Characterization of Forest Tree Seed Quality with Near Infrared Spectroscopy and Multivariate Analysis

Mulualem Tigabu
Department of Silviculture
Umeå

Doctoral thesis
Swedish University of Agricultural Sciences
Umeå 2003

Acta Universitatis Agriculturae Sueciae
Silvestria 274

ISSN 1401-6230
ISBN 91-576-6508-7

Printed by: SLU, Grafiska Enheten, Umeå, Sweden, 2003

Abstract

Mulualem Tigabu 2003. Characterization of Forest Tree Seed Quality with Near Infrared Spectroscopy and Multivariate Analysis. Doctoral dissertation.
ISSN 1401-6230 ISBN 91-576-6508-7

The thesis presents a rapid and non-destructive method for characterizing the genetic, physiological and technical qualities of both temperate and tropical tree species on single seed basis. It is based on 'cross fertilization' of near infrared technology and multivariate analysis. The result demonstrated that seed sources, mothers and fathers of *Pinus sylvestris* could be identified using near infrared spectroscopy (NIRS) with 100%, 93% and 71% classification accuracies, respectively. NIRS was employed to detect internal insect infestation in *Cordia africana* and a 100% classification of sound and insect infested seeds were achieved on the basis of insect cuticular components and moisture difference between the two fractions. An extension of this study was performed on *Picea abies* seeds differing in origin and year of collection. Detection of infested and uninfested seeds with NIRS was found insensitive to subtle differences in proteins, lipids and moisture provided that between-seed lot spectral variability is removed *a priori* with appropriate spectral pretreatment technique or the calibration model takes into account this natural variability. Sound and insect-damaged seeds of *Albizia schimperiana* were also successfully separated based on differences in relative water content. The moisture gradient between sound and insect-damaged seeds was intentionally created by soaking both fractions in water at room temperature for a specified time owing to the fact that the hard and impermeable seed coat of intact seeds does not allow the diffusion of water.

The application of NIRS for the discrimination of viable and empty seeds of *Pinus patula* was evaluated and the two fractions were discriminated with 100% accuracy on the basis of divergence in lipid and protein contents. A further study was made to simultaneously discriminate filled, empty and insect-infested seeds of three *Larix* species. The result demonstrated a 100% recognition of infested and empty seeds while the recognition rates of filled seeds ranged between 90% and 100%; the highest being for *Larix sukaczewii* followed by *Larix decidua* and *Larix gmelinii*, respectively. In seed vigour analysis, it was also possible to distinguish between vigorous and aged seeds with 100% classification accuracy.

The results reported in this thesis demonstrate the capability of NIRS combined with multivariate analysis as a tool for rapid and non-destructive analysis of several seed quality attributes. As establishment of new forest plantations shows an increasing trend globally, NIRS will play a pivotal role in upgrading seed lot quality through sorting of unproductive seeds, and hence facilitating single seed sowing for containerised seedling production in nurseries and/or direct sowing out in the field. Therefore, continued emphasis should be given towards developing simple, cost-effective and automated sorting system in the future.

Key words: empty seed, filled seed, genetic seed quality, insect infestation, NIR, seed origin, seed quality, seed vigour

Author's address: Mulualem Tigabu, Department of Silviculture, SLU, S-901 83 UMEÅ, Sweden. E-mail: Mulualem.Tigabu@ssko.slu.se.

Dedicated to

My parents,

My brother, Amsalu Tigabu

&

sisters, Fasik, Aynadis and Hanna Tigabu

Contents

Appendix

List of original papers

The present thesis is based on the following papers, which will be referred to in the text by their respective Roman numerals.

I. Tigabu, M., Odén, P.C. & Lindgren, D. 2003. Identification of seed source and parents of *Pinus sylvestris* L. using visible–near infrared reflectance spectra and multivariate analysis. TREES (Submitted).

II. Tigabu, M. & Odén, P.C. 2002. Multivariate classification of sound and insect-infested seeds of a tropical multipurpose tree, *Cordia africana*, with near infrared reflectance spectroscopy. J. NIR Spectrosc. 10: 45-51.

III. Tigabu, M., Odén, P.C., & Shen, T.Y. 2003. Application of near infrared spectroscopy for the detection of *Plemeliella abietina* – larva – infested and uninfested seeds of *Picea abies*. Can. J. For. Res. (submitted).

IV. Tigabu, M., & Odén, P.C. 2003. Near infrared spectroscopy-based method for separation of sound and insect-damaged seeds of *Albizia schimperiana*, a multipurpose legume. Seed Sci. & Technol. 31, 317-328.

V. Tigabu, M., & Odén, P.C. 2003. Discrimination of viable and empty seeds of *Pinus patula* Schiede and Deppe with near-infrared spectroscopy. New Forests 25: 163-176.

VI. Tigabu, M., & Odén, P.C. 2003. Simultaneous detection of filled, empty and insect-infested seeds of three *Larix* species with single seed near infrared transmittance spectroscopy. New Forests (in press).

VII. Tigabu, M., & Odén, P.C. 2003. Rapid and non-destructive analysis of vigour of *Pinus patula* seeds using single seed near infrared transmittance spectra and multivariate analysis. Seed Sci. & Technol. (submitted).

Introduction

Seed quality

The recent global forest account indicates that new forest plantation areas are being successfully established at the rate of 3.1 million hectares per year (Food and Agricultural Organization 2001). Industrial plantations, from which wood or fibre are supplied for wood processing industries, account for 48% and non-industrial plantations, such as fuelwood plantations, small scale wood lots, trees for conservation purposes, constitute 26% of the global plantation estate (Food and Agricultural Organization 2001). In terms of species composition, *Pinus* and *Eucalyptus* are the dominant genera, representing 20% and 10% of the global plantation estate, respectively. As many multipurpose trees are on endangered species list that necessitated conservation of germplasm (Hilton-Taylor 2000), there is an increasing trend towards planting indigenous species. The success of any sustainable reforestation program, among other things, hinges on a continuous supply of high quality seeds for the production of the desired quantity of seedlings in nurseries or for successful stand establishment by direct sowing out in the field.

What is seed quality then? Seed quality is defined as "a measure of characters or attributes that will determine the performance of seeds when sown or stored" (Hampton 2002). It is a multiple concept encompassing the physical, physiological, genetic, pathological and entomological attributes that affect seed lot performance (Basu 1995). Seed quality is often indexed using viability and vigour. Viability, synonymous with germination capacity, refers to the ability of a seed to germinate and produce a normal seedling. In other words, it denotes the degree to which a seed is alive, metabolically active and possesses enzymes capable of catalysing metabolic reactions needed for germination and seedling growth. Seed vigour is "the sum total of those properties of the seed which determine the level of activity and performance of the seed or seed lot during germination and seedling emergence" (Hampton and TeKrony 1995). As seed vigour is not a single measurable property, aspects of performance associated with seed vigour include rate and uniformity of seed germination and seedling growth, emergence ability of seeds under unfavourable environmental conditions, and performance after storage and transport, particularly the retention of the ability to germinate.

Several factors affect the production of high quality seeds, such as insect infestation (*e.g.* Barbosa and Wagner 1989, Wagner *et al.* 1991, El Atta 1993, Dajoz 2000, Bates *et al.* 2000, 2001), pollination failure and post-zygotic degeneration (*e.g.* Owens *et al.* 1990, El-Kassaby *et al.* 1993), infection by seed borne pathogens (Pritam and Singh 1997), environmental conditions during seed development (Gutterman 2000) as well as the genetic constitution (Bazzaz *et al.* 2000). Insect infestation reduces seed quality by damaging the embryonic axis, or consuming cotyledon or endosperm thereby exhausting the reserve food, and a seed severely attacked by feeding larvae will be empty of its contents (El Atta 1993, Bates *et al.* 2000, 2001). Early infestation often causes abortion of the

attacked ovule or the whole fruit; and attacks occurring later during fruit development result in empty seeds (Janzen 1972, Hedlin *et al.* 1981). In many conifers, such as pines and larches, the occurrence of a large quantity of empty but normal appearing seeds due to pollination and fertilization failures is well documented (Hakansson 1960, Hall and Brown 1976, 1977, Owens and Molder 1979, Kosinski 1986, 1987, Owens *et al.* 1994, Owens 1995). Obviously, such unproductive seeds should be detected and eliminated to enable single seed sowing in containerised seedling production or to ensure the success of emergence and establishment of seedlings by direct sowing.

Seed ageing or seed deterioration is also a well-known cause of reduced vigour and viability, which commences during physiological maturity and continues during harvest, processing and storage. Studies have shown that seed deterioration is accompanied by a cascade of physiological and biochemical perturbations (see reviews by Smith and Berjak 1995, Marcos-Filho and McDonald 1998, Walters 1998, MacDonald 1999) that eventually result in reduced overall germination performance, speed and uniformity of germination, inferior seedling emergence and growth, reduced storability, as well as susceptibility to environmental and biological stresses (*e.g.* Delouche and Baskin 1973, Kalpana and Madhava Rao 1995). Usually the loss of vigour precedes the loss of viability, and seed lots with similar total germination capacity can differ greatly in their rates of germination, emergency, and growth. The decline in seed vigour can be reversed using pretreaments such as priming (Winsa and Bergsten 1994, Sivritepe and Dourado 1995, Oluoch and Welbaum 1996, Usberti and Valio 1997, Shen 2000), hormonal treatments (Wang *et al.* 1996) and cold moist stratification (Jones and Gosling 1990, Poulsen 1996). Assessment of seed vigour is, thus, one of the seed testing routines to provide information regarding potential field performance and storability of a given seed lot as well as to decide whether a seed lot should be primed or not.

The genetic seed quality encompasses adaptability to the planting site, growth performance, tolerance to biological and environmental stresses, and the level of gene diversity within a seed lot. It is particularly important in seed lots of forest trees because any anomalies cannot be detected early owing to the long life span of tree growth. Establishment of seed orchards using superior or plus-trees is the most common and cost-effective way of ensuring sustainable supply of genetically improved seeds (Zobel and Talbert 1984, Varghese *et al.* 2000). In Sweden, for example, 574 ha of Scots pine and 234 ha of Norway spruce seed orchards have been established that have supplied 42.9 and 9.7 tonnes, respectively of genetically improved seeds over the period 1968 – 1994 (Hannerz *et al.* 2000). However, pollen contamination has been shown as the major hurdle for maintaining the genetic purity of orchard seeds (*e.g.*, Yazdani and Lindgren 1991, Pakkanen and Pulkkinen 1991, Wang *et al.* 1991, Harju and Nikkanen 1996, Kang *et al.* 2001). Knowledge of seed source is also crucial because effects of maternal environment (also called aftereffects) during seed development have been reflected on the performance of the progeny (Lindgren and Wang 1986, Dormling and Johnsen 1992, Lindgren and Wei 1994, Wei *et al.* 2001).

A variety of seed testing methods have been continuously developed for rapid assessment of seed quality. X-radiography is a standardized method for assessing the proportion of filled, empty, insect-infested and physically damaged seeds in a given seed lot while excised embryo and tetrazolium tests are employed to promptly determine the viability of seed samples, especially for seeds that germinate slowly or exhibit dormancy (International Seed Testing Association 2003). The most widely used methods for assessing seed vigour are measurement of germination rate and seedling growth rate, stress tests (*e.g.* cold test and accelerated ageing), and biochemical tests such as tetrazolium staining and leachate conductivity (Hampton and TeKrony 1995, Bonner 1998, Demelash *et al.* 2003a). Other approaches include measurements of respiratory activity (Bonner 1986), ATP content (Lunn and Madsen 1981, Siegenthaler and Douet-Orhant 1994), glutamic acid decarboxylase activity (Grabe 1964) and fumarase activity (Shen and Odén 2000, 2002). Molecular markers, such as allozymes, chloroplast and mitochondrial DNA, are adopted to estimate the extent of pollen contamination in seed orchards and putative seed origin (Wang *et al.* 1991, Stoehr *et al.* 1998, Wang and Szmidt 2001, Ribeiro *et al.* 2002). However, these methods have some limitations. For example, X-radiography is potentially hazardous for operators and the seed, and it requires highly experienced personnel to interpret X-ray images. The cutting and tetrazolium tests are destructive in nature and laborious. The various seed vigour tests are destructive, subjective (*e.g.* biochemical tests) or relatively slow for tree seeds (*e.g.* germination and seedling growth rate tests). On top of this, none of them renders the possibility of sorting low vigour seeds from a seed lot. The molecular techniques for determining the genetic quality of seed crops are also technically complex and expensive.

Likewise, a variety of seed sorting techniques have been developed to upgrade seed lot quality; notably, the Pressure-Vacuum (PREVAC) method for removing mechanically and insect-damaged seeds (Lestander and Bergsten 1985, Bergsten and Wiklund 1987) and the incubation, drying and separation (IDS) technique for sorting empty and dead-filled seeds of Scots pine (Simak 1981, 1984), which later applied on seed lots of several other conifers and broadleaved species (Donald 1985, Bergsten and Sundberg 1990, Sweeney *et al.* 1991, Vanangamudi *et al.* 1991, Downie and Bergsten 1991, Downie and Wang 1992, Singh and Vozzo 1994, Poulsen 1995, Falleri and Pacella 1997, Demelash *et al.* 2002, 2003b). Results from these studies, however, showed that the efficacy of these methods varies among species and seed lots and complete separation is still difficult to achieve for some species. This could be due to large inherent seed size variability (*e.g. Cupressus lusitanica*, Bergsten and Sundberg 1990), inadequacy of density gradient between sound and insect-damaged seeds (*e.g. Albizia schimperiana*, pers. observation), or insufficiency of the specific density of the flotation media. Furthermore, it has been shown that some flotation media have a detrimental effect on seed germination and storability (Barnett 1971, Simak 1973, Hodgson 1977). In Norway spruce seeds, Tillman-Sutela and Kauppi (1995a) have shown that the wax and crystal layers around the micropyle (the natural opening in the seed) restrict the imbibition process; thereby hindering the separation of viable and non-viable seeds with the IDS method. It was these limitations that have motivated the present thesis work.

Near Infrared Spectroscopy

Historical development

The near infrared (NIR) part of the electromagnetic spectrum is commonly defined as the region spanning wavelengths from 780 nm to 2500 nm. The NIR region was first discovered back in 1800 by Sir William Herschel while attempting to measure the heat energy of solar emission beyond the red portion of the visible spectrum. In honour of his historic discovery, the wavelength region between 780 and 1100 nm is termed as the 'Herschel infrared' (Davies 1990). After a long pause, Abney and Festing made the first serious NIR measurements and interpretations in 1881, followed by Coblenz in 1905. Further systematic studies on NIR spectra of organic compounds and assignment of bands to functional groups were undertaken between 1922 and 1929, in the period 1930 to 1945 as well as in the 1980s (Osborne *et al.* 1993).

The advent of the Second World War was a turning point in the historical development of NIR technology. During this time, photoelectric detector (lead sulphide) was discovered that eventually became a major detector for the NIR region. Following this advancement in instrumentation a great deal of work was carried out in the period 1955 to 1965. The foundations for modern NIR analysis was laid in the 1960s when Karl Norris and co-workers started using wavelengths in the NIR region for rapid quality assessment of agricultural commodities, such as moisture in grain and seed (Norris and Hart 1965, Ben-Gera and Norris 1968), ripeness of fruits (Bittner and Norris 1968) and defects in eggs (Norris and Rowan 1962). Norris has also designed and developed the first grain moisture meter (Norris 1962, 1964) and recognized the power of multivariate analysis for extracting quantitative information from complex NIR spectra (McClure 1994). As a result, Karl Norris is recognized as 'father' of modern NIR technology. Dickey-john developed the first commercial NIR instrument, the Grain Analysis Computer (GAC), in 1971. Since then, several companies and individuals involved in the development of NIR instruments; notably, the Swedish Foss Tecator Company developed NIR transmittance spectroscopy fully dedicated to the analysis of intact individual grains/seeds (this instrument was used to record spectra from individual seeds in this thesis). The NIR technology has continued to show greater advancements in terms of instrumentation, precision as well as data acquisition and processing. Today, NIR spectroscopy is one of the fastest growing analytical technologies in the world with an overwhelming application in virtually all fields of science (Williams and Norris 1987, Osborne *et al.* 1993, McClure 1994, Workman 1999, Burns and Ciurczak 2001, Blanco and Villarroya 2002). The history of NIR technology is far richer and fascinating than described here; further details can be found elsewhere (Osborne *et al.* 1993, McClure 1994, Hindle 2001).

Principle and Theory

NIR spectroscopy works on the principle of interaction of electromagnetic radiation with matter, which takes several forms (Figure 1). When a solid sample,

like a seed, is illuminated with monochromatic radiation emitted by NIR instrument, the incident radiation will be reflected by the outer surface (known as specular reflectance), traverses deep into the inner tissue of the sample and reflected back (diffuse reflectance), passes all the way through the sample (transmittance), will be absorbed completely (absorption) and part of it will be lost as internal refraction and scattering. If a sample absorbs none of the incident energy, total reflection occurs. In NIR spectroscopy, we are interested in the diffuse reflectance and transmittance, although the former includes the specular component. If the specular component dominates the reflectance spectra, the actual absorption information from the sample will be obscured. Thus, the specular reflectance together with the wide-angle deflection and scattering within the sample are considered as sources of systematic noise in the spectra and need to be carefully handled during pre-processing of the spectral data. Often, organic materials selectively absorb NIR radiation that yields information about the molecular bonds within the material being measured.

