Citation for the published paper:
Zhenjiang Zhou, Julien Morel, David Parsons, Sergey V. Kucheryavskiy, Anne-Maj Gustavsson. (2019) Estimation of yield and quality of legume and grass mixtures using partial least squares and support vector machine analysis of spectral data. *Computers and Electronics in Agriculture*.
Volume: 162, Number: July 2019, pp 246-253.
https://doi.org/10.1016/j.compag.2019.03.038

Estimation of yield and quality of legume and grass mixtures using partial least squares and support vector machine analysis of spectral data

Zhenjiang Zhou[a b#], Julien Morel[b#], David Parsons[b], Sergey V. Kucheryavskiy[c], Anne-Maj Gustavsson[b]*

[a] College of Biosystems Engineering & Food Science, Zhejiang University, 310058 Hangzhou, China

[b] Department of Agricultural Research for Northern Sweden, Swedish University of Agricultural Sciences, SE-901 83 Umeå, Sweden

[c] Department of Chemistry and Bioscience, Aalborg University, DK-6700 Esbjerg, Denmark

[*]Corresponding author: Anne-Maj Gustavsson

Email: Anne-Maj.Gustavsson@slu.se

[#] Shared first author

Email of authors:

Zhenjiang Zhou Email: zhenjiangz@zju.edu.cn

Julien Morel Email: julien.morel@slu.se;

David Parsons Email: david.parsons@slu.se;

Sergey V. Kucheryavskiy Email: svk@bio.aau.dk;

Anne-Maj Gustavsson Email: Anne-Maj.Gustavsson@slu.se;

Abstract

The project aim was to estimate N uptake (N_{up}), dry matter yield (DMY) and crude protein concentration (CP) of forage crops both during typical harvest times and at a very early developmental stage. Canopy spectral reflectance of legume and grass mixtures was measured in Sweden using a commercialized radiometer (400 – 1000 nm range). In total, 377 plant samples were tested in-situ in different grass and legume mixtures (6 grass species and 2 clover species) across two years, two locations and five N rates. Two mathematical methods, namely partial least squares (PLS) and support vector machine (SVM) were used to build prediction models between N_{up}, DMY and CP, and canopy spectral reflectance. Of the total 377 samples, 251 were randomly selected and used for calibration, and the remaining 126 samples were used as an independent dataset for validation. Results showed that the performance of SVM was better than PLS (based on mean absolute error (MAE) for both calibration and validation datasets) for the estimation of all investigated variables. Results for the validation set showed that the MAEs of PLS and SVM for N_{up} estimation were 17 and 9.2 kg/ha, respectively. The MAEs of PLS and SVM for DMY estimation were 587 and 283 kg/ha, respectively. The MAEs of PLS and SVM for CP estimation were 2.8 and 1.8 %, respectively. In addition, a subsample, which corresponded to an early developmental stage, was analysed separately with PLS and SVM as for the whole dataset. Results showed that SVM was better than PLS for the estimation of all investigated variables. The high performance of SVM to estimate legume and grass mixture N uptake and dry matter yield could provide support for varying management decisions including fertilization and timing of harvest.

Keywords: Dry matter yield; Forage crop; Grass; Hyperspectral reflectance; Nitrogen uptake; Nutritive value; Partial least squares; Red and white clover; Support vector machine.

1. Introduction

Estimation of forage biomass and quality in the field is important for livestock farmers to make decisions such as adjusting the stocking rate, fertilization rate and timing, and harvest time. Laboratory wet chemical analysis has been used as the conventional method to determine nutritive value. However, these methods are laborious and expensive, and lead to time delays for decision making. Alternatively, near-infrared reflectance spectroscopy (NIRS) has been applied as a faster method to estimate nutritive value (Norris et al., 1976), particularly fibre, crude protein (CP), and metabolisable energy (ME). However, laboratory-based NIRS still involves laborious cutting and processing of plant samples. A fast and less laborious method is needed to estimate nutritive value, in order to provide rapid information for supporting decision making.

The canopy spectral reflectance (CSR) method has been developed to estimate crop biomass and physiochemical properties (e.g. N concentration), based on the principle that CSR in visible and near-infrared (NIR) bands is primarily affected by physiochemical properties (e.g. chlorophyll concentration) and cell structure of the canopy (Campbell, 1996). Numerous studies have concluded that CSR in the red band correlates well with chlorophyll and N concentration (Heath, 1969; Hatfield et al., 2008), and that CSR in the NIR range correlates well with plant biomass and leaf area (Allen and Richardson 1968).

Field spectroscopy has been successfully used to estimate nutritive value of forage crop (Starks et al., 2004; Biewer et al., 2009b; Pullanagari et al., 2012) and yield (Biewer et al., 2009a). Currently, an already commercialized spectrometer named the Yara N-sensor (Yara International ASA, Oslo, Norway), has been widely applied in Northern Europe to estimate N status of winter wheat and assist N fertilization decisions. However, it has not been tested in forage crops, even though increasing interest has been raised in this area, and forages often dominate agricultural land use, such as in Sweden where the current study was undertaken.

In addition to the potential practical application, from a scientific perspective, there has been little research conducted to estimate forage crop yield and quality using field spectroscopy. The advantage of a field spectroscopy method is that it measures CSR in a broad range of bands, providing an abundant supply of information. Chemometric methods, which are used for the analysis of spectral data, include principal component analysis (PCA), partial least squares regression (PLS), and machine learning methods such as support vector machine (SVM) (Mutanga et al., 2005; Zhao et al., 2007; Karimi et al., 2008; Wang et al., 2013). For estimation of crop agronomic variables using spectral reflectance data, PLS has typically been used, for example to estimate winter wheat N concentration (Li et al., 2014a), and winter wheat leaf area index (Li et al., 2014b). PLS has also been shown to work well to estimate yield and quality in forage crops (Mutanga et al., 2005; Zhao et al., 2007; Biewer et al., 2009). In contrast to PLS, which is a linear mathematical method, SVM is a non-linear method and based upon statistical learning theory. The principle of SVM and its solution is described in Cristianini and Shawe-Taylor (2000). Non-linear multivariate models of SVM have been used for many applications, e.g. rice root density (Xu et al., 2017), weed and maize classification (Zheng et al., 2017), wheat plant density estimation (Jin et al., 2017), leaf area estimation (Durbha et al., 2007; Yang et al., 2008), crop N concentration (Karimi et al., 2008; Wang et al., 2013; Zhai et al., 2013). We are unaware of any studies conducted to predict yield or quality parameters of forage crops in the field using SVM. Therefore, the objective of this study was to test the performance of spectral reflectance data analyzed using PLS and SVM, for estimating forage yield and quality parameters.

