LEGUMES

Development of Near Infrared Reflectance Spectroscopy Calibrations to Estimate Legume Content of Multispecies Legume–Grass Mixtures

F. Locher, H. Heuwinkel,* R. Gutser, and U. Schmidhalter

ABSTRACT

Legume content in legume–grass mixtures is a key parameter for the quantification of N_2 fixation, forage, and diet quality. This study was conducted (i) to develop a near infrared reflectance spectroscopy (NIRS) based method to estimate the legume content in multispecies legume–grass mixtures as in widespread use in Western Europe, (ii) to compare end-points and artificial mixture calibration strategies and (iii) to evaluate the effect grinding may have on the NIRS predictions of legume content. Calibration samples were taken in 1999 and 2000 in legume–grass fields that comprised a broad variation of site conditions. The samples were hand-sorted, dried, and ground. End-points calibrations derived from sets of legume samples (=100% legume content) and sets of grass samples (=0% legume content) were compared with calibrations where 63 spectra of artificially mixed samples (increments of 5% legume content) were added to represent a continuum of possible values of legume content. The influence of the preparation protocol of defined dry mixtures was compared by preparing duplicate mixtures where one replicate was prepared from fresh material, dried, and ground as a mixture and the other mixed from dry, ground material. Log (1/*R*) (*R* = reflectance) spectra were taken of all samples. Partial least squares regression was applied to develop calibration algorithms in the spectral range of 7500 to 3950 cm^{-1} (1333–2532 nm). First derivative combined with vector normalization proved to be the best data pretreatment. For each strategy, three models were developed: One model was based on all samples validated with a one-leave-out cross-validation, and two models were based on half of the samples validated by the other half. Prediction errors were between 2.2 and 4.0%, and coefficients of determination of all validations were greater than 99% so that no remarkable differences between the models existed. At least 70% of the selected spectral regions were in common for all models. These regions do not describe legumes themselves but rather the information that discriminates them from grasses. It is emphasized that the calibrations introduced have the potential for a broad use that needs to be proved by further validations.

Biological N fixation (BNF) by legumes is an important source of N in agriculture. Estimates of BNF vary strongly between and within species (LaRue and Patterson, 1981). Site-specific variation of N_2 fixation is rarely reported (Androsoff et al., 1995; Hansen and Vinther, 2001; Stevenson et al., 1995) although it is important information for sustainable low-input farming systems. Fixation of N_2 can simply be measured by the N yield of the legume multiplied by the proportion derived from symbiotic N_2 fixation (N derived from atmosphere). At a low level of N supply, the proportion of N derived from atmosphere is expected to be close to 100% if the legumes are grown in mixtures with nonlegumes. Then, N_2 fixation is mainly determined by the N yield of the legumes (Boller, 1988; Peoples et al., 1995; Weißbach, 1995). To calculate the N yield of the legumes grown in mixture, their share of the total yield of the mixture is needed. The legume content of mixtures varies seasonally and spatially with regard to botanical composition, which affects BNF. We are not aware of any literature that has shown these effects on BNF at field scale over time. One reason for this is the lack of a suitable method to easily determine legume content of mixtures. Traditional methods to determine legume content of legume–grass mixtures (botanical analysis, visual estimates, point quadrat, dry weight rank, N content, constituent differential method) are not suitable for this purpose. Even for a limited area, some of them are too labor intensive, most need well-trained operators or do not apply under all circumstances, and the information gathered is usually proportional to the resources invested (Whalley and Hardy, 2000). Additionally, these methods do not offer the possibility for automation in harvesting machines for future use in site-specific farming. Thus, an easy-to-use method is needed that may serve these goals and ensure a high repeatability.

