy

and the prospect of successful implementation, for the accurate prediction and measurement of fruit quality, the following points are expected to be essential:

Data collection:

- Collect more data within experiments or in time with a wide set of biological variations.
- Induce artificial variation in the physical and chemical properties of fruit before experiment.
- Include data on, amongst others, varieties, locations, orchards, seasons, and vendors.

Reference methods:

- Create a matrix for the best suitable reference method per species.
- Improve repeatability of measurements – include replicable and clearly designed guidelines.
- Combine techniques or create models with multiple parameters.
- Consider automation of data collection.

Hardware and sensor calibration:

- Ensure proper calibration procedures for each experiment as recommended by the manufacturer of the sensor.
- Use optimally designed optical probes designed for interaction or reflection measurement.
- Use a dedicated stable light source for NIR spectroscopy applications as recommended by the manufacturer of the sensor.
- Consider automation of data collection.

Data modelling:

- Carefully identify outliers before data modelling.
- Use appropriate data-modelling approaches such as the new sequential and parallel regression approaches.

- Identify beforehand whether the task is regression or classification.
- Move towards non-linear data-modelling methods, as the NIR spectra are a result of complex light propagation that might be difficult to model with linear models such as PLS.
- Re-use NIR models between instruments by modelling the transfer functions between instruments.

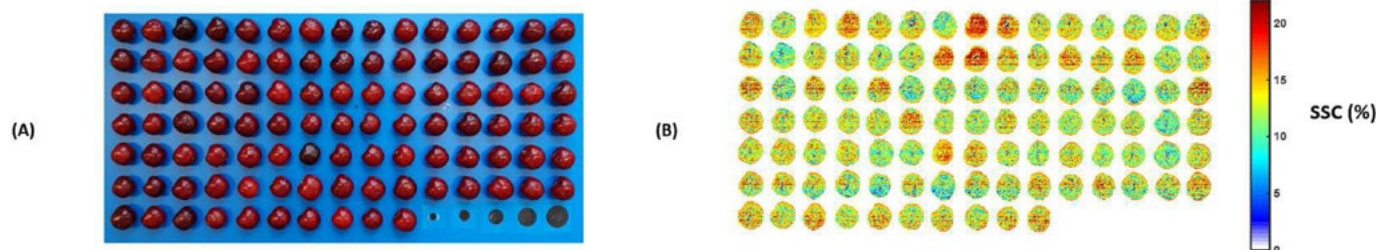
Model maintenance:

- The models should be checked occasionally for their robustness, with a few reference samples.
- If the model seems to diverge from its expected performance, it should be recalibrated or updated.