2. Materials and Methods

2.1. Site description and plant sampling

Field experiments were carried out in 2015 and 2016 at two sites in Sweden: Röbäcksdalen (63^o 48' N, 20^o 14' E) and Rådde (57^o 36' N, 13^o 15' E). Each site included different mixtures of grass and legume species (Table 1-2). The soil at Röbäcksdalen is a silt loam (1% clay, 76% silt, 23% sand) with 4 % organic matter. The soil at Rådde is also a silt loam (1 % clay, 53 % silt 46 % sand) with 6 % organic matter. The plot size was 1.5×13.5 m at Röbäcksdalen and 2.1×12.5 m at Rådde. The harvested area was 20.3 m^2 at Röbäcksdalen and 18.8 m^2 (1.5 ×12.5 m) at Rådde, which was used for fresh yield determination. A representative sub-sample was selected from the harvested material for dry matter determination and chemical analysis. Sub-samples were oven-dried at 60^{o}C for 48 hours until they reached a constant weight. Dried samples were ground to pass a 1-mm sieve. Nitrogen concentration of dry samples was determined using the Kjeldahl method.

Crops at each site were harvested three times each year. The harvest dates at Rådde were 4 June, 10 July and 24 August in 2015; and 27 May, 5 July and 22 August in 2016. The harvest dates at Röbäcksdalen in 2016 were 13 June, 18 July, and 2 September, however for the 1st harvest the YARA N-sensor data are not available due to equipment malfunction. In addition, in 2016 samples were also taken at an early developmental stage before the 1st harvest (corresponding to the time window of N fertilization), in order to investigate if crop N uptake can be estimated using field measured reflectance. The early cutting dates were 12 May and 3 June at Rådde and Röbäcksdalen, respectively.

Table 1 and 2

2.2. Canopy spectral reflectance measurement

On the same day as harvesting, canopy spectral reflectance was measured using a Yara N-sensor (Yara International ASA, Oslo, Norway) spectrometer, which is a commercialized instrument used for site-specific fertilization in Sweden and other European countries. It measures CSR at wavelengths of 400 – 1000 nm with resolution of 10 nm, a 25^{o} field of view and a zenith view angle of 45^{o} (Fig.1). The solar radiation was measured simultaneously for calculation of CSR reflectance by dividing the radiant exitance by the solar irradiance (the reader is referred to (Schaepman-Strub et al., 2006) for definitions). The sensor was held at a constant height of 1.0 m above the canopy. Measurements were taken around noon when the sun was not obscured by clouds. To take into consideration the effect of solar direction on CSR, measurements were taken at opposite ends of each plot, corresponding to facing the sun and opposite to the sun direction. The mean value of measurements in each plot was used for data analysis.

Fig. 1

2.3. Data analysis

Spectral reflectance data and laboratory measurements of the variables of interests were analysed with two mathematical methods, namely partial least squares (PLS) and support vector machine (SVM) to build regression models between CSR and variables of interest. Two datasets were used to build the models. The first one consisted of 377 samples that were randomly split into calibration (251 samples) and validation (126 samples) subsets. This dataset included all available data across sites, years, species, harvest dates and N fertilization rates. The aim was to test if a robust prediction model could be built for estimations of N_{up} (the nitrogen uptake, in kg/ha), DMY (the dry matter yield, in kg/ha) and CP (the crude protein concentration, in % of DM). In the second dataset, only the data acquired at an early developmental stage (i.e., sampled prior to the 1^{st} harvest) were considered. This dataset consisted of 78 samples split into calibration (52 samples) and validation (26 samples) subsets. The aim of only using data from an early developmental stage was to test if the variables of interest (especially N uptake) can be well estimated, as accurate estimation of N uptake could provide information on soil N mineralization at a very early stage and guide decision support on N fertilization.

2.3.1 Partial least squares

PLS is a prominent method which can effectively deal with multicollinearity in data and eliminate the less important or redundant variables. It is especially useful when the number of predictor variables is greater than the number of observations (Wold et al., 2001). PLS is used to linearly relate predictor variables (CSR in this study) and response variables (N_{up}, DMY and CP) by decomposing data matrices (predictors and response variables) simultaneously using a set of orthogonal latent variables (PLS-components). The orientation of latent variables is selected to maximize covariance between the predictors and response values represented by the components.

In this research we implemented PLS methods using the “mdatools” package in the R programming environment (R Core Team, 2016). Models were built using a calibration set and full cross-validation. The optimal number of components was identified by looking at the minimum root mean square error (RMSE) for the cross-validated predictions. External validation using an independent validation dataset was subsequently conducted to test the performance of the calibrated model.

2.3.2. Support vector machine

Support vector machine (SVM) is a supervised statistical learning algorithm developed by Vapnik (1982) that can be applied to both classification and regression tasks. It has gained increasing popularity for various purposes such as expression recognition (Lekdioui et al., 2017), water resources management (Deo et al., 2017) and medicine (Zheng et al., 2014). The general idea behind SVM is to map the vectors of covariates into a higher dimensional feature space using a kernel trick where a linear regression can be performed. Details of the theory of SVM are described in Cristianini and Shawe-Taylor (2000). The linearity of the relationship between N_{up}, CP and DMY and the spectral reflectance remains uncertain. With this consideration, a SVM, capable of constructing both linear and nonlinear inversion, was employed in this study. We used the *svm* function from the "e1071" package of the R software (R Core Team, 2016) to estimate N_{up}, CP and DMY of the experimental plots based on their spectral reflectance. Based on a radial basis kernel, we performed a grid search to set the optimal values of the hyper-parameters ε, C and γ, as they influence the accuracy and generalisation capabilities of the SVM (Cherkassky and Ma, 2004; Wang et al., 2003). ε is the insensitive-loss function that penalizes the prediction errors that fall within the $\pm\varepsilon$ range. C is the cost parameter that defines the penalty weight of deviations higher than $\pm\varepsilon$. Finally, γ is a radial basis kernel-specific parameter that controls the tradeoff between error due to bias and the variance in the adjusted model.