Near infrared reflectance spectroscopy has already proven its capability to determine legume content in legume–grass mixtures. It is an easy-to-use technique that can even be mounted onto harvesting machines (Dardenne and Féménias, 1999). Moore et al. (1990) found NIRS less susceptible to maturity stage effects than the constituent differential method and highlighted the strong reduction in sample preparation and measurement. Other authors proved the capability of NIRS to predict the legume content in binary forage mixtures (Petersen et al., 1987; Pitman et al., 1991; Shaffer et al., 1990; Wachendorf et al., 1999). More complex mixtures with several legumes and grasses were successfully tested by Coleman et al. (1990) and Pitman et al. (1991). However, there is not yet any study that shows the capability of NIRS in predicting legume content of multispecies legume mixtures in widespread use in Western Europe, i.e., with white clover (*Trifolium repens* L.), red clover (*Trifolium pratense* L.), and alfalfa (*Medicago sativa* L.) as dominating legumes in varying proportions. Wachendorf et

Dep. of Plant Sci., Technische Universität München, D-85350 Freising, Germany. Received 10 Sept. 2002. *Corresponding author (hauke@wzw.tum.de).

Published in Agron. J. 97:11–17 (2005).

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: BNF, biological nitrogen fixation; NIRS, near infrared reflectance spectroscopy; RMSEC, root mean square error of calibration; RMSECV, root mean square error of cross-validation; RMSEP, root mean square error of prediction (test-set validation).

al. (1999) developed separate methods for white clover and red clover in binary mixtures with ryegrass, but these may not be suitable for multispecies mixtures. Additionally, method development as published so far was mostly based on samples from well-defined plot experiments. Finally, the methods were rarely tested for their performance in natural stands as they were by, e.g., Pitman et al. (1991), whose focus was the determination of accompanying nonlegumes in tropical pastures.

In all cited studies, the error for NIRS prediction of legume content ranged between 2 and 14%, but the data can hardly be compared because plant material, measurement, and calibration conditions and even calibration strategy differed, which all affect the prediction error. The performance of a NIRS model strongly depends on the quality of the reference data (Naes et al., 2002). Further, sample preparation and calibration strategy affect the prediction error. In the case of legume content, the reference samples can be gained by different protocols: (i) Reference values may come from artificially mixed, weighted mixtures (e.g., Pitman et al., 1991) or from real samples taken adjacent to NIRS sampling points (e.g., Coleman et al., 1990; Wachendorf et al., 1999); and (ii) the pure samples used to create artificial mixtures may have been grown in mixtures or pure stands. The use of samples from pure stands to create artificial herbage mixtures for calibration was successful for the determination of composite grass or legume concentration but not for the determination of single species therein (Pitman et al., 1991). Coleman et al. (1990) generally restricted the use of this strategy to closed populations, i.e., calibration and validation samples are two random subsets of the same population. The use of the most simple calibration strategy, i.e., end-points calibration, again showed promise with composite grass and legume groups (Coleman et al., 1990; Pitman et al., 1991). But, Petersen et al. (1987) and Coleman et al. (1990) discuss that differences in particle size distribution may differ for the same species whether it was ground solely or in mixtures even if the same grinding conditions are used. These differences will affect light scatter and thus the near infrared spectra. Coleman et al. (1990) have shown that math treatment on the spectra largely eliminates these effects.

We developed NIRS models (i) to determine the legume content in multispecies legume–grass mixtures that are widely grown under temperate climate conditions in Europe and (ii) to compare end-points and artificial mixture calibration strategies. Since our standard procedure was based on the preparation of samples from ground legume or grass batches, we (iii) evaluated the effect grinding may have on the NIRS predictions of legume content by preparing duplicate mixtures, where one replicate was prepared from fresh material, dried, and ground as a mixture and the other mixed from dry, ground material according to the standard procedure. The aim was to achieve a method with general applicability by addressing all variation that may come from site conditions, plant age, and species composition and a prediction error less than 5% legume content.