When a molecule absorbs radiation in the infrared (IR) region, vibrations in the bonds occur either due to stretching or bending. Stretching is vibration in which there is a continuous change in the interatomic distance along the axis of the bond between the two atoms while vibration involving a change in bond angle is referred

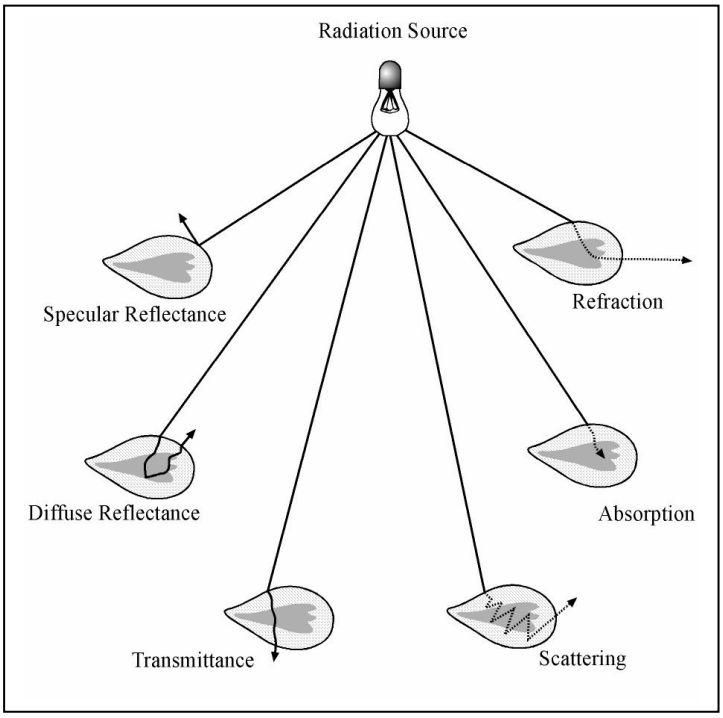

Figure 1. An illustration of the interaction of NIR radiation with seed samples.

to as bending or deformation (Figure 2).

The molecular bonds vibrate in a manner similar to a diatomic oscillator that can be explained using the quantum-mechanical model. According to the quantum selection rules, the only allowed vibrational transitions are those in which υ (the quantum number) changes by one ($\Delta\upsilon = \pm 1$). The harmonic oscillator model, thus, explains the absorption bands observed in the IR region due to fundamental modes of molecular vibration; but failed to explain the presence of overtone bands in the NIR. However, real molecules do not behave exactly as predicted by the law of simple harmonic motion and real bonds do not strictly obey Hook's law due to Coulombic repulsion between the two nuclei and dissociation of bonds beyond the limit of elasticity that levels off the potential energy (Figure 3). Consequently, the harmonic criterion is not fulfilled at higher vibrational states, and vibrations become rather anharmonic. Such anharmonic molecular vibrations allow energy transitions between more than one level, and thus creating overtone bands.

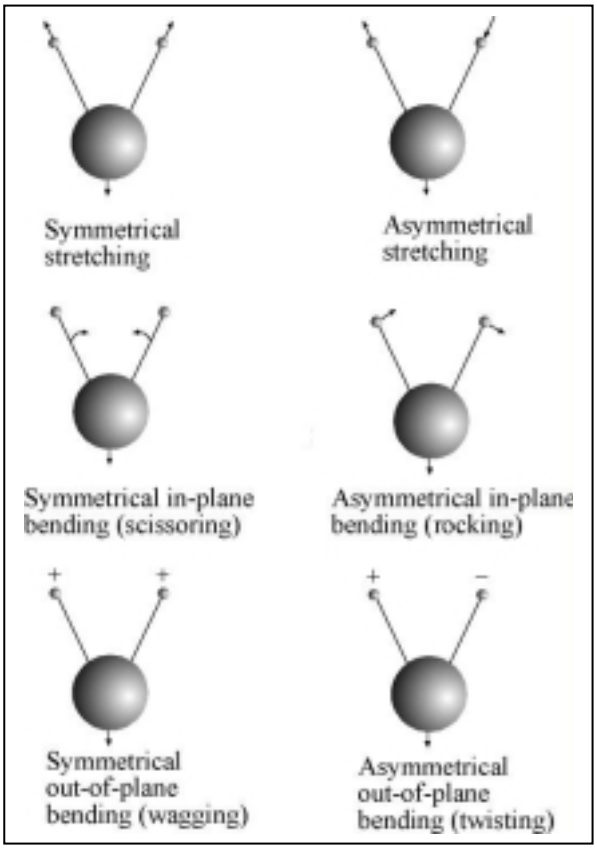

Figure 2. Modes of bond vibration for a hypothetical molecule AX_2.

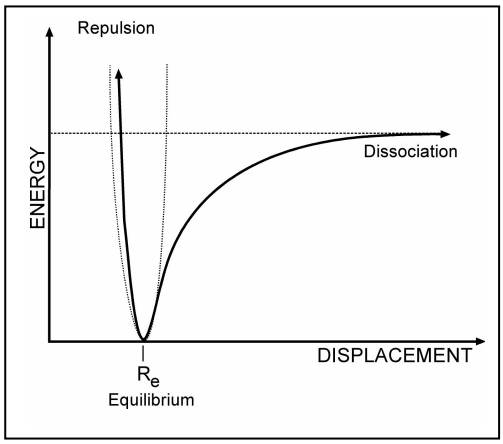

Figure 3. The energy of a diatomic molecule undergoing harmonic oscillation (dashed line) and anharmonic vibration (solid line) that explains absorption in the NIR region.

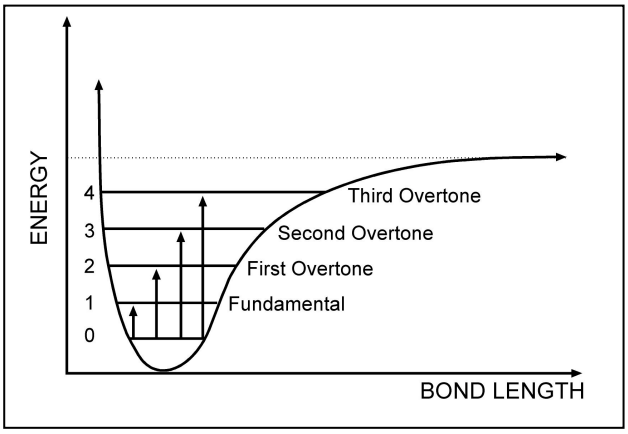

Figure 4. Energy transition levels creating overtone bands in the NIR region. The fundamental absorptions are the basis of IR spectroscopy.

The NIR spectrum contains overtones and combination bands, and the main bands typically observed in the NIR region correspond to bonds containing light atoms such as X – H, where X is carbon, nitrogen, oxygen or sulphure, and H is hydrogen that, in turn, are the major molecular moieties in virtually all organic materials. This is because the hydrogen atom is the lightest, and therefore exhibits the largest vibrations and the greatest deviations from harmonic behaviour. Other important functionalities in the NIR region include C = O, C – C, and C – Cl stretching vibrations, albeit the bands are much weaker (Shenk *et al.* 2001). The

overtone transitions are absorptions from the ground level to vibrational energy level 2 or higher (Figure 4) while combination bands arise from addition or subtraction of fundamental C – H, O – H and N – H vibrations. The overtones and combination bands are much weaker (often by a factor 10 or higher) than the fundamental absorption bands. This allows analysis of samples that are several millimetres thick (Bokobza 1998).

In addition to chemical information, NIR spectra contain physical information that can be used to determine physical properties, like bulk density in seeds (Velasco *et al.* 1998a, Font *et al.* 1999) and seed weight (Velasco *et al.* 1999). This is attributed to interactions between atoms in different molecules (such as hydrogen bonding and the dipole moment) that perturb vibrational energy states, thereby shifting the existing absorption bands and creating new ones through variation in crystal structure. This, in turn, allows crystal forms to be identified and physical properties determined (Blanco and Villarroya 2002).

Interpretation of NIR spectra is not as simple as that of IR spectrum owing to a large number of overlapping overtone and combination bands with broader peaks. In general, bonds with high dipole moments give the strongest overtone absorptions, and the Beer-Lambert law describes the quantitative aspect of absorption. The law states that the fraction of radiant energy absorbed by infinitesimal thickness of sample is proportional to the number of molecules in that thickness; *i.e.*, $A = \varepsilon Cl$, where A is absorbance, ε is the molar absorptivity, C the concentration and l is the path length. Since different materials absorb at different frequencies and exhibit different intensity of absorption, one is interested in determining the amount of various substances in a mixture based on measuring the relative amount of radiant energy absorbed at each frequency. Consequently, spectra measured as transmittance (T) is converted to absorbance (A) as follows:

$$A = \log (1/T) \text{ or } A = \log (T_0/T)$$

T_0 is 100% transmission. For practical reasons, the diffuse reflectance (R) is converted to absorbance according to the formula:

$$A = \log (1/R)$$

The intuitive argument for this relationship is that the diffuse reflectance is one in which the incident radiation is transmitted into the inner tissue of the samples and hence analogous to transmittance; except that the detector is repositioned to capture the diffuse reflectance (Birth and Hecht 1987). However, there are other more theoretical approaches to relate absorbance with concentration in diffuse reflectance spectrometry (see Osborne *et al.* 1993, Olinger *et al.* 2001). For an extensive coverage of the theory and principles of NIR technology see Murray and Williams (1987), Osborne *et al.* (1993), Ciurczak (2001) and Olinger *et al.* (2001)

Instrumentation

A host of NIR instrumentations is commercially available; ranging from laboratory and on-line systems to portable field instruments. A list of NIR spectrometer manufacturers and the type of commercially available instrumentation together with their typical characteristics as well as basic instrument specifications can be found in Workman and Burns (2001). The basic instrumental configuration in all NIR spectrometers includes: Radiation source, wavelength selector/ modulator, sample presentation, detector and output relay (Figure 5). Tungsten-halogen lamps with quartz envelopes are the major energy sources for NIR instruments. These lamps provide high-energy output (10 – 200 W) over the 360-3000 nm region and last longer due to a bathing effect of the halide inside the lamp. Light emitting diodes (LED), laser diodes and lasers are non-thermal or 'cold sources', in which most of the energy consumed appears as emitted radiation over a narrow range of wavelengths. As the emitting wavelengths are predetermined, instruments based on such devices are usually dedicated for specific analysis, such as determination of moisture in samples.

Radiation emitted from a source can be spectrally separated into individual wavelengths using different optical principles; namely, dispersive, interferometric and non-thermal (Osborne *et al.* 1993). A dispersive system is one where wavelengths of light are separated spatially and prisms were the classic dispersing elements in spectrometers for many years. However, prism is an inefficient arrangement with low and non-linear dispersion, and a large prism is often needed to achieve better performance. As a result, most scanning spectrometers used in laboratories and in industries today employ diffraction gratings and detector arrays for wavelength selection, which enable the detection of full spectrum simultaneously.

Another dispersive device incorporated into NIR spectrometers in recent years is *Acousto-optically tuneable filters* (AOTF). AOTF choose wavelengths by using radio-frequency signals to change the refractive index of a crystal made of TeO_2 (tellurium dioxide) in such a way that it transmits light of a given wavelength region or scans the whole spectral range. Since the AOTF is a monochromator with no moving parts (McClure 1994), it produces more reliable and reproducible wavelength scans than other devices, and is best suited for rugged on-line process environments. The second major optical principle used for wavelength selection in NIR spectroscopy is interferometry. This method, referred to as non-dispersive, does not cause angular dispersion, but instead uses filter, often known as interferometer, for wavelength differentiation. Among family of interferometric systems is the Michelson interferometer; the Fabry-Perot interferometer and Fourier transform NIR instruments. For more detail about interferometric systems refer to Osborne *et al.* (1993) and McClure (1994). The last category, the non-thermal system, involves the use of light emitting diodes, laser diodes or laser that can emit light in a narrow range of wavelengths. Laser diodes and lasers emit over an extremely narrow range and no pre-filtering of the radiation is required. Light

emitting diodes, however, emit over a relatively broader wavelength range, and interference filter is needed to narrow the radiation to the required bandwidth.

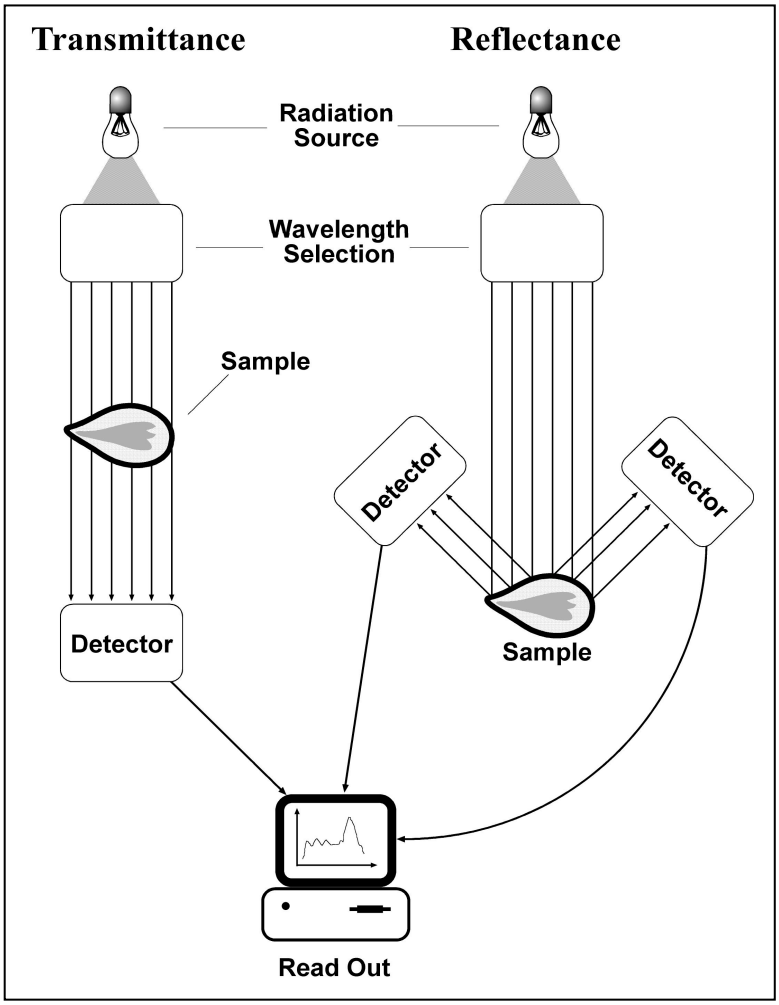

Figure 5. Basic components of NIR instrumentation operating in transmittance and reflectance modes.

Samples can be presented in a variety of forms for scanning by NIR spectrometers. Solid samples like seeds can be directly scanned using fibre optic probes. It can also be measured using standard sample holders that can be supplied by the manufacturer together with instrument, as in the case of Infratec 1225 Grain Analyzer. Ground samples can be scanned using standard sample cup made of quartz with glass windows. With minor modification to narrow down the window size, such a cup was used for measuring spectra from single seeds in this thesis.

Radiation transmitted through or reflected from a sample is detected using devices comprising of semiconductors. Lead sulphide (PbS) is the most widely

used detector in the NIR over the range of 1100-2500 nm while silicon sensors are used for the 360-1000 nm range (McClure 1994). In multi-channel system covering visible-NIR region (400-2500 nm), PbS detectors sandwiched with silicon photodiodes are often used to acquire spectral information over many wavelengths simultaneously. Another less common detector is a device composed of Indium gallium arsenide (InGaAs) that operates over the range of 1000-1800 nm with slightly better sensitivity than PbS (Osborne *et al.* 1993). Finally, computers are becoming an indispensable part of NIR instrumentation for capturing spectral data as well as for process monitoring and analysis of spectral data.

Multivariate analysis

NIR spectroscopic data are often recorded at several hundred-wavelength channels, *i.e.* multidimensional. They are also highly collinear, meaning that some of the variables can be written approximately as linear functions of other variables. On top of this, it is not always possible to use absorbance at a single wavelength to predict the concentration of one of the absorbers due to the overlapping nature of spectral peaks (the so called selectivity problem). Spectral interferences from other unidentified constituents in the sample and/or instrumental drifts, measurement errors *etc* also require special attention in order to get a good result. A number of multivariate projection methods (also called data compression methods) have been developed to extract the valuable information from the spectra (see Martens and Næs 1989). In essence, the projection methods will try to find a low-dimensional hyper-plane that represents the multidimensional data as well as possible and make interpretation of results easier. In this thesis, two related projection methods are employed: Principal Component Analysis (PCA) and Partial Least Squares Projection to Latent Structures (PLS).