In order to test the performance of different models, mean absolute error (MAE), root mean square error (RMSE) (Loague and Green, 1991), and R^2 were compared for the calibration and validation subsets.

$$MAE = \frac{\sum_{i=1}^{n} |P_i - O_i|}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (P_i - O_i)^2} \quad (2)$$

where P_i and O_i are predicted and observed values, respectively, and n is the number of samples.

3. Results

3.1. Statistics of nitrogen uptake, dry matter yield and crude protein concentration

At Rådde, the average N_{up}, DMY and CP were 84 kg/ha, 3131 kg/ha, and 17 %, respectively, and at Röbäcksdalen, 30 kg/ha, 1785 kg/ha, and 12 %, respectively (Fig.2). There was large variation within all response variables at both sites, providing a wide range of data for the calibration of the models.

Fig. 2

3.2. Spectral measurements

Examples of spectral reflectance signatures were plotted to show the typical effects of N rate and mixture type (Fig.3, Fig.4). In the visible range (400 – 700 nm), the reflectance was lower in the blue and red parts of the spectrum, and higher in the green part of the spectrum, due to the leaf-contained chlorophyll. The spectral signature showed a sharp increase between 680 and 740 nm (known as the red-edge domain) to reach a plateau in the measured near infrared (NIR) domain, from 740 to 1000 nm. A concave shape was systematically observed between 950 and 1000 nm, due to the leaf-contained water (Peñuelas et al., 1993). "Pure grass" showed a higher reflectance in the visible range compared to "grass and clover mixture", whereas in the NIR the reflectance was higher for the "grass and clover mixture" (Fig. 3). The N rate also had an influence on the spectral signature, as treatments fertilized with high N levels (300 kg N/ha) showed higher reflectance in the NIR domain and lower reflectance in the visible domain, when compared to the zero N treatment (Fig. 4). All spectra showed an increase of reflectance during the season, which is particularly noticeable in the NIR domain. This increase is more pronounced for the "grass and clover mixture" than for the "pure grass" (Fig. 3). The dynamics are similar for the two contrasted nitrogen levels (Fig. 4).

Fig. 3

Fig. 4

3.3. Calibration of PLS and SVM models

R^2 values obtained for PLS models were in the range of 0.65-0.77 for the whole dataset and 0.79-0.88 for the early cutting dataset (Table 3). R^2 values obtained for SVM were in the range of 0.90-0.96 for the whole dataset and 0.95-0.98 for the early cutting dataset.

For the whole dataset, MAE for PLS were 15.4 kg/ha, 512 kg/ha and 2.4 %-units for N uptake, dry matter yield and crude protein concentration, respectively (Table 3). These MAEs were higher than the corresponding statistics for SVM. For the early cutting dataset, the MAEs were lower than for the whole dataset, and were again lower for SVM than for PLS. The RMSE results followed the same trend as MAE.

Table 3

3.4. Validation of PLS and SVM model

Both PLS and SVM were validated using randomly selected sub-samples from either the whole dataset or early cutting dataset (Table 4). For the whole dataset, R^2 values for PLS were 0.68 – 0.69 and for SVM 0.84 – 0.92. For the early cutting dataset, R^2 values for PLS were 0.80 – 0.92 and for SVM 0.83 – 0.94. Typically, R^2 values for validation were lower than those for calibration, with some exceptions.

For the whole dataset, MAE of SVM were 9.2 kg/ha, 283 kg/ha and 1.8 %-units for N uptake, dry matter yield and crude protein concentration, respectively (Table 4). For the early cutting dataset, MAE of SVM were 7.5 kg/ha, 123 kg/ha and 2.0 %-units for N uptake, dry matter yield and crude protein concentration, respectively. Both MAE and RMSE of validation were smaller for SVM than for PLS, irrespective of the dataset or variables.

The measured and predicted variables of the validation subsets were plotted using different crop species and experimental sites as markers (Fig. 5 and Fig. 6). Across the whole dataset and early cutting dataset they are distributed evenly around the 1:1 lines. No distinct groups of species or site factors were detected, implying that forage species, cultivars and sites did not impose substantial influence on the prediction using PLS and SVM models. A stronger relationship between measured and predicted values was found in the early cutting dataset than the whole dataset.

Table 4

Fig. 5 and Fig. 6

4. Discussion

4.1. Reflectance measurements

The observed effects of N fertilization on reflectance are consistent with previous studies (e.g., Hinzman et al., 1986; McMurtrey et al., 1994; Yoder and Pettigrew-Crosby, 1995; Daughtry, 2000). These results are mainly related to (i) an increase in chlorophyll content that leads to a reduced reflectance in the visible range (especially in the green and red domains of the spectrum) and (ii) an increase of the leaf biomass that leads to an increased reflectance in the NIR range (Fig. 4).

A similar pattern is observed when clover is added in the crop mixture (Fig. 3). Clover is a leguminous plant, and as such, it is less sensitive to low N fertilization, due to its ability to fix atmospheric N. This can explain the differences in reflectance measured for pure grass vs clover-grass mixtures, although the differences in leaf chemistry and architecture between species might also impose influence on their spectral reflectance signatures.

An increase of the measured reflectance is observed for all samples (exemplified in Fig. 3 and Fig. 4). Such an increase can be explained by the concomitant increase of temperatures throughout the summer. The greater increase for the “grass and clover mixture” might suggest that there is either (i) a lower nitrogen limitation for the growth when compared to "pure grass" or (ii) that the canopy structure of the "grass and clover mixture" would induce a stronger light reflection when compared to "pure grass".

4.2. PLS and SVM

The aim of this study was to predict N_{up}, DMY and CP using reflectance in the range of 400 – 1000 nm measured using a commercialized tool. In this study, SVM performed better than PLS (in terms of MAE, RMSE and R^2) for estimation of all response variables. Advantages of SVM lie in its robustness and insensitivity to the number of dimensions (Brown et al., 2000; Wu et al., 2008; Yao et al., 2015). SVM is also better than other methods at coping with potential confounding factors for different varieties, sites and developmental stages (Yao et al., 2015). Our results indicate that a non-linear relationship might exist between leaf N and canopy spectral reflectance (which can be seen in the top-right sub-plot in Fig. 5). This can also be a reason that non-linear SVM was superior to linear PLS for estimating quality variables in this study. Our results corroborate those of Yao et al. (2015) and Du et al. (2016), who showed that SVM performed better than PLS to estimate wheat and rice N concentrations.