MATERIALS AND METHODS

Plant Material and Site Condition

The experimental sites were part of the FAM Research Station Scheyern, which is located in Southern Germany 40 km north of Munich (450–490 m above sea level, 803 mm mean annual precipitation, 7.4°C mean annual temperature). Samples were taken from two fields (2.3 and 2.4 ha in size) in 1999 and from a third field (3.5 ha) in 2000 (Table 1), all sown to a multispecies legume–grass mixture [alfalfa, white clover, red clover, orchardgrass (*Dactylis glomerata* L.), timothy (*Phleum pratense* L.), perennial ryegrass (*Lolium perenne* L.), tall fescue (*Festuca arundinacea* Schreb.), and oatgrass (*Arrhenatherum elatius* L.)]. Measurement plots (0.25 m^2 each) were established at selected sites that covered the range of soil heterogeneity found in the fields. Therefore, site-induced effects on the plant material were included. At the plots, aboveground biomass was harvested to gain material for the NIRS calibration procedure (Table 1) 1 or 2 d before cutting the whole field. Shoots were separated into the legume and grass fractions but not analyzed for their species composition because the aim was a robust calibration for the determination of the legume content as a whole. Yet, multispecies mixtures will always change their composition depending on management and site conditions, i.e., sometimes single species may dominate within the legume or grass fraction. We observed, e.g., in spring, a predominance of red clover in the legume fraction, whereas at other harvesting dates, alfalfa became more dominant. But mostly all legume species sown were present, with white clover regularly being the minor component in the legume fraction. This natural adaptation process cannot be precisely predicted in terms of species composition, and it will happen randomly in future samples, which is why we did not attempt to analyze any changes in species composition of the legume or grass batch separately. Weeds were removed because the calibration was aimed at predicting sam-

Table 1. Harvesting dates of the calibration samples. The field names at the FAM Research Station Scheyern, Germany, are given. Within each plot (>25 m^2), six to nine replicates (0.25 m^2) were established. Samples taken were all sorted by hand into the legume and grass fractions.

Field	Cutting date				n_{plots}†	n_{repl}‡	$n_{spectra}$§
A04	26 May 1999	18 July 1999	28 Aug. 1999	28 Oct. 1999	2	6	96
A09	26 May 1999	18 July 1999	28 Aug. 1999	7 Oct. 1999	2	9	144
	26 May 1999	18 July 1999	28 Aug. 1999	7 Oct. 1999	1	6	48
A13	8 May 2000				7	4	56
	8 May 2000				9	1	18
							Σ362

† Number of plots in the field.
‡ Number of replicates per plot.
§ Theoretical number of legume plus grass samples (= 2 × number of cutting dates × n_{plots} × n_{repl}).

ples predominantly free of weeds, and they usually occur infrequently in legume–grass crops. The dried samples (60°C > 72 h) were ground in a shear mill (BRABENDER, Duisburg, Germany) to pass a 1.5-mm screen. On a few plots, not enough material was harvested so that the actual number of spectra does not coincide with the theoretical number of samples (Table 1). The artificially mixed samples were prepared from the pooled pure legume and grass fractions of August 1999. Twenty-one defined standards were mixed at 5% increments to continuously represent the range of 0 to 100% legume content.

One restriction to our calibration procedure may be due to the sample preparation. As outlined above, preparing the calibration samples with pure batches of legumes and grasses may create different particle size distributions compared with the field samples, which are ground as mixtures. The influence the grinding protocol may have on the prediction accuracy was tested with samples from two harvesting dates in 2002. Samples were collected at legume–grass experiments that were run at four different locations throughout Bavaria. Twenty-one fresh samples from multispecies mixtures were hand-separated and a subsample of them recombined to a defined legume content simulating real fresh mixtures. They were dried and ground as a mixture, which is principally in contrast to the standard sample-handling protocol used in this study. The legume content of these samples ranged between 18 to 82% on dry matter basis. The remaining pure samples were used to create a replicate set of samples in the same manner as the artificial calibration standards, i.e., weighted mixtures from ground material. The influence of the grinding protocol was evaluated by comparing the predicted legume content of these principally duplicate but differently prepared samples.

