ISBN 978-82-575-1062-6
ISSN 1503-1667

Bioforsk Øst, Apelsvoll -
The Norwegian Institute for Agricultural and
Environmental Research Arable Crops Division, Nylinna 226
NO-2849 Kapp, Norway
Phone: +47 40 60 41 00
www.bioforsk.no, e-mail: apelsvoll@bioforsk.no

Norwegian University of Life Sciences
NO–1432 Ås, Norway
Phone +47 64 96 50 00
www.umb.no, e-mail: postmottak@umb.no

Stein Ivar Øvergaard

Norwegian University of Life Sciences • Universitetet for miljø- og biovitenskap
Department of Mathematical Sciences and Technology
Philosophiae Doctor (PhD) Thesis 2012:26

Philosophiae Doctor (PhD) Thesis 2012:26

Prediction of spring wheat yield and grain quality with remote VIS-NIR spectroscopy and multivariate data analysis

Prediksjon av avling og kvalitet i vårhvete ved hjelp av VIS-NIR spektroskopi og multivariat dataanalyse

Stein Ivar Øvergaard

Prediction of spring wheat yield and grain quality with remote VIS-NIR spectroscopy and multivariate data analysis

Prediksjon av avling og kvalitet i vårhvete ved hjelp av VIS-NIR spektroskopi og multivariat dataanalyse

Philosophiae doctor (PhD) avhandling

Stein Ivar Øvergaard

Bioforsk Øst Apelsvoll

Institutt for matematiske realfag og teknologi
Universitetet for miljø- og biovitenskap

Ås 2012

Avhandling nr 2012:26
ISSN 1503-1667
ISBN 978-82-575-1062-6

PhD evaluation committee:

Associate professor Ulf Geir Indahl (Committee Administrator)
Department of Mathematical Sciences and Technology,
Norwegian University of Life Sciences
Postbox 5003
NO-1432 Ås, Norway
E-mail: ulf.indahl@umb.no

Associate professor Bo Stenberg
Department of Soil and Environment,
Swedish University of Agricultural Sciences
Box 234
SE-532 23 SKARA
E-mail: bo.stenberg@mark.slu.se

Senior Researcher Dr. Vegard Segtnan
Nofima Mat
Postboks 210
NO-1431 Ås, Norway
E-mail: vegard.segtnan@nofima.no

PhD supervisors:

Associate professor Knut Kvaal
Department of Mathematical Sciences and Technology,
The Norwegian University of Life Sciences
Postbox 5003
NO-1432 Ås, Norway
E-mail: knut.kvaal@umb.no

Professor Tomas Isaksson
Department of Chemistry, Biotechnology and Food Science,
The Norwegian University of Life Sciences
Postbox 5003
NO-1432 Ås, Norway
E-mail: tomas.isaksson@umb.no

Researcher Dr. Audun Korsæth
Arable Crops Division
The Norwegian Institute for Agricultural and Environmental Research
NO-2849 Kapp, Norway
E-mail: audun.korsaeth@bioforsk.no

Acknowledgements

I wish to thank my supervisors Audun Korsaeth, Tomas Isaksson and Knut Kvaal for all their enthusiasm, encouragement and for helpful discussions. Your help was invaluable for completing this thesis.

The technical staff at Apelsvoll research center is greatly acknowledged for their high technical skill and their willingness to help me. Without you, all the field work would have been impossible to carry out. I will particularly express my thanks to Torkel Gaardløs for excellent technical skills and high working spirit.

The remaining staff at Apelsvoll research center, I greatly acknowledge for always supporting me and for providing a friendly and productive working environment.

My sincere gratitude goes to the research group of Pierre Dardenne at the Waloon Center for Agricultural Research (CRA-W) in Gembloux, Belgium, where I stayed for four weeks during 2009. Special thanks to Juan Antonio Fernández–Pierna and Vincent Baeten for giving me time to discuss with you and initiate the cooperation that led to paper II within this thesis.

The current PhD students at Apelsvoll deserve the warmest thanks for numerous discussions and problem-solving coffee breaks.

Thanks to Hugh Riley and Wendy Waalen for the proof reading of this thesis and several manuscripts.

Above all, I wish to thank my wife Guro, my children Gunhild, Mikael and Tarjei, my parents Mikael and Synnøve for all this support and understanding during the work with this thesis. Thanks to my aunt Kirsten who kindly let me stay in her house during my work periods at UMB in Ås.

Kapp, 16. April 2012 Stein Ivar Øvergaard

Contents

Papers I-IV (individual page numbering)

Abstract

Along with an increasing population, both globally and locally, demands on agriculture increase. A sustainable increase in cereal yields is particularly focused and in this context, precision agriculture is becoming increasingly important. Precision agriculture encompasses the use of present technology to tailor the treatment of agricultural crops to time- and site-specific conditions. Within the field of precision agriculture, prognosis of cereal yields and cereal quality is an important topic. Such prognoses can produce yield maps, which can be used to identify problem areas and to make plans for the next cropping season. If the prognosis can be carried out early in the season and at a sufficiently large scale, it can also be of interest for the agricultural authorities. This thesis presents improvements in the computation of prediction models for yield and protein in spring wheat (*Triticum aestivum L.)* by means of spectroscopy and multivariate data analysis contained in four papers. Paper I was a comparative study of three spectrometers in terms of the instruments ability to give good prediction models of grain yield and grain protein concentration in spring wheat. In the period 2007-2010, seven field experiments were carried out with altogether 976 plots of spring wheat. Spectra from all plots were measured five times during the cropping season at Zadoks (Z) stages Z31, Z59, Z65, Z87 og Z90. All plots were harvested individually and cereal samples were analyzed for yield and protein in the laboratory. Z65 was the most favourable time to take measurements for prediction of yield and protein. The Powered Partial Least Squares method (PPLS) was used to compute regression models between spectroscopic data and cereal analysis. This method turned out to have strong ability for variable selection, and in paper II, we chose to perform a comparative study on PPLS against other recent and established variable selection methods. The

study showed that variable selection had no effect on larger calibration sets (> approx. 150 samples), whereas for a smaller dataset, variable selection can potentially improve the predictive ability. PPLS showed an advantage over the other methods in this regard. The objective of paper III was to assess the robustness of yield and protein prediction over several sites and years, in order to compute a practically useful model that gives consistent predictions under real-life situations. However, approximately 10 % of the dataset suffered from severe lodging and these plots completely destroyed the predictive ability of the models. Hence, a classification model was computed on the spectra with Partial Least Squares Discriminant Analysis (PLS-DA), which was able to classify a spectral measurement in the classes “lodging” or “standing crop”. After removal of the lodged plots, spectra were conveyed to further analyses. There were large bias and skewness problems in the model when validated over several years and a single model comprising all years was impossible to compute. However, the PPLS models performed significantly better than traditional methods based on vegetation indices. In paper IV, the yield prediction models of paper III were augmented with meteorological data. The data fusion was performed with Canonical Powered Partial Least Squares (CPPLS) and 3 of 4 models gave significantly improved predictive ability compared with the models based on spectroscopic data alone. Hence, we have shown that fusion of spectroscopic and meteorological data can be a feasible way of computing practically useable yield prediction models.

Sammendrag

Sammen med en voksende befolkning både nasjonalt og globalt øker kravene til landbruket, spesielt i form av bærekraftig økning av kornavlinger. I denne sammenhengen kommer presisjonsjordbruk til å få økt betydning i årene som kommer. Presisjonsjordbruk innebærer at man bruker den til enhver tid tilgjengelige teknologi til å gjøre tids- og stedstilpasset behandling av jordbruksvekster. Innenfor presisjonsjordbruk er beregning av gode prognoser for kornavling og -kvalitet en viktig oppgave. Avlingskskart utarbeides fra slike prognoser og kan brukes til å kartlegge problemområder samt planlegge neste års vekstsesong. Hvis man kan beregne avlingsprognosene tidlig nok og i stor nok skala vil det også kunne ha interesse som beslutningsstøtteverktøy på myndighetsnivå. Denne avhandlingen presenterer forbedringer innen beregning av avlingsprognoser ved hjelp av spektrometri og multivariat dataanalyse gjennom fire artikler. Artikkel I presenterer en sammenlikenede studie av tre spektrometere sin evne til å gi gode avlings- og proteinmodeller i vårhvete (*Triticum aestivum L.)*. I løpet av perioden 2007-2010 ble det anlagt til sammen 7 feltforsøk med i alt 976 ruter. Spektra fra disse rutene ble målt fem ganger i hver sesong på Zadoks (Z) stadier Z31, Z59, Z65, Z87 og Z90. Alle ruter ble tresket og kornprøvene ble analysert for avling og protein i laboratorium. Z65 viste seg som det mest gunstige tidspunktet for innsamling av data for avling- og proteinprognoser. Powered Partial Least Squares (PPLS) ble valgt til å beregne regresjonsmodeller mellom spektroskopiske data og avling og protein. Denne metoden viste seg å ha sterke egenskaper innen variabelseleksjon. Derfor valgte vi i artikkel II å utføre en sammenliknende studie mellom PPLS og nyere samt etablerte variabelseleksjonsmetoder. Vi fant at variabelseleksjon ikke har noen postiv effekt i store datasett (> ca. 150 prøver), men i mindre datasett kan variabelseleksjon gi en signifikant forbedring i prediksjonsevnen og PPLS

viste seg som den sterkeste metoden i denne sammenhengen. Artikkel III tok sikte på å evaluere robustheten til avlings- og proteinmodellene over flere år for kunne gi en praktisk brukbar modell som gir gode prediksjoner under virkelige forhold. Ca. 10 % av det originale datasettet hadde imidlertid legde som ødela prediksjonsmodellene fullstendig. Derfor ble det beregnet en klassifikasjonsmodell med Partial Least Squares Discriminant Analysis (PLS-DA) som kan klassifisere en måling i klassene «legde» eller «stående vekst» kun ut fra spektral informasjon. Klassifikasjonsmodellen ble kjørt som et innledende steg til avlings- og proteinprediksjonen. Det viste seg imidlertid at det også var store nivåforskjeller mellom år, og en enkelt modell som dekket alle fire sesonger (2007-2010) viste seg umulig å beregne. Likevel ga PPLS-modellene signifikant bedre prediksjonsevne enn tradisjonelle modeller basert på vegetasjonsindekser. I artikkel IV ble avlingsmodellene fra tredje artikkel utvidet med meteorologiske data. For å kombinere de spektroskopiske og meteorologiske data ble Canonical Powered Partial Least Squares (CPPLS) valgt til å beregne modellene. Ved å beregne modeller med værdata inkludert ble 3 av 4 avlingsmodeller signifikant forbedret og skjevheter i modellene ble betydelig redusert. Vi har dermed vist at kombinasjon av værdata og spektroskopiske data er en praktisk brukbar metode for å beregne avlingsprognoser.

List of papers

Paper I

S. I. Overgaard, T. Isaksson, K. Kvaal, A. Korsaeth, Comparisons of two hand-held, multispectral field radiometers and a hyperspectral airborne imager in terms of predicting spring wheat grain yield and quality by means of powered partial least squares regression". Journal of Near Infrared Spectroscopy. **18,** 247 (2010)

Paper II

S. I. Overgaard, J. A. Fernández-Pierna, V. Baeten, P. Dardenne, T. Isaksson, Prediction error improvements using variable selection on small calibration sets - a comparison of some recent methods. Manuscript submitted for publication in Journal of Near Infrared Spectroscopy.

Paper III

S. I. Overgaard, A. Korsaeth, and T. Isaksson, "Prediction of wheat yield and protein using remote sensors on grain fields - Part I: Assessing NIR model robustness for year and site variations. Manuscript submitted for publication in Journal of Near Infrared Spectroscopy.

Paper IV

S. I. Overgaard, T. Isaksson, and A. Korsaeth, "Prediction of wheat yield and protein using remote sensors on grain fields - Part II: Improving prediction ability using data fusion. Manuscript submitted for publication in Journal of Near Infrared Spectroscopy.

1. Introduction

1.1 Precision agriculture in cereal production

Demands on cereal production increase steadily and are mainly driven by increased focus on environmental impact as well as the increasing global need for cereal products. In order to increase agricultural yields in a sustainable way, precision agriculture is an important subject. One widely used definition of precision agriculture is to use the most suitable technology to adjust the treatment of soil and crops to within-field variation. The goal of this optimization is to increase the efficiency of the input factors and thereby achieve better product quality with less variation, improved economic payoff and reduced negative environmental impact.

A crucial prerequisite for the utilization of technology in the above described fashion is the availability of Global Navigtion Satellite System (GNSS). Without the development of small, light and affordable GNSS receivers with high geometrical accuracy, precision agriculture would not be a feasible crop production tool.

To gather data about soil and crop in order to estimate relevant agricultural information, various instruments and instrument platforms have been used to optimize the production of cereals.

The best known and probably most technologically advanced branch of precision agriculture is the site-specific application of fertilizer (e.g. Lukina et al. 2001; Berntsen et al. 2006; Korsaeth et al. 2006; Reyniers et al. 2006; Zillmann et al. 2006; Jorgensen et al. 2007; Sripada et al. 2007).

There are several commercially available systems for site-specfic application of fertilizer, most of which gather data from optical instruments to estimate the crop fertilizer requirements. The scientific basis for such use of optical instruments is the relation between chlorophyll content in the leaves and the nitrogen requirement of the plant. The chlorophyll content in the leaves can be estimated by measurements of relevant wavelengths of the electromagnetic spectrum (e.g. Gitelson et al. 1996).

Estimation of soil and crop properties is another important branch of precision agriculture. Predictions of properties such as grain yield, quality, diseases and weeds can be input for a Geographical Information System (GIS) in order to assist planning of future cropping seasons, as well as establishing crop management zones according to heterogeneities in crop and soil. Several studies have reported successful predictions of various soil and plant variables. Hand-held electromagnetic instruments have been used for the estimation of crop and soil properties (e.g. Korsaeth 2005; Pettersson et al. 2006). Hand-held, airborne and even space-borne optical instruments have been used to estimate plant properties at different scales such as yield estimation from consumer-grade (Red Green Blue) RGB-camera (e.g. Reyniers et al. 2004; Jensen et al. 2007), yield and protein estimation from multi- and hyperspectral cameras (e.g. Hansen et al. 2002; Basnyat et al. 2005; Liu et al. 2006; Pettersson et al. 2006; Reyniers et al. 2006; Jensen et al. 2007; Xue et al. 2007; Feng et al. 2008), detection of plant stress (e.g. Karimi et al. 2005; Behrens et al. 2007; Jorgensen et al. 2007; Mirik et al. 2007), detection of cereal diseases (e.g. Huang et al. 2007), application of pesticide (e.g. Berge et al. 2007) as well as future prediction and forecasting of yields and quality (e.g. Liu et al. 2006; de Wit et al. 2007).

1.2 Current methods for prediction of plant properties

The predominant technique for prediction of plant properties appears to be the analysis of measured plant canopy reflectance by means of various spectroscopic instruments, both imaging and non-imaging. Two main ways of dealing with reflectance data have been developed in recent decades. Some studies are based on model inversion, which aims to reproduce the radiation field that created the reflectance data by solving the Radiative Transfer Equation (RTE) og using Kubelka-Munk theory (e.g. Atzberger 2004; Houborg et al. 2007). The solution of the radiation equations is rather complex for crop canopies and this approach is rarely considered for real-time applications. The other main branch in reflectance data analysis is the statistical-empirical approach that utilizes two datasets, one for calibrating a prediction model and another independent dataset to validate the model. The generalization of such models is dependent on how much variation is represented in the calibration and validation datasets. Hence, this approach needs large datasets in order to give models that can be generalized.

Originally developed in the field of remote sensing based on satellite images with a limited number of wavelengths, vegetation indices have found use in the topic of plant property prediction. A Vegetation Index (VI) is a simple algebraic operation performed on a few (usually one to three) wavelengths and the goal of this operation is to compute a new variable that contains more information on the variables to be predicted (the response). The most famous and widely used vegetation index is the Normalized Difference Vegetation Index (NDVI, Rouse et al. 1974). Through the years, numerous vegetation indices have been developed in order to predict various plant properties. Examples of widely used indices are Difference Vegetation Index (DVI, Jordan 1969), Green Normalized Difference Vegetation Index (GNDVI, Gitelson et al. 1996), Green Ratio

Vegetation Index (GRVI, Tucker 1979), Modified Soil Adjusted Vegetation Index (MSAVI, Qi et al. 1994), Red Edge Inflection Point (REIP, Guyot et al. 1988) and Derived Chlorophyll Concentration a and b Index (D-chl-ab, Gitelson et al. 1996). Several studies compute one relevant vegetation index, which serves as the input variable in a univariate linear regression (LR), in order to compute the final prediction model for the plant property of interest (e.g. grain yield, Wang et al. 2004; Moriondo et al. 2007). An even more common approach is the use of several vegetation indices as input to a Multiple Linear Regression (MLR) model to predict the response (e.g. Behrens et al. 2006; Babar et al. 2007; Pettersson et al. 2007; Xue et al. 2007). Regressions based on a latent variable structure like Partial Least Squares (PLS) are not so common, but examples do exist (e.g. Pettersson et al. 2006; Jensen et al. 2007). Even rarer is the use of multi-way PLS regression (N-PLS), but Hansen et al. have reported successful predictions of grain yield and grain protein concentration using N-PLS on multi-temporal reflectance data (Hansen et al. 2002).

Despite the numerous studies on plant property prediction based on reflectance data, there are still areas that lack scientific coverage.

The development of new instruments with high spectral resolution and wide spectral range gives a lot more data than did the early instruments. In order to extract as much information as possible from these instruments, the full spectrum should be used as input for prediction model computation. However, in many cases where hyperspectral reflectance data with possibly thousands of wavelengths are collected, only a very few of them are actually used for extracting useful information (e.g. Behrens et al. 2006; Babar et al. 2007; Prasad et al. 2007). This approach is possibly suboptimal because relevant information in the measured spectra may be overlooked.

A common way to validate computed prediction models is the use of two nearby field experiments in the same year. Data from one experiment is used to calibrate the model, whilst data from the other is used to validate the model (Wang et al. 2004; Zhao et al. 2005; Xue et al. 2007; Yang et al. 2008). Several studies report very high prediction accuracy, but the generalization of such models is doubtful because seasonal variability due to weather conditions and other season-specific factors are omitted.

1.3 Objectives

The objectives of the project were:

1. Compare some available spectroscopic instruments in terms of how well they perform in the task of plant property prediction.

2. Compute operational and practically useful prediction models for grain yield and grain quality.

4. Assess the robustness of such models.

3. Compare modern multivariate regression methods against traditional vegetation index-based methods.

2. Materials and methods

2.1 Field experiments

Three field trials (Photo 1) were established in the seasons 2007-2010, with two experiments at Apelsvoll research center and one at a nearby farm (Hoff). The experiments had 18-20 replicate blocks (limited by space in the field) and had 144-160 wheat plots each of size 2 m by 8 m. Every replicate block received randomized applications of six levels of nitrogen (N) fertilizer, corresponding to 0, 100, 125, 150, 175 and 200 kg nitrogen ha^{-1}.

Photo 1. An aerial view of the field experiment site A at Apelsvoll in 2008.

Plant protection (herbicides, fungicides and insecticides) was carried out according to current agricultural practice. All sites were not present in all years and the datasets comprise seven site-years of data with altogether

976 individually harvested, measured and analysed plots.

2.2 Instruments

Three instruments were used for data collection. A CropScan hand-held instrument with 13 wavelengths (485 nm - 1650 nm), a FieldSpec3 hyperspectral point spectrometer with 2150 wavelengths (350 nm – 2500 nm) and an airborne hyperspectral line scanner (imaging instrument) with 160 wavelengths (400 nm – 1000 nm). The two handheld instruments were operated in parallel (Photo 2) in such a way that the viewing geometries of the two were practically identical.

Photo 2. Operation of the CropScan and FieldSpec3 in parallel.

The airborne instrument was operated at 1000 m altitude and had a geometrical resolution of 20 cm x 20 cm pixels on the ground.

2.3 Measurements

Spectroscopic measurements were performed five times during the cropping season, at Zadoks (Z) stages Z31, Z59, Z65, Z87 and Z90 according to the Zadoks decimal code (Zadoks et al. 1974). Wheat plots were harvested and a grain sample was taken from each. All grain samples were analyzed gravimetrically for moisture and grain yield and with a FOSS Infratek grain analyzer for protein-, gluten- and starch concentrations as well as the Zeleny number (e.g. Jirsa et al. 2008). Soil samples were taken from every replicate block at all three sites.

Aggregated meteorological data were collected from the national yield prognosis programme (Korsaeth et al. 2009) for the weather station at Apelsvoll. The variables are short-wave radiation, precipitation, air temperature (2 m), wind speed and relative humidity. Potential evapotranspiration was calculated in accordance with Riley and Berentsen (2009). The daily weather data were then aggregated within four phenological phases: 1) sowing and seed emergence (Z00 to Z09), 2) leaf development and tillering (Z10 to Z29), 3) stem elongation and inflorescence emergence (Z30 to Z49), and 4) anthesis and ripening up to hard dough (Z50 to Z87).

2.3 Data analysis

Preprocessing of the spectroscopic data as performed by Multiplicative Scattering Correction (MSC, Isaksson et al. 1988) and differentiation by the Savtizky-Golay algorithm (Savitzky et al. 1964).

Pre-treatment of the meteorological data was performed with Principal Component Analysis (PCA). Two score vectors were extracted and used as derived meteorological variables.

For the computation of regression models between spectroscopic data and the response variables, we chose to use a recent extension of Partial Least Squares (PLS, Wold et al. 1983), known as Powered Partial Least Squares (Indahl 2005). This extension of PLS contains the traditional PLS solution within a continuum of infinitely many solutions. The algorithm can be directed to focus on predictors in the dataset that are highly correlated with the response or on predictors that have high variance. This is achieved through a reparameterization of the optimization criterion of PLS and the introduction of a method-specific parameter γ, which is limited to values in the interval [0,1]. The PPLS reduces to regular PLS by setting the γ to 0.5, whereas $\gamma=0$ makes PPLS focus exclusively on predictors with high variance. Setting $\gamma=1$ will direct the PPLS models to focus exclusively on predictors with high correlation with the response. The γ parameter can be set directly or it can be optimized over a user-specified range limited within the interval [0,1].

For analysis of models with several data types, such as spectroscopic and meteorological data, another recent extension of PLS, Canonical Partial Least Squares (CPLS, Indahl et al. 2009) was used. This extension enables the PLS algorithm to incorporate information from additional response variables. By optimization of the canonical correlation (rather than the

covariance) between the responses and the scores, the PLS solution is rotated to fit the main response in the best possible way while sacrificing some explanatory power of the additional responses This stands in contrast to the multi-response PLS (PLS2), which seeks to predict all responses equally well. The authors of CPLS encourage the combination of CPLS and PPLS into Canonical Powered Partial Least Squares (CPPLS). Since PPLS is used for analysis of the other models in this project, CPPLS was a natural choice of method.

For variable selection purposes in paper II, Backwards Variable Selection for Partial Least Squares (BVSPLS, Pierna et al. 2009) and Forward Stepwise Selection (FSS) is used in addition to PLS and PPLS. In paper I, Nearest Neighbour (NN, Fix et al. 1989) is used to select calibration samples. In paper II, the DUPLEX sample selection algorithm (Snee 1977) is used for the generation of representative calibration data sets.

Several model diagnostics were used to evaluate the results: Coefficient of Determination/explained variance (R^2), Root Mean Square Error of Prediction (RMSEP), bias, skewness (third moment about the mean of the residuals). Tests of RMSEP were performed with a Chi-square test in the sense of Indahl (2005), and testing of differences in predictive ability were performed by Cross-Validated Analysis of Variance (CVANOVA, Cederkvist et al. 2005)

3. Main results and discussion

3.1 Comparison of instruments (paper I)

The objective of paper I was to compare three available spectroscopic instruments in terms of how well they can predict properties of spring wheat. Collected data were used to compute PPLS regression models between all of the six analyzed constituents in the grain samples (grain yield, concentrations of protein, moisture, gluten and starch as well as the Zeleny number).

Models for all six constituents were computed on five occasions during the cropping season (only the 2007 data were present at this time). All models showed the same pattern regarding predictive ability as a function of time (Fig. 1). The most favourable time to perform spectroscopic measurements for prediction of wheat properties seemed to be Z65, where the explained variance peaked for most models. This is in agreement with other reported studies (e.g. Basnyat et al. 2004; Vicente-Serrano et al. 2006).

The five constituent variables protein, starch, gluten, moisture and Zeleny number were highly intercorrelated (R^2>0.85). Grain yield was less correlated with the other variables (R^2<0.70). Of the five other highly correlated variables, protein is the most attractive for practical prediction purposes because grain protein concentration is a critical measure of cereal quality that is used for allocating grain to different uses (Pettersson et al. 2006; Pettersson et al. 2007). Hence, we selected grain yield and grain protein concentration for further analysis.

For the grain yield case, the predictive abilities were at a high level for all instruments, with a peak R^2 of 0.96 for HySpex, 0.96 for CropScan and 0.97 for the FieldSpec3. The causal relationship between spectroscopic measurements and grain yield is largely due to the connection between the N content in the plant and the chlorophyll concentration in the leaves, which can be estimated by measurements of the spectral signature of the canopy (e.g. Jensen et al. 2007). Since all instruments had detectors in the spectral range of the chlorophyll absorption bands, the correlations with grain yield were almost independent of instrument.

For grain protein concentration, the differences between instruments were larger. The airborne instrument HySpex gave a peak R^2 of only 0.64, whereas CropScan and FieldSpec3 gave a peak R^2 of 0.88 and 0.94, respectively. Differences between the instruments become very clear in the protein models when inspecting the regression coefficients of each model. The FieldSpec3 instrument gave a model with non-zero regression coefficients in several parts of the spectrum. The region around 600 nm to 800 nm and 1200 nm to 1300 nm seemed to be especially important. None of the other instruments had detectors in the 1200 nm to 1300 nm spectral range, which mostly explains the large difference in prediction accuracy between the instruments. The causal relationship between spectroscopic data and protein is far more complex than that for grain yield. It is readily seen from the regression coefficients of the protein model and has also been suggested by other authors (e.g. Jensen et al. 2007).

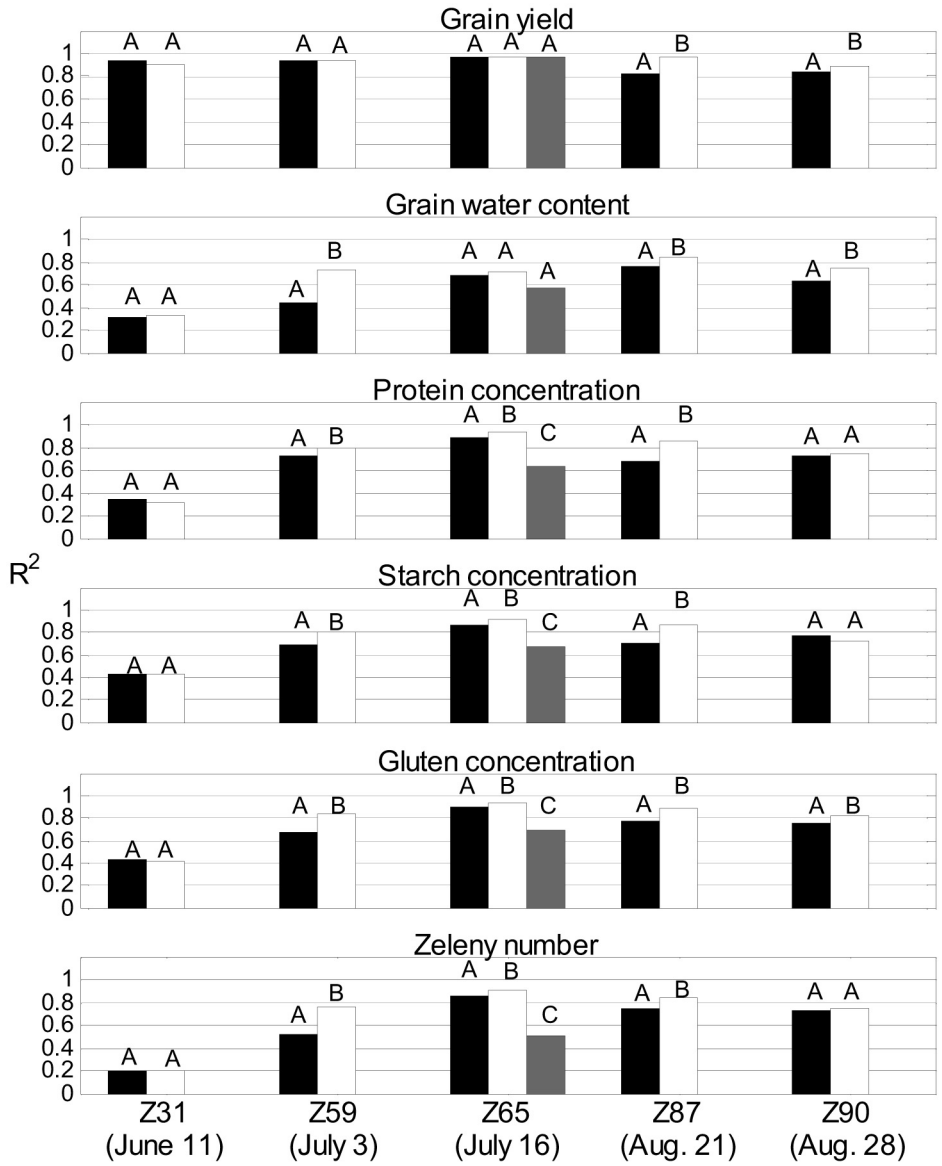


Figure 1. Relations between measurements and predictions of the validation-set data obtained with prediction model based measurements, conducted at five development stages of spring wheat by means of two hand-held radiometers, CropScan (filled bars) and FieldSpec3 (open bars) and an airborne hyperspectral scanner (stage Z65 only, grey bars). Bars with the same letter (within same day/stage) indicate non-significant differences in model fit (p=0.05).

For most constituents and measurement times, FieldSpec3 gave the best predictions (Fig. 1) and was hence selected as the instrument of choice for

further research. Prediction accuracy for both yield and protein were at least similar or better than results reported in the literature (e.g. Raun et al. 2001; Hansen et al. 2002; Basnyat et al. 2004; Liu et al. 2006; Pettersson et al. 2006; Xavier et al. 2006; Xue et al. 2007).

During analysis of the FieldSpec3 data, an interesting feature of the PPLS algorithm appeared. It seemed that by choosing a γ value near to 1 (i.e.>0.99) and truncating the loading weights less than $2.2204*10^{-17}$ (the relative numeric precision of MATLAB) as suggested by Indahl (2005), the predictor variable set could be reduced to very few wavelengths whilst still retaining most of the prediction accuracy for all constituents. Hence, models with as few predictors as possible were computed. The grain yield model was thus reduced to 3 predictors with R^2=0.95 and the grain protein concentration model was reduced to 5 predictors with R^2=0.90. Hence, the model complexity was reduced from 975 predictors to 3 and 5 predictors for grain yield and grain protein concentration, respectively. From this result, the idea arose to carry out a comparative study on PPLS against other methods for variable selection arose. This study was reported in paper II.

3.2 variable selection (paper II)

Three independent NIR datasets (fat/feed, fiber/maize and protein/maize) collected in the laboratory at the Waloon Center for Agricultural Reseasch (CRA-W) were used as testing grounds for PPLS as a variable selection method. The datasets are described in Pierna et al. (2009). We chose to test another recently suggested variable selection method, Backward Variable Selection for Partial Least Squares (BVSPLS), proposed by Pierna et al. (2009), in addition to PPLS. The reference methods were Forward

Stepwise Selection (FSS) and full spectrum Partial Least Squares (PLS). There are many examples of studies in variable selection (e.g. Chong et al. 2005; Anderssen et al. 2006; Andersen et al. 2010), but literature is scarce regarding studies on the recent methods PPLS and BVSPLS.

The datasets were split into validation sets and calibration sets. Most of the samples were put in the validation set, whilst 19 calibration sets of different size were selected, ranging from 20 samples to 200 samples with increments of 10 samples. Prediction models were made with all four methods for all 19 calibration sets and for all three main NIR datasets.

The results showed that variable selection had no effect in situations with calibration sets larger than 150 samples. All methods suffered from some degradation of the predictive ability as the calibration set size decreased. However, the variable selection methods, and especially PPLS, had less degradation of the prediction accuracy than that found using full spectrum PLS.

Three reasons for this behaviour were hypothesized:

1. Some predictor variables have only remote relevance to the response variable.

2. The signal to noise ratio (S/N) in some predictor variables may be so low that the elimination of these variables improves the model.

3. Some predictor variables may have a non-linear relationship to the response. Thus, elimination of these variables may give more parsimonious and linear prediction models and hence improve their prediction abilities.

Reference samples are often costly to obtain and a reduction of the number of calibration samples can potentially lead to economic savings. Variable selection techniques are one possible way of reducing the

calibration sample requirement whilst maintaining the predictive ability of the model.

In order to investigate the selection of predictors of each method, the selection frequencies of every predictor were computed and compared. Both FSS and BVSPLS had rather random selection patterns, whereas PPLS gave a more consistent selection of variables (i.e. selecting the same variable set for every size of calibration set, Fig. 2). Hence, the PPLS models were by far the easiest models to interpret.

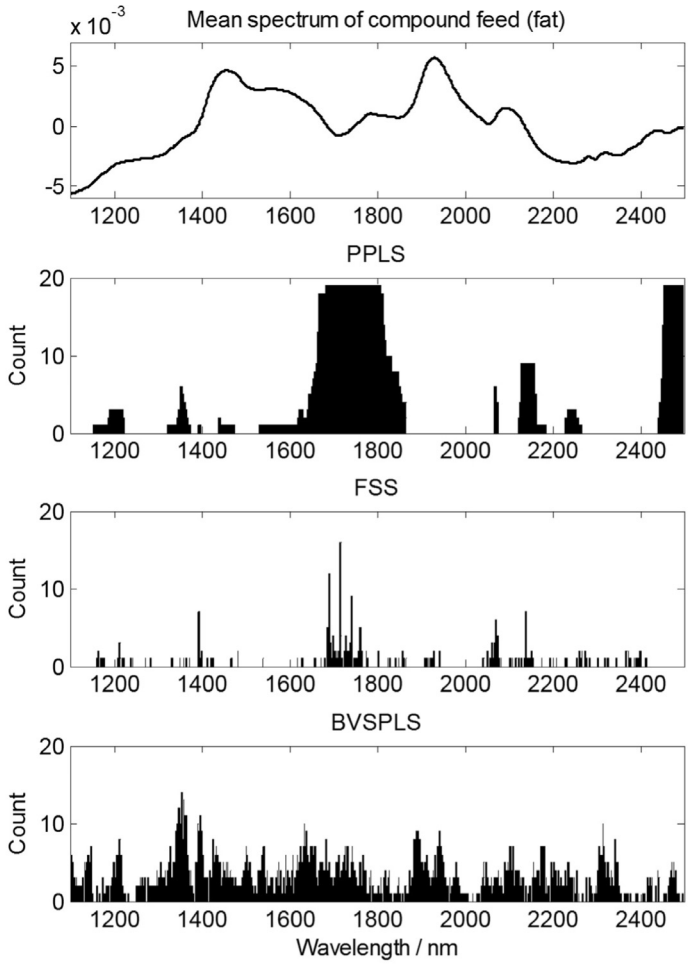


Figure 2. The number of times each variable were selected in the 19 fat prediction models plotted against the wavelength for Powered Partial Least Square (PPLS), Forward Stepwise Selection (FSS) and Backwards Variable Selection for PLS (BVSPLS). Mean spectrum (upper plot) is computed from the validation set for comparison.