Principal Component Analysis

Principal component analysis is a bilinear projection method that decomposes the original data matrix, **X** into " structure" and "noise" with few dimensional hyper-plane based on maximum variance directions (Esbensen 2000). Here the data matrix, **X**, denotes N samples or objects (*e.g.* individual seeds) upon which K variables (absorbance values at K wavelength channels) have been measured. The general PCA model can be expressed as:

$$\mathbf{X} = \mathbf{TP'} + \mathbf{E} = \sum \mathbf{t_a p'_a} + \mathbf{E}$$

T and **P'** denote a matrix of the scores and loadings, respectively after A dimensions while **E** represents the part of **X** left unexplained by the model. Scores, **T**, are coordinates of the objects projected down onto the hyper-plane and loadings, **P**, are directions of each dimension in the hyper-plane (*i.e.*, the cosine of the angle between the principal component and each of the original coordinate axes) and the residual is the distance between each point in K-space and its point

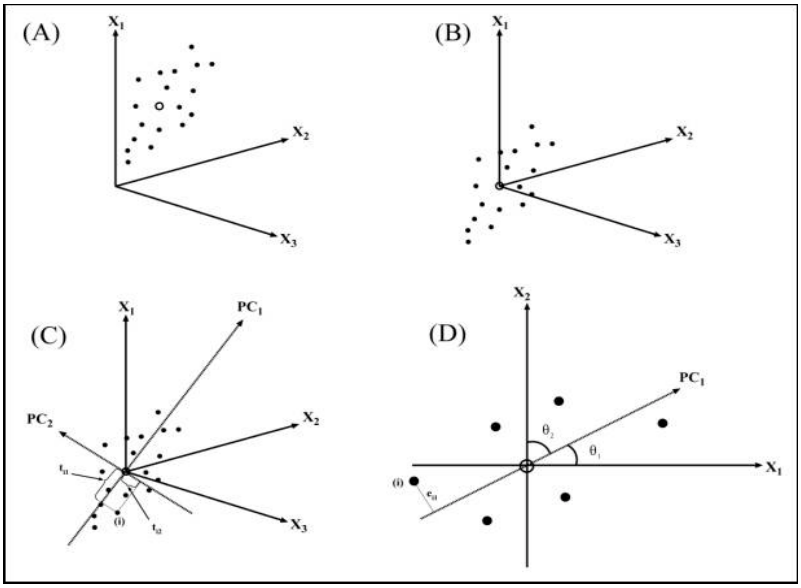

Figure 6. Geometric representation of PCA. A) Data plotted as a swarm of points in the variable space. Note that the open ring is the mean value. B) Mean centring of the data swarm that brings the original variable (also the PCs latter) into a common origin. C) The first PC, the maximum variance direction, which approximate the original data points as well as possible. The second PC lies along the second maximum variance direction and orthogonal to the first PC. The distance of each object, i, projected onto each PC to the centre is the score. D) The cosine of the angle between the PC and the original variable is the loading and the projected distance of each object to the PC is the residual.

on the plane (Figure 6). The scores, **T**, and the loadings, **P**, are derived by the NIPALS (Non-linear Iterative Partial Least Squares) algorithm that is described elsewhere (Martens and Næs 1989, Esbensen 2000). The computed principal components are always orthogonal to each other and they represent successively smaller and smaller variances. The maximum number of principal components that can be derived from an **X**-matrix equals to either N-1 or K, depending on which is the smaller. As higher order PCs usually describes smaller variation, one is interested in fewer significant components that can be determined by "eigenvalue" criterion or cross-validation. A component is considered significant if its normalized eigenvalue is larger than 2 or if the predictive power, Q^2, is larger than a significant limit.

Results from PCA are often presented as 2D (also 3D) plots of any pair of score and loading vectors. The most commonly used plot in multivariate data analysis is the score vector for PC1 versus the score vector for PC2. In fact, these are the two directions along which the swarm of data points displays the largest variation. The score plot (also referred to as map of samples) provides a useful guidance to identify outliers, to examine trends, clusters, and to explore similarity among objects. The loading plot (also called map of variables) gives us information about the relationship between the original **X**-variables and the principal components; *i.e.*, how much each variable contributes to the explanation of each PC. In addition, loading plots can be used to study how the original variables covary. Variables situated close together along a PC (having similar loadings) covary positively while those lying on opposite sides along a single PC are negatively correlated to each other.

PCA can also be used for more supervised classification purpose, known as Soft Independent Modelling of Class Analogy, SIMCA. SIMCA is a supervised multivariate classification approach based on a disjoint principal component analysis (PCA) for each class of similar observations (Wold 1976). A separate PCA model is computed for each class of samples. Based on the residuals of each sample from the PCA model, the standard deviation for each class (also called distance to the class model) is determined. This, in turn, is used to calculate the confidence interval or the critical distance to the model with an approximate F-test with degrees of freedom of the observation and the model at the 5% probability level. The unknown samples are then projected onto the existing PCA models and their residual standard deviations are compared to the confidence interval of each class. Finally, the unknown samples can be classified as: (1) belonging to a class, (2) belonging to several classes or (3) not belonging to any of the classes. A powerful graphical presentation of results from SIMCA analysis is to use the so called Coomans plot where class distances for two classes are plotted against each other in a scatter plot (see VII).

Partial Least Squares Projection to Latent Structures

PLS is the most widely used calibration technique in NIR spectroscopy owing to its capability to handle collinearity problems, its "built in" facility for outlier detection, the possibility to analyse multiple responses, the ease for visual interpretation of the data and its ability to cope with moderate missing data. Apart from quantitative analysis, PLS can be used for pattern recognition, the so-called Partial Least Squares-discriminant analysis, PLS-DA (Sjöström *et al.* 1986).

PLS analysis can be viewed as the regression extension of PCA. It establishes a relationship between the predictor block, **X**-matrix, and the response, **Y**, via an inner relation of their scores. The **X**-scores, **T**, describe the object variation in the predictor block (the spectral matrix in this case) and the corresponding variation in the response block by the **Y**-scores, **U**. What PLS does is to maximize the covariance between these inner variables (also called latent structures) **T** and **U**. A weight vector, **w***, is calculated for each PLS component that tells us the

contribution of each **X**-variable to the explanation of **Y** in that particular component. Thus, the matrix of weights, **W***, contains the structure in **X** that maximizes the covariance between **T** and **U** over all model dimensions. Finally, the corresponding matrix of weights for the **Y**-block, **C**, and the matrix of **X**-loadings, **P**, are calculated to perform the decomposition of **X** and **Y** as follows:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \quad \text{……. (1)}$$

$$\mathbf{Y} = \mathbf{UC}' + \mathbf{F} = \mathbf{TC}' + \mathbf{G} \quad \text{…….. (2)}$$

E, **F** and **G** are residual matrices for **X**, **Y** and the inner relation, respectively left unexplained by the model.

A matrix of regression coefficients, **B**, can then be computed according to the formula:

$$\mathbf{B} = \mathbf{W}^*\mathbf{C}' \quad \text{…… (3)}$$

From the above equations, the PLS model can be expressed as

$$\mathbf{Y} = \mathbf{XW}^*\mathbf{C}' = \mathbf{XB} + \mathbf{F} \quad \text{……. (4)}$$

Each new sample is predicted either using Eq. 4 or by computing the scores for the new samples and multiplying with the weight from the calibration model (Eq. 1 and 2).

The PLS parameters are derived by NIPALS algorithm for each component at a time. Given that the input variables, **X** and **y** are scaled and/or mean-centred, for a single **y** vector the following equations are used:

1) Estimate the loading weight, **w** as

$$\mathbf{w} = \mathbf{X'y} / \mathbf{(y'y)}$$

scale the **w** vector to length 1 using the factor, $(\mathbf{y'XX'y})^{-0.5}$

2) Estimate the score **t** as

$$\mathbf{t} = \mathbf{Xw}$$

3) Estimate the spectral loading **p** as

$$\mathbf{p} = \mathbf{X't} / \mathbf{(t't)}$$

4) Estimate the chemical loading **c** as

$$\mathbf{c} = \mathbf{y't} / \mathbf{(t't)}$$

5) Create new **X** and **y** residuals, **E** and **f**, as

$$\mathbf{E = X - tp'}$$

$$\mathbf{f = y - tc'}$$

For extracting the next component, use **X** = **E** and **y** = **f** and return to step 1. As a summary, the matrix relationship in PLS is shown in Figure 7.

PLS offers many parameters and diagnostics for model interpretation, and evaluation of model performance and relevance. The scores, **T** and **U**, contain information about the observations and their similarities or dissimilarities in relation to the problem at hand. PLS score plots of the **t**/**t**-type are used to uncover outliers in the descriptor matrix, **X**-space, while the **u**/**u**-type reveals deviation of observation in the responses matrix, **Y**-space. In addition, when PLS is used for classification/discrimination purposes, the **t**/**t**-type score plot for the descriptor matrix, **X**, is very useful to get an overview of the class discriminating ability of the computed PLS model. Finally, the **t**/**u**-type score plots are valuable tools to examine deviations from the dominating **X**/**Y** correlation structure as well as to identify departures from linearity between **X** and **Y**. A J-shaped curvature indicates that the response variables need transformation, such as logarithmic, and a curvature with inverse arching warrants transformation of **X**.

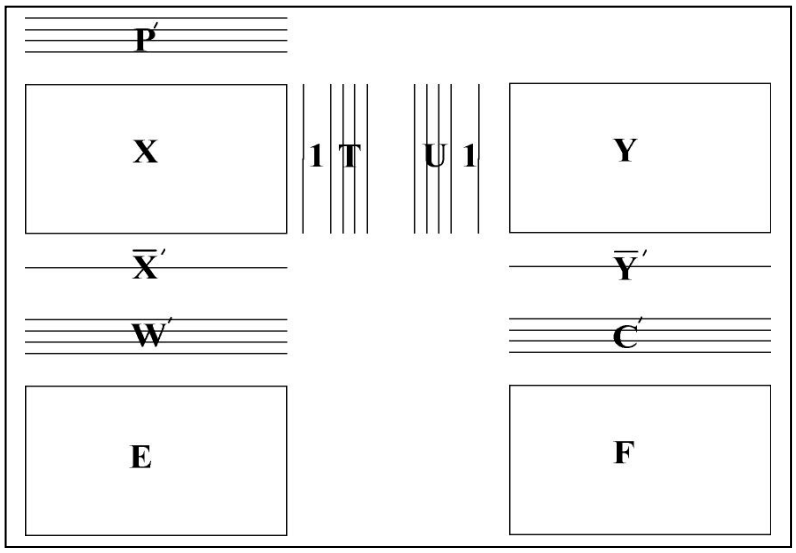

Figure 7. Summary of matrix relationship in PLS modelling. The vector 1 for X and Y denotes the variable averages, $1*\bar{X}'$ and $1*\bar{Y}'$, from the mean centering. The PLS scores are stored in T and U, the spectral loadings and weights (X) in P' and W', respectively, and the chemical loadings (Y) in C'. The variations in the data that were left unexplained by the PLS model form the E and F residual matrices.

Similarly, the variable related information is interpreted in several ways. A plot of **X**-weights shows how the original **X**-variables are linearly combined to form the score vectors, $\mathbf{t}_a$. Using **X**-weights, it is possible to understand which original variables are summarized by the new latent variable; *i.e.* **X**-variables that are highly correlated with **Y**-variables get higher weights. In NIR spectroscopy, line plot of **X**-weights is often used, as it allows analysis of which absorption peaks are modelled by each component. Interpreting a PLS model consisting of many components and covering a multitude of responses can be a challenging task. In such cases, a plot of PLS coefficients makes model interpretation less laborious and time consuming because they are summarized into one vector. Its drawback is that the information regarding the correlation structure among responses is lost when multiple responses are modelled simultaneously. To avert this problem, variable influence on projection (VIP) can be used. VIP is a weighted sum of squares of the PLS weights, **w***, taking into account the explained **Y**-variance of each model dimension. For a given model and problem, there will be one VIP-vector summarizing all components and **Y**-variables. Further information about the calculation of VIP parameters can be found in Eriksson *et al.* (1999). As a rule of thumb, predictors with a large VIP (> 1.0) influence the model substantially, and a cut-off around 0.7 to 0.8 is suggested to discriminate between relevant and irrelevant predictors.

The performance and relevance of PLS models are further evaluated by computing different statistics. The quantitative measure of *the goodness of fit* is given by the parameter R^2X and R^2Y, the explained variation for **X** and **Y**, respectively that can be computed as:

$$R^2X = 1 - SSX\,[A] / SSX\,[0]$$

$$R^2Y = 1 - SSY\,[A] / SSY\,[0]$$

SSX [A] is the sum of squares of the **X**-residuals, ($\sum \mathbf{e}^2_{ik}$), SSY [A] is the sum of squares of the **Y**-residual, ($\sum \mathbf{f}^2_{im}$), after extracting A components; SSX [0] and SSY [0] are total sums of squares for **X** and **Y**, respectively.

The prediction ability of the computed PLS model; *the goodness of prediction*, is also quantified by a parameter called the predicted variation, Q^2, using either cross validation or prediction sets. In all studies presented in the thesis, a seven-segment cross validation and prediction sets were employed to evaluate the prediction ability of the computed PLS models. The fraction of the total variation of the **Y**'s that can be predicted by a component, Q^2, is computed as:

$$Q^2 = 1 - PRESS/SS$$

PRESS is the prediction error sum of square ($\sum (Y - \hat{Y})^2$ and SS is the residual sum of squares of the previous dimension. This parameter is essential to determine the significance of each model dimension. According to Rule 1, if Q^2 for the whole data set due to cross validation is larger than a significant limit, the extracted dimension is considered significant. Q^2 can also be computed for each **Y**-variable,

and if it is larger than a significant limit, the tested dimension is significant according to Rule 2. The cumulative Q^2 for all extracted components can be computed as:

$$Q^2cum = (1.0 - \prod(PRESS/SS)a)$$

$\prod(PRESS/SS)a$ is the product of PRESS/SS for each individual component, a. Larger cumulative Q^2 value for a given response indicates that the model for that response is good. As a rule of thumb, a model with $Q^2 > 0.5$ is considered as good, $Q^2 > 0.75$ as very good and $Q^2 > 0.9$ as excellent. The ultimate objective of developing a calibration model is to make predictions in the future. In all the studies in the thesis, the computed calibration models were applied to predict new samples in the prediction sets that were kept aside during model building. The modelling error and the prediction ability are further evaluated by computing the root mean square error of calibration (RMSEC) and the root mean square error of prediction (RMSEP), respectively; and can be computed as follows:

$$RMSEC = \sqrt{\Sigma(\hat{y} - y)^2 / (N - A - 1)} \qquad RMSEP = \sqrt{\Sigma(\hat{y} - y)^2 / N}$$

$\hat{y}$ is the predicted value; y is the actual value; N is the number of samples in the validation sets (both for cross validation and test set) and A is the model dimension.

Spectral pretreaments

NIR spectra are not usually amenable for direct analysis due to unwanted systematic variation that has no correlation with the response variable. Light scattering, base line shift, instrumental drift, and path length differences are among the common sources of systematic noise in the spectra, which should be removed from the raw spectral signals. Spectral pretreaments, also called spectral filters, are mathematic functions for handling such interferences in order to avoid its dominance over the chemical signal. The commonest data pretreatment techniques in NIR spectroscopy are derivatives (Savitzky and Golay 1964), multiplicative signal correction (Geladi *et al.* 1985), standard normal variate transformation (Barnes *et al.* 1989) and orthogonal signal correction (Wold *et al.* 1998). In the thesis, they were applied, as deem necessary, to enhance the spectral features, and thereby developing robust models. Other approaches to handle systematic spectral variations are described in Næs *et al.* (2002).

Derivatives are intuitive ways of dealing with systematic variations in the spectra, and the first and second derivatives are often used to reduce additive baseline and scatter effects, respectively. The first derivative is the slope at each point of the original spectrum and calculated by taking differences between adjacent points and dividing by the wavelength gap, although the latter is not usually done as it only affects the scaling of the derivatives. Thus the first derivative at wavelength w could be computed as:

$$x_{1der} = x_w - x_{w-1}$$

x_w is absorbance at wavelength w in the sequence. The second derivative is the slope of the first derivative, and more similar to the original spectra; *i.e.*, having peaks in nearly the same locations but inverted in direction. The second derivative is computed as the difference of two adjacent first derivatives, yielding the second derivative formula:

$$x_{2der} = x_{w-1} - 2x_w + x_{w+1}$$

The major drawback of this simple approach is that derivatives reduce signals and amplify noise. To circumvent this problem, smoothing of the spectra prior to applying derivatives is essential. Savitzky and Golay (1964) described a more stringent approach based on fitting low-order polynomials.

Multiplicative signal correction (MSC) works primarily for cases where the scatter effect is the dominating source of variability, which is very typical in many applications of diffuse NIR reflectance spectroscopy. Assuming that each sample spectrum has an offset and a slope due to interference effects, one can correct for this if the variability is systematic; *i.e.*, constant over the spectral range. By plotting each spectrum, x_i, against the reference spectrum, the offset (a_i) and the slope (b_i) are calculated using least squares of the equation:

$$\mathbf{x_i = a_i + \bar{x}\, b_i}$$

Finally, the sample spectrum is corrected as follows:

$$\mathbf{x_{i,corr} = (x_i - a_i) / b_i}$$

The corrected spectra give a better prediction of the response not only due to removal of irrelevant information but also due to linearization of the relationship between the predictor and the response. An extension of the MSC approach is the piece-wise multiplicative scatter correction (PMSC), presented by Isaksson and Kowalski (1993). In essence, PMSC corrects non-linear additive and multiplicative scatter effects by fitting a linear regression in a local wavelength region. The assumption is that the scatter effects vary over the spectral range, and hence the scatter correction should be performed piece-wise using a moving window along the wavelength range.

The standard normal variate (SNV) transformation removes the multiplicative effect of scatter and particle size on an individual object basis (Barnes *et al.* 1989). It has an effect very much similar to MSC; the only difference is that SNV standardizes each spectrum using only data from that particular spectrum. The SNV transformation is performed according to the following general formula:

$$\mathbf{x^*_{ik} = (x_{ik} - m_i) / S_i}$$

$\mathbf{x^*_{ik}}$ = the transformed absorbance value for the ith object at the kth wavelength, $\mathbf{x_{ik}}$ = the original absorbance value at the kth wavelength for the ith sample, $\mathbf{m_i}$ = the mean of the K spectral measurements for sample i, $\mathbf{S_i}$ is the standard deviation of the same K measurements and K is the number of X-variables (wavelength channels). The actual pretreatment can be perceived as mean centring and scaling to unit variance in the object direction.

Orthogonal signal correction (OSC) is unique from the spectral pretreatments discussed above in one major aspect; it takes the response variable into account in its algorithm. OSC removes more general types of interferences in the spectra by removing components, latent variables, orthogonal to the response variable calibrated against. It is based on partial least squares regression, in which the weights in OSC are calculated to minimizing the covariance between the spectral data, **X**, and the response, **y**. Components orthogonal to **y** containing unwanted systematic variation are then subtracted from the original spectral data, **X**, to produce a filtered descriptor matrix.