Near Infrared Reflectance Spectroscopy

Log (1/*R*) spectra were taken with a Fourier Transform NIR spectrometer (Vector 22/N, Bruker, Ettlingen, Germany) coupled to an external integration module. A rotating sample cup (9 cm diam.) was used to present the samples (>10 g) to the measurement area (2.0 cm diam.). A metal stamp (822 g) was put on top of the sample to ensure a comparable sample density and to avoid any influence of external light. Spectra from diffuse reflection were recorded by a PbS detector between 10000 and 3500 cm^{-1}, i.e., 1000 to 2857 nm (Fig. 1A). All samples (calibration and validation) were each measured once apart from the 21 artificial mixtures, which were measured three times with repacking before the second and third measurement.

Measurement conditions were tested to ensure a high signal/noise ratio and a high resolution at an acceptable measurement duration, which is reflected by the number of scans. A resolution of 10 cm^{-1} and 30 scans was found ideal to determine a spectrum. The instrument settings combined with a rotating sample cup resulted in a scanned area of 44 cm^2 for each sample.

Calibration Procedure

Multivariate calibration was performed with partial least squares regression. The spectral region used for calibration by a chemometrical software (OPUS 3.1, Bruker, Ettlingen, Germany) was generally restricted to the range from 7500 to 3950 cm^{-1} (1333–2532 nm) because of noise above and below this range. This resulted in 833 data points per spectrum used for calibration. Two calibration strategies were used as shown in Fig. 2: Strategy A = end-points calibration only with pure legume and pure grass samples; Strategy B was extended by artificial, incremental mixtures. Both strategies were analyzed by two kinds of calibration–validation procedures to find out

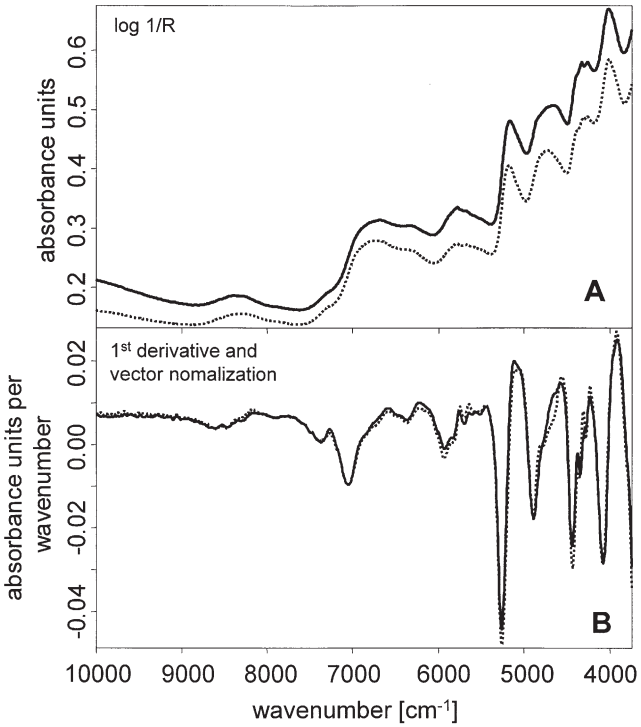

Fig. 1. Near infrared spectra of a legume sample (solid line) and a grass sample (dotted line) harvested in May 1999: (A) original absorbance spectra and (B) vector normalized first derivative spectra. The first derivative describes the slope of the original spectrum at a certain wave number. Therefore, the unit of the *y* axis in Fig. 1B is absorbance units per wave number.

the best model. The first kind was to take the whole data set from each strategy for calibration (A: 334 samples; B: 397 samples), resulting in Models A1 and B1. Each model developed during calibration was automatically tested by a one-leave-out cross-validation, i.e., each sample of the calibration set is estimated by a model based on all the other spectra of the data set. As the alternative, both data sets were split into roughly two halves, both designed to be as independent from each other as possible, i.e., samples of one sampling site were put either to calibration or validation. Both halves were once used for calibration and a second time as test set for validation resulting in two additional models for each strategy (A2, A3 and B2, B3). All models were developed during calibration by using an optimization routine offered by OPUS comparing various wave number regions and data pretreatments to determine the best calibration algorithm. Principally, the model with the lowest root mean square error of cross-validation (RMSECV) (Models A1 and B1) or root mean square error of prediction (RMSEP) (Models A2, A3, B2, and B3) was chosen, which in all cases was achieved by a combination of first derivative (Savitzky Golay algorithm, 17 smoothing points) and vector normalization. Then we removed spectral outliers, and the validation was repeated. Because the removal of outliers is a critical issue during model development—one could remove distant but valuable spectra (Shenk and Westerhaus, 1991)—this was only done once during each model development. Additionally, a one-leave-out cross-validation was run with the calibration samples of the Models A2, A3, B2 and B3 only to get comparative error figures to the Models A1 and B1 (but this is not necessary for model development). These data are reported along with the prediction errors (root mean square error, RMSE) calculated according to Eq. [1],