3.3 Robustness of predictions of grain yield and grain protein (paper III)

Since the models computed in the instrument comparison in paper I were based on only one season of data, the prediction errors in a future year are completely unknown. In order to test the robustness of the prediction models of grain yield and grain protein concentration, more data were needed. Hence, several datasets were collected in the same fashion as in paper I over the period 2007-2010 at three different sites, comprising 7 site-years of data with altogether 976 individually measured, harvested and laboratory analyzed plots of spring wheat.

It was soon evident that some of the data points were impossible to fit in any of the prediction models. The reason was quite severe lodging in about 10% of the data. Lodging changes the spectral signature of the wheat plants considerably and the relation between spectroscopic measurements and wheat properties is changed compared to measurements performed on a standing crop. Lodging is often encountered in practice and a prediction model must be able to handle such situations in order to be usable. Hence, the spectroscopic datasets were classified with a PLS-DA model into the classes “lodging” and “standing crop” based on the spectroscopic measurements alone. The total accuracy of the PLS-DA model was 98.3 %, so it was in practice possible to correctly classify a measurement as “lodging” or “standing crop” before that measurement entered the prediction model. Literature is scarce on studies that classify lodging in cereals, but spectroscopic data have been used in numerous studies to classify crop species with high accuracy (e.g. Congalton 1991; Thenkabail et al. 2004; Yang et al. 2004).

As in paper I, PPLS was used to compute regression models between spectroscopic data (lodging measurements removed with the classification model) and wheat sample properties. Since the objective of the study was to assess the robustness of the models, several validation strategies were chosen. Complete results from all the validation strategies are displayed in paper III. For evaluation of the practical usefulness of such prediction models, the leave-year-out cross-validation strategy (Fig. 3) is probably the most interesting. The explained variance for grain yield was high for all models (R^2=0.76 to R^2=0.94), whereas the values for the grain protein concentration model were lower, 0.18 in the worst case. It was hence clear that validation of the grain protein concentration model was not possible with these data. This stands in contrast to other studies reported in the literature, which often show high levels of correlation for both grain yield and grain protein concentration (e.g. Hansen et al. 2002; Wang et al. 2004; Zhao et al. 2005; Reyniers et al. 2006; Xue et al. 2007; Yang et al. 2008). Common to all these studies is the validation of model with data from the same year as the calibration data. Hence, the variance introduced by season-specific variables such as weather conditions and interactions between weather conditions and soil properties, are not accounted for. When such models were validated on a multi-season and multi-site dataset as in paper III, these problems become visible. Despite the high amount of explained variance of the grain yield predictions, all models except the 2008 validation suffered from severe bias and skewness. This became clearer when values for RMSEP, bias and skewness were inspected, since the R^2 value are practically independent of bias. The bias and skewness problems raise the need for further investigations of the

grain yield model.

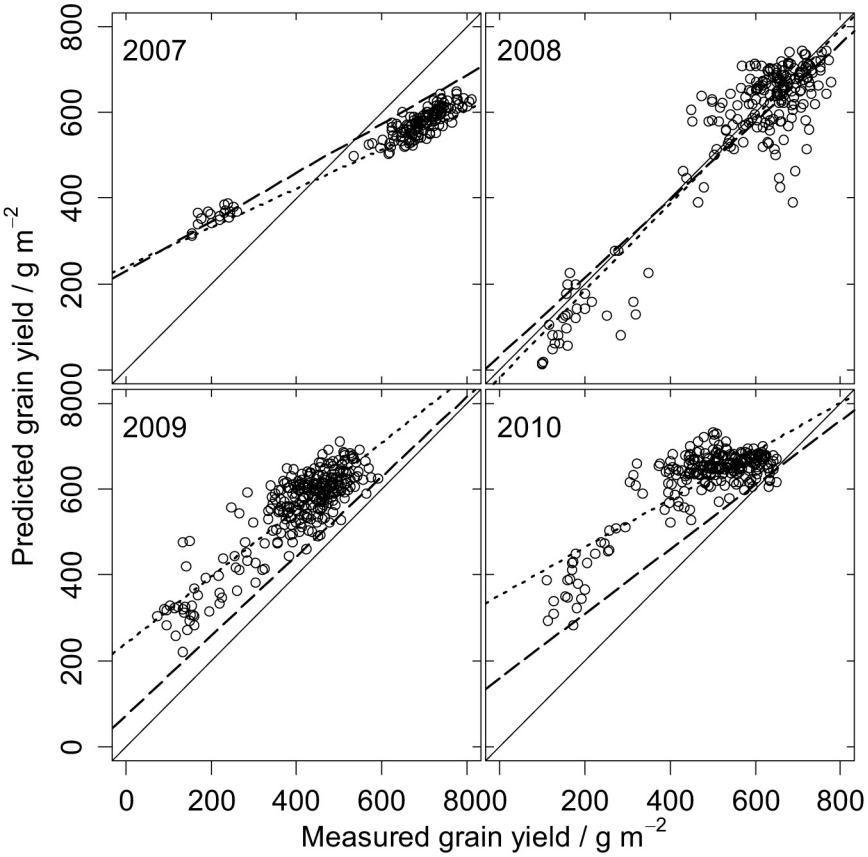


Figure 3. Measured grain yield plotted against grain yield predicted with the spectroscopic PPLS models, using data from each of the years 2007-2010 for validation (validation year denoted in each subplot), respectively. Target lines (solid) have zero intercept and slope 1. Regression lines from PPLS (dotted) and CPPLS (dashed) prediction models are displayed in each subplot.

The grain protein concentration model gave very poor predictions except for the computations based on one year only (Paper I). Since no external data exist at the plot-level, the statistical precision of this model cannot be improved. Hence, the grain protein concentration model was not subjected to further analysis based on the current spectroscopic dataset.

The grain yield model was compared to traditional yield estimation procedures based on four univariate linear regressions on four widely used

vegetation indices (NDVI, MSAVI, REIP and D-clh-ab). These indices were used to predict the grain yield data, and all of them failed in this task. Explained variances (R^2) were between 0.13 to 0.55, and both NDVI and REIP suffered from severe saturation at yield levels higher than 450 g m^{-2} (Fig. 4).

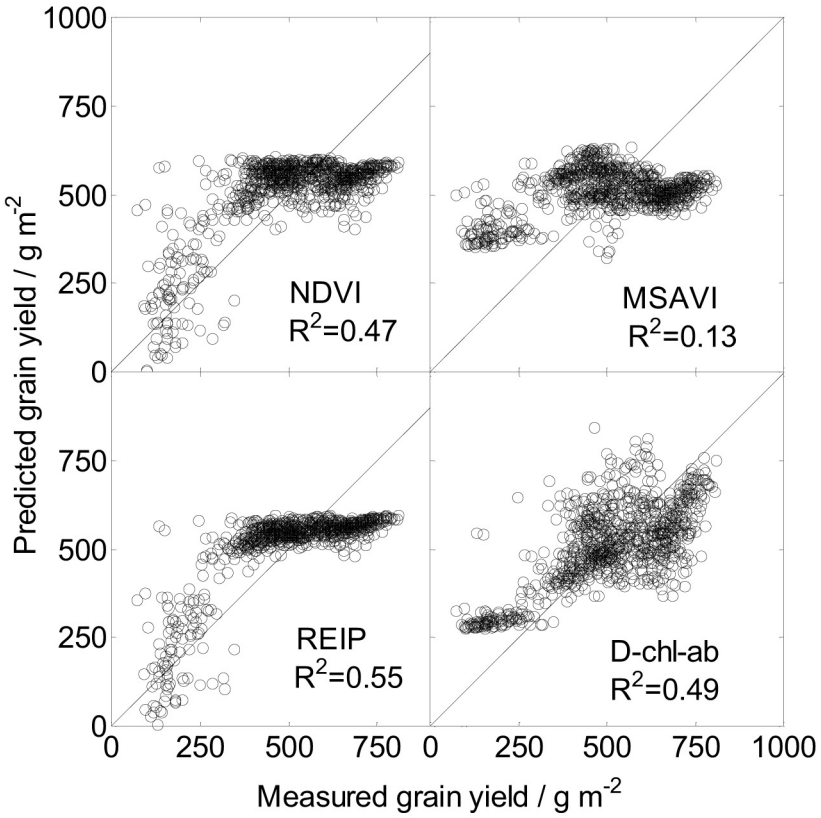


Figure 4. Measured grain yield (X-axis) plotted against predicted grain yield (Y-axis). Predictions were based on univariate least squares regressions between yield data and vegetation indices derived from all spectral measurements. All target lines have slope 1 and zero intercept. The indices used were Normalized Difference Vegetation Index (NDVI), Modified Soil Adjusted Vegetation Index (MSAVI), Red Edge Inflection Point (REIP) and D-chl-ab.

Despite the bias and skewness problems, the PPLS models clearly outperformed the traditional univariate method.

3.4 Fusion of spectroscopic data and meteorological data (paper IV)

In order to further improve the yield prediction model of paper III, the spectroscopic dataset was augmented to encompass also aggregated meteorological data. The weather data comprised 24 aggregated variables, which were reduced to 2 derived variables by the computation of PCA scores. These two derived variables were used as additional responses in a CPPLS model. The CPPLS model (Fig. 5) were cross-validated in the same way as the PPLS model in paper III, and the predictions were improved significantly for all years except 2008 according to the CVANOVA computations.

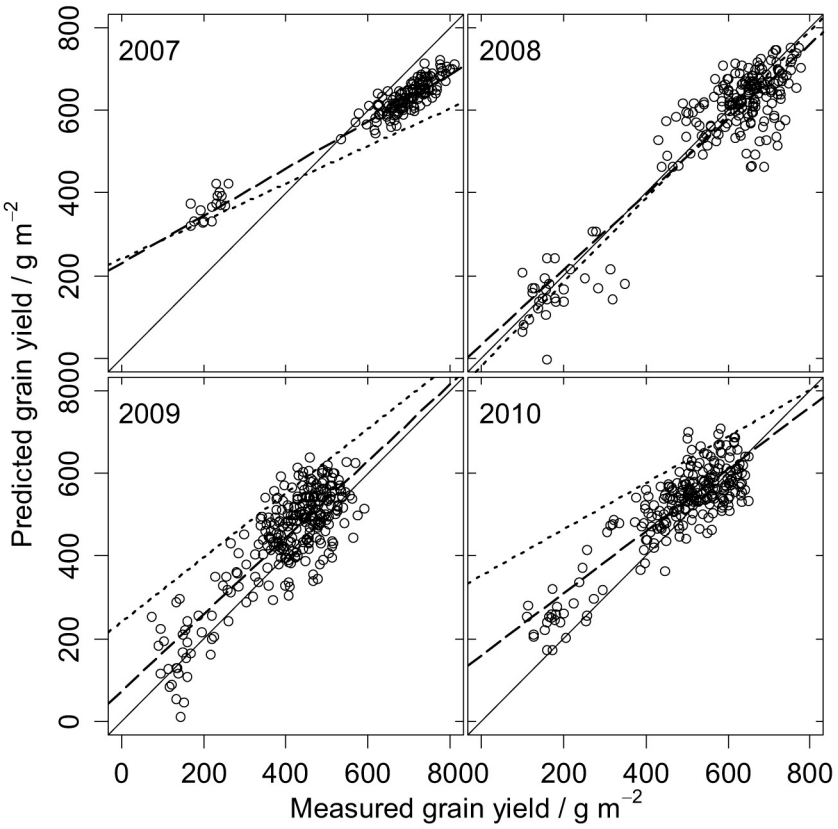


Figure 5. Measured grain yield plotted against grain yield predicted with the fusion CPPLS models, using data from each of the years 2007-2010 for validation (validation year denoted in each subplot), respectively. Target lines (solid) have zero intercept and slope 1. Regression lines from PPLS (dotted) and CPPLS (dashed) prediction models are displayed in each subplot.

Despite the fact that the predictions were significantly improved by the inclusion of additional data, there was still considerable skewness in the 2007 model (Fig. 5). Since weather variation is included in the CPPLS model, the skewness must have other causes. One cause for this could partly be that 2007 is the only year in which only experimental site was present. All the other years had two sites represented. Further research is needed to investigate this remaining skewness.

3. Main conclusions and outlook

Of the three instruments used, FieldSpec3 generated the most useful data for computation of prediction models for grain yield and grain protein concentration.

The PPLS methodology had very good ability to select relevant predictors and discard irrelevant predictors from spectroscopic data. This feature was especially pronounced in smaller calibration sets where it outperformed both FSS and full spectrum PLS. Moreover, PPLS gave the most easily interpretable models. The ability to calibrate models on small datasets is important because the need for costly reference sampling may be reduced. Future research on this topic should be augmented to the use of other data types rather than using spectroscopic data only.

Lodging can be classified into the two classes “lodging” and “standing crop” by means of a two-class PLS-DA model based on spectroscopic data alone. This is an important feature because measurements of crop lodging are not possible to predict reliably with the PPLS or CPPLS models. Hence, there is a need for automatic detection of measurements with lodging before they enter the actual prediction model. The PLS-DA model fulfilled this task, but further research on lodging classification requires a larger dataset and a comparative study should be performed on several classification methods such as k-Nearest Neighbour (kNN, Fix et al. 1989), Soft Independent Modeling of Class Analogy (SIMCA, Wold 1976), and Support Vector Machine (SVM, Cortes et al. 1995).

Multivariate grain yield prediction models clearly outperform the traditional prediction methods.

Fusion of spectroscopic data and aggregated meteorological variables by means of CPPLS, increases the robustness of the grain yield prediction models significantly, and is one step further towards such prediction models becoming a useful tool for the agricultural community. However, further research in this field is required and should focus on the inclusion of more prediction data and looking for other possible methodologies to perform the data fusion. A first step could be a comparative study on data fusion methods using the same data as presented here. Comparisons of the recent CPPLS could be performed against more established methods such as Hierarchical Principal Component Analysis (HPCA), Consensus PCA (CPCA), Hierarchical Partial Least Squares (HPLS) or Multiblock Partial Least Squares (MBPLS). All of these methods are described by Westerhuis et. al (1998).

4. References

Andersen, C. M. and R. Bro (2010). "Variable selection in regression-a tutorial." Journal of Chemometrics **24**(11-12): 728-737.

Anderssen, E., K. Dyrstad, et al. (2006). "Reducing over-optimism in variable selection by cross-model validation." Chemometrics and Intelligent Laboratory Systems **84**(1-2): 69-74.

Atzberger, C. (2004). "Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models." Remote Sensing of Environment **93**(1-2): 53-67.

Babar, M. A., M. van Ginkel, et al. (2007). "Heritability, correlated response, and indirect selection involving spectral reflectance indices and grain yield in wheat." Australian Journal of Agricultural Research **58**(5): 432-442.

Basnyat, P., B. McConkey, et al. (2004). "Optimal time for remote sensing to relate to crop grain yield on the Canadian prairies." Canadian Journal of Plant Science **84**(1): 97-103.

Basnyat, P., B. G. McConkey, et al. (2005). "Effectiveness of using vegetation index to delineate zones of different soil and crop grain production characteristics." Canadian Journal of Soil Science **85**(2): 319-328.

Behrens, T., J. Muller, et al. (2006). "Utilization of canopy reflectance to predict properties of oilseed rape (Brassica napus L.) and barley (Hordeum vulgare L.) during ontogenesis." European Journal of Agronomy **25**(4): 345-355.

Behrens, T., J. Muller, et al. (2007). "Optimizing a diode array VIS/NIR spectrometer system to detect plant stress in the field." Journal of Agronomy and Crop Science **193**(4): 292-304.

Berge, T. W., H. Fykse, et al. (2007). "Patch spraying of weeds in spring cereals: Simulated influences of threshold level and spraying resolution on spraying errors and potential herbicide reduction." Acta Agriculturae Scandinavica Section B-Soil and Plant Science **57**(3): 212-221.

Berntsen, J., A. Thomsen, et al. (2006). "Algorithms for sensor-based redistribution of nitrogen fertilizer in winter wheat." Precision Agriculture **7**(2): 65-83.

Cederkvist, H. R., A. H. Aastveit, et al. (2005). "A comparison of methods for testing differences in predictive ability." Journal of Chemometrics **19**(9): 500-509.

Chong, I. G. and C. H. Jun (2005). "Performance of some variable selection methods when multicollinearity is present." Chemometrics and Intelligent Laboratory Systems **78**(1-2): 103-112.
Congalton, R. G. (1991). "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." Remote Sensing of Environment **37**(1): 35-46.
Cortes, C. and V. Vapnik (1995). "Support-Vector Networks." Machine Learning **20**(3): 273-297.
de Wit, A. M. and C. A. van Diepen (2007). "Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts." Agricultural and Forest Meteorology **146**(1-2): 38-56.
Feng, W., X. Yao, et al. (2008). "Monitoring leaf pigment status with hyperspectral remote sensing in wheat." Australian Journal of Agricultural Research **59**(8): 748-760.
Fix, E. and J. L. Hodges (1989). "Discriminatory Analysis - Nonparametric Discrimination - Consistency Properties." International Statistical Review **57**(3): 238-247.
Gitelson, A. A. and M. N. Merzlyak (1996). "Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll." Journal of Plant Physiology **148**(3-4): 494-500.
Guyot, G., F. Baret, et al. (1988). "High spectral resolution : determination of spectral shifts between the red and near infrared." International Archives of Photogrammetry and Remote Sensing **11**: 750-760.
Hansen, P. M., J. R. Jørgensen, et al. (2002). "Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression." Journal of Agricultural Science **139**: 307-318.
Houborg, R., H. Soegaard, et al. (2007). "Combining vegetation index and model inversion methods for the extraction of key vegetation biophysical parameters using Terra and Aqua MODIS reflectance data." Remote Sensing of Environment **106**(1): 39-58.
Huang, W., D. W. Lamb, et al. (2007). "Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging." Precision Agriculture **8**(4-5): 187-197.
Indahl, U. (2005). "A twist to partial least squares regression." Journal of Chemometrics **19**(1): 32-44.
Indahl, U. G., K. H. Liland, et al. (2009). "Canonical partial least squares-a unified PLS approach to classification and regression problems." Journal of Chemometrics **23**(9-10): 495-504.
Isaksson, T. and T. Naes (1988). "The Effect of Multiplicative Scatter Correction (Msc) and Linearity Improvement in Nir Spectroscopy." Applied Spectroscopy **42**(7): 1273-1284.
Jensen, T., A. Apan, et al. (2007). "Detecting the attributes of a wheat crop using digital imagery acquired from a low-altitude platform." Computers and Electronics in Agriculture **59**(1-2): 66-77.

Jirsa, O., M. Hruskova, et al. (2008). "Near-infrared prediction of milling and baking parameters of wheat varieties." Journal of Food Engineering **87**(1): 21-25.
Jordan, C. F. (1969). "Derivation of Leaf-Area Index from Quality of Light on Forest Floor." Ecology **50**(4): 663-&.
Jorgensen, J. R. and R. N. Jorgensen (2007). "Uniformity of wheat yield and quality using sensor assisted application of nitrogen." Precision Agriculture **8**(1-2): 63-73.
Jorgensen, R. N., L. K. Christensen, et al. (2007). "Spectral reflectance at sub-leaf scale including the spatial distribution discriminating NPK stress characteristics in barley using multiway partial least squares regression." International Journal of Remote Sensing **28**(5): 943-962.
Karimi, Y., S. O. Prasher, et al. (2005). "Classification accuracy of discriminant analysis, artificial neural networks, and decision trees for weed and nitrogen stress detection in corn." Transactions of the Asae **48**(3): 1261-1268.
Korsaeth, A. (2005). "Soil apparent electrical conductivity (ECa) as a means of monitoring changes in soil inorganic N on heterogeneous morainic soils in SE Norway during two growing seasons." Nutrient Cycling in Agroecosystems **72**(3): 213-227.
Korsaeth, A. and T. Rafoss (2009). "Tidlige prognoser for kornavlingene ved bruk av værdata - Sluttrapport (in Norwegian)." Bioforsk RAPPORT **17**(4): 1-45.
Korsaeth, A. and H. Riley (2006). "Estimation of economic and environmental potentials of variable rate versus uniform N fertilizer application to spring barley on morainic soils in SE Norway." Precision Agriculture **7**(4): 265-279.
Liu, L. Y., J. J. Wang, et al. (2006). "Predicting winter wheat condition, grain yield and protein content using multi-temporal EnviSat-ASAR and Landsat TM satellite images." International Journal of Remote Sensing **27**(4): 737-753.
Lukina, E. V., K. W. Freeman, et al. (2001). "Nitrogen fertilization optimization algorithm based on in-season estimates of yield and plant nitrogen uptake." Journal of Plant Nutrition **24**(6): 885-898.
Mirik, M., G. J. Michels, et al. (2007). "Reflectance characteristics of Russian wheat aphid (Hemiptera : Aphididae) stress and abundance in winter wheat." Computers and Electronics in Agriculture **57**(2): 123-134.
Moriondo, M., F. Maselli, et al. (2007). "A simple model of regional wheat yield based on NDVI data." European Journal of Agronomy **26**(3): 266-274.
Pettersson, C. G. and H. Eckersten (2007). "Prediction of grain protein in spring malting barley grown in northern Europe." European Journal of Agronomy **27**(2-4): 205-214.

Pettersson, C. G., M. Soderstrom, et al. (2006). "Canopy reflectance, thermal stress, and apparent soil electrical conductivity as predictors of within-field variability in grain yield and grain protein of malting barley." Precision Agriculture **7**(5): 343-359.

Pierna, J. A. F., O. Abbas, et al. (2009). "A Backward Variable Selection method for PLS regression (BVSPLS)." Analytica Chimica Acta **642**(1-2): 89-93.

Prasad, B., B. F. Carver, et al. (2007). "Potential use of spectral reflectance indices as a selection tool for grain yield in winter wheat under great plains conditions." Crop Science **47**(4): 1426-1440.

Qi, J., A. Chehbouni, et al. (1994). "Amodified soil adjusted vegetation index." Remote Sensing of the Environment **48**: 119-126.

Raun, W. R., J. B. Solie, et al. (2001). "In-season prediction of potential grain yield in winter wheat using canopy reflectance." Agronomy Journal **93**(1): 131-138.

Reyniers, M. and E. Vrindts (2006). "Measuring wheat nitrogen status from space and ground-based platform." International Journal of Remote Sensing **27**(3): 549-567.

Reyniers, M., E. Vrindts, et al. (2004). "Optical measurement of crop cover for yield prediction of wheat." Biosystems Engineering **89**(4): 383-394.

Reyniers, M., E. Vrindts, et al. (2006). "Comparison of an aerial-based system and an on the ground continuous measuring device to predict yield of winter wheat." European Journal of Agronomy **24**(2): 87-94.

Riley, H. and E. Berentsen (2009). "Estimation of water use for irrigation in Norwegian agriculture." Bioforsk RAPPORT **4**(174): 1-52.

Rouse, J. W., R. H. Haas, et al. (1974). "Monitoring the Vernal Advancement of Retrogradation of Natural Vegetation." Greenbelt: NASA/GSFC: 1-371.

Savitzky, A. and M. J. E. Golay (1964). "Smoothing + Differentiation of Data by Simplified Least Squares Procedures." Analytical Chemistry **36**(8): 1627-&.

Snee, R. D. (1977). "Validation of Regression-Models - Methods and Examples." Technometrics **19**(4): 415-428.

Sripada, R. P., D. C. Farrer, et al. (2007). "Aerial color infrared photography to optimize in-season nitrogen fertilizer recommendations in winter wheat." Agronomy Journal **99**(6): 1424-1435.

Thenkabail, P. S., E. A. Enclona, et al. (2004). "Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications." Remote Sensing of Environment **91**(3-4): 354-376.

Tucker, C. J. (1979). "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation." Remote Sensing of Environment **8**(2): 127-150.

Vicente-Serrano, S. M., J. M. Cuadrat-Prats, et al. (2006). "Early prediction of crop production using drought indices at different time-scales and remote sensing data: application in the Ebro valley (North-East Spain)." International Journal of Remote Sensing **27**(3): 511-518.

Wang, Z. J., J. H. Wang, et al. (2004). "Prediction of grain protein content in winter wheat (Triticum aestivum L.) using plant pigment ratio (PPR)." Field Crops Research **90**(2-3): 311-321.

Westerhuis, J. A., T. Kourti, et al. (1998). "Analysis of multiblock and hierarchical PCA and PLS models." Journal of Chemometrics **12**(5): 301-321.

Wold, S. (1976). "Pattern-Recognition by Means of Disjoint Principal Components Models." Pattern Recognition **8**(3): 127-139.

Wold, S., H. Martens, et al. (1983). "The Multivariate Calibration-Problem in Chemistry Solved by the Pls Method." Lecture Notes in Mathematics **973**: 286-293.

Xavier, A. C., B. F. T. Rudorff, et al. (2006). "Hyperspectral field reflectance measurements to estimate wheat grain yield and plant height." Scientia Agricola **63**(2): 130-138.

Xue, L. H., W. X. Cao, et al. (2007). "Predicting grain yield and protein content in winter wheat at different N supply levels using canopy reflectance spectra." Pedosphere **17**(5): 646-653.

Yang, C., J. H. Everitt, et al. (2008). "Yield estimation from hyperspectral imagery using spectral angle mapper (SAM)." Transactions of the Asabe **51**(2): 729-737.

Yang, C. C., S. O. Prasher, et al. (2004). "Differentiation of crop and weeds by decision-tree analysis of multi-spectral data." Transactions of the Asae **47**(3): 873-879.

Zadoks, J. C., T. T. Chang, et al. (1974). "Decimal Code for Growth Stages of Cereals." Weed Research **14**(6): 415-421.

Zhao, C. J., L. Y. Liu, et al. (2005). "Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress." International Journal of Applied Earth Observation and Geoinformation **7**(1): 1-9.

Zillmann, E., S. Graeff, et al. (2006). "Assessment of cereal nitrogen requirements derived by optical on-the-go sensors on heterogeneous soils." Agronomy Journal **98**(3): 682-690.

PAPER I

Received: 17 February 2010 ■ *Revised: 30 July 2010* ■ *Accepted: 3 August 2010* ■ *Publication: 5 October 2010*

JNIRS JOURNAL OF NEAR INFRARED SPECTROSCOPY

Comparisons of two hand-held, multispectral field radiometers and a hyperspectral airborne imager in terms of predicting spring wheat grain yield and quality by means of powered partial least squares regression

Stein Ivar Øvergaard,[a,*] Tomas Isaksson,[b] Knut Kvaal[c] and Audun Korsaeth[a]

[a]The Norwegian Institute for Agricultural and Environmental Research, 2849 Kapp, Norway.
E-mail: SteinIvar.Overgaard@bioforsk.no

[b]The Norwegian University of Life Sciences, Department of Chemistry, Biotechnology and Food Science, 1430 Ås, Norway

[c]The Norwegian University of Life Sciences, Department of Mathematical Sciences and Technology, 1430 Ås, Norway

Three radiometric instruments were compared as tools for predicting crop yield and grain quality: a CropScan instrument with 13 photodiodes (485–1650 nm), a 2150-channel FieldSpec3 instrument (350–2500 nm) and a HySpex airborne hyperspectral line scanner with 160 image wavelength layers (400–1000 nm). The first two instruments are point spectroradiometers, while the HySpex is an imaging instrument with a pixel size of 20 × 20 cm on the ground when the instrument is used at an altitude of 1000 m. A spring wheat field experiment of 160 plots was measured five times during the 2007 growing season. At harvest, grain yield was measured on each plot and analysed for moisture, protein, gluten, starch concentration and Zeleny sedimentation value. A recent statistical method, powered partial least squares (PPLS), was used for modelling and variable selection. The predictive performance of the calibrated models was very good, with coefficients of determination for the validation data (r^2_{pred}) reaching 0.97 and 0.94 for grain yield and grain protein concentration, respectively. The predictions (r^2_{pred}) of the other grain quality variables were in the range of 0.88–0.92. The airborne HySpex did not perform as well as the other instruments, most likely due to its limited spectral range. FieldSpec3 was significantly better than CropScan in most cases, probably as the former instrument has wider spectral range, a larger number of wavelengths and higher spectral resolution than the latter. A PPLS variable selection was carried out, which reduced the analysed data set from 975 wavelengths to 3–5 wavelengths. Although the number of retained variables was very low, the reduced models still had almost the same predictive ability as the PPLS models based on the full data set. The obtained simplicity of the calibration models indicates that a very small and lightweight instrument could be suitable for crop monitoring. Lightweight instruments are crucial for the utilisation of small unmanned aerial vehicles (UAVs). UAV technology is evolving quickly and small, cost effective UAV platforms are already available on the market. The concept of combining a UAV with a specifically designed instrument could provide an extremely versatile and cost effective system for crop monitoring.

Keywords: precision agriculture, hyperspectral, variable selection, spectral reflectance, PPLS, FieldSpec, CropScan, HySpex

ISSN: 0967-0335
doi: 10.1255/jnirs.892

Introduction

The analysis of spectral reflectance measurements, in general, and near infrared spectral reflectance, in particular, has been used in a variety of fields in the recent decades, for example, in geology, chemistry, biology, agriculture and archaeology. In agriculture, remote sensing has several applications that are closely linked to the field of precision agriculture. The most widespread applications are monitoring of crop nitrogen status,[1–3] yield forecasting,[4,5] quality prediction,[6,7] disease detection[8] and monitoring plant stress.[9]

Several types of instrument have been used for remote sensing applications. The instrument platforms span from hand-held via airborne to space-borne and the corresponding viewing distances span from 10^{-3} m to 10^{6} m. The spectral range can be as small as 1 or just a few wavelengths,[10,11] or it may cover the whole range from ultraviolet radiation via visible light and well into the infrared zone of the electromagnetic spectrum.[12] Detectors vary from single-channel cameras and radiometers to hyperspectral cameras and spectroradiometers.

Different combinations of instrument platform, spectral range, spectral resolution and spatial resolution have different pros and cons, depending on the objective of the measurement. It is not possible, however, to determine beforehand which approach is optimum. Hence, there is a need for comparative studies. Many studies have been aimed at comparing two or more instruments,[10,13,14] but due to the ongoing development of new and improved instruments, new comparative studies are needed.

Literature is scarce with respect to comparative studies that involve several instruments with different spectral resolutions and data spacings. Reyniers *et al*.[13] included such instruments, but they did not discuss the reasons for the differences in predictive performance between the airborne and the ground-based systems tested.

In this study, three instruments were compared: two hand-held spectroradiometer instruments measuring at a point and one airborne imaging instrument. The two hand-held instruments represent completely different approaches to spectral reflectance measurement. The simplest instrument, CropScan, is a proven instrument that has been used for crop monitoring for many years,[2,15] while the newer FieldSpec3 is a continuous-range, scientific-grade spectroradiometer with wide spectral range, dense data spacing and high spectral resolution. The airborne instrument is a relatively new imaging hyperspectral line-scanner with high spatial and spectral resolution but with a more limited spectral range than the two hand-held instruments.

The operational characteristics of each instrument type are quite different. The point spectroradiometers can deliver spectra with high spectral resolution and dense data spacing, but at a quite slow pace and only in close proximity to the sample. On the other hand, the airborne system may produce spectra with almost the same spectral resolution and data spacing but at a much higher temporal frequency and at a greater distance from the sample.

The instruments were compared in terms of their ability to predict grain yield and various grain quality variables of spring wheat. The rationale for this was that yield and quality prediction is an area where spectral reflectance is widely used. Moreover, the oldest instrument used in the current study (CropScan) was shown by Hansen *et al*.[4] to be particularly suited for this purpose in spring wheat. The CropScan instrument used in the present study was equipped with similar photodiodes to those in the study by Hansen *et al*.[4]

An important part of spectral reflectance analysis is the process of converting the acquired spectral data into models suitable for predicting selected properties. In the present study, we have, for three reasons, chosen to use variable selection in order to achieve as simple models as possible. First, a simple model with a low number of predictors can use input data from an inexpensive, simple and lightweight instrument. Second, it is easier to interpret a simple model than a model using the full spectrum of a hyperspectral instrument. Last, by removing unimportant predictor variables, the signal-to-noise ratio (SNR) in the data increases and the resulting model will be potentially more robust. In this study, we use the inverse variation coefficient as a measure of SNR.

On the other hand, retaining the full spectrum may be an advantage for various model diagnostics and outlier detection methods. However, in the present study we have chosen to emphasise the model (and hence instrument) simplicity. A simple instrument may weigh less than a more complicated one. Lightness of weight is a critical factor for the use of such instruments in small unmanned aerial vehicles (UAVs). The UAVs have great potential for acting as instrument carriers in a multitude of small-scale remote sensing applications. This potential has been increased further by the recent development of programmable, autonomous flight controllers for use in helicopters.[16] The specification of a small and affordable remote sensing instrument and relevant prediction models will bring UAV-based crop management closer to practical usefulness.

Many approaches have been tried in order to select the most important variables for various remote sensing applications. The most common approaches are formation of traditional vegetation indices,[17–20] inspection of correlation matrices,[12] analysis of variation (ANOVA)[21] and various forms of stepwise linear regression.[12,21,22] In hyperspectral instruments, dense data spacing leads to many highly correlated variables in the collected spectra. Some of these variables may have weak or even no correlation to the response variable. The use of traditional statistical methods with such data can potentially lead to problems.[23] There is also a risk that a data set with large spectral range and possibly thousands of wavelengths will contain a lot of information that is not relevant to the reference variable of interest. This fact suggests that methods based on latent variable structures should be used, such as members of the partial least squares (PLS) family of regression methods.[24] The relatively new powered partial least squares (PPLS) method seems especially promising

in this context, due to its strong ability to weight variables that possess predictive ability with respect to the response variables.[23]

For the current study, two main objectives were stated. The first objective was to perform a comparative study of instruments in order to address how well each instrument can predict yield and quality variables. The second objective was to perform a variable selection to ensure that prediction is carried out with the simplest possible subset of the data.