The OSC algorithm starts with the calculation of the first principal component for the spectral data according to NIPALS. The first score vector, **t**, is then orthogonalized against **Y** as $\mathbf{(1-Y(Y'Y)^{-1}Y')t}$ to produce the orthogonal score vector, **t***. The PLS weights, **w,** are computed to make **Xw = t***, thereby minimizing the covariance between **X** and **Y**. The score vector, **t***, is then updated and give another score vector, **t****, which is then orthogonalized to **Y**, and the iteration proceeds until **t**** converges. The spectral data, **X**, can then be expressed as a product of the updated-orthogonalized score vector, **t****, and the corresponding loading vector, **p****, and a residual, **E**. The residual, **E**, constitutes the filtered data, **Xosc**, after removal of the first component orthogonal to **Y**.

$$\mathbf{E = X - t^{**}p^{**}}$$

$$\mathbf{Xosc = E}$$

With NIR diffuse reflectance spectra, two OSC- components are sometimes warranted (Wold *et al.* 1998). The second component can be removed by repeating the same procedure described above using the residual, **E** as **X**.

Prior to prediction, new samples in the prediction sets must be treated in the same way. To do this, the score vector, $\mathbf{t}_{test}$, is calculated using the weights derived from the calibration set and the new spectra, $\mathbf{X}_{test}$; *i.e.*, $\mathbf{t}_{test} = \mathbf{X}_{test}\mathbf{w}$. The residual, $\mathbf{E}_{test}$, constituting the filtered spectra can then be obtained by subtracting the spectral data, $\mathbf{X}_{test}$, from the product of the score vector, $\mathbf{t}_{test}$, and the loading vector from the calibration, **p****.

$$\mathbf{E_{test} = X_{test} - t_{test}p^{**}}$$

Analogously, if two components were removed from the calibration set, the same should be done in the test set. The residual from the first component is used as **X**, the weights from the second orthogonal component are used to calculate the

score vector for the second component, finally subtracting the second loading from the calibration multiplied by the computed score vector for the second component from the residual of the first component will yield the filtered descriptor matrix after two components orthogonal to **Y**. Basically, the OSC-treatment was developed to generate a robust prediction model for quantitative analyses through removal of interferences that have no relevance for the analyte at hand. However, in qualitative analysis where no true response variables exist, discrete values can be assigned to each class and used to perform OSC filtering (Wold *et al.* 1998). In this thesis, this is demonstrated in studies **I-III** and **V**.

Objectives

The principal objective of the research presented in this thesis is to evaluate the potential of NIR spectroscopy combined with multivariate analysis as a rapid and non-destructive method for characterizing forest tree seed quality. The study covered the genetic, technical and physiological aspects of seed quality of both temperate and tropical forest trees. The specific objectives were:

1) Identification of seed source and parents of *Pinus sylvestris* (**Paper-I**),

2) Detection of internal insect infestation in *Cordia africana* (**Paper-II**)

3) Examining whether detection of infested seeds of *Picea abies* is sensitive to seed origin and year of collection (**Paper-III**),

4) Separation of sound and insect-damaged seeds of *Albizia schimperiana* (**Paper-IV**),

5) Discrimination of viable and empty seeds of *Pinus patula* (**Paper-V**),

6) Simultaneously detection of filled, empty and insect-infested seeds of three *Larix* species (**Paper-VI**) and

7) Rapid analysis of seed vigour of *Pinus patula* (**Paper-VII**).

In all studies, the underlying hypothesis was that seeds in a certain quality class would have a unique spectral signature that can be utilized to build a discriminant multivariate model.

Materials and Methods

Tree species and sample preparation

Seeds of both temperate and tropical species were used to evaluate the potential of NIR spectroscopy as a rapid and non-destructive method to characterize various

quality attributes. The temperate species were *Pinus sylvestris* L., *Picea abies* (L.) Karst, *Larix decidua* Mill., *Larix gmelinii* Rupr., and *Larix sukaczewii* Dyl., which are highly esteemed for their timber value, adaptability to the harsh cold environment as well as a variety of environmental and recreational values (*e.g.* Holtmeier 1995, Martinsson 1995, Stener 1995, Schmidt and Shearer 1995, Polubojarinov *et al.* 2000). The tropical species include *Cordia africana* Lam., *Albizia schimperiana* Oliv., and *Pinus patula* Schiede and Deppe that are multipurpose and valuable timber species. The taxonomy, description, habitat conditions, geographic distributions and uses of these tropical species are reported elsewhere (Hunde and Thulin 1989, Teketay 1991, Valera and Kageyama 1991, Friis 1992, Bekele *et al.* 1993, Fichtl and Adi 1994).

In the study made to identify seed sources with visible and near infrared spectroscopy (**I**), seed samples were drawn from a single family (a cross of clones AC1005 and BD1178) growing in three localities in Sweden: Sävar (north, 63º54'N and 20º33'E), Röskär (central, 59º25'N and 18º12'E) and Degeberga (south, 55º47'N and 14º04'E) and harvested in 1982-83. For identifying parents (**I**), seeds from four mothers (clone no. AC1005, AC1014, BD1032 and BD1178) independently crossed with the same father (clone no. Y3020) and seeds from the same mother (clone no. AC1005) but separately crossed with four different fathers (clone no. AC1014, BD1032, BD1178 and Y3020) were used. To avoid the confounding effects of year of collection and environment, seeds from different fathers were drawn from those families grown in southern Sweden and harvested in 1982 while seeds from different mothers were sampled from those harvested in 1983 from a seed orchard in Sävar.

For the discrimination of filled/viable, empty and insect-infested seeds with NIRS (**II**, **III**, **V** and **VI**), seed samples from each species were initially sorted using X-radiography (43805 N X-ray system Faxitron Series Hewlett Packard) according to the international seed testing rule (International Seed Testing Association 2003). Seeds with visible embryonic axis and megagametophyte were recognized as filled/viable seeds while empty seeds were characterized by the absence of megagametophyte and embryo. Insect-infested seeds were those seeds with visible larvae enclosed within the seeds. To separate sound seeds from insect-damaged seeds with NIRS (**IV**), damaged seeds were initially sorted manually by inspecting the visible exit holes made by the emerging adults and then both fractions were soaked in 40 ml of de-ionised water for one, three, six, nine and twelve hours at room temperature. This enabled us to create moisture gradient between insect-damaged and sound seeds as we know *a priori* that sound seeds of *Albizia* and many other legumes do not absorb water because of the hard and impermeable seed coats (*e.g.* Teketay 1996, Teketay and Tigabu 1996, Tigabu and Odén 2001). They were surface dried on a blotting paper for 10 minutes before scanning by NIR spectrometer.

For the analysis of vigour using near infrared transmittance spectroscopy (**VII**), vigour classes were formed by exposing seeds to an accelerated ageing treatment at 41°C and ca. 100% relative humidity for three, seven or nine days while untreated seeds served as high vigour class. The seeds were spread in a single layer on the

surface of bronze wire mesh seed holder above 250 ml (1 cm deep) de-ionised water in plastic boxes (22.5x19x7 cm). The boxes were then tightly covered with lids and then placed in an ageing chamber. At each ageing time, 100 samples were drawn for NIR analysis after thoroughly rinsing with de-ionised water to remove fungal outgrowth and surface drying for 10 minutes. The accelerated ageing regimes adapted in this study reduced the overall germination capacity and mean germination time to 75% and 12.3 days after three days of ageing, to 55% and 13.6 days after seven days of ageing and to 13% and 19.1 days after nine days of ageing compared to 99% and 8.9 days, respectively for vigorous seeds.

Measurement of NIR spectra

NIR reflectance spectra, expressed in the form of log (1/R), were collected from single seeds with NIRSystems Model 6500 spectrometer (FOSS NIRSystems Inc., Silver Spring, Maryland, U.S.A.). NIR spectra were recorded on individual seeds using a fibre optic probe (**IV** and **V**) or a spinning sample cup (**I**, **II** and **III**). In the former case, individual seed was placed on a black metallic bar with an oval-shaped depression (ca. 2 x 1 mm), fixed on a stature and scanned by tightly screwing the fibre optic probe against each seed. In the latter case, individual seeds were placed in a modified spinning sample cup (diameter = 3.8 cm and depth = 0.9 cm) that allowed collection of radiation reflected from the entire surface of the seed. To narrow the sample cup window, a micro sample insert, black metallic ring with an oval slit in the middle (diameter = 0.7 cm and depth = 0.15 cm) was inserted into the spinning sample cup. Another micro sample insert without any slit was placed on top of each seed in order to avoid stray light reaching the cardboard cover that was used as a support. Since the background metallic bar had a negligible reflectance, such an arrangement enabled us to collect reflectance from individual seeds only. The instrument measures diffuse reflectance in the range 400 nm to 2500 nm at 2 nm resolutions. Thirty-two monochromatic scans were averaged from each seed and reference measurements were taken on a ceramic plate after every 10 scans.

NIR transmittance spectra, expressed in the form of log (1/T), were collected from single seeds with a 1225 Infratec analyser (FOSS Tecator, Sweden) from 850 to 1048 nm at 2 nm resolutions. Individual seeds were placed in a single seed cell at 20 fixed positions. Each seed sample was scanned 32 times and the average of 32 successive scans from each seed was recorded. Prior to scanning of every sample set (20 seeds at a time), reference measurement was taken using the standard built-in reference of the instrument.

Data analysis

To remove unwanted systematic noise in the spectra, the reflectance spectroscopy data sets (log 1/R) were treated using multiplicative signal correction (MSC), orthogonal signal correction (OSC) and/or first derivatives. Since no true y-values existed in our data set, discrete values were assigned for each class of observations. Depending on the nature of the data, one or two OSC components were extracted.

The transmittance spectroscopy data sets (log 1/T) were pre-treated with standard normal variate transformation (SNV) to remove the multiplicative effect of scatter and particle size on an individual object basis.

Prior to building the calibration models, 25-30% of the observations were excluded to make up the prediction sets. Initially, principal component analysis (PCA) was performed on calibration sets as a basis for outlier detection and to get an overview of the data. Subsequently, calibration models were derived with partial least squares regression using the digitised NIR spectra as descriptor matrix and a vector of artificial discrete values as regressand. Seed sources and parents were assigned with **y**-values from 1 to 3 and from 1 to 4 (BD1032, AC1014, BD1178 and AC1005 mothers; and fathers BD1032, Y3020, AC1014, and BD1178 were assigned with values 1 to 4 respectively). A value of 1 was assigned for filled, viable, sound, and vigorous seeds while –1 was assigned for empty, internally infested, insect-damaged, and aged seeds in each of the studies.

The number of significant PLS factors to build the model was determined by a seven-segment cross validation. A factor was considered significant if the ratio of the prediction error sum of squares (PRESS) to the residual sum of squares of the previous dimension (SS) was statistically smaller than 1.0, or if the predictive power (Q^2 = 1.0 – PRESS/SS) was larger than a significant limit. For a more comprehensive description of theories and applications of PLS regression in multivariate calibration and classification, see Martens and Næs (1989), Eriksson *et al.* (1999), Wold *et al.* (2001) and Næs *et al.* (2002).

Finally, the computed models were applied to predict new samples in the prediction sets. Prior to prediction, the new samples were automatically pre-treated with SNV, MSC and OSC by the software system (Simca-P, version 8, Copyright: Umetrics AB, Sweden) while the first derivatives of the spectra from the test samples were computed using Unscrambler 7.5 (Copyright: CAMO ASA, Norway). For all tests, the decision threshold was set either at 0.0 or ± 0.5 depending on the study. The classification accuracy (also referred to as classification rate and recognition rate in the thesis) for each model was computed as the ratio of number of samples in a given class predicted correctly to the total number in the prediction sets. All model calculations were made on mean-centred data sets.

In vigour analysis (**VII**), the SIMCA approach was also applied to classify vigorous and aged seeds. A separate PCA model was computed for vigorous and aged seeds. Based on the residuals of each samples from the PCA model, the standard deviation for each class was determined. This, in turn, was used to calculate the confidence interval or the critical distance to the model with an approximate F–test with degrees of freedom of the observation and the model at the 5% probability. The number of significant principal components to build the PC–models was determined by the 'eigenvalue' limit (EV) criterion as suggested by Eriksson *et al.* (1999) for large data tables and a component was considered significant if its normalized eigenvalue was larger than 2. PCs were also significant according to cross validation.

The unknown samples in the prediction set were then projected onto the existing PCA models and their residual standard deviations were compared to the critical distance of each class. Samples in the test set with a probability of class membership greater than 5% were classified as members of a given class, otherwise non-members. The classification results were summarized and presented in so called Cooman's plots where class distances for vigorous and aged seeds were plotted against each other in a scatter plot.

Results and discussion

Identification of seed source and parents

Visible (VIS) and near infrared (NIR) spectroscopy was employed in order to identify seed sources, mothers and fathers of *Pinus sylvestris* based on single seed spectra. Calibration models were computed using the entire range of VIS+NIR, the VIS and NIR regions as well as using raw, MSC- and OSC-treated data sets. The results showed that both the VIS and NIR spectra contained much information (R^2X ranging from 0.72 to 0.99), which in turn described the variation among seed sources considerably (R^2Y ranging from 0.75 to 0.99). The overall predictive power (Q^2 in the range from 0.72 to 0.99) according to cross validation was also high for all models. However, an OSC-treatment of the spectra reduced dimensional complexity (A = 1) of the computed models compared to the raw spectra and MSC-treated data set that utilized from three to nine components. For new samples in the prediction set, all calibration models (raw, MSC and OSC) in the VIS+NIR region successfully detected sources of Scots pine seeds with 100% accuracy, except the calibration model developed on raw data set where one sample was found at the limit for the northern and central seed sources. A similar result was found in the VIS region; but in the NIR region the MSC model resulted in higher average classification accuracy (99%) compared with the raw (84%) and OSC (89%) models.

For the identification of parents, calibration models derived from the VIS+NIR spectral region described more than 75.6% of the spectral variation and more than 93% of the between-mothers variation with an excellent prediction ability for the calibration set according to cross validation. The statistical summary for models developed in the VIS and NIR regions separately also showed an excellent overall fit of the models to the data. Between-fathers variability was better explained using the OSC-model derived from the VIS+NIR ($R^2Y = 0.94$) and visible spectra ($R^2Y= 0.92$) while the NIR spectra alone poorly described the between-fathers variability ($R^2Y = 0.18$ for both raw and OSC models), although the MSC-model was relatively good ($R^2Y =0.53$ and $Q^2 =0.5$). For identification of mothers, the highest average classification accuracy for the test samples was 93% using OSC-treated data sets in the VIS+NIR spectra. The OSC-model also resulted in better classification accuracy in the VIS region. The average classification accuracy was nearly similar among the three models in the NIR region. For identification of fathers, the highest average classification accuracy was achieved with MSC models

derived from full (71%) or visible range (70%) and the OSC model developed on visible spectra (70%).

Apparently, the visible region was effective in identifying seed sources and parents of Scots pine. This can be attributed to varying colour types often encountered in Scots pine seeds: Light, mixed and dark with varying proportion within- and between-provenances (Tillman-Sutela and Kauppi 1995b). These authors also claimed that seeds from the northernmost provenance were predominantly dark, observed little absorption in the visible range and minor differences in the quantitative colour characteristics of the seed coat extract between the light and dark seeds of various provenances. In our case, seeds from central Sweden were black and glossy while those from southern and northern Sweden were mottled and light brown (or grey), respectively. This might be due to the fact that the genotypes were not mixed up in our study unlike the seed samples in the Tillman-Sutela and Kauppi (1995b) study that were collected from forest stands and orchard grafts with *ca.* 45% pollen contamination. In addition, it should be noted that NIR spectroscopy is highly sensitive and sufficiently detects subtle differences while multivariate analysis is powerful to extract such information from the spectra unlike the univariate analysis (Martens and Næs 1989). Our finding accords with previous studies that have demonstrated the efficacy of reflectance spectra in the visible region for classify wheat kernels according to their colour (Delwiche and Massie 1996, Dowell 1997, 1998, Wang *et al.* 1999).

In the NIR region, absorption maxima were found at 900, 928, 1300, 1726, 1960, 2126 and 2310 nm that accounted for identifying seed sources (Figure 8A) while the peak at 1930 nm had the largest influence in both PLS factors for identifying mothers (Figure 8B). Other absorption maxima contributing to the identification of mothers appeared at 1126, 1204, 1440, 1506, 1520, 1724, 2190 and 2308 nm. The absorption peaks at 900 and 928 nm are characterised by the C – H stretching third overtones and the 1100-1300 nm wavelength region corresponds to the second overtone C – H stretching modes of vibration where several compounds show characteristic absorption; notably molecules with methyl and methylene structures and oil (Osborne *et al.* 1993, Shenk *et al.* 2001). Several studies have shown that the absorption bands in the 1700-1800 nm wavelength regions strongly correlate with fatty acids (tripalmitin, triolein and trilinolein) in seeds of several oil crops (*e.g.* Cho and Iwamoto 1989, Sato *et al.* 1991, 1995, 1998, Velasco *et al.* 1996, 1997, Daun and Williams 1997, Hourant *et al.* 2000). The longer wavelength region is characterized by N – H stretching vibration as well as combinations of C – H stretch and C – H deformation (Osborne *et al.* 1993, Shenk *et al.* 2001).

Evidently, the success of identifying seed source in the NIR region could be attributed to divergence in lipids. This is further corroborated by an earlier study that has documented an increase in total lipid content of Scots pine seeds with latitude (Tillman-Sutela *et al.* 1995). Furthermore, a wide difference between dark and light seed coats in the relative content of saturated and unsaturated fatty acids has been reported (Grzywacz and Rosochacka 1980); especially erucic acid is uniquely found in black-coloured seeds. Several other organic and inorganic

compounds have been observed from the seed coat extracts of varying colour types (Rosochacka and Grzywacz 1980). It appeared that the absorption peak at 1930 nm had the largest contribution for the explanation of the between-mothers variability. This band is characterized by a combination of O – H stretch and starch is responsible for absorption (Shenk *et al.* 2001). The possible explanation for the divergence in starch content in seeds from different mothers could be differences in the degree of maturity of the surface structures. Tillman-Sutela *et al.* (1998) have shown that the sacrotesta, the outermost layer of the seed coat, contains starch grains in seeds that are not fully matured while the matured seeds are empty or filled with brownish granular substances.