Further reading

- [1] K.B. Walsh, V.A. McGlone, D.H. Han, The uses of near infra-red spectroscopy in postharvest decision support: A review, *Postharvest Biology and Technology*, 163 (2020) 111139.
- [2] P.P. Subedi, K.B. Walsh, Non-invasive techniques for measurement of fresh fruit firmness, *Postharvest Biology and Technology*, 51 (2009) 297-304.
- [3] K.B. Walsh, Commercial adoption of technologies for fruit grading, with emphasis on NIRS. Information and technology for sustainable fruit and vegetable production, *FRUTIC*, 5 (2005) 12-16.
- [4] K.B. Walsh, J. Blasco, M. Zude-Sasse, X. Sun, Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use, *Postharvest Biology and Technology*, 168 (2020) 111246.
- [5] P. Mishra, S. Lohumi, H. Ahmad Khan, A. Nordon, Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches, *Computers and Electronics in Agriculture*, 178 (2020) 105780.
- [6] P. Mishra, M.S.M. Asaari, A. Herrero-Langreo, S. Lohumi, B. Diezma, P. Scheunders, Close range hyperspectral imaging of plants: A review, *Biosystems Engineering*, 164 (2017) 49-67.
- [7] P. Mishra, E. Woltering, B. Brouwer, E. Hogeveen-van Echtelt, Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach, *Postharvest Biology and Technology*, 171 (2021) 111348.
- [8] P. Mishra, E. Woltering, N. El Harchioui, Improved prediction of 'Kent' mango firmness during ripening by near-infrared spectroscopy supported by interval partial least square regression, *Infrared Physics & Technology*, 110 (2020) 103459.
- [9] P. Mishra, J.M. Roger, D.N. Rutledge, E. Woltering, Two standard-free approaches to correct for external influences on near-infrared spectra to make models widely applicable, *Postharvest Biology and Technology*, 170 (2020) 111326.
- [10] Y. Lu, W. Saeyes, M. Kim, Y. Peng, R. Lu, Hyperspectral imaging technology for quality and safety evaluation of horticultural products: A review and celebration of the past 20-year progress, *Postharvest Biology and Technology*, 170 (2020) 111318.
- [11] H. Croft, J.M. Chen, S. Liang, 3.09 - Leaf Pigment Content, *Comprehensive Remote Sensing*, Elsevier, Oxford, 2018, pp. 117-142.
- [12] P. Mishra, A. Biancolillo, J.M. Roger, F. Marini, D.N. Rutledge, New data preprocessing trends based on ensemble of multiple preprocessing techniques, *TrAC Trends in Analytical Chemistry*, 132 (2020) 116045.
- [13] P. Mishra, J.M. Roger, D.N. Rutledge, E. Woltering, SPORT pre-processing can improve near-infrared quality prediction models for fresh fruits and agro-materials, *Postharvest Biology and Technology*, 168 (2020) 111271.
- [14] P. Mishra, J.M. Roger, F. Marini, A. Biancolillo, D.N. Rutledge, Parallel pre-processing through orthogonalization (PORTO) and its application to near-infrared spectroscopy, *Chemometrics and Intelligent Laboratory Systems*, 212 (2020) 104190.
- [15] P. Mishra, J.M. Roger, D. Jouan-Rimbaud-Bouveresse, A. Biancolillo, F. Marini, A. Nordon, D.N. Rutledge, Recent trends in multi-block data analysis in chemometrics for multi-source data integration, *TrAC Trends in Analytical Chemistry*, 137 (2021) 116206.
- [16] P. Mishra, T. Verkleij, R. Klont, Improved prediction of minced pork meat chemical properties with near-infrared spectroscopy by a fusion of scatter-correction techniques, *Infrared Physics & Technology*, 113 (2021) 103643.
- [17] P. Mishra, S. Lohumi, Improved prediction of protein content in wheat kernels with a fusion of scatter correction methods in NIR data modelling, *Biosystems Engineering*, 203 (2021) 93-97.
- [18] P. Mishra, J.M. Roger, D.N. Rutledge, A. Biancolillo, F. Marini, A. Nordon, D. Jouan-Rimbaud-Bouveresse, MBA-GUI: A chemometric graphical user interface for multi-block data visualisation, regression, classification, variable selection and automated pre-processing, *Chemometrics and Intelligent Laboratory Systems*, 205 (2020) 104139.
- [19] P. Mishra, A. Nordon, J.-M. Roger, Improved prediction of tablet properties with near-infrared spectroscopy by a fusion of scatter correction techniques, *Journal of Pharmaceutical and Biomedical Analysis*, 192 (2020) 113684.
- [20] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: A basic tool of chemometrics, *Chemometrics and Intelligent Laboratory Systems*, 58 (2001) 109-130.
- [21] S. Wold, S. Hellberg, T. Lundstedt, M. Sjöström, PLS Modeling with Latent Variables in Two or More Dimensions, *Partial Least Squares Model Building: Theory and Application*, Frankfurt am Main, 1987.
- [22] S. Wold, C. Albano, W.J. Dunn, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg, M. Sjöström, *Multivariate Data Analysis in Chemistry*, in: B.R. Kowalski (Ed.), *Chemometrics: Mathematics and Statistics in Chemistry*, Springer Netherlands, Dordrecht, 1984, pp. 17-95.
- [23] Z. Xin, S. Jun, T. Yan, C. Quansheng, W. Xiaohong, H. Yingying, A deep learning based regression method on hyperspectral data for rapid prediction of cadmium residue in lettuce leaves, *Chemometrics and Intelligent Laboratory Systems*, 200 (2020) 103996.

