Rapid and non-destructive analysis of intact potato tubers

Hurtig og ikke-destruktiv analyse av potetmateriale

Morten Bøhler Pedersen

Norwegian University of Life Sciences
Department of Chemistry, Biotechnology and Food Science
Master Thesis 30 credits 2013

Acknowledgments

This master thesis was conducted at Nofima AS, in the period from January 2013 to May 2013, and constitutes 30 study points.

A great thanks to my supervisors, Elling-Olav Rukke and Nils Kristan Afseth. You have been immensely supportive and have guided me with steady hands. I feel proud that I had the chance to work with you on my master thesis.

The employees at Nofima are a wonderful group of people, who create a fantastic working environment. I've felt welcome every single day, and the interest many of you have shown in my work, have pushed me ever onwards. Thank you Nofima.

Maarud, Bama and Buer have made my work much easier by supplying the samples I needed for my study. Thank you for your great cooperation. I hope the results of this study will help you in future endeavors.

And of course my family and friends deserve a big thanks for helping and supporting me all these years at the University of Life Sciences. Thank you, you have made this period of my life a fantastic experience.

Ås, 14 May 2013

Morten B. Pedersen

Abstract

The purpose of this study was to study and develop NIR spectroscopy for the prediction of dry matter content in intact unpeeled potato tubers. The techniques ability to measure dry matter content was also tested under different conditions, i.e. in contact with the NIR instrument, non-contact and while the samples were moving.

Today the industry use the specific gravity to determine the dry matter content in a small sample set of potato tubers, and it is assumed that this measurement is representative for the entire potato batch. Potatoes are highly heterogeneous and can show significant variation in the constituents, not only between tubers, but also within one tuber. Creating a tool to measure dry matter content quantitative, on-line and non-destructive would be of tremendous help for the industry. This would make it possible for producers of potato products to acquire potato tubers with dry matter content suited for their purposes.

The selection of potato varieties were based on their use in the industry, and variation of skin and flesh color. They were provided by Bama, Maarud and Buer. Data from NIR measurements of the potato tubers were connected with corresponding dry matter data. This data set was combined with a data set from an earlier study (Helgerud et al. 2012), and a regression model was created using PLSR. The explained variance was $R^2 = 0,92$ with an RMSECV = 1,15 % and the model used 6 PLS factors. Using the PLSR model, prediction of dry matter was done on 1194 potato tubers of the variety Sava and Rafaela was done successfully.

Comparing the PLSR models for measurements done with contact, non-contact and in movement, there was relatively little difference when looking at the R^2, RMSECV and the amount of factors used. This confirms that it is possible to use NIR spectroscopy as a fast and reliable way to predict the dry matter content in potato tubers.

Sammendrag (Norwegian Abstract)

Formålet med oppgaven var å utvikle en modell, ved bruk av NIR spektroskopi, som kunne predikere tørrstoffinnholdet i poteter. Dette ble også testet under forskjellige forhold, dvs. i kontakt med NIR instrumentet, ikke-kontakt og måling ved bevegelig potet.

I dag benytter potet industrien seg av egenvekten til poteter for å bestemme mengden tørrstoff i et lite uttak av poteter. Dette antar de at gjelder for resten av partiet. Poteter er svært heterogene, og kan variere mye, ikke bare i mellom poteter, men også innad en potet. Å utvikle et verktøy for å måle tørrstoff raskt, ikke destruktivt og i store kvanta ville være til stor hjelp for industrien. Dette ville muliggjøre for produsenter av potetprodukter å kunne kjøpe poteter som har tørrstoff som er egnet til deres produkter.

Valget av potetsortene var basert på deres alminnelige bruk i industrien, og pga. variasjonen av skall og kjøtt. Potetene ble levert av Bama, Maarud og Buer. Data fra NIR målingene av potene, ble koblet med respektive tørrstoffmålinger. Data settet ble kombinert med et data sett fra et tidligere forsøk (Helgerud et al. 2012), og en regresjons modell ble lagd ved hjelp av PLSR. Variansen for modellen var R^2 = 0,92 med en RMSECV = 1,15 % og modellen brukte 6 PLS faktorer. Modellen ble brukt, vellykket, for å predikere mengden tørrstoff i 1194 poter, av sorten Sava og Rafaela.

Sammenligning av PLSR modellen av målingen gjort i kontakt, ikke-kontakt og ved bevegelse, viste liten forskjell når det ble sett på R^2, RMSEVC og antall faktorer brukt. Dette bekrefter at det er mulig å bruke NIR spektroskopi som en rask og pålitelig metode for å måle tørrstoff i poteter.

Table of content

Abbreviations

DM – Dry Matter

NIR – Near Infrared

PCA – Principal Component Analysis

PLSR – Partial Least Square Regression

RC – Regression coefficient

RMSECV – Root Mean Squared Error of Cross Validation

RMSEP – Root Mean Squared Error of Prediction

SG – Specific gravity

SNV – Standard Normal Variate

1. Aim of paper

Potatoes are an important part of the basic food supply in the world. In 2011 it was produced 374,4 million metric tons in the world (Food and Agriculture Organization of the United Nations 2013). For consumption, the potato is often processed in some manner, e.g. boiled, fried or baked. It is also a very popular snack, and potatoes are often deep fried (chips, french fries etc.). In the industry it is very important that the products they sell are the same each and every time. If producing chips, all the chips need to be the same color, and have the same crispness, just to mention a couple of quality parameters. Basically all the products need to be the same. Considering potatoes this can pose a challenge. Potatoes are very heterogeneous when it comes to dry matter (DM) content. This applies to within a single tuber, between tubers from the same variety, from year to year, and even potatoes from the same batch.

Today the established way of measuring DM in potatoes are by measuring the specific gravity (SG) of 2 x 5 kg of potatoes for every 10 metric tons (Lunden 1956). The results obtained here are assumed to apply for the entire batch. Considering how much two tubers can differ from each other, it can be assumed that this is not an effective and representative method. Furthermore since this method is not suitable for measuring DM content in single tubers, there are no good industrial data on the variation of DM within a batch of potatoes.

NIR (near infrared) spectroscopy is used to quickly acquire data and predict the content of different constituents in products and raw materials in a non-destructive way (Isaksson et al. 1996; Sirisomboon et al. 2009). If this method can quickly, continuously and non-destructively predict the DM in potatoes, it would increase the quality of products using potatoes. The producers of potato products can then buy potatoes in a specific range of DM content, suited for their products. The main goal of the present study is to see if it’s possible to develop a model to predict the DM in potato tubers under different conditions.

The first aim of this study was to calibrate a NIR instrument as a tool for rapid, non-destructive on-line measurement of potato tubers, and to test the robustness of the model. Secondly, the model developed was used to predict the DM content in a large amount of potato tubers. The last aim was to investigate how robust the NIR instrument was under different variations of measuring, i.e. sample in contact with the NIR instrument, with space between the instrument and sample, and measuring while the sample was moving.

The study was accomplished in three parts with four goals:

Part 1. Establishing a calibration for dry matter prediction in intact potato tubers and testing the strength of the NIR instrument.

Goal 1. To expand on an already existing calibration for measuring stationary potato tubers.

Goal 2. Test the strength of the obtained calibration model.

Part 2.

Goal 3. Predict the variation of DM in unsorted potatoes that the potato industry (Bama) receives from the farmers, using the calibration developed in part 1.

Part 3.

Goal 4. Investigate if it is possible to measure the DM in tubers not in contact with the machine, and also while moving

2. Literature

2.1 The Potato

The potato is part of a genus consisting of over 2000 species. Of these there are about only 10 % which produce tubers. Moreover, there is only one species that is more or less exclusively grown throughout the world: *Solanum tuberosum L.* (Burton 1989).

2.1.1 Morphology

The potato plant can be divided into two parts: the below- and above-ground parts. Above the ground the stem, leaves and flowers grow, whereas we find below the ground, the roots, stolon and tubers (Burton 1989).

Wild growing specimens will develop stolons from the roots of the plant, which in turn develop tubers where sprouts and eventually stems grow. Planted seed tubers will also develop sprouts and stems in the same manner. Each potato can develop one sprout for each of its eyes. These sprouts are in the beginning of the growth phase sharing resources from the same tuber, but eventually each stem will develop roots and leafs and will compete against each other for light and other resources (Struik 2007). The stolons will start to grow out of the stem closest to the mother tuber, and the first stolons will grow longer and faster than stolons developed at a later stage (P. C. Struik & van Voorst 1986).

The tubers are developed from swollen parts of the stolons, and can be described as being globular to elliptic in shape, with variation in size, shape, skin color and texture. They contain a high amount of starch and proteins (Struik 2007). When growth of tubers are initiated an increase of starch content can be observed in the stolons, and a decrease of glucose and fructose (Helder 1994).

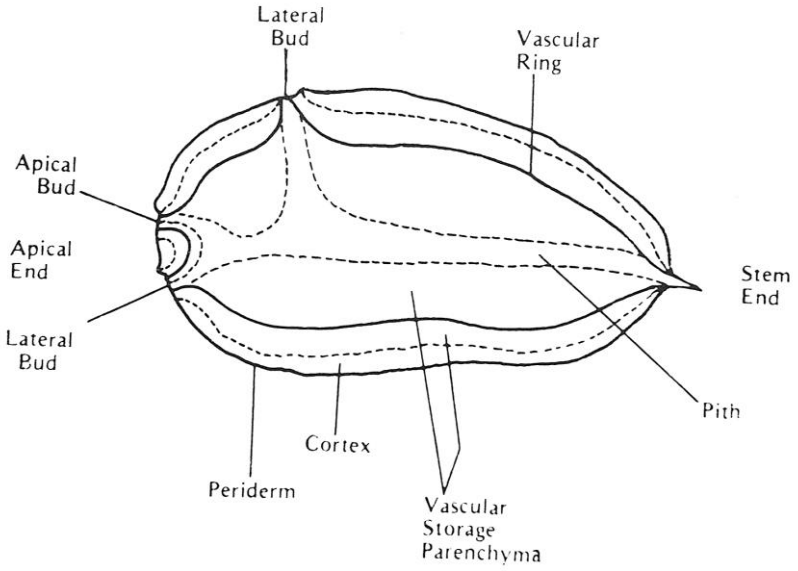


Figure 1 A representation of the anatomy in a potatot tuber (Salunkhe et al. 1991).

From the outside to the inside of the potato tuber the different layers (figure 1) are the skin, the cortex, the vascular system, the storage parenchyma and the pith (Struik 2007). Each of these layers has different DM content and composition (Anzaldúa-Morales et al. 1992), which makes NIR measurements a challenge.

2.1.2 Storage

During storage of plant materials, it is important to consider the respiration of the plant, which is directly related to the plant shelf life. Compared to many other plant commodities, potatoes are considered as a low-respiration plant with long shelf life (Brecht et al. 2008). Even so, it is important to store potato tubers correctly, regarding temperature and humidity, to increase the maximum storage time. Temperature is what influences this the most (Brecht et al. 2008).

Potatoes stored at a low temperature (2,5 °C) will have a higher respiration rate, compared to storage at higher temperatures (Schippers 1977; Dwelle & Stallknecht 1978). The respiration rate can be increased if the potatoes are damaged during field harvesting (Pisarczyk 1982). It is also well known that longer storage at a low temperature, will produce potato tubers with a much higher sugar content (Dwelle & Stallknecht 1978). This is caused by a degradation of starch to sucrose and other reducing sugars (Brecht et al. 2008). Ewing et al. (1981) showed that storing tubers at a cold temperature, (1 °C) even for a short duration (four days or more), gives an increase of sugar content. This can be somewhat counteracted by storing the tubers at 19 °C afterwards.

2.1.3 Dry matter

An important part of the potato, both considering nourishment and processing of potatoes, is the DM content. DM is affected by many factors, among those soil quality (e.g. temperature, moisture, nutrients), weather variations (e.g. sun and rain) and time of harvest. In table 1 the average DM content in potatoes are shown.

Table 1 Average content (%) in potato tubers (Matvaretabellen).

	Water content	Fat	Carbohydrates	Starch	Fiber	Protein	Energy (kJ)
Raw Potato, Autumn	82	0,1	15,3	12,4	1,6	1,7	306

As table 1 show, water is the main content of the potato. The DM consists mostly of starch and other carbohydrates (sugars). The content of the DM can vary greatly within each tuber (Sharma et al. 1958), between tubers from the same farm and within the same variety from year to year (Cole 1975).

2.1.3.1 Starch

Starch is by far the most used energy storing substance by plants. Starch consists of two carbohydrates, amylose and amylopectin (figure 2b and 2c respectively), which are packed together in granules (Bemiller & Huber 2008). The ratio of the two carbohydrates in the starch varies from plant to plant, but amylopectin is most often the dominant one. In potato, starch makes up about 15 - 20 % of the weight (Bertoft & Blennow 2009), whereas amylopectin constitutes about 70 -80 % of the starch weight (Hoover 2001).

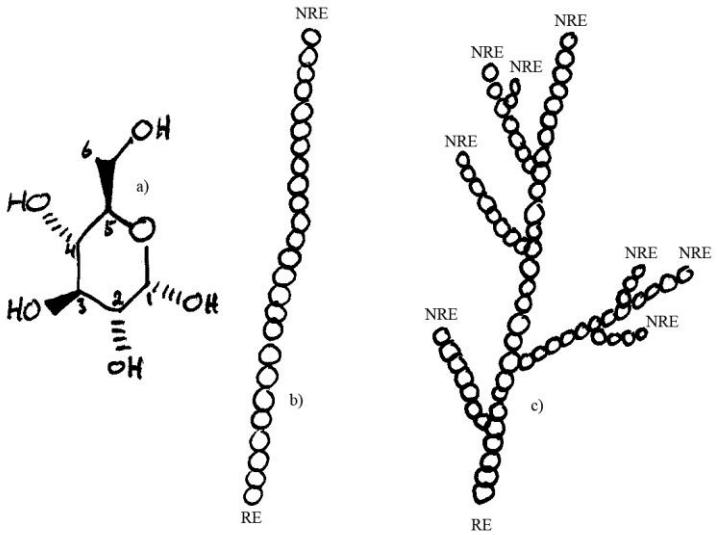


Figure 2 Simple representation of the two carbohydrates in starch. a) α-D-glucose b)Amylose c) Amylopection. NRE are Non reducing ends and RE are reducing ends of glucose.

Amylose is the smaller of the two carbohydrates, and consists of long linear chains, linked by (1→4) α-D-glucose units. There are a few branches, but they are so few and far apart, and either very short or very long in length, that amylose can be considered a linear molecule (Bemiller & Huber 2008). Amylopectin on the other hand, is one of the largest natural occurring molecules known. They are chained in the same way as amylose, but consist of several more branches, linked by α-D (1→6) linkages. The chains are located at random points, and constitutes about 4 - 5 % of the total linkages in amylopectin. In comparison, these linkages only make up 0,3 - 0,5 % of the total linkages in amylose (Bemiller & Huber 2008).

2.1.3.2 Proteins

Proteins are made up of 21 different amino acids (AA), but not all proteins are necessarily made up of all 21 (Mathews et al. 2000). It is the combination and sequence of AA, which gives proteins their properties. This could be size, bond strength and structure. It will also effect if a protein is lipo- or hydrophilic, or even both. Of the 21 AA, 10 of these are considered essential, i.e. humans cannot synthesize these and must therefore get them through the diet (Mathews et al. 2000).

Potatoes are considered a good source for AA (Kapoor et al. 1975), which can easily be combined with other basic foods (e.g. pasta and rice), to cover the entire range of essential AA. Gelder and Vonk (1980) quantized the amount of AA in proteins in 34 potato varieties. An excerpt of the results is shown in table 2, which is a general indication of the expected AA content in a potato tuber.