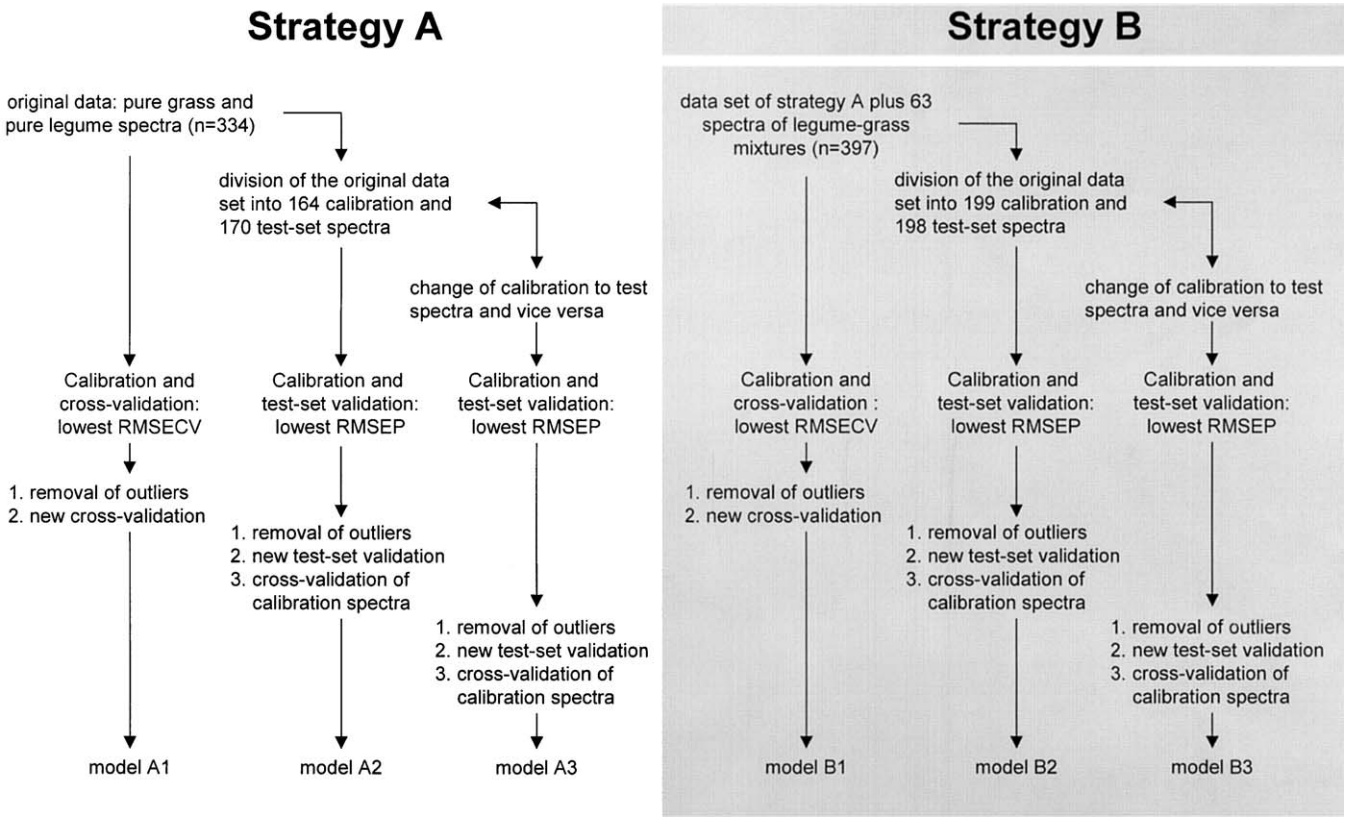

Fig. 2. Flow chart for the development of the different calibration algorithms for the determination of the legume content in multispecies mixtures. Strategy A results in calibrations derived only from grass mixtures (0%) and legume mixtures (100%). For Strategy B, artificially mixed standards were added. RMSECV, root mean square error of cross-validation; RMSEP, root mean square error of prediction (test-set validation).

[2], and [3] whether they are from calibration (C), cross-validation (CV), or test-set validation (P).

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_{C_i} - y_i)^2}{n}} \quad [1]$$

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_{CV_i} - y_i)^2}{n}} \quad [2]$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_{P_i} - y_i)^2}{n}} \quad [3]$$

where $\hat{y}_i$ = the NIRS predicted values, y_i = the true values, and n gives the number of tested samples.

RESULTS AND DISCUSSION

Prediction of Legume Content

The prediction error for the determination of legume content in multispecies legume–grass mixtures did not exceed 4% legume content for all models (Table 2). These error figures are in the same range as in the most similar studies published by Coleman et al. (1990) and Pitman et al. (1991). Compared with the variability found in field (in our study, it ranged from 14 to 98% legume content), the error is negligible. The minor change of the prediction errors from calibration to validation and from cross-validation to test-set validation (Table 2) underline that most of the variation found was accounted for in all models. Then, even the narrower calibrations (less than 200 samples) were sufficient for a good prediction of legume content, confirming the findings of Shaffer et al. (1990). But, a closer look at the data reveals some differences and effects that may be of relevance for future samples.

