REFLECTANCE CHARACTERISTICS OF BULK GRAINS USING A SPECTROPHOTOMETER

A Thesis
Submitted to the Faculty of Graduate Studies
The University of Manitoba
in partial fulfilment of the requirements for the degree of

Master of Science

by

Ming Tee Eu

Department of Biosystems Engineering
University of Manitoba
Winnipeg, Canada
R3T 5V6

0-612-23296-4

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

REFLECTANCE CHARACTERISTICS

OF BULK GRAINS

USING A SPECTOROPHOTOMETER

BY

MING TEE EU

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University

of Manitoba in partial fulfillment of the requirements of the degree

of

MASTER OF SCIENCE

Dedicated to my loving wife Corinna

ABSTRACT

The automated cleaning, grading, and monitoring of grain throughout the grain handling system would maintain, if not improve, Canada's ability to be successful in the global grain market. A machine vision system is currently being developed for use with such systems in the Department of Biosystems Engineering, University of Manitoba. One measurement characteristic that is relatively easy to use is the reflectance characteristic of grains. Reflectance characteristics of 8 cereals, 3 oilseeds, 8 pulse seeds, and 27 specialty seeds were measured using a spectrophotometer (Model: Cary 5, Varian Canada Inc., Mississauga, ON). Using Canada Western Red Spring (CWRS) wheat samples, the effects on reflectance characteristics of growing region, moisture content, grade, and amount of foreign material were quantified. To assess the capability of reflectance features for grain classification, thirteen features were extracted from the reflectance data based on slope-ratio, ratio, and normalized area. Discriminant analysis using the hold-out method was used to determine the classification accuracies. Procedure STEPDISC was used to determine the contribution of each feature to the model. Reflectance characteristics successfully classified (100% accuracy) the oilseeds, seven of the eight classes of cereals, five of the eight classes of pulses, and twenty of the twenty-seven classes of specialty seeds. Ratio features contributed more to the classification accuracies than did the slope-ratios or the area under the reflectance curve features. Based on the intuitive selection of features, the wavelengths that best classified the bulk grain samples were 800, 1050, and 1250 nm. Classification accuracies for cereals and pulses were higher when normal estimation was used. Reflectance characteristics did not successfully classify the grading characteristics of CWRS wheat.

ACKNOWLEDGEMENT

I thank Dr. D.S. Jayas for his guidance, support, and the opportunity to work with him.

I thank Dr. N.R.Bulley and Dr. S.J. Symons for serving on my thesis committee.

I thank Prince Rupert Grain Ltd., Natural Sciences and Engineering Research Council of Canada, and Agriculture and Agri Food Canada for partial funding of this study.

Many thanks to Brett Young Seeds for providing the specialty seeds and to the Canadian Grain Commission for providing the cereals, oilseeds, and pulse seeds.

I thank Jack Putnam, Matt MacDonald, and Dale Bourns for their technical assistance.

I dedicate this thesis to my wife for her constant love and encouragement, and my parents for their loving support throughout my studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Canada is recognized as a world leader in supplying quality grains, especially wheat (Canada Grains Council 1982). This reputation is maintained through application of continuing research and technology to its grading system to meet the expectations of its end-users. Current research indicates that a machine vision system could accurately and quickly identify the characteristics of a grain sample (Majumdar et al. 1996a, 1996b; Shatadal et al. 1995a). A machine vision system could increase automation and objective measurement and is expected to improve the efficiency of handling, storing, and shipping grain, and to provide a more competitive marketing tool.

Machine vision refers to the automated process of using a physical image sensor (instead of the human eye) and dedicated computing hardware (instead of the human brain). The sensor "sees" specific characteristics, such as morphology, color, texture, and reflectance, and relays the images to the "brain" which analyzes the image based on a predefined goal ("memory"). The machine vision system is capable of fast, precise, and accurate measurement. One characteristic that is relatively easy to apply in a field situation is reflectance characteristics. Reflectance characteristics in the near-infrared spectrum have been used since 1976 to segregate wheat based on protein level (Panford 1987). Research into the classification of seed types and quality measurements has produced much information but most remains suitable only for the laboratory situation, not for a commercial application. The use of reflectance characteristics in the near ultraviolet to near infrared spectrum to identify bulk samples of grains has not been extensively studied.

The present work is part of a project to design and assemble a prototype machine vision system to classify bulk grains at grain handling facilities. The objectives of this thesis research were:

1. to determine the reflectance characteristics of various bulk samples of seeds in the 350 to 1850 nm spectrum using a spectrophotometer;

2. to determine the effects of growing region, moisture content, grade, and amount of foreign material on the reflectance characteristics of Canada Western Red Spring (CWRS) wheat; and

3. to investigate the potential of the reflectance data for the classification of various seed types.

CHAPTER 2: LITERATURE REVIEW

2.1 Canadian Grading System

Canada produced about 57 Mt of grains which consisted of wheat (54%), barley (28%), canola (9%), oats (8%), flaxseed (1%), and rye (0.5%) for the 1995/1996 year (Statistics Canada 1997). Exports of these grains were 22.9 Mt. Sixteen million metric tonnes of wheat were exported, or 53% of all wheat produced in Canada. The primary export market was the United States which imported 59% of Canada's export wheat. Canada's market share of the world wheat market was 15%.

The grading system is a means of maximizing net returns for the producer from the market place (Canada Grains Council 1982). This objective can be met through efficient handling, storage, and shipment of grain, as well as matching the buyer's demands for a particular set of characteristics to a specific quality level. The grade determines the price that the producer receives and serves as a marketing tool for grain marketers. The grading system is most developed for wheat reflecting its large production and export volumes, and its multi-uses.

Wheat is used for many different purposes, including bread, pasta, noodles, porridge, prepared breakfast foods, and animal feed preparations (Canada Grains Council 1982). Wheat used to make bread represents the greatest proportion of wheat traded on the world market. Each type of bread has its own characteristics which are derived from the composition of the flour and ultimately the wheat kernel itself.

The grade for wheat is determined through application of grading standards by trained grain inspectors at transfer and terminal elevators. The grading standards are based on official grade definitions published by the Canadian Grain Commission (1987) and by reference to standard samples prepared every year by the Canadian Grain Commission. This method of grading uses both objective and subjective measurements. Objective measurements include test weight, moisture content, and protein content. However they are not always a direct measurement of the end-use value characteristic. For example, test weight only indirectly measures flour yield. Subjective measurements include variety, vitreousness (an objective method is the particle-size index but it is time-consuming), and soundness. For example, soundness is an indicator of various forms of damage such as sprouting (which can be objectively measured with the alpha amylase test to give a falling number), insect damage, and damaged kernels. Because many measurements for grade determination are subjective, disputes over grade and errors of grade occur. Errors result from eye fatigue, color and shape memory, personal bias, personal experience, and lighting conditions (Howarth et al. 1990). Bevilacqua (1987) found that 10-14% of the reinspected grain samples (about 1% of the total samples) were given a different grade from that originally assigned.

2.2 Machine Vision System

The grain industry is interested in machine vision technology to use for an on-line monitoring system (Shatadal et al. 1995b). The advantages of automation include objective measurement; continual evaluation and monitoring; and efficient, fast, and accurate results.

The results can be used to ensure that sorting and grading decisions are objective and consistent, that optimum cleaning strategies and appropriate bin decisions are made, and that complete automation of cleaning equipment and railcar unloading is achieved.

Machine vision is the integration of computers and cameras into a system designed to replace and enhance human visual judgement of agricultural products. The system has two main components; a camera and a computer. The camera can be equipped to 'see' various characteristics of the sample, such as morphology, color, texture, and reflectance. The image picked up by the camera is fed to the computer for analysis. The computer then determines the composition and physical characteristics of the sample using, for example, digital image analysis. The results of the analysis can then be used to direct the sorting and grading equipment.

Machine judgement is expected to produce better graded products and to improve productivity by being able to work consistently, continuously and with a constant level of accuracy (Howarth et al. 1990). Although the human eye is much quicker than a machine in judging the same characteristics, human judgement tends to be fallible and subject to fatigue especially for repetitive functions such as sorting and grading.

Machine vision for inspection and grading has been applied to apples, bell peppers, carrots, peaches, potatoes, tomatoes, and soybeans (Tao et al. 1995). Color has been the primary focus for sorting and grading of fruits and vegetables because of its significance to consumer acceptance and point-of-sale value. Color is also an indicator of quality as it reflects maturity and the presence of blemishes. Commercial systems are available for sizing, color sorting, and grading, but most tend to be expensive (Heinemann et al. 1994).

They are not widely used because of real-time constraints and problems with fruit handling, orientation, illumination, defect identification, and color resolution. (Singh and Delwiche 1994).

Machine vision for classification and grading of grains is being developed (Shatadal et al. 1995b; Majumdar et al. 1996a). The primary focus has been on morphological characteristics of the sample (Chen et al. 1989; Neuman et al. 1987; Sapirstein et al. 1987; Shatadal et al. 1995b; Symons and Fulcher 1988a, 1988b; Thomson and Pomeranz 1991; Zayas et al. 1986). Morphological differences of intact kernels have most commonly been analyzed with digital imaging (Chen et al. 1989; Majumdar et al. 1996a; Neuman et al. 1987; Sapirstein and Kohler 1995; Shatadal et al. 1995a; Symons and Fulcher 1988a, 1988b; Thomson and Pomeranz 1991; Zayas et al. 1985, 1986). The potential for separation of similar seed morphologies has been demonstrated by Churchill et al. (1993), Delwiche and Norris (1993), Shatadal et al. (1995a), and Slaughter et al. (1992). Few researchers have used color (Neuman et al. 1989a, 1989b; Winter et al. 1996) or textural (Majumdar et al. 1996b) characteristics. Reflectance characteristics have been little utilized until recently (Casady et al. 1993; Hawk et al. 1970; Howarth et al. 1990; Majumdar et al. 1996a). The current interest in reflectance spectroscopy technology reflects its advantages: it is rapid, does not require much sample preparation, can be used in field measurements, is well-accepted throughout the grain industry, and is advantageous over digital imaging with respect to equipment cost and computational processing time. Near-infrared and infrared reflectance spectroscopy has shown potential in measuring important grading characteristics (Delwiche 1995; Law and Tkachuk 1977; Stermer et al. 1977; Williams 1979).

2.3 Reflectance Characteristics

Reflectance characteristics have been used to analyze the mineral composition of rocks, the chemical composition of food, and the structural integrity of soils. Infrared reflectance spectroscopy for bulk grains was first used extensively in the analysis of the oil, protein, and moisture content of soybeans (Williams 1975). Near-infrared reflectance spectroscopy is used for bakery flour quality control by Quality Bakers of America Laboratory and in continuous on-line monitoring of flour mill streams (Panford 1987). The Kellogg Corporation uses near-infrared technique to monitor incoming raw material and to determine moisture, protein, and sugar content of various food products. Near-infrared reflectance spectroscopy is also used in the remote sensing of vegetation (Brown et al. 1994). Areas of potential application include barley genotype breeding programs; amino acid composition in wheat and barley breeding programs; malt extract measurement of barley plants; chemical analysis of corn, rice, forages, oilseeds, tobacco, pulses, meat and meat products, dairy products, and processed cereal foods; and determination of stage of maturity of various fruits and vegetables.

Near-infrared technology has the potential to furnish a series of objective tests for the determination of end-use processing quality of cereals, pulses, and oilseeds. Reflectance properties of whole grain instead of ground grain are of much interest because they offer the possibility of providing the most convenient testing procedure (Delwiche et al. 1995). Such a procedure has ultimate benefits in terms of convenience, speed, and cost.

Reflectance characteristics are used to determine the spectrum range or wavelength(s) that can be used to help classify a sample. The wavelength(s) that can be used to distinguish two grain samples is that area of the graph where the reflectance characteristics for those grains are the most divergent. Alternatively, the percent reflectance ratios or the ratios of slopes of percent reflectance curves at a specified wavelength(s) can also be used. Reflectance characteristics are measured using a spectrophotometer.

2.4 Spectrophotometer

A spectrophotometer measures the flux of the reflected light from a sample for a given wavelength. The reflectance characteristic is the ratio of the flux reflected by the sample to that reflected by the standard surface. Light that strikes a sample can be dispersed as specular and diffuse reflectance, and can be transmitted and absorbed. Specular reflectance is what prevents you from seeing inside a house from the outside on a sunny day. If you shade your eyes so that direct light is no longer reflected into them, you can see the objects in the house by diffuse reflectance. Transmitted light (transmittance) is the portion of light that is transmitted through the sample and absorbed light (absorbance) is the portion of light that is absorbed by the sample.

The spectrophotometer has a diffuse reflectance accessory (DRA) equipped with an integrating sphere which collects the diffusely reflected light. The optical design of the DRA is illustrated in Fig. 2.1. The sample beam (light source) enters the DRA and is focused onto the sample via mirrors (M_1, M_2), an offset lens, the sample beam entrance, and sample measurement port. At the sample measurement port, the beam reflects off the sample in the sample holder and is diffused throughout the sphere. At the same time, the reference beam (standard) enters directly into the sphere from the reference port and is also diffused

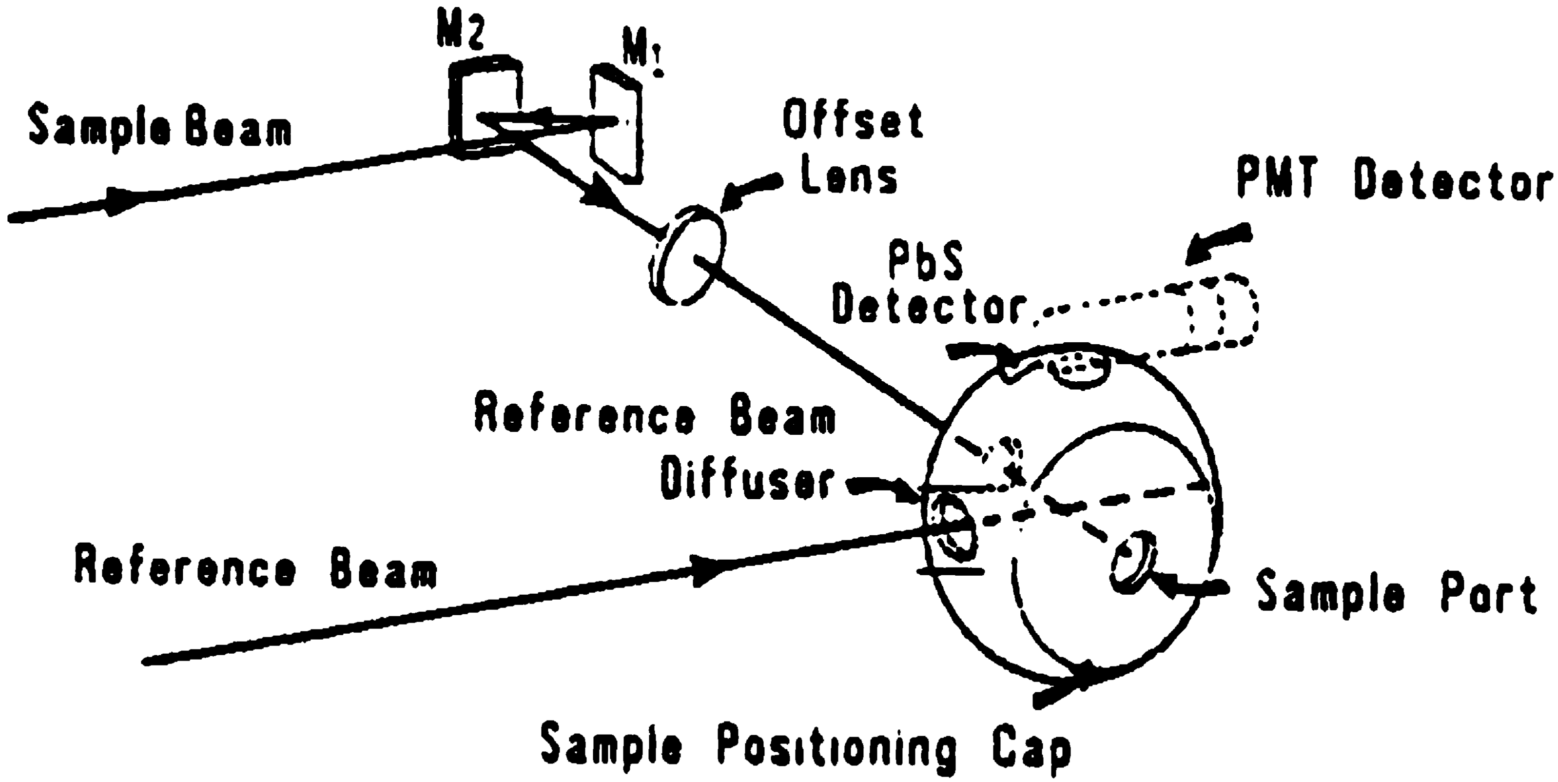

Fig.2.1 The optical design of the diffuse reflectance accessory of the spectrophotometer

throughout the sphere. The reference beam is obtained by collecting a baseline after all the scan parameters are set using a calibration reference disk coated with polyperfluoroethylene. The diffusely reflected light of the sample and reference beams are collected by the integrating sphere and presented to the detectors as an integrated signal. The wavelength that is picked up is determined by two different detectors: the ultraviolet and visible regions are detected by a photomultiplier tube while a lead sulfide photocell detects the near-infrared region. In contrast, the human eye is able to distinguish differences only in the visible range.

The accuracy and precision of the spectrophotometer can be affected due to: (i) deterioration of the coating on the reference disk from contact with smoke, plastic materials, or other contaminants; (ii) incorrect sample placement causing loss of reflected light; and (iii) the escape of diffused light through the ports which reduces the signal to noise ratio.

2.5 Reflectance Characteristics Research

Massie and Norris (1965) developed instrumentation to obtain reflectance curves for corn, oats, wheat, soybeans, rice, alfalfa seed, and milled rice in the 400 to 2000 nm range. Their use of the reflectance data was in the design of infrared grain dryers which required knowledge of the spectral reflectance and transmittance properties of the grain. They were interested in those areas of the curve which showed low reflectance as this meant that the energy was being used to dry the grain and not simply reflected back. They found that the moisture content did not have any significant effect on reflectance and transmittance properties of the grains and that reflectance was more important than transmittance for infrared drying.

The visible spectrum appears to be the most useful in sorting and grading of produce and grains. Reflectance values in the 400 to 700 nm region were the most important in characterizing appearance in apples, peaches, and pears (Bittner and Norris 1968). Bilanski et al. (1984) determined that spectral reflectance scans in the 390 to 398 nm wavelength were most effective in distinguishing bruised from good tissue of peeled apples. Bruised and good whole Red Delicious apples were best distinguished in the 720 to 840 nm range (Upchurch et al. 1990). Spectral sensitivities of peach surface defects centered around 650, 720, and 815 nm (Miller and Delwiche 1991). Various grains were best separated using the visible region (Hawk et al. 1970), as were damaged from undamaged soybean seeds (Casady et al. 1993), normal from damaged peanut kernels (Dowell 1992), and three of four defects in carrots (Howarth et al. 1990).

Initial interest in reflectance characteristics centered on measuring color, partly because of limited technology. Vegetables and fruits and single kernels of grains were therefore analyzed. As the technology improved, the range of application increased including classifying bulk samples of individual grains, and admixtures of grains. The following studies, presented in chronological order, represent some of the research work in the field of reflectance characteristics as part of an automated machine vision system.

Bittner and Norris (1968) used a spectroreflectometer to characterize reflectance properties of apples, peaches, and pears as they grew and matured. A wide wavelength region (250 to 2100 nm) was chosen to provide a better understanding of the optical properties of fruit. Major changes occurred in the visible (400 to 700 nm) range. Pears and Golden Delicious apples could be evaluated at one wavelength in the 550 to 620 nm region to predict stage of maturation. Peaches and red varieties of apples required a reflectance ratio measurement using one wavelength in the 550 to 590 nm range and the other in the 600 to 640 nm range as the simplest method of predicting stage of maturation.

One of the first studies to determine the reflectance characteristics of grains was conducted by Hawk et al. (1970). A spectroreflectometer was used to analyze twelve types or classes of grains, including wheat, oats, rye, barley, sorghum, corn, soybeans, and flaxseed. Differences in the ultraviolet region were too small to warrant further testing. Analysis of the slight differences in the near infrared region showed that different grains could not be classified. The greatest differences in reflectance data occurred between 450 and 700 nm. Reflectance data indicated hard red spring wheat could not be distinguished from hard red winter wheat or barley from oats. They concluded that it may be feasible to identify the primary grain in a sample from other grains that may be present using some combination of two wavelengths but additional investigation is required.