Table 2 The average AA content measured from 34 potato varities (Gelder & Vonk 1980).

Amino Acid	Average Content (g/100 g)	Amino Acid	Average Content (g/100 g)
Threonine[1]	5,42	**Aspartic Acid**	12,64
Valine[1]	6,42	**Serine**	5,40
Methionine[1]	2,15	**Glutamic acid**	10,23
Isoleucine[1]	5,29	**Proline**	4,83
Leucine[1]	10,28	**Glycine**	5,03
Phenylalanine[1]	6,53	**Alanine**	4,73
Lysine[1]	7,64	**½ Cystine**	0,77
Histidine	2,06	**Tyrosine**	5,62
Arginine	4,95		

[1]Essential AA

These AA is the basis of the proteins in a potato tuber. There are three main soluble protein groups in potatoes: patatins, protease inhibitors and other proteins, which constitutes about 40-60 %, 20-30 % and 20-30 % of the total protein content respectively (Pots et al. 1999).

2.1.4 Processing of Potato Tubers

Potato tubers are rarely consumed in a raw a state. Boiled, fried, deep-fried, baked or grilled are common ways to serve potatoes. Taste and feel (texture, mealiness, consistency etc.) is very important to the experience of the meal. The content of DM has a great influence on the textural and rheological feel of the tubers (J. Singh et al. 2008; Kirkpatrick et al. 1951). Producers of "ready-to-eat" meals have a goal to produce the same product each and every time. With the standardized preparation methods, each potato is treated in the same manner without regard to the DM content.

2.1.4.1. Maillard Reaction and Acrylamide

A maillard reaction is a non-enzymatic browning between reducing sugars and a primary amino group, which is catalyzed when heating (Bemiller & Huber 2008). This will create a brown, dark color which is desirable in some foods, but not in all. Chips and fries are examples of commodities where the consumers expect a yellow-white color. If potatoes, planned to be fried or deep fried, are stored in the wrong way, resulting in an increase of sugars, they might go through the maillard reaction and get a dark color (Ewing et al. 1981; Khanbari & Thompson 1993).

Another downside with the maillard reaction is the formation of acrylamides. Acrylamide has been found to be a human neurotoxin, and should be avoided (Calleman et al. 1993). Heat treatment above 120 °C (Bemiller & Huber 2008) may start a reaction between reducing sugars and AA (particularly methionine and asparagine), producing acrylamide (Stadler et al. 2002). For further reading about the topic of acrylamide and health safety, an extensive review by Friedman (2003) is recommended.

The challenge with maillard reaction and acrylamide development in potato production is a clear example as to why the industry needs a method to measure the DM content fast, quantitative and on-line.

2.1.5 Specific Gravity

Specific gravity (SG) is the established method of measuring DM content in potato tubers today. SG is usually calculated from the weight of the potato tuber in air, and in water, using equation (1):

(1) $$Specific\ gravity = \frac{weight\ in\ air}{weight\ in\ air - weight\ in\ water}$$

This is the official equation, used in the EU, to calculate SG and predict the amount of starch and DM content in a batch of potato tubers (Anonymous 1999). After measuring the SG, a producer either uses a predefined table or an equation to predict the DM content.

There have been disagreements on how to define the relationship between SG, underwater weight, DM content and starch content. This has caused several different relationships to be developed (Von Scheele et al. 1937; Nissen 1955; Porter et al. 1964; Houghland 1966). Simmonds (1977) looked at eight different relationships, and found that the difference between the methods was apparent, but negligible. However, he pointed out that using a relationship based on empirical linear regression is the most realistic predictor. Earlier works by Nissen (1955, 1967) showed that each measurement will be affected by the air content in each tuber. Nissen recommended using vacuum to empty the intercellular spaces in the tubers, and fill them with water, as this would increase the accuracy of the prediction. This was supported by later works by Wilson & Lindsay (1969), which concluded that the accepted regression line can't be used for every variety of potatoes, as the amount of air in each tuber will vary greatly.

2.2 Vibrational Spectroscopy

2.2.1 Electromagnetic energy

Electromagnetic energy is sorted based on the wavelength. Radio waves have the longest wavelengths, from 1,0 cm up to over 100 000 km, while gamma rays are the shortest, starting at 10^{-11} m at its longest (Elert n.d.) .The visible light range is somewhere between these two, ranging from about 400 – 800 nm. The whole spectrum can be seen in figure 3.

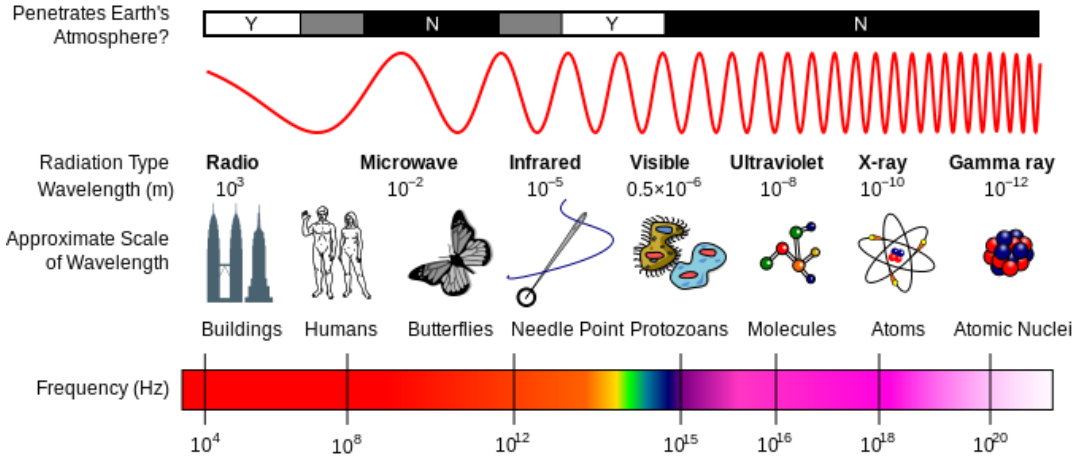


Figure 3 A figurative representiation of the electromagnetic spectrum. It shows the wavelength (nm) and frequency (Hz) for some of the different energies (Wikimedia Commons 2013).

When molecules are exposed to electromagnetic energy they may absorb some of it. The effect of the energy absorption differs, based on what range of electromagnetic energy is used. Radio waves will affect the spin orientation in the magnetic field around the atom or molecule, while more energy rich gamma rays, from the other side of the spectrum, will affect the nucleus of each atom (Merritt & Settle 1981). This absorption effect can be used to quantify the content of different substances in matters, by measuring the non-absorbed energy. One of the great advantages with this method, is that it can be done on unprepared products, in gas, liquid or solid form (Li-Chan 2010).

2.2.2 Molecular vibrations

The infrared spectrum ranges from 0,7 to 500 µm, which includes the NIR region at 0,8 to 2,5 µm, and mid-infrared region at 2,5 – 15,4 µm (Merritt & Settle 1981). When this energy is absorbed by atoms, there is a transition of vibrational energy (D. Williams & Fleming 2008). Figure 4 shows some of the vibrational movements the molecules can have.

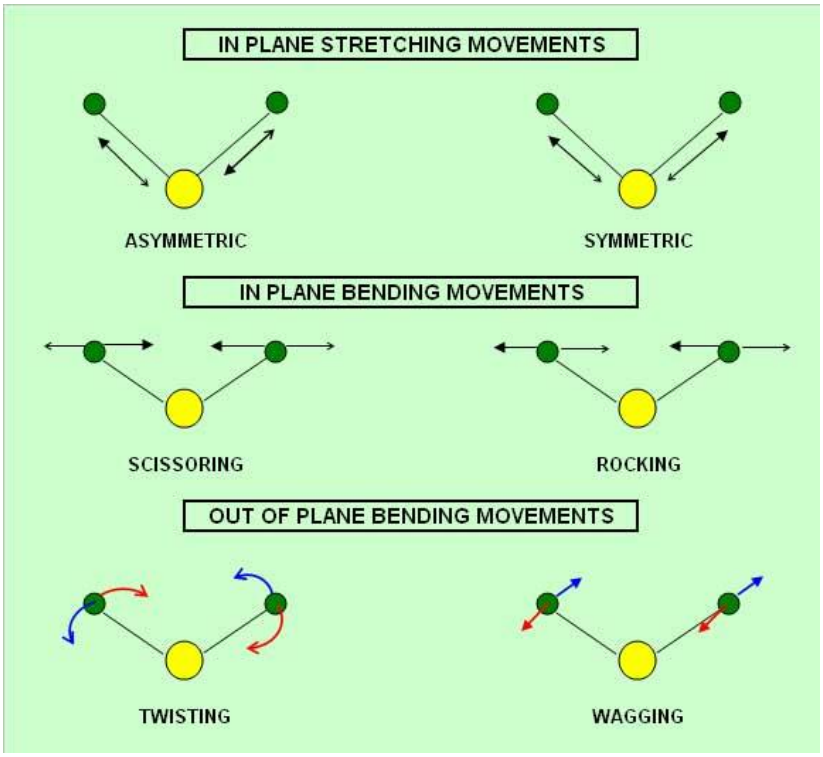


Figure 4 The diagram shows some of the movements a molecule can go through when exposed to electromagnetic energy. The movements are defined as either bending or stretching **(Intertek Plastics Technology Laboratories n.d.).**

The kind of transition depends on what kind of atoms are in a molecule, and the bindings in between them. This means that it can be predicted what kind of movement happened, based on the energy which was not absorbed. Which in turn, can give information on what kind of molecules and functional groups are or aren't present in the sample (Merritt & Settle 1981).

2.2.3 Vibrational Transitions

Assuming that the atoms obey Hooke's law, vibrational transitions can be explained by the following equation:

(2) $$V_{iv} = h\upsilon_i \left(v_i + \frac{1}{2}\right)$$

Where V is the vibrational states, h is Planck's constant, υ_i is the fundamental frequency of the particular mode, and v_i is the vibrational quantum number of the *i*th mode (Griffiths 2010). The energy difference between vibrational transitions (i.e. the difference between the ground state $v_i = 0$ and the number of excited states ($v_i = 1, 2, 3 \ldots i$)), is an important part of the information that can be gathered from spectroscopy. The first transition, between $v_i = 0$ and $v_i = 1$, is called the first excited state, and is associated with mid-IR. Transitions from $v_i = 0$ and $v_i \geq 2$ is called overtones and is associated with NIR (Griffiths 2010).

2.2.4 NIR Spectroscopy

NIR spectroscopy is mostly associated with measurements of the overtones of C-H, O-H and N-H stretching vibrations (Griffiths 2010). A great advantage with using NIR, compared to mid-IR, is that it's easier to acquire data from raw or processed material, and it can study samples in all phases (gases, liquids and solids). The instruments for NIR are also more sensitive, which means that spectra with high signal-to-noise ratio can be measured in less than one second (Griffiths 2010).

NIR spectroscopy consists of several techniques of sampling. Transmission, transflection and interactance method are a few of the available techniques. In this study interactance has been used, because of its superior ability to penetrate deeper into tissues. NIR interactance spectroscopy is a method where illumination and detection are placed side by side (figure 5).

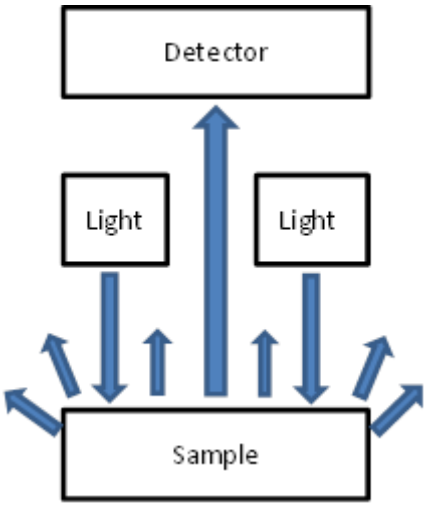


Figure 5 The light enters the sample, and some of it is absorbed while some leave the sample. The light that leaves is measured in a spectroscopic reader that detects which wavelengths are absorbed by the sample.

One set of fiber-optics illuminates the sample, which interacts with the sample. When emerging from the sample, another set of fiber-optics detects light coming out, transferring the data to software that can interpret the measurements (Chalmers & Griffiths 2010). An absorbance plot in the NIR spectrum is shown in figure 6.

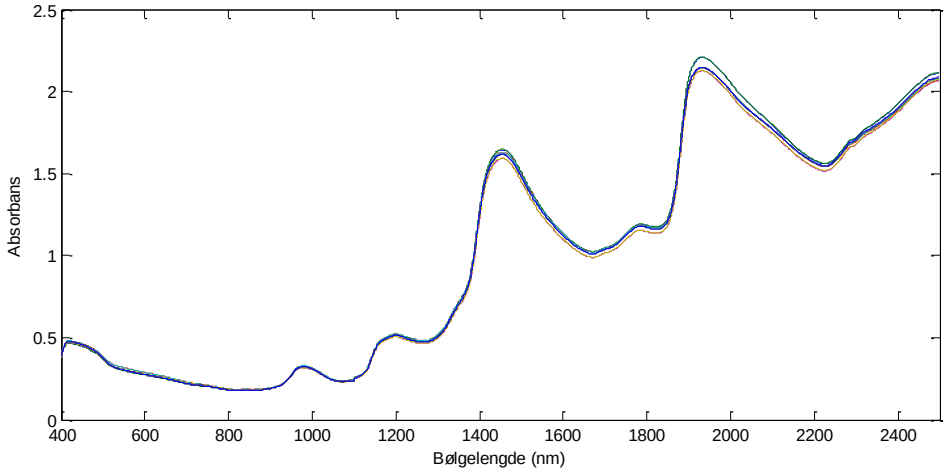


Figure 6 Shows a absorbance line plot in the NIR spectrum (850 – 2500 nm) (Bølgelengde = Wavelength).

NIR spectroscopy is a versatile tool, used for many purposes, especially within food analysis. As early as in 1996 Isaksson et al. successfully used an online NIR reflectance analysis system to predict the fat, water and protein content in grinded meat (RMSECV 0,73 – 1,50 %, 0,75 – 1,33 % and 0,23 – 0,32 % respectively). It has also been found that it's possible to use NIR spectroscopy to predict the fat content in live fish (Folkestad et al. 2008). This method is not limited to measuring animal tissue, but it can also be used to determine different qualities in plant material. For instance non-destructive predictions of the DM content in greenhouse grown tomatoes (Khuriyati et al. 2004), or sorting out soya beans with defects (Sirisomboon et al. 2009). NIR spectroscopy is an excellent tool to determine different qualities in many different materials, without the need to destroy the samples.

2.2.5 NIR Spectroscopy of Potato Tubers

One of the earliest studies on the effectiveness of NIR spectroscopy measurements of DM content in potato tubers, was done by Dull et al.(1989). They measured whole potato tubers, thin and thick sliced tubers, using two NIR spectrophometers, in the range of 800 - 1000 nm. Dull et al. compared the predicted DM content with the actual DM content, and found a correlation (R) of 0,9749, 0,9520 and 0,9178 for thin slices, thick slices and intact tubers respectively. Since then several other studies have been performed on the feasibility of measuring DM with NIR spectroscopy, but most researchers have used mashed or sliced potatoes in their research (Scanlon et al. 1999; Haase 2011). As well as estimating DM content, research have shown that it is possible to use NIR to estimate the starch content in potato tubers as well (Hartmann & Büning-Pfaue 1998; Haase 2003, 2006, 2011). In 2010 Brunt et al. designed an in-line NIR system for measuring DM content in potato tubers.