Detection of internal insect infestation

The potential of NIR spectroscopy for the detection of internal insect infestation in seeds of *Cordia africana* was investigated. The result showed that the calibration model derived from OSC-spectra was excellent in terms of model complexity (A = 2), explaining the variation between infested and sound seeds ($R^2Y = 0.94$) and the overall predictive power of the calibration set ($Q^2 = 0.94$) compared with raw (A = 5, $R^2Y = 0.75$, $Q^2 = 0.71$), first derivative (A = 4, $R^2Y = 0.83$, $Q^2 = 0.80$) and MSC (A = 3, $R^2Y = 0.70$, $Q^2 = 0.69$) spectra. The possible sources of systematic variation could be light scattering due to the rough surface of the seed and path length difference arising from positioning of individual seeds during scanning. Seeds of *Cordia* have round oval shape and a larva within a seed was sometimes positioned on top and sometime below during scanning, and hence creating path length difference. For samples in the test set, the classification rates for both sound and infested seeds were 100%. The raw-model resulted in slightly better classification rate for both classes (92% and 88% for sound and infested seeds, respectively) compared with the first derivative and MSC models.

A plot of PLS weights showed analogous profile with the difference spectrum obtained by subtracting the average spectrum of infested seeds from those of sound seeds. Major absorption peaks appeared at 1360, 1380, 1830, 1870 and 1902 nm (Figure 8C). This indicates that the chemical signal from insect larva was the basis for the classification of infested and sound seeds. Insect cuticular lipids are composed mainly of fatty acids, alcohols, esters, glycerides, sterols, aldehydes, ketones and hydrocarbons (Lockey 1988) as well as protein, catachols, pigments and oxalates (Kramer *et al.* 1995). The observed absorption peaks in the 1300 – 1400 nm corresponds to C – H combinations and O – H first overtone (Shenk *et al.* 2001) while the 1820 – 1880 nm wavelength region corresponds to C – H deformation (Murray and Williams 1987). Functional groups responsible for absorbance in these regions are CH_2 and CH_3, which are the common chemical moieties in fats and oils, which in turn are the major components of insect cuticle. Dowell *et al.* (1998) analysed spectra of the chitin hexamer (β-(1-4)-linked hexasaccharide of 2-acetamido-2-deoxy-D-glucopyranoside) and ground insect cuticle; and found absorption peaks around 1178 and 1500 nm, which are not distinctively seen in the present study, but still contributed to the discrimination of infested and sound seeds fairly well.

The 1900 – 1960 nm region corresponds mainly to O – H first overtone, which could be attributed to high moisture in infested seeds due to respiratory metabolism of hidden larvae (note that hidden larvae were alive). The shorter wavelength region has also some smaller peaks around 770, 808, 874 and 938 nm that correspond to N – H and C – H third overtones (Osborne *et al.* 1993). Structures typical of protein and lipid were responsible for absorption in this region, which in turn could be due to some proteins and lipids in insect cuticle. Absorption bands reported here agree with those determined by Ridgway and Chambers (1996, 1998), Ghaedian and Wehling (1997) and Dowell *et al.* (1998, 1999, 2000).

An extension of the study on internal insect infestation was conducted on *Picea abies* seeds in order to examine whether discrimination of uninfested and infested seeds by NIR spectroscopy is sensitive to seed origin and year of collection. Calibration models were developed on five seed lots collected from Sweden, Finland and Belarus at different years. Prior to modelling, between-seed lot spectral variation that had no relevance for discriminating the two fractions was removed using OSC treatment. Calibration models developed on each seed lot after extracting two OSC components described efficiently the variation between uninfested and infested seeds ($R^2Y \geq 0.917$) with an excellent overall predictive power ($Q^2 \geq 0.900$) according to cross validation. In all cases, the spectral information was summarized with one significant PLS factor, which concurs with the actual phenomenon in the data (either uninfested or infested). Each single lot model resulted in 100% classification rate for samples drawn from the same seed lot used to build the discriminant models, except calibration model derived from stand seeds of Belarus that misclassified 5% of uninfested seeds. New samples drawn from other seed lots were also discriminated with nearly 100% accuracy (Table 1). Pretreatment of the spectra with OSC prior to model building was paramount to remove subtle differences in reserve compounds (total lipid and protein contents) as well as moisture among seed lots, thereby generating a robust single lot model. A similar result was reported earlier where variation in moisture among samples significantly reduced the detection of internal insect larvae in wheat kernels while the levels of protein showed little effect (Dowell *et al.* 1998).

For comparison, discriminant models were developed by pooling calibration sets of each seed lot. The results showed that the discriminant model computed using raw data set explained 83.6% of the variation between infested and uninfested seeds (R^2Y) and 82.1% of the predicted variation with six significant factors according to cross validation. With two significant factors, however, the OSC- model explained 92.1% and 91.9% of the between-class and predicted variations, respectively. Both models completely detected infested seeds in the test set. However, the raw-model misclassified 4% of uninfested seeds while the OSC-model resulted in a 100% classification rate for uninfested seeds. As a whole, the classification accuracy using either single lot model or pooled model is similar, suggesting that calibration model developed on a single seed lot can be used for rapid assessment of infestation rate in other seed lots irrespective of their origin or year of harvest.

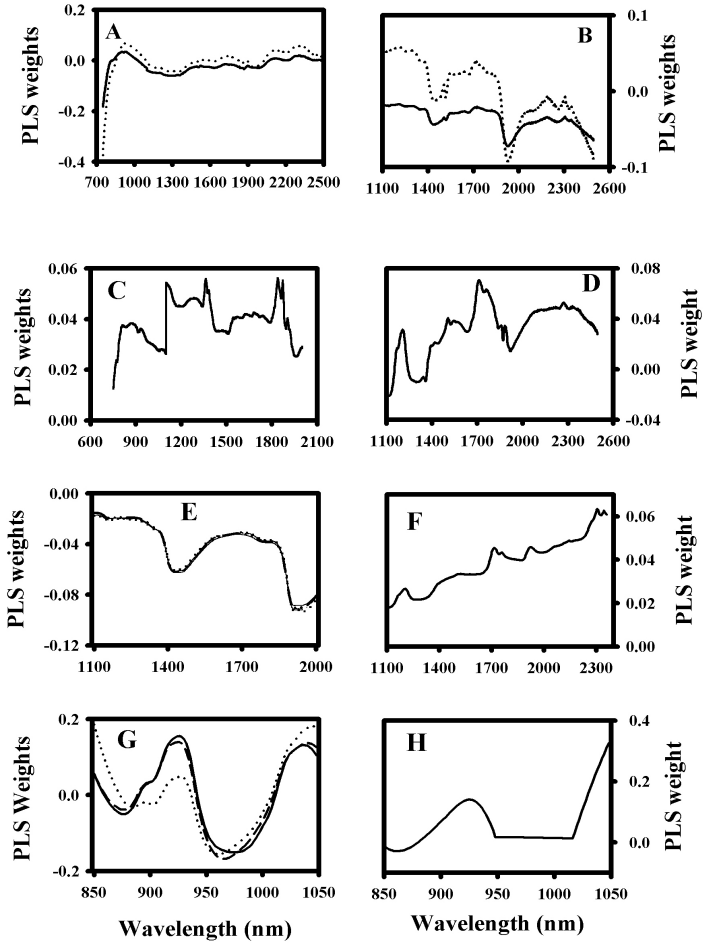

Figure 8. PLS weight plots depicting wavelength regions that influenced the identification of seed sources (panel A), mothers of *Pinus sylvestris* (panel B), discrimination of sound and infested seeds of *Cordia africana* (panel C) and *Picea abies* (panel D), sound and insect-damaged seeds of *Albizia schimperiana* (panel E), viable and empty seeds of *Pinus patula* (panel F), filled, empty and infested seeds of three *Larix* species (panel G) and vigorous and aged seeds of *Pinus patula* (panel H). Note in panels A & B the solid line is weight spectrum for the first factor and dotted line for the second factor; in Panel E the weight spectra from the different imbibition times are highly overlapped. In panel G, the solid, dashed and dotted lines stand for *L. decidua*, *L. sukaczewii* and *L. gmelinii*, respectively.

Unlike the previous study on *Cordia africana* (**II**), the origin of spectral difference between infested and uninfested seeds is attributed to storage reserves that are depleted in the former by the feeding larvae. The difference spectrum, computed by subtracting the average spectrum of uninfested seeds from that of infested ones, revealed major absorption peaks at 1210, 1506, 1710, 1760 and

2276 nm. These peaks had the largest PLS weights and hence highly influenced the discriminant model (Figure 8D). Absorption bands in these regions mainly correlate with lipids and proteins due to C – H second overtone, N – H stretch first overtone and C – H stretch first overtone (Osborne *et al.* 1993, Shenk *et al.* 2001). Lipids are the major storage reserve in spruce seeds, accounting 28.33% (**III**) followed by proteins, constituting 17.43% of the chemical composition of the seed. The dominant fatty acids in spruce seeds are linoleic (C18: 2n-6), trienoic (C18: 3 5c9c12c) and oleic (C18: 1n-8) acids, which represent 49, 25 and 12 mol% of the total fatty acids respectively (Tillman-Sutela *et al.* 1995, Wolff *et al.* 2001).

Table 1. *Classification rate (%) of uninfested (US) and infested (IS) seeds of Picea abies in the external test sets by single lot models. Note bold-faced values are classification rates for test samples drawn from the same seed lots used to develop the calibration models*

	Classification rates									
	Sweden-O		Sweden-S		Finland-O		Finland-S		Belarus	
Models*	US	IS	US	IS	US	IS	US	IS	US	IS
PLS-SO	**100**	**100**	100	100	100	100	100	100	100	100
PLS-SS	100	95	**100**	**100**	100	100	100	100	100	100
PLS-FO	100	100	100	100	**100**	**100**	100	100	95	100
PLS-FS	100	100	100	100	100	100	**100**	**100**	95	100
PLS-B	100	100	95	100	100	100	100	100	**95**	**100**

* *PLS-SO and PLS-SS are calibration models developed using orchard and stand seeds from Sweden, respectively; PLS-FO and PLS-FS are models derived from orchard and stand seeds from Finland, respectively and PLS-B is model developed using stand seeds from Belarus. The letters O and S after each country denotes orchard and stand seeds.*

Separation of sound and insect-damaged seeds

NIR spectroscopy was used to separate sound and insect-damaged seeds of *Albizia schimperiana* soaked in water for one, three, six, nine and twelve hours at room temperature. The calibration models described more than 95% of the spectral variation (R^2X) with few significant PLS factors according to cross validation irrespective of the imbibition time. Nonetheless, a calibration model calculated from one-hour imbibition treatment poorly described the variation between sound and damaged seeds ($R^2Y = 0.407$, $Q^2 = 0.398$). With increasing imbibition time,

the computed models explained more than 79% of the variation between the two groups with an excellent overall prediction ability ($Q^2 > 0.795$) for the calibration set.

Classification rates of sound and damaged seeds in test sets by PLS models computed on full spectrum and selected absorption bands are shown in Table 2. Although 96% of sound seeds were recognized correctly, damaged seeds imbibed for one hour were poorly predicted. The classification rate of damaged seeds was improved with increasing imbibition time, and a complete separation of sound and damaged seeds was achieved after six, nine and twelve hours of imbibition. Classification by selected NIR absorption bands gave more or less similar results with 'full' spectrum models. Six and twelve hours of imbibition resulted in consistent classification rate in both 'full' spectrum and selected absorption band models.

Table 2. *Classification rates (%) of sound (SS) and insect-damaged (DS) seeds of Albizia schimperiana in the test set using calibration models developed on 1100 – 2000 nm, 1400 – 1500 nm and 1900 – 2000 nm wavelength regions*

	Wavelength regions					
	1100 – 2000		1400 – 1500		1900 – 2000	
Treatments	SS	DS	SS	DS	SS	DS
1 hr-soaked	96	56	100	56	96	59
3 hr-soaked	100	96	100	96	100	92
6 hr-soaked	100	100	100	100	100	100
9 hr-soaked	100	100	100	96	100	100
12 hr-soaked	100	100	100	100	100	100

Analysis of PLS weight plots revealed two broad absorption bands in the 1400 – 1500 nm and 1900 – 2000 nm wavelength regions with peaks *ca*. at 1450 nm and 1940 nm (Figure 8E). Pure water has absorption peaks at 1450 and 1940 nm due to O – H stretch first overtone and combination bands involving O – H stretch and O – H bend although these bands are subject to shift as a result of variation in temperature and in hydrogen bonding when water is in a solvent or solute admixture (Osborne *et al.* 1993). The broad absorption peaks found in this study are similar to NIR absorption peaks of pure water. As expected, the classification models utilized mainly spectral difference attributed to relative water content to distinguish sound and damaged seeds.

Discrimination of empty and viable seeds

The potential of NIR transmittance (log 1/T) and reflectance (log 1/R) spectroscopy for the discrimination of viable and empty seeds of *Pinus patula* was investigated. Using transmittance spectra, the calibration model explained 93% of the between-class variation (R^2Y) and 92.8% of the predicted variation (Q^2) with two significant PLS factors according to cross-validation. However, the calibration model developed on reflectance spectra described 85.2% and 85.1% of the class and predicted variations, respectively with one significant PLS factor. Calibration models derived from selected absorption bands also substantially described the between-class variability with very good prediction ability for the calibration as well as test sets. For new samples in the test set, the calibration model developed on log 1/T data set classified viable and empty seeds with 100% accuracy. In contrast, a PLS model computed based on 'full' NIR reflectance spectra (1100 – 2360 nm) resulted in 96% and 88% classification rates for viable and empty seeds, respectively.

The difference spectra indicated that viable seeds absorbed more of the incident radiation than empty seeds; and absorption peaks typical of viable seeds were found at 926, 1170, 1206, 1716, 1760, 2308 and 2346 nm. For the reflectance spectra, the PLS weight plot that revealed identical absorption maxima with that of the difference spectrum is shown in Figure 8F. Evidently, the origin of spectral difference between the two fractions is attributed to differences in the availability of reserve compounds. It should be noted that empty seeds are devoid of the storage organ and the embryo, and hence no deposition of reserve compounds. The major reserve compounds in pine seeds are oil, protein and carbohydrate (mainly starch), which account 48, 35 and 6% of the total seed composition, respectively (Bewley and Black 1994, Miquel and Browse 1995). The dominant fatty acid compositions of the oil from *Pinus patula* seeds are linoleic, 9,12-18:2 (46.85%), pinolenic, 5,9,12-18:3 (19.96%), and oleic, 9-18:1 (16.26%) acids (Wolff *et al.* 1997). These authors also reported several Δ5-olefinic acids, the sum of which accounts *ca.* 26.33% of the total fatty acids; and two major saturated acids: Palmitic (16:0) and stearic (18:0) acids.

The observed absorption peaks correlate with fatty acids due to C – H stretching vibration of various functional group: CH_2, CH_3, CH = CH (Osborne *et al.* 1993, Shenk *et al.* 2001). Osborne *et al.* (1993) has described that the major absorption band in fat or oil is due to a long chain fatty acid moiety that gives rise to CH_2 second overtone at 1200 nm; and the band near 1180 nm has been assigned as the second overtone of the fundamental C – H absorption of pure fatty acids containing *cis* double bonds, *e.g.* oleic acid, (Sato *et al.* 1991). Several authors have extensively studied the absorption bands in the 1700 – 1800 nm wavelength region in relation to fatty acid and oil characterization (*e.g.* Chow and Iwamoto 1998, Reinhardt and Röbbelen 1991, Sato *et al.* 1995, 1998, Velasco *et al.* 1996, 1997, 1998b, Daun and Williams 1997 and Hourant *et al.* 2000). The absorption maxima in the vicinity of 2308 and 2346 nm are characteristics of CH_2 stretch and bend combinations as well as other vibrational modes and a positive correlation

between total polyunsaturated fatty acids (18:2 and 18:3) and absorbance in the 2050 – 2230 nm region have been documented (Hourant *et al.* 2000). It was also reported that the NIR spectra of groundnut oil and liquid paraffin showed typical absorption bands at 2310 and 2345 nm (Osborne *et al.* 1993). Apparently, the absorption bands observed in our study could be correlated to the dominant fatty acids in *Pinus patula* seeds; linoleic, pinolenic and oleic acids as well as several polyunsaturated fatty acids, such as Δ5-olefinic acids.

Simultaneous detection of filled, empty and insect-infested seeds

PLS discriminant models were developed to distinguish empty and insect infested seeds from filled seeds of three *Larix* species based on NIR transmittance spectra. The computed models for each species as well as the composite model explained more than 80% of the variation between seed fractions (R^2Y) with 3 significant PLS factors according to cross validation. The overall prediction ability for calibration sets ($Q^2cv \geq 0.796$) as well as for test sets ($Q^2test \geq 0.858$) was excellent for all models. This shows that the NIR spectroscopy data contained much information that can be used to discriminate filled seeds from empty and insect infested seeds. For new samples in the test sets, discriminant models computed for each species separately resulted in 100% recognition of empty and insect infested seeds for all species. The recognition rate of filled seeds was, however, varied between species; the highest being for *Larix sukaczewii* (100%) followed by *Larix decidua* (97%) and *Larix gmelinii* (90%). The composite model developed by combining spectra of all species also resulted in 100% recognition rates for empty and insect infested seeds. Although the recognition rate of filled seeds of *L. gmelinii* and *L. sukaczewii* remained unchanged, the full spectrum composite model slightly reduced the recognition rate of filled seeds of *L. decidua* (95%) compared to the model developed separately for each species.