Bilanski et al. (1984) used a spectrophotometer with a wavelength range of 350 to 700 nm to examine four different quality indices: average reflectance, reflectance at one narrow bandwidth wavelength, reflectance at two wavelengths, and derivative values of reflectance at two wavelengths. Reflectance scans were obtained on five varieties of peeled apples with the objective of being able to predict bruise depth and to differentiate between good and bruised tissue. Bruise depth is an important component in the efficient automated removal of bruised apple tissue because bruised apples are unacceptable in processed products. Reflectance properties were not able to reliably predict bruise depth. Differentiation between good and bruised apple tissue was not effective using average reflectance but was effective for the three other quality indices. The least accurate index was the reflectance at one narrow bandwidth wavelength. For four of the apple varieties, that wavelength was between 390 and 398 nm; and for the fifth variety it was 458 nm. The most accurate index was the derivative values of reflectance at two wavelengths. Index of reflectance values at two wavelengths was intermediate in accuracy. Correct identification of bruised tissue ranged from 93 to 99% depending on variety and quality index.

Upchurch et al. (1990) used a spectrophotometer in the 400-1000 nm wavelength region to investigate the applicability of using reflectance properties to distinguish bruised from nonbruised fruit for unpeeled Red Delicious apples. Five models were developed from the reflectance spectra data: single wavelength; two wavelength difference; two wavelength ratio; normalized two wavelength difference; and a derivative model. The wavelength region of 720 to 840 nm produced the best models for discrimination of bruised from nonbruised apples. The single wavelength and two wavelength difference models misclassified 50% and 49.5% of the samples, respectively. The ratio, normalized difference, and derivative models were more successful, misclassifying only 2.5%, 2.5%, and 3.5% of the samples, respectively.

Howarth et al. (1990) used a spectrophotometer with wavelengths from 486 to 1043 nm to obtain reflectance data for carrots as part of a project using a machine vision inspection system to convey, detect, and sort carrots. The carrots were normal or had one of four defects: dry rot, soft rot, black crown, or cavity spots. Except for cavity spots, the defects were adequately identified and a band pass filter centered at 630 nm was most appropriate.

Miller and Delwiche (1991) measured the spectral reflectance of six peach surface defects (bruise, cut, scar, brown rot, and wormhole) and normal tissue in the 350 to 1200 nm range. The objective was to determine which of the seven indices was the best in discriminating between defective and normal peach surface based on Mahalanobis distance. The indices were: reflectance at a single wavelength; ratio of reflectance of one wavelength to another wavelength; ratio of one to two wavelengths; ratio of one to three wavelengths; a combination of reflectance at a single wavelength and ratio of one to two wavelengths; a combination of two ratios of one to three wavelengths; or a combination of reflectance at a single wavelength and two ratios of one to three wavelengths. An index with a single wavelength component (reflectance at a single wavelength; reflectance at a single wavelength and a ratio of one to two wavelengths; reflectance at a single wavelength and two ratios of one to three wavelengths) was better than an index without a single wavelength component over the range of peach defects. The highest Mahalanobis distances were obtained with the latter index using three wavelengths at 650, 720, and 815 nm.

Dowell (1992) investigated four optical sensing methods for the identification of damaged (freeze, questionable, damaged) and undamaged (questionable, undamaged) peanut kernels. The methods compared were monochromatic machine vision system, contact colorimeter, noncontact colorimeter, and spectrophotometer. The classification accuracies for the four methods, respectively, were 62.7, 79.1, 85.8, and 94.1% of the damaged kernels and 100% of the undamaged kernels. The best method was the spectrophotometer with 0% (freeze damaged) to 12% (questionably damaged) misclassification of damaged kernels. When the wavelength ranges and slopes were optimized only 1% of damaged kernels were misclassified. Misclassification of kernels for the other methods were the least for the freeze damaged (0, 6, and 4% respectively) and the greatest for the questionably undamaged (68, 32, and 21% respectively).

Casady et al. (1993) used a spectroradiometer in the 300 to 850 nm spectrum for color imaging of soybeans in identifying the amount of fungi damage. Fungi damage causes the soybean to change color and is a good indicator of type of fungi and extent of damage. The wavelength spectrum of 436 to 724 nm was the most useful but no single wavelength or combination (using two different bandpass filters) could be used to linearly separate all classes of damaged soybean seeds.

Heinemann et al. (1994) evaluated an image-analysis machine-vision system in the visible spectrum for the identification of four quality features of mushrooms (shape, color, stem cut, and veil opening). Misclassification ranged from 8 to 56% depending on the quality feature being evaluated. Disagreement between human inspectors ranged from 14 to 36% depending on the quality feature being evaluated.

Singh and Delwiche (1994) used an image-analysis machine-vision system in the near infrared region to evaluate five peach defects. The classification error was 28.6% and due primarily to natural variability in features. The correlation coefficient between machine and manual systems was 0.75 for bruise defect and 0.72 for scar defect, due primarily to surface curvature and manual measurement errors.

Saputra et al. (1995) used near infrared reflectance in the 1400 to 1975 nm to determine if the dominant sugar and acid content of Gedong mangoes could predict their taste (sweet, sweet-sour, sour). The traditional method of classifying mango by taste has been by aroma and experience. A HPLC analysis identified the dominant sugar as sucrose and the dominant acid as malic acid. The results of a discriminant analysis showed that sucrose and malic acid can be used to predict taste. The prediction by near-infrared reflectance for sucrose and malic acid concentration was calibrated by HPLC analysis. Malic acid concentration and sucrose concentration were found to be correlated with days after harvest. A higher malic acid concentration and lower sucrose concentration was found 90 days after harvest and a lower malic acid concentration and higher sucrose concentration was found 100 days after harvest. This finding was correlated with a sourer tasting fruit at 90 days and a sweeter tasting fruit after 100 days. The stepwise method selected nine optimal wavelengths for prediction. The wavelengths selected to predict sucrose concentration 100 days after harvest, for example, were 1461, 1550, 1589, 1685, 1717, 1725, 1789, 1869, and 1909 nm. The correlation coefficients 90 and 100 days after harvest for malic acid were 0.95 and 0.98, respectively, and 0.96 and 0.98 for sucrose concentration.

Tabil et al. (1996) applied near-infrared reflectance to alfalfa cubes. Near-infrared reflectance has the potential to be a quicker and more reliable method than the unofficial grading method of manually breaking the cubes to observe the leaf/stem content. The near-infrared product analyzer was calibrated using known leaf fractions to estimate leaf fractions of cubes. The three grades of commercial alfalfa cubes were successfully classified based on leaf fraction.

Majumdar et al. (1996a) used a spectrocolorimeter to obtain data in the 400 to 700 nm wavelength region. Tristimulus values, color solid scale values, percent reflectance ratios, and ratios of slopes of percent reflectance values were used to test the feasibility of identifying bulk grain samples. They concluded that identification was effective in most cases but not all regardless of parameter used. However, discriminant analysis using above parameters collectively gave significant improvement in classification accuracy.

The literature review indicates that reflectance characteristics have been researched widely for applications in the agriculture industry for increased automation and quality control. The visible spectrum appears to be the most effective in classifying various grains, vegetables, and fruits. Application to fruits and vegetables has focused primarily on color but as the technology has become more advanced and less expensive, other features are being examined. This technology is of interest to the grain industry for automation of some of its functions. Applications are still in the research stage but as more data are gathered and field trials are conducted, reflectance characteristics as part of an automated sorting and grading system may someday be implemented in the grain industry.

CHAPTER 3: MATERIALS AND METHODS

3.1 Spectrophotometer

The hardware for the acquisition of reflectance data consisted of a spectrophotometer (Model: Cary 5, Varian Canada Inc., Mississauga, ON) equipped with an integrating sphere for reflectance measurements and a photomultiplier tube that permitted measurement in the visible and ultraviolet regions while a lead-sulfide detector took measurements in the near-infrared region. This change of detectors will introduce a shift at 850 nm on the reflectance characteristics curve (see Appendix A). A 486 Dx 33 MHZ personal computer was loaded with window-based software (Varian) to interface with the spectrophotometer. The reference material used to set the baseline and to calibrate the instrument was polyperfluoroethylene. Total reflectance (specular and diffuse) was measured using the spectrophotometer. Transmittance and absorbance were assumed negligible.

3.2 Bulk Grain Samples

Bulk grain samples of various seed types were collected from different growing regions of Western Canada in the 1995 growing season. Eight types of cereals, 3 types of oilseeds, and 8 types of pulses were obtained from the Industry Services Division of the Canadian Grain Commission, Winnipeg, MB and 27 types of specialty seeds were obtained from Brett Young Seeds, Winnipeg, MB. The cereal grains were: Canada Western Red Spring (CWRS) wheat, 2-row barley, 6-row barley, oats, Canada Western Amber Durum(CWAD) wheat, rye, buckwheat, and soft white spring wheat (SWSW). The oilseeds tested were: sunflowers, canola, and flaxseed. The pulse seeds used were: lentils, pea beans, peas, pinto beans, green beans, black beans, dark green speckled lentils, and faba beans. The specialty seeds were: alfalfa, birdsfoot trefoil, bromegrass, meadow brome, canary seed, alsike clover, red clover, sweet clover, creeping red fescue, reed canary grass, orchard grass, annual ryegrass, perennial ryegrass, creeping bentgrass, meadow fescue, Kentucky bluegrass, crown millet, brown mustard, oriental mustard, yellow mustard, sorghum sudangrass, timothy, crested wheatgrass, intermediate wheatgrass, tall wheatgrass, dahurian wild rye, and Russian wild rye. The samples were kept in a deep freezer at -18°C.

Canada Western Red Spring wheat was collected from 20 growing regions representing various climatic subdivisions of the Canadian prairies (Luo 1997). The wheat samples from one of the regions (Ab4s) were conditioned to 10, 12, 14, 16, and 18% nominal moisture content. The samples were conditioned by measuring the initial moisture content of the grain and by adding the appropriate amount of water to achieve a given moisture content. The moisture contents were verified after conditioning as 10.09, 12.03, 13.80, 15.85, and 17.36%, respectively.

The wheat samples from one of the regions (Ab4s) were conditioned to have 3, 6, 9, 12, and 15% foreign material. The initial samples were considered pure. Samples were prepared on a mass basis. For example, 3 g of foreign material were added to 97 g of pure sample to give 3% foreign material mixture. The foreign material was collected from the Canadian Grain Commission, Winnipeg, MB.

To examine the potential of correctly classifying different grades of CWRS wheat, samples of grades 1, 2, and 3 were obtained from the Canadian Grain Commission, Winnipeg, MB.

3.3 Sampling and Analysis Technique

Samples of about 0.5 kg of various grains were obtained. A randomly selected subsample of 4 to 5 g was loaded onto the sample holder which was attached to the opening of the integrating sphere of the spectrophotometer. The spectrophotometer took three measurements at each 0.5 nm interval from 350 to 1850 nm and averaged the values. The data were then plotted as one point for each wavelength. The variation in reflectance data for repeated measurement on an undisturbed sample was within ± 0.3 percentage points (figure is not presented). This procedure was replicated five times by placing the previous sample back in the bag and taking a new random sample from the same bag. A total of 79 samples, each replicated five times, were analyzed. A typical variation among five replicates of the same sample is shown in Fig. A1 (Appendix A) for CWRS wheat.

The percent reflectance data are difficult to use in an industrial application because they are affected by intensity of light, image background, dust, and aging of the light source (Majumdar et al. 1996a). A more robust method is to use the ratios of slopes of percent reflectance values and the ratios of percent reflectance values at different wavelengths. The specific wavelengths were intuitively chosen by observing the reflectance data of different classes of cereal grains and choosing wavelengths that gave the maximum variation among classes. These wavelengths were confirmed by randomly selecting reflectance data from the CWRS wheat regions data and plotting them on the same graph.

Thirteen features were extracted for each sample using reflectance data between 750 and 1600 nm inclusive. This near infrared region was chosen because the mean reflectance data showed the greatest variation between reflectance curves. This is in contrast to Hawk et al. (1970) finding of greater variation in the visible spectrum (450 to 700 nm). Their result may have been a function of the equipment and of technology as instruments for measuring reflectance characteristics have become more advanced. Features 1-6 were defined as the slope-ratios and features 7-12 were defined as ratios:

$$Feature\ 1:\ slope\text{-}ratio = \frac{\%\ reflectance\ at\ 1500\ nm\ -\ \%\ reflectance\ at\ 1600\ nm}{\%\ reflectance\ at\ 1200\ nm\ -\ \%\ reflectance\ at\ 1300\ nm}$$

$$Feature\ 2:\ slope\text{-}ratio = \frac{\%\ reflectance\ at\ 1000\ nm\ -\ \%\ reflectance\ at\ 1100\ nm}{\%\ reflectance\ at\ 750\ nm\ -\ \%\ reflectance\ at\ 850\ nm}$$

$$Feature\ 3:\ slope\text{-}ratio = \frac{\%\ reflectance\ at\ 1500\ nm\ -\ \%\ reflectance\ at\ 1600\ nm}{\%\ reflectance\ at\ 1000\ nm\ -\ \%\ reflectance\ at\ 1100\ nm}$$

$$Feature\ 4:\ slope\text{-}ratio = \frac{\%\ reflectance\ at\ 1200\ nm\ -\ \%\ reflectance\ at\ 1300\ nm}{\%\ reflectance\ at\ 750\ nm\ -\ \%\ reflectance\ at\ 850\ nm}$$

$$Feature\ 5:\ slope\text{-}ratio = \frac{\%\ reflectance\ at\ 1500\ nm\ -\ \%\ reflectance\ at\ 1600\ nm}{\%\ reflectance\ at\ 750\ nm\ -\ \%\ reflectance\ at\ 850\ nm}$$

$$Feature\ 6:\ slope\text{-}ratio = \frac{\%\ reflectance\ at\ 1200\ nm\ -\ \%\ reflectance\ at\ 1300\ nm}{\%\ reflectance\ at\ 1000\ nm\ -\ \%\ reflectance\ at\ 1100\ nm}$$

$$Feature\ 7:\ ratio = \frac{\%\ reflectance\ at\ 1550\ nm}{\%\ reflectance\ at\ 1250\ nm}$$

$$Feature\ 8:\ ratio = \frac{\%\ reflectance\ at\ 1050\ nm}{\%\ reflectance\ at\ 800\ nm}$$

$$Feature\ 9:\ ratio = \frac{\%\ reflectance\ at\ 1550\ nm}{\%\ reflectance\ at\ 1050\ nm}$$

$$Feature\ 10:\ ratio = \frac{\%\ reflectance\ at\ 1250\ nm}{\%\ reflectance\ at\ 800\ nm}$$

$$Feature\ 11:\ ratio = \frac{\%\ reflectance\ at\ 1550\ nm}{\%\ reflectance\ at\ 800\ nm}$$

$$Feature\ 12:\ ratio = \frac{\%\ reflectance\ at\ 1250\ nm}{\%\ reflectance\ at\ 1050\ nm}$$

The normalized area under the reflectance curve was calculated using the trapezoidal rule:

$$Area = \frac{h}{2}\ [\ f\ (a)\ +\ 2\sum_{i=1}^{n-1} f\ (x_i)\ +\ f\ (b)\ ]$$

where h = step size or interval,
a = first point of the curve,
b = last point of the curve,
I = integer, and
n = maximum number of data.

Procedure DISCRIM of SAS (1990) was used to classify the various bulk grains using the thirteen features. The analysis was done using the hold-out method with normal and non-parametric estimations. Four of the five replicates were used for the training set and the remaining replicate was used for the testing set. Each replicate was used as a test set, giving 5 analyses for each group.

The discriminant analyses were done on different groups: cereals (8 classes), oilseeds (3 classes), pulses (8 classes), and specialty seeds (27 classes). The analyses were also done on a group of samples composed of cereals, oilseeds, and pulses. For CWRS wheat, the analyses were conducted for groups of samples based on growing region, moisture content, foreign material, and grade.

Procedure STEPDISC of SAS (1990) was used to determine the level of contribution of each feature to the classification. Individual contribution to the classification was determined by removing the best feature, redoing the analysis with the remaining features. This process was repeated until only one feature was left or the feature(s) did not meet the criterion and hence had no discriminating power. All five replicates for each group were used for the analyses.

CHAPTER 4: RESULTS AND DISCUSSION

Note that the discontinuity, or shift, of the reflectance characteristic curves (Appendices A and B) at 850 nm is due to a change in detectors of the spectrophotometer (from a photomultiplier tube for measuring the ultraviolet and visible spectrum to a lead-sulfide detector for measuring the near-infrared spectrum). The legend of each figure is listed in corresponding order based on the end of each curve.

4.1 Cereals

The classification accuracies (Table 4.1) were determined by the reflectance characteristics (Appendix B as Fig. B1) and the confusion matrices (Appendix C as Tables C1a and C1b).

A classification accuracy of 100% was obtained for 7 (2 row barley, 6 row barley, CWRS wheat, CWAD wheat, buckwheat, oats, and rye) of 8 classes of cereals using normal estimation. Soft white spring wheat was misclassified as CWAD wheat for one of the replications, giving a classification accuracy of 80%.

Using non-parametric estimation, a classification accuracy of 100% was obtained for 6 (2 row barley, 6 row barley, CWRS wheat, buckwheat, oats, and rye) of 8 classes of cereals. The remaining two classes, SWSW and CWAD wheat, had classification accuracies of 80%. One SWSW replicate was misclassified as CWAD wheat and one CWAD wheat replicate was misclassified as SWSW.

Normal estimation was slightly more accurate than non-parametric estimation (one more class was correctly classified) indicating that the samples might be normally distributed. The one replicate of SWSW that was misclassified as CWAD wheat may be due to the small number of replicates such that if one replicate was slightly deviated from the others it may not have been correctly classified.

Reflectance characteristics were highly successful in classifying cereals. These results show that the use of reflectance characteristics in classifying grains are promising as part of an automated system for garin identification. Further research will be needed to implement the use of these reflectance characteristics in the automated system.

The individual rankings of the features are listed in Table 4.2. The most significant feature used in the model was feature 11 (ratio of % reflectance at 1550 nm to % reflectance at 800 nm). The least significant feature was feature 2 (slope ratio of % reflectance at 1000-1100 nm to % reflectance at 750-850 nm).

Table 4.1 Classification accuracies (%) for eight types of cereals using normal and non-parametric estimation with five replications.

Cereal	Normal	Non-parametric
2 row barley	100	100
6 row barley	100	100
CWRS wheat	100	100
CWAD wheat	100	80
Buckwheat	100	100
Oats	100	100
Rye	100	100
SWSW	80	80

Table 4.2. Individual rankings of features using STEPDISC analysis with eight classes of cereals in the model.

Number	Selected features	Average squared canonical correlation	Partial r^2
1	Feature 11	0.1395	0.9763
2	Feature 12	0.1389	0.9727
3	Feature 9	0.1379	0.9655
4	Feature 10	0.1372	0.9601
5	Feature 7	0.1366	0.9559
6	Feature 1	0.1364	0.9549
7	Feature 8	0.1353	0.9474
8	Feature 13	0.1321	0.9246
9	Feature 4	0.1199	0.8393
10	Feature 3	0.1146	0.8025
11	Feature 5	0.1107	0.7752
12	Feature 6	0.0927	0.6488
13	Feature 2	0.0732	0.5125

4.2 Oilseeds

Oilseeds were 100% correctly classified using normal and non-parametric estimations. The reflectance characteristics are shown in Fig. B2 of Appendix B. The confusion matrices are presented in Table C2 of Appendix C. The individual rankings of the features are listed in Table 4.3. The most significant feature used in the model was feature 11 (ratio of % reflectance at 1550 nm to % reflectance at 800 nm). The least significant feature was feature 13 (area under the reflectance curve).

Reflectance characteristics were completely successful in classifying oilseeds. This implies that reflectance characteristics can be used for classifying oilseeds as part of an automated system to identify oilseeds.

Table 4.3. Individual rankings of features using STEPDISC analysis with three classes of oilseeds in the model.

Number	Selected features	Average squared canonical correlation	Partial r^2
1	Feature 11	0.4990	0.9979
2	Feature 12	0.4965	0.9930
3	Feature 7	0.4951	0.9903
4	Feature 10	0.4945	0.9889
5	Feature 9	0.4942	0.9884
6	Feature 5	0.4801	0.9602
7	Feature 6	0.4800	0.9602
8	Feature 4	0.4620	0.9239
9	Feature 2	0.4523	0.9046
10	Feature 8	0.4507	0.9014
11	Feature 1	0.3888	0.7775
12	Feature 3	0.3546	0.7092
13	Feature 13	0.3197	0.6394

4.3 Pulses

The classification accuracies (Table 4.4) were determined by the reflectance characteristics (Fig. B3 of Appendix B) and the confusion matrices (Tables C3a and C3b of Appendix C).