This system included systems for pulping the tubers, a sulfite dosage system to prevent enzymatic browning, and measuring the SG. The downside with this method is that its capacity is about 12 potato samples per hour and it's a destructive method. Subedi & Walsh published in 2009 the first NIR interactance investigation on intact potato tubers. Using a NIR-enhanced spectrometer (310 – 1133 nm), they achieved an explained variance (R^2) of 0,87, for intact tubers. For peeled potatoes the R^2 was 0,92. In the same investigation they measured sliced tubers, both motionless and in motion. The R^2 was 0,92 and 0,85 respectively.

Helgerud et al. (2012) have like Subedi & Walsh measured the DM content of moving potato tubers, but unlike Subedi & Walsh, they used a commercially available 2D NIR interactance system and whole, unpeeled potato tubers. Helgerud et al. also used a prototype 1D visible/NIR interactance instrument to measure stationary intact potato tubers, which is the same system that is used in the current study. An excerpt of the results from the research is shown in table 3.

Table 3 The explained variance (R^2) and RMSECV (Root Mean Squared Error of Cross Validation) of Helgeland et als. (2012) study.

	R^2	RMSECV (%)
1D Interactance Measurements	0,95	0,91
2D Interactance Measurements	0,83	1,68

Comparing the results from the 1D and 2D measurements, it is clear that 1D interactance has a higher explained variance. Helgerud et al. explained the poorer prediction performance with lower penetration depth and shorter measurement time with the 2D NIR interactance system.

There is no published literature where an on-line system for measuring DM content in potato tubers, has gotten satisfactory low RMSEP. Also, there is no commercially available system for an on-line system.

2.3 Chemometrics

Spectroscopic analysis provides large amounts of data in the form of spectra. These spectra can consist of tens, hundreds, or even thousands of wavelengths. Several of these wavelengths may often be linear dependent on each other, a feature often denoted as colinearity (Næs 2002). To be able to obtain reliable data from the spectra, there is a need for a mathematical toolbox for e.g. utilizing information from collinear variables, compressing huge amounts of data, pretreatment, calibration against reference measurements, and interpretation of results. This toolbox is often denoted chemometrics.

Chemometrics is "the use of mathematics and statistics on chemical data" (Martens & Næs 1989). In spectroscopy, chemometrics is used both for qualitative analysis (e.g. classification and clustering) and for quantitative analysis (e.g. calibration and prediction). When using NIR spectroscopy a calibration is needed for the data to be reliable. Measuring and interpreting all the wavelengths without calibration, is an undertaking which is overwhelming and will probably yield confusing results. Calibration turns the wavelengths into precise and relevant information which can be used (Martens & Næs 1989).

2.3.1 Preprocessing

According to Barnes et al. (1989) the variation of information acquired from NIR readings originates from three sources; "nonspecific scatter of radiation, variable spectral path length through the sample, and chemical composition of the sample". Thus, there are multiple sources of noise in the spectra apart from the "pure" chemical information, and it is necessary to separate physical noise, leaving only information that affects the chemical composition. In other words, the purpose of preprocessing is to transform the data so that it is suitable for analysis (Esbensen 2009). There are several methods available to achieve this, but in this study Standard Normal Variate (SNV) transformation have been used. The physical noise created by variation in particle sizes and light scattering is removed by using equation (3) (Barnes et al. 1989):

(3) $$SNV_{(1-700)} = (y_{1-700} - \bar{y}) / \sqrt{\frac{\sum(y_{1-700} - \bar{y})^2}{n-1}}$$

Here, SNV denotes the corrected spectrum, y is the individual wavelengths in a spectrum, $\bar{y}$ is the mean of all the wavelengths in a spectrum, and n is the number of wavelengths in a spectrum. In effect, SNV correction reduces some of the slope effects seen in the spectra and center them closer to the mean linear slope (Barnes et al. 1989).

2.3.2. Calibration

In quantitative spectroscopic calibration the main goal is often to find the relationship between spectroscopic measurements (x) and a given reference value (y). There are generally two main approaches of calibration, namely, univariate calibration and multivariate calibration. In univariate calibration, only one measurement variable is used, and the calibration finds the linear relationship between this variable and a given reference value. In univariate calibration there are two equations (equation 1 and 2) that are used to estimate the value of *y*, from the measured value of *x* (Næs 2002).

(4) $$y = b_0 + b_1 x + error \;\rightarrow\; \hat{y} = \hat{b}_0 - \hat{b}_1 x$$

(5) $$x = a_0 + a_1 x + error \rightarrow\; \hat{y} = -\left(\frac{\hat{a}_0}{\hat{a}_1}\right) + \left(\frac{1}{\hat{a}_1}\right) x$$

where y is the reference measurement, x is the rapid measurement while a and b are fixed values. In spectroscopy, as in many other types of measurement data, it is often more appropriate to include several measurement variables in the calibration, i.e. perform a multivariate calibration. (Næs 2002). This type of calibration can take all the wavelengths used into account, and create a model of prediction which often has a much higher reliability than a model based on a single wavelength.

2.3.2.1 Principal Component Analysis (PCA)

The data collected from spectroscopic experiments are usually sorted in an **X**-matrix with n objects and p variables as shown in equation 6:

(6) $$\boldsymbol{X} = \{X[p, n]\}$$

In NIR spectroscopy *n* are the number of samples, and *p* is the number of wavelengths used for each sample (Esbensen 2009). Since spectroscopy often has tens or hundreds of wavelengths it is near impossible to represent the data in a graph that is limited to three dimensions (Næs 2002). PCA compresses this data into principal components (Esbensen 2009) which is the components that explains the data the most. The form of a PCA model can be expressed as:

(7) $$X = TP' + E$$

where T is the score-plot, P is the loading matrix representing the regression coefficients of X, and E is the residual matrix representing any noise data (measurement noise, operator mistakes etc.) (Martens & Næs 1989). The first component explains most of the variation, while the second explains the variation second most etc. It removes the effect of colinearity,

and shows more clearly any outliers that might be in the data. Outliers are measurements that lies so far out from the rest of the data, that it isn't considered normal variation, but instead noise. Removing an outlier from further data treatment is a difficult decision, because the outlier could be a rare reading and therefore a part of the variance. This means that removing outliers might remove important data.

2.3.2.2 Partial Least Squares Regression (PLSR)

PLSR is considered, like PCA, a bilinear modeling technique (Martens & Næs 1989). Unlike PCA, PLSR takes into account the responses from the Y-matrix. By looking at the X-matrix and finding what information here is relevant for the Y-matrix, the relationship between these two values are found. This is called covariance (Esbensen 2009). This results in a model with fewer components, where the first X-matrix component explains the most of the variation in the Y-matrix. The second component explains the second most, etc.

2.3.5 Validation

Validation is a technique that tests the strength of the model created with PLSR. This is done using data from other samples, acquired in a similar manner as the data in the model. The simplest method would be to repeat the data collection with new samples, i.e. so called test-set validation, but this may be costly and more samples may not be available (Esbensen 2009). A solution for this is to use full cross validation. One sample from the model is removed and estimated with the remainder of the model, and the result is compared with the reference value. In the next step, the data is replaced and a different sample is removed and used to test the model in the same manner. This continuous until all samples have been kept out once (Stone 1974). This is not as good as creating another set of data, but is a necessary method when there is not enough resources for creating an entire new data set for the validation (Esbensen 2009).

Two common methods of expressing the strength of the model are R^2 and Root Mean Square Error of Prediction (RMSEP). R^2 is a dimensionless number between 0 and 1,00 that says something about the correlation between the x- and y-values. The closer to 1,00 R^2 is, the more of the variation in the y-values can be explained by the measured x-values (Esbensen 2009). RMSEP on the other hand, says something about the error expected from any predictions made by using the model (Esbensen 2009). RMSEP is expressed in equation (6):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad (8)$$

where n are numbers of samples, y_i is a given sample, and $\hat{y}_i$ is the prediction of the given sample.

2.3.2.3 Regression Coefficients

A PLSR model is a complex model explaining the relationship between the measured x values and the responses y. Linear regression model is a simple model where single spectral peaks are correlated to single responses y. The model can be expressed as:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (9)$$

where the constants β_0 and β_1 are an unknown intercept and slope (regression coefficients) respectively, and ε is an random error component (Montgomery et al. 2001). Regression coefficients (RC) denote which wavelengths are important in a given PLSR model.

3. Materials and Methods

3.1 Materials

3.1.1 NIR Instrument

The machine used for this study was a prototype VIS/NIR (visible/near infrared) interactance instrument, which could register 30 different, equally spaced, wavelengths of light in the region of 460-1040 nm. The potatoes were lit by two 50 watt halogen lamps (OSRAM, Augsburg, Germany), that illuminated two equally large squares on the potatoes. The distance between the two light sources was 14 mm and the backscattered light was collected through a collection tube. The instrument was able to collect $^{80\ spectra}/_{1sec}$. From an earlier experiment it was found that the light could penetrate up to 20 mm into unpeeled potatoes (Helgerud et al. 2012).

3.1.2 Potato Tubers

Eight different varieties of potato tubers were chosen based on their common use in industrial processing, and variation of skin and flesh color. All of the potatoes were stored at 4 °C before preparation. Which company provided them is shown in table 4, along with their use in the industry and description. When possible, tubers with a diameter of 45 – 55 mm were chosen.

Table 4 An overview of the varieties used in the study, and which companies that provided them, along with the color of flesh and skin for each variety (Fagforum Potet n.d.; Bundesssortenamt 2012).

Potato	Origin	Use in industry	Skin Color	Flesh Color
Saturna	Maarud	Chips	White	Light Yellow
Bruse	Maarud	Chips	Red	Light Yellow
Sava	Bama Moss	Sous vide (ready to eat meals)	White	Yellow
Fakse	Bama Moss	Sous vide (ready to eat meals)	White	Light Yellow
Folva	Bama Moss	Sous vide (ready to eat meals)	White	Light Yellow
Mandel	Buer	Potato cakes	White	Yellow
Asterix	Bama Moss	French Fries	Red	Light Yellow
Rafaela	Bama Moss	Sous vide (ready to eat meals)	Yellow	Yellow

3.1.3 Software

The data acquired from the NIR measurements were collected within the software MATLAB, V 7.12 data analyzer software (The MathWorks, Inc. Natick, MA, USA), and imported into Unscrambler X, V 10.1 statistical analysis software (CAMO PROCESS AS, Oslo, Norway), where the data analyses were performed. The reference values (actual DM content) was treated in Excel (2010, Microsoft).

3.2 Methods

3.2.1 Reference Method

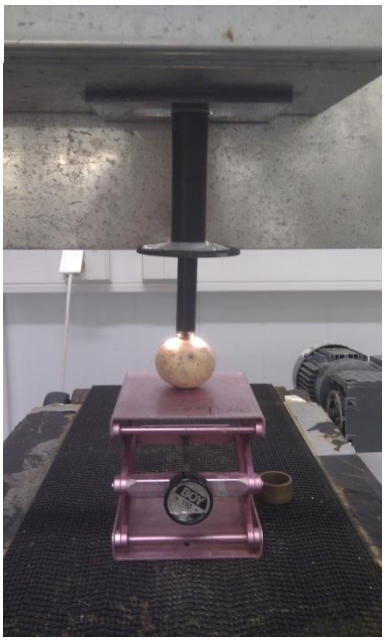
The method to acquire dry matter data from the potatoes was tested to see if it gave stable enough results. Five potatoes was chosen randomly and scrubbed gently for dirt and excess water was dried of. Afterwards they were cut into 2,0 x 2,0 cm strips with a Hallde RG-100 table cutter (AB Hällde Maskiner, Sweden). The length of the strips, varied with the size of the potatoes. The strips were kept in small plastic buckets with lids, to decrease water loss to the air. Strips from each of the potatoes were weighed and put in a heating cabinet to dry for at least 48 hours at 105 °C. Four replicates from each potato were taken, each sample weighing 20 ± 1 g. After drying the samples were put in an exikator, for about 5 minutes, for cooling. The samples were weighed separately to keep them from absorbing moisture from the air. The DM content and the standard deviation (SD) were calculated in Excel (Microsoft, 2010). This method was used for all reference testing in this study.

3.2.2 NIR Measurements

The NIR measurements were done in three different ways:

1. Contact: This was done with the potato in contact with the collection tube, which meant that only the light from within the potato would be registered (figure 7a). The acquisition time was 2 sec, with three replicates per potato. Between each replicate the potato was turned about 30 °.
2. Non-contact: The second test was done in the same manner, but with a distance at about 10 mm from the device (figure 7b). This was measured from the thickest potato, which meant that most potatoes would have a longer measurement distance.
3. Movement: The last test was done with the same distance, with the potatoes moving on a conveyor belt (figure 7b). The engine was of custom build and was fixed at a speed of 250 RPM (rounds per minute). In this part the potatoes were not turned between each replicate, and the acquisition time was 5 sec.

All of the measurements were done in the center of the longest axis of each tuber. Pritchard and Scanlon (1997) found that this position had a dry matter content which represented the average dry matter content closest. This has been confirmed later (Helgerud et al. 2012). A computer using the software Matlab registered the data from the measurements.

a) 

b) 

Figure 7 The setup of the measurements. a) Contact measurements. b) Non-contact measurements and movement measurements (passing potatoes).

3.3 Description of the experiments

The work of this study can be diveded into two main parts. The first part mainly consisted of building a model with 470 different potato tubers (data set C) and to define its strength. 356 tubers were measured in this study (data set A), while 114 tubers were used from an earlier study (Helgerud et al. 2012) (data set B). This model was used to predict the variation in DM of 1194 potato tubers from the industry, and the temperature data from ten potatoes.

Another part of the study used a subset of 240 potato tubers from the data set A as described above. In addition to contact NIR measurements, the tubers were also measured at a distance from the collection tube, and while in movement.

An overview of the data sets (with denotation and source) and the main workflow is illustrated in the flowchart in figure 8.

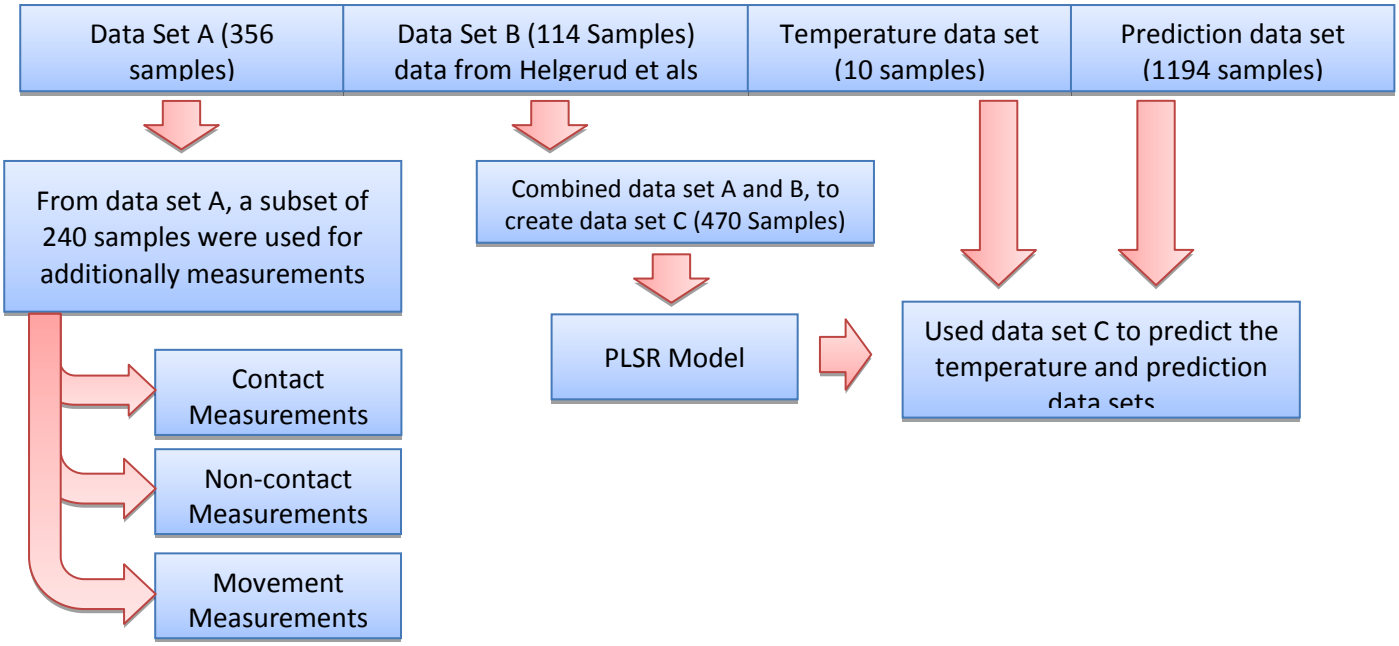


Figure 8 The flowchart shows the data sets created in the study, and what they were used for.