Simultaneous discrimination of filled, empty and insect infested seeds of three *Larix* species based on two selected NIR absorption bands, 890 – 940 nm and 1000 – 1048 nm, was also successful. A 3-factor PLS model explained more than 72% of the class variation (R^2Y) with a very good predictive power for the calibration sets ($Q^2 > 72\%$) in both wavelength regions. For all species, empty and insect infested seeds were completely distinguished from filled seeds using either of the selected absorption bands. The truncated spectra in the 890 – 940 nm wavelength region misclassified 3% of filled seeds of *L. sukaczewii* while the 1000 – 1048 nm wavelength range resulted in complete recognition. While the shorter wavelength region (92% cf. 90% in the full spectrum model) slightly increased the recognition rate of filled seeds of *L. gmelinii*, the longer wavelength region of the spectra resulted in complete detection. The recognition rate of filled seeds of *L. decidua* was also slightly improved (97%) by selected absorption bands compared with that of the full spectrum model. As a whole the results indicate the possibility of developing filter type sorting instrument for large-scale seed cleaning operations that would be less expensive than monochromatic grating based equipment.

The difference spectra revealed that filled seeds of all three species showed unique absorption in the 900 – 950 nm and 1000 – 1048 nm regions with small bumps at 898 nm. While empty and insect infested seeds absorbed more of the NIR radiation in the 850 – 900 nm and 950 – 1000 nm regions with peaks at 874 and 974 nm. The discriminant models had the largest weights in this region, too (Figure 8G). The observed absorption bands from filled seeds correlate with lipids and proteins (Murray and Williams 1987, Osborne *et al.* 1993). Previous studies on chemical content of *Larix* seeds have showed that lipids are the dominant reserve compounds as in the case of many other conifers (Wolff *et al.* 1997, 2001). The variation in reserve compounds between filled and empty seeds is obvious as storage organs, and hence storage reserves, are absent in the latter. However, the difference in the amount of storage reserves between filled and insect infested seeds could arise either from complete depletion of reserve compounds by the feeding larva or because the attack might have occurred early during seed development, and hence resulted in empty seeds. By cutting a sample of insect infested seeds and examining its contents, we observed that both the embryo cavity and the megagametophyte were totally absent, and it is only the larvae that were found enclosed within the seed coats. Studies on seeds of other conifers, such as *Pseudostuga menziesii*, have shown apparent depletion of lipids and proteins by the feeding larvae, and a seed severely attacked by feeding larvae was empty of its contents (Bates *et al.* 2000, 2001). The absorption bands typical of empty and infested seeds could be related to the chemical composition of the seed coat or testa, mainly phenolics (Bewley and Black 1994, Copeland and McDonald 2001) and to moisture from the respiratory activity of feeding larvae. This has been reported earlier where infested kernels of wheat showed high absorption in the region that correlates to moisture (Baker *et al.* 1999, Ridgway *et al.* 1999).

Rapid analysis of seed vigour

NIR transmittance spectroscopy was employed to classify vigorous and aged seeds, thereby serving as a tool for rapid and non-destructive analysis of vigour. The SIMCA analysis showed a clear differentiation of vigorous and aged seeds for new observations in the prediction set (Figure 9). No samples were misclassified as member of the other class, although few samples from each class were outside the 95% prediction confidence limit of the respective class. The number of outlying samples was higher for aged seeds than vigorous seeds. This is, in fact, expected because individual seeds do not behave in a similar manner to the ageing treatment. Small seeds usually deteriorate more rapidly than large seeds in the high relative humidity environment due to their high surface to volume ratio (MacDonald 1999).

Discrimination of vigorous and aged seeds as well as among classes of aged seeds was performed using PLS regression. As the model computed to simultaneously differentiate the four vigour classes was poor ($R^2Y = 0.27$ and $Q^2 = 0.26$), a separate PLS model was computed to discriminate between vigorous and aged seeds (PLS-all), between seeds aged for three and seven days (PLS-37), and between seeds aged for three and nine days (PLS-39). Discrimination of seeds aged for seven and nine days was not successful and hence not reported. The PLS-all

model explained 74% of the variation between vigorous and aged seeds (R^2Y) with one significant factor (A) using the 'full' spectral region. The prediction ability (Q^2 = 0.73) due to cross validation was also very good. A similar result was achieved using truncated spectra, 890 – 948 nm, where lipids show characteristic absorption. The PLS-37 model developed on 'full' spectra described 51% of the class variation, although the prediction ability for the calibration set was low (Q^2 = 0.45). Discriminant analysis performed using truncated spectra slightly improved the prediction ability (Q^2 = 0.50) without changing the explained class variation much. The PLS-39 model computed on both 'full' and truncated spectra explained more than 55% of the variation between three- and nine-day aged seeds and more than 52% of the predicted variation according to cross validation.

For new samples in the prediction sets, a complete recognition of vigorous seeds, seeds aged for seven or nine days was achieved while one sample aged for three days was misclassified as vigorous seed. Among aged seeds, it was also possible to differentiate seeds aged for three, seven and nine days with 80%, 90% and 75% accuracies, respectively. It appeared that the misclassification of seeds aged for nine days was much higher than seeds aged for seven days. Again, this could be associated with the rate of deterioration of individual seeds during the ageing process. In general, the result demonstrates that NIR spectroscopy has a great potential for rapid analysis of vigour as well as for sorting of vigorous and aged seeds. The success of differentiating seeds exposed to different accelerated ageing durations highlights the possibility of monitoring the progress of deterioration provided that other sources of variation are minimized or controlled.

The weight plots for the first and second PLS factors revealed that the 850 – 880, 890 – 940, 1010 – 1030 nm wavelength regions influenced the discrimination of vigorous and aged seeds as well as between aged seeds (Figure 8H). The absorption band in the 850 – 880 nm corresponds to the third overtone C – H stretching vibration and molecular structures responsible for absorption are benzene and chloromethane with absorption maxima at 874 nm and 880 nm, respectively (Osborne *et al.* 1993). This might be attributed to volatile compounds, such as aldehydes, alcohols, ketones, esters, and terpenes that evolve during storage and artificial ageing (Zhang *et al.* 1993, 1995, MacDonald 1999). The absorption peaks at 928 nm and 1020 nm have been assigned to oil and protein, respectively (Osborne *et al.* 1993).

There are growing evidences supporting the reduction in the quantity of unsaturated fatty acids, such as linoleic (18:2) and linolenic acid (18:3), as well as total lipid and phospholipid contents with accelerated and natural ageing of seeds (*e.g.* Pukacka and Kuiper 1988, Pukacka 1991, Marquez-Millano *et al.* 1991, Kalpana and Madhava Rao 1996, Thapliyal and Conner 1997). A decrease in total protein content and soluble protein in pigeonpea, peanut and sal seeds with accelerated ageing has also been documented (Nautiyal *et al.* 1985, Jeng and Sung 1994, Kalpana and Madhava Rao 1997). Apparently, the divergence in lipids and proteins is the basis for the discrimination of vigorous seeds from aged seeds with NIR spectroscopy. This region was also useful for discriminating between aged seeds that could be attributed to the rate of depletion of lipids and proteins with duration of ageing.

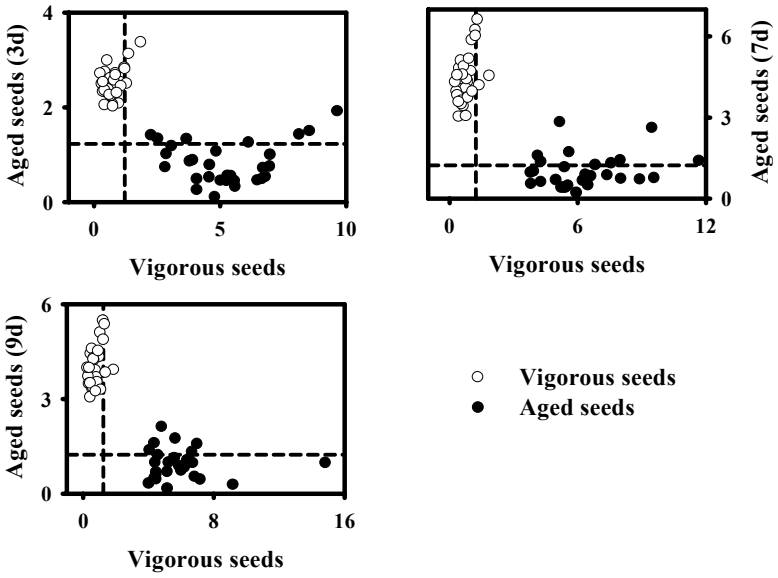

Figure 9. Classification of vigorous and aged seeds in the prediction set using SIMCA. Dashed lines denote the 95% prediction confidence interval for class membership for vigorous (vertical) and aged (horizontal) seeds.

Conclusions

The potential of NIR spectroscopy combined with multivariate analysis was evaluated for the characterization of forest tree seed quality. The results reported in this thesis demonstrate the capability of the technique for rapid and non-destructive analysis of the physiological, technical and genetic attributes of seed quality. As

establishment of new forest plantation shows an increasing tendency globally, NIRS will play a pivotal role in upgrading seed lot quality through sorting of unproductive seeds, and hence facilitating single seed sowing for containerised seedling production in nurseries and/or direct sowing out in the field.

A promising result was achieved in identifying seed source and parents of *Pinus sylvestris* using VIS+NIR spectroscopy. The result highlights the potential of the technique for routine authentication of putative seed origin of pine seeds and for characterizing and sorting seeds according to their genotype. Detection of internal insect infestation in forest tree seed using NIRS was also successful. The results show that subtle differences in protein and lipid contents as well as moisture among seed lots did not affect the classification accuracy provided that the calibration model takes into account these natural variability (pooled calibration model) or such variability is removed *a priori* with appropriate spectral pretreatment such as OSC. NIRS has demonstrated the capability for separating sound and insect-damaged seeds of *Albizia schimperiana* based on differences in relative water content between the two fractions. Since the method is based on a universal phenomenon, *i.e.*, seeds with hard and impermeable seed coats do not absorb water unless the surface is punctured in some way, it can easily be extended to several legumes and other species known to have hard seed coats. The specific imbibition time for each species should, however, be empirically determined.

Classification of viable and empty seeds of *Pinus patula* using near infrared transmittance and reflectance spectroscopy was successful. The technique is rapid and more efficient as it takes a fraction of a minute to scan a single seed, and no sample preparation is needed unlike, for example, the IDS technique. In addition, it can easily be extended to other species as the principle is based on a universal phenomenon, *i.e.*, reserve compounds that are found only in viable seeds are detected by NIRS. Filled, empty and insect infested seeds of three *Larix* species were successfully detected using NIR spectroscopy. Thus, the result highlights the potential of NIR spectroscopy as a rapid and non-destructive technique to upgrade seed lot quality of *Larix* species by sorting out empty and infested seeds simultaneously. Furthermore, the technique can offer a unique opportunity for seed orchard managers to rapidly evaluate the efficacy of artificial pollination and the success of cultural treatments in reducing the quantity of empty seed production in seed orchards.

NIR spectroscopy has demonstrated a great potential as a rapid and non-destructive method for vigour test as well as sorting deteriorated seeds from a seed lot thereby enhancing its performance. Although the SIMCA analysis resulted in complete classification of vigorous and aged seeds, the PLS models were more accurate in classifying not only vigorous and aged seeds, but also among classes of aged seeds. Interestingly, the classification is based mainly on the underlying biochemical changes associated with seed deterioration.

Classification and/or discrimination of seeds based on their quality attributes were also evaluated using selected absorption bands. In many cases, similar results were achieved with that of the full spectrum models. This underscores the prospect

of developing simple and less expensive instrument based on filters, diode arrays or lasers for commercial purpose. Therefore, continued emphasis should be given towards developing automated sorting equipment for large-scale seed cleaning and/or upgrading seed lot performance. Optimistically, NIR spectroscopy will become one of the seed testing methods in the foreseeable future.

References

Baker, J.E., Dowell, F.E. and Throne, J.E. 1999. Detection of parasitized rice weevils in wheat kernels with near-infrared spectroscopy. *Biological Control* 16: 88-90.

Barbosa, P. and Wagner, M.R. 1989. *Introduction to Forest and Shade Tree Insects*. Academic press Inc., San Diego, California, pp. 277-290.

Barnes, B.J., Dhanoa, M.S. and Lister, S.J. 1989. Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra. *Applied Spectroscopy* 43: 772-777.

Barnett, J.P. 1971. Flotation in ethanol reduced storability of Southern pine seeds. *Forest Science* 17: 50-54.

Basu, R.N. 1994. Seed viability. In *Seed Quality: Basic Mechanisms and Agricultural Implications* (ed. A.S Basra), pp. 1-44. Food products press, New York.

Bates, S.L., Borden, J.H., Savoie, A. and Blatt, S.E. 2000. Impact of feeding by *Leptoglossus occidentalis* (Hemiptera: Coreidae) on the major storage reserves of mature Douglas-fir (Pinaceae) seeds. *Canadian Entomologist* 132: 91-102.

Bates, S.L., Lait, C.G., Borden, J.H. and Kermode, A.R. 2001. Effect of feeding by the western conifer seed bug, *Leptoglossus occidentalis*, on the major storage reserves of developing seeds and on seedling vigour of Douglas-fir. *Tree Physiology* 21: 481-487.

Bazzaz, F.A., Ackerly, D.D. and Reekie, E.G. 2000. Reproductive allocation in plants. In *Seeds: the Ecology of Regeneration in Plant Communities* (ed. M. Fenner), 2nd ed., pp. 1-29. CABI Publishing, Wallingford, UK and New York, USA.

Bekele, A., Birnie, A. and Tengnäs, B. 1993. *Useful Trees and Shrubs for Ethiopia*. Regional soil conservation unit, Swedish International Development Authority, Nairobi, Kenya. 474 pp.

Ben-Gera, I. and Norris, K.H. 1968. Determination of moisture content in soybeans by direct spectrophotometry. *Israel Journal of Agricultural Research* 18: 125-132.

Bergsten, U. and Wiklund, K. 1987. Some physical conditions for removal of mechanically damaged *Pinus sylvestris* L. seeds by using the PREVAC method. *Scandinavian Journal of Forest Research* 2: 315-323.

Bergsten, U. and Sundberg, M. 1990. IDS-sedimentation of *Cupressus lusitanica* seeds. In *Tropical Tree Seed Research* (ed. J.W. Turnbull). pp. 99-102. Proc. of an international workshop held at the Forestry Training Centre,

Gympie, Qld, Australia, 21-24 Aug. 1989, ACIAR proceedings No. 28, Canberra, Australia.

Bewley, J.D. and Black, M. 1994. *Seeds: Physiology of Development and Germination*. 2nd edition. Plenum press, New York. 445pp.

Birth, G.S. and Hecht, H.G. 1987. The physics of near infrared reflectance. In *Near-Infrared Technology in the Agricultural and Food Industries* (eds. P.C. Williams and K.H. Norris). pp 1-15. American Association of Cereal Chemists, St. Paul Minnesota.

Bittner, D.R. and Norris, K.H. 1968. Optical properties of selected fruits VS maturity. *Transaction of the American Society of Agricultural Engineers* 11: 534-536.

Blanco, M. and Villarroya, I. 2002. NIR spectroscopy: a rapid-response analytical tool. *Trends in Analytical Chemistry* 21: 240-250.

Bokobza, L. 1998. Near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* 6: 3-17.

Bonner, F.T. 1986. Measurement of seed vigour for loblolly and slash pines. *Forest Science* 32: 170-178.

Bonner, F.T. 1998. Testing tree seeds for vigour: a review. *Journal of Seed Technology* 20: 5-17.

Burns, D.A. and Ciurczak, E.W. 2001. *Handbook of Near-Infrared Analysis*. 2nd edition, Marcel Dekker Inc., New York. 803 p.

Cho, R.K. and Iwamoto, M. 1989. The purity identification of sesame oil by near infrared reflectance spectroscopy. *Proceedings of the 2nd International NIRS conference*, Tsukuba, Japan.

Ciurczak, E.W. 2001. Principles of Near-infrared Spectroscopy. In *Handbook of Near-Infrared Spectroscopy* (eds. D.A. Burns and E.W. Ciurczak). 2nd ed., pp. 7-18. Marcel Dekker Inc., New York.

Copeland, L.O. and McDonald, M.B. 2001. *Principles of Seed Science and Technology*. Kluwer Academic Publishers, Boston/ Dordrecht/ Lonodon. 467 pp.

Dajoz, R. 2000. *Insects and Forests: The role and diversity of insects in the forest environment*. Intercept Ltd, London-Paris-New York. 668 pp.

Daun J.K., and Williams P.C. 1997. Near Infrared Analysis of Oilseeds – Current status and future directions. In *New Techniques and Applications in Lipid Analysis* (eds. R. MacDonald and M. Mossoba). pp 266-282. AOCS Press, Champaign.

Davies, A.M.C. 1990. Subdivisions of the infrared region. *Applied Spectroscopy* 44: 14A.

Delouche, J.C. and Baskin, C.C. 1973. Accelerated ageing techniques for predicting the relative storability of seed lots. *Seed Science and Technology* 1: 427 -452.

Delwiche S.R. and Massie D. R. 1996. Classification of wheat by visible and near-infrared reflectance from single kernels. *Cereal Chemistry* 73: 399-405

Demelash, L., Tigabu, M. and Odén, P. C. 2002. Separation of empty and dead-filled seeds from a seed lot of *Pinus patula* with IDS technique. *Seed Science and Technology* 30: 677-681.

Demelash, L., Tigabu, M. and Odén, P. C. 2003a. Evaluating the relative storability of IDS-treated and untreated *Pinus patula* seeds by accelerated aging. *Journal of Tropical Forest Science* (in press).