The classification accuracies were 100% for 5 (black beans, dark green speckled lentils, faba beans, lentils, and pinto beans) of 8 classes of pulses using normal estimation. The remaining three classes (green beans, pea beans, and peas) had an accuracy of 80%. Green beans were misclassified as peas, pea beans as green beans, and peas as lentils.

For non-parametric estimation, the classification accuracies were 100% for 3 (dark green speckled lentils, lentils, and pinto beans) of the 8 classes of pulses. Three (black beans, faba beans, and green beans) of the classes had an 80% classification accuracy. Black beans in one replicate were identified as dark green speckled lentils; faba beans were identified as dark green speckled lentils; and green beans were identified as peas (same as normal estimation). The remaining two classes, pea beans and peas, had 60% classification accuracy. Pea beans were misclassified twice as green beans. Peas had one replicate misclassified as green beans and another replicate misclassified as pea beans.

Normal estimation was more effective indicating that the samples were normally distributed. The misclassification of green beans, pea beans, and peas may be due to the small number of replicates and the similarity of their reflectance curves.

Reflectance characteristics were highly successful in classifying pulses. These results show that the use of reflectance characteristics in classifying pulses are promising as part of an automated system for seed identification. Further research will be needed to implement the use of these reflectance characteristics in the automated system.

The individual rankings of the features are listed in Table 4.5. The most significant feature used in the model was feature 11 (ratio of % reflectance at 1550 nm to % reflectance at 800 nm). The least significant feature was feature 2 (slope ratio of % reflectance at 1000-1100 nm to % reflectance at 750-850 nm).

Table 4.4. Classification accuracies (%) for eight types of pulses using normal and non-parametric estimation with five replications.

Cereal	Normal	Non-parametric
Black beans	100	80
Dark green speckled lentils	100	100
Faba beans	100	80
Green beans	80	80
Lentils	100	100
Pea beans	80	60
Peas	80	60
Pinto beans	100	100

Table 4.5. Individual rankings of features using STEPDISC analysis with eight classes in the model.

Number	Selected features	Average squared canonical correlation	Partial r^2
1	Feature 11	0.1350	0.9453
2	Feature 8	0.1344	0.9405
3	Feature 10	0.1341	0.9387
4	Feature 1	0.1306	0.9140
5	Feature 12	0.1238	0.8664
6	Feature 9	0.1230	0.8611
7	Feature 7	0.1190	0.8327
8	Feature 13	0.1178	0.8249
9	Feature 5	0.0970	0.6790
10	Feature 4	0.0880	0.6155
11	Feature 3	0.0793	0.5550
12	Feature 6	0.0712	0.4986
13	Feature 2	0.0504	0.3530

4.4 Cereals, Oilseeds, and Pulses

The classification accuracies are given in Table 4.6 and the confusion matrices are presented in Tables C4a and C4b of Appendix C.

The classification accuracies for the combined analyses of cereals, oilseeds, and pulses resulted in fewer correct classifications for cereals and pulses than when they were analyzed separately. Only faba beans, using non-parametric estimation, had a better classification; increased from 80% to 100%. Oilseeds were not affected.

Those seeds which were 100% correctly classified when in separate groups but only 80% correctly classified in the combined grouping were cereals: 2-row barley (both estimations), CWAD wheat (normal estimation), and rye (both estimations). Those seeds which were 80% correctly classified when in separate groups but only 60% correctly classified in the combined grouping were one cereal and three pulses: CWAD wheat (non-parametric estimation), green beans (normal estimation), pea beans (normal estimation), and peas (normal estimation). Green beans, non-parametric estimation, fell from 80% to 40% correct classification. Pea beans, non-parametric estimation, fell from 60% to 20% correct classification.

The reduction in classification accuracy from the individual group analyses may be due to the greater number of classes available to the discriminant analyses. This introduces more classes that may be similar in their reflectance characteristics and hence, increase the misclassifications.

Reflectance characteristics were still successful in classifying oilseeds. These results show that the use of reflectance characteristics for classifying oilseeds among cereals and pulses are promising. Further research will be needed to implement the use of these reflectance characteristics in an automated system for seed identification.

The individual rankings of the features are listed in Table 4.7. The most significant feature used in the model was feature 12 (ratio of % reflectance at 1250 nm to % reflectance at 1050 nm). The least significant feature was feature 3 (slope ratio of % reflectance at 1500-1600 nm to % reflectance at 1000-1100 nm).

Table 4.6 Classification accuracies (%) for cereals, oilseeds, and pulses analyzed separately and together using normal and non-parametric estimations with five replications.

Grain	Separate Groups		Combined Group	
	Normal	Non-para.	Normal	Non-par.
Cereals				
2-row barley	100	100	80	80
6-row barley	100	100	100	100
CWRS wheat	100	100	100	100
CWAD wheat	100	80	80	60
buckwheat	100	100	100	100
oats	100	100	100	100
rye	100	100	80	80
SWSW	80	80	80	80
Oilseeds				
canola	100	100	100	100
flaxseed	100	100	100	100
sunflower	100	100	100	100
Pulses				
black beans	100	80	100	80
dark grn speckled lentils	100	100	100	100
faba beans	100	80	100	100
green beans	80	80	60	40
lentils	100	100	100	100
pea beans	80	60	60	20
peas	80	60	60	60
pinto beans	100	100	100	100

Table 4.7 Individual rankings of features using STEPDISC analysis with nineteen classes in the model.

Number	Selected features	Average squared canonical correlation	Partial r^2
1	Feature 12	0.0553	0.9945
2	Feature 9	0.0550	0.9900
3	Feature 11	0.0550	0.9899
4	Feature 7	0.0544	0.9793
5	Feature 10	0.0538	0.9692
6	Feature 1	0.0528	0.9501
7	Feature 8	0.0527	0.9483
8	Feature 13	0.0503	0.9048
9	Feature 4	0.0491	0.8841
10	Feature 5	0.0464	0.8361
11	Feature 2	0.0414	0.7449
12	Feature 6	0.0410	0.7387
13	Feature 3	0.0366	0.6593

4.5 Specialty Seeds

The classification accuracies (Table 4.8) were determined by the reflectance characteristics (Fig. B4a to B4d of Appendix B) and the confusion matrices (Tables C5a and C5b of Appendix C). Note that the reflectance characteristics of crested wheatgrass (crest) and creeping red fescue (crfes) appear on each graph to act as a reference between all four graphs.

The classification accuracies were 100% for 18 (alsike clover, brown mustard, canary seed, creeping red fescue, creeping bentgrass, crown millet, intermediate wheatgrass, Kentucky bluegrass, orchard grass, oriental mustard, perennial ryegrass, red clover, reed canary grass, sorghum sudangrsss, sweet clover, timothy, birdsfoot trefoil, and yellow mustard) of the 27 classes of specialty seeds using normal estimation. Six classes (alfalfa, annual ryegrass, crested wheatgrass, meadow fescue, meadow brome, and tall wheatgrass) had a classification accuracy of 80%. Alfalfa was misclassified as timothy for one replicate. Annual ryegrass was misclassified as meadow fescue, crested wheatgrass as Russian wild rye, meadow fescue as dahurian wild rye, meadow brome as meadow fescue, and tall wheatgrass as crested wheatgrass.

Russian wild rye had a classification accuracy of 60% using normal estimation, and misclassified twice as crested wheatgrass. Bromegrass and dahurian wild rye had classification accuracies of 40%. Bromegrass was misclassified once as creeping red fescue, once as dahurian wild rye, and once as meadow brome. Dahurian wild rye was misclassified once as annual ryegrass, once as meadow fescue, and once as perennial ryegrass.

Using non-parametric estimation, the classification accuracies were 100% for 19 of 27 classes of specialty seeds. The classes were: meadow fescue, Russian wild rye, and the same classes as above for normal estimation except for intermediate wheatgrass. Four classes (alfalfa, annual ryegrass, meadow brome, and intermediate wheatgrass) had a classification accuracy of 80%. Alfalfa was misclassified as timothy (same as normal estimation), annual ryegrass as meadow fescue (same as normal estimation), meadow brome as other, and intermediate wheatgrass as other. Crested wheatgrass had a 60% classification accuracy being misclassified once as Russian wild rye and once as other. Dahurian wild rye and tall wheatgrass both had a classification accuracy of 40%. Dahurian wild rye was misclassified once as perennial ryegrass and twice as other; tall wheatgrass was misclassified once as crested wheatgrass and twice as other. Bromegrass was only correctly identified once, giving a 20% classification accuracy. It was misclassified once as creeping red fescue, once as other, and twice as meadow brome.

Between the two estimation methods, 20 classes had 100% classification accuracies. Of the 27 classes, only 6 classes had different classification accuracies depending on the estimation used for analyses.

Reflectance characteristics were somewhat successful in classifying specialty seeds. These results show that the use of reflectance characteristics in classifying specialty seeds needs further research to be implemented in an automated system for seed identification.

The individual rankings of the features are listed in Table 4.9. The most significant feature used in the model was feature 12 (ratio of % reflectance at 1250 nm to % reflectance

at 1050 nm). The least significant feature was feature 5 (slope ratio of % reflectance at 1500-1600 nm to % reflectance at 750-850 nm).

Table 4.8 Classification accuracies (%) for twenty-seven specialty seeds using normal and non-parametric estimations with five replications.

Specialty Seed	Normal Estimation	Non-parametric Estimation
Alfalfa	80	80
Annual ryegrass	80	80
Alsike clover	100	100
Brown mustard	100	100
Bromegrass	40	20
Canary seed	100	100
Crested wheatgrass	80	60
Creeping red fescue	100	100
Creeping bentgrass	100	100
Crown millet	100	100
Dahurian wild rye	40	40
Meadow fescue	80	100
Intermediate wheatgrass	100	80
Kentucky bluegrass	100	100
Meadow brome	80	80
Orchard grass	100	100
Oriental mustard	100	100
Perennial ryegrass	100	100
Red clover	100	100
Reed canary grass	100	100
Russian wild rye	60	100
Sorghum sudangrass	100	100
Sweet clover	100	100
Tall wheatgrass	80	40
Timothy	100	100
Birdsfoot trofoil	100	100
Yellow mustard	100	100

Table 4.9 Individual rankings of features using STEPDISC analysis with twenty-seven classes in the model.

Number	Selected features	Average squared canonical correlation	Partial r^2
1	Feature 12	0.0379	0.9857
2	Feature 9	0.0378	0.9837
3	Feature 7	0.0377	0.9811
4	Feature 13	0.0375	0.9756
5	Feature 11	0.0374	0.9732
6	Feature 6	0.0374	0.9727
7	Feature 4	0.0372	0.9696
8	Feature 10	0.0371	0.9647
9	Feature 8	0.0368	0.9573
10	Feature 1	0.0366	0.9532
11	Feature 2	0.0363	0.9431
12	Feature 3	0.0344	0.8952
13	Feature 5	0.0341	0.8861

4.6 Canada Western Red Spring (CWRS) Wheat Grading Characteristics

Reflectance data features were not able to distinguish with 100% accuracy the growing regions, moisture contents, grades, or amount of foreign material levels of CWRS wheat. Moisture contents and grades were classified with a better accuracy than growing regions or amount of foreign material levels. These results suggest that the reflectance characteristics of wheat grown in different regions across Western Canada are similar. The presence of foreign material has negligible effect on reflectance characteristics of wheat; moisture content and grade of wheat affect the reflectance characteristics slightly.

4.6.1 Region

The classification accuracies for growing regions (Table 4.10) were determined by the reflectance characteristics (Fig. B5a to B5d of Appendix B) and the confusion matrices (Tables C6a and C6b of Appendix C). The reflectance characteristics are presented in four graphs for easier reading: one for Alberta growing regions, two for Saskatchewan growing regions, and one for Manitoba growing regions.

Region differences were not separable using reflectance characteristics. Classification accuracies ranged from a high of 80% for two regions, Ab6 (both estimations) and Sk9aq (normal estimation only), to a low of 0% for 15 regions using non-parametric estimation. Normal estimation resulted in more replicates being correctly classified but only one or two replicates per growing region.

The individual rankings of the features are listed in Table 4.11. The most significant feature used in the model was feature 8 (ratio of % reflectance at 1050 nm to % reflectance

Table 4.10 Classification accuracies (%) of CWRS wheat grading characteristics using normal and non-parametric estimations with five replications.

	Grading Characteristic	Normal Estimation	Non-parametric Estimation
Region:	Ab2s*	20	0
	Ab3s	0	0
	Ab4s	20	0
	Ab5s	0	0
	Ab6	80	80
	Ab7	20	40
	Mb1	0	0
	Mb3s	40	0
	Mb4s	20	20
	Mb7v	0	0
	Sk1b	0	0
	Sk2b	20	0
	Sk3ast	0	0
	Sk3bnt	0	0
	Sk4a	0	0
	Sk5bq	20	0
	Sk6b	20	0
	Sk8b	20	20
	Sk9aq	80	40
	Sk9b	0	0
Moisture Content:	10	40	40
	12	80	60
	14	20	20
	16	60	60
	18	40	40
Grade:	1	60	60
	2	60	60
	3	40	40
Foreign Material:	3	60	40
	6	20	0
	9	20	20
	12	0	0
	15	0	20

*Ab=Alberta, Mb=Manitoba, Sk=Saskatchewan at 800 nm). The least significant feature was feature 12 (ratio of % reflectance at 1250 nm to % reflectance at 1050 nm). This is in contrast to earlier results for classification of grains. Feature 12 was important to the classification of grains but would not necessarily be expected to be important to the classification of wheat by growing regions. The reflectance characteristics of the wheat by growing regions were very similar and thus feature 12 was not listed as important. These findings are a benefit to the design of an automated system because region will not likely affect the classification of CWRS wheat.

Table 4.11 Individual rankings of features using STEPDISC analysis with twenty classes (regions) in the model.

Number	Selected features	Average squared canonical correlation	Partial r^2
1	Feature 8	0.0361	0.6861
2	Feature 10	0.0328	0.6247
3	Feature 11	0.0323	0.6141
4	Feature 3	0.0289	0.5498
5	Feature 7	0.0284	0.5393
6	Feature 9	0.0271	0.5158
7	Feature 6	0.0266	0.5052
8	Feature 1	0.0256	0.4859
9	Feature 13	0.0217	0.4132
10	Feature 2	0.0202	0.3837
11	Feature 4	0.0186	0.3544
12	Feature 5	0.0152	0.2894
13	Feature 12	0.0146	0.2769

4.6.2 Moisture Content

The classification accuracies for moisture content (Table 4.10) were determined by the reflectance characteristics (Fig. B6 of Appendix B) and the confusion matrices (Tables C7a and C7b of Appendix C).

The best classified moisture content class was 12% moisture at 80% accuracy for normal and non-parametric estimations. The next best classification was 16% moisture content at 60% accuracy for both estimations. Normal and non-parametric estimations produced the same results except that normal estimation misclassified one 12% moisture content replicate as 14% moisture content, and non-parametric estimation misclassified one 12% moisture content replicate as 10% moisture content. All classes had at least one replicate correctly classified.

The poor misclassification could be caused by non-optimum selection of wavelength features. Law and Tkachuk (1977), using wavelengths of 1930 and 2100 nm, determined that near-infrared reflectance spectroscopy was an accurate and reliable method of moisture content estimation. Misclassification may also be due to excessive noise generated by the spectrophotometer as shown in Fig. A2 (Appendix A) for 14% moisture content. (Excessive noise was not likely a factor in the other grading characteristics.) The 14% moisture content level was the least accurate (20% classification accuracy). As well, there was slight variation (see section 3.2) in actual moisture content when verified after conditioning.

The individual rankings of the features are listed in Table 4.12. The most significant feature used in the model was feature 11 (ratio of % reflectance at 1550 nm to % reflectance

at 800 nm). The least significant feature was feature 13 (area under the reflectance curve). Feature 1 did not enter the model.

Table 4.12 Individual rankings of features using STEPDISC analysis with five classes (moisture contents) in the model.

Number	Selected features	Average squared canonical correlation	Partial r^2
1	Feature 11	0.2074	0.8296
2	Feature 10	0.2001	0.8004
3	Feature 9	0.1979	0.7914
4	Feature 7	0.1898	0.7590
5	Feature 8	0.1789	0.7157
6	Feature 12	0.1753	0.7012
7	Feature 2	0.1495	0.5978
8	Feature 6	0.1292	0.5167
9	Feature 4	0.1053	0.4211
10	Feature 5	0.0983	0.3934
11	Feature 3	0.0829	0.3318
12	Feature 13	0.0782	0.3129
-	Feature 1	-	-

4.6.3 Grade

The classification accuracies for grade (Table 4.10) were determined by the reflectance characteristics (Fig. B7 of Appendix B) and the confusion matrices (Table C8 of Appendix C).

Grade classes were classified the same for both estimations. Grade 1 had a classification accuracy of 60% with misclassification of two replicates as grade 3. Grade 2 had a classification accuracy of 60% with misclassification of two replicates as grade 1. Grade 3 had a classification accuracy of 40% with misclassification of three replicates as grade 1. No grades were misclassified as grade 2.

Reflectance characteristics were not successful in classifying grade of CWRS wheat, however the results show that an automated system to classify different grain types will probably identify different grades of wheat as wheat. Such a system can be used for the automation of rail car unloading.

The individual rankings of the features are listed in Table 4.13. The most significant feature used in the model was feature 9 (ratio of % reflectance at 1550 nm to % reflectance at 1050 nm). The least significant feature was feature 1 (slope ratio of % reflectance at 1500-1600 nm to % reflectance at 1200-1300 nm). Features 2 to 6 inclusive did not enter the model. The reflectance characteristic curves (Fig. A6) were similar in their slopes. The slope-ratio features were too close due to the similarity of the nature of the slopes of the curves for the discriminant analysis to distinguish and thus did not enter into the model.

Table 4.13 Individual rankings of features using STEPDISC analysis with three classes (grades) in the model.

Number	Selected features	Average squared canonical correlation	Partial r^2
1	Feature 9	0.3398	0.6797
2	Feature 7	0.3290	0.6581
3	Feature 11	0.3222	0.6444
4	Feature 12	0.3087	0.6174
5	Feature 13	0.2964	0.5928
6	Feature 10	0.2700	0.5399
7	Feature 8	0.2382	0.4765
8	Feature 1	0.2082	0.4165
-	Feature 2	-	-
-	Feature 3	-	-
-	Feature 4	-	-
-	Feature 5	-	-
-	Feature 6	-	-

4.6.4 Foreign material

The classification accuracies for foreign material content (Table 4.10) were determined by the reflectance characteristics (Fig. B8 of Appendix B) and the confusion matrices (Tables C9a and C9b of Appendix C).

The amount of foreign material in a sample was not accurately predicted using the wavelength features extracted from the reflectance data. The best classification accuracy was 60% for 3% foreign material content using normal estimation. The worst classification accuracy using normal estimation was zero for 12% and 15% content; and, using non-parametric estimation, zero for 6% and 12% content. These results may be due to the sample preparation as there was wide variation between the reflectance characteristic curves of the replicates. The subsample may not have had the same percent of foreign material for each replicate due to the settling and size of the chaff. These results show that the use of reflectance characteristics in classifying grains as part of the machine vision system need further research to be implemented.

The individual rankings of the features are listed in Table 4.14. The most significant feature used in the model was feature 8 (ratio of % reflectance at 1050 nm to % reflectance at 800 nm). The least significant feature was feature 13 (area under reflectance curve). All other features did not enter the model. The reflectance characteristic curves (Fig. B7, Appendix B), were very close, there were no specific trends found on each of the curves, and most of the features were too close due to the similarity of the nature of the curves. This result may be due to the reflectance characteristics of CWRS wheat rather than the foreign material content grading characteristic of CWRS wheat.

Table 4.14 Individual rankings of features using STEPDISC analysis with five classes (foreign materials) in the model.

Number	Selected features	Average squared canonical correlation	Partial r^2
1	Feature 8	0.0801	0.3206
2	Feature 13	0.0733	0.2934
-	Feature 1	-	-
-	Feature 2	-	-
-	Feature 3	-	-
-	Feature 4	-	-
-	Feature 5	-	-
-	Feature 6	-	-
-	Feature 7	-	-
-	Feature 9	-	-
-	Feature 10	-	-
-	Feature 11	-	-
-	Feature 12	-	-

CHAPTER 5: SUMMARY

The use of reflectance data to characterize bulk samples of grain is very promising (Table 5.1). Oilseeds were correctly classified 100% of the time regardless of method of estimation or of grouping used for analyses. This result is likely due to the wide variation in the reflectance curves for sunflowers, canola, and flaxseed, as well as to the features chosen from the reflectance data for the discriminant analysis. The normal estimation method of analyses correctly classified more cereals and pulses than did the non-parametric estimation. This result may be due to the sample having a normal, or near normal distribution. Analyses of cereals, pulses, and oilseeds together did not result in a higher classification accuracy for cereal and pulses. This result may be due in part to the similarity of the reflectance curves between some cereals and some pulses. The classification accuracies for the specialty seeds were about equal between the two estimation methods. This result may be due in part to the similarity of the reflectance curves between different seeds types, and to the small number of replicates used for the trainning set.