Materials and methods

A comparison between two hand-held spectroradiometers (CropScan and FieldSpec3) and an airborne hyperspectral scanner (HySpex) was performed in 2007 in an ongoing field trial with spring wheat (*Triticum aestivum* L., var. "Bjarne") in central SE Norway, at Bioforsk Apelsvoll arable crops division.

Field experiment

Apelsvoll (60° 42″ N, 10° 51″ E, 250 m above sea level) has a mean annual precipitation of 600 mm, a mean annual temperature of 3.6°C and a mean growing season (May–September) temperature of 12°C. The experimental area, which slopes 3–6% eastwards, is on an imperfectly drained brown earth (Gleyed melanic brunisoil, Canada Soil Survey[25]) with predominantly loam and silty sand textures.

The field trial was established in 2006, on an area of 26 × 160 m, which was divided into 20 replicate blocks (Figure 1). Six nitrogen level treatments were represented within each block: 0 kg, 100 kg, 125 kg, 150 kg, 175 kg and 200 kg nitrogen ha^{-1} (designated N0, N100, N125, N150, N175 and N200, respectively). The border plots on either side of the randomised plots received 100 kg nitrogen ha^{-1}, thus giving three plots with N100 per replicate block. Plant protection (herbicides, fungicides and insecticides) was carried out according to the current practice.

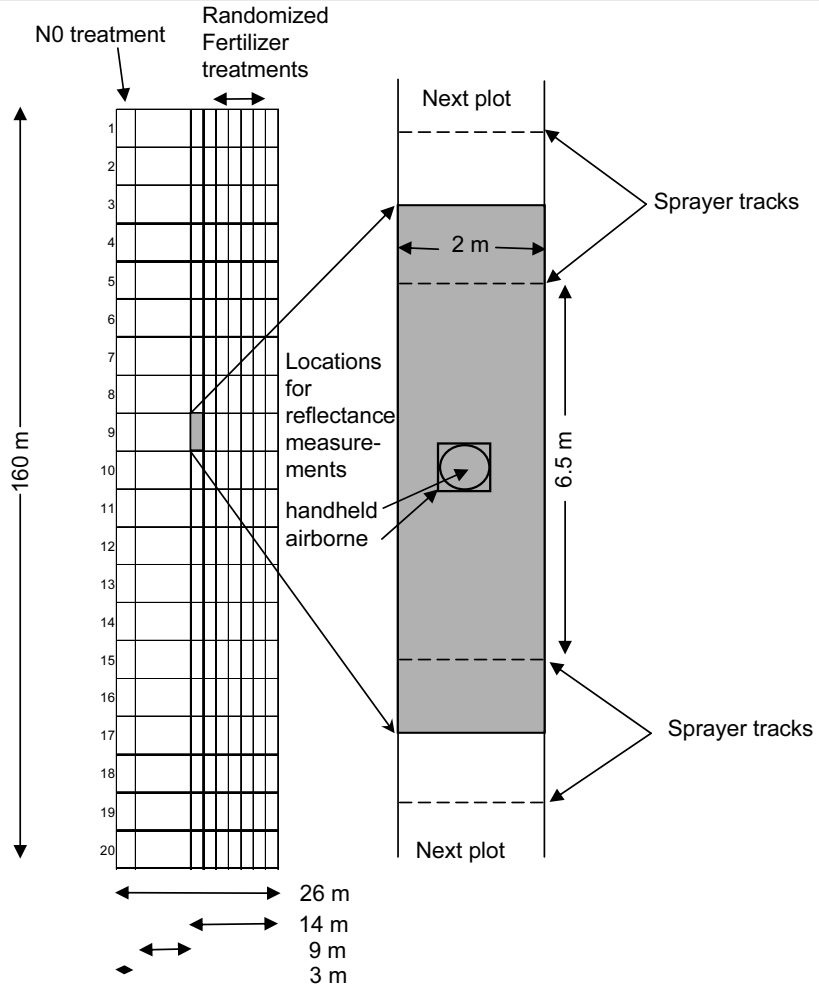


Figure 1. Experiment layout and sampling locations. Overall experiment layout to the left and one example plot enlarged to the right. Measurement areas for the respective instruments are shown with legends on the right-hand plot.

Table 1. Instrument specifications.

Specification	CropScan	FieldSpec3	HySpex
Number of variables	13	2150	160
Spectral range (nm)	485–1650	350–2500	400–100
Spectral resolution (nm)	10–12	2.8–12	10
Data spacing (nm)	1–700	1.4–2.0	3.7
Imager	No	No	Yes
Reflectance calibration method	Continouous dual-optic	Spectralon panel	Post processing software
Field of view (°)	28	25	17 (total image area)
Pixel size on the ground	Diameter 0.4–0.6 m	Diameter 0.4–0.6 m	20 × 20 cm throughout the whole image area
Number of pixels	1	1	1600 across flight track
Operating altitude (m)	~1.8	~1.8	1000

Instruments

For a summary of instrument specifications, see Table 1. All spectral resolutions are manufacturer specifications given in the sense of full width at half maximum (FWHM).

CropScan

The hand-held device CropScan (version MSR16R; CropScan Inc., USA, www.cropscan.com) has the feature of dual optics which measure simultaneously both incoming irradiance and sample radiance (upwards and downwards pointing detectors) in all 13 wavelengths.

FieldSpec3

FieldSpec3 (version 3; Analytical Spectral Devices Inc., USA, www.asdi.com) is a hand-held device. It does not possess dual optics and calibration was performed on a Spectralon panel provided and calibrated by LabSphere (LabSphere Inc, USA, www.labsphere.com).

HySpex

The airborne device used in this study is a hyperspectral line scanner (HySpex VNIR-1600; Norsk Elektro Optikk AS, Norway, www.neo.no). Data from this instrument is calibrated to reflectance through extensive software processing, explained in the next section.

Measurements

Canopy radiance spectra were measured with the two hand-held radiometers, on all 160 plots, five times during the growth season in 2007: 11 June, 3 July, 16 July, 21 August and 28 August. These dates coincided with growth stages Z31, Z59, Z65, Z87 and Z90, respectively, given by Zadoks *et al*.[26] The two hand-held spectroradiometers were mounted on the same pole and operated in parallel. The measurements were taken as near to simultaneously as possible. Two replicate measurements were made on each sample plot and with each radiometer. The pole must be kept in constant orientation with respect to the horizon and the sun[27] in order to achieve the most stable reflectance measurements. To achieve this, we used a pole-mounted spirit level of the same type as used in land surveying equipment and by visually aiming the pole towards the instantaneous position of the sun.

The FieldSpec3 was calibrated after measuring each replicate block in the field experiment, so that the time between each calibration did not exceed five minutes. The measuring optic of the instruments were kept constant at 1.8 m above the ground, which corresponds to a spot diameter of approximately 0.4–0.6 m for the two hand-held spectroradiometers (Figure 1). At Z31, the spot diameter was close to 0.6 m, while at harvest (Z90) the diameter had shrunk to 0.4 m, due to higher plant canopy. Hyperspectral image data were recorded on 16 July (Z65), at the same time as the spectroradiometer measurements. During data collection, the instrument was mounted in a Piper PA31-350 Chieftain aeroplane flown at 1000 m altitude. The HySpex instrument was co-mounted with an airborne laser scanner (Leica ALS50-II; Leica Geosystems AG, Switzerland, www.leica-geosystems.com), which provided data to create a digital surface model (DSM) with accuracy < 30 cm of the HySpex data recording area. The instantaneous position and attitude (i.e. the spatial orientation with respect to an Earth-fixed global coordinate system) of the aircraft was provided by a real time kinematic (RTK) global positioning system (GPS) receiver (NovAtel OEM4; NovAtel Inc., Canada, www.novatel.com) feeding an inertial measurement unit (IMU) (Honeywell Micro IRS IMU; Honeywell Inc., USA, www.honeywell.com) with position data. Together, the GPS and IMU provide dynamic position accuracy of 1 cm + 1 ppm (i.e. 1 cm absolute accuracy plus 1 ppm of the distance between the aeroplane and the GPS reference station) and attitude accuracy at 0.1 degree level. Operation and set-up of the aeroplane, instrument and positioning equipment was performed by TerraTec (TerraTec AS, Norway, www.terratec.no). A combined orthorectification and georeferencing software, parametric geocoding and

orthorectification for airborne optical scanner data (PARGE), was used to correct the images obtained by the instrument (PARGE; Remote Sensing Applications, Switzerland, www.rese.ch). The DSM, position and attitude data were used as input to PARGE for orthorectification and georeferencing of the HySpex images. Further, atmospheric correction was performed using the ATCOR-4 software (ATCOR-4; Remote Sensing Applications, Switzerland, www.rese.ch). The same software also allowed for reflectance processing of the radiance images. This processing is based on a MODTRAN-derived (MODerate resolution atmospheric TRANsmission),[28] database of typical solar geometries and flight altitudes, where the solution is interpolated for the actual data recording geometry,[29,30] thus excluding the need for a calibration panel on the ground. An image excerpt of nine pixels from the central area of each plot was extracted and averaged. The nine pixels were chosen in order to reduce image edge-effects on the plot borders. Averaging was performed to account for within-plot spectral variation. At maturity (Z90), a plot harvester was used to harvest 6.5 × 1.5 m plots between the sprayer tramlines (Figure 1). The grain samples were analysed gravimetrically for dry matter (DM) and moisture. Zeleny sedimentation value and concentrations of protein, starch and gluten were analysed by near-infrared transmission (Infratec 1241 Grain Analyzer; FOSS Tecator, Denmark, www.foss.dk). All reference variables were measured on a dry weight basis. The Infratec calibration model had serial number WH182126. This model is based on artificial neural network (ANN) calibrations performed by the instrument manufacturer. For an overview of the reference variables measured see Table 2.

Data treatment

All data analyses were carried out using the software package MATLAB (version R2007b; MathWorks Inc., USA, www.mathworks.com). Data from the FieldSpec3 and Hyspex were treated with the standard multiplicative scattering correction (MSC) algorithm.[31,32] Differentiation and data smoothing were carried out with the Savitzky–Golay algorithm[33] (differentiation order 1, polynomial order 2, window width 15). The more discrete nature of the CropScan data did not allow for these pre-treatments. All spectral data sets were mean centred. After preprocessing, a good multivariate calibration could be obtained by a smaller number of latent variables than without MSC and differentiation (FieldSpec3 and HySpex data). Additionally, multivariate calibrations were performed using undifferentiated data from the FieldSpec3 and HySpex. The regression coefficients resulting from these calibrations allowed for a better visual comparison between instruments in terms of important wavelengths (i.e. peaks in regression coefficients). In the present study, the word peak is used for a sharp, local minimum/maximum in regression coefficients. Moreover, the regression coefficients were used as approximations for "integrated values" of the set of coefficients originating from the differentiated data, thus allowing easier interpretation of peaks in absorption/reflection.

For HySpex, all the spectra had a considerable amount of noise in the upper part of the range, above 850 nm, and were discarded from further analysis. Preliminary tests revealed that all the FieldSpec3 data from the spectral region above 1325 nm also contained much noise (SNR < ~2). All prediction models improved when data from the highest spectral range was removed. Consequently, the range above 1325 nm was omitted.

Calibration sample selection

All data sets (160 samples) were divided equally between a calibration set and a validation set. In order to make the two data sets as equal as possible, nearest neighbour (NN) clustering, based on the ideas of Fix and Hodges,[34] was used. Using the algorithm, 80 pairs (clusters) were created for each dataset (containing 160 samples). The distance calculations were performed on score values from a principal component transformation of the original data matrix to reduce the computational load of the algorithm. One sample from each pair (cluster) was then selected for the calibration set, whereas the remaining samples were used in the validation set.

Model building and selection

To create models relating the reflectance data (X) to the reference variables (y), the PPLS algorithm[23] was used. PPLS is a generalisation of the traditional PLS1 algorithm.[24] The algorithm is relatively new, but has proven to be a useful tool for extracting relevant information from NIR spectra, thus making it suitable for our application. The selection of the numbers of PPLS components was performed as a conservative

Table 2. Descriptive statistics of the measured reference variables, comprising minimum, maximum and mean values, standard error (*SE*) and assumed reference standard error (SE_{REF}).

Variable	Minimum	Mean	Maximum	*SE*	SE_{REF}
Grain yield (g m^{-2})	151	638	812	172	5.0
Moisture (g 100 g^{-1})	14.3	18.6	24.4	2.62	0.13
Protein (g 100 g^{-1})	10.0	12.8	16.7	1.53	0.15
Starch (g 100 g^{-1})	63.9	65.9	68.3	1.06	1.0
Gluten (g 100 g^{-1})	19.4	30.9	40.2	5.28	1.5
Zeleny (–)	28.5	42.6	65.0	11.3	5.0

chi-squared test. One major difference between PPLS and PLS is the introduction of the control parameter γ, which has the ability to direct the focus of the PPLS algorithm. A γ value of 0.5 degenerates PPLS to the PLS solution, whereas choosing values close to unity, the algorithm focuses almost exclusively on the predictors with strong predictive ability. Values below 0.5 make the algorithm focus on predictors with high variance. When using the PPLS algorithm for modelling, the γ parameter was optimised with golden section search and parabolic interpolation. To determine the significance of differences between prediction errors, chi-squared testing was performed in accordance with Indahl.[23] The PPLS algorithm was also used for variable selection.

Creating simple multiple least squares regression models: an alternative approach

In the sequence described above, a complex method relating reflectance data to reference variables is described. By using multivariate methods on the entire spectra, the method aims to explore all information contained in the spectral measurement.

As an alternative to the complex method above, we wanted to test whether it would be possible to achieve similar results by simply performing a variable selection (i.e. pinpointing the three to five wavelengths which contain most of the information in the data) and applying MLR on the retained variables. Variable selections based on PPLS models between FieldSpec3 data and the reference variables were carried out. All PPLS models from Z65 were run with $\gamma = 1$, which corresponds to variable selection based on correlation only. For easier interpretability of the selected variables, all models in this section were run on the undifferentiated spectral data. The wavelengths that had loading weights smaller than the relative floating point accuracy in Matlab (i.e. 10^{-16}) were discarded from the computation, as proposed by Indahl.[23] The remaining variables were used as input to an MLR modelling procedure. As with previous models, the regression coefficients were calculated with a calibration set (50% of the samples) and validated with a validation set (the other 50% of the samples). The r^2 coefficient for each model was computed and reported along with the selected wavelengths of each model. All regression coefficients were tested at 95% significance level.

Results

The measured reference variables were intercorrelated (Table 3), particularly protein concentration, starch concentration, gluten concentration and Zeleny sedimentation value. The starch concentration component was highly negatively correlated with all the other variables. The weakest correlation was found between grain yield and the Zeleny sedimentation value. The experimental error, expressed as the coefficient of variation (*CV*) for the N100 plots of each replicate block, was low (*CV* < 8.5%).

Shortly after heading of the spring wheat (Z65), radiometric measurements were performed with all the three instruments. Strongest relations between validation set predictions and reference measurements were found when using data obtained with the hand-held instruments (Figure 2, Table 4). The method based on airborne data acquisition yielded overall the poorest results, except for the prediction of grain yield and moisture, which did not differ between instruments (Tables 4 and 5).

The reference variables were generally predicted best when using the models developed using the data from the FieldSpec3 (Figure 2 and Table 5), and the predictions were significantly better at Z65 for protein, starch and gluten concentration and for Zeleny sedimentation value (Figure 2 and Table 5). The CropScan models gave significantly superior predictions of the same variables relative to the HySpex models (Figure 2 and Table 5). The model performances did not differ significantly between any of the three instruments for grain yield and moisture at Z65 (Figure 2 and Table 5).

The models derived from FieldSpec3 and HySpex data gave high γ values (Table 4). The models fitted to the CropScan data differed, giving γ values equal or close to 0.50. The regression models derived from the 13-wavelength CropScan instrument gained very little predictive ability when using more than two PPLS components, unlike the models derived from the hyperspectral instruments (FieldSpec3 and HySpex). The latter could, in some cases, benefit from up to four and three

Table 3. Correlation matrix (*r* values) for all combinations of the reference variables.

Variable	Protein ($g\,g^{-1}$)	Starch ($g\,g^{-1}$)	Gluten ($g\,g^{-1}$)	Zeleny (–)	Moisture ($g\,g^{-1}$)	Grain yield ($g\,m^{-2}$)
Protein ($g\,100\,g^{-1}$)	1	—	—	—	—	—
Starch ($g\,100\,g^{-1}$)	–0.98	1	—	—	—	—
Gluten ($g\,100\,g^{-1}$)	0.99	–0.98	1	—	—	—
Zeleny (–)	0.97	–0.94	0.95	1	—	—
Moisture ($g\,100\,g^{-1}$)	0.86	–0.85	0.87	0.83	1	—
Grain yield ($g\,m^{-2}$)	0.71	–0.75	0.78	0.60	0.70	1

Entries in the table show correlation coefficients for the corresponding row and column labels. All *P* values for testing the hypothesis of no correlation were less than 10^{-16}.

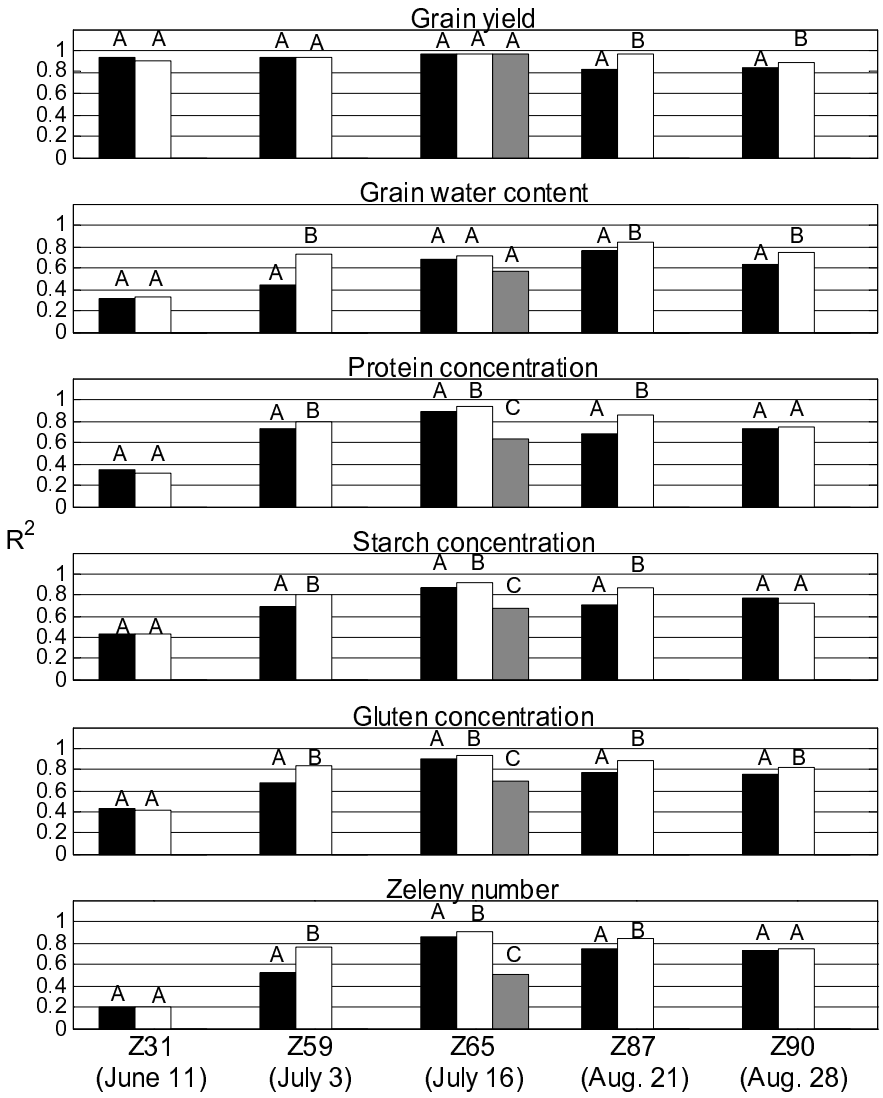


Figure 2. Relations (r^2) between measurements and predictions of the data for the validation set, with prediction models based on measurements conducted at five development stages of spring wheat by means of two hand-held radiometers, CropScan (filled bars) and FieldSpec3 (open bars) and an airborne hyperspectral scanner (stage Z65 only, grey bars). Bars with same letter (within the same day/stage) indicate non-significant differences in model fit (P=0.05).

PPLS components, respectively (Table 4). The growth development stage affected the results. Best overall model performance was obtained when the radiometric measurements were performed at Z65, regardless of the instrument used (Figure 2). This coincided in time with the only data acquisition performed with the airborne instrument. The models tended to improve their prediction ability with development stage of the wheat, from the first measurements at Z31 and up to the peak at Z65 (both hand-held instruments). From Z65 onwards, the predictions were poorer, except for moisture, which was best predicted with data measured at Z87. Grain yield was best predicted of all the reference variables (Table 4). The yield measurements (and the corresponding predictions) appeared to have two different clusters (Figure 3, left plots). The data points in the low-yielding data all represented plots which were not fertilised. The predictions of the grain quality data were overall poorer than the corresponding predictions of grain yield (Table 4). Removing the N0-treatment from the Z65 FieldSpec3-models reduced the r^2 to 0.73 for grain yield, 0.67 for moisture, 0.90 for protein concentration, 0.87 for starch concentration, 0.90 for gluten concentration and 0.89 for Zeleny sedimentation value.

The regression coefficients of the FieldSpec3 and the HySpex grain yield models appeared to have a comparable pattern (Figure 3). Both instruments gave large peaks in the regression coefficients (undifferentiated data) above and

Table 4. Relations between measurements and estimates based on sensor readings performed at Z65, and related model diagnostics for each of the three radiometric instruments.

Sensor and variable	r^2_{cal} [a]	r^2_{pred} [b]	*RMSEC*[c]	*RMSEP*[d]	γ[e]	#comp[f]
CropScan						
Grain yield ($g\,m^{-2}$)	0.96	0.98	30.2	25.0	0.50	2
Moisture ($g\,100\,g^{-1}$)	0.68	0.68	1.46	1.43	0.50	2
Protein ($g\,100\,g^{-1}$)	0.88	0.82	0.50	0.64	0.50	2
Starch ($g\,100\,g^{-1}$)	0.87	0.83	0.37	0.43	0.50	2
Gluten ($g\,100\,g^{-1}$)	0.89	0.84	1.64	2.06	0.50	2
Zeleny (–)	0.85	0.79	4.26	5.15	0.54	2
FieldSpec3						
Grain yield ($g\,m^{-2}$)	0.97	0.96	31.8	32.9	0.95	1
Moisture ($g\,100\,g^{-1}$)	0.71	0.74	1.42	1.30	0.88	3
Protein ($g\,100\,g^{-1}$)	0.94	0.92	0.39	0.40	0.95	3
Starch ($g\,100\,g^{-1}$)	0.92	0.91	0.30	0.29	0.89	4
Gluten ($g\,100\,g^{-1}$)	0.94	0.92	1.32	1.36	0.91	3
Zeleny (–)	0.90	0.88	3.69	3.70	0.93	3
HySpex						
Grain yield ($g\,m^{-2}$)	0.96	0.95	33.6	35.6	0.50	3
Moisture ($g\,100\,g^{-1}$)	0.57	0.60	1.59	1.73	0.95	2
Protein($g\,100\,g^{-1}$)	0.63	0.71	0.89	0.82	0.95	1
Starch ($g\,100\,g^{-1}$)	0.67	0.74	0.60	0.54	0.95	1
Gluten ($g\,100\,g^{-1}$)	0.69	0.75	2.87	2.61	0.95	1
Zeleny (–)	0.51	0.63	7.75	6.97	0.95	1

[a]Coefficient of determination from the calibration set
[b]Coefficient of determination from the prediction set
[c]Root mean square error of calibration
[d]Root mean square error of prediction
[e]Optimised γ parameter from the PPLS algorithm
[f]Final number of PPLS components.

below 530 nm and around 740 nm. With the CropScan grain yield model, the largest peaks were at 1650 nm, 660 nm and 560 nm. In contrast, both the FieldSpec3 and the HySpex models had a local minimum at 660 nm. The FieldSpec3 model (undifferentiated data), however, also had peaks at 800–900 nm, 980 nm and 1070 nm.

All the protein models had a very pronounced peak at 740 nm, along with two much smaller peaks, at 660 nm and 690 nm

Table 5. Test values of the chi-squared test for differences between instrument performances using data obtained at Z65.

Variable	FS/CS[a]	FS/HS[b]	HS/CS[c]
Grain yield ($g\,m^{-2}$)	0.948 (NS)	0.948 (NS)	0.899 (NS)
Moisture ($w\,w^{-1}$)	0.971 (NS)	0.891 (NS)	0.915 (NS)
Protein ($w\,w^{-1}$)	0.771	0.434	0.564
Starch ($w\,w^{-1}$)	0.803	0.498	0.620
Gluten ($w\,w^{-1}$)	0.801	0.459	0.572
Zeleny (–)	0.865	0.476	0.550

[a]FieldSpec vs. CropScan.
[b]FieldSpec vs. HySpex
[c]HySpex vs. CropScan
NS, Non-significant differences with a critical test value of 0.869 (P=0.05)

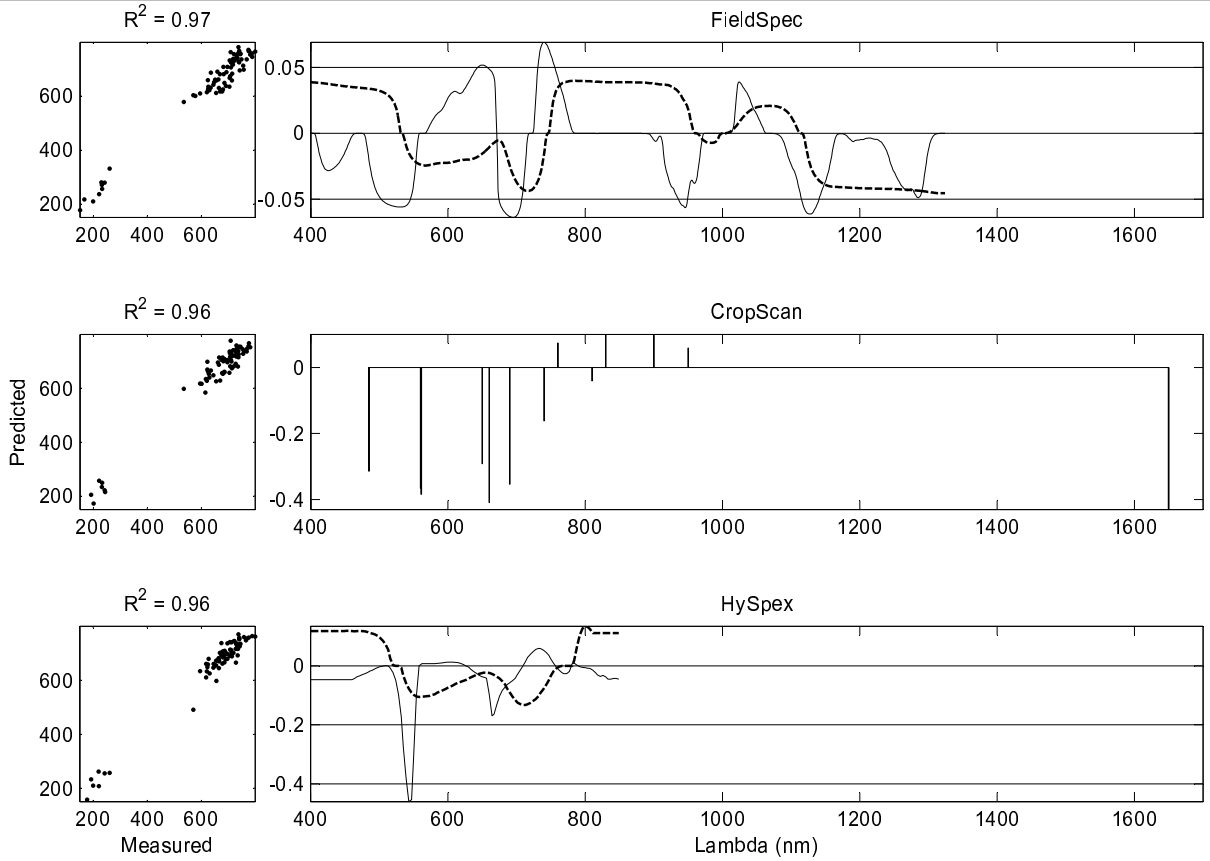


Figure 3. Validation-set measurements of grain yield (g m^{-2}) at harvest versus predictions based on radiometric measurements performed at Z65 (left-hand plots), and regression coefficients of the selected prediction models (right-hand plots, solid lines). The pre-treatment included differentiation (first derivatives) of the FieldSpec3 and HySpex data, but not of the CropScan data. Regression coefficients (normalised values) of models based on undifferentiated data are indicated for FieldSpec3 and Hyspex (broken lines).

(Figure 4, undifferentiated data). The FieldSpec3 had unique peaks at 715 nm, 990 nm and 1100 nm. The HySpex model, however, had two distinct peaks at 490 nm and 630 nm. The 490 nm peak corresponded with a peak in the CropScan model, but the peak at 630 nm was unique for the HySpex model.

The relatively simple MLR models, using predictors (wavelengths) selected by a PPLS variable selection procedure, resulted in validation set predictions which were almost as good as those given by the full PPLS-models (Tables 4 and 6). There was no significant difference in model performance for any of the reference variables. For grain yield and moisture, the PPLS procedure selected three wavelengths, while the models for protein, starch, gluten and Zeleny sedimentation value consisted of five wavelengths (Table 6). All models, except protein and gluten, differed in terms of selected wavelengths.

Discussion

In this study, two hand-held spectroradiometers and one airborne hyperspectral scanner were compared in terms of their usefulness for predicting selected properties of spring wheat grain. A recent multivariate method (PPLS) was tested in the analyses and compared with the traditional PLS1 method. The PPLS method was first used to create models using the full data set (i.e. all measured wavelengths). As a simpler alternative, the PPLS algorithm was used to select the three to five wavelengths containing most of the information in each dataset, followed by MLR, in order to test their combined prediction ability.

General points

The model performances changed during plant development (Figure 2). Since all the reference variables were measured at harvest time, one could assume that the predictive performance of the models should increase with decreasing time to harvest. This was not the case here, as the most favourable time for spectral measurements appeared to be after heading but before the yellow ripening stage. This is in agreement with Hansen *et al.*[4] who predicted yield and protein content of spring wheat under Danish conditions. For winter wheat grown in China, the most favourable growth

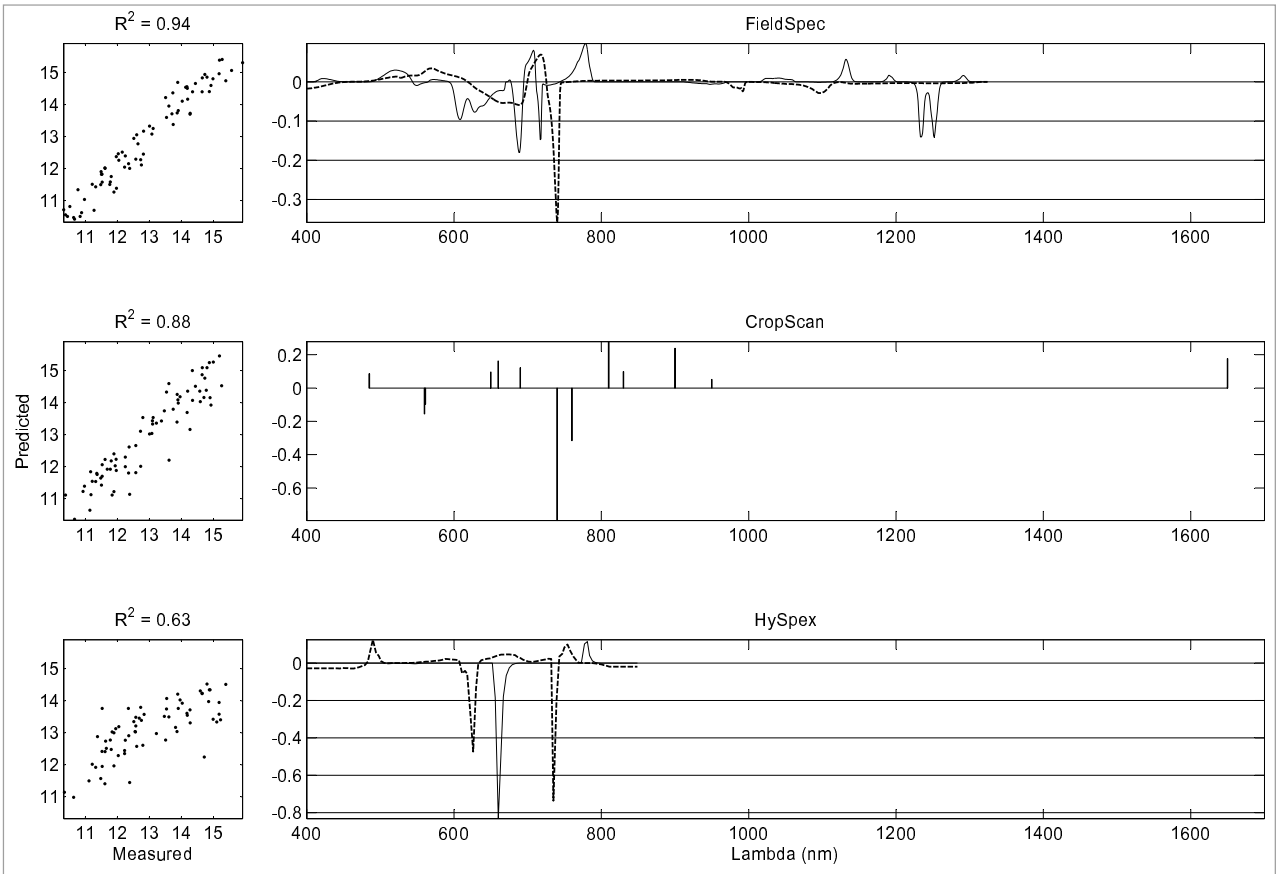


Figure 4. Validation-set measurements of grain protein concentration (g 100 g^{-1}) at harvest versus predictions based on radiometric measurements performed at Z65 (left plots), and regression coefficients (r^2) of the selected prediction models (right plots, solid lines). The pre-treatment included differentiation (first derivatives) of the FieldSpec3 and HySpex data, but not of the CropScan data. Regression coefficients (normalised values) of models based on undifferentiated data are indicated for FieldSpec3 and Hyspex (broken lines).

stage for predicting grain protein content has been reported to be at mid-filling.[7]

During the growth season, spectral signatures of the wheat plants change considerably, reflecting the change in colour and structure as the plants go through the various development stages. The results of this study and others[4,7] show that for predictions of grain quality properties, spectral measurements should be performed when the wheat plants are in their generative stage. It appears, however, that information obtained long before grain filling is completed may be relevant for the final quality features such as grain protein concentration.