The systematic effects on bias and intercept found during the regression of the NIRS predicted on the true legume content (Table 2) are a hint that the test sets and calibration sets of the smaller models (A2, A3, B2, B3) were not thoroughly congruent. This was expected because we selected calibration and test sets to a maximum of independence, i.e., no site was present in both sets. But, species composition, plant age, and seasonal effects were randomly distributed across both sets. Obviously, this caused a slight decrease in the prediction power of the smaller models compared with Models A1 and B1. But, on the other hand, the small differences confirm the general applicability of the models. There

Table 2. Calibration and validation statistics of six near infrared reflectance spectroscopy (NIRS) models [A1, A2, A3, B1, B2, and B3 (q.v. Fig. 2)] to determine the legume content of multispecies legume–grass mixtures. Data pretreatment for each model: first derivative and vector normalization.

Model	Number of spectra for Calibration	Number of spectra for Validation	Spectral regions†	Selected data points‡	Factors§	RMSEC¶	RMSECV#	RMSEP††	R_C^2‡‡	R_V^2§§	Bias¶¶	Intercept##	Slope##
			cm^{-1}			legume content, %			%		legume content, %		
A1	334	320	7147–6433, 6082–5369, 5018–3950	650	9	2.2	2.4		99.8	99.8	0.0	0.1	0.99
A2	164	163	7147–5369, 5018–3950	740	9	2.4	2.7	3.3	99.8	99.5	−1.3	3.2	0.97
A3	170	162	7502–7143, 6792–5369, 5018–4659, 4308–3950	652	5	4.0	3.3	3.7	99.4	99.5	1.1	−1.7	1.01
B1	397	388	7502–6433, 6082–5369, 5018–3950	742	10	2.3	2.5		99.8	99.7	0.0	0.13	0.99
B2	199	192	7147–5369, 5018–4659, 4308–3950	650	8	2.5	2.6	3.2	99.7	99.5	−0.7	2.5	0.97
B3	198	193	7502–5369, 5018–4659, 4308–3950	742	8	2.5	2.6	2.9	99.7	99.6	0.8	−1.3	1.01