4.3. Application of oblique optics sensors and its limitations

Many studies using spectral reflectance for N estimation have been conducted with handheld sensors that measured CSR in the nadir direction. One novelty of this study is that the optical lens had an oblique view of 45 degrees (Fig. 1). The design of the instrument imitates the commercialised tractor-mounted version (Fig. 1), which is already widely used in northern Europe. Therefore, the results from this study are highly transferable to the tractor-mounted spectral sensor (e.g. by encoding SVM methodology to a portable and cheap computer that is connected to the spectral sensor). The advantage of the oblique sensor is that it can be more easily fitted to a tractor (Fig.1) than a nadir-directed sensor. However, oblique sensors can potentially cause the problem of measurement variation depending on tractor driving direction with respect to the sun position. For example, the spectral reflectance from the oblique sensor facing to the sun-lit side of the plant will be different from the one facing the shaded side of the plant. This problem was addressed in our study by measuring spectral reflectance from each end of the experimental plots. For the application of this method on the tractor-mounted sensor, the variation of reflectance caused by tractor driving direction and sun position can be addressed by configuring the device in a way that fibre optics are oriented in four directions (two fibre optics on each side of the tractor). Such a configuration can take measurement in four directions, encompassing both the sun-lit and shaded sides of the plant. As the interaction between canopy geometry, solar elevation angle, solar azimuth angle and viewing angle influence observed reflectance (Jackson et al., 1979; Ranson et al., 1985), further studies are needed to clarify the response of reflectance to different combinations of solar position and sensor viewing angle for legume and grass mixture fields. A field bi-directional reflectance distribution function method could be used to build such a relationship. Despite the problems related to the oblique sensors mentioned above, it has been demonstrated that oblique sensors allow a more accurate estimation of field traits compared to nadir sensors. The potential of an oblique sensor for N status estimation has been investigated by Mistele et al. (2004) and Mistele and Schmidhalter (2008), who measured canopy reflectance from four oblique quadrilateral-views. This makes the measured reflectance almost independent of the solar zenith angle since both the sun-lit and shaded sides of plants can be captured by the sensor. Another disadvantage related to nadir measurement is that it only measures the reflectance of the plant upper layer. Oblique sensors can capture more information of plant structure by measuring reflectance deeper in the canopy, thus improving the estimation accuracy of plant characteristics (Diner et al., 1999). The advantages of oblique sensors are demonstrated in several studies. For example, Gianelle and Guastella (2007) showed that oblique measurement is more appropriate than nadir for determining grass dry matter yield. Aparicio et al. (2004) found that the wheat dry matter yield can be more precisely estimated by oblique measurement when LAI is high. Perbandt et al. (2011) demonstrated that oblique measurement is better than nadir for DMY and metabolisable energy of maize, while CP is better estimated by nadir measurement.

The oblique sensor in this study showed good estimation of N uptake during the early developmental stage, which is the most important period for N application. Mounting oblique sensors on a tractor offers the possibility to scan the crop beside the tractor and apply N fertilizer accordingly.

4.4. Generalization of the models

Heterogeneous sward structures, varying number of species and varieties, different locations, and different atmospheric conditions may impose confounding effects on the relationship between spectral reflectance and nutritive value. For example, Biewer et al. (2009b) built separate prediction models for grass and legumes, resulting in enhanced prediction accuracy for CP. The intention of our study was to estimate forage yield and quality by spectral methods across sites, mixture types and developmental stages rather than have to develop specific models for different sites, developmental stage and species mixtures. Neither the performances of PLS or SVM models were strongly influenced by those factors (Fig. 5). Thus, the model developed from the whole dataset in this study might be more generally useful compared with those based on a given mixture type or developmental stage, with the caveat that with more data collection such models can be improved and refined. One of the constraints of this study is that spectral method was only tested ton CP, while other forage nutrition variables were not measured, nonetheless, there is a high correlation in forage crops between CP, fibre content, ash, lignin, and metabolisable energy (Pullanagari et al., 2012), indicating potential for extending the method from this study for estimation of other nutrition variables.

5. Conclusion

The SVM method was better than PLS for forage yield and quality estimation. The performance of SVM models for estimating forage yield and quality was consistent among the calibration and validation datasets. These results imply that SVM is a promising tool to analyze on-field acquired spectral data for the estimation of forage crop quality and yield, especially when considering the tractor-mounted Yara-N sensor. Mounting oblique sensors in the very early crop developmental stage on a tractor offers the possibility to scan the crop and apply N fertilizer accordingly. The results presented in this study need to be confirmed with further measurements that would provide a more comprehensive database of sun and sensor interactions, including measurements during cloudy days, various row directions and solar positions. Nevertheless, as the models showed a good robustness and accuracy for various site and mixtures combinations, it is reasonable to assume that the models developed in this study could provide farmers with real time information on-the-go during the harvest period, informing the farmer about forage yield and quality.

Acknowledgement

The work was supported by the FORMAS, MISTRA and Lantmännen foundation. The study was conducted using data and material from Röbäcksdalen, SITES (Swedish Infrastructure for Ecosystem Science), a national coordinated infrastructure, supported by the Swedish Research Council.

References

Allen, W.A., Richardson, A.J., 1968. Interaction of light with a plant canopy. J. Opt. Soc. Am. 58, 1023-1028.

Aparicio, N., Villegas, D., Royo, C., Casadesus, J., & Araus, J. L., 2004. Effect of sensor view angle on the assessment of agronomic traits by groundlevel hyperspectral refelctance measurements in durum wheat under contrasting Mediterranean conditions. International Journal of Remote Sensing, 25, 1131-1152.

Biewer, S., Erasmi, S., Fricke, T., Wachendorf, M., 2009a. Prediction of yield and the contribution of legumes in legume-grass mixtures using field spectrometry. Precision Agriculture 10, 128-144.

Biewer, S., Fricke, T., Wachendorf, M., 2009b. Development of canopy reflectance models to predict forage quality of legume–grass mixtures. Crop Science 49, 1917-1926.