Ratios of percent reflectance values were more important features than slope-ratios or the area under the reflectance curve in classifying bulk grains (Table 5.2). This result may be due to the similarity of the reflectance curves such that the slope-ratio is unable to distinguish differences between seed types. One slope-ratio feature became important at the fourth level for pulses and at the sixth level for cereals, oilseeds, and specialty seeds. The area under the reflectance curve first became significant at the fourth level for specialty seeds

Table 5.1 Summary of the number of classes per classification accuracy for cereals, oilseeds, pulses, and specialty grasses using normal and non-parametric estimation with results of combined analyses of cereals, oilseeds, and pulses in parentheses.

	Classification Accuracy (%) (based on 5 replicates)	Cereals [8 classes]	Oilseeds [3 classes]	Pulses [8 classes]	Specialty Seeds [27 classes]
Normal estimation:	100	7 (4)	3 (3)	5 (5)	18
	80	1 (4)	0	3 (3)	6
	60	0	0	0	1
	40	0	0	0	2
	20	0	0	0	0
	0	0	0	0	0
Non-parametric estimation:	100	6 (4)	3 (3)	3 (4)	19
	80	2 (3)	0	3 (1)	4
	60	0 (1)	0	2 (1)	1
	40	0	0	0 (1)	2
	20	0	0	0 (1)	1
	0	0	0	0	0

Table 5.2 Summary of individual rankings of features listing most significant feature to least significant feature that entered the model.

Bulk Grain Samples					Grading Characteristics of CWRS			
Cereals	Oilseeds	Pulses	Specialty Seeds	Combined Analyses	Growing Region	Moisture Content	Grade	Foreign Material
11	11	11	12	12	8	11	9	8
12	12	8	9	9	10	10	7	13
9	7	10	7	11	11	9	11	
10	10	1	13	7	3	7	12	
7	9	12	11	10	7	8	13	
1	5	9	6	1	9	12	10	
8	6	7	4	8	6	2	8	
13	4	13	10	13	1	6	1	
4	2	5	8	4	13	4		
3	8	4	1	5	2	5		
5	1	3	2	2	4	3		
6	3	6	3	6	5	13		
2	13	2	5	3	12			

and at the eighth level for cereals and pulses. The area under the reflectance curve was the least significant feature in the oilseeds analyses.

Based on the features that I chose, the four most important features for cereals, oilseeds, and specialty seeds were features 11, 12, 9, and 7 which were ratios with a wavelength of 1550 or 1250 nm for the numerator and a wavelength of 1050 or 800 nm for the denominator. The three most important features for pulses were features 11, 8, and 10 which were ratios with a wavelength of 800 nm for the denominator. These results indicate that three different kinds of filters could be used on the machine-vision system design to obtain the classification information as the grains are passed through the line at the grain elevator.

Reflectance characteristics were not sucessful in classifying the grading characteristics of CWRS wheat. Wheat grown in different regions across Western Canada showed similar reflectance characteristics. Based on the wavelengths used here, the machine vision system would not have to be recalibrated for wheat grown in different regions of Canada. Moisture content and grade of wheat affected the reflectance characteristics slightly. Moisture content misclassification was likely due to non-optimum wavelength selection, excessive noise, and slight variation in conditioned to actual moisture content. The presence of foreign material had negligible effect on reflectance characteristics of wheat. This result may be due to the sample preparation method because the variation between replicates was quite large. Further research is needed in the application of reflectance characteristics to grading characteristics of CWRS wheat.

The features important to the classification of the grading characteristics were similar for growing region, moisture content, and grade. Only two features had any discriminating ability for foreign material content. This result is likely due to the reflectance characteristics of CWRS wheat rather than the grading characteristics.

CHAPTER 6: CONCLUSIONS

The research in this thesis contributes part of the information required to develop a machine vision system for the automated classifying and sorting of bulk grains. Thirteen features were extracted from the reflectance characteristics data: six slope-ratios, six ratios, and the area under the reflectance curve. These features were used to determine the feasibility of using reflectance data for classification of bulk grains including cereals, oilseeds, pulses, and specialty seeds. The reflectance characteristics were also used to determine the feasibility of classifying CWRS wheat grading characteristics including growing regions, moisture contents, grades, and amount of foreign materials.

The following results were obtained:

1. Reflectance characteristics successfully classified (100% accuracy) the oilseeds: flax, canola, and sunflower.

2. Reflectance characteristics successfully classified (100% accuracy) seven of the eight classes of cereals: 2-row barley, 6-row barley, CWRS wheat, CWAD wheat, buckwheat, oats, and rye. Soft white spring wheat was misclassified once as CWAD wheat.

3. Reflectance characteristics successfully classified (100% accuracy) five of the eight classes of pulses: black beans, dark green speckled lentils, faba beans, lentils, and pinto beans. Green beans were misclassified once as peas, pea beans once as green beans, and peas once as lentils.

4. Reflectance characteristics successfully classified (100% accuracy) twenty of the twenty-seven classes of specialty seeds: alsike clover, brown mustard, canary seed, creeping red fescue, creeping bentgrass, crown millet, meadow fescue, intermediate wheatgrass, Kentucky bluegrass, orchard grass, oriental mustard, perennial ryegrass, red clover, reed canary grass, Russian wild rye, sorghum sudangrsss, sweet clover, timothy, birdsfoot trefoil, and yellow mustard.

5. Reflectance characteristics did not successfully classify grading characteristics of CWRS wheat. Classification accuracies ranged from 0 to 80%.

6. Classification accuracies for cereals and pulses were higher when normal estimation was used than when non-parametric estimation was used.

7. Ratio features contributed more to the classification accuracies than did the slope-ratios or the area under the reflectance curve features.

8. The wavelengths that best classified the bulk grain samples were 800, 1050, and 1250 nm.

In conclusion, reflectance data are a viable characteristic to be implemented in an industry applied machine vision system for the classification of various types of bulk grains. Further research is required in the classification of grading characteristics of CWRS wheat.

REFERENCES

Bevilacqua, R.T. 1987. Grading of grains and oilseeds in Canada. Lecture presented at the 22nd International Grain Industry Course. Canadian International Grain Institute, Winnipeg, MB, Canada.

Bilanski, W.K., C.L.Pen, and D.R.Fuzzen. 1984. Apple bruise detection using optical reflectance parameters. *Canadian Agricultural Engineering* 26(2):111-114.

Bittner, D.R. and K.H.Norris. 1968. Optical properties of selected fruits vs maturity. *Transactions of the ASAE* 11(4):534-536

Brown, R.B., J.P.G.A.Steckler, and G.W.Anderson. 1994. Remote sensing for identification of weeds in no-till corn. *Transactions of the ASAE* 37(1):297-302.

Canada Grains Council. 1982. *Grain Grading for Efficiency and Profit.* Canada Grains Council, Winnipeg, MB.

Canadian Grain Commission. 1987. *Official Grain Grading Guide.* Canadian Grain Commission, Winnipeg, MB.

Casady, W.W., M.R.Paulsen, and J.B.Sinclair. 1993. Optical properties of damaged soybean seeds. *Transactions of the ASAE* 36(3):943-947.

Chen, C., Y.P.Chiang, and Y.Pomeranz. 1989. Image analysis and characterization of cereal grains with a laser range finder and camera contour extractor. *Cereal Chemistry* 66(6):466-470.

Churchill, D.B., D.M.Bilsland, and T.M.Cooper. 1993. Separation of mixed lots of tall fescue and ryegrass seed using machine vision. *Transactions of the ASAE* 36(5):1383-1386.

Delwiche, S.R. 1995. Single wheat kernel analysis by near-infrared transmittance: protein content. *Cereal Chemistry* 72(1):11-16.

Delwiche, S.R. and K.H.Norris. 1993. Classification of hard red wheat by near-infrared diffuse reflectance spectroscopy. *Cereal Chemistry* 70(1):29-35.

Delwiche, S.R., Y.-R.Chen, and W.R.Hruschka. 1995. Differentiation of hard red wheat by near-infrared analysis of bulk samples. *Cereal Chemistry* 72(3):243-247.

Dowell, F.E. 1992. Identifying undamaged and damaged peanut kernels using tristimulus values and spectral reflectance. *Transactions of the ASAE* 35(3):931-937.

Hawk, A.L., H.H.Kaufmann, and C.A.Watson. 1970. Reflectance characteristics of various grains. *Cereal Science Today* 15(11):381-384.

Heinemann, P.H., R.Hughes, C.T.Morrow, H.J.Sommer III, R.B.Beelman, and P.J.Wuest. 1994. Grading of mushrooms using a machine vision system. *Transactions of the ASAE* 37(5):1671-1677.

Howarth, M.S., S.W.Searcy, and G.S.Birth. 1990. Reflectance characteristics of fresh-market carrots. *Transactions of the ASAE* 33(3):961-964.

Law, D.P. and R.Tkachuk. 1977. Determination of moisture content in wheat by near-infrared diffuse reflectance spectrophotometry. *Cereal Chemistry* 54(4):874-881.

Luo, X. 1997. Color image analysis for cereal grain classification. Unpublished Ph.D. thesis, Department of Biosystems Engineering, University of Manitoba, Winnipeg, MB.

Majumdar, S., D.S.Jayas, J.L.Hehn, and N.R.Bulley. 1996a. Classification of various grains using optical properties. *Canadian Agricultural Engineering* 38(2): 139-144.

Majumdar, S., D.S.Jayas, S.J.Symons, and N.R.Bulley. 1996b. Textural features for automated grain identification. CSAE Paper No. 96-602. Saskatoon, SK: CSAE.

Massie, D.R. and K.H.Norris. 1965. Spectral reflectance and transmittance properties of grain in the visible and near infrared. *Transactions of the ASAE* 8(4):598-600.

Miller, B.K. and M.J.Delwiche. 1991. Spectral analysis of peach surface defects. *Transactions of the ASAE* 34(6):2509-2515.

Neuman, M., H.D.Sapirstein, E.Shwedyk, and W.Bushuk. 1989a. Wheat grain color analysis by digital image processing: I. Methodology. *Journal of Cereal Science* 10:175-182.

Neuman, M., H.D.Sapirstein, E.Shwedyk, and W.Bushuk. 1987. Discrimination of wheat class and variety by digital image analysis of whole grain samples. *Journal of Cereal Science* 6:125-132.

Neuman, M., H.D.Sapirstein, E.Shwedyk, and W.Bushuk. 1989b. Wheat grain color analysis by digital image processing: II. Wheat class discrimination. *Journal of Cereal Science* 10:183-188.

Panford, J.A. 1987. Application of Near-Infrared Reflectance Spectroscopy in North America. In *Near-Infrared Technology in the Agricultural and Food Industries.* Williams, P. and K.Norris (ed.) American Association of Cereal Chemists, Inc., St. Paul, Minnesota, USA. 330p.

Sapirstein, H.D. and J.M.Kohler. 1995. Physical uniformity of graded railcar and vessel shipments of Canada Western Red Spring wheat determined by digital image analysis. *Canadian Journal of Plant Science* 75:363-369.

Sapirstein, H.D., M.Neuman, E.H.Wright, E.Shwedyk, and W.Bushuk. 1987. An instrumental system for cereal grain classification using digital image analysis. *Journal of Cereal Science* 6:3-14.

Saputra, D., I.W.Budiastra, and H.K.Purwadaria. 1995. Classification of mango by near infrared diffuse reflectance. *Proceedings of the FPAC IV Conference* 197-203, St. Joseph, MI: ASAE.

SAS. 1990. *SAS User's Guide: Statistics.* Raleigh, NC: Statistical Analysis System, Inc.

Shatadal, P., D.S.Jayas, and N.R.Bulley. 1995a. Digital image analysis for software separation and classification of touching grains: II. Classification. *Transactions of the ASAE* 38(2):645-649.

Shatadal, P., D.S.Jayas, J.L.Hehn, and N.R.Bulley. 1995b. Seed classification using machine vision. *Canadian Agricultural Engineering* 37(3):163-167.

Singh, N. and M.J.Delwiche. 1994. Machine vision methods for defect sorting stonefruit. *Transactions of the ASAE* 37(6):1989-1997.

Slaughter, D.C., K.H.Norris, and W.R.Hruschka. 1992. Quality and classification of hard red wheat. *Cereal Chemistry* 69(4):428-432.

Statistics Canada. 1997. *Grain Trade of Canada.* Minister of Industry, Statistics Canada, Ottawa, ON. 92p.

Stermer, R.A., Y.Pomeranz, and R.J.McGinty. 1977. Infrared reflectance spectroscopy for estimation of moisture of whole grain. *Cereal Chemistry* 54(2):345-351.

Symons, S.J. and R.G.Fulcher. 1988a. Determination of wheat kernel morphological variation by digital image analysis: I. Variation in Eastern Canadian milling quality wheats. *Journal of Cereal Science* 8:211-218.

Symons, S.J. and R.G.Fulcher. 1988b. Determination of wheat kernel morphological variation by digital image analysis: II. Variation in cultivars of soft white winter wheats. *Journal of Cereal Science* 8:219-229.

Tabil, L.G.,Jr., R.T.Patil, S.Sokhansanj, and K.K.Krishna. 1996. Estimation of leaf fraction in alfalfa chops and cubes by near infrared product analyzer. CSAE Paper No. 96-301. Saskatoon, SK: CSAE.

Tao, Y., P.H.Heinemann, Z.Varghese, C.T.Morrow, and H.J.Sommer III. 1995. Machine vision for color inspection of potatoes and apples. *Transactions of the ASAE* 38(5):1555-1561.

Thomson, W.H. and Y.Pomeranz. 1991. Classification of wheat kernels using three-dimensional image analysis. *Cereal Chemistry* 68(4):357-361.

Upchurch, B.L., H.A.Affeldt, W.R.Hruschka, K.H.Norris, and J.A.Throop. 1990. Spectrophotometric study of bruises on whole, 'Red Delicious' apples. *Transactions of the ASAE* 33(2):585-589.

Williams, P.C. 1975. Application of near infrared reflectance spectroscopy to analysis of cereal grains and oilseeds. *Cereal Chemistry* 52:561-576.

Williams, P.C. 1979. Screening wheat for protein and hardness by near infrared spectroscopy. *Cereal Chemistry* 56(3):169-172.

Winter, P., S.Sokhansanj, C.C.Wood, and W.Crerar. 1996. Quality assessment and grading of lentils using machine vision. CSAE Paper No. 96-310. Saskatoon, SK: CSAE.

Zayas, I., Y.Pomeranz, and F.S.Lai. 1985. Discrimination between Arthur and Arkan wheats by image analysis. *Cereal Chemistry* 62(6):478-480.

Zayas, I., F.S.Lai, and Y.Pomeranz. 1986. Discrimination between wheat classes and varieties by image analysis. *Cereal Chemistry* 63(1):52-56.

APPENDIX A: Reflectance Characteristics of the Five Replicates for Selected Samples.

The legend of each figure is listed in corresponding order based on the end of each curve.

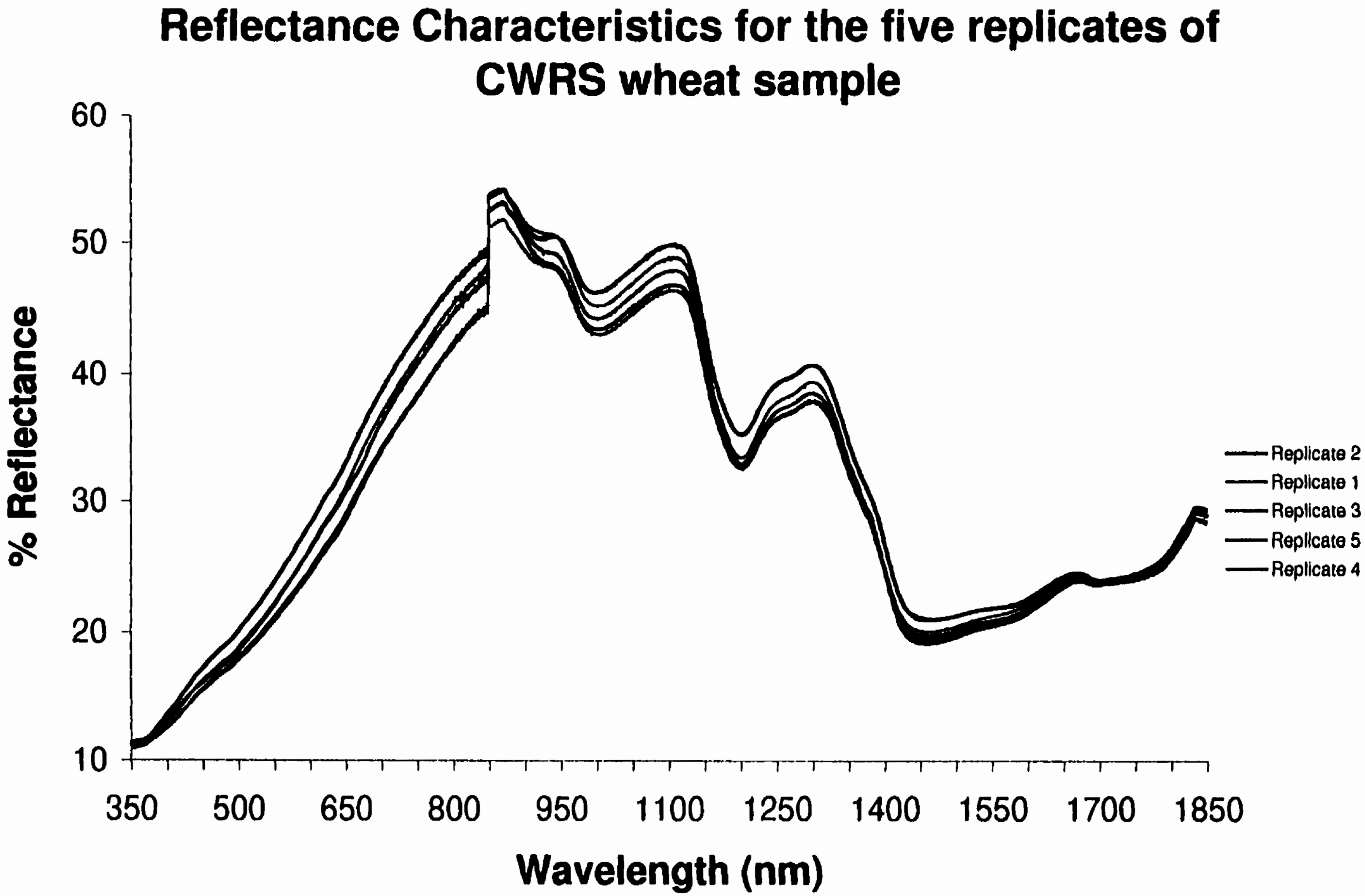

Figure A1. Five replicates of CWRS wheat showing the variability of measurements.