3.3.1 Part 1: Establishing the Calibration

Eight varieties of potatoes were used for data set A. Table 5 shows an overview of the different potato varieties, how many of each variety was used and how long they were in storage (4 °C) before preparation.

Table 5 An overview of the potato varieties used in the model (n = number of samples used), and how long they were stored at 4 °C before preparation.

	Saturna	Bruse	Fakse	Sava	Folva	Mandel	Asterix	Rafaela
n	40	40	40	90	30	30	30	60
Storage time (Days)	5 (12)[1]	6 (12)[2]	5	7 (0)[3]	6	5	5	0

[1] 30 Saturna potatoes were stored for five days, ten were stored for 12 days.
[2] 30 Bruse potatoes were stored for six days, ten were stored for 12 days.
[3] 30 Sava potatoes were stored for seven days, 60 were stored for zero days.

3.3.1.1 NIR Measurements

30 potatoes between 45 – 55 mm were selected, the day before experimentation. Two extra tubers were selected for measuring core temperature before NIR measurements. The potatoes were cleaned and gently scrubbed; making sure the skin was as little damaged as possible. Afterwards they were stored at room temperature (18 – 19 °C) overnight, to dry off, and increase and stabilize the core temperature. The next day the potatoes were measured with NIR spectroscopy as described in chapter 3.2.2. Non-contact and movement measurements were done on 240 potatoes, while contact measurements were done on all 360 potatoes.

3.3.1.2 Dry Matter Reference Measurements

After NIR measurements the potatoes were treated and dried as described in chapter 3.2.1 with exception that three, instead of four, replicates were taken. The Asterix potatoes were of small size, and it was decided to weigh only 15 ± 1 g for each replicate.

3.3.1.3 Temperature Sensitivity

10 potatoes with a core temperature of 4 °C were measured, as described in the contact test in chapter 3.2.2. This was repeated for every 2 ° increments, until 16 °C. The DM was then measured in the tubers, as described in chapter 3.2.1.

3.3.2 Part 2: Quantitative DM measurements of Potato Tubers

1200 potatoes were measured over the course of two weeks, as described in the contact test in chapter 3.2.2. All the potatoes were provided by Bama Moss, from four different batches, with 300 tubers in each batch. Two batches were of the variety Sava, and the two remaining batches were the variety Rafaela. The samples were chosen based on size and appearance,

using the same criteria for obtaining the model. These tubers were taken directly from the producers, washed and stored overnight at room temperature before measurements.

3.3.3 Part 3: Comparison of Different Sampling Conditions

While gathering data for the model, additional measurements were done on a subset of 240 potato tubers. The extra measurements were non-contact and movement measurements as described in chapter 3.2.2. Table 6 gives an overview of the varieties and samples used for this part of the study.

Table 6 An overview of the potato varieties used in the second part of the study, and how long they were stored at 4 °C before preparation.

	Saturna	**Bruse**	**Fakse**	**Sava**	**Folva**	**Mandel**	**Asterix**
n	40	40	40	30	30	30	30
Storage time (Days)	5 (12)[1]	6 (12)[2]	5	7	6	5	5

[1] 30 Saturna potatoes were stored for five days, ten were stored for 12 days.
[2] 30 Bruse potatoes were stored for six days, ten were stored for 12 days.

3.4 Data Analysis

3.4.1 Analysis of the X-matrix (NIR-Spectra)

For each potato tuber, there were three replicates of spectra per method. All spectra were converted in Matlab, for further use in Unscrambler. The visible range was removed, so as only the NIR spectrum (760 – 1040 nm) was remaining. The spectra were plotted in a line plot to reveal any outliers, which was evaluated for removal from further analysis. An average of the spectra from each tuber, was created, and then SNV transformed (Barnes et al. 1989). Plotting the averages in a PCA plot identified any other outliers, which was evaluated for removal. The entire analysis was done both for the data used in the calibration, and for the quantitative NIR measurements to be predicted.

3.4.2 Analysis of the y-values (DM reference values)

Each potato tuber had three replicates for DM content. DM content was expressed as percentage (%) of total weight of the samples. The data was treated in Excel 2010, where the standard deviation (SD) for each tuber was calculated. Based on the SD, possible outliers were evaluated for removal.

3.4.3 Calibration and Prediction

A regression model, was made by correlating the X-matrix and y-values using PLSR (Martens & Næs 1989) for data set C. Full cross validation was applied to test the strength of the calibration model, and the models were evaluated based on the correlation (R^2), RMSECV and interpretation of the RC. Using the PLSR model, predictions were made of the prediction and temperature data sets.

4. Results and Discussion

4.1 Part 1: Establishing the Calibration

4.1.1 DM Reference Test

Five potatoes of the variety Saturna were chosen for validating the reference test. Four replicates were made for each potato, and the mean and SD were calculated in Excel. The results are shown in table 7.

Table 7 The average DM content and SD for each potato used in the reference test.

Potato	Mean (%)	SD
1	33,32	0,469
2	32,87	0,273
3	33,72	0,473
4	34,29	0,593
5	31,20	0,337

With exception of potato 4, all the potatoes had a SD below 0,5. After reviewing the results it was decided that the reference method could be used further in the study.

4.1.2 DM Reference Data

The following results show the variation of the DM of the potato tubers measured in data set A, with data on the highest, lowest, average and SD of the DM content within one variety of tubers, and in total. These results are shown in table 8.

Table 8 An overview of the data set gathered from the reference data. It shows the samples (n) used, SD, min, max and average DM content for each variety.

Variety	Min % DM	Average % DM	Max % DM	SD DM	n
Saturna	21,8	25,4	28,3	1,42	40
Bruse	25,4	28,1	31,5	1,34	40
Fakse	17,7	20,3	22,6	1,30	40
Asterix	14,7	19,1	23,5	1,98	30
Folva	17,1	20,3	23,8	1,42	30
Sava	14,6	19,4	22,4	1,74	90
Mandel	21,9	25,9	29,3	2,08	30
Rafaela	11,8	16,5	18,6	1,42	60
Total	11,8	21,9	31,5	4,08	360

From each tuber three replicates were made. If the SD of the DM content was above 1,0 the replicates were evaluated for removal. If a replicate deviated much from its two counterparts, it was removed. This process led to the removal of 25 replicates.

The remaining DM data was combined with data set B. The summary of the DM data from that data set is shown in table 9.

Table 9 An overview of the DM measurements in data set B (Helgerud et al. 2012).

Variety	Min % DM	Average % DM	Max % DM	SD DM	n
Asterix	18,0	20,1	21,8	1,0	19
Bruse	21,4	26,4	30,5	1,9	19
Celine	14,8	18,1	20,3	1,6	19
Folva	14,4	16,6	19,7	1,4	29
Saturna	22,2	24,4	27,0	1,4	28
Total	14,4	21,0	30,5	4,1	114

4.1.3 NIR Spectra

For each potato tuber, three replicates of measurements were made, which gave a total of 1080 spectra in the visible and NIR range (460 – 740 nm and 760 – 1040 nm respectively). The visible range of the data was removed from further data treatment, and all the spectra where plotted in a line plot. Figure 9 shows a NIR spectrum obtained from the study.

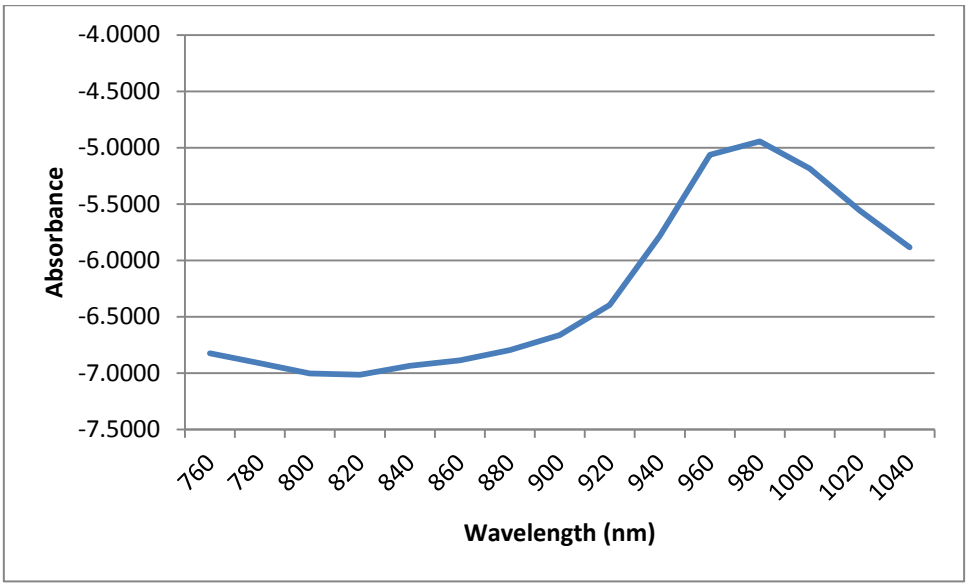


Figure 9 The NIR spectrum for a potato tuber.

The highest absorption is at wavelengths 960 nm and 980 nm. Considering a potato tuber may contain around 80 % of water, the high absorption is probably due to the second OH-water overtone (958 nm and 978 nm)(P. C. Williams & Norris 1990).

If a spectrum deviated from the general trend of the spectra, it would be considered a possible outlier. Two such spectra were found, BR_057c and RA_321a, and were removed. The remainder of the plots was calculated into 360 averages, and SNV corrected.

4.1.3.1 PCA

A PCA plot was created of the spectra in the NIR range (figure 10). The plot shows the main variation and any outliers. Four outliers were identified and evaluated for removal. During data collection, BR_054 and MA_225 was notable different from other potato tubers, and this may be the cause for the large discrepancies. Also FA_89 and 84 were found to be deviating from the main population in the PCA plot. These two tubers had extreme spectra which advocated for removal. This may be because the tubers had physical damages, which were not noticed during data collection, or it may have been human or instrumental error. The four outliers were removed from further analysis. The remaining 356 spectra was denoted data set A.

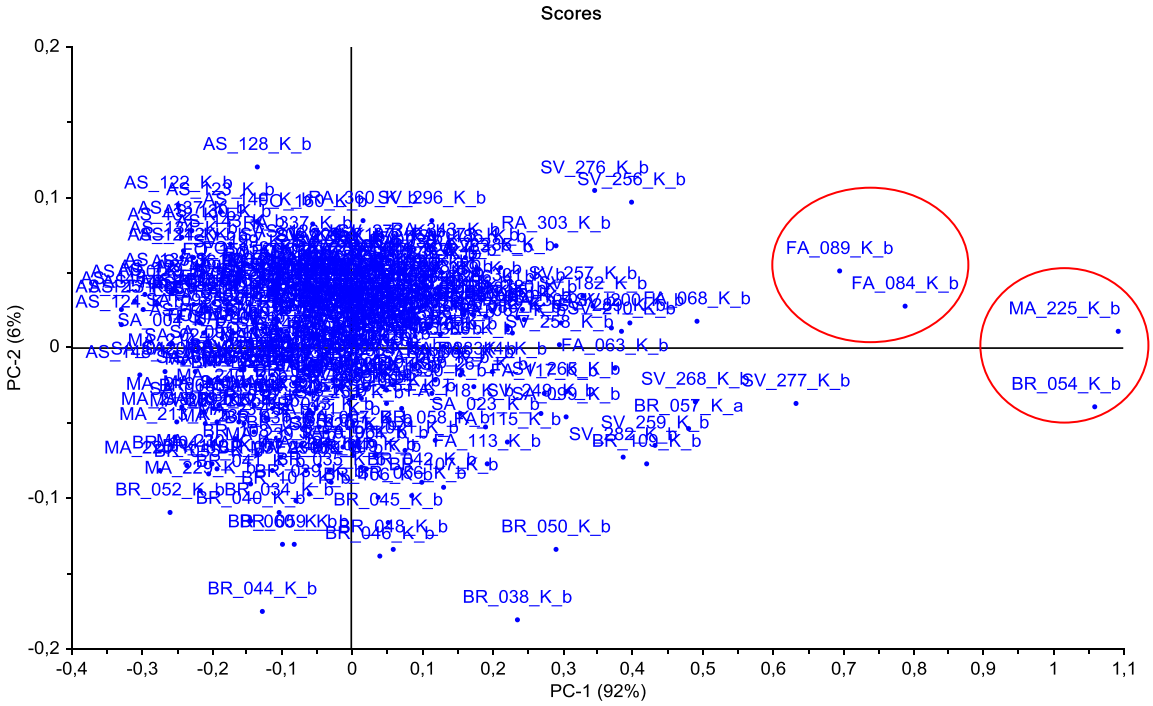


Figure 10 PCA-plot of the average SNV corrected spectra for data set A. Four outliers were found, and are marked in red.

4.1.4 Regression Analysis (PLSR)

For data set A, the preprocessed NIR spectra were connected to the corresponding DM data using PLSR. For validation, a full cross validation was used. A representation of the analysis is shown in figure 11. The R^2 for the model is 0,92, with an RMSECV = 1,17 %. Five PLS factors were used in this model.

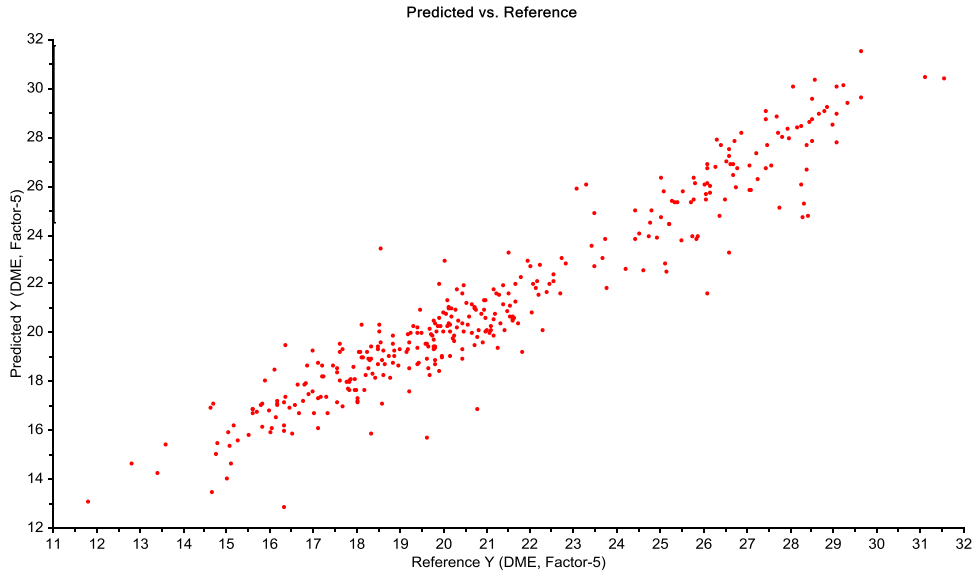


Figure 11 Estimated vs. reference –plot for DM content in percentage of total weight for data set A.