The accumulation and transfer of plant nitrogen into the grain is the key feature for determining the final grain protein content[5] and high plant nitrogen status during grain filling is thus an indicator of high final grain protein content. High correlations between plant nitrogen status and leaf chlorophyll concentration have been reported[2,35,36] and the regression coefficients of the protein models showed some peaks which were in the chlorophyll absorption bands (Figure 4). Towards maturity of the wheat plants, there was a continuous reduction in leaf chlorophyll concentration observed as "greenness". Our data indicate that the spectral signature of the maturing canopy contained less information related to the grain cereal quality than that of more indirect indicators such as plant N status during grain filling.

The observed peaks in the regression coefficients of all the models (Figures 3 and 4) may be explained by chlorophyll absorbance (peaks at 490 nm, 630 nm, 660 nm and 690 nm), whereas peaks in the area 710–740 nm are most likely due to the rapid rise of plant spectral reflectance above 700 nm, namely the red-edge phenomenon.[37] The FieldSpec3 models showed additional peaks in areas of the spectrum where water is known to absorb radiation (for example, the peaks at 980–990 nm), and were other chemical bonds may have affected the measurements (for example, the peaks at 800–900 nm, C–H stretch and/or amines; at 1100 nm, C–H stretch; and at 1070 nm, N–H stretch). The relationships between the spectral measurements and the reference variables are quite indirect. Hence, further interpretation of the spectral peaks was difficult and, therefore, omitted.

Table 6. Coefficients of determination (r^2) for the validation set from multiple least squares regression models, using only wavelengths (nm) selected by a powered partial least squares variable selection algorithm as predictors.

Variable, and regression coefficient	r^2_{val}	λ1	λ2	λ3	λ4	λ5	λ6	λ7	λ8	λ9	λ10	λ11	λ12
Grain yield (g m^{-2})	0.95		487			716		922					
Coefficient			−17,542			−1925.1		1394.2					
Moisture (w w^{-1})	0.72						739			993			1105
Coefficient							−223			212			−43.4
Protein (w w^{-1})	0.90	388		533	673		740					1019	
Coefficient		477		−75.9	−47.2		−93.8					63.6	
Starch (w w^{-1})	0.89			533			739		970	989			1109
Coefficient				190			127		100	−76.0			−35.2
Gluten (w w^{-1})	0.89	382		533	673		739					1016	
Coefficient		165		−89.2	−314		−318					203	
Zeleny (–)	0.86			534			741		968	990	1000		
Coefficient				73.3			−132		278		−244		

There were strong correlations between several reference variables (Table 3). In particular, the protein concentration, starch concentration, gluten concentration and Zeleny sedimentation value were strongly intercorrelated ($|r| > 0.94$). The correlations were reflected in the regression coefficients of the models (i.e. they were quite similar). This means that successful prediction of one of the correlated variables most likely involves successful prediction of all the four quality variables. Hence, the question arises of what is really being measured. Does the method react to central characteristic bonds of the chemical structure of the grain protein components or do we detect some other features which happen to be highly correlated with grain protein? The spectral region of interest for predicting protein in single wheat kernels is usually 850–1050 nm.[38] The five FieldSpec3 wavelengths selected for protein were 388 nm, 533 nm, 673 nm, 740 nm and 1019 nm (Table 6). Hence, only the latter wavelength was within the range of typical protein-related absorption bands. The other selected wavelengths may be related mostly to chlorophyll and moisture absorption. Although the quality variables were highly intercorrelated, the wavelengths selected for grain yield, moisture, protein, starch and Zeleny sedimentation value diverged, indicating several different indirect causal relationships between spectral reflectance measurements and the reference variables.

Instrument comparison

The FieldSpec3 performed better than the CropScan in most cases (Figure 2) and the two hand-held instruments were superior to the airborne instrument for all the reference variables, except grain yield and moisture (Table 5). This may be explained by differences between instruments in their spectral range, spectral resolution, data spacing, recording altitude, data acquisition time span and calibration procedures.

Both the hand-held instruments measured radiation well into the shortwave infrared range of the electromagnetic spectrum. When inspecting the regression coefficients of the prediction models, we noticed several peaks for the hand-held instruments in the spectral area beyond 850 nm (Figures 3 and 4). This implies that this spectral area contained information relevant for predicting the quality variables, information unavailable from the HySpex models.

The instruments compared diverged not only in spectral range, but also in spectral resolution and data spacing. Inspection of the FieldSpec3 regression coefficients of the Z65 models (Figures 3 and 4, right hand plots) revealed several sharp peaks, indicating that there were some very local phenomena in the underlying spectral measurements. The presence of such local phenomena favours instruments with high spectral resolution and dense data spacing, such as FieldSpec3. The superiority of the FieldSpec3 relative to the CropScan (Figure 2) may be partly due to the large data spacing in the CropScan spectra between 950 nm and 1650 nm.

The recording altitudes of the instruments varied greatly. Higher recording altitude exposes the instrument data to more atmospheric influence. These effects were corrected

for in the ATCOR-4 atmospheric correction procedure of the HySpex data. The reflectance spectra of the HySpex data showed more noise than the corresponding spectra of the hand-held instruments (data not shown). One possible reason for the noise in the HySpex data could originate from the correction procedure, since violation of the model assumptions made in the ATCOR-4 software may cause erroneous reflectance spectra.[30] During HySpex data recording, a 2 × 2 m panel with known spectral reflectance was placed on the ground next to the field experiment. A regression equation was fitted between the raw HySpex data of the reference panel and the known reflectance spectrum of the reference panel. The resulting regression equation was then applied to the remainder of the raw HySpex data to achieve locally calibrated reflectance data. The resulting reflectance spectra showed the same behaviour as the reflectance values from ATCOR-4 ($R^2 > 0.96$). Thus, we can assume that the noise in the data did not originate from the ATCOR-4 correction procedure.

The data acquisition time of the HySpex lasted only a few seconds, whereas the hand-held spectroradiometer data was collected over a time span of approximately 2 h. The sun lighting conditions inevitably varied during this period, but frequent calibrations of the spectroradiometers should minimise that effect. The calibration procedures differed, however, between the instruments. The CropScan had a dual optic system which constantly corrected for changing solar irradiance and was thus assumed to be the instrument which adapted best to changing conditions. The FieldSpec3 was calibrated on a Spectralon panel and relied completely on frequent calibrations in order to provide good data. During data collection with the FieldSpec3 instruments, consecutive calibration readings were taken on the Spectralon panel and, if more than 5% spectral drift was observed, the previous measurement cycle was repeated. In this study, the measurement times were carefully selected with respect to stable weather conditions and as little cloud as possible and the observed spectral drift of the FieldSpec3 was generally low. The HySpex was calibrated using the ATCOR-4[29,30] procedure. The favourable weather conditions at the time of measurement (Z65) combined with the short data acquisition time of the HySpex instrument, give reason to believe that the difference in calibration method was not an important reason for poorer results of the HySpex instrument.

Another reason for weaker model performance in the last part of the growing season may be poorer weather conditions. The two last dates for spectroradiometer measurements (Z87 and Z90) had slightly less favourable weather and lighting conditions (i.e. increasing cloud cover during measurement) than the first three dates. Variable cloud cover may have caused an increase in the measurement noise, particularly for the FieldSpec3, which was not equipped with the dual optic calibration feature as the CropScan was. Chang *et al*.[39] reported that the CropScan instrument was less prone to the effects of variable cloud cover than the FieldSpec instrument. The FieldSpec3 models performed better, however, than the CropScan models during the last two measurements, which indicates that the weather and lighting conditions were not of great importance for the results.

The data acquisition times with the hand-held instruments were in the range of 2 h for a field experiment of 160 plots. Evidently, these kinds of instrument are not suitable for large-scale remote sensing applications due to their slow operation. At the other extreme, HySpex measured the whole field experiment in a matter of seconds. If it is acceptable to sacrifice some predictive performance for the advantage of high productivity, HySpex would be the obvious choice for large-scale remote sensing of spring wheat properties.

The main difference regarding productivity between the two hand-held instruments is the calibration procedure. CropScan has no need for recalibration during a measurement run and is also less sensitive to rapidly changing light conditions than FieldSpec3, due to the dual optic calibration feature.[39] From an operational and productivity point of view, CropScan would be the most desirable hand-held instrument. On the other hand, FieldSpec3 provides a spectrum with high spectral resolution and with a very broad spectral range. If the objective is to conduct a variable selection for isolation of important wavelengths (predictors) in an unknown prediction task beforehand, FieldSpec3 is superior by far to CropScan. The outstanding performance of the FieldSpec3 comes at the cost of more cumbersome handling, operation and calibration than the less sophisticated, but more user-friendly, CropScan.

Method comparison

For the Z65 models, the PPLS algorithm was superior to the PLS1 algorithm in all but six cases (indicated by $\gamma \neq 0.50$, Table 4). For five of these exceptions, CropScan data were used. The 13 wavelengths measured by the CropScan instrument were selected on the basis of the results obtained by Hansen *et al*.[4] who used CropScan for predictions of grain yield and grain protein in wheat. Thus, it came as no surprise that the selected CropScan setup was also well suited for the prediction tasks in the present study. Inspection of the regression coefficients (Figures 3 and 4, right plots) showed that all the CropScan wavelengths contributed to the models (large coefficients). This was, however, not the case for the FieldSpec3 and the HySpex instruments, for which several channels did not contribute to the models (coefficients near zero). The spectral data in the areas not contributing to the model would have a certain amount of variance and, thus, contribute to the y–X covariance computed in the PLS1 algorithm.[24] Having such data in the models is known to cause problems with the PLS1 algorithm in cases where there is high variance in predictors that simultaneously have little correlation with the reference variables.[23] The PPLS algorithm reduces this problem by weighting variance and correlation differently, which may explain the observed superiority of this method for model building based on dense datasets (i.e. many wavelengths).

Predictions based on variable selection and multiple least squares regression

After the PPLS variable selection, the grain yield and moisture could be predicted with only three wavelengths in the lower NIR region. The selected sets of wavelengths differed, however. Protein concentration, starch concentration, gluten concentration and Zeleny sedimentation value could be predicted with five wavelengths in the visible/NIR region. Although some wavelengths of the models predicting the quality variables were the same, one to three out of the five selected wavelengths were unique for each quality variable, except for the protein and gluten models, which were similar (Table 6). This implies that three of the four strongly correlated variables can still, to some extent, be separated. All computed r^2 values from the final MLR models had the same level as the r^2 values from the full PPLS models, although the MLR models were computed without using the first derivatives. Thus, the PPLS algorithm seemed to have very strong properties for variable selection. Several of the wavelengths selected were similar, or close, to an existing CropScan wavelength, confirming that many of the wavelengths in the CropScan were selected quite well for crop monitoring purposes, although there is apparently room for improvement. In contrast, several of the PPLS-selected wavelengths were outside the effective spectral range of theHySpex (400–850 nm). Thus, extending the effective spectral range of this instrument to approximately 1240 nm would probably improve its performance considerably.

All models presented in this study performed very well compared to numerous other studies aimed at predicting grain yield and grain protein.[4,40,41–43] Considering the models from Z65, only two models had predicted r^2 below 0.8. All the other models had r^2 considerably above 0.8 and five models achieved r^2 well above 0.9. However, predictions at almost the same level have been presented,[4,43] but these were the only studies that utilised PLS and PLS-like methods. Numerous studies utilise vegetation indices such as the normalised difference vegetation index (NDVI).[17] The predictions based on vegetation indices hardly ever reach the same level of accuracy as predictions made with the PLS family of methods. A commonly used argument for using vegetation indices is that they are more robust.[20,40,41,44] Since this study aims to be an instrument comparison, the robustness of the models was not tested in this study. However, independent validation datasets are currently being collected and model robustness will be addressed in a future study. Preliminary results from the ongoing work indicate that the models presented here are robust.

Conclusion

The most favourable time to predict wheat grain yield and grain quality was after heading but before yellow ripening. The results imply that monitoring and forecasting of wheat yield and quality by means of spectral reflectance can be carried out relatively early in the season and with high prediction accuracy.

The FieldSpec3 was the instrument with the best overall performance. Radiometric data comprising a wide spectral range, a large number of wavelengths and a high spectral resolution have an advantage over simpler datasets as a basis for building predictive models. Instruments providing such data are, however, often expensive and cumbersome to handle, but they represent a good starting point for identifying key variables.

Good predictions can be achieved with a very low number of variables. Thus, screening of a dense radiometric dataset with the PPLS method appears to be a promising tool for developing simple and robust prediction models. This concept should be investigated further in other research areas.

Acknowledgements

Thanks are due to Daniel Schläpfer for reflectance processing of the HySpex images with the ATCOR-4 software, to Ulf Indahl for making the PPLS MATLAB code available, to Torkel Gaardløs and Terje Tandsæther for invaluable help with the field data collection and to Juan Antonio Fernández Pierna, Vincent Baeten and Pierre Dardenne at the Waloon Centre of Agricultural Research, Belgium, for many good ideas and interesting discussions.

References

1. J.R. Jorgensen and R.N. Jorgensen, "Uniformity of wheat yield and quality using sensor assisted application of nitrogen", *Precision Agric.* **8(1/2),** 63 (2007). doi: 10.1007/s11119-006-9029-2
2. M. Reyniers and E. Vrindts, "Measuring wheat nitrogen status from space and ground-based platform", *Int. J. Remote Sensing* **27(3),** 549 (2006). doi: 10.1080/01431160500117907
3. E. Zillmann, S. Graeff, J. Link, W.D. Batchelor and W. Claupein, "Assessment of cereal nitrogen requirements derived by optical on-the-go sensors on heterogeneous soils", *Agron. J.* **98(3),** 682 (2006). doi: 10.2134/agronj2005.0253
4. P.M. Hansen, J.R. Jørgensen and A. Thomsen, "Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression", *J. Agric. Sci.* **139,** 307 (2002). doi: 10.1017/S0021859602002320
5. L.Y. Liu, J. Wang, Y. Bao, W. Huang, Z. Ma and C. Zhao, "Predicting winter wheat condition, grain yield and protein content using multi-temporal EnviSat-ASAR and Landsat TM satellite images", *Int. J. Remote Sensing* **27(4),** 737 (2006). doi: 10.1080/01431160500296867
6. Z.J. Wang, J.H. Wang, L.Y. Liu, W.H. Huang, C.J. Zhao and C.Z. Wang, "Prediction of grain protein content in winter wheat (*Triticum aestivum* L.) using plant pigment

ratio (PPR)", *Field Crops Res*. **90(2/3),** 311 (2004). doi: 10.1016/j.fcr.2004.04.004

7. L.H. Xue, W.X. Cao and L.Z. Yang, "Predicting grain yield and protein content in winter wheat at different N supply levels using canopy reflectance spectra", *Pedosphere* **17(5),** 646 (2007). doi: 10.1016/S1002-0160(07)60077-0

8. W. Huang, D.W. Lamb, Z. Niu, Y. Zhang, L. Liu and J. Wang, "Identification of yellow rust in wheat using *in-situ* spectral reflectance measurements and airborne hyperspectral imaging", *Precision Agric*. **8(4/5),** 187 (2007). doi: 10.1007/s11119-007-9038-9

9. S.M Vicente-Serrano, "Evaluating the impact of drought using remote sensing in a Mediterranean, semi-arid region", *Natural Hazards* **40(1),** 173 (2007). doi: 10.1007/s11069-006-0009-7

10. K.L. Castro-Esau, G.A. Sanchez-Azofeifa and B. Rivard, "Comparison of spectral indices obtained using multiple spectroradiometers", *Remote Sensing of Environment* **103(3),** 276 (2006). doi: 10.1016/j.rse.2005.01.019

11. M. Shibayama, "Radiometric estimation of canopy leaf inclination angles of various crop species using multiband polarization and reflectance", *Plant Prod. Sci*. **9(2),** 156 (2006). doi: 10.1626/pps.9.156

12. P.S. Thenkabail, E.A. Enclona, M.S. Ashton and B. Van Den Meer, "Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications", *Remote Sensing of Environment* **91(3/4),** 354 (2004). doi: 10.1016/j.rse.2004.03.013

13. M. Reyniers, E. Vrindts and J. De Baerdemaeker, "Comparison of an aerial-based system and an on the ground continuous measuring device to predict yield of winter wheat", *Eur. J. Agron*. **24(2),** 87 (2006). doi: 10.1016/j.eja.2005.05.002

14. D.B. Lobell and G.P. Asner, "Comparison of Earth Observing-1 ALI and Landsat ETM+ for crop identification and yield prediction in Mexico", *IEEE Trans. Geosci. Remote Sensing* **41(6),** 1277 (2003). doi: 10.1109/TGRS.2003.812909

15. R.E.E. Jongschaap, "Run-time calibration of simulation models by integrating remote sensing estimates of leaf area index and canopy nitrogen", *Eur. J. Agron*. **24(4),** 316 (2006). doi: 10.1016/j.eja.2005.10.009

16. F. Kendoul, D. Lara, I. Fantoni-Coichot and R. Lozano, "Real-time nonlinear embedded control for an autonomous quadrotor helicopter", *J. Guidance Control Dynamics* **30(4),** 1049 (2007). doi: 10.2514/1.27882

17. C.F Jordan, "Derivation of leaf-area index from quality of light on the forest floor", *Ecology* **50(4),** 663 (1969). doi: 10.2307/1936256

18. A.A. Gitelson and M.N. Merzlyak, "Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll", *J. Plant Physiol*. **148(3/4),** 494 (1996).

19. C.J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation", *Remote Sensing of Environment* **8(2),** 127 (1979). doi: 10.1016/0034-4257(79)90013-0

20. N.H. Broge and E. Leblanc, "Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density", *Remote Sensing of Environment* **76(2),** 156 (2001). doi: 10.1016/S0034-4257(00)00197-8

21. D.H. Zhao, J.L. Li and J.G. Qi, "Identification of red and NIR spectral regions and vegetative indices for discrimination of cotton nitrogen stress and growth stage", *Comput. Electron. Agric.* **48(2),** 155 (2005). doi: 10.1016/j.compag.2005.03.003

22. C. Yang, J.H. Everitt and J.M. Bradford, "Airborne hyperspectral imagery and yield monitor data for estimating grain sorghum yield variability", *Trans. ASAE* **47(3),** 915 (2004).

23. U. Indahl, "A twist to partial least squares regression", *J. Chemometr.* **19(1),** 32 (2005). doi: 10.1002/cem.904

24. S. Wold, H. Martens and H. Wold, "The multivariate calibration-problem in chemistry solved by the PLS method", *Lecture Notes in Mathematics* **973,** 286 (1983). doi: 10.1007/BFb0062108

25. A. Korsaeth, "Soil apparent electrical conductivity (ECa) as a means of monitoring changes in soil inorganic N on heterogeneous morainic soils in SE Norway during two growing seasons", *Nutr. Cycl. Agroecosyst.* **72(3),** 213 (2005). doi: 10.1007/s10705-005-1668-6

26. J.C. Zadoks, T.T. Chang and C.F. Konzak, "Decimal code for growth stages of cereals", *Weed Res*. **14(6),** 415 (1974). doi: 10.1111/j.1365-3180.1974.tb01084.x

27. D.R. Peddle, "Reflectance processing of remote sensing spectroradiometer data", *Computers Geosci*. **27(2),** 203 (2001). doi: 10.1016/S0098-3004(00)00096-0

28. A. Berk, L.S. Berstein, G.P. Anderson, P.K. Acharya, D.C. Robertson, J.H. Chetwynd and S.M. Adler-Golden, "MODTRAN cloud and multiple scattering upgrades with application to AVIRIS", *Remote Sensing of Environment* **65(3),** 367 (1998). doi: 10.1016/S0034-4257(98)00045-5

29. D. Schlapfer and R. Richter, "Geo-atmospheric processing of airborne imaging spectrometry data. Part 1: Parametric orthorectification", *Int. J. Remote Sensing*, **23(13),** 2609 (2002). doi: 10.1080/01431160110115825

30. R. Richter and D. Schlapfer, "Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: Atmospheric/topographic correction", *Int. J. Remote Sensing*, **23(13),** 2631 (2002). doi: 10.1080/01431160110115834

31. P. Geladi, D. Macdougall and H. Martens, "Linearization and scatter-correction for near-infrared reflectance spectra of meat", *Appl. Spectros*. **39(3),** 491 (1985). doi: 10.1366/0003702854248656

32. T. Isaksson and T. Naes, "The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy", *Appl. Spectros*. **42(7),** 1273 (1988). doi: 10.1366/0003702884429869

33. A. Savitzky and M.J.E. Golay, "Smoothing + differentiation of data by simplified least squares procedures", *Anal. Chem.* **36(8),** 1627 (1964). doi: 10.1021/ac60214a047
34. E. Fix and J.L. Hodges, "Discriminatory analysis—nonparametric discrimination—consistency properties", *Int. Stat. Rev.* **57(3),** 238 (1989). doi: 10.2307/1403797
35. N. Rama Rao, P.K. Garg, S.K. Ghosh and V.K. Dadhwal "Estimation of leaf total chlorophyll and nitrogen concentrations using hyperspectral satellite imagery", *J. Agric. Sci.* **146,** 65 (2008). doi: 10.1017/S0021859607007514
36. C.J. Zhao, L. Lin J. Wang, W. Huang, X. Song and C. Li, "Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress", *Int. J. Appl. Earth Observation and Geoinformation*, **7(1),** 1 (2005). doi: 10.1016/j.jag.2004.10.002
37. D.N.H. Horler, M. Dockray and J. Barber, "The red edge of plant leaf reflectance", *Int. J. Remote Sensing* **4(2),** 273 (1983). doi: 10.1080/01431168308948546
38. S.R. Delwiche, "Single wheat kernel analysis by near-infrared transmittance–protein content", *Cereal Chem.* **72(1),** 11 (1995).
39. J. Chang, S.A. Clay and D.E. Clay, "Clouds influence precision and accuracy of ground-based spectroradiometers", *Commun. Soil Sci. Plant Anal.* **36(13/14),** 1799 (2005). doi: 10.1081/CSS-200062449
40. J. Marti, J. Bort, G.A. Slafer and J.L. Araus, "Can wheat yield be assessed by early measurements of normalized difference vegetation index?", *Ann. Appl. Biol.* **150(2),** 253 (2007). doi: 10.1111/j.1744-7348.2007.00126.x
41. L. Wall, D. Larocque and P.M. Leger, "The early explanatory power of NDVI in crop yield modelling", *Int. J. Remote Sensing* **29(8),** 2211 (2008). doi: 10.1080/01431160701395252
42. Y. Zhu, X. Yao, Y.-C. Tian, X.-J. Liu and W.-X. Cao, "Analysis of common canopy vegetation indices for indicating leaf nitrogen accumulations in wheat and rice", *Int. J. Appl. Earth Observation and Geoinformation* **10(1),** 1 (2008). doi: 10.1016/j.jag.2007.02.006
43. T. Jensen, A. Apan, G. Young and L. Zeller, "Detecting the attributes of a wheat crop using digital imagery acquired from a low-altitude platform", *Computers and Electronics in Agriculture* **59(1/2),** 66 (2007). doi: 10.1016/j.compag.2007.05.004
44. N. Aparicio, D. Villegas, J. Casadesus, J.L. Araus and C. Royo, "Spectral vegetation indices as nondestructive tools for determining durum wheat yield", *Agron. J.* **92(1),** 83 (2000). doi: 10.2134/agronj2000.92183x<

PAPER II

Prediction error improvements using variable selection on small calibration sets – a comparison of some recent methods

Stein Ivar Øvergaard[1,3,*], Juan Antonio Fernández-Pierna[2], Vincent Baeten[2], Pierre Dardenne[2], Tomas Isaksson[4]

[1]The Norwegian Institute for Agricultural and Environmental Research, 2849 Kapp, Norway

[2]Walloon Agricultural Research Centre, Valorisation of Agricultural Products department, Food and Feed Quality Unit, Gembloux, Belgium

[3]The Norwegian University of Life Sciences, Department of Mathematical Sciences and Technology, 1430 Ås, Norway

[4]The Norwegian University of Life Sciences, Department of Chemistry, Biotechnology and Food Science, 1430 Ås, Norway

[*]Corresponding author: stein.ivar.overgaard@bioforsk.no

Keywords:

BVSPLS; FSS; NIR; PLS; PPLS; Variable selection

Abstract:

Variable selection provides useful information about the most important predictors in the dataset, information which is not always available in the beginning of an analysis. Two recent variable selection methods, Backward Variable Selection for Partial Least Squares (BVSPLS) and Powered Partial Least Squares (PPLS), were compared against each other and against benchmark methods in terms of their ability to produce accurate prediction models in NIR spectroscopy data. These two variable selection methods were compared to the benchmarks Forward Stepwise Selection (FSS) and full spectrum Partial Least Squares (PLS). All four regression methods were studied using three different NIR datasets. PPLS and BVSPLS gave good prediction results in all three datasets even with a very limited number of calibration samples available (<40). All methods gave similar prediction results when the number of calibration samples was higher (>150). PPLS gave the best predictive performance of all methods and also gave the selections of variables that were most easily assigned to specific chemical bonds. Hence, the PPLS models were more easily interpretable than the other models. This study quantifies differences between the two recent variable selection methods as well as the differences between recent methods and established benchmark methods.

1. Introduction

Regression is probably the most widely studied and applied statistical analysis method in the chemometric literature. The aim is to develop models which can be used to predict properties of interest based on measurements of the chemical system, such as spectroscopic data. Multivariate calibration techniques such as Multiple Linear Regression (MLR), Principal Component Regression (PCR) and f Partial Least Squares regression (PLS)[1] can then be used to compute a mathematical model. It correlates the multivariate measurement (spectrum) to the concentration of the analyte of interest, and such a model can be used to predict the concentrations of new samples.

When the number of measured predictor variables is large and it is not known beforehand which specific predictors are most influential on the responses, selection of variables could be feasible. Variable selection tries to identify a subset of variables that still possess the sufficient features to build a robust regression model. Moreover, due to a number of practical and statistical reasons (e.g. to avoid collinearity, reduce computational load), a large set of variables should be reduced to a smaller, more manageable set. The main goal of any variable selection technique is to obtain a small subset of variables that gives a model with the prediction and generalization abilities better or at least equivalent to a model based on the original set of variables. Variable selection in regression is a difficult part of model building because the number of subsets to be considered grows exponentially with the number of candidate variables. The advantages of variable selection are the exclusion of irrelevant and redundant variables leading to better signal to noise ratio, better data visualization and model interpretability, reduction of measurement requirements as well as increased prediction accuracy and precision. Subsequently, these properties could induce the development of cheaper instruments, cheaper analysis as well

as faster prediction models. Moreover, an increase in the model robustness can be achieved by the application of a variable selection technique. One possible drawback of doing variable selection is that certain outlier detection methods may be more difficult to undertake. Numerous methods have been developed for variable selection such as varieties of subset selection methods[2, 3], stepwise regression, jack-knife or bootstrapping algorithms[4-7], evolutionary algorithms[8] genetic algorithms[9-12] and thresholding algorithms[13]. Recently, a Backward Variable Selection method for PLS regression (BVSPLS) has been proposed[14]. Another relatively recent method is the Powered Partial Least Squares[15], which is a generalization of the traditional Non-linear Iterative Partial Least Squares (NIPALS) algorithm. PPLS can also be used for variable selection purposes. The development of new variable selection methods is constantly evolving, and the need for comparative studies is raised. There are several studies aimed at comparing various methods[16], but due to the ongoing development of new methods, comparative studies will always be needed.

The objective of this paper is to compare between BVSPLS and PPLS and to some established benchmark methods. The most established, simplest and most pragmatic method for variable selection is Forward Stepwise Selection of Variables (FSS). There are several examples in the literature where varieties of FSS have been used as a reference methods[2, 16]. Moreover, the FSS algorithm is implemented in a multitude of data analysis software and is hence widely used. For this reasons, FSS is a natural choice of benchmark method for this comparative study. Moreover, the traditional PLS solution without any variable selection should also be included as a benchmark method in order to address the question whether variable selection itself has a positive effect for the predictive ability of the models. In order to validate the feasibility of the methods, we chose to compute prediction models based on small datasets.

2. Materials and methods

2.1 Datasets, preprocessing and sample selection

Three large datasets used in Fernández Pierna et al.[14] were also used in this study: fat/feed (Feed) , fiber/maize (Maize I) and protein/maize (Maize II). All datasets had spectral range from 100 nm to 2498 nm, with every second wavelength removed, thus containing 700 variables each. See Fernández Pierna et al.[14] for more details. There were some duplicate samples in the datasets and in order to achieve a proper validation, duplicate response variable values and their corresponding spectra were removed from the datasets. Hence, the number of samples in each dataset (N) was reduced to 2721 for fat/feed, 2488 for Maize I and 1349 for Maize II. All datasets were preprocessed with the Standard Normal Variate procedure SNV, [17]. Splitting of the data into calibration and validation sets was done with the DUPLEX algorithm [18]. This algorithm splits a dataset (i.e. the spectra) into two parts by means of a Euclidean distance measure. The algorithm goal is to create two datasets with homogeneous statistical properties for calibration and validation purposes. A subset of 200 samples from each dataset were selected with DUPLEX and reserved for calibration purposes. The remaining samples were allocated as validation set. Since we chose to work with smaller calibration sets, the 200 selected samples were further decimated to 20 samples in 19 steps with the DUPLEX algorithm. The first step selected 190 samples out of the original 200 samples and the second step selected 180 samples. For each successive step, the number of selected samples was decreased by 10. Thus, 19 calibration sets (200,190,…,20 samples) and one validation set (all samples except the 200 calibration samples) were calculated from each main dataset.

2.2 Software

All analyses were performed using MATLAB (version R2007b, MathWorks Inc., USA, www.mathworks.com) with PLS Toolbox (version 4.2, Eigenvector Research Inc., USA, www.eigenvector.com). The BVSPLS and DUPLEX algorithms were implemented in MATLAB code by the authors. The PPLS algorithm was implemented in MATLAB by the authors based on the original code from Ulf Indahl.

2.3 Variable selection methods

2.3.1 Forward Stepwise selection (FSS)

The FSS algorithm is a simple and widely used procedure for variable selection. Three different basic varieties of stepwise regression are commonly used: forward selection, backward elimination and stepwise method. Forward selection sequentially introduces new predictors into the model one at a time while the backward loop eliminates predictors one at a time from the current variable set. The stepwise method is a hybrid between forward selection and backward elimination. It starts as forward selection, but for each selection step it runs an elimination step to compute the need for deleting predictors. The algorithm uses a Fisher F-statistic in order to decide when variables should be removed or included. To construct the final prediction model, we used a PLS algorithm on the retained variables. We chose to set the inclusion and removal values such that the FSS algorithm selected 40-80 variables. The p-values for inclusion and removal were both set to 0.11 in order to let the FSS select approximately 40-80 variables in the used datasets. Selecting that many variables will almost certainly introduce some multicollinearity between the variables, but the PLS algorithm will handle this more robustly than for instance the least squares method in the MLR algorithm.

143

144 2.3.2 Backwards Variable Selection for PLS (BVSPLS)

145

146 The BVSPLS is a recently proposed method and is a backward elimination method. Unlike
147 the other algorithms in this study, the BVSPLS needs three datasets. In addition to the
148 usual calibration and validation set, a dedicated dataset (the stop-set) for decision on
149 which variables to retain is needed. The first algorithm step is to compute a PLS model
150 based on the full calibration dataset with all variables included. Consequently, one
151 variable is removed each time the algorithm loop executes. For each loop execution, the
152 RMSEP (Root Mean Square Error of Prediction) of the stop-set, is computed. When all
153 variables have been discarded, a plot of the RMSEP from the stop-set against the number
154 of variables can be presented. The algorithm then chooses the number of variables
155 corresponding to the minimum RMSEP. This subset of variables is then retained for the
156 final model which is a traditional PLS algorithm. For further details, see Fernández Pierna
157 et al. [14]. In order to provide a stop to BVSPLS, we chose to let 10% of the calibration
158 samples form the stop-set. Hence, for a 100 samples calibration set, 10 of the samples
159 (selected with DUPLEX) were put in the stop-set.

160

161 2.3.3 Powered Partial Lest Squares (PPLS)

162

163 The PPLS is a generalization of the traditional NIPALS algorithm. Rather than optimizing
164 the covariance between the predictors and the response, the PPLS splits the covariance
165 expression in the weight vector optimization criterion into a variance part and a
166 correlation part. The user can then choose the weighting between the variance component
167 and correlation component through an additional control parameter, gamma (γ). The
168 algorithm can be used both for modelling and variable selection through the choice of γ. A
169 γ value of 0.5 makes the PPLS solution equivalent to the traditional PLS solution, whereas

values close to 0 or 1 makes the algorithm select variables based on predictor variance and correlation with the response, respectively. It is also possible to let the algorithm optimize the y value within a predefined numerical range using an optimization procedure that maximises the correlation between PPLS scores and the response. In this study, we chose to let the PPLS work with y values optimized from the interval [0.99,1]. The algorithm hence focuses almost exclusively on the variables with strong corrrelation to the response and also possibly strong predictive ability. As suggested by Indahl[15], the variables that had loading weights less than the relative numerical resolution of MATLAB (2.2204×10^{-17}) were discarded. See Indahl[15] for further details.

2.4 Selection of optimal number of PLS/PPLS factors

All methods tested in this study have the feature of latent variables. Hence, model complexity has to be selected by the user. To make the resulting models more comparable, we chose to perform the selection of latent variables just once for each combination of method and dataset. For each dataset, the model complexity was determined on the basis of the complete calibration set of 200 samples and the number of factors was held constant throughout the whole range of calibration sets. Selection of the number of PLS/PPLS components was carried out as a conservative chi-square test. The main idea is to consider the minimum mean square error of cross-validation (MSECV) as a realization of the true model error variance, σ_0^2. Using the chi-square power function, an acceptance region for MSECV can be computed. The model with the fewest number of components that also have an MSECV inside the acceptance region is then selected as the final model. See Indahl [15] for further mathematical details. However, several numbers of factors for each model were computed and compared, but the differences between models were only modestly affected by the choice of number of factors.