Brown, M.P., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M., Haussler, D., 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc. Natl. Acad. Sci. USA, 97, 262-267.

Campbell, J.B., 1996. Introduction of remote sensing. 2nd ed. The Guilford Press, London.

Cherkassky, V., Ma, Y., 2004. Practical selection of SVM parameters and noise estimation for SVM regression. Neural Netw. 17, 113-126. https://doi.org/10.1016/S0893-6080(03)00169-2

Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511801389

Daughtry, C., 2000. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. Remote Sens. Environ. 74, 229-239. https://doi.org/10.1016/S0034-4257(00)00113-9

Deo, R.C., Kisi, O., Singh, V.P., 2017. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmospheric Res. 184, 149-175. https://doi.org/10.1016/j.atmosres.2016.10.004

Diner, D.J., Asner, G.P., Davies, R., Knyazikhin, Y., Muller, J.-P., Nolin, A.W., Pinty, B., Schaaf, C.B., Stroeve, J., 1999. New directions in earth observing: Scientific applications ofMultiangle remote sensing. Bulletin of the American Meteorological Society 80, 2209-2228.

Du, L., Shi, S., Yang, J., Sun, J., Gong, W., 2016. Using Different Regression Methods to Estimate Leaf Nitrogen Content in Rice by Fusing Hyperspectral LiDAR Data and Laser-Induced Chlorophyll Fluorescence Data. Remote Sens. 8, 526.

Durbha, S.S., King, R.L., Younan, N.H., 2007. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. Remote Sensing of Environment 107, 348-361.

Gianelle, D., Guastella, F., 2007. Nadir and off-nadir hyperspectral field data: Strengths and limitations in estimating grassland biophysical characteristics. International Journal of Remote Sensing, 28, 1547-1560.

Hatfield, J., Gitelson, A.A., Schepers, J.S., Walthall, C., 2008. Application of spectral remote sensing for agronomic decisions. Agron. J. 100, 117-131.

Heath, O.V.S., 1969. Physiological aspects of photosynthesis. http://agris.fao.org/agris search/search.do?recordID=US201300595563

Hinzman, L., Bauer, M., Daughtry, C., 1986. Effects of nitrogen fertilization on growth and reflectance characteristics of winter wheat Remote Sens. Environ. 19, 47–61. https://doi.org/10.1016/0034-4257(86)90040-4

Jackson, R.D., Pinter, P.J., Idso, S.B., Reginato, R.J., 1979. Wheat spectral reflectance: Interactions between crop confguration, sun elevation, and azimuth angle. Appl. Opt. 18, 3730-3733.

Jin, X., Liu, S., Baret, F., Hemerlé, M., Comar, A., 2017. Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. Remote Sensing of Environment 198, 105-114.

Karimi, Y., Prasher, S., Madani, A., Kim, S., 2008. Application of support vector machine technology for the estimation of crop biophysical parameters using aerial hyperspectral observations. Canadian Biosystems Engineering 50, 13-20.

Lekdioui, K., Messoussi, R., Ruichek, Y., Chaabi, Y., Touahni, R., 2017. Facial decomposition for expression recognition using texture/shape descriptors and SVM classifier. Signal Process. Image Commun. 58, 300–312. https://doi.org/10.1016/j.image.2017.08.001

Li, F., Mistele, B., Hu, Y., Chen, X., and Schmidhalter, U., 2014a. Reflectance estimation of canopy nitrogen content in winter wheat using optimised hyperspectral spectral indices and partial least squares regression. European Journal of Agronomy 52, 198-209.

Li, X., Zhang, Y., Bao, Y., Luo, J., Jin, X., Xu, X., Song, X., and Yang, G., 2014b. Exploring the best hyperspectral features for LAI estimation using partial least squares regression. Remote Sensing 6, 6221-6241.

Loague, K., and Green, R., E., 1991. Statistical and graphical methods for evaluating solute transport models: overview and application. Journal of contaminant hydrology 7, 51-73.

McMurtrey, J., Chappelle, E., Kim, M., Meisinger, J., Corp, L., 1994. Distinguishing nitrogen fertilization levels in field corn (Zea mays L.) with actively induced fluorescence and passive reflectance measurements. Remote Sens. Environ. 47, 36–44. https://doi.org/10.1016/0034-4257(94)90125-2

Mistele, B., Gutser, R., Schmidhalter, U., 2004. Validation of field-scaled spectral measurements of the nitrogen status in winter wheat. In: Mulla, D.J. (Ed.), Proceedings of the 7th International Conference on Precision Agriculture and other precision resources management. Minneapolis, Minnesota, USA, (CD ROM), pp. 1187-1195.

Mistele, B., Schmidhalter, U., 2008. Spectral measurements of the total aerial N and biomass dry weight in maize using a quadrilateral-view optic. Field Crops Research, 106, 94-103.

Mutanga, O., Skidmore, A., K., Kumar, L., & Ferwerda, J., 2005. Estimating tropical pasture quality at canopy level using band depth analysis with continuum removal in the visible domain. International Journal of Remote Sensing, 26 (6), 1093-1108.

Norris, K.H., Barnes, R.,F., Moore, J.,E., and Shenk, J.,S., 1976. Predicting forage quality by infrared reflectance spectroscopy. J. Anim. Sci. 43: 889–897.

Peñuelas, J., Filella, I., Biel, C., Serrano, L., Savé, R., 1993. The reflectance at the 950-970 nm region as an indicator of plant water status. Int. J. Remote Sens. 14, 1887–1905. https://doi.org/10.1080/01431169308954010

Perbandt, D., Fricke, T., Wachendorf, M., 2011. Off-nadir hyperspectral measurements in maize to predict dry matter yield, protein content and metabolisable energy in total biomass. Precision Agriculture 12, 249-265.

Pullanagari, R., Yule, I., Tuohy, M., Hedley, M., Dynes, R., King, W., 2012. In-field hyperspectral proximal sensing for estimating quality parameters of mixed pasture. Precision Agriculture 13, 351-369.

R Core Team, (2016), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL. https://www.R-project.org/.

Ranson, K.J., Daughtry, C.S.T., Biehl, L.L., Bauer, M.E., 1985. Sun-view angle effects on reflectance factors of corn canopies. Remote Sens. Environ. 18,147-161.