The regression coefficient (RC) for the amount of DM in % of total weight is shown in figure 12. The wavelengths between 900 – 920 nm seem to be the most important regions for estimating the DM.

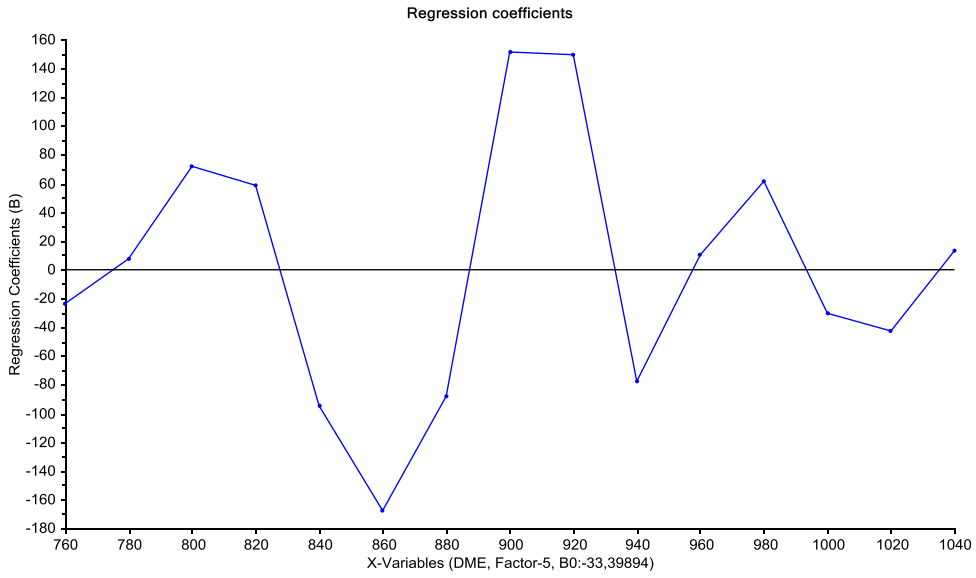


Figure 12 Regression coefficent for the PLSR model of data set A.

4.1.5 Combining the Data Sets

The PLSR plot from data set B is shown in figure 13. The model has a $R^2 = 0,95$, RMSECV = 0,92 % and 5 PLS factors were used.

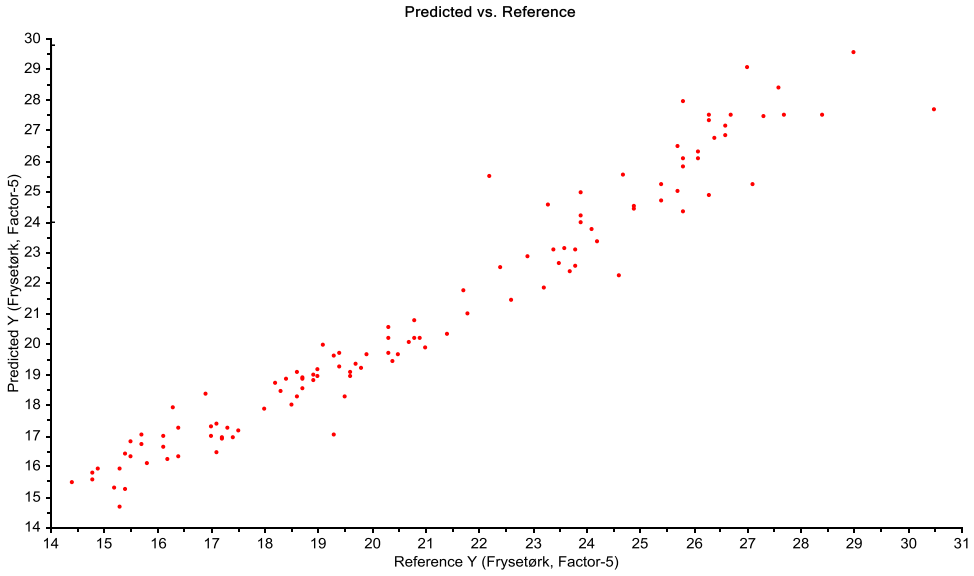


Figure 13 Estimated vs. reference –plot for DM content in percentage of total weight, for data set B.

The RC for the amount of DM in % of total weight for the PLSR model of data set B is shown in figure 14. The model seems to have about the same intensity in tops and bottoms of the plot, and at the same wavelengths (900 – 920 nm) as with the PLSR model of data set A.

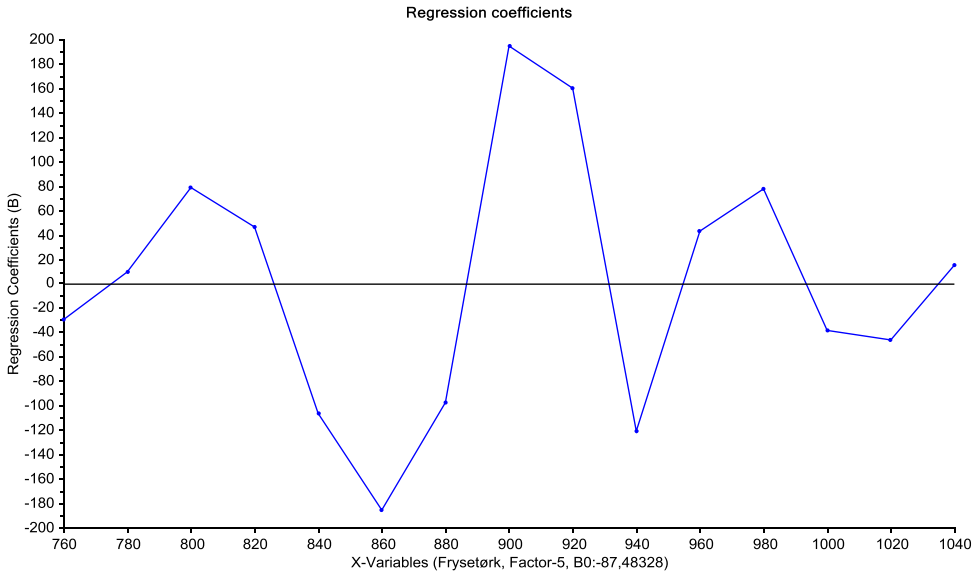


Figure 14 Regression coefficent for the PLSR model of data set B

The two data sets were combined into data set C in Unscrambler, and the PLSR plot for the data set is shown in figure 15. Data set C has in total 470 samples, R^2 = 0,92, a RMSECV = 1,15 % and is a factor 6 model.

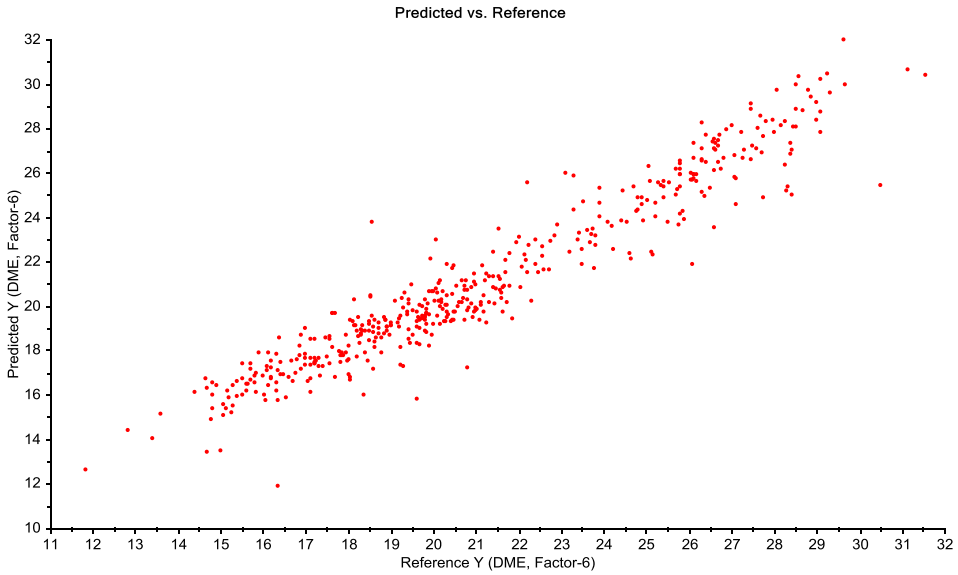


Figure 15 Estimated vs. reference –plot for DM content in percentage of total weight, for data set C.

The RC from the PLSR model of data set C is shown in figure 16. The intensity of the different peaks in the RC are different compared to the separate models, but the same pattern emerges in all three models plots.

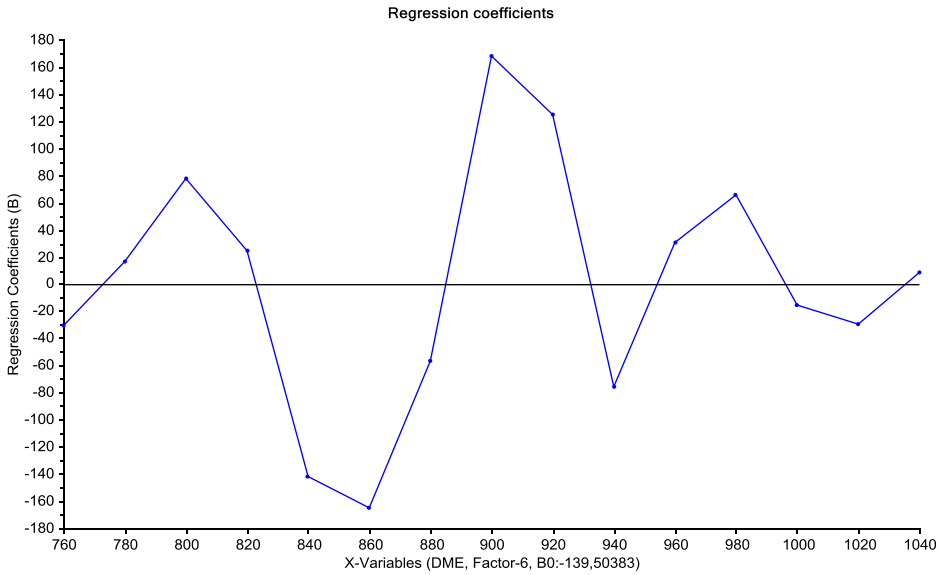


Figure 16 Regression coefficent for the PLSR model of data set C.

The most important data from the three models are collected in table 10.

Table 10 The table shows the explained variance (R^2), RMSECV, how many factors used in the model and how many samples each model consisted of.

Model	R^2	RMSECV (%)	Factor	n
Data Set A	0,92	1,17	5	356
Data Set B	0,95	0,91	5	114
Data Set C	0,92	1,15	6	470

Data set B's model is a bit stronger than data set A and data set C's model, but the differences between the three models are miniscule. On the other hand, the amount of samples used in the combined model makes it more robust then either model by itself. Further research is done with data set C's PLSR model.

4.1.6 Model Strength

To gain a wider understanding of the models strength, a second validation was used. All the data for one potato variety was removed from the model and predicted with the remainder of the model. This was done once for each potato variety. The R^2 values and the number of samples of each variety are presented in table 11. Most of the tubers had a high correlation rate, with notable exception of Fakse.

Table 11 Each variety was removed from the model and predicted with the remaining samples. This was repeated until all varieties had been predicted. The table shows the explained variance (R^2) and the number of samples from each variety.The R^2 for Fakse was not available (NA).

Potato Variety	Saturna	Bruse	Fakse	Asterix	Folva	Sava	Mandel	Rafaela	Celina
R^2	0,91	0,97	NA	0,89	0,91	0,92	0,97	0,97	0,94
n	68	58	38	49	59	59	29	29	19

4.1.6.1 Temperature Data

Ten potato tubers were measured with a core temperature at 4 °C. This was repeated for every two degrees up to 16 °C. Using the reference method, the DM content was measured. PCA plots for all the replicates, at every temperature for each potato was made. In figure 17 the PCA plot for RA_06 is shown.

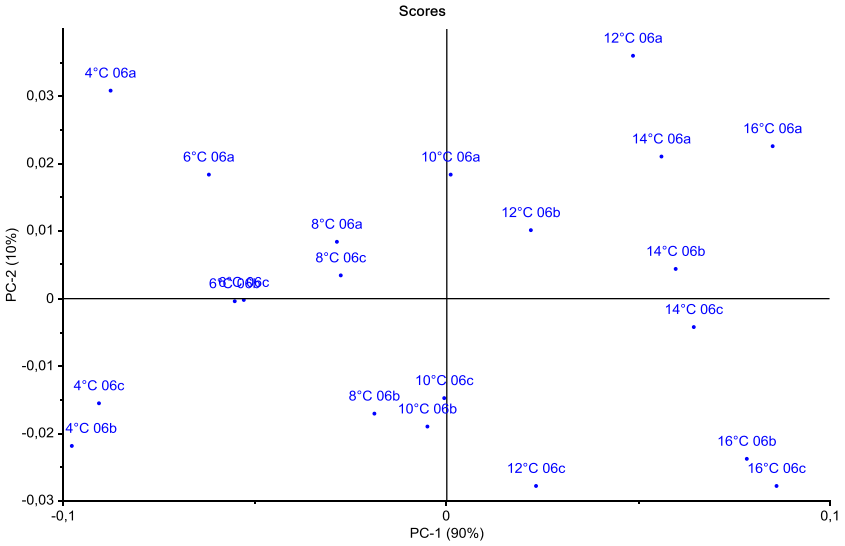


Figure 17 PCA-plot of the average SNV corrected spectra for RA_06 from the temp data set. There is a tendecy for the measurements at the same temperature to group together, when looking at PC1.

The PCA plot shows that there is a tendency that the measurements are affected by the core temperature of potatoes, with 4 °C to left of the plot and 16 °C to the right. Looking at the loadings of the plot (figure 18) it is shown that most of the variation effects are in the wavelength range of 940-1020 nm.

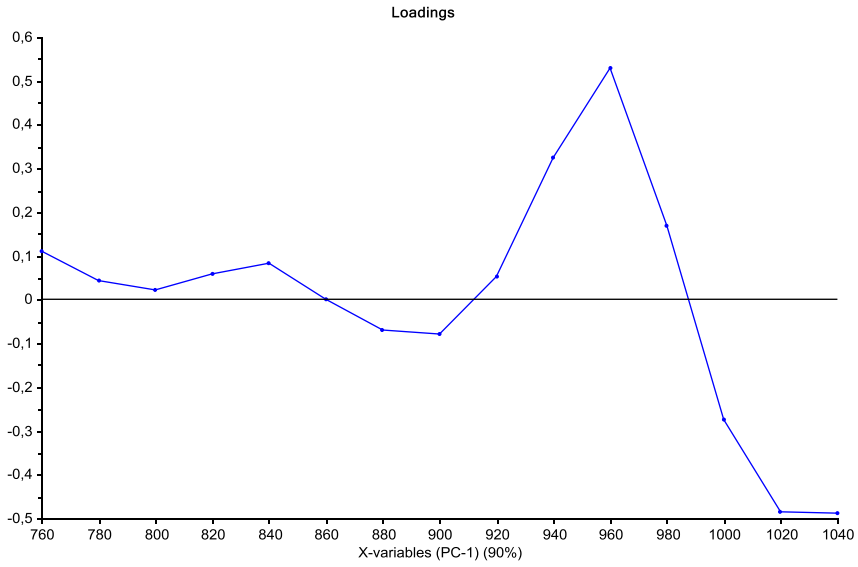


Figure 18 The loading plot for RA_06 from the temp data set. It shows a high effect at wavelenght 960 nm and 1020 nm.

The DM of the potatoes was predicted using the PLSR model from data set C, and compared with the actual DM content (figure 19).