2.5 Validation procedure

Each of the 19 calibration sets was used to construct a prediction model, which was used to predict the validation set, which was the same for all 19 models and for all four regression methods. We chose to use a dedicated validation set instead of cross-validation for comparison of the predictive ability of each model because several studies have pointed out that cross-validation can lead to severe over-fitting and over-optimistic estimated of the models diagnostic measures[2, 19, 20]. The test procedure was performed in the following way:

1. The number of PLS factors was determined by computing a model with the full calibration set of 200 samples. The same number of factors was used for every variable selection method.
2. Each variable selection algorithm was executed once on the 19 smaller calibration sets (200,190,…,20 samples), and predictions for the validation set were computed each time.
3. Based on these validation set predictions, coefficient of determination (R^2) between measured and predicted constituents for each method were computed and reported.
4. Step 2 to 3 was repeated for each main dataset (fat/feed, Maize I and Maize II).

Hence, R^2 for 20 to 200 calibration samples were obtained in a comparable way with 10 samples increments.

220 3. Results

221

222 Figures 2-3 show the coefficient of determination between measured and predicted
223 constituents from the validation-set predictions as a function of the number of calibration
224 samples for the Feed, Maize I and Maize II samples, respectively. For all datasets, variable
225 selection gave little or no prediction error improvement over the PLS algorithm for
226 calibration sets larger than approximately 60 samples (Feed, Fig. 1), 120 samples (Maize I,
227 Fig, 2) and 160 samples (Maize II, Fig. 3). For calibration datasets smaller than this, all
228 variable selection techniques gave better predictive performance compred to full spectrum
229 PLS. Especially for low sample numbers, the BVSPLS and PPLS gave better performance
230 than both PLS and FSS.
231 In the Feed dataset (Fig. 1), all models were stable at a high R^2 above 50 samples
232 calibration sets. From 50 to 20 calibration samples, BVSPLS and PLS started to show a
233 decrease in performance. Especially the PPLS had an advantage over the other methods for
234 calibration sets smaller than 50 samples.
235 For the Maize I case (Fig. 2), the situation was similar in the sense that all solutions were
236 stable at a relatively high level of explained variance for calibration sets larger than 120
237 samples. Here, the PPLS gave the best predictions for all datasets smaller than 120
238 samples with the BVSPLS slightly lower prediction ability. FSS had performance between
239 PLS and PPLS/BVSPLS for calibration sets smaller than 120 samples.
240 In the last dataset, the Maize II (Fig. 3), all methods performed poorer than in the first two
241 datasets. Full spectrum PLS gave the lowest prediction ability for datasets smaller than
242 170 samples. For datasets larger than 170 samples, there were just small differences
243 between the methods. For smaller datasets than 170 samples, the PPLS gave slightly better
244 performance than FSS and BVSPLS.

245

246 For all datasets, the PPLS and BVSPLS selected more variables for inclusion in the
247 prediction models than the FSS did (Fig. 4-6).
248
249 All variable selection methods differed markedly with regard to frequency of the selected
250 variables. The PPLS and FSS selected some variables in the spectrum more often than other
251 variables and especially PPLS gave a quite clear and structured image of which variables
252 that contributes positively to the prediction models. BVSPLS, however, selected variables
253 more evenly spread throughout the spectrum (Fig. 7-9).
254
255 To illustrate the improvement of variables selection (Fig. 10), an example for 60
256 calibration samples of the Feed dataset is illustrated (Fig. 1). We chose the PPLS and the
257 PLS models for this case and plotted the predicted fat content data against the measured
258 fat content from each method.

4. Discussion

Two recent methods, BVSPLS and PPLS, were compared against each other and against FSS and PLS methods. Even after extensive literature reviews, we could not find comparative studies on variable selection methods similar to BVSPLS and PPLS.

All variable selection techniques gave improvements over PLS in some cases and the improvements were more pronounced at smaller calibration sets. The numbers of PLS factors were determined using the full 200 samples calibration sets and the numbers of components found were held constant for all other calibration sets. The actual numbers of components were determined with a conservative Chi-square-test. We tried, however, several other model dimensionalities, but the general results and improvements in prediction ability were only slightly affected by this.

To explain the differences in predictive ability, we have pointed out three reasons. 1. Some predictor variables have only remote relevance to the response variable. 2. The signal to noise ratio (S/N) in some predictor variables may be so low that the elimination of those variables improves the model. 3. Some predictor variables may have a nonlinear relationship to the response. Thus, elimination of these variables may give more parsimonious and linear prediction models and hence improve the prediction abilities. All, or a subset of these reasons could explain the prediction error improvements that we have presented, but further research is needed to exploit the details in the mechanisms behind this phenomena.

FSS and especially PPLS selected very interpretable sets of variables (Fig. 7-9). For the Feed dataset, the PPLS and FSS emphasized strongly the C-H stretch bands in the 1700 nm range (Fig. 7). This tendency also held in the Maize I dataset where the PPLS and FSS selected many variables in the O-H stretch band at 1450 nm and the C=O stretch band from

2000 nm to 2200 nm (Fig 8). An even stronger interpretability was found in the Maize II dataset where the FSS and PPLS focused very strongly on the N-H stretch bands in the 1800 nm, 2000 nm and 2400 nm regions (Fig. 9). For the BVSPLS, the picture was more difficult to interpret because the algorithm selected variables almost evenly spread out in the measured spectrum (Fig. 7-9) but still with better prediction results than those obtained with PLS and similar as those of PPLS.

5. Conclusion

In this paper, several validation procedures were conducted on the recent variable selection methods BVSPLS and PPLS and compare these to the bechmarks FSS and ful spectrum PLS. The comparisons were carried out on three different NIR spectroscopic datasets predicting fat in compound feed (Feed), fiber in maize(Maize I) and protein in maize (Maize II). We have drawn three conclusions from this study.

1. Variable selection gave a positive effect on the prediction ability of small calibration sets. Since calibration samples are often costly to collect, this may be important finding in order to make the best regression models out of few calibration samples.

2 The results clearly showed a consistent and well-structured selection of variables. FSS and BVSPLS gave not so consistent variable selections as PPLS.

3. Both PPLS and BVSPLS showed very high ability to compute good prediction models on small datasets, clearly better than more established regression methods. This shows that variable selection techniques are evolving and requires continued comparisons with existing algorithms.

6. Acknowledgement

In 2009, the first author spent four weeks working at the Walloon Agricultural Research Centre (CRA-W) in Gembloux, Belgium. Thanks are due to all the staff at the CRA-W for their generosity many fruitful discussions. Also thanks to mr. Audun Korsaeth at The Norwegian Institute for Agricultural and Environmental Research (Bioforsk) for the funding that made this project possible.

References

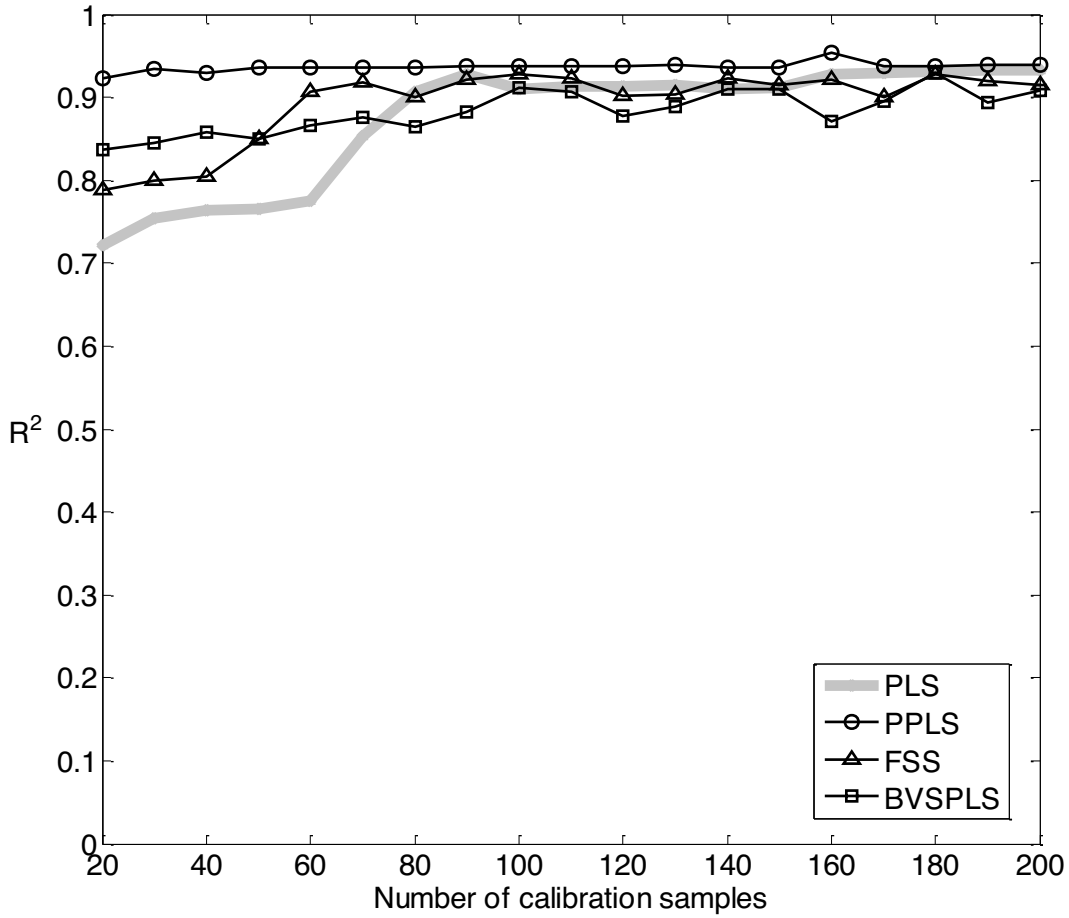
1. S. Wold, H. Martens, and H. Wold, "The Multivariate Calibration-Problem in Chemistry Solved by the Pls Method". *Lecture Notes in Mathematics*. **973,** 286 (1983)
2. L. Norgaard, et al., "Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy". *Applied Spectroscopy*. **54,** 413 (2000)
3. A. Hoskuldsson, "Variable and subset selection in PLS regression". *Chemometrics and Intelligent Laboratory Systems*. **55,** 23 (2001)
4. F. Westad and H. Martens, "Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression". *Journal of near Infrared Spectroscopy*. **8,** 117 (2000)
5. H. Martens and M. Martens, "Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR)". *Food Quality and Preference*. **11,** 5 (2000)
6. N. M. Faber, "Uncertainty estimation for multivariate regression coefficients". *Chemometrics and Intelligent Laboratory Systems*. **64,** 169 (2002)
7. R. Wehrens, H. Putter, and L. M. C. Buydens, "The bootstrap: a tutorial". *Chemometrics and Intelligent Laboratory Systems*. **54,** 35 (2000)
8. M. Shamsipur, et al., "Ant colony optimisation: a powerful tool for wavelength selection". *Journal of Chemometrics*. **20,** 146 (2006)
9. R. Leardi, M. B. Seasholtz, and R. J. Pell, "Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data". *Analytica Chimica Acta*. **461,** 189 (2002)
10. R. Leardi, "Application of genetic algorithm-PLS for feature selection in spectral data sets". *Journal of Chemometrics*. **14,** 643 (2000)
11. D. Broadhurst, et al., "Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry". *Analytica Chimica Acta*. **348,** 71 (1997)
12. D. Whitley, "An overview of evolutionary algorithms: practical issues and common pitfalls". *Information and Software Technology*. **43,** 817 (2001)
13. S. Saebo, et al., "ST-PLS: a multi-directional nearest shrunken centroid type classifier via PLS (vol 22, pg 54, 2008)". *Journal of Chemometrics*. **22,** 423 (2008)
14. J. A. F. Pierna, et al., "A Backward Variable Selection method for PLS regression (BVSPLS)". *Analytica Chimica Acta*. **642,** 89 (2009)
15. U. Indahl, "A twist to partial least squares regression". *Journal of Chemometrics*. **19,** 32 (2005)

16. I. G. Chong and C. H. Jun, "Performance of some variable selection methods when multicollinearity is present". *Chemometrics and Intelligent Laboratory Systems*. **78,** 103 (2005)
17. R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard Normal Variate Transformation and De-Trending of near-Infrared Diffuse Reflectance Spectra". *Applied Spectroscopy*. **43,** 772 (1989)
18. R. D. Snee, "Validation of Regression-Models - Methods and Examples". *Technometrics*. **19,** 415 (1977)
19. E. Anderssen, et al., "Reducing over-optimism in variable selection by cross-model validation". *Chemometrics and Intelligent Laboratory Systems*. **84,** 69 (2006)
20. C. M. Andersen and R. Bro, "Variable selection in regression-a tutorial". *Journal of Chemometrics*. **24,** 728 (2010)

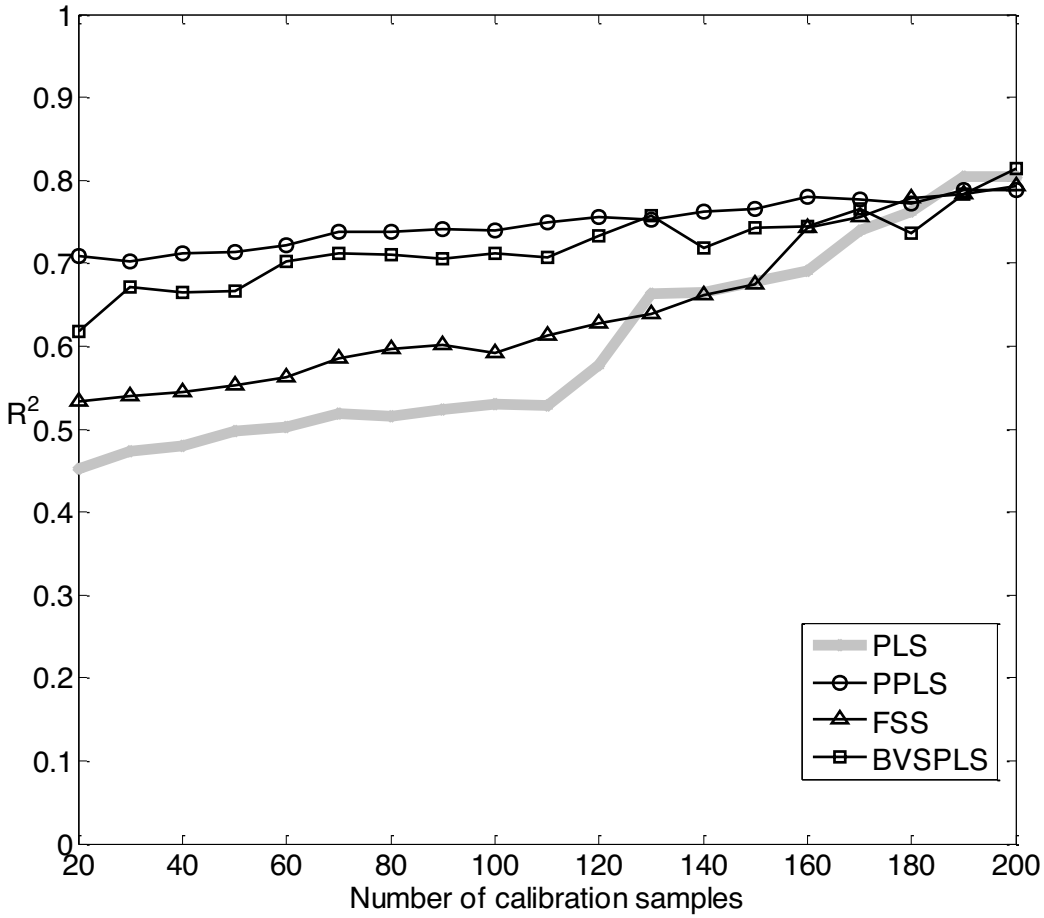
382 Table 1. Overview of the datasets (calibration and validation set together)

Product	Constituent	Range[1] [w/w - %]	STD[1] [w/w - %]	No. of validation samples	S_{REF}[2]
Feed	Fat	0.660 – 33.9	5.07	2521	0.20
Maize I	Fiber	24.3 – 67.3	6.82	2288	0.60
Maize II	Protein	4.02 – 13.7	1.60	1149	0.20

383
384 [1]Standard deviation for calibration and validation set together
385 [2]Standard error of reference method



386
387 Figure 1. Coefficient of determination (R^2) between measured and predicted constituents
388 from the validation-set predictions of the Feed dataset as a function of the number of
389 calibration samples for the fat content in feed mixture. The regression models used was
390 full spectrum Partial Least Squares (PLS), Partial Powered Least Squares (PPLS), Forward
391 Stepwise Selection (FSS) and Backwards Variable Selection for PLS (BVSPLS). The number of
392 PLS factors was 7 for all models.



393

394 Figure 2. Coefficient of determination (R^2) between measured and predicted constituents
395 from the validation-set predictions of the Maize I dataset as a function of the number of
396 calibration samples for the fiber content in maize. The regression models used was full
397 spectrum Partial Least Squares (PLS), Partial Powered Least Squares (PPLS), Forward
398 Stepwise Selection (FSS) and Backwards Variable Selection for PLS (BVSPLS). The number of
399 PLS factors was 8 for all models.

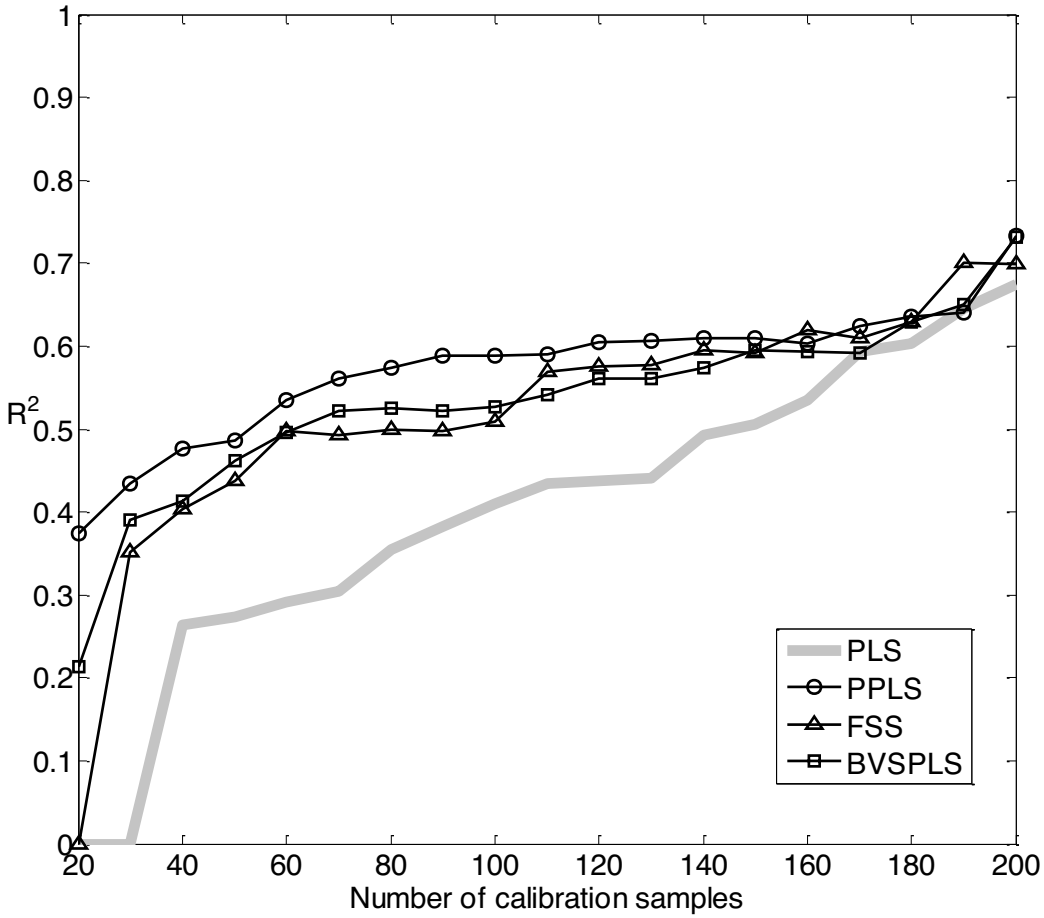
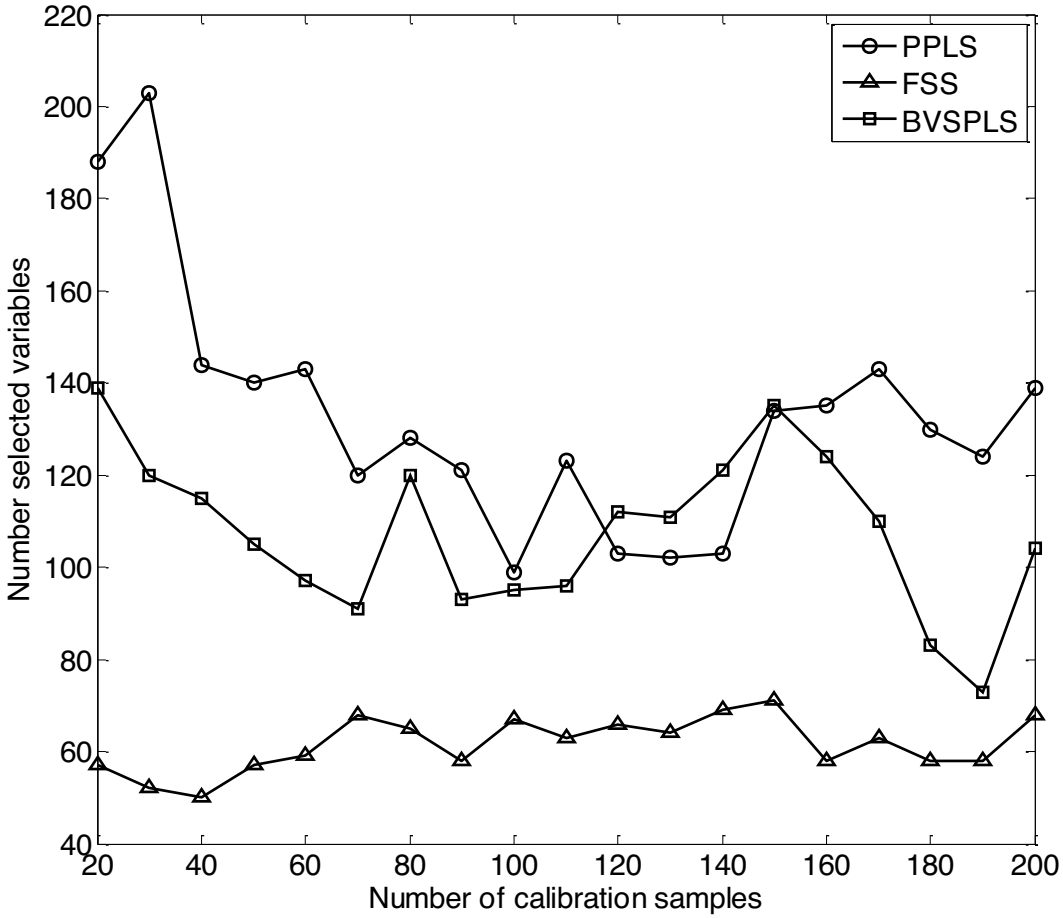
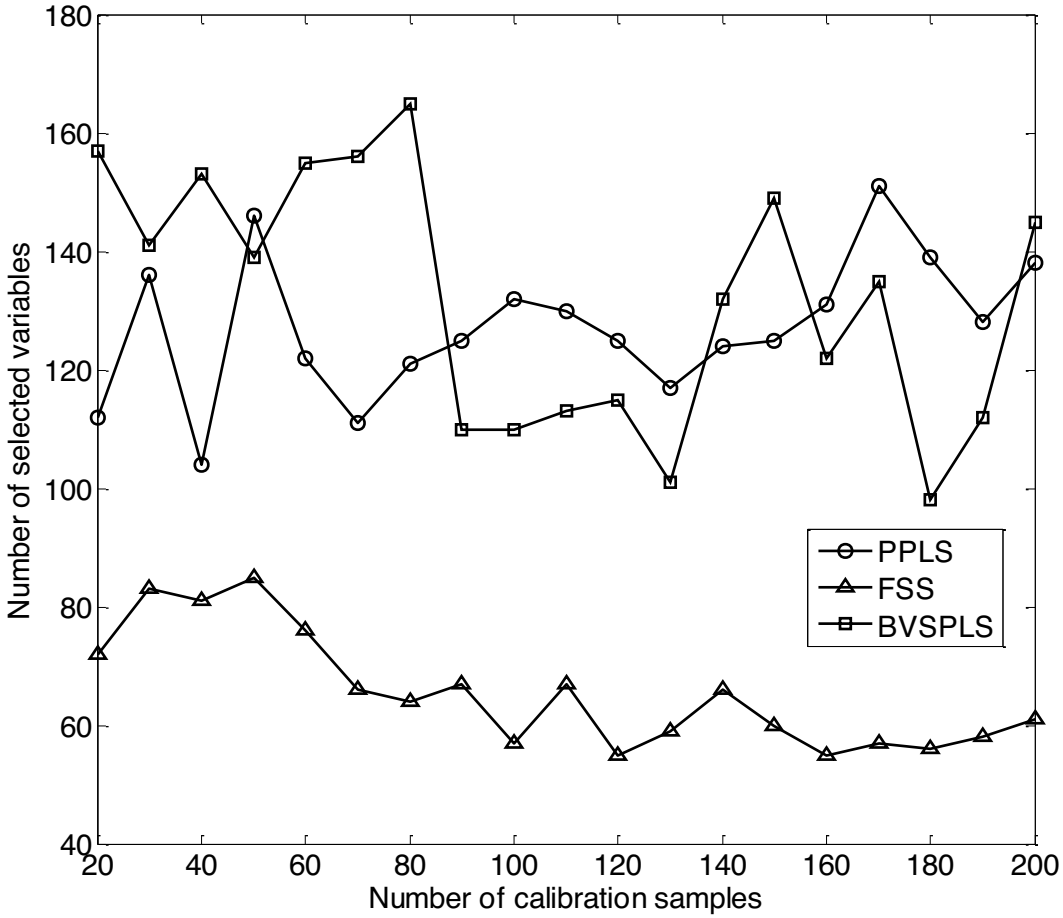


Figure 3. Coefficient of determination (R^2) between measured and predicted constituents from the validation-set predictions of the Maize II dataset as a function of the number of calibration samples for the protein content in maize. The regression models used was full spectrum Partial Least Squares (PLS), Partial Powered Least Squares (PPLS), Forward Stepwise Selection (FSS) and Backwards Variable Selection for PLS (BVSPLS). All models had 5 PLS factors.



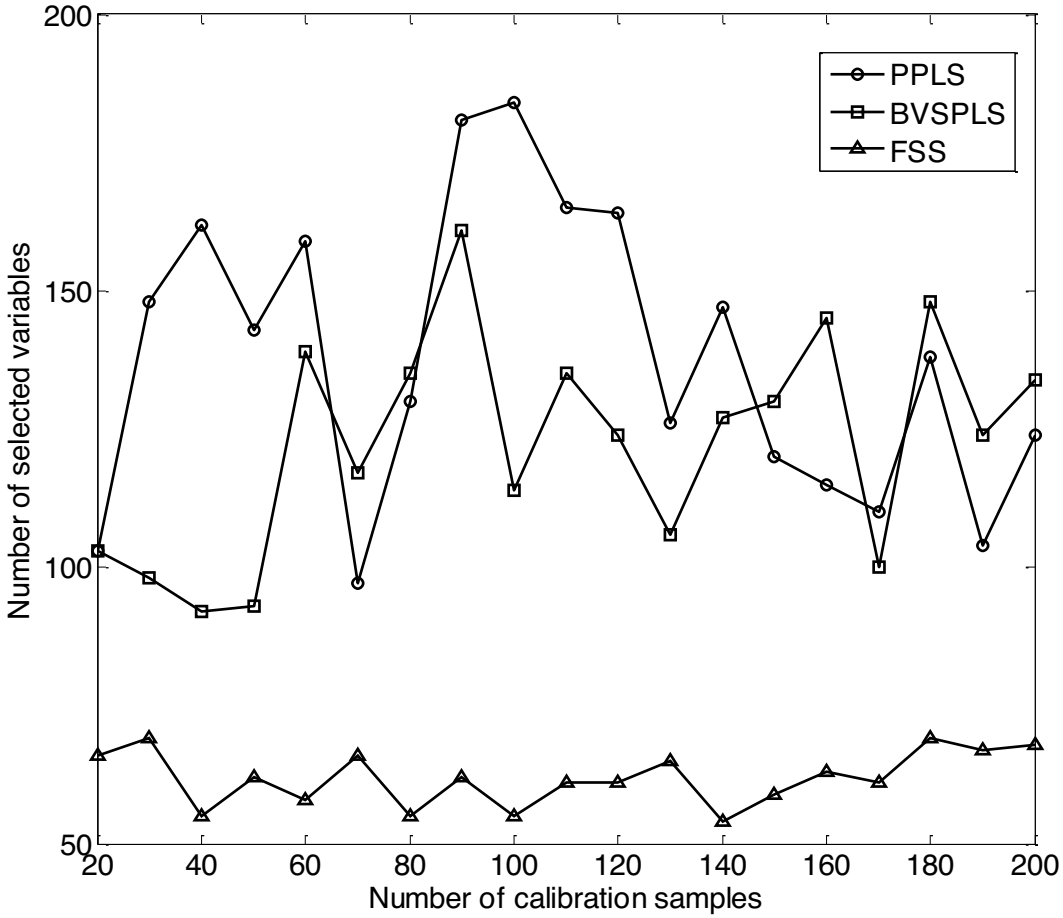
407

408 Figure 4. Number of selected variables for the three variable selection methods Powered
409 Partial Least Squares (PPLS), Forward Stepwise Selection (FSS) and Backwards Variable
410 Selection for PLS (BVSPLS) as a function of the number of calibration samples for fat
411 content prediction in the Feed dataset.



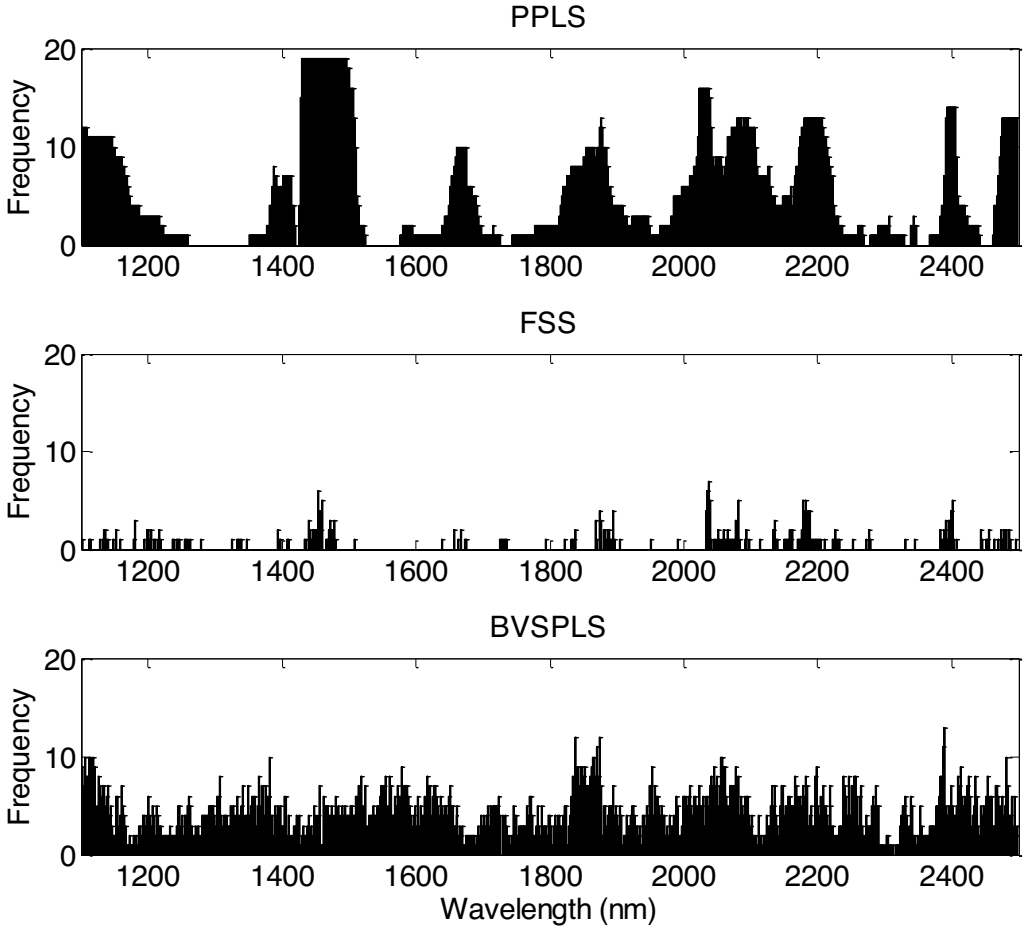
412

413 Figure 5. Number of selected variables for the three variable selection methods Powered
414 Partial Least Squares (PPLS), Forward Stepwise Selection (FSS) and Backwards Variable
415 Selection for PLS (BVSPLS) as a function of the number of calibration samples for fiber
416 content prediction in the Maize I dataset.



417

418 Figure 6. Number of selected variables for the three variable selection methods Powered
419 Partial Least Squares (PPLS), Forward Stepwise Selection (FSS) and Backwards Variable
420 Selection for PLS (BVSPLS) as a function of the number of calibration samples for protein
421 content prediction in the Maize II dataset.



422

423 Figure 7. Frequencies of the selected variables for Powered Partial Least Square (PPLS),
424 Forward Stepwise Selection (FSS) and Backwards Variable Selection for PLS (BVSPLS).
425 Histogram of the selected variables in each method (i.e. the number of times each variable
426 was selected of in total 19 models) versus the wavelength for fat content in the Feed
427 dataset.