† Spectral regions that were found to give the best calibration.
‡ Number of data points in the model selected from 850 data points.
§ Number of factors needed for validation.
¶ RMSEC, root mean square error of calibration.
# RMSECV, root mean square error of cross-validation.
†† RMSEP, root mean square error of prediction.
‡‡ R_C^2, coefficient of determination of calibration; it usually refers to test-set validation, otherwise to cross-validation.
§§ R_V^2 , coefficient of determination of validation; it usually refers to test-set validation, otherwise to cross-validation.
¶¶ Bias is the mean difference between the true and NIRS predicted legume content values as derived from validation.
Intercept and slope are from the regression line of NIRS predicted on true legume content values as derived from validation.

is a striking difference in the number of factors used during calibration between Model A3 and all the other models (Table 2). This highlights two contrasting aspects. If several algorithms result in similar levels of error, it is recommended to choose the one with the lowest number of factors included in the algorithm (Naes et al., 2002). However, the slight increase in the prediction error of Model A3 could be a hint that not all variation was included to predict legume content. Further independent validation has to be done to judge the relevance of these differences. So far we conclude that, although there were no relevant differences between the models found, future samples will be better predicted by Models A1 and B1.

Comparison of Calibration Strategy

Martens and Naes (1987) and Brereton (2000) recommend a calibration design that covers the whole range of possible values. They argue that end-points calibrations (like Strategy A) assume linearity in the relationship between the spectra and the reference values. If this is not fulfilled, the prediction of intermediate samples would be erroneous. Our results are in line with others (Coleman et al., 1990; Pitman et al., 1991): There was no benefit observed in calibrating with a continuum of intermediate samples compared with an end-points calibration, whether judged by calibration or validation errors (Table 2). We tested this constraint by predicting the artificially mixed, intermediate samples with Model A1 (end-points calibration), i.e., the tested samples were not present in the calibration set. The relationship between the true values and the NIRS predicted values (Fig. 3) was almost unbiased, with a slope close to 1 and a negligible intercept. The prediction error as derived from this data set (RMSEP) was 2.3% legume content, which was no larger than RMSECV. Therefore, a linearity between spectra and legume content, as stated by Coleman et al. (1990) and Pitman et al. (1991), was confirmed. Based on this, the use of end-points calibrations seems to be justified for future model development.

Sample Preparation—Data Pretreatment

Particle size distribution strongly affects the spectra since NIRS mainly gathers information of solid samples from their surface. Even with grinding, particle size will vary and may lead to different scatter coefficients of grasses and legumes depending on, e.g., species composition and plant age. This could be a problem in the prediction of the legume content (Coleman et al., 1990) because legumes are ground to a finer degree than grasses even though the same screen size is applied (Petersen

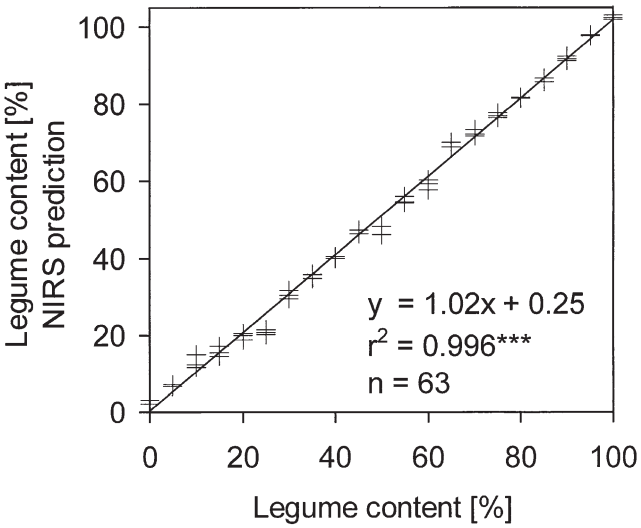

Fig. 3. Near infrared reflectance spectroscopy (NIRS) prediction of the legume content of artificially mixed samples using Model A1 (q.v. Fig. 2), which was developed based on an end-points calibration.