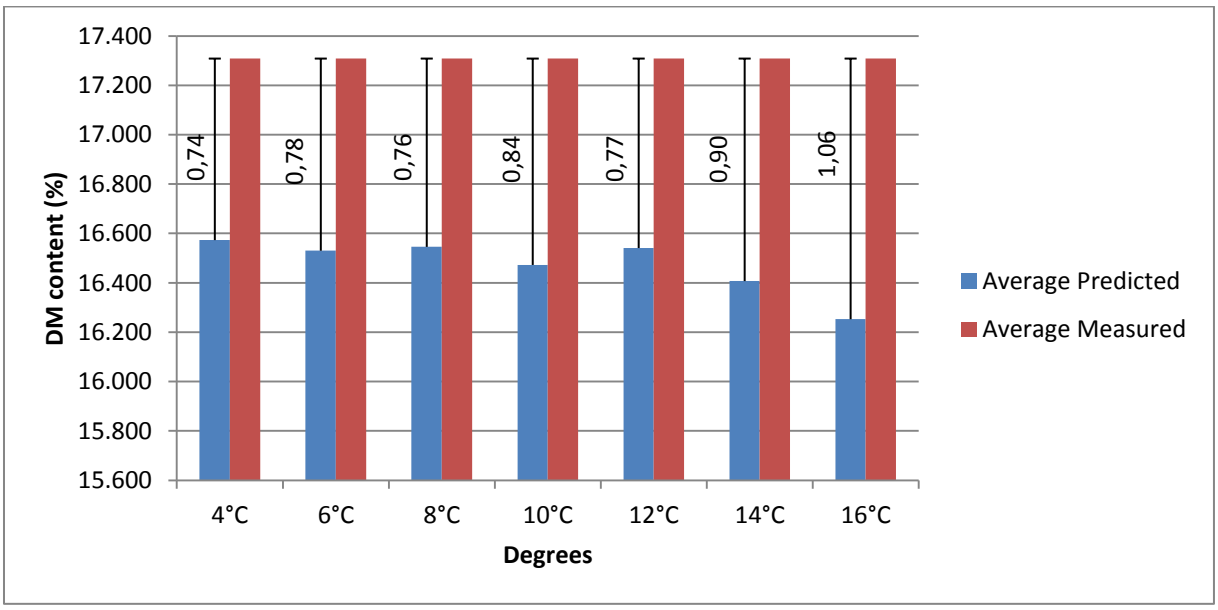


Figure 19 Shows the predicted vs. measured DM content. The black lines shows the difference between the two values in percentage points.

In general all the predictions are lower than the actual DM content, but well inside the RMSECV of the model. The differences in prediction between the temperatures are very small,

4.2 Discussion Part 1

The tubers used in this study was stored at the same temperature, but at different storage times. Even though the DM composition in the tubers may have changed during storage, it should not affect the study, as the DM content was measured right after the NIR measurements. All the measurements were done on tubers with a core temperature between 18 – 19 °C.

4.2.1 DM Data

The DM matter measurements of the potato tubers in data set A, ranged from 11,8 % - 31,5 %, with an average at 21,9 %. This covers most of the natural variation found in tubers (Burton 1989).Comparing the results with the data from data set B, the average DM content is much alike, with exception of Folva.

The maximum and minimum values of the varieties seem to be higher and lower respectively, in data set A compared to data set B. This is to be expected as more samples and varieties were used in data set A, which increases the probability to find extreme DM values. The total SD and average DM content are much alike in both data sets, which implies that the variation is somewhat similar, but with greater extremes. This is positive for the model, as a greater variation will assure a more robust model.

4.2.2 NIR Measurements

A PCA plot was used to show the main variation in the samples gathered, and it showed four outliers. Two of the outliers were deemed safe to remove, as the samples were already commented upon, as odd looking, during data gathering. The remaining two could contain important information about the main variation, but it was observed that the two samples were close to the aforementioned outliers. This may indicate that they had smaller defects, which were not detected earlier. When also considering that the data set was already large, it was decided that removal would not hurt the model.

Using PLSR, the spectra were related with the DM reference data. Comparing the regression models for data set A and B, the variation, RMSECV and the number of factors used, are very similar. Also, when looking at the RC for the models, both models have the most contribution from the wavelengths between 900 – 920 nm, and a slight contribution from wavelength 980 nm. Starch is the highest DM fraction in potato tubers, and it is associated with the second overtone for the carbohydrate OH-stretch at wavelength 901 nm. The same OH-stretch is

found at wavelength 978 nm, which probably explains the slight influence at this point. Starch is also associated with absorbance at wavelength 918 nm (P. C. Williams & Norris 1990).

In the timeframe between the present and Helgerud et als. study, respectively, some adjustments had been made on the NIR instrument. Also, Helgerud used a different method of collecting reference data. This may have caused some differences in the results, but the similarities in both the RC and NIR spectra, indicates that this probably had little effect and the data sets could be combined.

Combining the data sets gave a PLSR regression model with a variance (R^2=0,92) and RMSECV (1,15 %), which were closer to the values from the PLSR model of data set A. This is to be expected as there are more samples in data set A. It should be noted that the combined model uses six factors, compared to data set A and B with five factors. It should still be more reliable as it has more samples in total. The RC shows that the contribution from the wavelengths is unchanged, and it's decided that the PLSR model of data set C is to be used for all predictions.

4.2.2.1 Validation

Full cross validation was used to validate the PLSR models. It was compared with segmented cross validation, using 10 % of the samples for each segment. The difference between the validation methods was miniscule (data not shown), and full cross validation was used for all validations.

4.2.3 Model Strength

Overall the predicted vs. actual DM content is very good. An apparent exception is Fakse, with no measureable R^2 in the regression model. This is a clear indicator that if a NIR measurement system is to be used in the industry; some varieties might either need their own models, or a high presence in a model, for correct prediction. This is something that has to be investigated for each variety.

4.2.3.1 Temperature Data

The PCA plot (figure 17) for the temperature data for RA_06, shows that there is an effect of temperature on the measurements. The data collected at 4 °C is more prominent at wavelength 1020 nm (figure 18), while data collected at 16 °C is more prominent at wavelengths 960-980 nm. Predicting the DM content in the tubers at different temperatures, always gives a lower DM content then there actually is, while at the same time there is little difference between the predictions. The low predictions imply that there is a systematic error in the prediction, as

there should be an equal amount of too high predictions. The small difference between predictions related to temperature, can be explained by the RC for data set C's PLSR model. The most important wavelengths for predicting DM in potato tubers are between 900 -920 nm and temperature affects the area at 960 – 1020 nm. This indicates that the model is probably not affected by temperature. Even though very few samples were used for this experiment, the tendency is clear, but even so, another test with a larger data set should be done. Further research could examine if an increase or decrease in water content in tubers, or if predicting samples of varieties which are not in the model, will make the model more sensitive for temperature.

4.3 Part 2: Prediction of Tubers

4.3.1 Pretreatment

The NIR spectra of 1199 tubers were analyzed in a line plot and samples SV_149a-c and SV_276a outliers were removed. After calculating the average of each tuber, and SNV correction, the remaining spectra were plotted in a PCA plot (figure 20). Four outliers were identified, and evaluated for removal. During data collection, the tubers SV_0557 and SV_0488 were notable different from the other samples, and were removed. SV_0579 and SV_0489 were showing tendencies for being extreme values in the PCA plots, and were removed with the aforementioned outliers.

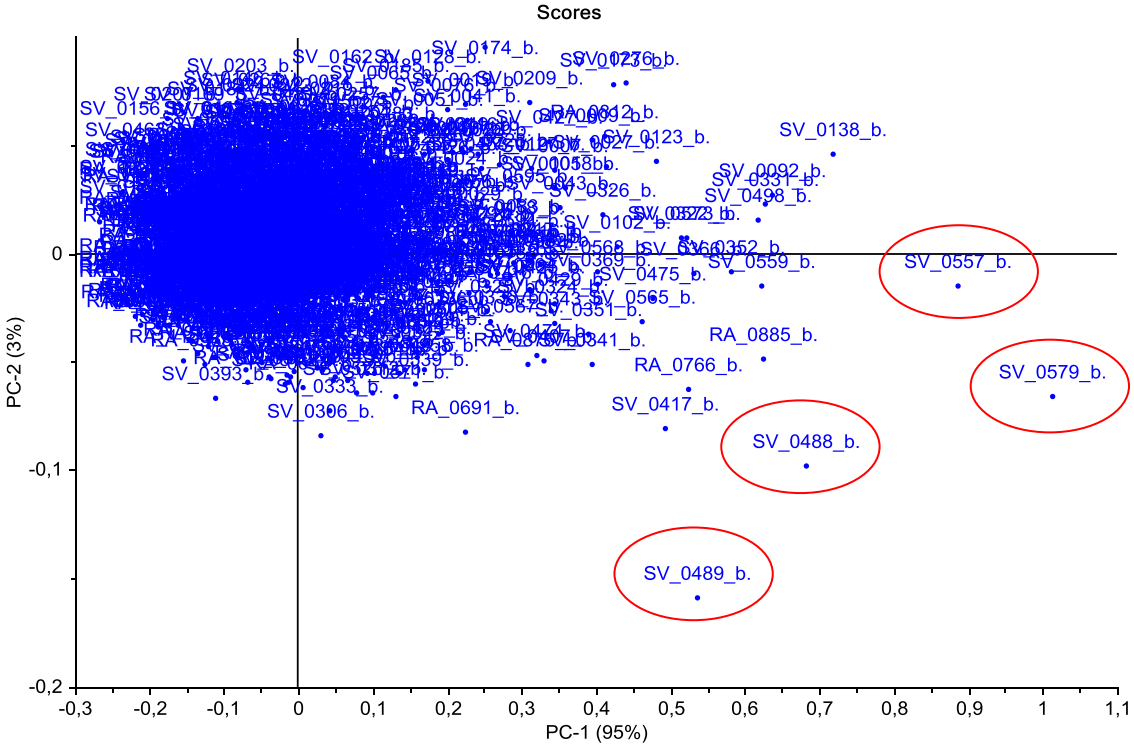


Figure 20 21 PCA-plot of the average SNV corrected spectra for the prediction data set. Four outliers were found (marked in red) and were removed.

The remaining 1194 spectra were sorted into their respective batches, and predicted with regression analysis using the PLSR model of data set C. Histograms of the DM content in the potato tubers for the different batches are shown in figure 22a – d.

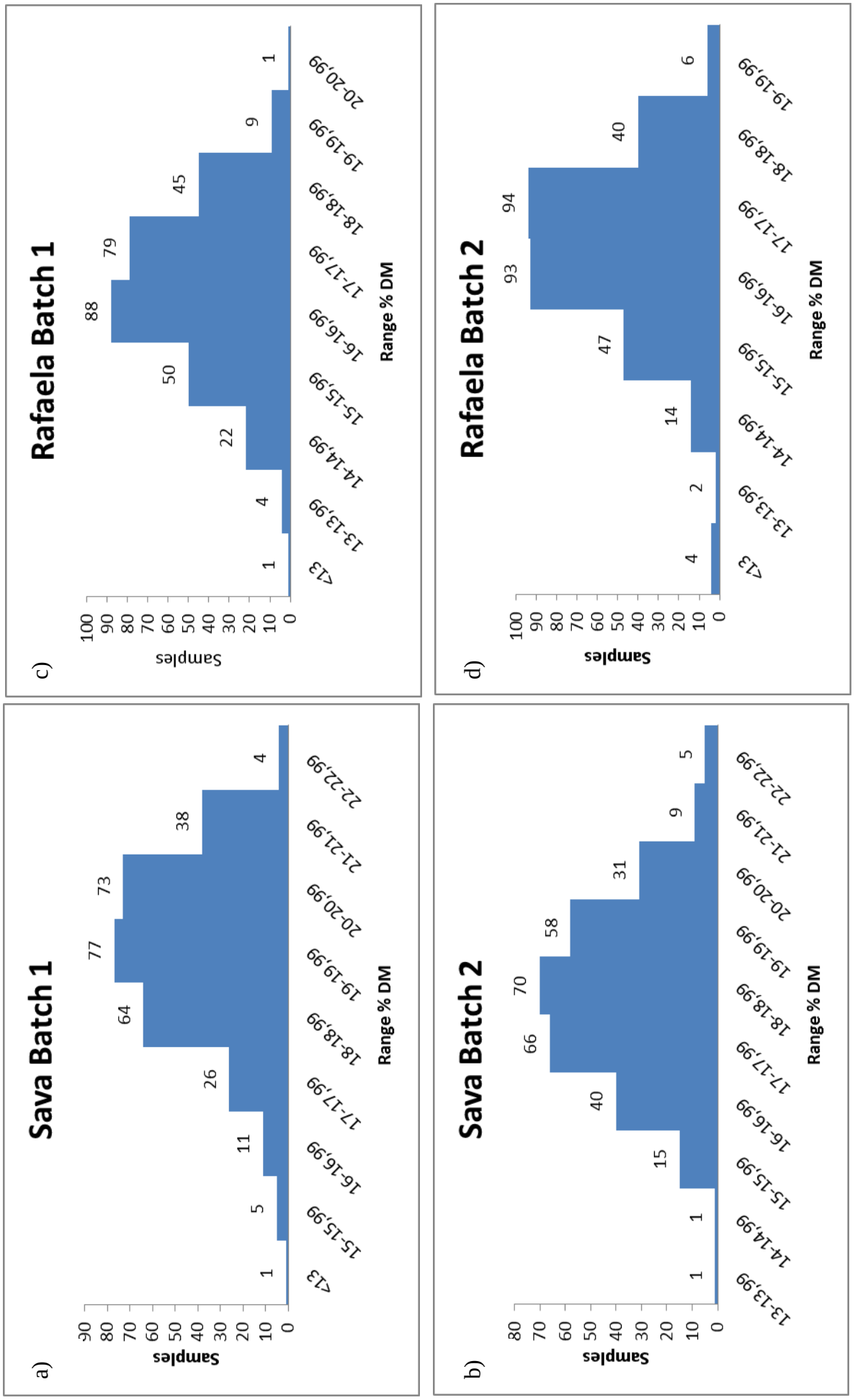


Figure 22 Histogramplots of the DM content variation for the predicted batches of potatoes. a) DM content variation for Sava batch 1 b) DM content variation for Sava batch 2 c) DM content variation for Rafaela batch 1 d) DM content variation for Rafaela batch 2

The Sava batches differ somewhat in the DM distribution, but all four histograms show that the DM distribution is wide.

A short summation of the DM data for the different batches of potato tubers are shown in table 12.

Table 12 An overview of the DM content in the different batches predicted.

	Min	Average	Max	SD	n
Sava 1	12,1	19,4	22,7	1,5	299
Sava 2	14,0	18,4	22,9	1,6	296
Rafaela 1	9,2	16,8	20,0	1,3	299
Rafaela 2	11,9	16,8	19,3	1,2	300

Predicting the different batches of potatoes gave SD's which deviated little from batch to batch. Comparing the averages of the two Sava batches, gives a small difference, but this is within the RMSECV of the model. The variation of the DM content is different, with Sava batch 1 having more samples in a higher range. This difference can also be seen in the Rafaela batches. Three of the batches show tendencies for normal distribution. The exception, Rafaela batch 2, seems to have a more narrow spread of DM content.

4.4 Discussion Part 2

The prediction of the potato tubers has shown that there is a high variation of DM content in potato tubers. This becomes more important when considering the averages of the different batches, within varieties, did not differ as much. In the industry, they predict the DM content in a small outtake from every batch, and it is assumed that these results apply to the entire batch. If applying the same principal, and assume that the average is correct for the entire batch, only about a quarter of the samples would fall within range of 1 % of the average. Considering that the industry uses even smaller outtakes, there is a good chance that an even smaller part of the batch is within 1 % of the average. The tendencies for normal distribution in the histograms, implies that the model used is robust.