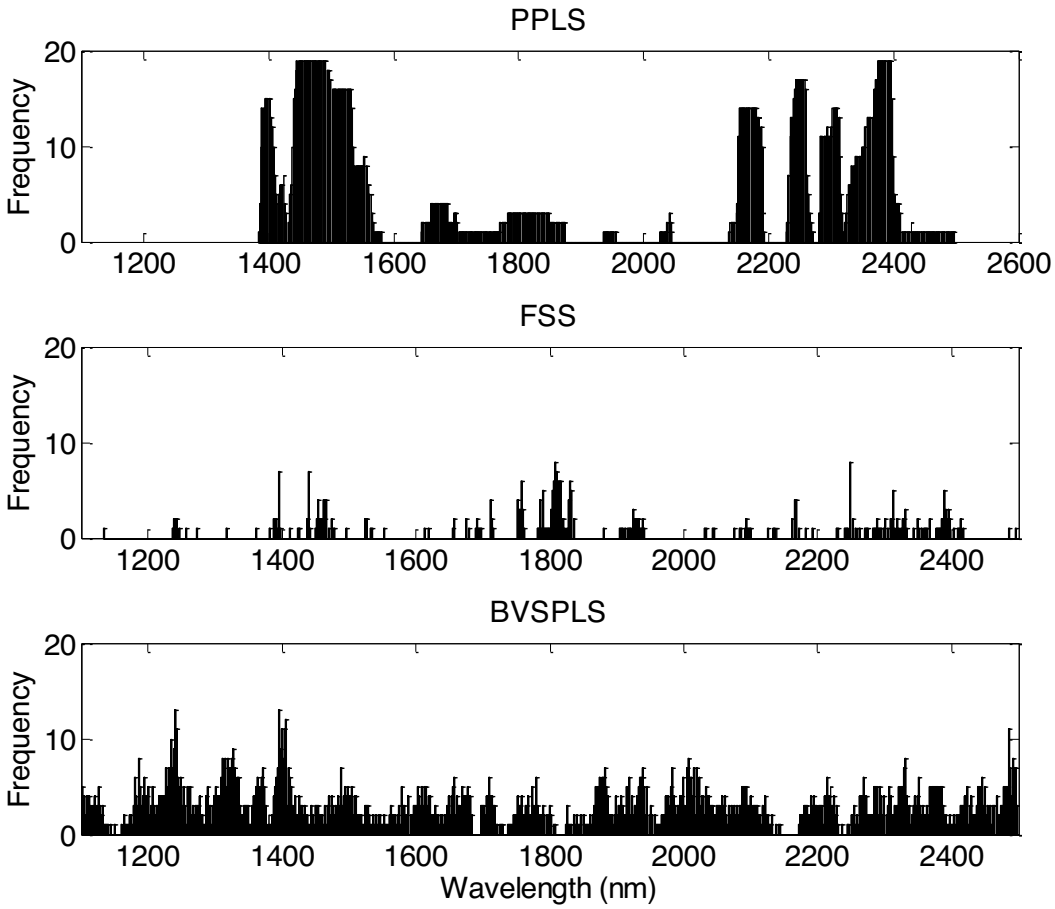
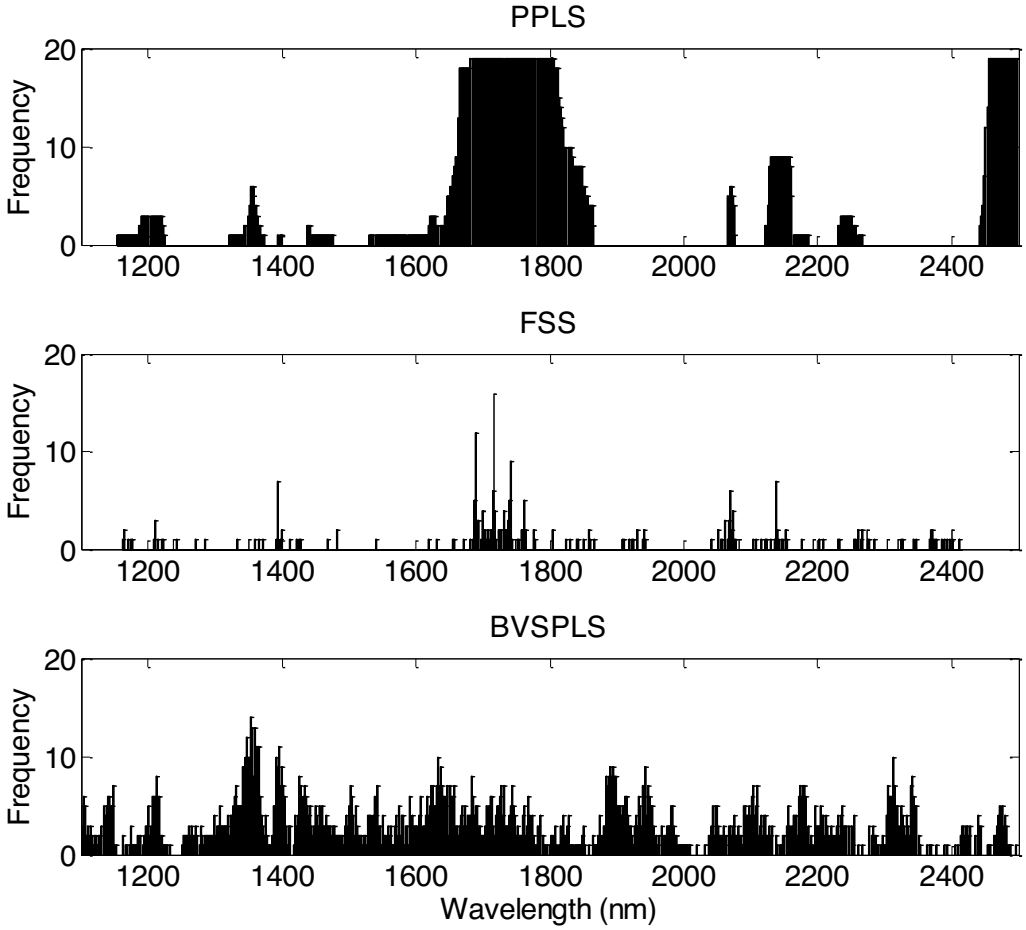


Figure 8. Frequencies of the selected variables for Powered Partial Least Square (PPLS), Forward Stepwise Selection (FSS) and Backwards Variable Selection for PLS (BVSPLS). Histogram of the selected variables in each method (i.e. the number of times each variable was selected of in total 19 models) versus the wavelength for the Maize I dataset.



433

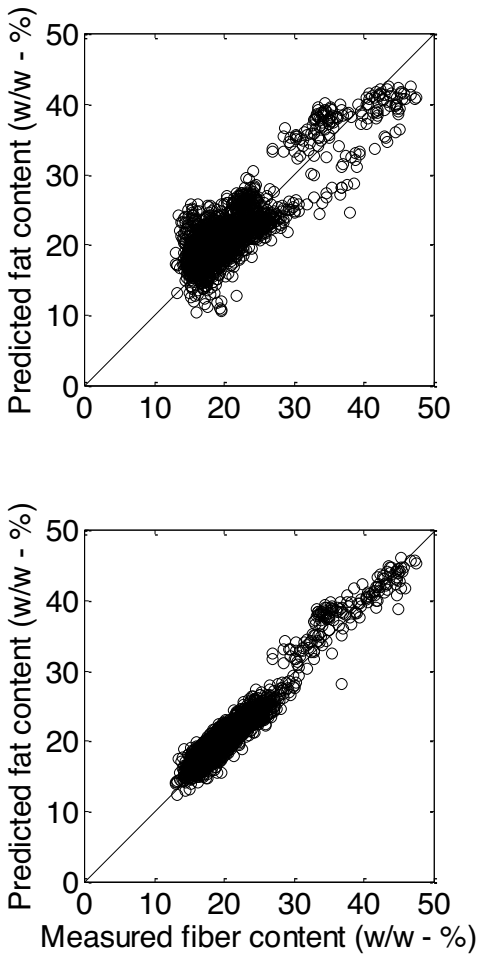
434 Figure 9. Frequencies of the selected variables for Powered Partial Least Square (PPLS),

435 Forward Stepwise Selection (FSS) and Backwards Variable Selection for PLS (BVSPLS).

436 Histogram of the selected variables in each method (i.e. the number of times each variable

437 was selected of in total 19 models) versus the wavelength for the Maize II dataset.

438



439

440 Figure 10. Predicted fat content from Partial Least Squares (PLS, upper plot) and Partial
441 Least Squares (PPLS, lower plot) in feed mix plotted against the measured fat content. The
442 PLS gave a coefficient of determination (R^2) between measured and predicted constituent
443 of 0.78, whereas the PPLS model gave R^2=0.93. Both models were calibrated using a
444 calibration set with 60 samples and validated on the full validation set (N=2521, Fig. 1).

PAPER III

Prediction of wheat yield and protein using remote sensors on grain fields - Part I: Assessing NIR model robustness for year and site variations

Øvergaard, S.I[1,3,*], Isaksson, T.[2], Korsaeth, A.[1]

[1]The Norwegian Institute for Agricultural and Environmental Research

[2]The Norwegian University of Life Sciences, Department of Chemistry, Biotechnology and Food Science, 1430 Ås, Norway

[3]The Norwegian University of Life Sciences, Department of Mathematical Sciences and Technology, 1430 Ås, Norway

Keywords

FieldSpec; Hyperspectral; Model robustness; Near infrared spectroscopy; Powered Partial Least Squares; Precision agriculture; Spectral reflectance; Variable selection; Vegetation index

16 Abstract

17

18

19 FieldSpec3, a portable field spectroradiometer was used to measure canopy
20 reflectance in spring wheat. Spectral reflectance data was collected on three
21 different experiment locations in up to four different years in the period
22 2007-2010, so that seven unique site-years were included, comprising altogether
23 976 individual plots. Prediction models for grain yield and grain protein
24 concentration were computed by means of the recent statistical method Powered
25 Partial Least Squares (PPLS). Several datasets had moderate to severe lodging,
26 which had a markedly negative influence on the prediction results. To correct for
27 this problem, a classification model for the classes “lodging” and “standing crop”
28 was calibrated from the spectral data. The model gave a total classification
29 accuracy of 98.3 %. Models were calibrated and validated on several combinations
30 of the spectral datasets in order to reveal spatial and temporal effects on the
31 prediction performance. The grain yield predictions explained 94 %
32 (RMSEP = 156 g m^{-2}) of the variance and the predictions of grain protein
33 concentration explained 67 % (RMSEP = 1.51 g DM 100 g^{-1}) of the variance. The
34 model performance generally increased with increasing variation in the calibration
35 data, both in time (i.e. more years included) and space (i.e. more sites included),
36 and the study showed that one year of spectral measurements is not sufficient for
37 building fully operational models for cereal property predictions. The performance
38 of the grain yield PPLS models was compared with that of models based on some
39 widely used vegetation indices (Normalized difference Vegetation Index (NDVI),
40 Modified Soil Adjusted Vegetation Index (MSAVI), Red Edge Inflection Point (REIP)

and d-chl-ab). The explained variance of the vegetation indices approach did not exceed 55 %, and this method was thus clearly inferior.

43 1. Introduction

44

45

46 Precision agriculture is concerned with using new technology in order to observe
47 and respond to within-field variation. A technology driven area which has expanded
48 rapidly in recent years, is reflectance remote sensing by the use of various forms of
49 spectroradiometers and cameras [1-6]. Several properties, covering a range of crops,
50 have been addressed using such techniques, with grain yield and protein
51 concentration of cereals being among the most frequently reported[7-10]. These are
52 properties of great practical and economical interest. Systems for yield mapping
53 are requested as a foundation for planning the next cropping season and for
54 locating potential problem areas (i.e. low yielding zones). Remote sensing
55 techniques represent a promising alternative to current yield mapping systems,
56 which are based on various monitors mounted on the combine harvester. The latter
57 systems rely heavily on challenging correction procedures, due to the many errors
58 that often occur during monitoring/data aquisition[11]. An operational system for
59 mapping grain protein concentration could be used for targeted harvesting of
60 cereals in order to maximize the quality and economic benefits from the crop.

61

62 Prediction models for such use have a high demand on robustness in order to be
63 fully operational, and they have to handle perturbed situations that are frequently
64 encountered in practice (e.g. lodging and various cases of misgrowth). Most studies
65 present prediction models which are based on only 1-2 years of data [5, 12-14]. Hence,

66 the models do not reflect the large variation between years that are often
67 encountered in practical agriculture. Other studies validate prediction models on
68 sites that lie in the same place and on the same type of soil as the calibration
69 experiment [15-18]. These models do not take into account the variations in soil type
70 and weather conditions during crop growth. The effects of variation between sites
71 and years on model calibration and their validation is as yet poorly understood, and
72 this appears to be a limiting factor for the conversion of scientific findings into
73 real-life applications.

74 In this study we have chosen to use the statistical-empirical approach, where a
75 model is calibrated from one dataset and validated on another, in an attempt to
76 make the model applicable to other situations[9, 15]. An alternative could have been
77 to use model inversion based on either the Radiative Transfer Equation (RTE) or
78 Kubelka-Munk theory[19, 20]. The model inversion approach was considered less
79 suitable for the purpose of this study, due to its high requirement of ground truth
80 data for calibration.

81

82 Several statistical-empirical studies rely on the use of previously developed
83 vegetation indices for the prediction of plant properties[5, 21-23]. Historically, such
84 indices were computed from 2-3 wavelengths in order to maximize the extraction
85 of data from a limited number of wavelengths, typically available from the first
86 satellite images. However, the use of such indices are still frequently reported[5, 24,
87 25] and an often cited rationale for this is that vegetation indices give more stable
88 and robust predictions of plant properties (i.e. under varying conditions). A very
89 limited number of studies utilize a larger part of the spectral signature without the

formation of vegetation indices[7, 8]. Literature is scarce on direct comparisons between these two main approaches.

We had two main objectives for this study. The first objective was to develop prediction models for grain yield and grain protein concentration in spring wheat, and to test how the model performances were affected by variation between years and sites. Our second objective was to compare the prediction performance of models based on some widely used vegetation indices with that of models based on full spectrum data. To achieve this, data from three experimental sites were collected in four different years, comprising seven unique site-years with altogether 976 individual plots.

2. Materials and methods

Spectroradiometer measurements were carried out in specially designed field-trials with spring wheat (*Triticum aestivum* L., var. 'Bjarne') in SE Norway, at Bioforsk Arable Crops Division. Reflectance measurements, grain yield data and grain protein concentration data were collected from three different experimental sites during the growing seasons 2007-2010. Not all of the sites were used each year, so that there were data from seven site-years altogether. On the basis of these data, prediction models for grain yield and grain protein concentration were computed and validated.

2.1 Field trials

Almost identical field trials were established at sites A, B and C in 2007, 2008 and 2010, respectively. The trials covered 26 x 160 m (Site C: 26 x 144 m), and comprised 20 (Site C: 18) replicate blocks (Fig. 1). Each block consisted of eight 2 x 8 m plots, with fixed locations in all years. The sites were selected so as to maximise variation in soil texture, organic matter and/or drainage status between blocks and to minimize it within replicate blocks. Six nitrogen (N) level treatments were applied within each block: 0, 100, 125, 150, 175 and 200 kg nitrogen ha^{-1} (designated N0, N100, N125, N150, N175 and N200, respectively). The border plots on either side of the randomized plots received 100 kg N ha^{-1}, thus giving three N100 plots per replicate block. Plant protection (herbicides, fungicides and insecticides) was carried out according to the current practice. Soil samples were

126 taken at 2 depths (0-25 cm and 25-60 cm) from each replicate block at site A and
127 B. At site C, samples were taken only from 0 25 cm. Soil samples were analysed for
128 total C content by dry combustion (CHN analyzer, Eurofins, Norway) and total N
129 (Kjeldahl, Eurofins, Norway). Moreover, the soil samples were analysed for particle
130 size , ignition-loss, pH, P-AL (AL denotes the ammonium lactate/acetic acid
131 mixture used for extraction[26]), Mg-AL, Ca-AL, Na-AL, K-AL and K-HNO_3. An overview
132 of these soil properties is given in Table 1.
133
134 Sites A and C were located at Apelsvoll farm (60° 42" N, 10° 51" E, 250 m asl),
135 which is part of Bioforsk Arable Crops Division, and which has a mean annual
136 precipitation of 600 mm, a mean annual temperature of 3.6° C and a mean growing
137 season (May-September) temperature of 12° C. The trials were established in 2007
138 (site A) and 2010 (site C) on imperfectly drained brown earth (Gleyed melanic
139 brunisol, Canada Soil Survey) with dominantly loam and silty sand textures, and
140 moderate soil variation (see table 1 for details).
141
142 The experimental area of site A slopes 3-6 % eastwards. Measurements from this
143 site were performed during all cropping seasons in the period 2007-2010. Due to
144 unfavourable weather conditions during some measurements sequences and
145 reduced labour capacity in 2008, some datasets do not contain measurements from
146 all 160 plots (Table 2). Data from 2007 have been published previously[27].
147
148 Site B is located on a nearby farm, Hoff (60° 41" N, 10° 51" E, 250 m asl), where
149 the experimental area slopes 1-7 % southwards. The weather conditions at site B
150 are comparable with those at sites A and C, but the site has a soil ranging from

151 imperfectly drained brown earth (Gleyed melanic brunisol, Canada Soil Survey) to
152 humified peaty gley (Terric humisol, Canada Soil Survey). The variations in the
153 measured soil properties were largest for site B, and this applied particularly to
154 ignition-loss and total N and C (table 1). The trial at site B was established in 2008
155 and was run for two years. In both 2008 and 2009, there was some severe lodging
156 due to the combination of occasional heavy rain and probable high N mineralization
157 in the blocks dominated by peaty soil.
158
159 The experimental area of site C, which slopes 2-5 % eastwards, is located approx.
160 700 m north of site A. The trial contains 18 replicate blocks (not 20 as those at site
161 A and B), due to field size limitations. Hence a total of 144 plots were present at
162 this trial.
163
164 2.2 Instrumentation and measurements
165
166 2.2.1 Radiometric field measurements
167
168 FieldSpec3 (version 3, Analytical Spectral Devices Inc., USA, www.asdi.com) is a
169 portable field spectroradiometer measuring spectra at one single spot in space,
170 whose area depends on the choice of foreoptic and measurement height. For the
171 current study, the foreoptic was operated at 1.1 m above the crop canopy with a
172 field of view angle of 25 °, corresponding to a spot area of 0.2 m^2. In order to
173 transform the measured signal into reflectance data, calibrations were performed
174 on a Spectralon panel provided and calibrated by LabSphere (LabSphere Inc, USA,
175 www.labsphere.com). The instrument measures 2150 wavelengths in the spectral

range from 350 nm to 2500 nm with a spectral resolution (Full Width at Half Maximum: FWHM) between 2.0 nm and 2.8 nm. Highest resolution is achieved at the shortest wavelength. For further details on instrument specifications and modes of operation, see[27]. The measurement regime used in[27] were duplicated for all seasons and experiments, i.e. reflectance spectra were collected from the experiments five times during each growing season at growth stages Z31, Z59, Z65, Z87 and Z90 according to the Zadoks decimal code[28]. All measurements were taken in duplicate and the instrument was recalibrated after every 40 measurements (approx. every 10 minutes). If a spectral drift > 5 % was observed between calibrations, the previous measurement cycle was repeated. Based on a screening performed at an earlier stage[27], only the data acquired at Z65 were used for computing prediction models in the current study. All measurements were taken as reflectance spectra and all models were also computed in reflectance units.

2.2.2 Measurements of yields and protein concentration

At maturity, each plot was harvested and a grain sample was taken. The grain moisture and dry matter (DM) were analysed gravimetrically and grain protein concentration was analysed with a FOSS Infratek Grain Analyser 1241 (FOSS Tecator, Denmark, www.foss.dk), using the calibration model number WH182126 provided by the manufacturer. For an overview of the grain properties, see Table 2.

2.3 Data analysis

All data analyses were carried out using MATLAB (version 2007b, MathWorks Inc., USA, www.mathworks.com). PLS Discriminant Analysis (PLS-DA) and data preprocessing routines were provided by PLS Toolbox (version 4.2, Eigenvector Research Inc., USA, www.eigenvector.com). The Powered Partial Least Squares (PPLS) algorithm was programmed by the authors based on the original code[29].

2.3.1 Preprocessing

The raw data were corrected for the detector splice point discontinuity at 1000 nm with a software routine provided by the instrument manufacturer (ViwSpec Pro, version 3.9, Analytical Spectral Devices Inc., USA, www.asdi.com). All spectral measurements were preprocessed with the Multiplicative Scattering Correction (MSC) method[30]. Thereafter, first derivatives were computed with the Savitzky-Golay algorithm[31] (differentiation order: 1, polynomial order: 2, window width: 15). All data were mean centered prior to analysis.

2.3.2 PPLS

For prediction model computation, the recent method Powered Partial Least Squares[29] was used. This variant of PLS has the ability to point its focus on variables with either high X variance or high X-y correlation through optimization of a method-specific parameter, called γ. When the γ value is preset to 0.5, the PPLS solution degenerates to the traditional PLS solution, whereas values below 0.5 focus on variables with high variance and values above 0.5 focus on variables with

high correlation with the response. All PPLS models in this study were computed with γ allowed to float in the interval [0.5,0.99]. During the computation, the PPLS algorithm chooses the γ value that gives the highest correlation between the PPLS scores and the response variable. The number of PLS factors was selected with a conservative Chi-square test based on the cross-validated Root Mean Square Error (RMSECV). For further details on the PPLS algorithm and selection of model complexity, see ref.[29].

2.3.3 Data classification

Approximately 7 % of the full dataset included moderate to severe crop lodging. Due to a very low signal-to-noise ratio of plots with lodging, it was necessary to exclude such plots from the data. Lodging is often encountered in practice, and the usefulness of the prediction models is strongly dependent on their ability to handle situations with lodging. In order to make the prediction models more operational, a pre-analysis step of "lodging-detection" was constructed in the form of a classification model. The spectral measurements were classified in the classes "standing crop" and "lodging" by PLS discriminant analysis (PLS-DA)[32], using field observation of lodging as the response variable. This response variable was a binary dummy variable comprising the two values 0 for "standing crop" and 1 for "lodging". A plot was considered as lodged if more than 20 % of the plants was not standing upright (as visually judged). The lodging model was cross validated and the number of PLS factors was selected with a conservative Chi-square test based on the RMSECV[29]. All predictions were then run using the lodging model as a pre-analyses step to remove plots with lodging from the data set. Prediction models

computed from the classified dataset had significantly higher predictive ability than those computed from the datasets containing lodging plots. This implies that the prediction models presented were not able to provide any predictions of yields or grain protein concentrations in areas with lodging.

2.3.4 Validation of prediction models for yields and protein concentration

A stepwise validation strategy was chosen for the robustness assessment of the prediction models. The nomenclature for naming the datasets is *year* directly followed by *site* identification letter (i.e. 2007A means data from site A in 2007). In validation strategy 1, completely independent validation data were used. Site C only had data from 2010 and dataset 2010C was thus selected as the severest validation set possible for models based on the other sites and years. The first model was calibrated on only one site-year (2007A), whereas the succeeding models included one more site-year in the calibration set until all 6 available calibration site-years were used (Table 5). In validation strategy 2, all site-years except 2007A were initially used for validation. Then a successive translocation of data from validation set to calibration set was performed. In validation strategy 3, we carried out a "leave-year-out" cross-validation. Data from each of the four years were successively used alone for validation, whilst data from the remaining years were used for calibration. In validation strategy 4, a "leave-site-out" cross-validation scheme was followed. Data from each of the three sites was successively used alone for validation, whereas data from the remaining sites were used for calibration. In all strategies, model performance was evaluated by four diagnostic measures between measured and predicted grain yield and protein concentration

values. The diagnostics were coefficient of determination (R^2), root mean square error of prediction (RMSEP), standard error of performance (SEP) and bias. In order to reveal differences in yield and protein level between sites and years, an Analysis of variance (ANOVA) of the response variables were performed in combination with Tukey's multiple comparison at a p-level of 5 %.

2.3.6 Vegetation indices

The PPLS models presented here use information from a large number of spectral bands. It was of interest to compare our method with a commonly used alternative, an approach based on vegetation indices. A vegetation index is an algebraic operation on a small number (often 2 or 3) carefully selected wavelengths. In the literature, several indices have been developed and many of these have been used for prediction of plant properties[5, 15, 24, 33]. Here, a subset of four indices was selected: Normalized Difference Vegetation Index (NDVI)[34], Modified Soil Adjusted Vegetation Index (MSAVI)[35], Red Edge Inflection Point (REIP)[36] and d-chl-ab[37] (Table 3). All of these have been used with some success for predicting properties of wheat, especially grain yield[15]. Each vegetation index was computed from the full spectral dataset (based on the equations given in Table 3), and successively used as predictors in univariate least squares regressions, with either grain yield or grain protein concentration as the response variable.

297 3. Results

298

299 There were significant ($p < 0.001$) yield differences between years, except
300 between 2008 and 2010. For protein, there was significant ($p < 0.001$) differences
301 between all years except 2008 and 2009. The yield levels differed significantly
302 between all sites. This was also the case for protein, except for in 2008 and 2009,
303 when there were no significant difference. The grain yield and protein
304 concentration levels also differed markedly within sites (Table 2). Lodging was
305 particularly pronounced in 2009 at site B, where moderate to severe lodging was
306 observed in 29 % of the plots. Initial model runs showed that lodging had a
307 markedly negative effect on the performance of the prediction models (data not
308 shown). The PLS-DA classification model was able to classify 98 % of the lodging
309 plots correctly (Table 4 and Fig. 2).

310

311 In validation strategy 1, using the 2010C data only for validation (i.e. the most
312 independent data), the grain yield and protein models showed in general quite
313 similar behavior, although the amount of variance explained by the protein models
314 was considerably lower than that explained by the grain yield models (Table 5). As
315 the number of calibration samples increased (from model 1 to 6), all models
316 became less biased and the number of significant PPLS factors increased. Using the
317 2007A data only for calibration, 83 % and 66 % of the variation in grain yield and
318 grain protein concentration was explained, respectively (Table 5, model 1). The
319 prediction performances decreased slightly when calibration data from 2008 and
320 2009 were included. When the 2010A data were added to the calibration set, the
321 highest prediction performances in strategy 1 were reached, as more than 89 % and

68 % of the variation in grain yield and grain protein concentration was explained, respectively (Table 5, model 6).

Strategy 2 utilized all available data in each model. The first model, which was based on the 2007A dataset for calibration, did not perform very well (Table 6, model 1). There was a tendency in strategy 2 that the yield model was improved with increasing amount of calibration data, whereas the explained variances in the protein models generally remained very low and did not exceed 50 % until the 2010A data (i.e. all remaining data) were included in the calibration set (Table 6, model 6). The number of significant PPLS factors increased along with the number of calibration samples. Compared with strategy 1, the model biases varied more in strategy 2.

Neither of the two first strategies took full advantage of the temporal and spatial distributions of the data. Therefore, a calibration approach with a "leave-year-out" cross-validation procedure was carried out in strategy 3 (i.e. site-years from one season at the time were used for validation and the remaining site-years were used for calibration). The grain yield models that were validated on the 2007 data had the best overall performance, accounting for more than 94 % of the variation (Table 7, model 1). Despite the high explained variance, the measured to predicted plot appeared to have a different slope than that of the other models (Fig. 3, upper left plot). The lowest bias occurred in the 2008 model (-14 g m^{-2}), whereas the 2009 and 2010 models both had greater, positive biases (Table 7 and Fig. 3). All protein models showed poor (2008, 2009 and 2010) to modest (2007) predictive performance (Table 7).

In strategy 4, a “leave-site-out” cross-validation procedure was performed (i.e. site-years from one site at the time were used for validation, and the remaining site-years for calibration). Performance of the grain yield prediction models were modest when validated on data from site B (Table 8, model 2), but were markedly increased when validated on the data from either site A (model 1) or site C (model 3). Using site C for validation also explained the highest amount of variance in protein, whereas the use of validation data from sites A or B gave very poor protein prediction (Table 8).

Common to all four strategies was that the protein models had a higher number of significant PPLS factors and thus a higher degree of complexity than the grain yield models. Only the grain yield models were examined further.

As an alternative to PPLS regression on the full spectra, univariate least squares regressions between some widely used vegetation indices and the grain yield data were carried out. All regressions showed generally low prediction performance (Fig. 4). The NDVI and REIP indices were especially poor predictors of grain yield, as they tended to saturate at high yields.

4. Discussion

In this study, data from seven site-years were combined in various validation strategies in order to assess the robustness and usefulness of prediction models for grain yield and grain protein concentration. The results showed in general that the predictions improved with increasing variance, both in time and space, of the calibration data. Grain yield was predicted best, with overall explained variance exceeding 80 % in most cases, reaching a maximum of 94 %. In contrast, none of the protein models obtained a degree of explanation above 70 %.

4.1 Effect of year

Model performance differed when the leave-year-out cross-validation procedure as carried out (Table 7), and it appeared to be beneficial to have calibration data from the same year as that in which the predictions were made for (Table 5, model 6). This finding was, however, not consistent, as the overall best model was calibrated from 2008-2010-data and validated on 2007-data (Table 7, model 1, grain yield). Since 2007A was the only field trial in 2007, it was not possible to check for a possible improvement of including other 2007-data in the calibrations.

One rather general explanation of the benefit of having data from the same year in the calibration as in the validation data is that there are seldom two growing seasons that have the same conditions for crop growth. Changing micro-meteorological conditions between years may, for example, cause diverging relations between spectral signatures of the canopy and plant properties of

391 interest. This effect could be direct (e.g. differing harvesting quotients) or indirect
392 (e.g. unfavourable conditions for grain filling and ripening occurring after the
393 spectral measurements). More specific reasons for the observed effect of year on
394 the results may be altered weather conditions between times of the radiometric
395 measurements, variations in weed density and lodging.
396
397 On several occasions during data collection, especially during the 2009 and 2010
398 seasons, the weather conditions were not optimal with respect to spectral
399 measurement (i.e. variable cloud cover, which causes spectral drift in the
400 measurements). The conditions at times of data acquisition can certainly explain
401 some of the observed year-effect. We followed, however, the same
402 spectroradiometer calibration regime as in ref.[11] in which acceptable results were
403 achieved even in conditions with quite severe cloud cover. Hence, we do not
404 believe that this was a major contributing factor in the current study. It should be
405 emphasized that high robustness towards changes in the measuring environment is
406 a prerequisite for systems designed for use under practical conditions when such
407 changes are more the rule than the exception.
408
409 All field trials received plant protection according to current practice, but weeds
410 were still present in some cases. This was particularly so in 2009, when the density
411 of couch grass (Elymus Repens L.) was relatively high, mainly due to a poor effect
412 of herbicide spraying the previous autumn (2008). The weed density was probably
413 not so high that it would have significantly influenced grain yield or grain protein
414 concentration, according to studies on the impact of weeds on wheat grain yield[38].
415 However, we assume that the spectral signatures may have been affected in some

cases. Other studies have showed that differences in spectral signature between cereals, shrubs, grasses and weeds were sufficient to separate between these groups[39]. Hence, spectral noise caused by the weeds present may explain the drop in model performance (grain yield) when 2009-data alone was used for validation (Table 7, model 3). Variation in weed density is normally encountered in real-life situations, and the results suggest that our approach may tackle such variation reasonably well. Nevertheless, greater weed densities than those observed in this study and/or the presence of other weed species, may potentially reduce the usefulness of the method.

The application of a pre-treatment classification model successfully removed most of the lodging plots from the data. Hence, lodging was probably not an important factor that could account for the difference in prediction performance between years. Lodging often occurs in practice. In an operational system, the classification model developed could be used as an initial step to separate between areas with and without lodging. It is very difficult to predict yields and yield quality in cases with lodging, partly because the plant canopy may be covered by other plant parts, which radically changes the spectral signature, and partly because it is questionable to what extent the grain will actually be picked up by the combine harvester (depending on e.g. severity of lodging, time and weather conditions until harvest, and cereal variety). Hence, areas detected as being lodged could be discarded or the prediction result flagged as “unreliable”, whereas the remaining measurements could enter the prediction model and be used for a prediction.

4.2 Effect of site

441 Prediction results differed markedly between the three sites A, B and C (Table 8).
442 The predictions performed for site B were especially poor. This may largely be
443 explained by differences in the soil properties between the sites. Whereas
444 sites A and C, which were localized relatively close, had rather similar soil
445 properties, site B diverged from the others by having the highest average topsoil
446 contents of clay, soil organic matter and all of the measured nutrients, except
447 K-HNO_3 (Table 1). Moreover, this site also had the largest within-field soil variation
448 of the three sites, thus representing the largest potential for site-specific variation
449 in crop growth.
450
451 The higher nutrient content of site B was reflected by the larger yield level of the
452 unfertilized plots, indicated by enhanced minimum yields (Table 2). The mean
453 yields and the yield variation within site did not, however, reflect the measured
454 soil properties equally well. One reason for this was that the frequency of plots
455 with lodging was higher at site B than at the other sites, resulting in reduced yields
456 particularly on plots with high yield potential. One result of this was a smaller
457 numerical range in the yield values of the datasets from site B than sites A and C
458 (Table 2), which subsequently may account for some of the difference in prediction
459 performance between the three sites (Table 8). Another factor which may explain
460 the poor predictions for site B was the larger weed pressure at this site (see
461 previous section).
462
463 4.3 Diverging performance between yield and grain protein concentration models
464

The grain yield models performed overall much better than the models predicting grain protein concentration (Table 7 and 8). Similar performance differences have been showed by other authors[7, 23].This is probably due to a more complex causal relationship between the spectral measurements and grain protein concentration than that between spectral measurements and grain yield[8]. The grain protein concentration models generally had a higher number of PPLS factors than the grain yield models, which supports our assumption. In a previous study[27], we found that the spectral variables related to grain protein content were not the same as those commonly used to predict grain protein concentration directly on wheat kernels. This indicates that there is another, and probably more complex link between grain protein content and spectral measurements in the field as compared to that found for individual wheat kernels in a laboratory environment, which basically rely on the characteristic chemical bounds for protein[40]

4.4 Biases of the grain yield models

Why does the general yield level apparently affect the model bias? We believe that this may at least in part be caused by a change in the fraction of the total above ground biomass allocated to the grain (i.e. harvest index, HI) between years. Since the amount of biomass has been shown to affect the spectral signature significantly[41], it is likely that a change in the HI would affect the yield predicting performance of the model. Harvest index, calculated for 25 plots at site A in 2008, 2009 and 2010, was 0.49, 0.43 and 0.44, respectively (data not shown). The overestimation (positive bias) of the yield models in 2009 and 2010 could thus be a result of relatively lower HI in those years compared to that in 2008.

490

491 4.5 Vegetation indices

492 The vegetation indices showed very poor performance compared to the PPLS
493 models based on the full spectrum (Figs. 3 and 4). Most vegetation indices are
494 based merely on 2-3 wavelengths and are often designed for extracting as much
495 information as possible from a limited number of wavelengths (for instance a
496 Landsat Thematic Mapper (TM) image with only 5 wavelengths). Several indices
497 have previously been used with some success for crop property prediction[15, 22, 24, 42].
498 However, a common problem with most indices is the occurrence of saturation at
499 high yield levels, making them less useful in such situations[9, 37, 43]. The saturation
500 effect is especially evident in Figure 4 (left plots). Literature is scarce on the topic
501 of crop property prediction using latent variable regression techniques on full
502 spectrum data, but good results have been reported using PLS regression on colour
503 image data to predict grain yield[8]. Another interesting approach is the use of multi-
504 way PLS (N-PLS) and repeated reflectance measurements to predict grain
505 properties[7]. In recent years, development of small and affordable full-spectrum
506 radiometers for field use has exploded. This development calls for more advanced
507 data analysis techniques, but nevertheless, as our study shows, the potential
508 benefit of combining full-spectrum data with multivariate analyses significantly
509 outweighs the limitations of traditional approaches.