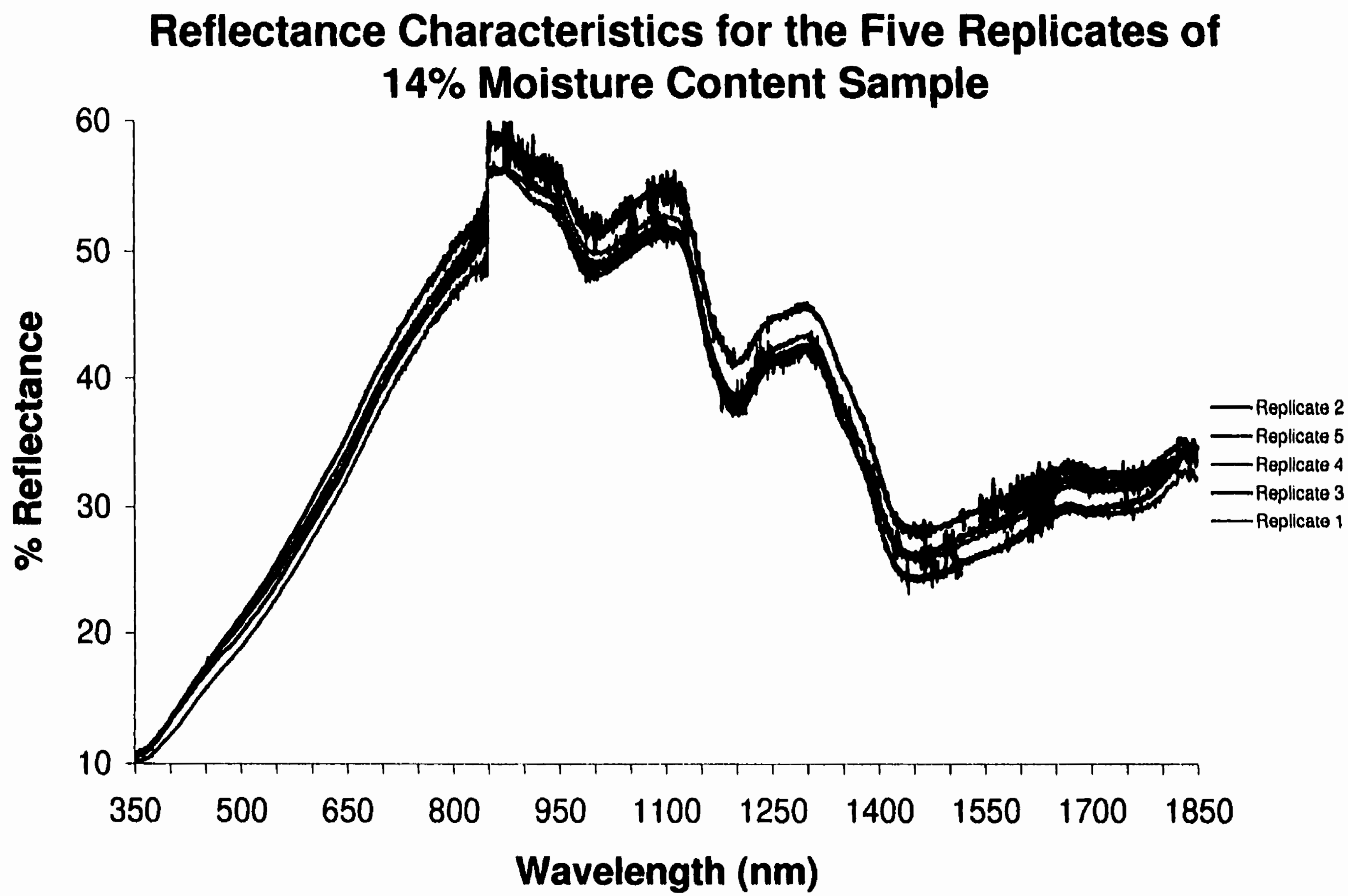

Figure A2. Five replicates of 14% moisture content showing the present of noise and variability of measurement.

APPENDIX B: Average Reflectance Characteristics of Each Sample.

The legend of each figure is listed in corresponding order based on the end of each curve.

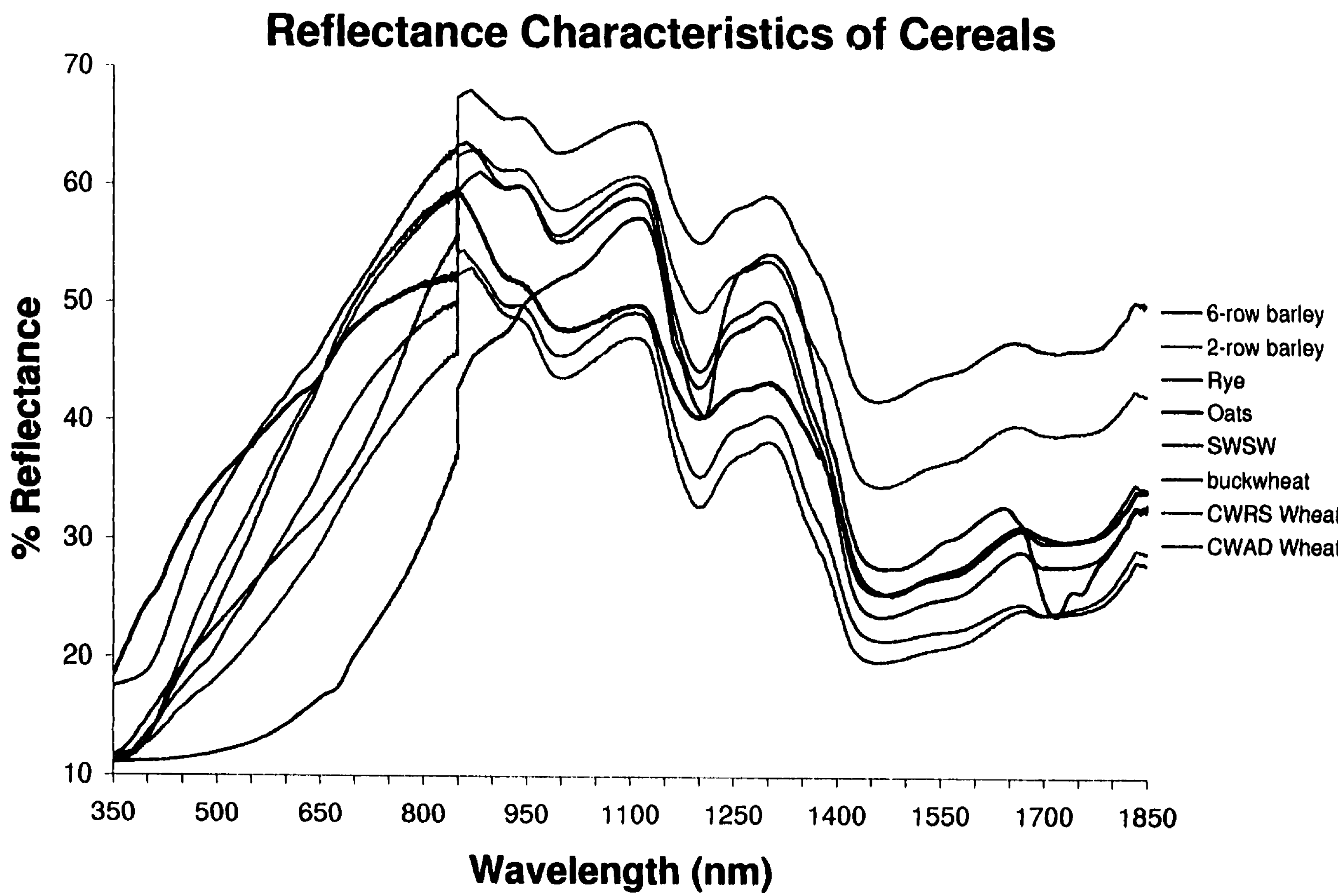

Figure B1. Average of five replicates of each sample of cereals.

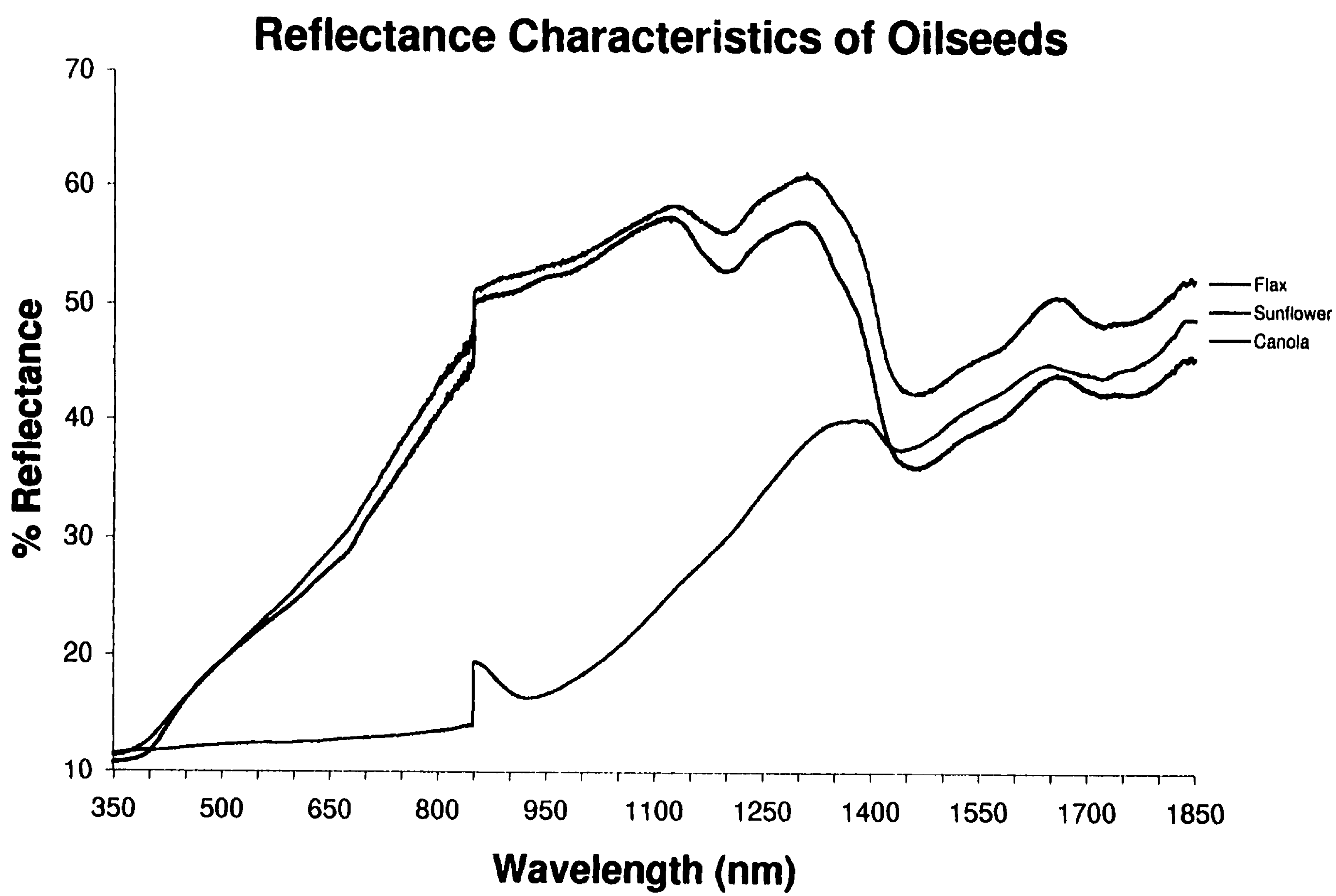

Figure B2. Average of five replicates of each sample of oilseeds.

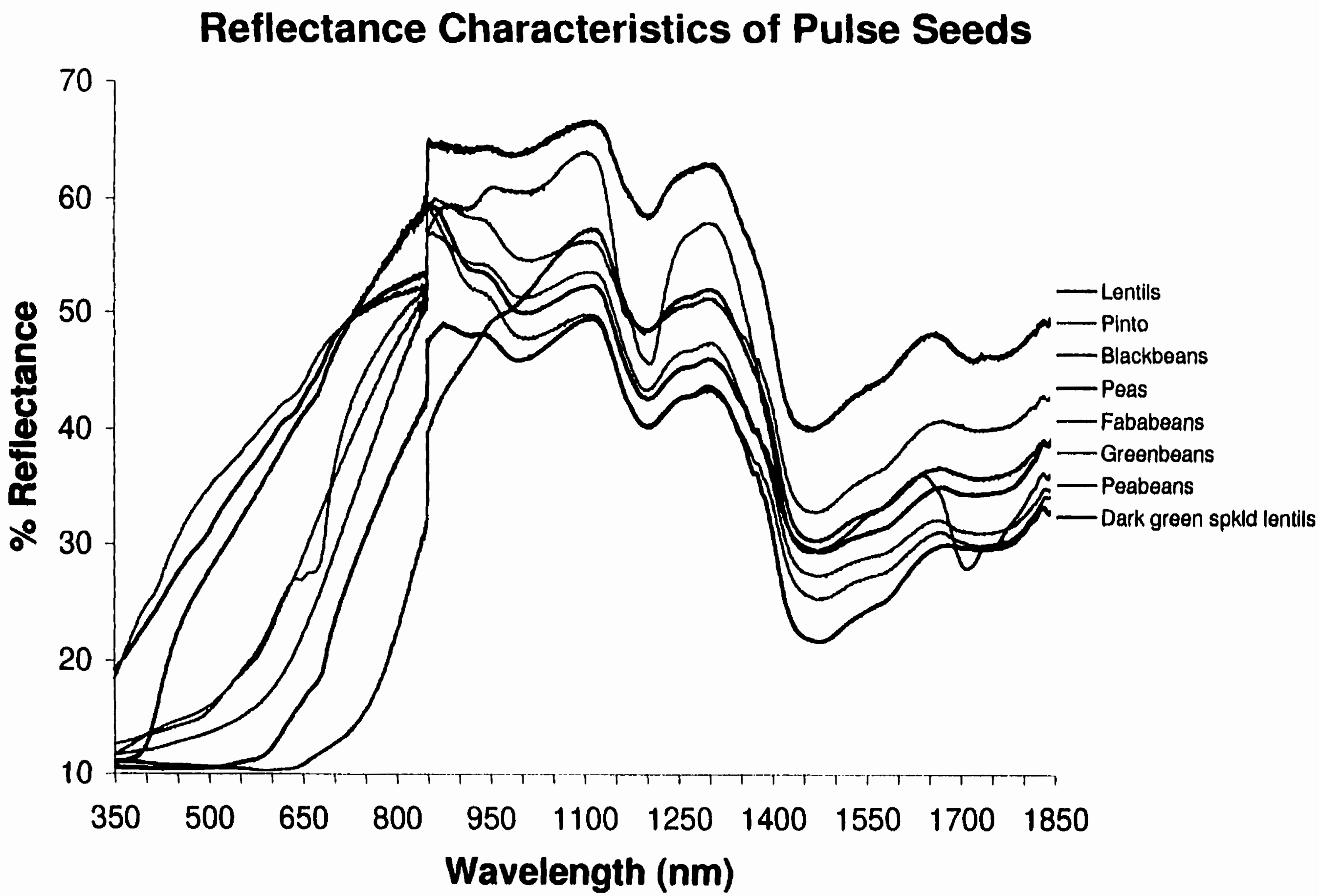

Figure B3. Average of five replicates of each sample of pulse seeds.

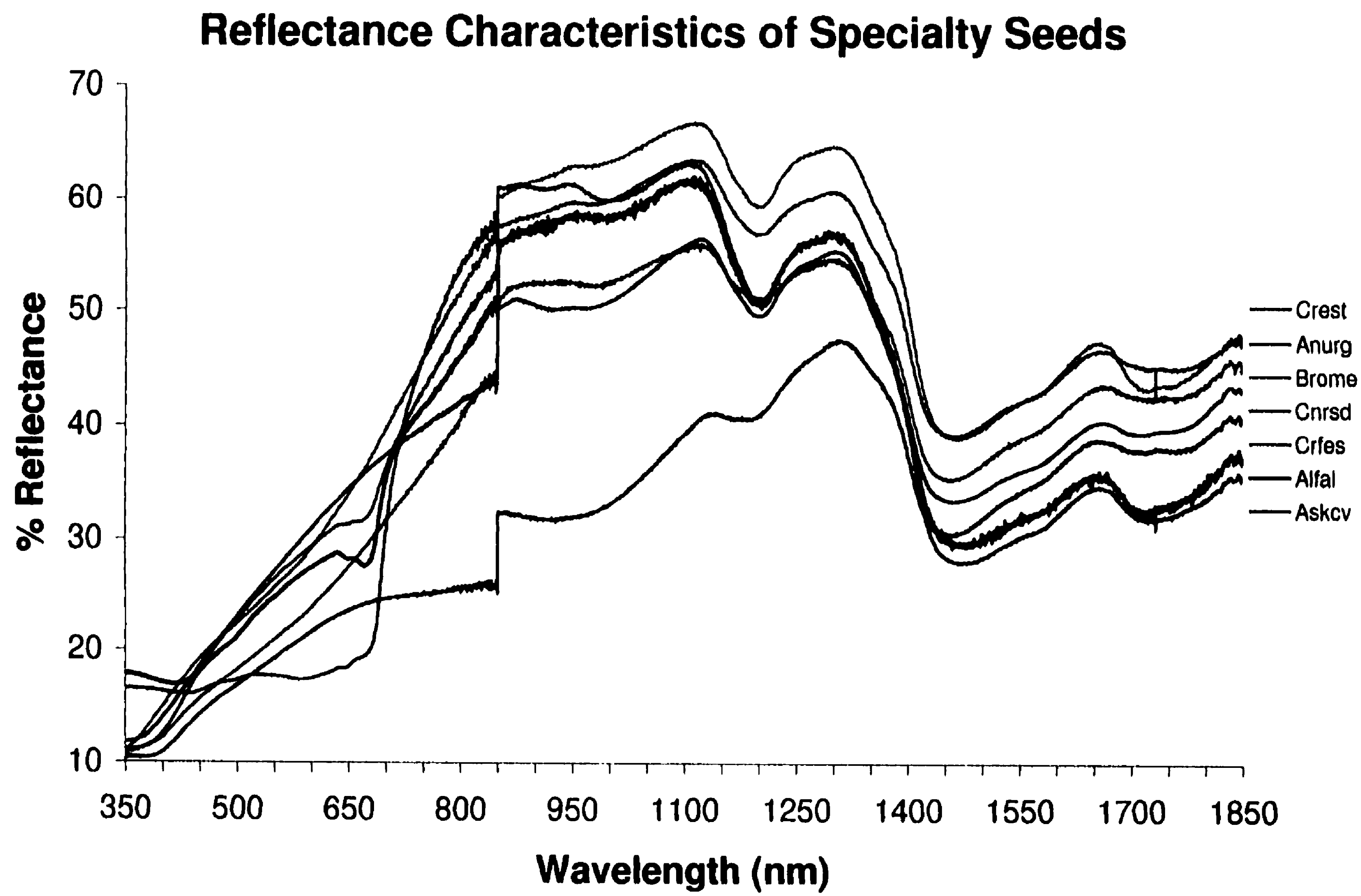

Figure B4a. Average of five replicates of seven samples of specialty seeds.

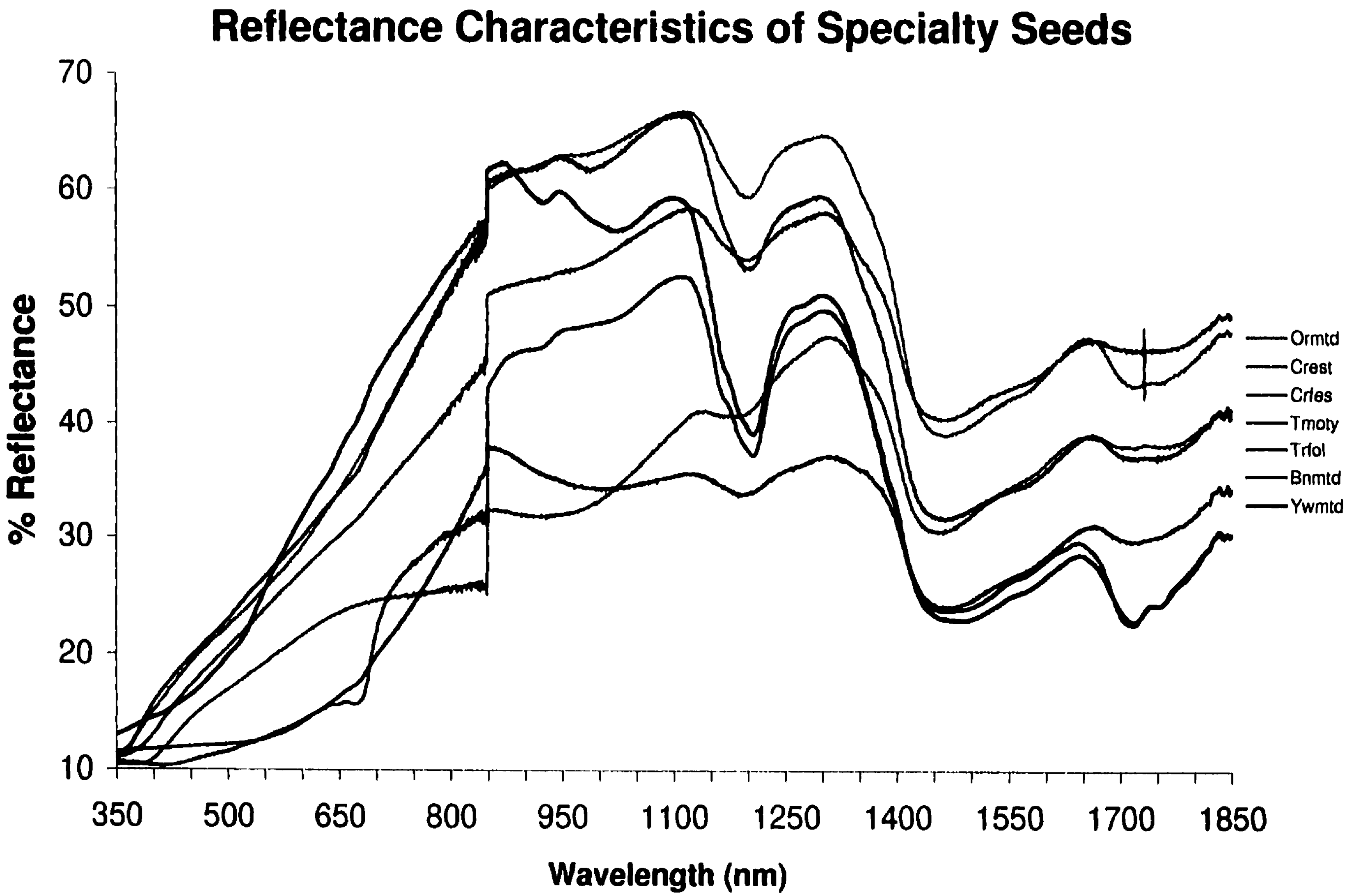

Figure 4b. Average of five replicates of seven samples of specialty seeds.