Schaepman-Strub, G., Schaepman, M., Painter, T.H., Dangel, S., Martonchik, J.V., 2006. Reflectance quantities in optical remote sensing-Definitions and case studies. Remote sensing of environment 103, 27-42.

Starks, P.J., Coleman, S.W., Phillips, W.A., 2004. Determination of forage chemical composition using remote sensing. Journal of range management 57, 635-640.

Starks, P.J., Zhao, D., Phillips, W.A., Coleman, S.W., 2006. Herbage mass, nutritive value and canopy spectral reflectance of bermudagrass pastures. Grass Forage Sci. 61, 101–111.

Vapnik, V., 1982. Estimation of Dependences Based on Empirical Data: Springer Series in Statistics (Springer Series in Statistics). Springer-Verlag New York, Inc.

Wang, W., Xu, Z., Lu, W., Zhang, X., 2003. Determination of the spread parameter in the Gaussian kernel for classification and regression. Neurocomputing 55, 643–663. https://doi.org/10.1016/S0925-2312(02)00632-X

Wang, F., Huang, J., Wang, Y., Liu, Z., Zhang, F., 2013. Estimating nitrogen concentration in rape from hyperspectral data at canopy level using support vector machines. Precision Agriculture 14, 172-183.

Wold, S., Sjostrom, M., Eriksson, L., 2001. Pls-regression: A basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58(2), 109–130.

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Philip, S.Y., 2008. Top 10 algorithms in data mining. Knowledge and information systems 14, 1-37.

Xu, S.X., Zhao, Y.C., Wang, M.Y., Shi, X.Z., 2017. Determination of rice root density from Vis-NIR spectroscopy by support vector machine regression and spectral variable selection techniques. Catena 157, 12-23.

Yang, X.-H., Huang, J.F., Wang, X.Z., Wang, F.M., 2008. The estimation model of rice leaf area index using hyperspectral data based on support vector machine. Spectroscopy and Spectral Analysis 28, 1837-1841.

Yao, X., Huang, Y., Shang, G., Zhou, C., Cheng, T., Tian, Y., Cao, W., Zhu, Y., 2015. Evaluation of six algorithms to monitor wheat leaf nitrogen concentration. Remote Sens. 7, 14939-14966.

Yoder, B.J., Pettigrew-Crosby, R.E., 1995. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400–2500 nm) at leaf and canopy scales. Remote Sens. Environ. 53, 199–211. https://doi.org/10.1016/0034-4257(95)00135-N

Zhai, Y., Cui, L., Zhou, X., Gao, Y., Fei, T., Gao, W., 2013. Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: comparison of partial least-square regression and support vector machine regression methods. Int. J. Remote Sens. 34, 2502-2518.

Zhao, D., Starks, P. J., Brown, M. A., Phillips, W. A., & Coleman, S. W., 2007. Assessment of forage biomass and quality parameters of Bermuda grass using proximal sensing of pasture canopy reflectance. Grassland Science, 53(1), 39-49.

Zheng, B., Yoon, S.W., Lam, S.S., 2014. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Syst. Appl. 41, 1476–1482. https://doi.org/10.1016/j.eswa.2013.08.044

Zheng, Y., Zhu, Q.B., Huang, M., Guo, Y., Qin, J.W., 2017. Maize and weed classification using color indices with support vector data description in outdoor fields. Computers and Electronics in Agriculture 141, 215-222.

Table 1

Species, cultivars (cv), and seed rates used in the Röbäcksdalen field experiment.

Mixture treatment	Seed rate of different species and varieties (kg ha^{-1})						
	Timothy cv. Grindstad	Meadow fescue cv. Revansch	Tall fescue cv. Swaj	Festulolium cv. Hykor	Perennial ryegrass mixture of cvs.	Red clover cv. Torun	White clover cv. Undrom
TiMF	12.0	10.0					
TiTall	12.0		10.0				
TiHyk	12.0			10.0			
TiPerHyk	10.4			9.1	2.5		
TiTall-clover	9.7		7.9			3.3	1.1

Table 2

Species, cultivars (cv), and seed rates used in the Rådde field experiment.

Mixture treatment	Seed rate of different species and varieties (kg ha^{-1})							
	Timothy cv. Switch	Perennial ryegrass cv. Foxtrot	Perennial ryegrass cv. Kentaur	Festu-lolium cv. Hykor	Red clover cv. Vicky	White clover cv. Klondike	Festu-lolium cv. Felopa	Meadow fescue cv. Minto
TiPerHyk	10.4	1.25	1.25	9.1				
TiPerHyk-clover	8.4	1.0	1.0	7.3	3.3	1.1		
TiPerHyk-Fel	10.4	1.25	1.25	4.6			4.5	
TiMF	12.1							9.9

Table 3

Calibration statistics of the prediction of N uptake, crude protein concentration and dry matter yield by partial least squares regression (PLS) and support vector machine (SVM).

Dataset	Variables	n	Mean	PLS model			SVM model		
				MAE	RMSE	R^2	MAE	RMSE	R^2
All harvests	N uptake (kg/ha)	251	65	15.4	19	0.75	6.8	11	0.92
	Dry matter yield (kg/ha)	251	2696	512	667	0.77	200	284	0.96
	Crude protein (% of DM)	251	15	2.4	3.1	0.65	1.2	1.7	0.90
Early cutting	N uptake (kg/ha)	52	30	8.7	10	0.86	3.1	4	0.98
	Dry matter yield (kg/ha)	52	822	176	209	0.88	114	141	0.95
	Crude protein (% of DM)	52	19	2.1	2.8	0.79	1.2	1.3	0.95

n indicates the number of measurements, Mean is the average of measurements. MAE is mean absolute error. RMSE is root mean square error.

Table 4

Validation statistics of the prediction of N uptake, crude protein concentration and dry matter yield by partial least squares regression (PLS) and support vector machine (SVM).