510

511 4.6 Practical consequences and outlook

512 The grain yield models presented here had good performance in most cases and
513 could be used for several practical purposes. At the national agricultural authority
514 level, early yield predictions are of interest for market regulating purposes. At

515 farm level, robust yield maps are in demand as a foundation for planning the next
516 cropping season and for locating potential problem areas (i.e. low yielding zones).
517 Current yield mapping systems, based on various monitors mounted on the combine
518 harvester, rely heavily on challenging correction procedures. The yield prediction
519 models presented here had some degree of bias and skewness. We are currently
520 working on a method to correct these problems by the use of data fusion between
521 NIR reflectance spectra and aggregated climate variables. The results from this will
522 be published as part II of the current study.
523
524 Models for grain protein concentration did not have the same degree of accuracy as
525 the grain yield models. A high-performance model for grain protein concentration
526 could be used for protein-mapping at field level and targeted harvesting, i.e.
527 where zones high in protein concentration are harvested separately from zones
528 with poorer quality, with the class intervals adjusted in accordance with the
529 current quality payment regime. The models presented here did not, however, give
530 performance sufficient to realize targeted harvesting. Improving the protein
531 models is clearly one of the topics for further research work in this area.
532
533 The models were calibrated for spring wheat, variety 'Bjarne'. We assume that the
534 grain yield model would need re-calibration in order to be applicable for other crop
535 varieties and species.
536
537 In this study the spectral measurements were obtained by hand-held equipment. If
538 the model is to be applied under practical conditions at the field or farm level,
539 another platform for measurements would be a prerequisite. The airborne

approach, as described in an earlier study[27], appears to be promising, and we are currently exploring the use of Unmanned Aerial Vehicles (UAVs) as a flexible alternative.

5. Conclusions

As expected, single year or single site calibrations were not robust enough for the prediction of other years and/or sites. However, when a few or more years and sites were included, good predictions of grain yield were obtained. The results clearly indicate that the prediction performance and robustness both over year and site variations are of practical use for the farmers and/or agricultural authorities.

The utilization of wide, multiband, spectral signatures in combination with the multivariate regression method PPLS outperforms the more traditional approach using vegetation indices and MLR, in terms of predictive ability.

The rapid development of small and affordable full-spectrum radiometers for field use holds promise for the successful utilization of our results in future systems designed for practical precision agriculture.

559 Acknowledgement

560

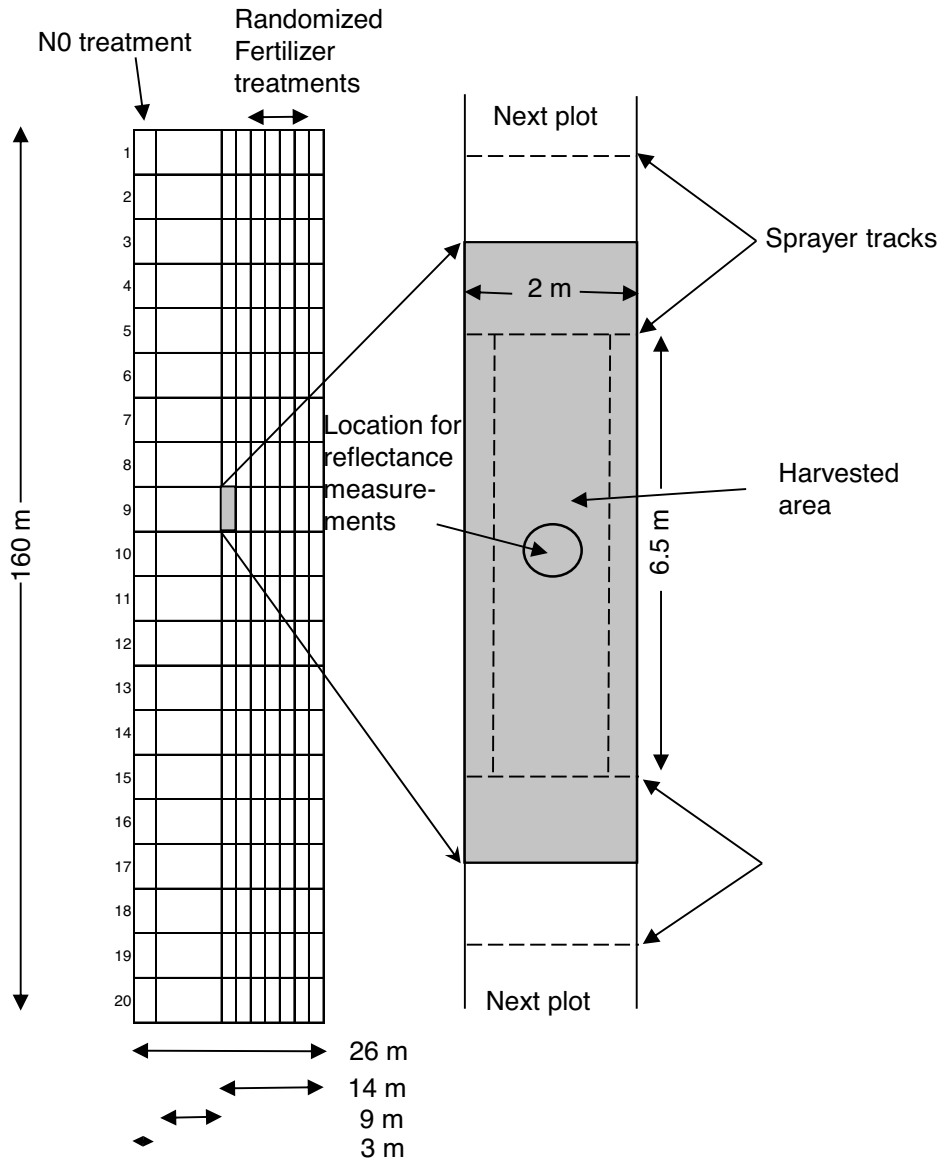
561 T. Gaardløs is gratefully acknowledged for his skilled technical assistance and
562 H. Riley for comments and corrections to a previous version of this manuscript. The
563 project was funded by the Norwegian Institute for Agricultural and Environmental
564 Research.

References

1. F. Alvaro, L. F. G. Del Moral, and C. Royo, "Usefulness of remote sensing for the assessment of growth traits in individual cereal plants grown in the field". *International Journal of Remote Sensing*. **28,** 2497 (2007)
2. O. Beeri, et al., "Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie". *Remote Sensing of Environment*. **110,** 216 (2007)
3. M. C. Belanger, et al., "Comparison of reflectance and fluorescence spectroscopy for the detection of nitrogen deficiency in potato plants". *Canadian Journal of Remote Sensing*. **33,** 69 (2007)
4. W. Feng, et al., "Monitoring leaf pigment status with hyperspectral remote sensing in wheat". *Australian Journal of Agricultural Research*. **59,** 748 (2008)
5. M. Moriondo, F. Maselli, and M. Bindi, "A simple model of regional wheat yield based on NDVI data". *European Journal of Agronomy*. **26,** 266 (2007)
6. N. R. Rao, et al., "Estimation of leaf total chlorophyll and nitrogen concentrations using hyperspectral satellite imagery". *Journal of Agricultural Science*. **146,** 65 (2008)
7. P. M. Hansen, J. R. Jørgensen, and A. Thomsen, "Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression". *Journal of Agricultural Science*. **139,** 307 (2002)
8. T. Jensen, et al., "Detecting the attributes of a wheat crop using digital imagery acquired from a low-altitude platform". *Computers and Electronics in Agriculture*. **59,** 66 (2007)
9. L. Y. Liu, et al., "Predicting winter wheat condition, grain yield and protein content using multi-temporal EnviSat-ASAR and Landsat TM satellite images". *International Journal of Remote Sensing*. **27,** 737 (2006)
10. M. Reyniers, E. Vrindts, and J. De Baerdemaeker, "Comparison of an aerial-based system and an on the ground continuous measuring device to predict yield of winter wheat". *European Journal of Agronomy*. **24,** 87 (2006)
11. J. Chang, S. A. Clay, and D. E. Clay, "Clouds influence precision and accuracy of ground-based spectroradiometers". *Communications in Soil Science and Plant Analysis*. **36,** 1799 (2005)
12. N. Aparicio, et al., "Spectral vegetation indices as nondestructive tools for determining durum wheat yield". *Agronomy Journal*. **92,** 83 (2000)
13. M. A. Babar, et al., "The potential of using spectral reflectance indices to estimate yield in wheat grown under reduced irrigation". *Euphytica*. **150,** 155 (2006)
14. P. C. Doraiswamy, et al., "Crop yield assessment from remote sensing". *Photogrammetric Engineering and Remote Sensing*. **69,** 665 (2003)

15. L. H. Xue, W. X. Cao, and L. Z. Yang, "Predicting grain yield and protein content in winter wheat at different N supply levels using canopy reflectance spectra". *Pedosphere*. **17,** 646 (2007)
16. Z. J. Wang, et al., "Prediction of grain protein content in winter wheat (Triticum aestivum L.) using plant pigment ratio (PPR)". *Field Crops Research*. **90,** 311 (2004)
17. C. J. Zhao, et al., "Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress". *International Journal of Applied Earth Observation and Geoinformation*. **7,** 1 (2005)
18. C. Yang, J. H. Everitt, and J. M. Bradford, "Yield estimation from hyperspectral imagery using spectral angle mapper (SAM)". *Transactions of the Asabe*. **51,** 729 (2008)
19. R. Houborg, H. Soegaard, and E. Boegh, "Combining vegetation index and model inversion methods for the extraction of key vegetation biophysical parameters using Terra and Aqua MODIS reflectance data". *Remote Sensing of Environment*. **106,** 39 (2007)
20. C. Atzberger, "Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models". *Remote Sensing of Environment*. **93,** 53 (2004)
21. J. Marti, et al., "Can wheat yield be assessed by early measurements of Normalized Difference Vegetation Index?" *Annals of Applied Biology*. **150,** 253 (2007)
22. B. Prasad, et al., "Potential use of spectral reflectance indices as a selection tool for grain yield in winter wheat under great plains conditions". *Crop Science*. **47,** 1426 (2007)
23. C. G. Pettersson, M. Soderstrom, and H. Eckersten, "Canopy reflectance, thermal stress, and apparent soil electrical conductivity as predictors of within-field variability in grain yield and grain protein of malting barley". *Precision Agriculture*. **7,** 343 (2006)
24. L. Wall, D. Larocque, and P. M. Leger, "The early explanatory power of NDVI in crop yield modelling". *International Journal of Remote Sensing*. **29,** 2211 (2008)
25. P. Basnyat, et al., "Effectiveness of using vegetation index to delineate zones of different soil and crop grain production characteristics". *Canadian Journal of Soil Science*. **85,** 319 (2005)
26. H. Égner, H. Riehm, and W. M. Domingo, "Untersuchung über die chemische Boden-Analyse als Grundlage für die Beurteilung des Nährstoffzustandes der Boden. (Using chemical soil analysis as a basis for assessing soil nutrient status)". *Kungl. Lantbrukshögskolans Annaler*. **26,** 199 (1960)
27. S. I. Overgaard, et al., "Comparisons of two hand-held, multispectral field radiometers and a hyperspectral airborne imager in terms of predicting spring wheat grain yield and quality by means of powered partial least squares regression". *Journal of near Infrared Spectroscopy*. **18,** 247 (2010)
28. J. C. Zadoks, T. T. Chang, and C. F. Konzak, "Decimal Code for Growth Stages of Cereals". *Weed Research*. **14,** 415 (1974)
29. U. Indahl, "A twist to partial least squares regression". *Journal of Chemometrics*. **19,** 32 (2005)

30. T. Isaksson and T. Naes, "The Effect of Multiplicative Scatter Correction (Msc) and Linearity Improvement in Nir Spectroscopy". *Applied Spectroscopy*. **42,** 1273 (1988)
31. A. Savitzky and M. J. E. Golay, "Smoothing + Differentiation of Data by Simplified Least Squares Procedures". *Analytical Chemistry*. **36,** 1627 (1964)
32. M. Barker and W. Rayens, "Partial least squares for discrimination". *Journal of Chemometrics*. **17,** 166 (2003)
33. M. Reyniers and E. Vrindts, "Measuring wheat nitrogen status from space and ground-based platform". *International Journal of Remote Sensing*. **27,** 549 (2006)
34. J. W. Rouse, et al., "Monitoring the Vernal Advancement of Retrogradation of Natural Vegetation". *Greenbelt: NASA/GSFC*. 1 (1974)
35. J. Qi, et al., "Amodified soil adjusted vegetation index". *Remote Sensing of the Environment*. **48,** 119 (1994)
36. G. Guyot, F. Baret, and D. J. Major, "High spectral resolution : determination of spectral shifts between the red and near infrared". *International Archives of Photogrammetry and Remote Sensing*. **11,** 750 (1988)
37. A. A. Gitelson and M. N. Merzlyak, "Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll". *Journal of Plant Physiology*. **148,** 494 (1996)
38. B. Melander, "Modeling the Effects of Elymus-Repens (L) Gould Competition on Yield of Cereals, Peas and Oilseed Rape". *Weed Research*. **34,** 99 (1994)
39. P. S. Thenkabail, et al., "Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications". *Remote Sensing of Environment*. **91,** 354 (2004)
40. S. R. Delwiche, "Single Wheat Kernel Analysis by near-Infrared Transmittance - Protein-Content". *Cereal Chemistry*. **72,** 11 (1995)
41. R. E. E. Jongschaap, "Sensitivity of a crop growth simulation model to variation in LAI and canopy nitrogen used for run-time calibration". *Ecological Modelling*. **200,** 89 (2007)
42. J. R. Jorgensen and R. N. Jorgensen, "Uniformity of wheat yield and quality using sensor assisted application of nitrogen". *Precision Agriculture*. **8,** 63 (2007)
43. J. Casadesus, et al., "Using vegetation indices derived from conventional digital cameras as selection criteria for wheat breeding in water-limited environments". *Annals of Applied Biology*. **150,** 227 (2007)



699

700 Figure 1. Experiment layout and sampling locations. Overall experiment layout to
701 the left and one example plot enlarged to the right. Measurement area is indicated
702 with a circle and harvested area is indicated with dashed lines on the right-hand
703 plot.

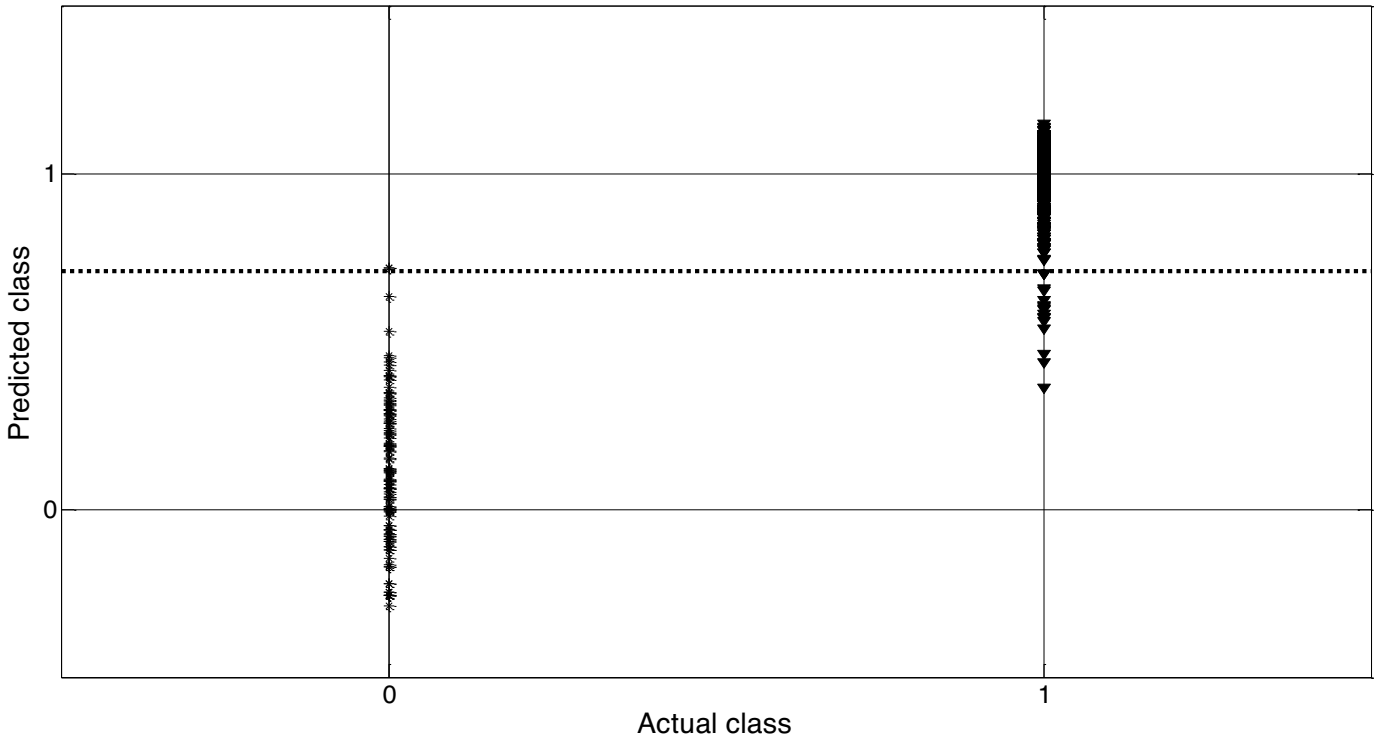


Figure 2. Score plot from a PLS-DA classification of all spectral measurements. The 976 data points were divided between the classes lodging (asterisks) and non-lodging (triangles). The dashed line denotes the decision boundary line.

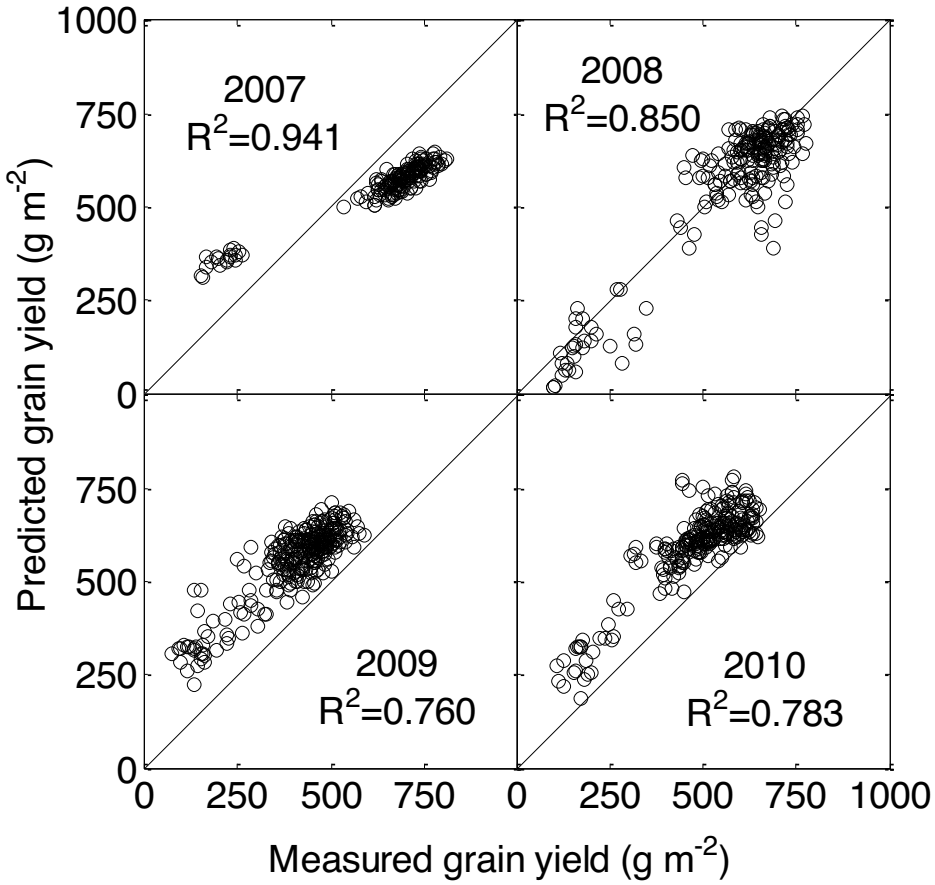
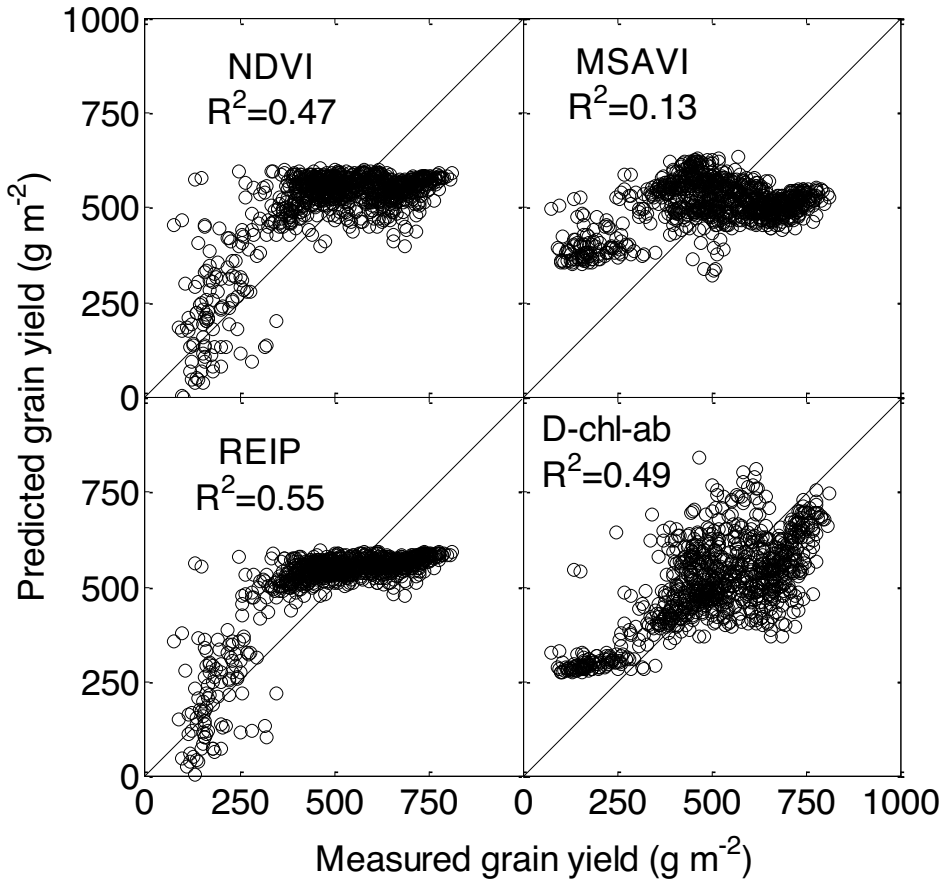


Figure 3. Measured grain yield (X-axis) plotted against predicted grain yield from the validation set (Y-axis). Prediction models were calibrated with Powered Partial Least Squares (PPLS) on spectral measurements from 3 seasons and validated on measurements from one season (denoted inside each subplot along with the coefficient of determination, R^2). The target lines have slope 1 and zero intercept.



715

716 Figure 4. Measured grain yield (X-axis) plotted against predicted grain yield
717 (Y-axis). Predictions were based on univariate least squares regressions
718 between yield data and vegetation indices derived from all spectral
719 measurements. All target lines have have slope 1 and zero intercept. The
720 indices used were Normalized Difference Vegetation Index (NDVI), Modified
721 Soil Adjusted Vegetation Index (MSAVI), Red Edge Inflection Point (REIP) and
722 D-chl-ab.

723 Table 1. Measured soil properties at all three experiment sites

Soil properties[1]	Depth	SITE A		SITE B		SITE C	
	cm	mean	CV[2]	mean	CV	mean	CV
P-AL	0-25	7.1	20	7.4	30	6.1	25
K-AL	0-25	7.8	14	11	26	6.5	215
K-HNO3	0-25	41	76	35	15	44	13
Mg-AL	0-25	6.8	19	20	25	10	17
Ca-AL	0-25	121	27	761	27	228	15
Na-AL	0-25	1.3	14	1.8	23	-[3]	-
Total N	0-25	0.14	19	1.0	43	0.20	21
Total C	0-25	1.8	17	15	44	-	-
Ignition loss	0-25	5.1	29	28	43	4.5	31
Ignition loss	25-60	5.1	35	27	104	-	-
pH	0-25	5.7	3.9	6.4	3.8	6.4	3.1
Gravel > 2 mm	0-25	72	38	5.3	28	8.6	38
Gravel > 2 mm	25-60	81	38	6.2	63	-	-
Sand 0.06 - 2 mm	0-25	58	4.5	45	12	51	10
Sand 0.06 – 2 mm	25-60	56	8.9	43	34	-	-
Silt 0.002 – 0.06 mm	0-25	30	5.0	36	9.4	34	8.2
Silt 0.002 – 0.06 mm	25-60	30	6.0	42	31	-	-
Clay < 0.002 mm	0-25	12	18	20	19	15	25
Clay < 0.002 mm	25-60	14	32	15	33	-	-

724 [1]Units: Gravel given in g kg^{-1} bulk soil, Total N, C, ignition-loss and soil texture fractions in g 100 g^{-1}

725 fine earth (< 2 mm), other nutrients in mg 100 g^{-1} fine earth

726 [2]Coefficient of Variation, (%).

727 [3]Not measured

728

729 Table 2. Descriptive statistics of the crop response variables for each year and site.
730 The p-values were computed from a nested analysis of variance of the yield and
731 protein data. Standard errors of reference method were 5.0 g m^{-2} for grain yield
732 and 0.15 g DM 100 g^{-1} for grain protein concentration

Variable	Dataset (number of plots)	min.	mean	max.	stdv.
Grain yield	2007A (160)	151	638	812	172
	2008A (120)	99	531	771	185
	2008B (120)	157	588	778	155
	2009A (160)	74	402	590	121
	2009B (132)	91	407	570	97
	2010A (140)	113	513	681	146
	2010C (144)	110	444	629	120
p-values	Site: < 0.001, Year: < 0.001				
Grain Protein	2007A (160)	10.0	12.8	16.7	1.5
	2008A (120)	10.7	14.3	17.0	1.6
	2008B (120)	9.9	14.7	19.9	1.9
	2009A (160)	9.5	14.4	17.2	1.8
	2009B (132)	9.9	14.0	17.6	1.6
	2010A (140)	9.4	12.1	15.3	1.5
	2010C (144)	9.4	12.9	15.9	1.8
p-values	Site: < 0.001, Year: < 0.001				

733

734 Table 3. Vegetation indices used to predict grain yield and grain protein
735 concentration. The letter R indicates reflectance value, scaled between 0 and 1,
736 whilst the subscripts indicate the wavelength at which the reflectance values are
737 taken

Name	Equation	Reference
NDVI	$\frac{R_{810} - R_{690}}{R_{810} + R_{690}}$	[34]
MSAVI	$\frac{1}{2}\left[2(R_{810} + 1) - \sqrt{(2\,R_{810} + 1)^2 - 8(R_{810} - R_{690})}\right]$	[35]
REIP	$700 + 40\frac{\frac{R_{810} - R_{690}}{2} - R_{740}}{R_{760} - R_{740}}$	[36]
d-chl-ab	$\frac{\frac{R_{760} - R_{740}}{2}}{R_{560}}$	[37]

738

739 Table 4. Confusion matrix for the PLS-DA classification model for all spectral
740 measurements. Entries are the counts of correctly and incorrectly classified data
741 points according to the actual lodging recordings. Total accuracy was 98.3 %

		Predicted class		Total
		Standing crop	Lodging	
Actual class	Standing crop	891	34	892
	Lodging	1	68	69
	Total	892	102	976

742

743 Table 5. Performance of PPLS prediction models for grain yield and grain protein
744 concentration using spectral measurements from an increasing number of site years
745 (sites A and B) as calibration data and spectral measurements in 2010 from the
746 independent site C as validation data (validation strategy 1, see M&M section)

Variable	Dataset/Diagnostic	1	2	3	4	5	6
	Calibration set	2007A	2007A 2008A	2007A 2008A 2008B	2007A 2008A 2008B 2009A	2007A 2008A 2008B 2009A 2009B	2007A 2008A 2008B 2009A 2009B 2010A
	Validation set	2010C	2010C	2010C	2010C	2010C	2010C
Grain yield	R^2	0.834	0.808	0.834	0.825	0.854	0.891
	RMSEP	133	189	112	123	117	108
	Bias	97	130	64	86	72	71
	SEP	91	137	91	88	93	82
	Number of factors	2	2	3	4	4	4
Grain rotein	R^2	0.662	0.627	0.612	0.610	0.655	0.684
	RMSEP	1.7	2.2	2.1	2.1	2.0	1.6
	Bias	0.1	0.5	0.8	1.0	1.0	0.7
	SEP	1.7	2.1	1.9	1.8	1.7	1.4
	Number of factors	3	4	4	5	5	5

747

748 Table 6. Performance of PPLS prediction models for grain yield and grain protein
749 concentration using spectral measurements from an increasing number of site years
750 (sites A and B) as calibration data and a decreasing number of site years
751 (sites A, B and C) as validation data (validation strategy 2, see M&M section)

Variable	Dataset/Diagnostic	1	2	3	4	5	6
	Calibration set	2007A	2007A 2008A	2007A 2008A 2008B	2007A 2008A 2008B 2009A	2007A 2008A 2008B 2009A 2009B	2007A 2008A 2008B 2009A 2009B 2010A
	Validation set	2008A 2008B 2009A 2009B 2010A 2010C	2008B 2009A 2009B 2010A 2010C	2009A 2009B 2010A 2010C	2009B 2010A 2010C	2010A 2010C	2010C
Grain yield	R^2	0.576	0.513	0.647	0.601	0.783	0.891
	RMSEP	195	161	169	126	176	108
	Bias	160	115	132	102	82	71
	SEP	112	113	107	75	156	82
	Number of factors	2	2	4	5	5	4
Grain protein	R^2	0.167	0.198	0.194	0.243	0.264	0.684
	RMSEP	2.1	4.0	2.9	2.2	1.8	1.6
	Bias	-0.5	2.7	1.0	-1.0	0.6	0.7
	SEP	2.1	3.0	2.7	2.0	1.7	1.4
	Number of factors	4	3	4	4	6	5

752

753

754 Table 7. Performance of PPLS prediction models for grain yield and grain protein
755 concentration using spectral measurements from all available sites within each of the
756 years 2007-2010 separately as validation data and all remaining site-years as calibration
757 data in each case (validation strategy 3, see M&M section).

Variable	Dataset/Diagnostic	1	2	3	4
	Calibration	2008A 2008B 2009A 2009B 2010A 2010C	2007A 2009A 2009B 2010A 2010C	2007A 2008A 2008B 2010A 2010C	2007A 2008A 2008B 2009A 2009B
	Validation	2007A	2008A 2008B	2009A 2009B	2010A 2010C
Grain yield	R^2	0.941	0.850	0.760	0.783
	RMSEP	156	76	218	176
	Bias	-89	-14	148	82
	SEP	128	74	160	156
	Number of factors	4	4	7	5
Grain Protein	R^2	0.669	0.158	0.268	0.264
	RMSEP	1.5	3.5	3.0	1.8
	Bias	-0.9	1.2	1.7	0.6
	SEP	1.2	3.2	2.5	1.7
	Number of factors	4	4	5	6

758

759 Table 8. Performance of PPLS prediction models for grain yield and grain protein
760 concentration using spectral measurements from each site separately as validation
761 data and data from the other sites as calibration data (validation strategy 4, see
762 M&M section).

Variable	Dataset/Diagnostic	1	2	3
	Calibration	2008B 2009B 2010C	2007A 2008A 2009A 2010A 2010C	2007A 2008A 2008B 2009A 2009B 2010A
	Validation	2007A 2008A 2009A 2010A	2008B 2009B	2010C
Grain yield	R^2	0.802	0.526	0.891
	RMSEP	153	207	108
	Bias	102	156	71
	SEP	114	136	82
	Number of factors	2	4	4
Protein	R^2	0.183	0.297	0.684
	RMSEP	1.8	1.7	1.6
	Bias	-1.0	0.5	0.7
	SEP	1.5	1.6	1.4
	Number of factors	4	6	5

763

PAPER IV

Prediction of wheat yield and protein using remote sensors on grain fields - Part II: Improving prediction ability using data fusion

Øvergaard, S.I[1,3,*], Isaksson, T.[2], Korsaeth, A.[1]

[1]The Norwegian Institute for Agricultural and Environmental Research, Arable Crops Division, NO-2849 Kapp, Norway

[2]The Norwegian University of Life Sciences, Department of Chemistry, Biotechnology and Food Science, 1430 Ås, Norway

[3]The Norwegian University of Life Sciences, Department of Mathematical Sciences and Technology, 1430 Ås, Norway

Keywords

Bias; Canonical Powered Partial Least Squares; FieldSpec; Hyperspectral; Near infrared spectroscopy; Precision agriculture; Spectral reflectance; Spring wheat

Abstract

Fusion of NIR spectra and weather data by means of Canonical Powered Partial Least Squares (CPPLS) was used to correct yield prediction models of spring wheat grain yields. The data comprised seven unique site-years and 976 individual plots. In part I of the study, in which models based on NIR spectra alone were used, we achieved a high degree of explained yield variance (up to 94 %), but encountered large problems with bias and skewness in the computed regression models. We hypothesized that weather variation between seasons as the governing process behind this behaviour. In the present part of the study (part II), the bias and skewness problems were significantly reduced by the inclusion of aggregated weather variables as additional predictors. Average bias and skewness of the final predictions amounted to X and Y, respectively, which corresponds to reductions of 38 % and 23 %. The weather data used are available about 4 weeks prior to harvest, which implies that the approach would allow early yield prediction. In conclusion, this study shows the potential for improving NIR spectra based prediction models by including season-specific information, such as weather data, which is becoming more readily available through improvements in meteorological services.

1. Introduction

Along with rising demands on agricultural production, both in terms of high yields as well as reduced environmental impact, the field of precision agriculture is becoming increasinglys important. Precision agriculture encompasses the use of available technology to tailor the crop treatment to the current crop situation at any site in a field (e.g. site-specific application of fertilizer based on real-time processing of reflectance measurements[1]). Similar technology may also be used for yield predictions. Accurate yield maps are useful for identifying potential agronomic problem areas[2], which may be seen as low yielding zones, or for quantifying site-specific nutrient off-take, information which may be used for a balanced nutrient replacement in the following growing season[3]. Moreover, yield prognoses are high demanded at the agricultural authority level for market regulation purposes[4].