4.5 Part 3: Comparison of Different Sampling Conditions

Using a subset of 240 samples from data set A, three extra data sets were made: contact data, non-contact data and movement data.

4.5.1 Contact Data

4.5.1.1 PCA

The contact data from the NIR measurements gave 720 spectra. The visible range of the data was removed from further data treatment, and the NIR spectra where plotted in a line plot. This exposed any outliers in the data, and the replicate BR_057c was removed. The average from the remaining spectra was transformed with SNV and the data was shown in a PCA plot (figure 23).

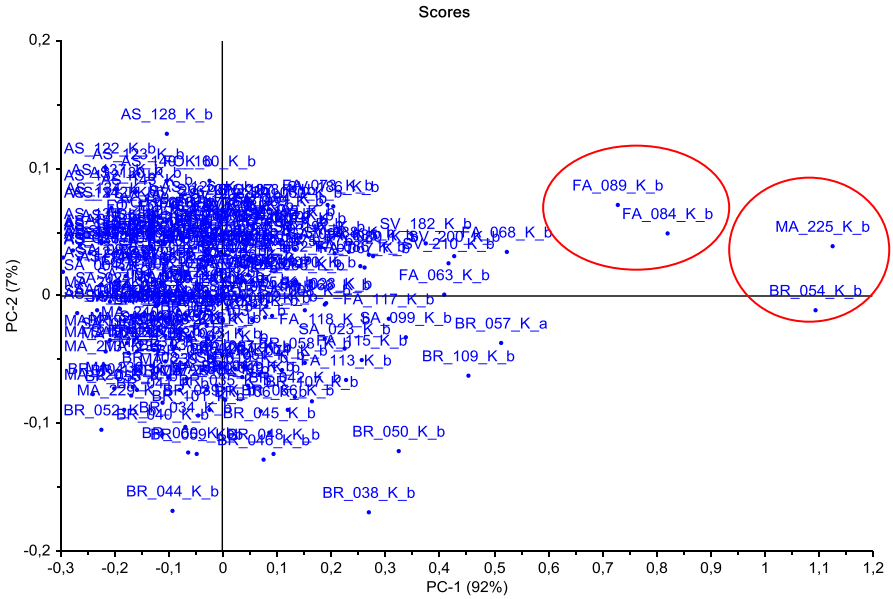


Figure 23 PCA-plot of the average SNV corrected spectra for the contact data set. Four outliers were found (marked in red) and were removed.

The PCA plot shows the entire absorption spectrum (760 nm – 1040 nm), and points out more clearly any outliers. Four outliers were found which clearly stood out from the rest of the data. MA255 and BR054 were noted during NIR measurements to have dark or odd skin that probably explains the deviation. FA089 and FA084 deviated so much from the main spectra, they were considered to be extreme and were removed. The potatoes could have had the same problem as aforementioned outliers, but there is also a possibility for a human or instrumental error.

4.5.1.2 PLSR

Figure 24 shows the PLSR model of the remaining 236 samples. The model had a R^2 = 0,89, a RMSECV = 1,19 % and was a factor 4 model.

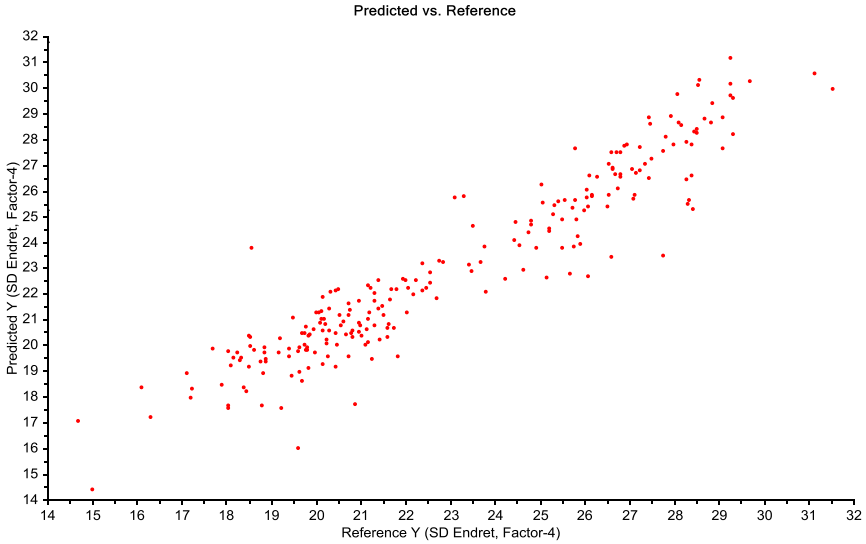


Figure 24 Estimated vs. reference –plot for DM content in percentage of total weight, for contact data set.

4.5.2. Non-Contact Data

4.5.2.1 PCA

The NIR measurements of the non-contact data gave 720 spectra. They were treated in the same manner as the contact data, and MA_225c was removed based on the line plots. The averages of the remaining spectra were plotted in a PCA plot figure 25.

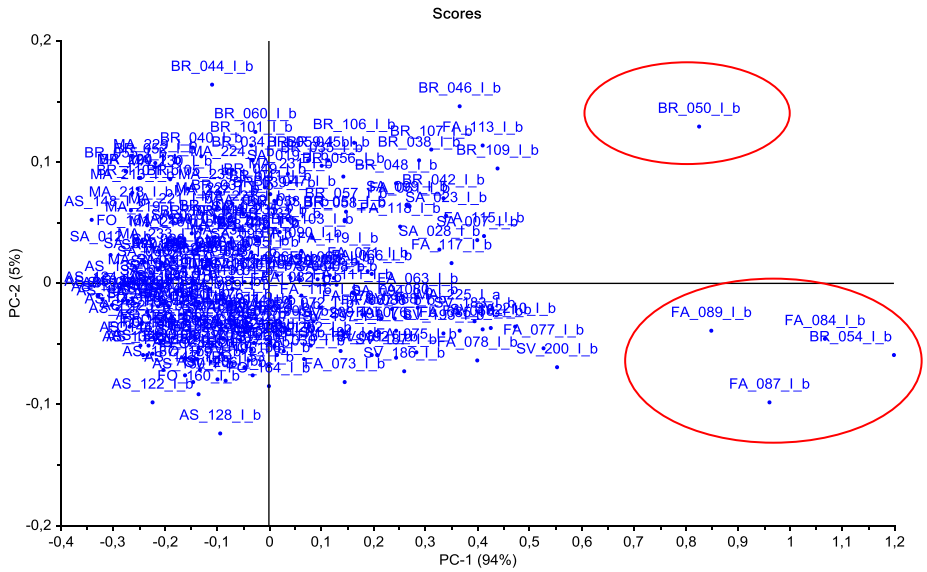


Figure 25 PCA-plot of the average SNV corrected spectra for the non-contact data set. Five outliers were found (marked in red) and were removed

The same four samples that were identified as outliers in the PCA contact data were found to be outliers here also. In addition BR_050 was found to be an outlier. All the samples were found to be far removed from the main population, and it was decided that they should be removed from further treatment.

4.5.2.2 PLSR

Multivariate calibration with PLSR was performed on the remaining 235 samples. The model is shown in figure 26 and had a R^2 = 0,89, a RMSECV = 1,23 % and used 5 factors.

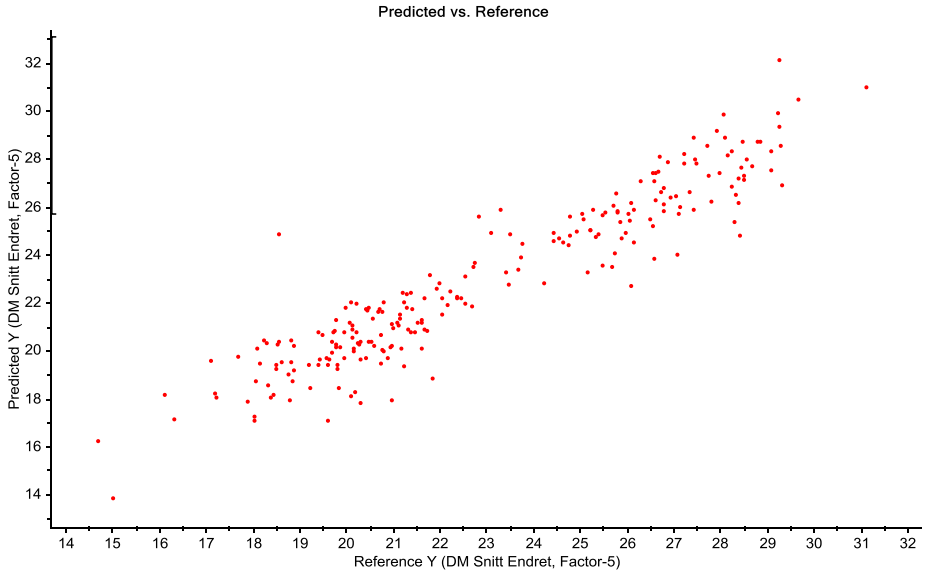


Figure 26 Estimated vs. reference –plot for DM content in percentage of total weight, for the non-contact data set**.**

4.5.3. Movement Data

The spectra acquired from the moving potato tubers differ from the spectra acquired from the earlier methods. 5 seconds of measurements were done, which gave much noise spectra from the conveyor belt. An example of the spectra acquired can be seen in figure 27.

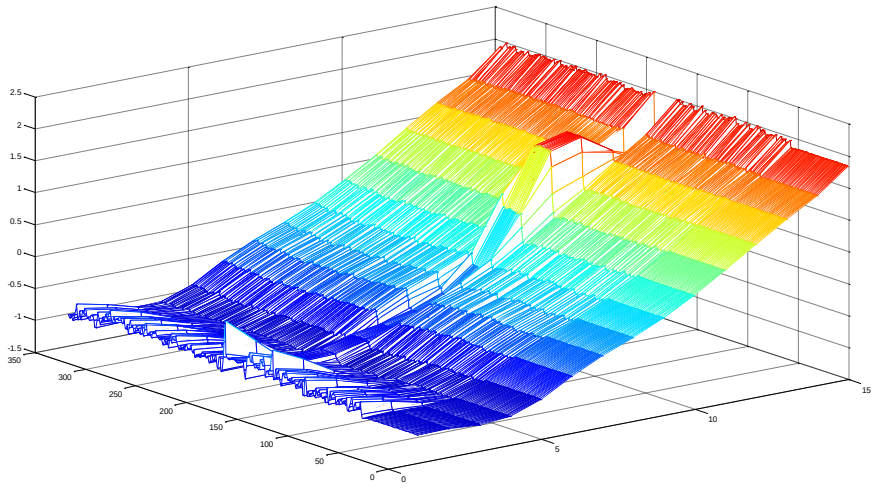


Figure 27 The spectra gathered from one tuber. Most of the spectra are from the conveyor belt, but the small wave in middle is where the potato tuber passed the collection tube.

Most of the lines seem to be more or less the same, with exception of a group of lines in the center. This area is where the tuber passed the lights and collection tube of the instrument. There were about 20 spectra in this area. Removing all other data and calculating the average of the 20 spectra, would give one replicate. The process gave satisfactory replicates for each tuber, except for MA_225a-c and FA_120c which were removed for further analysis.

4.5.3.1 PCA

The NIR measurements of the movement data gave 716 spectra. Plotting the spectra in line plots revealed no outliers. All the spectra was then averaged and plotted in a PCA plot (figure 28).

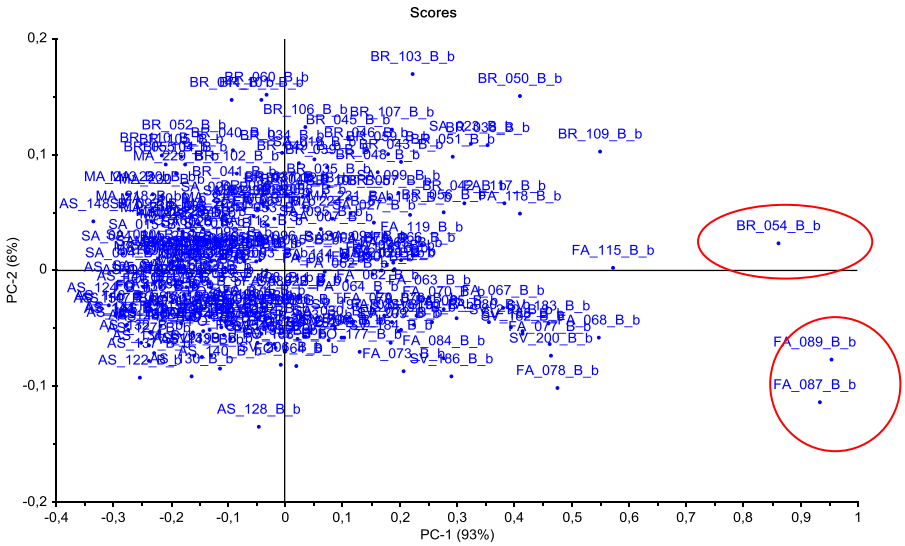


Figure 28 PCA-plot of the average SNV corrected spectra for the movement data set. Three outliers were found (marked in red) and were removed.

With the exception of MA_225, the same samples that were identified as outliers earlier were also identified here. As with the earlier PCA plots, the outliers were removed.

4.5.3.2 PLSR

Multivariate calibration with PLSR was performed on the remaining 235 samples. The model had a R^2 = 0,92, a RMSECV = 1,06 % and was a factor 5 model. The PLSR model is shown in figure 29.

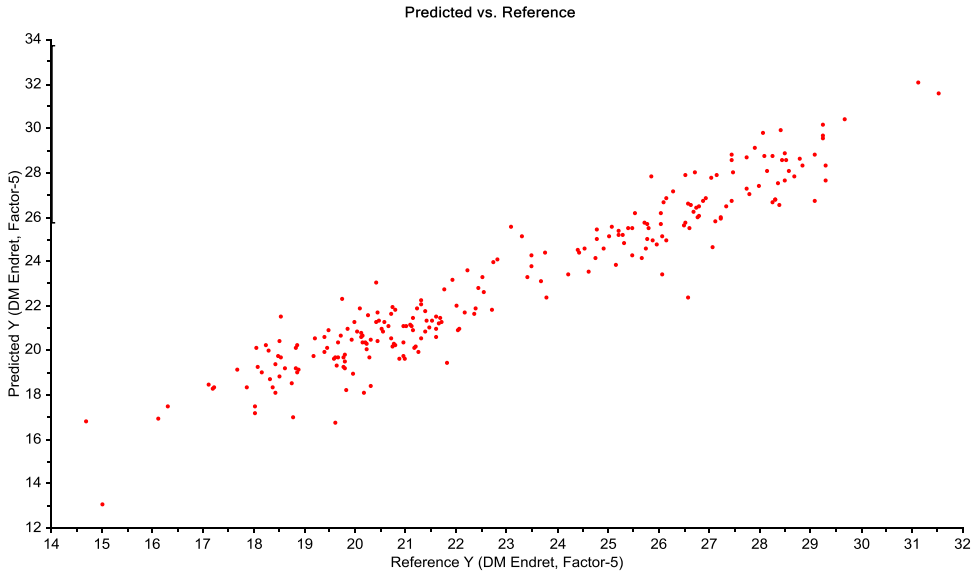


Figure 29 Estimated vs. reference –plot for DM content in percentage of total weight, movement data set.

4.5.4 Summation

The data collected from the three PLSR regression analyses are collected in table 13.

Table 13 An overview of the data from the PLSR models from each of the data sets.

	Contact Data	Non-Contact Data	Movement Data
R^2	0,89	0,89	0,92
RMSECV (%)	1,19	1,23	1,06
Factor	4	5	5
n	236	235	235

The R^2 values are slightly better in the movement data, which may be explained that a greater surface area of the tuber was measured when passing the light beam.