Dataset	Variables	n	Mean	PLS model			SVM model		
				MAE	RMSE	R^2	MAE	RMSE	R^2
All harvests	N uptake (kg/ha)	126	67	17	21	0.69	9.2	12.6	0.89
	Dry matter yield (kg/ha)	126	2660	587	732	0.68	283	374	0.92
	Crude protein (% of DM)	126	16	2.8	3.5	0.69	1.8	2.5	0.84
Early cutting	N uptake (kg/ha)	26	39	7.9	9.7	0.90	7.5	9.9	0.89
	Dry matter yield (kg/ha)	26	972	153	182	0.92	123	156	0.94
	Crude protein (% of DM)	26	22	2.5	3.2	0.80	2.0	2.7	0.83

n indicates the number of measurements, Mean is the average of measurements. MAE is mean absolute error. RMSE is root mean square error

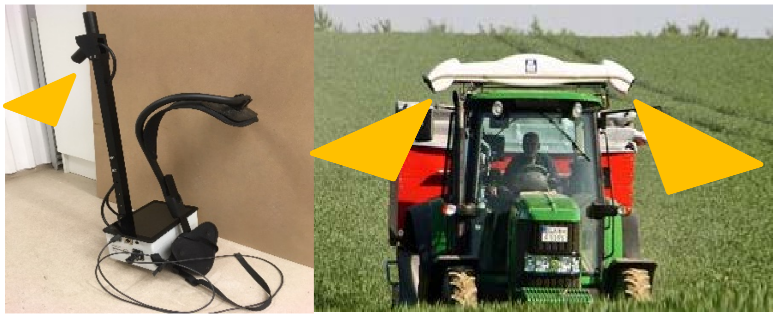

Fig. 1. The photo on the left shows the hand-held sensor used in this project. The photo on the right shows the tractor-mounted Yara-N sensor and is from the Yara company homepage. The white instrument on the roof of the tractor is the N sensor, which has an oblique view angle, enabling the estimation of N status on the crop on both sides of the tractor. The yellow triangle in both photos is a depiction of the field of view of the reflectance measurement.

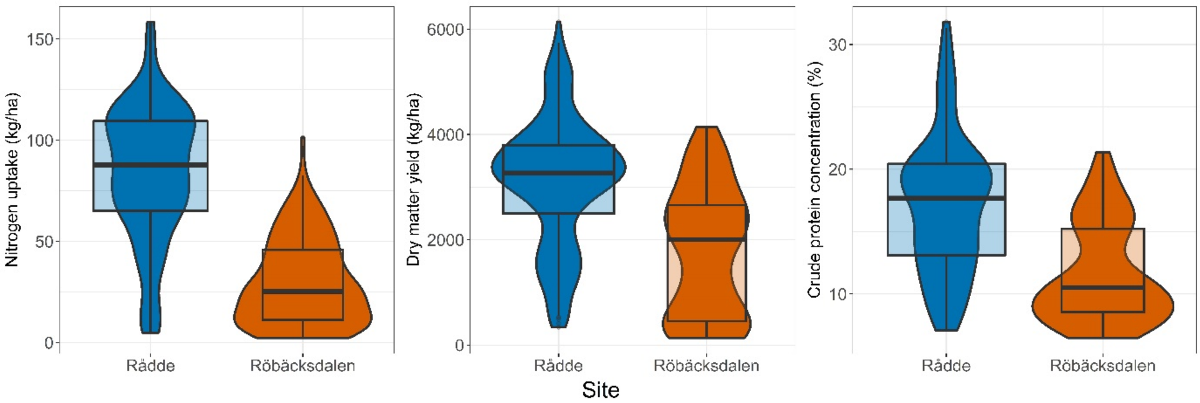

Fig. 2. Variation of agronomy variables of the whole dataset (377 samples) at two sites in 2015 and in 2016 for nitrogen uptake, dry matter yield and crude protein concentration. The horizontal lines indicate the minimum,

first quartile, median, third quartile, and maximum of the dataset. Sample distribution is indicated by the width of the colored area.

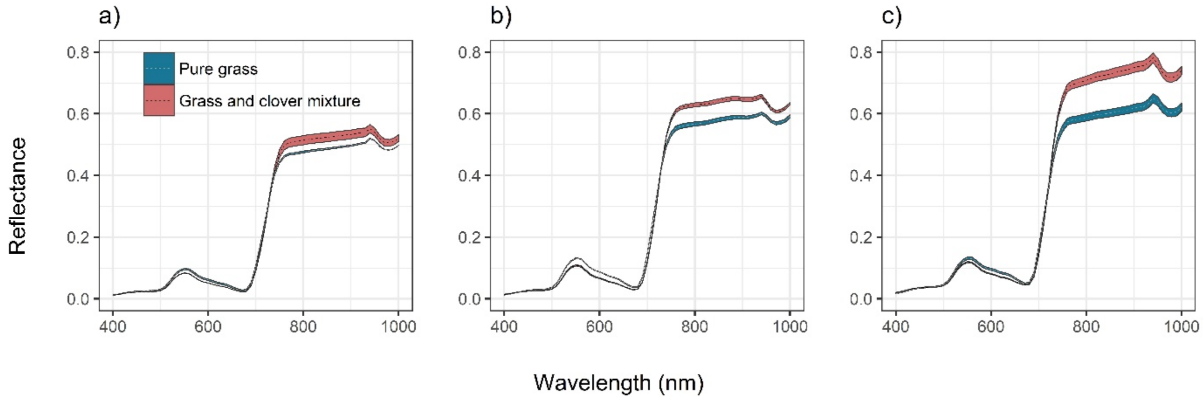

Fig. 3. Comparison of the spectral signatures of pure grass (TiTall) and grass and clover mixture (TiTall-clover) for three dates: a) 03/06/2016, b) 18/07/2016, c) 02/09/2016 at Röbäcksdalen. Only the zero N treatment was used for plotting the spectral signature.