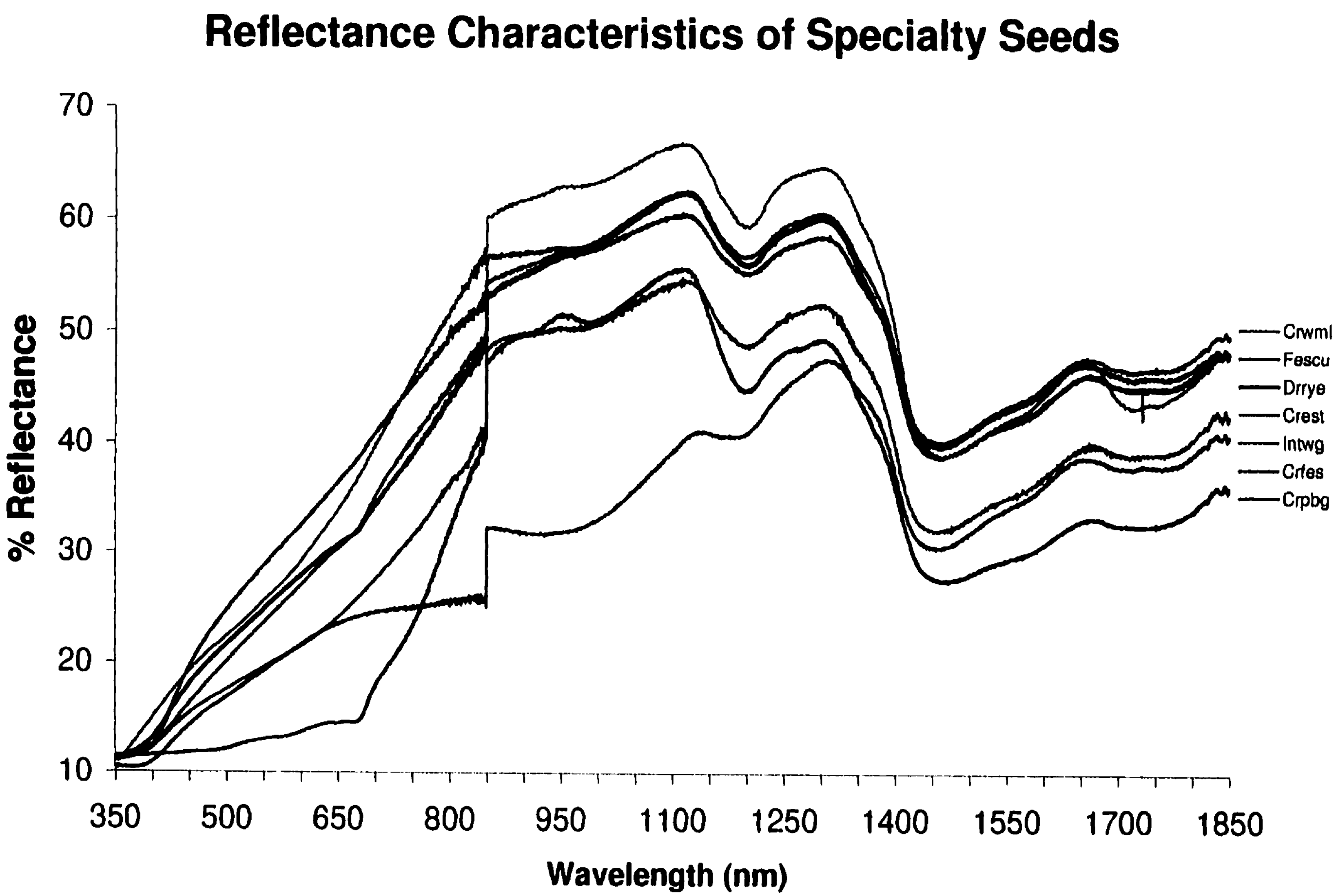

Figure B4c. Average of five replicates of seven samples of specialty seeds.

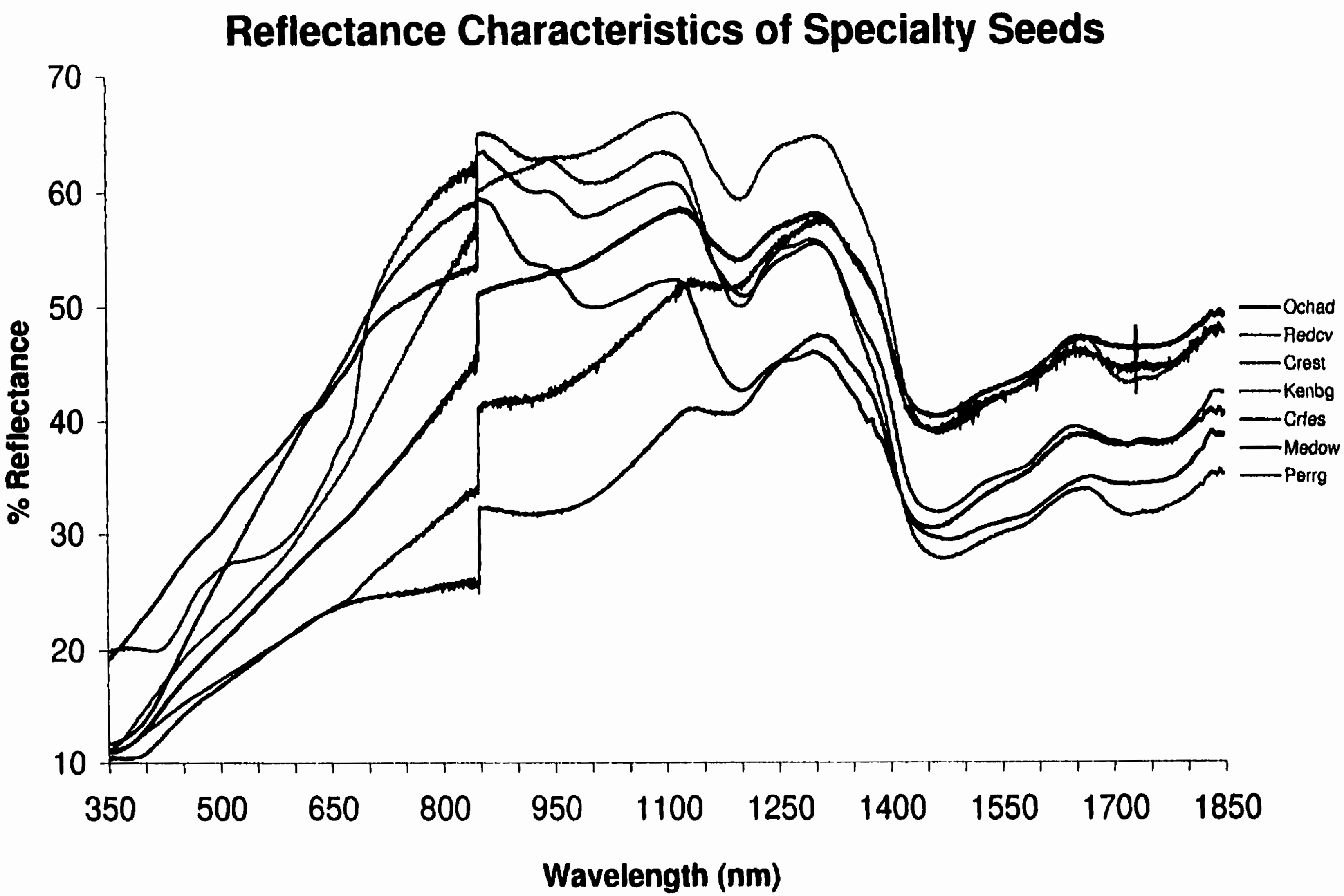

Figure B4d. Average of five replicates of seven samples of specialty seeds.

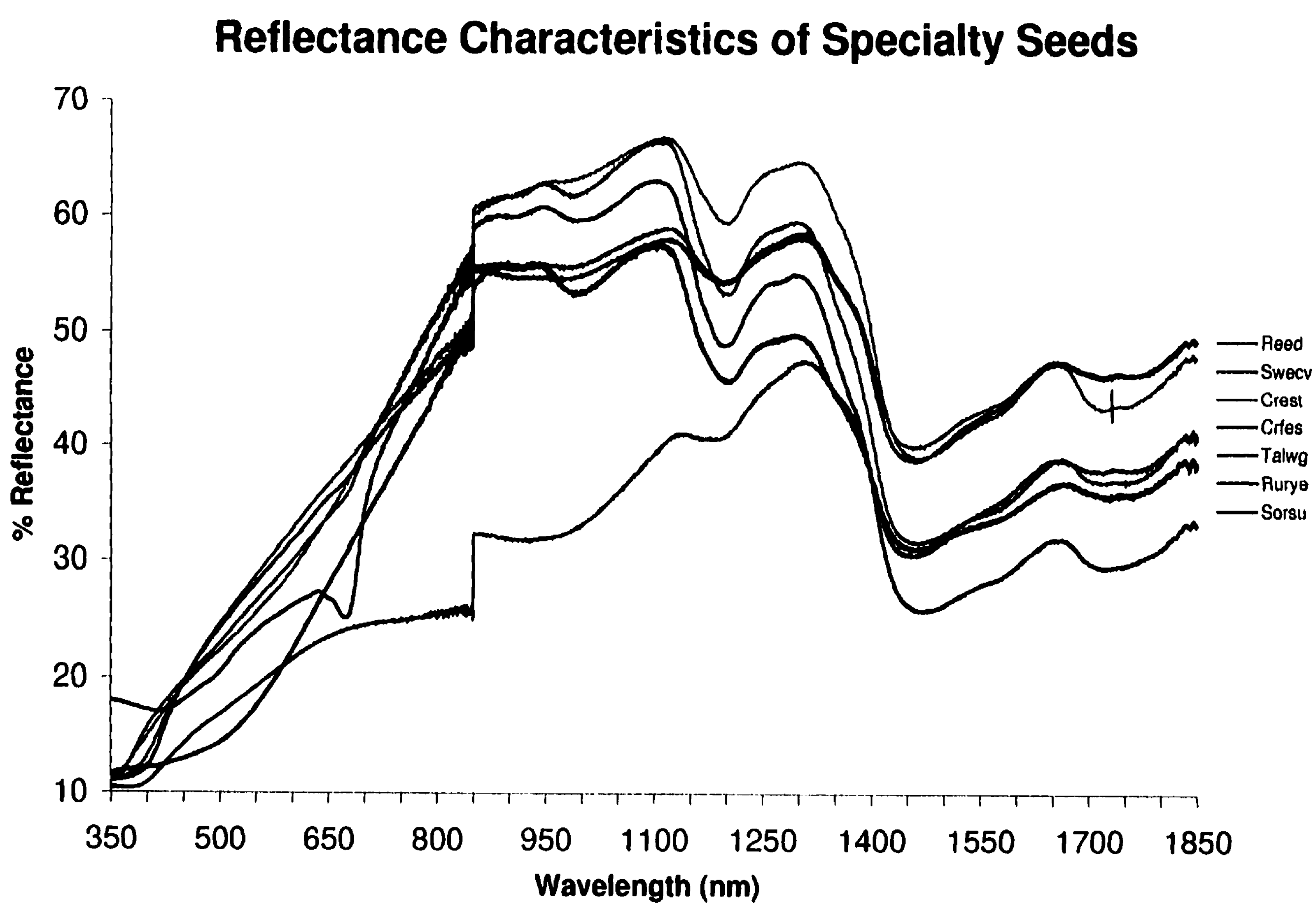

Figure B4e. Average of five replicates of seven samples of specialty seeds.

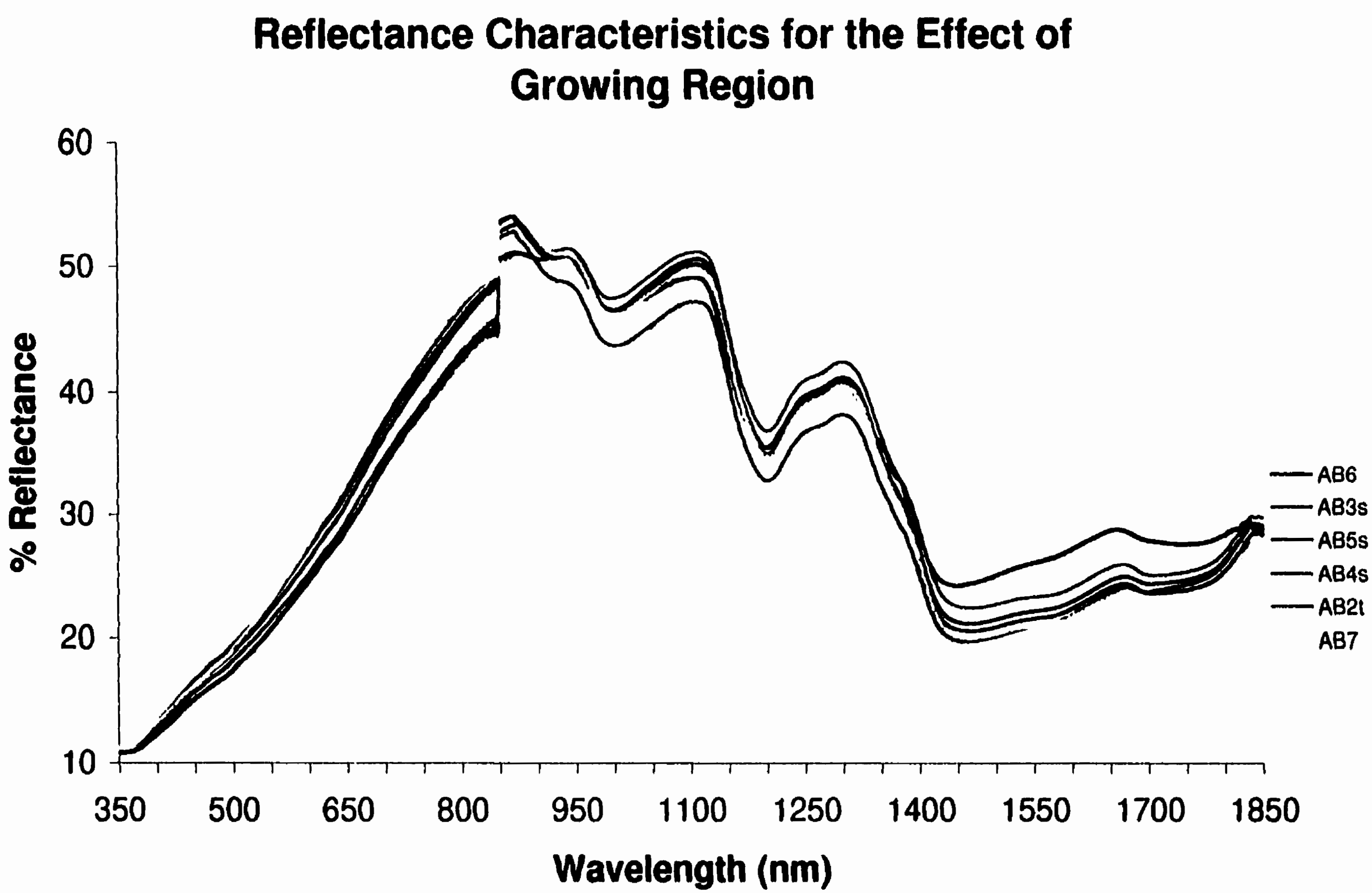

Figure B5a. Average of five replicates of each CWRS wheat sample from Alberta.

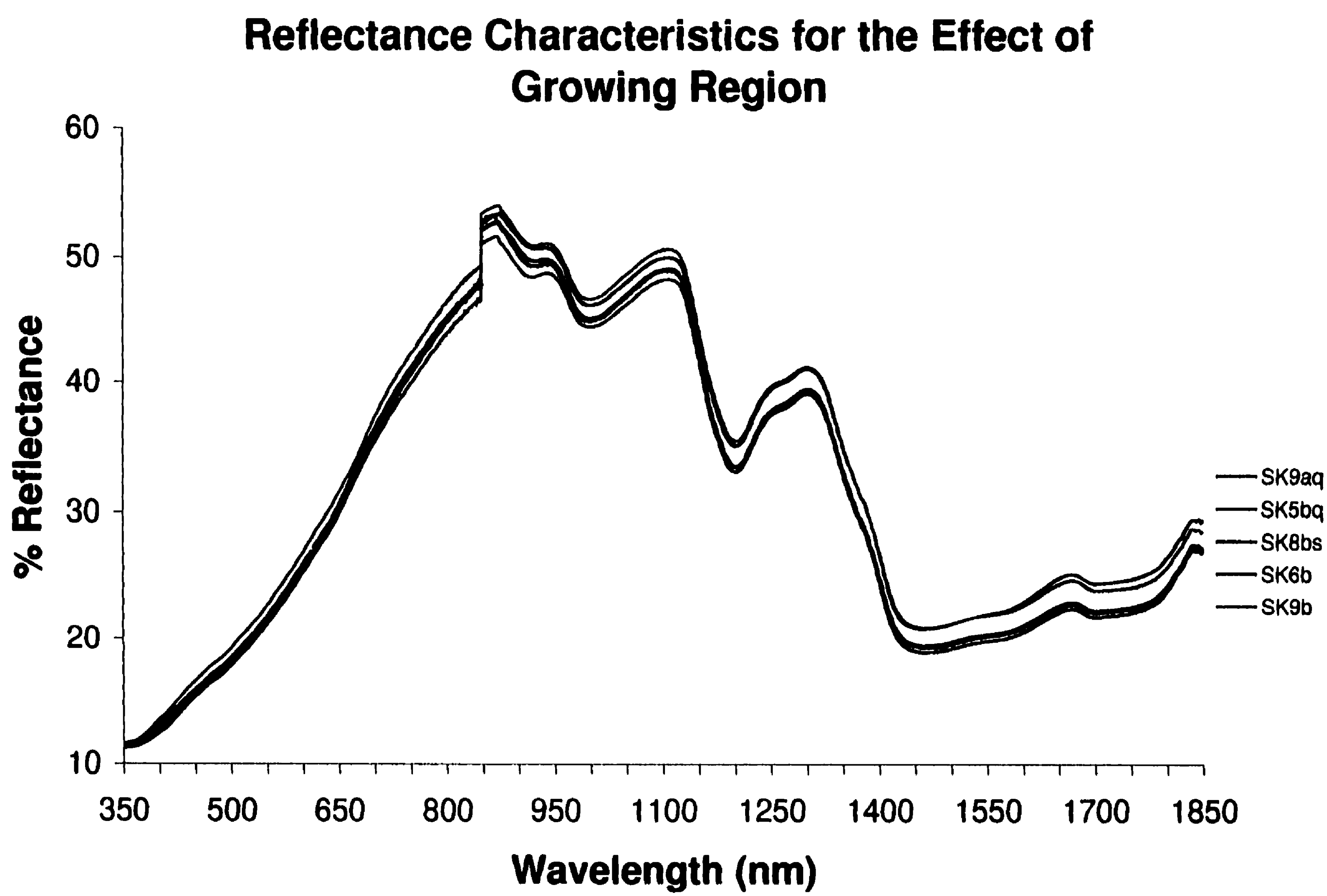

Figure B5b. Average of five replicates of five samples of CWRS wheat from Saskatchewan.