4.6 Discussion Part 3

With one exception, the same outliers were found in all three PCA plots, which suggest that the different conditions have little to say for the data collection. The three regression models showed little difference both in R^2 and RMSECV, but the movement model was marginal better. This is despite of only registering approximately 20 spectra when each potato is passing. A possible explanation for the better result may be that since the potatoes passed the collection tube, a larger surface area was measured. The results are a clear indicator that it is possible to use NIR spectroscopy on moving samples. Comparing the RC of the tree models supports this, as they are more or less the same (data not shown). Based on the number of spectra acquired, the speed of the conveyor belt was about 3-4 potatoes per second.

Comparing the R^2 values from the movement data with earlier work by Subedi & Walsh (2009) the model in this study is stronger (0,92 vs. 0,85 respectively). It should be mentioned that they used fewer samples (50) and sliced tubers. Since the light cannot penetrate the potato further then 20 mm, and a tuber is highly heterogeneous, it could be assumed that measurements covering the entire diameter of the tuber would yield better results. But having many samples in a model seems to counteract this effect. It could also have something to do with the varieties used.

An interesting observation was done during data interpretation regarding the model and contact data. When using PLSR regression on the contact data in part 2, the PLSR model was a factor 4 model. Comparing this with the PLSR model of data set A, which had more samples, achieved a factor 5 model. Adding more samples to the model, by combining data set A and B, the PLSR regression returned a model using 6 factors. This implies that the combined model has a very high variation, which further implies that the model created here is robust.

5. Summary and Future Work

The PLSR model of data set C had very good results (R^2 = 0,92 and RMSECV = 1,15 %), and it was shown that NIR spectroscopy can be used to quickly gather a large database of DM variation in tubers. The predictions indicated that tubers from the same batch can have a large range of DM content. This coincides with earlier research (Cole 1975). The difference in average DM content between the respective batches was small, but the variation of DM content was slightly different. Compared to the total size of the batch, the outtake was very small. This may mean that the variation is actually much larger. Considering that the industry use even smaller sets of outtakes, indicates that the assumed average DM content in a batch of potatoes is not correct. Using NIR to continually measure the DM in potatoes would remove this uncertainty. It would be interesting if this method could be implemented in the industry, to gain a broader knowledge of the variation of DM content in potatoes. If this was to be done, it would also be natural to develop a method were not only the DM content will be predicted, but also the amount of different constituents in potatoes.

This study has shown that it is possible to develop a working model for non-destructive, on line and rapid NIR measurements of DM content in potato tubers. The R^2 and RMSECV of the movement PLSR model, are equally good as the PLSR model of data set C. This is in spite of acquiring fewer spectra per reading, and having a lower amount of samples. Comparing the movement model with an earlier In-line model developed by Brunt et al. (2010) using SG, shows that their model had higher R^2 (0,985). Still, the advantage of using NIR spectroscopy is that is non-destructive, and the movement data has shown that it is possible to measure 3 – 4 potatoes per sec (against Brunt et als 12 samples per hour). Furthermore, increasing the amount of samples in the model, will probably increase the robustness of the model. How many samples NIR spectroscopy can measure per second, is something for future research, and is necessary for the industry to be able to utilize this instrument.

References

Anonymous, (1999) The commission of the european communities. *Official Journal of the European Union*, 42 (L327/37 (21.12.1999)).

Anzaldúa-Morales, A., Bourne, M.C. & Shomer, I., (1992) Cultivar, Specific Gravity and Location in Tuber Affect Puncture Force of Raw Potatoes. *Journal of food science*, 57 (6): 1353 – 1356.

Barnes, R.J., Dhanoa, M.S. & Lister, S.J., (1989) Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43 (5): 772–777.

Bemiller, J.N. & Huber, K.C., (2008) *Fennema's Food Chemistry* 4th ed. S. Damodaran, K. L. Parkin, & O. R. Fennema, eds., Boca Raton :CRC Press

Bertoft, E. & Blennow, A., (2009) *Advances in Potato Chemistry and Technology* 1st ed. J. Singh & L. Kaur, eds., Elsevier

Brecht, J.K. et al., (2008) *Fennema's Food Chemistry* 4th ed. S. Damodaran, K. L. Parkin, & O. R. Fennema, eds., CRC Press

Brunt, K., Smits, B. & Holthuis, H., (2010) Design, Construction, and Testing of an Automated NIR In-line Analysis System for Potatoes. Part II. Development and Testing of the Automated Semi-industrial System with In-line NIR for the Characterization of Potatoes. *Potato Research*, 53 : 41–60.

Bundesssortenamt, (2012) Beschreibende Sortenliste Kartoffeln. Available at: http://www.bundessortenamt.de/internet30/fileadmin/Files/PDF/bsl_kartoffeln_2012.pdf [Accessed May 9, 2013].

Burton, W.G., (1989) *The Potato* 3rd ed., New York :Longman Group UK

Calleman, C.J. et al., (1993) A nonlinear dosimetric model for hemoglobin adduct formation by the neurotoxic agent acrylamide and its genotoxic metabolite glycidamide. *Environmental health perspectives*, 99 : 221–223.

Chalmers, J.M. & Griffiths, P.R., (2010) *Application of Vibrational Spectroscopy in Food Science Volume I* 1st edt. E. C. Y. Li-Chan, P. R. Griffiths, & J. M. Chalmers, eds., Chichester, UK :John Wiley & Sons

Cole, C.S., (1975) Variation in dry matter between and within potato tubers. *Potato Research*, 18 (May 1974): 28–37.

Dull, G.G., Birth, G.S. & Leffler, R.G., (1989) Use of Near Infrared Analysis for the Nondestructive Measurements of Dry Matter in Potatoe. *American Potato Journal*, 66 : 215–225.

Dwelle, R.B. & Stallknecht, G.F., (1978) Respiration and Sugar Content of Potato Tubers as Influenced by Storage Temperatures. *American Potato Journal*, 55 : 561–571.

Elert, G., Electromagnetic Spectrum. *The Physics Hypertextbook*. Available at: http://physics.info/em-spectrum/ [Accessed April 5, 2013].

Esbensen, K.H., (2009) *Multivariate Data Analysis* 5th edt., Esbjerg, Aalborg University :Camo Software AS

Ewing, E.E., Senesac, A.H. & Sieczka, J.B., (1981) Effects of Short Periods of Chilling and Warming on Potaot Sugar Content and Chipping Quality. *American Potato Journal*, 58 : 633–647.

Fagforum Potet, Potetsortsinformasjon. Available at: http://www.potet.no/sortsinformasjon/ [Accessed February 21, 2013].

Folkestad, A. et al., (2008) Rapid and non-invasive measurements of fat and pigment concentrations in live and slaughtered Atlantic salmon (Salmo salar L.). *Aquaculture*, 280 : 129–135.

Food and Agriculture Organization of the United Nations, (2013) FAOSTAT. Available at: http://faostat3.fao.org/home/index.html#VISUALIZE_TOP_20 [Accessed May 9, 2013].

Friedman, M., (2003) Chemistry , Biochemistry , and Safety of Acrylamide . A Review. *Journal of Agricultur and Food Chemistry*, 51 : 4504–4526.

Gelder, W.M.J. van & Vonk, C.R., (1980) Amino acid composition ofcoagulable protein from tubers of 34 potato varieties and its relationship with protein content. , 23 : 427–434.

Griffiths, P.R., (2010) *Application of Vibrational Spectroscopy in Food Science Volume I* 1st edt. E. C. Y. Li-Chan, P. R. Griffiths, & J. M. Chalmers, eds., Chichester, UK :John Wiley & Sons

Haase, N.U., (2003) Estimation of Dry Matter and Starch Concentration in Potatoes by Determination of Under-water Weight and Near Infrared Spectroscopy. *Potato Research*, 46 : 117–127.

Haase, N.U., (2011) Prediction of Potato Processing Quality by Near Infrared Reflectance Spectroscopy of Ground Raw Tubers. *Journal of Near Infrared Spectroscopy*, 19 : 37–45.

Haase, N.U., (2006) Rapid Estimation of Potato Tuber Quality by Near-Infrared Spectroscopy. *Starch/Stärke*, 58 : 268–273.

Hartmann, R. & Büning-Pfaue, H., (1998) NIR Determination of Potato Constituents. *Potato Research*, 41 : 327–334.

Helder, J., (1994) *Tuber formation in the wild potato species Solanum demissum Lindl* . Wageningen University.

Helgerud, T. et al., (2012) Near-infrared Spectroscopy for Rapid Estimation of Dry Matter Content in Whole Unpeeled Potato Tubers. *Journal of Food Research*, 1 (4): 55–65.

Hoover, R., (2001) Composition, molecular structure, and physicochemical properties of tuber and root starches: a review. *Carbohydrate Polymers*, 45 (3): 253–267. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0144861700002605.

Houghland, G.V.C., (1966) New Converstion Talbe for Specific Gravity, Dry Matter, and Starch in Potatoes. *American Potato Journal*, 43 : 1966.

Intertek Plastics Technology Laboratories, Picture of "Atomic Stretching and Bending". Available at: http://www.ptli.com/testlopedia/tests/FTIR-E168andE1252-more.asp [Accessed February 20, 2013].

Isaksson, T. et al., (1996) On-line, proximate analysis of ground beef directly at a meat grinder outlet. *Meat science*, 43 : 245–53.

Kapoor, a. C., Desborough, S.L. & Li, P.H., (1975) Potato tuber proteins and their nutritional quality. *Potato Research*, 18 (3): 469–478.

Khanbari, O.S. & Thompson, a. K., (1993) Effects of amino acids and glucose on the fry colour of potato crisps. *Potato Research*, 36 (4): 359–364.

Khuriyati, N., Matsuoka, T. & Kawano, S., (2004) Precise near infrared spectral acquisition of intact tomatoes in interactance mode. *Journal of Near Infrared Spectroscopy*, 395 : 391–395.

Kirkpatrick, M.E., Mountjoy, B.M. & Albright, L.C., (1951) Cooking Quality, Specific Gravity, and Reducing-Sugar Content of Early-Crop Potatoes. *United States Department of Agriculture*, Circ., 872 .

Li-Chan, E.C.Y., (2010) *Application of Vibrational Spectroscopy in Food Science Volume I* 1st edt. E. C. Y. Li-Chan, P. R. Griffiths, & J. M. Chalmers, eds., Chichester, UK :John Wiley & Sons

Lunden, A.P., (1956) Undersøkelser Over Forholdet Mellom Potetenes Spesifikke Vekt og Deres Tørrstoff- og Stivelsesinnhold. *Forskning og forsøk i landbruket*, 7 : 81–107.

Martens, H. & Næs, T., (1989) *Multivariate Calibration*, Guildford :John Wiley & Sons 1st chapter

Mathews, C.K., Van Holde, K.E. & Ahern, K.G., (2000) *Biochemistry* 3rd ed., San Francisco :Addison-Wesley Publishing Company

Matvaretabellen, Matvaretabellen. Available at: http://www.matvaretabellen.no/potatoes-g6.1 [Accessed March 22, 2013].

Merritt, W. & Settle, D., (1981) *Instrumental Methods of Analysis* 6th edt., Belmont, Ca. :Wadesworth Publishing Company

Montgomery, D.C., Peck, E.A. & Vining, C.G., (2001) *Introduction to linear regression analysis* 3rd ed., New York, USA :John Wiley & Sons

Nissen, M., (1955) The Weight of Potatoes in Water. *American Potato Journal*, 32 : 332–339.

Nissen, M., (1967) The weight of potatoes in water: Further studies on the relation between the dry matter and starch content. *European Potato Journal*, 10 (2): 85–99.

Næs, T., (2002) *A user-friendly guide to multivariate calibration and classification*, Chichester :NIR Publications

Pisarczyk, J.M., (1982) Field harvest damage affects potato tuber respiration and sugar content. *American Potato Journal*, 59 (5): 205–211.

Porter, W. l., Fitzpatrick, T.J. & Talley, E.A., (1964) Studies of the Relationship of Specific Gravity to Total Solids of Potatoes. *American Potato Journal*, 41 : 329–336.

Pots, M. et al., (1999) The effect of storage of whole potatoes of three cultivars on the patatin and protease inhibitor content ; a study using capillary electrophoresis and MALDI-TOF mass spectrometry. *Journal of the Science of Food and Agriculture*, 79 : 1557–1564.

Pritchard, M.K. & Scanlon, M.G., (1997) Mapping dry matter and sugars in potato tubers for prediction of whole tuber process quality. *Canadian Journal of Plant Science*, 77 (3): 461–467.

Salunkhe, D.., Kadam, S.S. & Jadhav, S.J., (1991) *Potato: Production, Processing and Products* 1st ed., Boca Raton :CRC Press

Scanlon, M.G., Pritchard, M.K. & Adam, L.R., (1999) Quality Evalutation of Processing Potatoes by Near Infrared Reflectance. *Journal of the Science of Food and Agriculture*, 79 : 763–771.

Von Scheele, C., Svensson, G. & Rasmusson, J., (1937) Die Bestimmung des Stärkegehalts und der Trockensubstanz der Kartoffel mit Hilfe des spezifischen Gewichts. *Landwirtschaftliche Versuchsstation*, (127): 67–96 (cited from Burton, W.G. 1989).

Schippers, P.A., (1977) The rate of respiration of potato tubers during storage . 3 . Relationships between rate of respiration , weight loss and other variables. , 20 : 321–329.

Sharma, M.K., Isleib, D.R. & Dexter, S.T., (1958) Specifiv Gravity of Different Zones within Potato Tubers. *American Potato Journal*, 35 : 784–788.

Simmonds, N.W., (1977) Relations between specific gravity , dry matter content and starch content of potatoes. *Potato Res.*, 20 (2): 137–140.

Singh, J. et al., (2008) Rheological and Textural Characteristics of Raw and Par-Cooked Taewa (Maori potatoes) of New Zealand. *Journal of Texture studies*, 39 : 210–230.

Sirisomboon, P., Hashimoto, Y. & Tanaka, M., (2009) Study on non-destructive evaluation methods for defect pods for green soybean processing by near-infrared spectroscopy. *Journal of Food Engineering*, 93 : 502–512.

Stadler, R.H. et al., (2002) Acrylamide from Maillard reaction products. *Nature*, 419 (6906): 449–50.

Stone, M., (1974) Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society*, 36 (2): 111–147.

Struik, P.C., (2007) Above-Ground and Below-Ground Plant Development. In *Potato Biology and Biotechnology Advances and Perspectives*. Oxford :Elsevier, Chp 11.

Struik, P.C. & Van Voorst, G., (1986) Effects of drought on the initiation , yield , and size distribution of tubers of Solanum tuberosum L . cv . Bintje. *Potato Research*, 29 : 487–500.

Subedi, P.P. & Walsh, K.B., (2009) Assessment of Potato Dry Matter Concentration Using Short-Wave Near-Infrared Spectroscopy. *Potato Research*, 52 : 67–77.

Wikimedia Commons, (2013) EM Spectrum Properties. Available at: http://en.wikipedia.org/wiki/File:EM_Spectrum_Properties_edit.svg [Accessed February 1, 2013].

Williams, D. & Fleming, I., (2008) *Organic Chemistry* 6th edt., New York :McGraw-Hill Education

Williams, P.C. & Norris, K.H., (1990) Qualitative Applications of Near-Infrared Reflectance Spectroscopy. In P. Williams & K. Norris, eds. *Near-Infrared Technology in the Agricultural and Food Industries*. St. Paul, Minnesota :American Association of Cereal Chemists, 241 –246.

Wilson, J.H. & Lindsay, A.M., (1969) The relationship between specific gravity and dry matter vontent of potato tubers. *American Potato Journal*, 46 (9): 323–328.