Demelash, L., Tigabu, M. and Odén, P. C. 2003b. Enhancing germinability of *Schinus molle* L. seed lot from Ethiopia with specific gravity and IDS techniques. *New Forests* (in press).

Donald, D.G.M. 1985. The separation of full dead seed from live seed in *Pinus elliottii*. In *International Symposium on Nursery Management Practices for the Southern Pine* (ed. D. South). pp. 83-87. Auburn University, Ala.

Dormling, I. and Johnsen, Ø. 1992. Effects of the parental environment on full-sib families of *Pinus sylvestris*. *Canadian Journal of Forest Research* 22: 88-100.

Dowell, F.E. 1997. Effect of NaOH on visible wavelength spectra of single wheat kernels and colour classification efficiency. *Cereal Chemistry* 74: 617-620.

Dowell, F.E. 1998. Automated colour classification of single wheat kernels using visible and near-infrared reflectance. *Cereal Chemistry* 75: 142-144.

Dowell, F.E., Throne, J.E. and Baker, J.E. 1998. Automated non-destructive detection of internal insect infestation of wheat kernels by using near-infrared reflectance spectroscopy. *Journal of Economic Entomology* 91: 899-904.

Dowell, F.E., Throne, J.E. and Baker, J.E. 1999. Identifying stored-grain insects using near-infrared spectroscopy. *Journal of Economic Entomology* 92: 165-169.

Dowell, F.E., Broce, A.B., Xie, F., Thorne, J.E. and Baker, J.E. 2000. Detection of parasitized fly puparia using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* 8: 259-265.

Downie, B. and Bergsten, U. 1991. Separating germinable and non-germinable seeds of eastern white pine (*Pinus strobes* L.) and white spruce (*Picea glauca* (Moebch) Voss) by the IDS technique. *Forestry Chronicle* 67: 393-396.

Downie, B. and Wang, B.S.P. 1992. Upgrading germinability and vigour of jack pine, logdepole pine, and white spruce by the IDS technique. *Canadian Journal of Forest Research* 22: 1124-1131.

El Atta, H.A. 1993. The effect of *Caryedon serratus* Oliver (Col., Bruchidae) on viability and germination of seeds of *Acacia nilotica* (L. Willd. Ex Del.) in the Sudan. *Forest Ecology and Management* 57: 169-177.

El-Kassaby, Y.A., Chaisurisri, K., Edwards, D.G.W. and Taylor, D.W. 1993. Genetic control of germination parameters of Douglas-fir, Stika spruce, western red cedar and yellow cedar and its impact on container nursery production. In *Dormancy and Barriers to Germination* (ed. D.G.W. Edwards). pp. 37-42. Proc. on an international symposium of IUFRO project Group P2.04-00 (Seed problems), Victoria, British Colombia, Canada, April 23-26, 1991.

Eriksson, L., Johansson, E., Wold, N.K. and Wold, S. 1999. *Introduction to Multi and Megavariate Analysis Using Projection Methods (PCA and PLS)*. Umetrics AB, Umeå, Sweden. 490 pp.

Esbensen, K.H. 2000. *Multivariate Data Analysis in Practice: An Introduction to Multivariate Data Analysis and Experimental Design*. 4th edition, CAMO ASA, Norway. 598 pp.

Falleri, E. and Pacella, R. 1997. Applying the IDS method for the removal of empty seeds in Platanus X acerifolia. *Canadian Journal of Forest Research* 27: 1311-1315.

Fichtl, R. and Adi, A. 1994. *Honey bee flora of Ethiopia*. Weikersheim: Margraf Verlag. 510 pp.

Font, R., Rio, M.D., Frnandez-Martinez, J.M., Haro, A.D., del Rio, M., and De Haro, A. 1999. Using near infrared spectroscopy (NIRS) for the determination of bulk density in Indian mustard seed. *Cruciferae Newsletter* 75-76.

Food and Agricultural Organization. 2001. *Global Forest Resources Assessment: 2000 main report*. Rome, Italy, 479 pp.

Friis, I. 1992. *Forests and Forest Trees of North East Tropical Africa*. Kew Bulletin, additional series XV. 396 pp.

Geladi, P., MacDougall, D. and Martens, H. 1985. Linearization and scatter-correction for near infrared reflectance spectra of meat. *Applied Spectroscopy* 39: 491-500.

Ghaedian, A.R. and Wehling, R.L. 1997. Discrimination of sound and granary-weevil-larva-infested wheat kernels by near-infrared diffuse reflectance spectroscopy. *Journal of AOAC International* 80: 997-1005.

Grabe, D.F. 1964. Glutamic acid decarboxylase activity as a measure of seedling vigour. *Proceedings of the Association of Official Seed Analysts*, 54: 100-109.

Grzywacz, A.P. and Rosochacka, J. 1980. The colour of *Pinus sylvestris* L. seed and their susceptibility to damping-off. I The colour and quality of seeds and fatty acid content of the seed coat. *European Journal of Forest Pathology* 10: 138-144.

Gutterman Y 2000. Maternal effects on seeds during development. In *Seeds: the Ecology of Regeneration in Plant Communities* (ed. M. Fenner). 2nd ed. pp 59-84. CABI Publishing, Wallingford New York.

Hakansson, A. 1960. Seed development in *Larix*. *Botanical Notes* 113: 29-40.

Hall, J.P. and Brown, I.R. 1976. Microsporogenesis, pollination and potential yield of seed of *Larix* in NE Scotland. *Silvae Genetica* 25: 132-137.

Hall, J.P. and Brown, I.R. 1977. Embryo development and yield of seed in *Larix*. *Silvae Genetica* 26: 77-84.

Hampton, J.G. 2002. What is seed quality? *Seed Science and Technology* 30: 1-10.

Hampton, J.G. and TeKrony, D.M. 1995. *Handbook of Vigour Test Methods*. International Seed Testing Association, Zurich, Switzerland. 117 pp.

Hannerz, M., Eriksson, U., Wennström, U. and Wilhelmsson, L. 2000. *Scots pine and Norway spruce seed orchards in Sweden – a description with an analysis of future seed supply*. The Forestry Research Institute of Sweden, Report 1. 40 pp.

Harju, A.M. and Nikkanen, T. 1996. Reproductive success of orchard and non-orchard pollens during different stages of pollen shedding in a Scots pine seed orchard. *Canadian Journal of Forest Research* 26: 1096-1102.

Hedlin, A.F., Yates, H.O., Tavor, D.C., Ebel, B.H., Koerber, T.W. and Merkel, E.P. 1981. *Cone and Seed Insects of North American Conifer*. Canadian For. Serv., For. Serv. U.S. Dept. Agric., and Secretaria de Agricultura y Recursos Hidraulicos, Mexico. 122 pp.

Hilton-Taylor, C. 2000. *IUCN Red List of Threatened Species*. IUCN, Gland, Switzerland and Cambridge, UK. 61 pp.

Hindle, P.H. 2001. Historical development. In *Handbook of Near-Infrared Spectroscopy* (eds. D.A. Burns and E.W. Ciurczak). 2 edition. pp. 1-6. Marcel Dekker Inc., New York.

Hodgson, T.J. 1977. Effects of organic flotation media on the germination of *Pinus elliottii*, *P. patula* and *P. taeda* seed. *South Africa Forestry Journal* 102: 17-21.

Holtmeier, F.K. 1995. European larch in middle Europe with special reference to the central alps. In *Ecology and Management of Larix Forests: A Look Ahead* (comps. W.C. Schmidt and K.J. McDonald). pp. 41-49. USDA, Forest Serv., Intermountain Research Sta., General Technical Report, GTR-INT-319.

Hourant, P., Baeten, V., Morales, M.T., Meurens, M. and Aparicio, R. 2000. Oil and fat classification by selected bands of Near-Infrared Spectroscopy. *Applied Spectroscopy* 54: 1168-1174.

Hunde. A and Thulin. M. 1989. Mimosoidea. In *Flora of Ethiopia* (eds. I. Hedberg and S. Edwards). Vol.3. The National Herbarium, Addis Ababa.

International Seed Testing Association 2003. *International Rules for Seed Testing*. International Seed Testing Association, Bassersdorf, CH-Switzerland.

Isaksson, T. and Kowalski, B. 1993. Piece-wise multiplicative scatter correction applied to near-infrared diffuse transmittance data from meat products. *Applied Spectroscopy* 47: 702-709.

Janzen, D.H. 1972. Escape in space by *Sterculia apetala* seeds from the bug *Dysdercus fasciatus* in a Costa Rican deciduous forest. *Ecology* 53: 350-361.

Jeng, T.L. and Sung, J.M. 1994. Hydration effect on lipid peroxidation and peroxide-scavenging enzymes activity of artificially age peanut seed. *Seed Science and Technology* 22: 531-539.

Jones, S.K. and Gosling, P. 1990. The successful redrying of imbibed, or imbibed plus prechilled stika spruce seeds. *Seed Science and Technology* 18: 541-547.

Kalpana, R. and Madhava Roa, K.V. 1995. On the ageing mechanism in pigeonpea (*Cajanus cajan* (L) Millsp.) seeds. *Seed Science and Technology* 23: 1-9.

Kalpana, R. and Madhava Roa, K.V. 1996. Lipid changes during accelerated ageing of seeds of pigeonpea (*Cajanus cajan* (L.) Millsp.) cultivars. *Seed Science and Technology* 24: 475-485.

Kalpana, R. and Madhava Rao, K.V. 1997. Protein metabolism of seeds of pigeonpea (*Cajanus cajan* (L.) Millsp.) cultivars during accelerated ageing. *Seed Science and Technology* 25: 271-279.

Kang, K.S., Lindgren, D. and Mullin, T.J. 2001. Prediction of genetic gain and gene diversity in seed orchard crops under alternative management strategies. *Theoretical and Applied Genetics* 103: 1099-1107.

Kosinski, G. 1986. Megagametogenesis, fertilization and embryo development in *Larix decidua*. *Canadian Journal of Forest Research* 16: 1301-1309.

Kosinski, G. 1987. Empty seed production in European larch (*Larix decidua*). *Forest Ecology and Management* 19: 241-246.

Kramer, K.J., Hopkins, T.L. and Shaefer, J. 1995. Application of solid NMR to the analysis of insect sclerotized structures. *Insect Biochemistry and Molecular Biology* 25: 1067-1080.

Lestander, T. and Bergsten, U. 1985. PREVAC-en metod för att avlägsna mekaniskt skadat frö. [PREVAC- a method for the removal of mechanically damaged seeds]. *Sverige Skogsvårdsförbunds Tidskrift* 1: 35-42.

Lindgren, D. and Wang, Q. 1986. Are genetic results influenced by the environment during seed maturation? In *Breeding Theory, Progeny Testing and Seed Orchards* (ed. R.J. Weir). pp 192-199. Proc. of IUFRO conference, 13-17 Oct 1986, Williamsburg, VA.

Lindgren, D. and Wei, R.P. 1994. Effects of maternal environment on mortality and growth in young *Pinus sylvestris* in field trials. *Tree Physiology* 14: 323-327.

Lockey, K.H. 1988. Lipids of the insect cuticle: origin, composition and function. *Comparative Biochemistry and Physiology. B* 89: 595-645.

Lunn, G. and Madsen, E. 1981. ATP-levels of germinating seeds in relation to vigour of Cauliflower, summer rape, sugerbeets. *Physiologia Plantarum* 53: 164-169.

MacDonald, M.B. (1999). Seed deterioration: physiological, repair and assessment. *Seed Science and Technology* 27: 177-237.

Marcos-Filho, J. and McDonald, M.B. 1998. Sensitivity of RAPD analysis, germination and vigour tests to detect the intensity of deterioration of naturally and artificially aged soybean seeds. *Seed Science and Technology* 26: 141-157.

Marquez-Millano, A., Elam, W.W. and Blanche, C.A. 1991. Influence of accelerated ageing on fatty acid composition of slash pine (*Pinus elliottii* Engelm. var. Elliottii) seeds. *Journal of Seed Technology* 15: 29-41.

Martens, H. and Næs, T. 1989. *Multivariate Calibration*. John Wiley and Sons Ltd, Guildford, Great Britain. 419 pp.

Martinsson, O. 1995. Yield and Productivity of Siberian larch (*Larix sukaczewii* Dyl.) in Northern Sweden. In *Ecology and Management of Larix Forests: A Look Ahead* (Comps. W.C. Schmidt and K.J. McDonald). pp. 244-258. USDA, Forest Serv., Intermountain Research Sta., General Technical Report, GTR-INT-319.

McClure, W.F. 1994. Near-Infrared Spectroscopy: the giant is running strong. *Analytical Chemistry* 66: 43A-53A.

Miquel, M. and Browse, J. 1995. Lipid biosynthesis in developing seeds. In *Seed Development and Germination* (eds. J. Kigel and G. Galili). pp. 169-193. Marcel Dekker Inc., New York.

Murray, I. and Williams P.C. 1987. Chemical properties of near-infrared technology. In *Near-Infrared Technology in the Agricultural and Food Industries* (eds. P.C. Williams and K.H. Norris). pp 17-34. American Association of Cereal Chemists, St. Paul Minnesota.

Næs, T., Isaksson, T., Fearn, T. and Davies, T. 2002. *A User Friendly Guide to Multivariate Calibration and Classification*. NIR Publications, Chichester, UK. 344 pp.

Nautiyal, A.R., Thapliyal, A.P. and Purohit, A.N. 1985. Seed viability in sal. IV. Protein changes accompanying loss of viability in *Shorea robusta*. *Seed Science and Technology* 13: 83-86.

Norris, K.H. 1962. Instrumentation of infrared radiation. *Transaction of the American Society of Agricultural Engineers* 5: 17-20.

Norris, K.H. and Rowan, J.D. 1962. Automatic detection of blood in eggs. *Agricultural Engineering* 43: 154-159.

Norris, K.H. 1964. Reports on the design and development of a new moisture meter. *Agricultural Engineering* 45: 370-372.

Norris, K.H. and Hart, J.R. 1965. Direct spectrophotometric determination of moisture content of grain and seeds. In *Principles and Methods of Measuring Moisture Content in Liquids and Solids* (ed. A. Waxler). pp. 19-25. Vol. IV, Reinhold Publishing Corporation, New York.

Olinger, J.M., Griffiths, P.R. and Burger, T. 2001. Theory of diffuse reflection in the NIR region. In *Handbook of Near-Infrared Spectroscopy* (eds. D.A. Burns and E.W. Ciurczak). pp. 19-52. Marcel Dekker Inc., New York.

Oluoch, M.O. and Welbaum, G.E. 1996. Effect of post harvest washing and post-storage priming on viability and vigour of six-year-old muskmelon (*Cucumis melo* L.) seeds from eight stages of development. *Seed Science and Technology* 24: 75-87.

Osborne, B.G., Fearn, T. and Hindle, P.H. 1993. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis.* 2nd edition, Longman Scientific and Technical, Harlow, UK. 227 p.

Owens, J.N. 1995. Constraints to seed production: temperate and tropical forest trees. *Tree Physiology* 15: 477-484.

Owens, J.N. and Molder, M. 1979. Bud development in *Larix occidentalis* II. Cone differentiation and early development. *Canadian Journal of Botany* 57: 1557-1572.

Owens, J.N., Colangeli, A.M. and Morris, S.J. 1990. The effect of self-, cross, and no pollination on ovule, embryo, seed, and cone development in western red cedar (*Thuja plicata*). *Canadian Journal of Forest Research* 20: 66-75.

Owens, J.N., Morris, S.J. and Catalano, G.L. 1994. How the pollination mechanism and prezygotic and postzygotic events affect seed production in *Larix occidentalis*. *Canadian Journal of Forest Research* 24: 917-927.

Pakkanen, A. and Pulkkinen, P. 1991. Pollen production and background pollination levels in Scots pine seed orchards of northern Finnish origin. In *Pollen Contamination in Seed Orchards* (ed. D. Lindgren). pp. 14-21. Proc. of the Meeting of the Nordic Group for Tree Breeding, August 1991, Sweden.

Polubojarinov, O.I., Chubinsky, A.N. and Martinsson, O. 2000. Decay resistance of Siberian larch wood. *Ambio* 26: 352-353.

Poulsen, K.M. 1995. Application of the IDS-method to *Pinus caribaea* seed. *Seed Science and Technology* 23: 269-275.

Poulsen, K.M. 1996. Prolonged cold, moist pretreatment of conifer seeds at controlled moisture content. *Seed Science and Technology* 24: 75-87.

Pritam, S. and Singh, P. 1997. Forest tree seed pathogens and their management in sustainable forestry. *Journal of Mycology and Plant Pathology* 27: 138-147.

Pukacka, S. 1991. Change in membrane lipid components and antioxidant levels during natural ageing of seeds of *Acer platanoides*. *Physiologia Plantarum* 82: 306-310.

Pukacka, S. and Kuiper, P.J.C. 1988. Phospholipid composition and fatty acid peroxidation during ageing of *Acer platanoides* seeds. *Physiologia Plantarum* 72: 89-93.

Reinhardt, T-C. and Röbbelen, G. 1991. Quantitative analysis of fatty acids in intact rapeseeds by near infrared reflectance spectroscopy. In *the 8th International Rapeseed Conference* (ed. Groupe Consultatif international de Recherche sur le Colza). pp. 1380-1384. Saskatoon, Canada, 9-11 July 1991.

Ribeiro, M.M., LeProvost, G., Gerber, S., Vendramin, G.G., Anzidei, M., Decroocq, S., Marpeau, A., Mariette, S. and Plomion, C. 2002. Origin identification of maritime pine stands in France using chloroplast simple-sequence repeats. *Annals of Forest Science* 59: 53-62.

Ridgway, C. and Chambers, J. 1996. Detection of external and internal insect infestation in wheat by near-infrared reflectance spectroscopy. *Journal of the Science of Food and Agriculture* 71: 251-264.