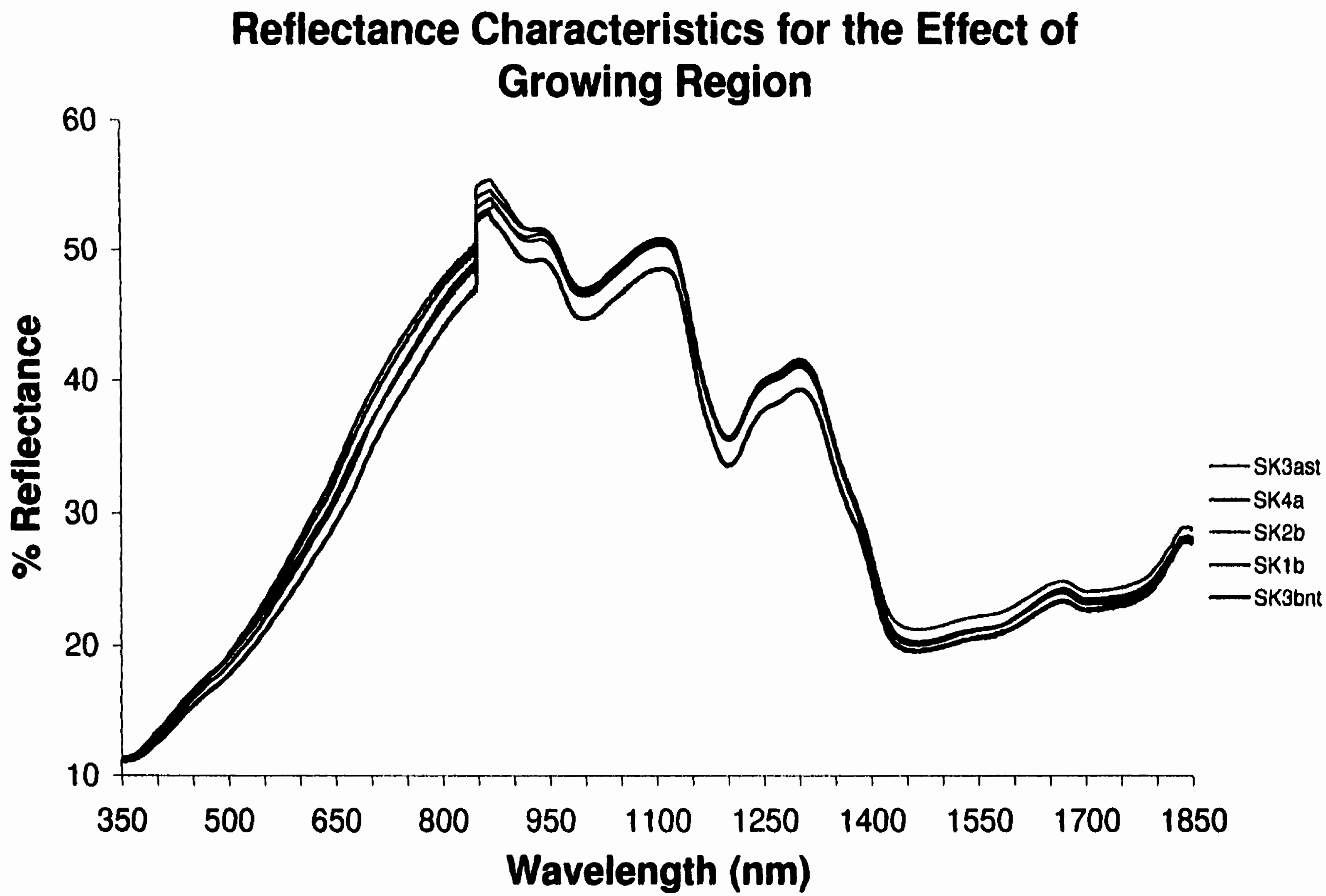

Figure B5c. Average of five replicates of five samples of CWRS wheat from Saskatchewan.

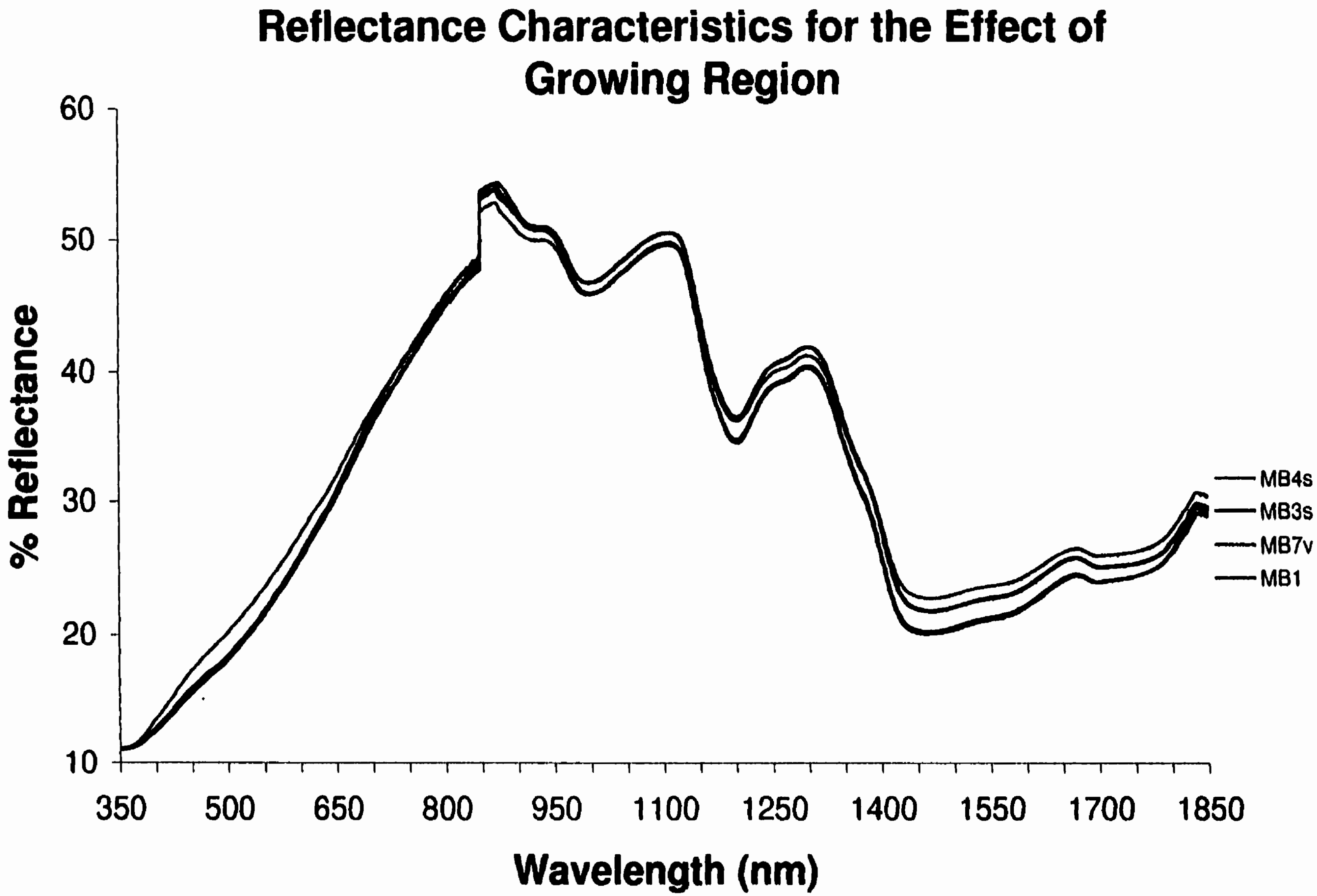

Figure B5d. Average of five replicates of each samples of CWRS wheat from Manitoba.

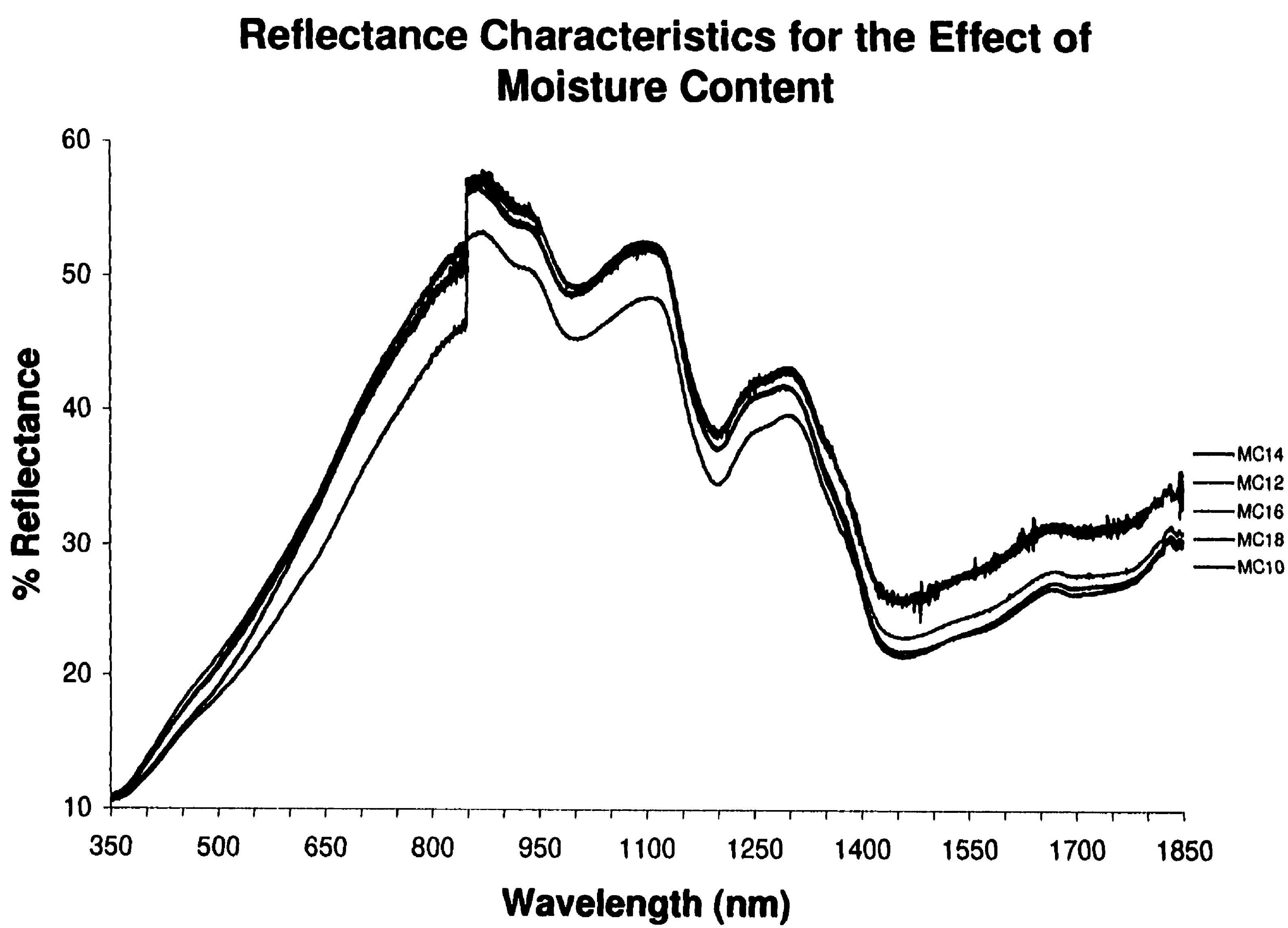

Figure B6. Average of five replicates of each CWRS wheat sample for different moisture contents.

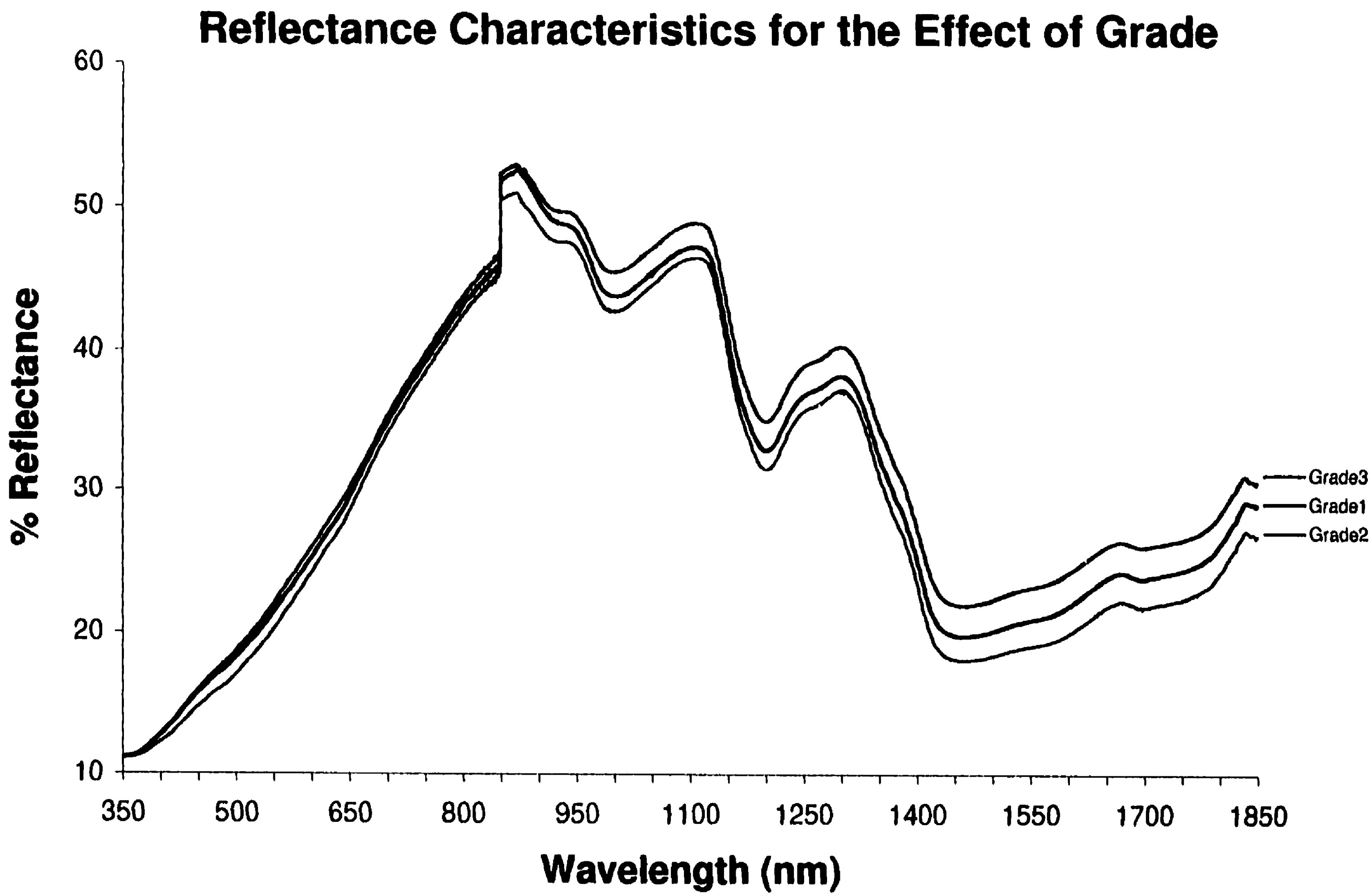

Figure B7. Average of five replicates of each CWRS wheat sample for different grades.

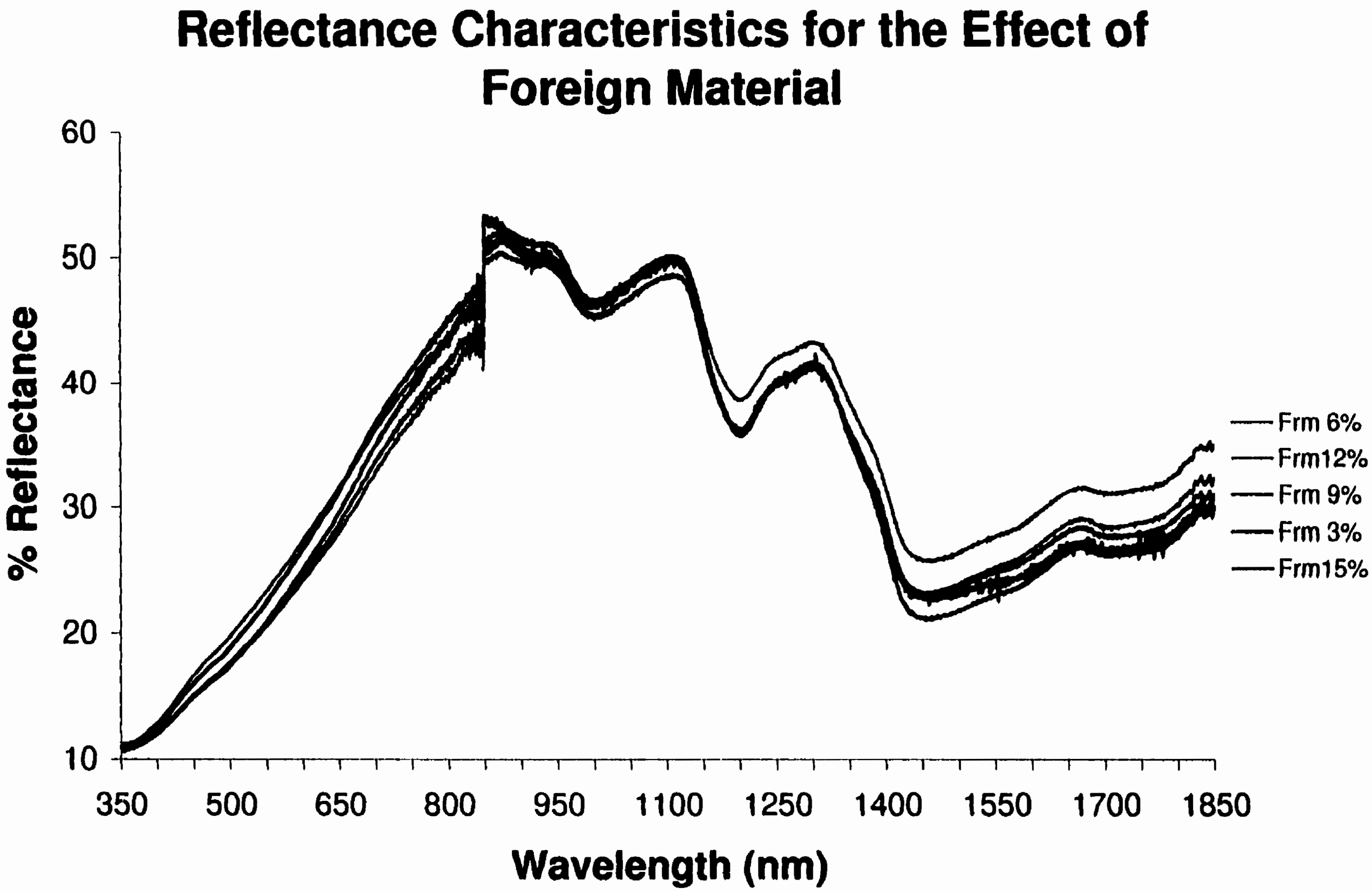

Figure B8. Average of five replicates of each CWRS wheat sample for different foreign material levels.

APPENDIX C: Classification Accuracies

Table C1a. Confusion matrix of the thirteen features model with eight classes for the hold-out method (Normal estimation) for cereals[1].

Class (to) → (from) ↓	2-row barley	6-row barley	CWRS wheat	CWAD wheat	buck-wheat	oats	rye	SWSW
2-row barley	5 (100%)	0	0	0	0	0	0	0
6-row barley	0	5 (100%)	0	0	0	0	0	0
CWRS wheat	0	0	5 (100%)	0	0	0	0	0
CWAD wheat	0	0	0	5 (100%)	0	0	0	0
buckwheat	0	0	0	0	5 (100%)	0	0	0
oats	0	0	0	0	0	5 (100%)	0	0
rye	0	0	0	0	0	0	5 (100%)	0
SWSW	0	0	0	1 (20%)	0	0	0	4 (80%)

[1]The row figures are the number of replicates (with percentages) identified with each column heading.

Table C1b. Confusion matrix of the thirteen features model with eight classes for the hold-out method (Non-parametric estimation) for cereals[1].

Class (to) → (from) ↓	2-row barley	6-row barley	CWRS wheat	CWAD wheat	buck-wheat	oats	rye	SWSW
2-row barley	5 (100%)	0	0	0	0	0	0	0
6-row barley	0	5 (100%)	0	0	0	0	0	0
CWRS wheat	0	0	5 (100%)	0	0	0	0	0
CWAD wheat	0	0	0	4 (80%)	0	0	0	1 (20%)
buckwheat	0	0	0	0	5 (100%)	0	0	0
oats	0	0	0	0	0	5 (100%)	0	0
rye	0	0	0	0	0	0	5 (100%)	0
SWSW	0	0	0	1 (20%)	0	0	0	4 (80%)

[1]The row figures are the number of replicates (with percentages) identified with each column heading.

Table C2. Confusion matrix of the thirteen features model with three classes for the hold-out method (Normal and Non-parametric estimation) for oilseeds[1].