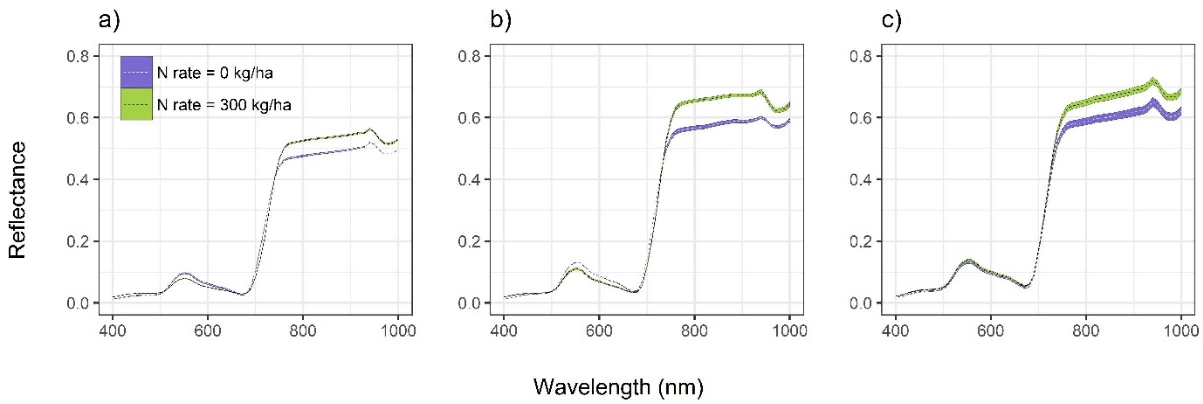

Fig. 4. Comparison of the spectral signatures of two contrasting nitrogen treatments for three dates: a) 03/06/2016, b) 18/07/2016, c) 02/09/2016 at Röbäcksdalen. Only the pure grass (TiTall) treatment was used for plotting the spectral signature.

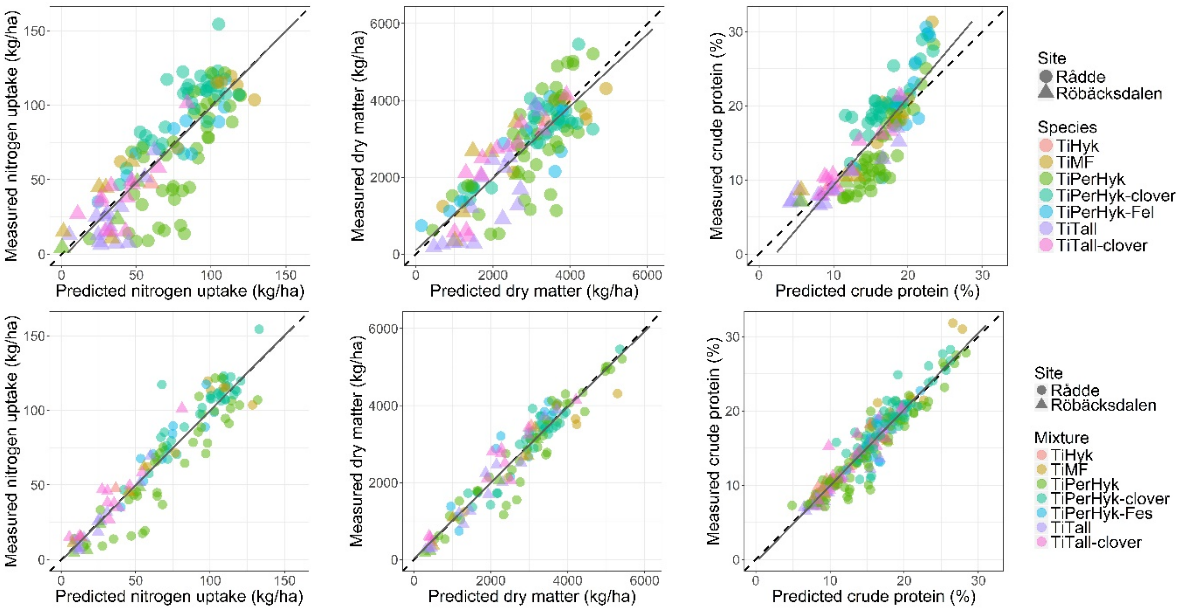

Fig. 5. Validation of partial least squares (upper graphs) and support vector machine (lower graphs) models for nitrogen uptake, dry matter yield and crude protein concentration using data from all harvests. Abbreviations for species mixtures are described in Tables 1 and 2.

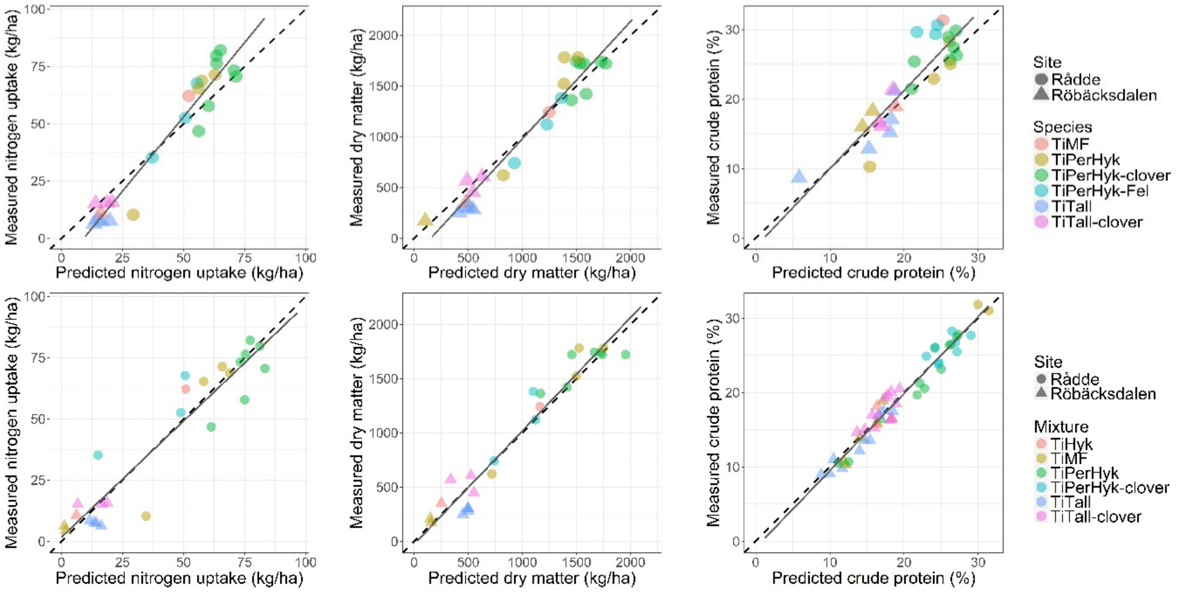

Fig. 6. Validation of partial least squares (upper graphs) and support vector machine (lower graphs) models for nitrogen uptake, dry matter yield and crude protein concentration using data from the early cutting only. Abbreviations for species mixtures are described in Tables 1 and 2.