- [24] X.J. Yu, H.D. Lu, D. Wu, Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflection imaging, *Postharvest Biology and Technology*, 141 (2018) 39-49.
- [25] N.T. Anderson, K.B. Walsh, J.R. Flynn, J.P. Walsh, Achieving robustness across season, location and cultivar for a NIR spectroscopy model for intact mango fruit dry matter content. II. Local PLS and nonlinear models, *Postharvest Biology and Technology*, 171 (2021) 111358.
- [26] N.T. Anderson, K.B. Walsh, P.P. Subedi, C.H. Hayes, Achieving robustness across season, location and cultivar for a NIR spectroscopy model for intact mango fruit dry matter content, *Postharvest Biology and Technology*, 168 (2020) 111202.
- [27] J.-M. Roger, F. Chauchard, V. Bellon-Maurel, EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits, *Chemometrics and Intelligent Laboratory Systems*, 66 (2003) 191-204.
- [28] X.D. Sun, P. Subedi, K.B. Walsh, Achieving robustness to temperature change of a NIRS -PLSR model for intact mango fruit dry matter content, *Postharvest Biology and Technology*, 162 (2020) 111117.
- [29] J.J. Workman, A review of calibration transfer practices and instrument differences in spectroscopy, *Applied Spectroscopy*, 72 (2017) 340-365.
- [30] B. Malli, A. Birlutiu, T. Natschläger, Standard-free calibration transfer – An evaluation of different techniques, *Chemometrics and Intelligent Laboratory Systems*, 161 (2017) 49-60.
- [31] P. Mishra, R. Nikzad-Langerodi, Partial least square regression versus domain invariant partial least square regression with application to near-infrared spectroscopy of fresh fruit, *Infrared Physics & Technology*, (2020) 103547.
- [32] W. Saeys, N.N. Do Trong, R. Van Beers, B.M. Nicolai, Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review, *Postharvest Biology and Technology*, 158 (2019).
- [33] B.M. Nicolai, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron, J. Lammertyn, Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review, *Postharvest Biology and Technology*, 46 (2007) 99-118.
- [34] P. Mishra, F. Marini, B. Brouwer, J.M. Roger, A. Biancolillo, E. Woltering, E.H.-v. Echtelt, Sequential fusion of information from two portable spectrometers for improved prediction of moisture and soluble solids content in pear fruit, *Talanta*, 223 (2021) 121733.



Visible and near-infrared spectroscopy for the measurement of fresh-produce quality parameters

More information?

Authors



Dr. Puneet Mishra
E puneet.mishra@wur.nl
T +31 317 489 118

Dr. Patrick Spoelstra*

For more information, please feel free to contact



ir. EH Westra
E eelke.westra@wur.nl
T +31 317 480 208



ir. PHI Goethals MSC
E paul.goethals@wur.nl
T +31 317 485 226

Wageningen Food & Biobased Research

P.O. Box 17
NL-6700 AA Wageningen
Tel: +31 (0)317 480 084
E-mail: info.fbr@wur.nl
Internet: www.wur.nl/foodandbiobased-research

© Wageningen Food & Biobased Research, institute within the legal entity Stichting Wageningen Research

* Dr. Patrick Spoelstra's contributions to the whitepaper are from 2018-19, as an employee of Wageningen Food & Biobased Research.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system of any nature, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher. The publisher does not accept any liability for inaccuracies in this report.