Ridgway, C. and Chambers, J. 1998. Detection of insects inside wheat kernels by NIR imaging. *Journal of Near Infrared Spectroscopy* 6: 115-119.

Ridgway, C. Cowe, I.A. and Chambers, J. 1999. Detection of grain weevils inside single wheat kernels by a very near infrared two-wavelength model. *Journal of Near Infrared Spectroscopy* 7: 213-221.

Rosochacka, J. and Grzywacz, A.P. 1980. The colour of *Pinus sylvestris* L. seed and their susceptibility to damping-off. II The colour of seed coats and their chemical composition. *European Journal of Forest Pathology* 10: 193-201.

Sato, T., Kawano, S. and Iwamoto, M. 1991. Near-infrared spectral patterns of fatty acid analysis from fats and oils. *Journal of the American Oil Chemists' Society* 68: 827-833.

Sato, T., Takahata, Y., Noda, T., Yanagisawa, T., Morishita, T. and Sakai, S. 1995. Non-destructive determination of fatty acid composition of husked sunflower (*Helianthus annua* L.) seeds by near infrared spectroscopy. *Ibid* 71: 1177-1183.

Sato, T., Uezono, I., Morishita, T. and Tetsuka, T. 1998. Non-destructive estimation of fatty acid composition in seeds of *Brassica napus* L. by near-infrared spectroscopy. *Journal of the American Oil Chemists' Society* 15: 1877-1881.

Savitzky, A. and Golay, M.J.E. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36: 1627-1639.

Schmidt, W.C. and Shearer, R.C. 1995. *Larix occidentalis*: a pioneer of the North American West.. In *Ecology and Management of Larix Forests: A Look Ahead* (Comps. W.C. Schmidt and K.J. McDonald). pp. 33-37. USDA, Forest Serv., Intermountain Research Sta., General Technical Report, GTR-INT-319.

Shen, T.Y. 2000. *Forest Tree Seed Quality Determination based on Enzyme Activities*. PhD Dissertation, Swedish University of Agricultural Sciences, Department of Silviculture, Umeå, Sweden. 36 pp.

Shen, T.Y. and Odén, P.C. 2000. Fumarase activity as a quicker vigour test for Scots pine (*Pinus sylvestris* L.) seeds. *Seed Science and Technology* 28: 825-835.

Shen, T.Y. and Odén, P.C. 2002. Relationship between seed vigour and fumarase activity in *Picea abies*, *Pinus contorta*, *Betula pendula* and *Fagus sylvatica*. *Seed Science and Technology* 30: 177- 186.

Shenk J.S., Workman J.J. and Westerhaus M.O. 2001. Application of NIR spectroscopy to agricultural products. In *Handbook of Near-Infrared Spectroscopy* (eds. D.A. Burns and E.W. Ciurczak). 2nd ed. pp 419-474. Marcel Dekker Inc, New York.

Siegenthaler, P.A. and Douet-Orhant, V. 1994. Relationship between the ATP content measured at three imbibition times and germination of onion seeds during storage at 3, 15 and 30ºC. *Journal of Experimental Biology* 45: 1365-1371.

Simak, M. 1973. Separation of forest seed through flotation. *International Symposium on Seed Processing*, Vol. I, Paper No. 16, Bergen, Norway.

Simak, M. 1981. Bortsortering av matat-dött frö ur ett fröparti. [Removal of filled-dead seeds from a seed bulk.]. *Sverige Skogsvårdsförbunds Tidskrift* 5: 31-36.

Simak, M. 1984. A method for the removal of filled-dead seeds from a sample of *Pinus contorta*. *Seed Science and Technology* 12: 767-775.

Singh, R.V. and Vozzo, J.A. 1994. Application of the Incubation, Drying and Separation method to *Pinus roxburghii* seeds. *General Technical report, SO-101, 7-11*. South forest experiment station, New Orleans, Louisiana.

Sivritepe, H.O. and Dourado, A.M. 1995. The effect of priming treatments on the viability and accumulation of chromosomal damage in aged pea seeds. *Annals of Botany* 75: 165-171.

Sjöström, M., Wold, S. and Söderström, B. 1986. PLS discriminant plots. In *Pattern Recognition in Practice II*. Elsevier Science Publisher B.V., Holland. pp. 461-470.

Smith, M.T. and Berjak, P. 1995. Deteriorative changes associated with the loss of viability of stored desiccation-tolerant and desiccation sensitive seeds. In *Seed Development and Germination* (eds. J. Kigel and G. Galili). pp. 701-746. Marcel Dekker, Inc, New York.

Stener, L.G. 1995. Larch seed orchards in Sweden. In *Larch Genetics and Breeding: Research Findings and Ecological-Silvicultural Demands* (O. Martinsson). pp. 185-188. Proc. of IUFRO working party S2.02-07, Remningstorp and Siljansfors, Sweden.

Stoehr, M.U., Orvar, B.L., Vo, T.M., Gawley, J.R., Webber, J.E. and Newton, C.H. 1998. Application of a chloroplast DNA marker in seed orchard management evaluations of Douglas-fir. *Canadian Journal of Forest Research* 28: 187-195.

Sweeney, J.D., El-Kassaby, A., Taylor, D.W., Edwards, D.G.W. and Miller, G.E. 1991. Applying the IDS method to remove seeds infested with the seed

chalcid, *Megastigmus spermotrophus* Wachtl, in Douglas fir, *Pseudostuga menziesii* (Mirb.) Franco. *New Forests* 5: 327-334.

Teketay, D. 1991. Shade trees of coffee in Harerghe, Eastern Ethiopia. *International Tree Crops Journal* 7: 17-27.

Teketay, D. 1996. Germination ecology of twelve indigenous and eight exotic multipurpose leguminous species from Ethiopia. *Forest Ecology and Management* 80: 209-223.

Teketay, D. and Tigabu, M. 1996. The effect of pre-sowing seed treatments, temperature and light on the germination of *Tamarindus indica* L., a multipurpose tree. *Journal of Tropical Forestry* 12: 73-79.

Thapliyal, R.C. and Connor, K.F. 1997. Effect of accelerated ageing on viability, Leachate exudation, and fatty acid content of *Dalbergia sissoo* Roxb. Seeds. *Seed Science and Technology* 25: 311-319.

Tigabu, M. and Odén, P.C. 2001. Effect of scarification, gibberellic acid and temperature on seed germination of two multipurpose *Albizia* species from Ethiopia. *Seed Science and Technology* 29: 11-20.

Tillman-Sutela, E. and Kauppi, A. 1995a. The significance of structure for imbibition of seeds of the Norway spruce, *Picea abies* (L.) Karst. *Trees* 9: 269-278.

Tillman-Sutela, E. and Kauppi, A. 1995b. The morphological background to imbibition in seeds of *Pinus sylvestris* L. of different provenances. *Trees* 9: 123-133.

Tillman-Sutela, E., Johansson, A. and Laakso, P. 1995. Triacylglycerols in the seeds of northern Scots pine, *Pinus sylvestris* L., and Norway spruce, *Picea abies* (L.) Karst. *Trees* 10: 40-45.

Tillman-Sutela, E., Kauppi, A. and Sahlén, K. 1998. Effect of disturbed photoperiod on the surface structures of ripening Scots pine (*Pinus sylvestris* L.) seeds. *Trees* 12: 499-506.

Usberti, R. and Valio, F.M. 1997. Osmoconditioning effects on germination of guinee grass (*Panicum maximum*) seeds. *Seed Science and Technology* 25: 303-310.

Valera, F.P. and Kageyama, P.Y. 1991. *Pinus patula*. Seed leaflet 8a. Danida Forest seed centre. Humlebaek, Denmark. 25p

Vanangamudi, K., Zope, J.S., Vozzo, J.A. and Elam, W.W. 1991. Incubation, drying and separation method to separate viable and non-viable, artificially aged seeds of *Pinus roxburghii* Sarg. In *Dormancy and Barriers to Germination* (ed. D.G. Edwards). pp. 141-148. Proceeding of International Symposium, IUFRO Project P.2.04-00, Seed Problems. Forestry Canada, Pacific Forestry Centre, Victoria, B.C.

Varghese, M., Nicodemus, A., Nagarajan, B., Siddapa, K.R.S., Bennet, S.R.R. and Subramanian, K. 2000. *Seedling Seed Orchards for Breeding Tropical Trees*. Scroll Press, Coimbatore, India. 126 pp.

Velasco, L., Fernández-Martinez, J. and De Haro, A. 1996. Screening Ethiopian mustard for erucic acid by near infrared reflectance spectroscopy. *Crop Science* 36: 1068-1071.

Velasco, L., Fernández-Martinez, J. and De Haro, A. 1997. Determination of the fatty acid composition of oil in intact seed mustard by Near-infrared

reflectance spectroscopy. *Journal of the American Oil Chemists' Society* 74: 1595-1602.

Velasco, L., Fernández-Martinez, J., Haro, A.D. and De Haro, A. 1998a. Application of near infrared spectroscopy to estimate the bulk density of Ethiopian mustard seeds. *Journal of the Science of Food and Agriculture* 77: 312-318.

Velasco, L., Schierholt, A. and Becker, H.C. 1998b. Performance of near infrared reflectance spectroscopy (NIRS) in routine analysis of C18 unsaturated fatty acids in intact rapeseed. *Fett/Lipid* 100: 44-48.

Velasco, L., Möllers, C. and Becker, H.C. 1999. Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseeds (*Brassica napus* L.) by near-infrared reflectance spectroscopy. *Euphytica* 106: 79-85.

Wagner, M.R., Atuahene, S.K.N. and Cobbinah, J.R. 1991. *Forest Entomology in West Tropical Africa: Forest Insects of Ghana.* Kluwer Academic Publishers, The Netherlands. 200 pp.

Walters, C. 1998. Understanding the mechanisms and kinetics of seed ageing. *Seed Science Research* 8: 223-244.

Wang, X.R. and Szmidt, A.E 2001. Molecular markers in population genetics of forest trees. *Scandinavian Journal of Forest Research* 16: 199-220

Wang, Q., Zhang, F. and Smith, D. 1996. Application of GA3 and kinetin to improve corn and soybean seedling emergence at low temperature. *Environmental and Experimental Botany* 36: 377-383.

Wang, D., Dowell, F.E, and Lacey, R.E. 1999. Single wheat kernel size effects on near-infrared reflectance spectra and colour classification. *Cereal Chemistry* 76: 34-37.

Wang, X.R., Lindgren, D., Szmidt, A.E. and Yazdani, R. 1991. Pollen migration into a seed orchard of *Pinus sylvestris* L. and the methods of its estimation using allozyme markers. *Scandinavian Journal of Forest Research* 6: 379-385.

Wei R.P., Lindgren, K. and Lindgren, D. 2001. Parental environmental effects on cold acclimation and height growth in lodgepole pine seedlings. *Silvae Genetica* 50: 252-257.

Williams, P.C. and Norris, K.H. 1987. *Near-Infrared Technology in the Agricultural and Food Industries*. American Association of Cereal Chemists, Inc., St. Paul, Minnesota, USA. 330 pp.

Winsa, H. and Bergsten, U. 1994. Direct seeding of *Pinus sylvestris* using microsite preparation and invigorated seed lots of different quality: 2-year results. *Canadian Journal of Forest Research* 24: 77-86.

Wold, S. 1976. Pattern recognition by means of disjoint principal components models. *Pattern Recognition* 8: 127-139.

Wold, S., Antii, H., Lindgren, F. and Öhman, J. 1998. Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory System* 44: 175-185.

Wold, S., Sjöström, M. and Eriksson, L. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory System* 58: 109-130.

Wollf, R.L., Comps, B. and Marpeau, A.M. 1997. Taxonomy of Pinus species based on the seed oil fatty acid compositions. *Trees* 12: 113-118.

Wolff, R.L., Lavialle, O., Pédrono, F. Pasquier, E., Deluc, L.G., Marpeau, A.M., and Aitzetmüller, K. 2001. Fatty Acid Composition of Pinaceae as taxonomic Markers. *Lipids* 36: 439-451.

Workman, J.J. 1999. Review of process and non-invasive near-infrared and infrared spectroscopy: 1993 – 1999. *Applied spectroscopy review* 34: 1-89.

Workman, J.J. and Burns, D.A. 2001. Commercial NIR Instrumentation. In *Handbook of Near-Infrared Spectroscopy* (eds. D.A. Burns and E.W. Ciurczak). pp. 53-70. Marcel Dekker Inc., New York.

Yazdani, R. and Lindgren, D. 1991. Variation of pollen contamination in a Scots pine seed orchard. *Silvae Genetica* 40: 243-246.

Zhang, M., Lui, Y., Torii, I., Sasaki, H. and Esashi, Y. 1993. Evolution of volatile compounds by seeds during storage periods. *Seed Science and Technology* 21: 359-373.

Zhang, M., Yajima, H., Umezawa, Y., Nakagawa, Y., and Esashi, Y. 1995. GC-MS identification of volatile compounds evolved by dry seeds in relation to storage conditions. *Seed Science and Technology* 23: 59-68.

Zobel, B., and Talbert, J. 1984. *Applied Forest Tree Improvement*. John Wiley and Sons, New York, USA. 505 pp.

Acknowledgements

I am very much grateful to my supervisor Professor Per Christer Odén for accepting me as a PhD student in Forest Seed Science and teaching me critical thinking and scientific reasoning. Had it not been for his generous support and consistent encouragement through out the course of the study period both at professional and personal levels, it would have been impossible to write this thesis. His assistance in planning experiments, meticulously and swiftly reviewing several drafts of manuscripts, as well as, openly sharing his wealth of knowledge in seed biology and technology are all instrumental for the successful compilation of this thesis.

Dr. Demel Teketay, my assistant supervisor, has aroused my interest in Forest Seed Science, particularly in seed ecology, since 1992 when I was doing my senior research project at Alemaya University of Agriculture. Since then he has been a continuous support for my academic progress in various ways. His splendid support during seed collection, practical arrangements and unfailing encouragement during the course of the PhD study contributed much to reach this stage.

I would like to take this opportunity to thank Prof. Dag Lindgren for marvellous co-authorship, Dr. Owe Martinsson and Mr. Rudolf Kollenmark for providing seeds of *Larix* species, Dr. Ola Rosvall and Mr. Erik Walfridsson at the Forest Research Institute in Sävar and Mr. Olle Rosenberg in Uppsala, Sweden for valuable discussions, for providing us with *Picea abies* seeds and for the assistance in the identification of insects. Dr. Tong Yun Shen deserves special thanks not only for a splendid co-authorship but also for enthusiastic discussion on several matters and as a kind and co-operative colleague. I owe Ali Soltani (my Iranian friend) a debt of gratitude for his incredible assistance in computer related matters with patience and enthusiasm. I would also like to thank Dillon Chrimes, Daniel Tiveau, and Dr. Robert Nygård for reading manuscripts and Hans Sjögren for support with personal matter and developing collaborative research. I am sincerely indebted to Dr. Louis Sawadogo for welcoming me to Burkina Faso, for productive research collaborations and unforgettable dinner in his place in Koudougou. I am also very much grateful to all colleagues in our research group: the late Leulseged Demelash, Maria Schelin, Didier Zida, Sandra Razanamandranto, Catherine Dembele and Joakim Fjellström for interesting discussions in the field of forest reproduction and seed handling. My deepest gratitude should be extended to Margareta Söderström for her skilful assistance in the laboratory as well as Rune Johansson. Not least, I am very much indebted to the department of Silviculture for providing me a conducive working environment; special thanks are due to Ann-Kathrin Persson, Inga-Lis Johansson and Karin Folmerz for their practical and administrative assistances.

The Research Group for Chemometrics, Department of Chemistry, Umeå University is highly acknowledged for allowing me to use their NIRSystems Model 6500 spectrophotometer. I am sincerely grateful to Prof. Michael Sjöström for

reviewing almost all manuscripts despite his tight schedules. I am also indebted to Dr. Floyd E. Dowell at the U.S. Department of Agriculture, Grain Marketing and Production Research Centre, Kansas, USA for critically reviewing a manuscript and for valuable suggestions.

I thank my brother, Amsalu Tigabu, and sisters Fasika, Aynadis and Hanna Tigabu, for their morale support and encouragement. Many colleagues, both in Ethiopia and in Sweden, have been a constant support over the last four years. In Ethiopia, Dr. Girma Balcha, Dr. Yonah Yemshaw, Dr. Zebene Asfaw, as well as staff of the Forestry Research Centre, the National Tree Seed Centre and Wondo Genet College of Forestry are acknowledged. My best friends Fantaw Yimer and Jemal Ibrahim as well as the late Yemisirach Zenebe deserve special thanks for nurturing my sisters in my absence and for spiritual support; and Dr. Tarekegn Abebe for sharing me his wealth of wisdom in social and intellectual matters. I also enjoyed the fabulous company of Dr. Tarekegn during our stay in Sweden with whom I had lots of fun. Dr. Getachew Eshete and Dr. Zewdu Eshetu are acknowledged for various fruitful discussions during their time in Sweden. In Sweden, Alemayehu Mulugeta, Genet Abrham, Ayalew Mulugeta, Mahlet Wada, Amsale Assefa, Eden Fissihaye, Nasir, Getachew Wanna, Abinet Tsegay, Tadelech Wanna, Matios, Mesay, Addis, Abebe, Yohannes, Tenagne, Belay Kassa, Serkalem, Atsede and many others made my stay hospitable for which I convey my heartfelt gratitude. My love Emebet Wada deserves special thanks for standing by my side all the time and for her strong persevering support. Alex and Emu, you are simply great!! Thanks for everything that I cannot explain in words due to space limitation.

The Swedish Council for Forestry and Agricultural Research, and the J.C. and Seth M. Kempe foundations are gratefully acknowledged for financial support.