Various methods for yield prediction have been reported over the last decades, but the use of remotely sensed reflectance data seems to be the dominant technique. Several studies have been carried out with relatively simple analytical approaches, based on vegetation indices and linear regression (e.g. references[5-7]), but more advanced techniques, such as Partial Least Squares (PLS) and Multi-way Partial Least Squares (N-PLS), have also been reported (e.g. references[8, 9]).

When evaluating a model, a thorough validation procedure is essential in order to assess its robustness to changes in the environment, relative to that in which the model was calibrated. Nevertheless, a common approach used in model validation has been to calibrate the model with one year's data from one site, and to validate it using data from a nearby experimental site in the same year, e.g. references[9-11]. Numerous studies have reported large variations in crop growth between years (e.g.[12]), and such seasonal variations have been shown to have a markedly negative effect on the predictive ability of reflectance-based yield prediction models[13]. Literature is scarce on studies that have attended to correct for seasonal variations in such models.

In part I of the current study[13], we developed yield prediction models based on remotely sensed reflectance data covering four seasons and three sites (seven site-years). We found quite severe bias and skewness in our prediction models, and we speculated whether variations in weather conditions between the measurement seasons (2007 to 2010) could explain these observations. To test our hypothesis, we decided to conduct a second study (part II, reported here) of the study, using available weather variables as additional predictors.

Fusion of diverging data types, such as NIR spectra and weather data, can be achieved in a number of ways, for example by means of Hierarchical Principal Component Analysis (HPCA), Consensus PCA (CPCA), Hierarchical Partial Least Squares (HPLS) or Multiblock Partial Least Squares (MBPLS)[14]. In this study, we chose to use a recent extension of the Powered Partial Least Squares (PPLS)[15],

called the Canonical Powered Partial Least Squares (CPPLS)[16], since we had already successfully used PPLS in Part I of this study[13] and in two previous, related studies[17, 18].

The main objective of this study was to test whether data fusion of season-specific weather variables and remotely sensed reflectance data may reduce bias and skewness in yield prediction models, and thus improve this robustness of such models.

2. Materials and methods

In this study we used the same yield and NIR data as obtained in Part I[13]. A brief description follows of the experimental setup, instruments and measurements.

2.1 Experimental sites

Experiments with spring wheat (*Triticum aestivum* L., var. 'Bjarne') were performed at three different sites during the period 2007-2010 (altogether seven site-years). The sites were located at the Bioforsk Arable Crops Division at Apelvoll farm in SE Norway (site A and C) and at a nearby farm (site B).

Sites A and B had 20 replicate blocks, whereas site C had 18 replicate blocks, each consisting of 8 plots with size 2 x 8 m. Six nitrogen (N) level treatments were applied within each block: 0, 100, 125, 150, 175 and 200 kg nitrogen ha^{-1}. The border plots on either side of the randomized plots received 100 kg N ha^{-1}, thus giving three N100 plots per replicate block. Including the border plots, Site A and B comprised 160 plots each and site C 144 plots.

2.2 Instrumentation

All spectral reflectance measurements were taken with a FieldSpec3 spectroradiometer (version 3, Analytical Spectral Devices Inc., USA,

109 www.asdi.com). The instrument was operated on a pole 1 m above the crop canopy
110 when measuring (reflectance mode).

111

112 2.3 Measurements

113

114 One reflectance spectrum was collected from each plot 5 times during each season
115 at Zadoks (Z) growth stages[19] Z31, Z59, Z65, Z87 and Z90. As in part I of the study,
116 only reflectance data from Z65 were used for further analysis because earlier
117 studies have showed this to be the most favourable time to measure reflectance
118 for yield prediction purposes[17].

119 At maturity, the plots were harvested with a combine harvester, and yields were
120 determined gravimetrically.To find suitable weather variables, we utilized an
121 existing system for yield prognoses. In Norway, the cereal yield level (crop specific)
122 in each municipality is estimated each year per August 1 as part of the national
123 cereal yield prognoses[4]. The system is based on meteorological observations from a
124 network of weather stations, including the station at Apelsvoll (close to all sites
125 involved in this study). Hence, daily values of global sun radiation, precipitation,
126 air temperature (2 m), wind speed and relative humidity were taken from the
127 Apelsvoll weather station. Potential evapotranspiration was calculated in
128 accordance with Riley and Berentsen[20]. The daily weather data was then
129 aggregated within four phenological phases: 1) sowing and seed emergence (Z00 to
130 Z09), 2) leaf development and tillering (Z10 to Z29), 3) stem elongation and
131 inflorescence emergence (Z30 to Z49), and 4) anthesis and ripening to hard dough

(Z50 to Z87). An overview of these 24 variables (six weather variables times four phenological phases) is given in Table 1.

2.4 Data preprocessing

Spectral data were subjected to the same preprocessing as in Part I[13] of this study. In brief, raw data were corrected for splice point discontinuity with ViewSpec Pro (version 3.9, Analytical Spectral Devices Inc., USA, www.asdi.com). Multiplicative Scattering Correction (MSC)[21] as well as computation of first derivatives with the Savitzky-Golay algorithm[22] (differentiation order: 1, polynomial order: 2, window width: 15) were performed with MATLAB (version 2007b, MathWorks Inc., USA, www.mathworks.com). Only the part of the data that was considered free of lodging was included in the current study.

Some of the weather variables had partly quite severe multicollinearity, which led to numerical problems in the subsequent analyses. To eliminate this problem, we performed a Principal Component Analysis (PCA) of the weather data (Fig. 1). The resulting score values of the first two components, representing 83 % of the overall variation in the 24 initial variables, were used in the subsequent analyses. Extraction of more than two PCA components from the weather data did not further improve the cross-validated Root Mean Square Error (RMSECV) of the PCA model. Computation of the PCA scores was performed with the R function

"prcomp"[23] with mean centring and normalization by division with the standard deviation of each variable.

2.5 Data analysis

All statistical models and graphics were performed with R[23] and the pls-package[24]. Recently, two important extensions of the PLS methodology have been proposed. The first extension is Powered Partial Least Squares (PPLS)[15], which enables the PLS to focus on predictors with high variance or predictors with high correlation with the response through the choice of a method-specific parameter γ. The second extension is Canonical Partial Least Squares (CPLS)[16], which lets the PLS work with additional response variables and uses optimization of canonical correlation to reduce systematic effects in the main model response. This represents an interesting new way of achieving data fusion and the approach seems suitable for our application. The authors of CPLS and PPLS encourage the merging of the two methodologies and since we previously have shown that PPLS worked well on our data[13, 17], the use of Canonical Powered Partial Least Squares (CPPLS) was considered a natural choice in our data fusion task. Hence, all models presented in this study were computed with CPPLS with the γ parameter optimized over the closed interval [0.5,0.95] as in our previous yield prediction studies[13, 17]. Since the objective of the study was to quantify and reduce the impact of seasonal variation on the yield predictions, we selected a leave-year-out validation strategy. This procedure is similar to validation strategy 3 used in Part I of the current study. Models were assessed by means of the diagnostic measures the Coefficient of

178 Determination (R^2), Root Mean Error of Prediction (RMSEP), bias, skewness (third
179 moment about the mean of the residuals) and Standard Error of Performance(SEP).
180 A test of the prediction ability between models from NIR only and the CPPLS
181 models were performed as a paired Student's t-test on the absolute values of the
182 model residuals, following the ideas of Cederkvist et al.[25].

183

184 In Part I, we concluded that all the protein prediction models were unusable for
185 predictions due to very low explained variance. In the current study (Part II), we
186 aim at correcting seasonal, systematic variations of the predictions in Part I, and
187 not at improving the explained variance. Hence, the protein models of Part I were
188 not included here.

189

3. Results

Measured grain yields were in the range of 73.5 g m^{-2} to 811.8 g m^{-2}.

Grain yield predicted with the weather-corrected CPPLS models differed significantly from the grain yield predictions of the uncorrected, initial, PPLS models, except when validated on the 2008 data, which is seen from the p-values of the paired t-test on the absolute values of the residuals (Table 2).

All the selected model diagnostic measures (RMSEP, SEP, bias and skewness) were decreased by the corrections, except for the bias in 2008, which was initially very low (Table 2). The average reduction in bias was 38 %, whereas the reduction in skewness was 23 %. The reductions in bias and skewness were reflected in the regression lines for PPLS and CPPLS models (Fig. 2 and Fig. 3). Skewness was, however, still considerable for the 2007 predictions (Table 2, Fig. 2, upper left plot).

Averaged over years, the RMSEP of the CPPLS models was reduced from 157 to 85 kg ha^{-1}.

The amount of explained variance was overall high, ranging from 76 % to 94 % (Table 2).

4. Discussion

All the PPLS prediction models for grain yield used in Part I of this study[13] were improved by using season–specific weather data for correcting the NIR based models, as this reduced both RMSEP, SEP, bias and skewness. Least progress (none for bias) was observed when using the 2008 data for validation, most likely due to the high performance of the initial model, which gave little room for improvements.

In spite of a distinct improvement in all other model diagnostics, skewness was still considerable in the 2007-validation, even with the CPPLS model (Fig. 3). We speculate that this could be explained, at least partly, by the size of the validation dataset. In 2007, only about half the number of validation samples was available compared with that of the other years, since only one experimental site was present in 2007.

Overall, our the prediction results were better than those found in previous studies, e.g.[26, 27]. The present study demonstrates a feasible approach to meet the challenges often encountered under practical conditions, where models are applied for situations not reflected in their calibrations, such as future cropping seasons.

233 Considering the need for models which cope with such real life situations, there
234 are surprisingly many studies which do not explain how their models tackle
235 variations between years. In the literature on wheat grain yield prediction, a
236 common validation approach is to run two nearby field experiments, using one for
237 calibration and the other for validation of the prediction model, e.g.[5, 8, 28-31]. Such
238 a validation regime does not take into account variations between years, thus
239 leaving the model's prediction ability of a future cropping season unknown.

240

241 A study which did account for year to year variation was that of Lukina et al.[32],
242 who estimated winter wheat yield from spectral reflectance measurements during
243 three years (1998, 1999 and 2000). As in our study, there were differences in
244 measured yield pattern between years. When they calibrated a model to each year
245 separately, there were some quite large shifts in model variables and predicting
246 accuracy, pointing towards a need for some season-specific correction. In order to
247 perform such a correction, Lukina et al.[32] included cumulative growing degree days
248 (GDD) from planting until radiometric measurement as an extra variable. This did
249 not, however, lead to any significant improvement in the wheat yield predictions.

250

251 Another study which tried to use GDD for model correction was that of Raun et
252 al.[33]. They constructed a model for predicting winter wheat yields by combining
253 radiometric measurements, performed at two different growth stages, with GDD
254 cumulating between the two dates of measurement. When using the model with
255 data from six locations across a 2-yr period, they reported a high degree of
256 explained yield variance (83 %), but RMSE appeared to be large (as visually judged

from their plots). The real, predictive model performance was, however, difficult to evaluate, as only model calibrations were presented by Raun et al.[33].

How well a model fits a calibration data set does not reveal any information on how good the model performance will be for other data, i.e. for real predictions. In our study, we validated the model performances for one year strictly on data which were collected during other years. Such a validation belongs, however, to the exceptions, as many studies have been published without any validation procedure (e.g. Prasad et al.[26]) or with a cross-validation only (e.g. Pettersson and Eckersten[34]). As a result, it is not possible to evaluate the practical usefulness of such models, even when they are calibrated on data covering several years.

Literature covering yield predictions based on radiometric measurements appears to be dominated by studies relying on vegetation indices (i.e. a simple algebraic operation on one or two wavelengths), e.g. [26, 27, 32-36]. Historically, this approach was developed basically as a result of the very limited number of spectral bands available in the early days of remote sensing. An underlying assumption is that a vegetation index is more robust to spectral variation in time and space than are the raw reflectance data. In the first part of this study[13], we showed that our PPLS models clearly outperformed models based on some common vegetation indices, even before we introduced the corrections presented here. Considering the present availability of instruments which cover a large number of spectral bands, we question the persisting use of such index-based approaches.

Weather data used in this study are available at stage Z87, after the measurements time of the NIR spectra, but early enough to make yield predictions well ahead of harvest. The availability of interpolated, and thus site-specific, weather data is increasing through the appearance of several internet-based services for the distribution of meteorological data. These new services contribute greatly towards the practicality of using aggregated weather variables along with spectral measurements.

5. Conclusion

The combination of NIR reflectance data and aggregated weather data may substantially reduce the bias and skewness problems in yield prediction models. Reduction of these problems is one step further in the direction of being able to compute reliable prediction models of grain yield based on remotely sensed data, thus bringing the technique closer to practical use.

295 Acknowledgement

296 T. Gaardløs and the other technical staff at Apelsvoll are gratefully acknowledged
297 for their skilled technical assistance. Thanks to Hugh Riley for the proof reading of
298 this paper. The project was founded by the Norwegian Institute for Agricultural and
299 Environmental Research.

References

1. E. Zillmann, et al., "Assessment of cereal nitrogen requirements derived by optical on-the-go sensors on heterogeneous soils". *Agronomy Journal*. **98,** 682 (2006)
2. D. W. Lamb, "The use of qualitative airborne multispectral imaging for managing agricultural crops - a case study in south-eastern Australia". *Australian Journal of Experimental Agriculture*. **40,** 725 (2000)
3. J. V. Stafford, R. M. Lark, and H. C. Bolam, *Using yield maps to regionalize fields into potential management units*, in *Proceedings of the Fourth International Conference on Precision Agriculture, Pts a and B*. 1999. p. 225.
4. A. Korsaeth and T. Rafoss, "Tidlige prognoser for kornavlingene ved bruk av værdata - Sluttrapport." (in Norwegian) *Bioforsk RAPPORT*. **17,** 1 (2009)
5. C. Yang, J. H. Everitt, and J. M. Bradford, "Yield estimation from hyperspectral imagery using spectral angle mapper (SAM)". *Transactions of the Asabe*. **51,** 729 (2008)
6. C. G. Yang, J. H. Everitt, and J. M. Bradford, "Airborne hyperspectral imagery and linear spectral unmixing for mapping variation in crop yield". *Precision Agriculture*. **8,** 279 (2007)
7. P. Basnyat, et al., "Optimal time for remote sensing to relate to crop grain yield on the Canadian prairies". *Canadian Journal of Plant Science*. **84,** 97 (2004)
8. P. M. Hansen, J. R. Jørgensen, and A. Thomsen, "Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression". *Journal of Agricultural Science*. **139,** 307 (2002)
9. T. Jensen, et al., "Detecting the attributes of a wheat crop using digital imagery acquired from a low-altitude platform". *Computers and Electronics in Agriculture*. **59,** 66 (2007)

10. C. Yang, J. H. Everitt, and J. M. Bradford, "Airborne hyperspectral imagery and yield monitor data for estimating grain sorghum yield variability". *Transactions of the Asae*. **47,** 915 (2004)
11. A. C. Xavier, et al., "Hyperspectral field reflectance measurements to estimate wheat grain yield and plant height". *Scientia Agricola*. **63,** 130 (2006)
12. A. Korsaeth and H. Riley, "Estimation of economic and environmental potentials of variable rate versus uniform N fertilizer application to spring barley on morainic soils in SE Norway". *Precision Agriculture*. **7,** 265 (2006)
13. S. I. Overgaard, A. Korsaeth, and T. Isaksson, "Prediction of wheat yield and protein using remote sensors on grain fields - Part I: Assessing NIR model robustness for year and site variations. Manuscript submitted for publication." *Journal of near Infrared Spectroscopy*. (2012)
14. J. A. Westerhuis, T. Kourti, and J. F. Macgregor, "Analysis of multiblock and hierarchical PCA and PLS models". *Journal of Chemometrics*. **12,** 301 (1998)
15. U. Indahl, "A twist to partial least squares regression". *Journal of Chemometrics*. **19,** 32 (2005)
16. U. G. Indahl, K. H. Liland, and T. Naes, "Canonical partial least squares-a unified PLS approach to classification and regression problems". *Journal of Chemometrics*. **23,** 495 (2009)
17. S. I. Overgaard, et al., "Comparisons of two hand-held, multispectral field radiometers and a hyperspectral airborne imager in terms of predicting spring wheat grain yield and quality by means of powered partial least squares regression". *Journal of near Infrared Spectroscopy*. **18,** 247 (2010)
18. S. I. Overgaard, et al., "Prediction error improvements using variable selection on small calibration sets - a comparison of some recent methods. Manuscript submitted for publication." *Journal of near Infrared Spectroscopy*. (2012)
19. J. C. Zadoks, T. T. Chang, and C. F. Konzak, "Decimal Code for Growth Stages of Cereals". *Weed Research*. **14,** 415 (1974)
20. H. Riley and E. Berentsen, "Estimation of water use for irrigation in Norwegian agriculture." *Bioforsk RAPPORT*. **4,** 1 (2009)
21. T. Isaksson and T. Naes, "The Effect of Multiplicative Scatter Correction (Msc) and Linearity Improvement in Nir Spectroscopy". *Applied Spectroscopy*. **42,** 1273 (1988)
22. A. Savitzky and M. J. E. Golay, "Smoothing + Differentiation of Data by Simplified Least Squares Procedures". *Analytical Chemistry*. **36,** 1627 (1964)
23. R Development Core Team, "*R: A language and environment for statistical computing*", in Place: Published.
24. B.-H. Mevik, R. Wehrens, and K. Liland, "*pls: Partial Least Squares and Principal Component regression*", in Place: Published.
25. H. R. Cederkvist, A. H. Aastveit, and T. Naes, "A comparison of methods for testing differences in predictive ability". *Journal of Chemometrics*. **19,** 500 (2005)
26. B. Prasad, et al., "Potential use of spectral reflectance indices as a selection tool for grain yield in winter wheat under great plains conditions". *Crop Science*. **47,** 1426 (2007)
27. M. A. Babar, et al., "The potential of using spectral reflectance indices to estimate yield in wheat grown under reduced irrigation". *Euphytica*. **150,** 155 (2006)

28. M. Reyniers, E. Vrindts, and J. De Baerdemaeker, "Comparison of an aerial-based system and an on the ground continuous measuring device to predict yield of winter wheat". *European Journal of Agronomy*. **24,** 87 (2006)
29. L. H. Xue, W. X. Cao, and L. Z. Yang, "Predicting grain yield and protein content in winter wheat at different N supply levels using canopy reflectance spectra". *Pedosphere*. **17,** 646 (2007)
30. Z. J. Wang, et al., "Prediction of grain protein content in winter wheat (Triticum aestivum L.) using plant pigment ratio (PPR)". *Field Crops Research*. **90,** 311 (2004)
31. C. J. Zhao, et al., "Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress". *International Journal of Applied Earth Observation and Geoinformation*. **7,** 1 (2005)
32. E. V. Lukina, et al., "Nitrogen fertilization optimization algorithm based on in-season estimates of yield and plant nitrogen uptake". *Journal of Plant Nutrition*. **24,** 885 (2001)
33. W. R. Raun, et al., "In-season prediction of potential grain yield in winter wheat using canopy reflectance". *Agronomy Journal*. **93,** 131 (2001)
34. C. G. Pettersson, M. Soderstrom, and H. Eckersten, "Canopy reflectance, thermal stress, and apparent soil electrical conductivity as predictors of within-field variability in grain yield and grain protein of malting barley". *Precision Agriculture*. **7,** 343 (2006)
35. N. Aparicio, et al., "Spectral vegetation indices as nondestructive tools for determining durum wheat yield". *Agronomy Journal*. **92,** 83 (2000)
36. M. Moriondo, F. Maselli, and M. Bindi, "A simple model of regional wheat yield based on NDVI data". *European Journal of Agronomy*. **26,** 266 (2007)

416 Table 1. Overview of the weather variables.

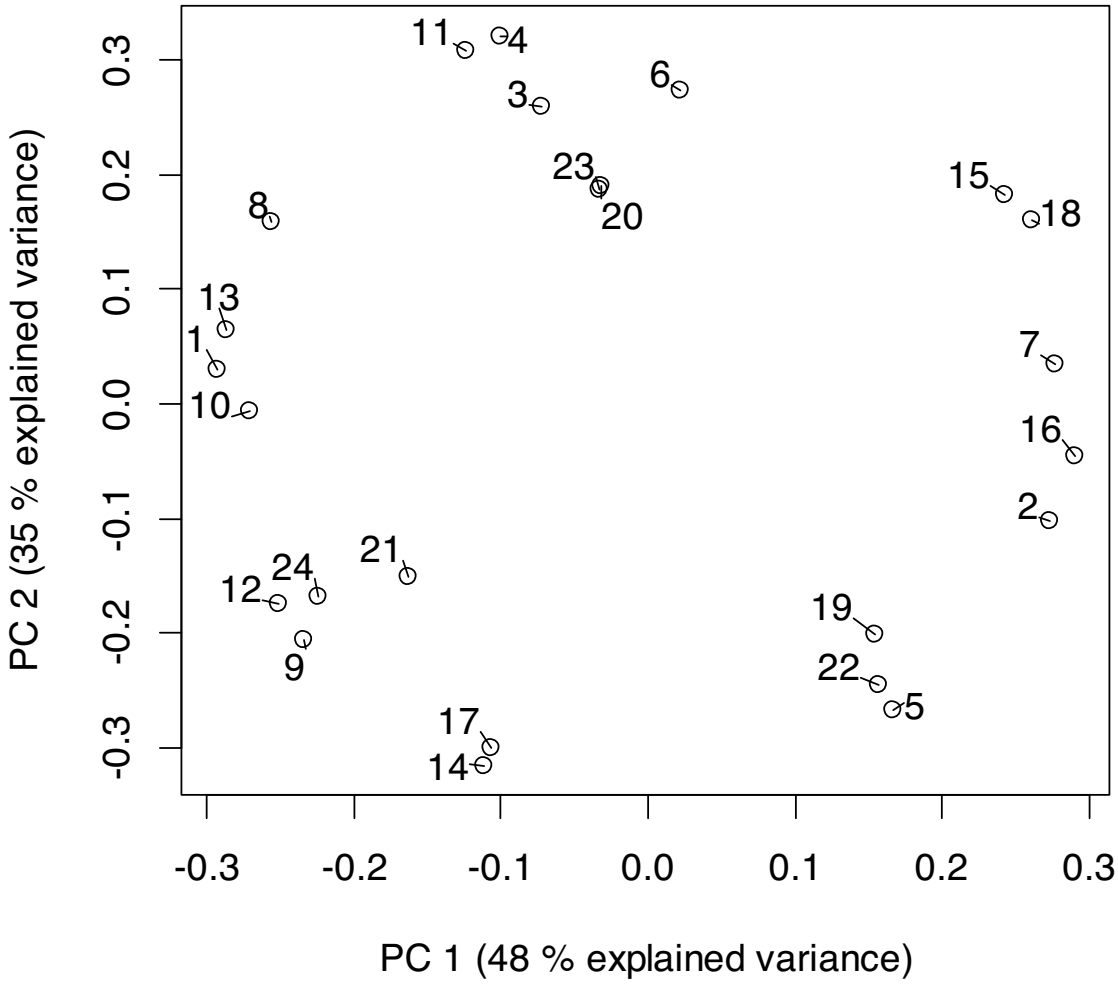
Variable number	Variable (phase[1])	Phenological phase[1]	Unit	Mean	Range
1	Temperature	1	°C	8.5	7.6 – 10.3
2	Precipitation	1	mm	12.2	0.0 – 18.7
3	Radiation	1	$W \cdot m^{-2}$	170.0	148.9 – 183.6
4	Windspeed	1	$m \cdot s^{-1}$	1.1	0.8 – 1.3
5	Humidity (P3)	1	%	59.4	54.0 – 70.0
6	Evapotranspiration	1	mm	20.2	17.9 – 22.1
7	Temperature	2	°C	11.9	9.3 -- 13.0
8	Precipitation	2	mm	27.4	15.6 – 40.2
9	Radiation	2	$W \cdot m^{-2}$	305.5	281.7 -358.3
10	Windspeed	2	$m \cdot s^{-1}$	0.9	0.7 – 1.3
11	Humidity (P3)	2	%	58.7	54.8 – 62.4
12	Evapotranspiration	2	mm	36.4	32.7 – 46.2
13	Temperature	3	°C	14.4	14.1 – 14.9
14	Precipitation	3	mm	68.2	39.9 – 97.8
15	Radiation	3	$W \cdot m^{-2}$	637.1	567.8 – 671.9
16	Windspeed	3	$m \cdot s^{-1}$	0.7	0.5 – 0.8
17	Humidity (P3)	3	%	60.7	57.8 – 64.2
18	Evapotranspiration	3	mm	820	70.6 – 86.3
19	Temperature	4	°C	15.5	14.6 – 16.5
20	Precipitation	4	mm	113.7	62.2 – 154.3
21	Radiation	4	$W \cdot m^{-2}$	385.3	321.4 – 456.8
22	Windspeed	4	$m \cdot s^{-1}$	0.7	0.6 – 0.8
23	Humidity	4	%	73.7	67.0 – 78.9
24	Evapotranspiration	4	mm	57.4	52.1 – 66.4

417 [1]According to the definition in section 2.3.

418 Table 2. Model diagnostics for the leave-year-out cross-validation: Root Mean
419 Square Error of Prediction (RMSEP), bias, Standard Error of Performance (SEP),
420 skewness[1], P-values for the test of differences in prediction ability in the
421 spectroscopic-only (PPLS) and the fusion models (CPPLS), explained variance (R^2)
422 and the number of CPPLS factors in each model.

Diagnostic measure	2007		2008		2009		2010	
	PPLS	CPPLS	PPLS	CPPLS	PPLS	CPPLS	PPLS	CPPLS
RMSEP	156	85	76	70	218	80	176	103
Bias	-89	-42	-14	-14	148	42	82	61
SEP	128	74	74	69	160	68	156	83
Skewness	1.98	1.87	0.59	0.54	0.38	0.23	0.30	0.18
P-value	<0.01		0.09		<0.01		<0.01	
R^2	0.94		0.85		0.76		0.78	
Number of factors	4		4		7		5	

423 [1]Third moment about the mean of the resiudals.



424

425 Figure 1. Principal Component Analysis (PCA) loadings for the 24 weather variables
426 on the first (abscissa) and second (ordinate) PCA component. The data point labels
427 correspond with the numbering in Table 1.

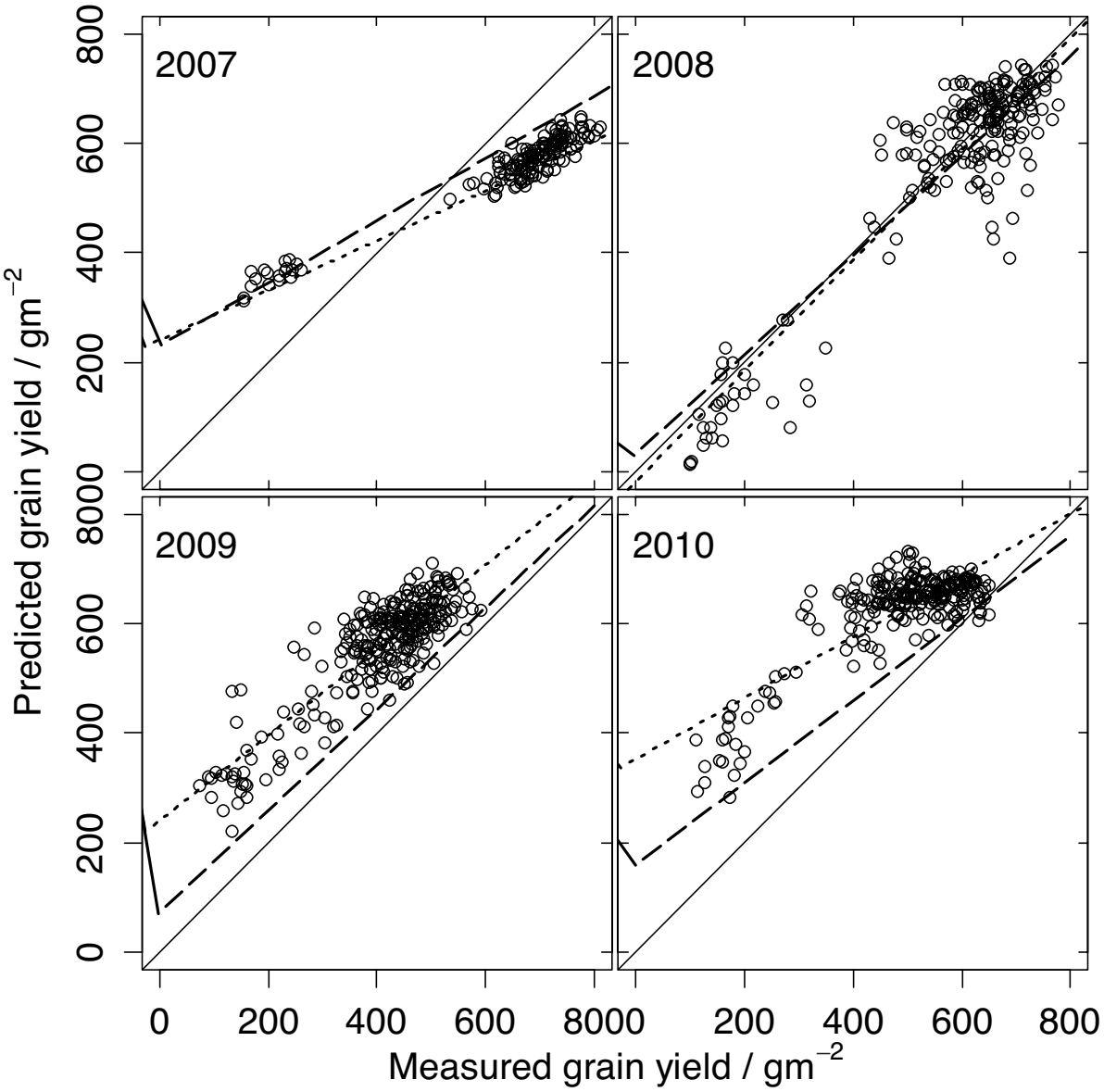
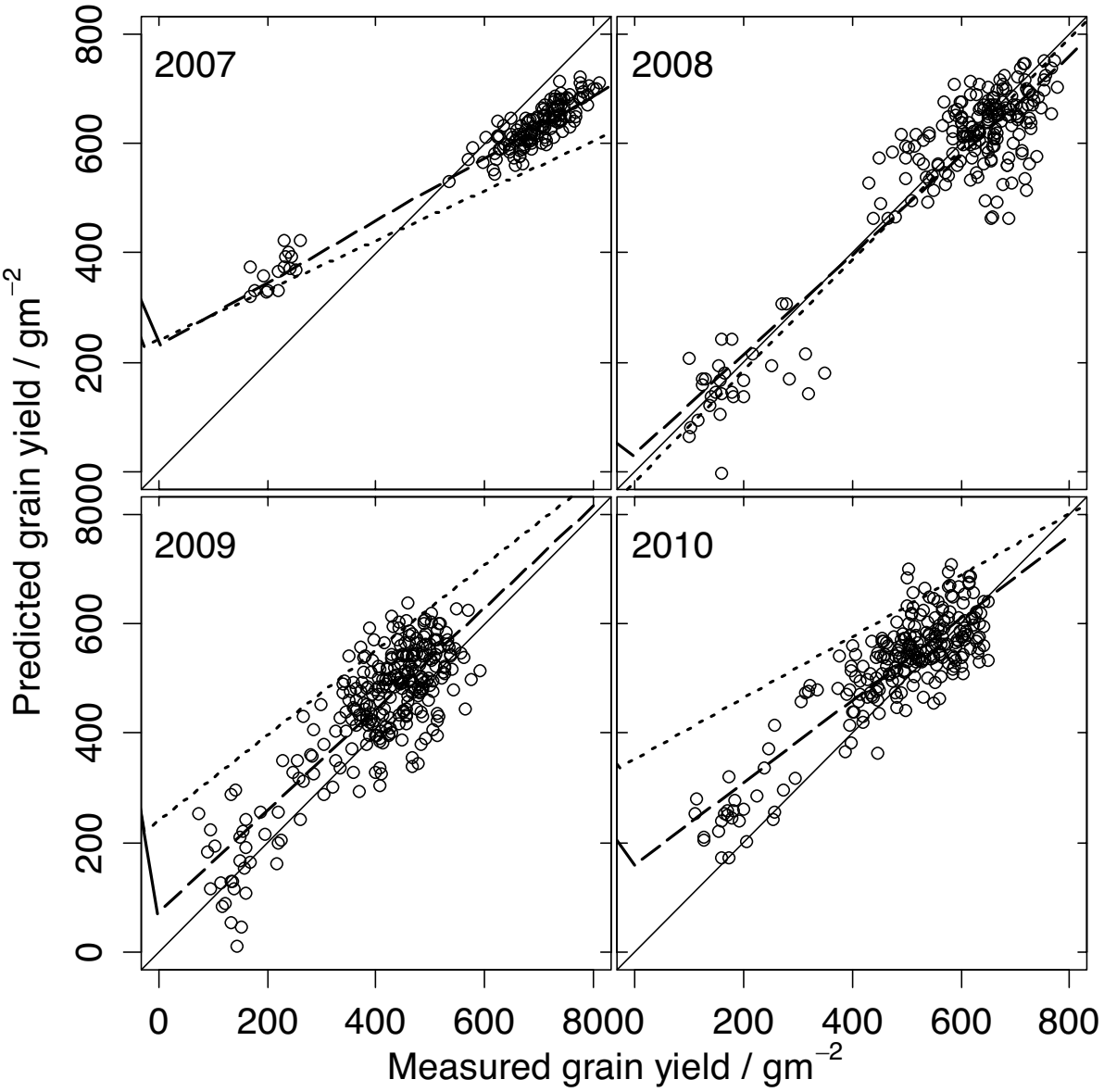


Figure 2. Measured grain yield plotted against grain yield predicted with the spectroscopic PPLS models, using data from each of the years 2007-2010 for validation (validation year denoted in each subplot), respectively. Target lines (solid) have zero intercept and slope 1. Regression lines from PPLS (dotted) and CPPLS (dashed) prediction models are displayed in each subplot.



435

436 Figure 3. Measured grain yield plotted against grain yield predicted with the fusion
437 CPPLS models, using data from each of the years 2007-2010 for validation
438 (validation year denoted in each subplot), respectively. Target lines (solid) have
439 zero intercept and slope 1. Regression lines from PPLS (dotted) and CPPLS (dashed)
440 prediction models are displayed in each subplot.