Class (to) → (from) ↓	Canola	Flax	Sunflower
Canola	5 (100%)	0	0
Flax	0	5 (100%)	0
Sunflower	0	0	5 (100%)

[1]The row figures are the number of replicates (with percentages) identified with each column heading.

Table C3a. Confusion matrix of the thirteen features model with eight classes for the hold-out method (Normal estimation) for pulses[1].

Class (to) → (from) ↓	black beans	dark green spkld lentils	faba beans	green beans	lentils	pea beans	peas	pinto beans
black beans	5 (100%)	0	0	0	0	0	0	0
dark green speckled lentils	0	5 (100%)	0	0	0	0	0	0
faba beans	0	0	5 (100%)	0	0	0	0	0
green beans	0	0	0	4 (80%)	0	0	1 (20%)	0
lentils	0	0	0	0	5 (100%)	0	0	0
pea beans	0	0	0	1 (20%)	0	4 (80%)	0	0
peas	0	0	0	0	1 (20%)	0	4 (80%)	0
pinto beans	0	0	0	0	0	0	0	5 (100%)

[1]The row figures are the number of replicates (with percentages) identified with each column heading.

Table C3b. Confusion matrix of the thirteen features model with eight classes for the hold-out method (Non-parametric estimation) for pulses[1].

Class (to) → (from) ↓	black beans	dark green spckd lentils	faba beans	green beans	lentils	pea beans	peas	pinto beans	other
black beans	4 (80%)	1 (20%)	0	0	0	0	0	0	0
dark green speckled lentils	0	5 (100%)	0	0	0	0	0	0	0
faba beans	0	1 (20%)	4 (80%)	0	0	0	0	0	0
green beans	0	0	0	4 (80%)	0	0	1 (20%)	0	0
lentils	0	0	0	0	5 (100%)	0	0	0	0
pea beans	0	0	0	2 (40%)	0	3 (60%)	0	0	0
peas	0	0	0	1 (20%)	0	1 (20%)	3 (60%)	0	0
pinto beans	0	0	0	0	0	0	0	5 (100%)	0

[1]The row figures are the number of replicates (with percentages) identified with each column heading.

Table C4a. Confusion matrix of the thirteen features model with nineteen classes for the hold-out method (Normal estimation) for cereals, oilseeds, and pulses[1].

Class (to) → (from) ↓	2-row barley	6-row barley	CWRS wheat	CWAD wheat	buckwheat	oats	rye	SWSW	canola	flax	sunflower	black beans	dark green spkld lentils	faba beans	green beans	lentils	pea beans	peas	pinto beans
2-row barley	4 (80%)	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6-row barley	0	5 (100%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CWRS wheat	0	0	5 (100%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CWAD wheat	0	0	1 (20%)	4 (80%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
buckwheat	0	0	0	0	5 (100%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
oats	0	0	0	0	0	5 (100%)	0	0	0	0	0	0	0	0	0	0	0	0	0
rye	0	0	0	0	0	0	4 (80%)	0	0	0	0	0	0	0	0	1 (20%)	0	0	0
SWSW	0	0	1 (20%)	0	0	0	0	4 (80%)	0	0	0	0	0	0	0	0	0	0	0
canola	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0	0	0	0	0	0	0
flax	0	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0	0	0	0	0	0
sunflower	0	0	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0	0	0	0	0
black beans	0	0	0	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0	0	0	0

Table C4a continued. Confusion matrix of the thirteen features model with nineteen classes for the hold-out method (Normal estimation) for cereals, oilseeds, and pulses.

Class (to) → (from) ↓	2-row barley	6-row barley	CWRS wheat	CWAD wheat	buckwheat	oats	rye	SWSW	canola	flax	sunflower	black beans	dark green spkd lentils	faba beans	green beans	lentils	pea beans	peas	pinto beans
dark green spkld lentils	0	0	0	0	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0	0	0
faba beans	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0	0
green beans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3 (60%)	0	1 (20%)	1 (20%)	0
lentils	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100%)	0	0	0
pea beans	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	3 (60%)	0	0
peas	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3 (60%)	0
pinto beans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100%)

[1]The row figures are the number of replicates (with percentages) identified with each column variable.

Table C4b. Confusion matrix of the thirteen features model with nineteen classes for the hold-out method (Non-parametric estimation) for cereals, oilseeds, and pulses[1].

Class (to) → (from) ↓	2-row barley	6-row barley	CWRS wheat	CWAD wheat	buckwheat	oats	rye	SWSW	canola	flax	sunflower	black beans	dark green spkld lentils	faba beans	green beans	lentils	pea beans	peas	pinto beans	other
2-row barley	4 (80%)	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6-row barley	0	5 (100%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CWRS wheat	0	0	5 (100%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CWAD wheat	0	0	1 (20%)	3 (60%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)
buckwheat	0	0	0	0	5 (100%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
oats	0	0	0	0	0	5 (100%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
rye	0	0	0	0	0	0	4 (80%)	0	0	0	0	0	0	0	0	1 (20%)	0	0	0	0
SWSW	0	0	0	0	0	0	0	4 (80%)	0	0	0	0	0	0	0	0	0	0	0	1 (20%)
canola	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0	0	0	0	0	0	0	0
flax	0	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0	0	0	0	0	0	0
sunflower	0	0	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0	0	0	0	0	0
black beans	0	0	0	0	0	0	0	0	0	0	0	4 (80%)	0	0	0	0	0	0	0	1 (20%)

Table C4b continued. Confusion matrix of the thirteen features model with nineteen classes for the hold-out method (Non-parametric estimation) for cereals, oilseeds, and pulses.

Class (to) → (from) ↓	2-row barley	6-row barley	CWRS wheat	CWAD wheat	buckwheat	oats	rye	SWSW	canola	flax	sunflower	black beans	dark green spkld lentils	faba beans	green beans	lentils	pea beans	peas	pinto beans	other
dark green spckled lentils	0	0	0	0	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0	0	0	0
faba beans	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0	0	0
green beans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2 (40%)	0	2 (40%)	1 (20%)	0	0
lentils	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100%)	0	0	0	0
pea beans	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	3 (60%)
peas	2 (40%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3 (60%)	0	0
pinto beans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100%)	0

[1]The row figures are the number of replicates (with percentages) identified with each column variable.

Table C5a. Confusion matrix of the thirteen features model with twenty-seven classes for the hold-out method (Normal estimation) for specialty seeds[1].

Class (to) → (from) ↓	alfalfa	annual ryegrass	alsike clover	brown mustard	bromegrass	canary seed	crested wheatgrass	creeping red fescue	creeping bentgrass	crown millet	dahurian wild rye	meadow fescue	intermd wheatgrass	Kentucky bluegrass	meadow brome	orchard grass	oriental mustard	perennial ryegrass	red clover	reed canary grass	Russian wild rye	sorghum sudangrass	sweet clover	tall wheatgrass	timothy	birdsfoot trefoil	yellow mustard
alfalfa	4 (80%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0
annual ryegrass	0	4 (80%)	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
alsike clover	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
brown mustard	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bromegrass	0	0	0	0	2 (40%)	0	0	1 (20%)	0	0	1 (20%)	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0
canary seed	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
crested wheatgrass	0	0	0	0	0	0	4 (80%)	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0
creeping red fescue	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
creeping bentgrass	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
crown millet	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
dahurian wild rye	0	1 (20%)	0	0	0	0	0	0	0	0	2 (40%)	1 (20%)	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	0
meadow fescue	0	0	0	0	0	0	0	0	0	0	1 (20%)	4 (80%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
intermediate wheatgrass	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kentucky bluegrass	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0

Table C5a continued. Confusion matrix of the thirteen features model with twenty-seven classes for the hold-out method (Normal estimation) for specialty seeds.

Class (to) → (from) ↓	alfalfa	annual ryegrass	alsike clover	brown mustard	bromegrass	canary seed	crested wheatgrass	creeping red fescue	creeping bentgrass	crown millet	dahurian wild rye	meadow fescue	intermd wheatgrass	Kentucky bluegrass	meadow brome	orchard grass	oriental mustard	perennial ryegrass	red clover	reed canary grass	Russian wild rye	sorghum sudangrass	sweet clover	tall wheatgrass	timothy	birdsfoot trefoil	yellow mustard
meadow brome	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	4 (80%)	0	0	0	0	0	0	0	0	0	0	0	0
orchard grass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0
oriental mustard	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0
perennial ryegrass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0
red clover	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0
reed canary grass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0
Russian wild rye	0	0	0	0	0	0	2 (40%)	0	0	0	0	0	0	0	0	0	0	0	0	0	3 (60%)	0	0	0	0	0	0
sorghum sudangrass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0
sweet clover	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0
tall wheatgrass	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4 (80%)	0	0	0
timothy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0
birdsfoot trefoil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0
yellow mustard	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)

[1]The row figures are the number of replicates (with percentages) identified with each column heading.

Table C5b. Confusion matrix of the thirteen features model with twenty-seven classes for the hold-out method (Non-parametric estimation) for specialty seeds[1].

Class (to) → (from) ↓	alfalfa	annual ryegrass	alsike clover	brown mustard	bromegrass	canary seed	crested wheatgrass	creeping red fescue	creeping bentgrass	crown millet	dahurian wild rye	meadow fescue	intermd wheatgrass	Kentucky bluegrass	meadow brome	orchard grass	oriental mustard	perennial ryegrass	red clover	reed canary grass	Russian wild rye	sorghum sudangrass	sweet clover	tall wheatgrass	timothy	birdsfoot trefoil	yellow mustard	other
alfal	4 (80%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	0
anurg	0	4 (80%)	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
askcv	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bnmtd	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
brome	0	0	0	0	1 (20%)	0	0	1 (20%)	0	0	0	0	0	0	2 (40%)	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)
cnrsd	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
crest	0	0	0	0	0	0	3 (60%)	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	1 (20%)
crfes	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
crpbg	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
crwml	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
drrye	0	0	0	0	0	0	0	0	0	0	2 (40%)	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	0	2 (40%)
fescu	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
intwg	0	0	0	0	0	0	0	0	0	0	0	0	4 (80%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)
kenbg	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table C5b continued. Confusion matrix of the thirteen features model with twenty-seven classes for the hold-out method (Non-parametric estimation) for specialty seeds.

Class (to) → (from) ↓	alfalfa	annual ryegrass	alsike clover	brown mustard	bromegrass	canary seed	crested wheatgrass	creeping red fescue	creeping bentgrass	crown millet	dahurian wild rye	meadow fescue	intermd wheatgrass	Kentucky bluegrass	meadow brome	orchard grass	oriental mustard	perennial ryegrass	red clover	reed canary grass	Russian wild rye	sorghum sudangrass	sweet clover	tall wheatgrass	timothy	birdsfoot trefoil	yellow mustard	other
mdw brome	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4 (80%)	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)
orchrd grass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0	0
orientl mustd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0	0
perenl ryegrs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0	0
red clover	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0	0
rd can grass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0	0
Russn w rye	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0	0
srghm sdngrs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0	0
sweet clover	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0	0	0
tall whtgr	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2 (40%)	0	0	0	2 (40%)
timthy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0	0
brdsft trefoil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0	0
yellw mustd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100)	0

[1]The row figures are the number of replicates (with percentages) identified with each column heading.

Table C6a. Confusion matrix of the thirteen features model with twenty classes for the hold-out method (Normal estimation) of CWRS wheat growing regions[1].

Class (to) → (from) ↓	Ab2s	Ab3s	Ab4s	Ab5s	Ab6	Ab7	Mb1	Mb3s	Mb4s	Mb7v	Sk1b	Sk2b	Sk3ast	Sk3bnt	Sk4a	Sk5bq	Sk6b	Sk8b	Sk9aq	Sk9b
Ab[1]2s	1 (20%)	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	1 (20%)	0	0	0	1 (20%)	0	1 (20%)
Ab3s	0	0	1 (20%)	1 (20%)	0	0	0	0	0	0	0	1 (20%)	0	0	1 (20%)	0	0	0	1 (20%)	0
Ab4s	0	0	1 (20%)	0	0	0	1 (20%)	0	0	0	0	0	1 (20%)	0	1 (20%)	0	1 (20%)	0	0	0
Ab5s	1 (20%)	1 (20%)	0	0	0	0	0	0	1 (20%)	0	0	0	2 (40%)	0	0	0	0	0	0	0
Ab6	1 (20%)	0	0	0	4 (80%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ab7	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	2 (40%)	0	1 (20%)	0	1 (20%)	0
Mb[2]1	1 (20%)	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	2 (40%)	0	0	0	0	1 (20%)	0
Mb3s	0	0	1 (20%)	0	0	0	0	2 (40%)	0	0	0	0	0	0	0	1 (20%)	0	0	1 (20%)	0
Mb4s	0	0	1 (20%)	0	0	0	0	0	1 (20%)	0	0	1 (20%)	0	0	0	1 (20%)	1 (20%)	0	0	0
Mb7v	2 (40%)	0	1 (20%)	0	0	0	1 (20%)	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0
Sk[3]1b	1 (20%)	2 (40%)	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	1 (20%)

Table C6a continued. Confusion matrix of the thirteen features model with twenty classes for the hold-out method (Normal estimation) of CWRS wheat growing regions.

Class (to) → (from)↓	Ab2s	Ab3s	Ab4s	Ab5s	Ab6	Ab7	Mb1	Mb3s	Mb4s	Mb7v	Sk1b	Sk2b	Sk3ast	Sk3bnt	Sk4a	Sk5bq	Sk6b	Sk8b	Sk9aq	Sk9b
Sk2b	1 (20%)	0	0	0	0	0	0	0	0	0	1 (20%)	1 (20%)	0	0	1 (20%)	0	0	1 (20%)	0	0
Sk3ast	0	0	1 (20%)	1 (20%)	0	0	0	0	1 (20%)	0	0	0	0	0	0	1 (20%)	0	0	0	1 (20%)
Sk3bnt	2 (40%)	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	1 (20%)	0	0	1	0
Sk4a	0	0	1 (20%)	0	0	1 (20%)	0	0	0	0	0	1 (20%)	0	1 (20%)	0	0	1 (20%)	0	0	0
Sk5bq	0	0	0	1 (20%)	0	0	0	0	0	0	0	1 (20%)	0	0	0	1 (20%)	1 (20%)	0	0	1 (20%)
Sk6b	0	0	0	0	0	1 (20%)	0	0	1 (20%)	0	0	0	0	0	1 (20%)	0	1 (20%)	1 (20%)	0	0
Sk8b	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	0	1 (20%)	1 (20%)	0	2 (40%)
Sk9aq	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	4 (80%)	0
Sk9b	1 (20%)	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	2 (40%)	0	1 (20%)	0	0	0	0

[1]The row figures are the number of replicates (with percentages) identified with each column heading. Ab=Alberta, Mb=Manitoba, Sk=Saskatchewan

Table C6b. Confusion matrix of the thirteen features model with twenty classes for the hold-out method (Non-parametric estimation) of CWRS wheat growing regions[1].

Class (to) → (from) ↓	Ab2s	Ab3s	Ab4s	Ab5s	Ab6	Ab7	Mb1	Mb3s	Mb4s	Mb7v	Sk1b	Sk2b	Sk3ast	Sk3bnt	Sk4a	Sk5bq	Sk6b	Sk8b	Sk9aq	Sk9b	Other
Ab2s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100%)
Ab3s	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	1 (20%)	0	0	1 (20%)	0	2 (40%)
Ab4s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5 (100%)
Ab5s	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4 (80%)
Ab6	0	0	0	0	4 (80%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)
Ab7	0	0	0	0	0	2 (40%)	0	0	0	0	0	0	0	0	1 (20%)	0	1 (20%)	0	0	0	1 (20%)
Mb1	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	1 (20%)	0	0	3 (60%)
Mb3s	0	0	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	3 (60%)
Mb4s	0	0	0	0	0	0	0	0	1 (20%)	0	0	1 (20%)	0	0	0	0	0	0	0	0	3 (60%)
Mb7v	1 (20%)	0	1 (20%)	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	1 (20%)
Sk1b	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	3 (60%)

Table C6b continued. Confusion matrix of the thirteen features model with twenty classes for the hold-out method (Non-parametric estimation) of CWRS wheat growing regions.

Class (to) → (from) ↓	Ab2s	Ab3s	Ab4s	Ab5s	Ab6	Ab7	Mb1	Mb3s	Mb4s	Mb7v	Sk1b	Sk2b	Sk3ast	Sk3bnt	Sk4a	Sk5bq	Sk6b	Sk8b	Sk9aq	Sk9b	Other
Sk2b	0	0	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	1 (20%)	0	0	3 (60%)
Sk3ast	1 (20%)	0	1 (20%)	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	1 (20%)	0	0	1 (20%)
Sk3bnt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	4 (80%)
Sk4a	0	0	1 (20%)	0	0	1 (20%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3 (60%)
Sk5bq	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	0	0	0	0	0	0	0	0	4 (80%)
Sk6b	0	0	0	0	0	0	0	0	0	1 (20%)	0	1 (20%)	0	0	0	0	0	0	0	0	3 (60%)
Sk8b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	1 (20%)	0	3 (60%)
Sk9aq	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (20%)	2 (40%)	0	2 (40%)
Sk9b	0	0	0	0	0	0	0	0	0	0	1 (20%)	1 (20%)	0	1 (20%)	0	0	0	0	0	0	2 (40%)

[1]The row figures are the number of replicates (with percentages) identified with each column heading. Ab=Alberta, Mb=Manitoba, Sk=Saskatchewan

Table C7a. Confusion matrix of the thirteen features model with five classes for the hold-out method (Normal estimation) for moisture content of CWRS wheat[6].

Class (to) → (from) ↓	MC10	MC12	MC14	MC16	MC18
MC10[1]	2 (40%)	2 (40%)	1 (20%)	0	0
MC12[2]	0	4 (80%)	1 (20%)	0	0
MC14[3]	2 (40%)	2 (40%)	1 (20%)	0	0
MC16[4]	0	0	0	3 (60%)	2 (40%)
MC18[5]	0	1 (20%)	0	2 (40%)	2 (40%)

Table C7b. Confusion matrix of the thirteen features model with five classes for the hold-out method (Non-parametric estimation) for moisture content of CWRS wheat[6].

Class (to) → (from) ↓	MC10	MC12	MC14	MC16	MC18
MC10[1]	2 (40%)	2 (40%)	1 (20%)	0	0
MC12[2]	1 (20%)	4 (80%)	0	0	0
MC14[3]	2 (40%)	2 (40%)	1 (20%)	0	0
MC16[4]	0	0	0	3 (60%)	2 (40%)
MC18[5]	0	1 (20%)	0	2 (40%)	2 (40%)

[1]10% moisture content, [2]12% moisture content, [3]14% moisture content, [4]16% moisture content, [5]18% moisture content, [6]The row figures are the number of replicates (with percentages) identified with each column heading.

Table C8. Confusion matrix of the thirteen features model with three classes for the hold-out method (Normal and Non-parametric estimation) for grade of CWRS wheat[1].

Class (to) → (from) ↓	Grade1	Grade2	Grade3
Grade1	3 (60%)	0	2 (40%)
Grade2	2 (40%)	3 (60%)	0
Grade3	3 (60%)	0	2 (40%)

[1]The row figures are the number of replicates (with percentages) identified with each column heading.

Table C9a. Confusion matrix of the thirteen features model with five classes for the hold-out method (Normal estimation) for foreign material in CWRS wheat[6].

Class (to) → (from) ↓	FM3	FM6	FM9	FM12	FM15
FM3[1]	3 (60%)	0	2 (40%)	0	0
FM6[2]	0	1 (20%)	1 (20%)	1 (20%)	2 (40%)
FM9[3]	2 (40%)	1 (20%)	1 (20%)	1 (20%)	0
FM12[4]	1 (20%)	0	3 (60%)	0	1 (20%)
FM15[5]	1 (20%)	4 (80%)	0	0	0

Table C9b. Confusion matrix of the thirteen features model with five classes for the hold-out method (Non-parametric estimation) for foreign material in CWRS wheat[6].

Class (to) → (from) ↓	FM3	FM6	FM9	FM12	FM15	Other
FM3[1]	2 (40%)	0	1 (20%)	0	0	2 (40%)
FM6[2]	0	0	1 (20%)	1 (20%)	3 (60%)	0
FM9[3]	2 (40%)	0	1 (20%)	1 (20%)	0	1 (20%)
FM12[4]	1 (20%)	0	2 (40%)	0	1 (20%)	1 (20%)
FM15[5]	1 (20%)	3 (60%)	0	0	1 (20%)	0

[1]3% foreign material, [2]6% foreign material, [3]9% foreign material, [4]12% foreign material, [5]15% foreign material ,[6]The row figures are the number of replicates (with percentages) identified with each column heading.