et al., 1987). Therefore, our standard procedure to prepare the calibration standards, i.e., mixing them from dry and ground legume and grass batches, may systematically affect particle size distribution compared with real samples, which will be ground as natural mixtures. To evaluate this effect, we predicted the legume content of duplicate mixtures where one replicate was prepared from fresh material, dried, and ground as a mixture and the other mixed from dry, ground material according to the standard procedure. We did not measure particle size distribution, but we assumed differences to occur as stated by Petersen et al. (1987). This was supported by visible differences observed in the occurrence of coarse particles in our samples. However, in the prediction of legume content of these duplicate samples (Fig. 4), there was no difference found. But, this does not exclude any influence of the partner in the mixture on the particle size distribution, which we assume did evolve from grinding, it only proves that it was of no relevance for the calibrations developed. There are two explanations why the calibration was robust to particle size effects. First, even the legume and grass batches themselves had wide variation in particle size because they represented a broad range of species composition and plant age. Second, the data pretreatment, a combination of first derivative and vector normalization (Fig. 1B), largely reduced the effects of scatter caused by different particle size and particle orientation. This data pretreatment is supported by other authors who found derivatives (Coleman et al., 1985; Deaville and Flinn, 2000; Moore et al., 1990; Petersen et al., 1987; Pitman et al., 1991) or normalization (Wachendorf et al., 1999) best.

During the preparation of the fresh mixtures, we found it difficult to ensure a high repeatability, which in turn will affect the quality of the reference data. Only 80% of the fresh prepared mixtures were within a limit of ±5% of their expected dry weight, which was explained by the difficulty of obtaining a representative sample from fresh, chopped plant material compared with dry, ground plant material. Therefore, our findings provide evidence that preparing standards from dry, ground material is superior to preparing them from fresh samples because it ensures high reference data quality.

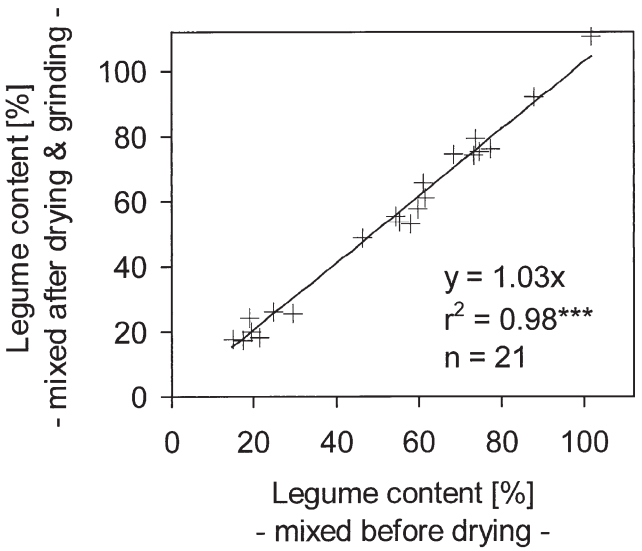

Fig. 4. Comparison of the predicted legume content of duplicate legume–grass mixtures where one replicate was prepared from fresh material, dried, and ground as a mixture and the other mixed from dry, ground material according to the standard procedure. For both sample sets, legume content was predicted using Model A1 (q.v. Fig. 2).

Can We Spectrally Define Legumes and Are the Calibrations Robust Enough for Broad Use?

The spectral regions selected during calibration represent relevant information that discriminates between grasses and legumes. If the models are similar in their predictive power, as was stated above, one could assume that the spectral regions common for all models do describe the relevant information to discriminate legumes from grasses. Actually, not more than 70% of the spectral regions were the same for all models. The coefficient of determination for prediction of legume content from the spectra was almost the same for any of the models as referred to their spectral range. If the models were restricted to the spectral regions common in all models (6792–6433, 6082–5369, 5018–4659, 4308–3950 cm^{-1} = 1472–1555, 1644–1863, 1993–2146, 2321–2532 nm), the prediction error slightly increased but did not exceed 4.5% legume content. In conclusion, this spectral range represented most of the variation necessary to discriminate between legume and grasses for the samples tested. From this point of view, the introduced calibrations do have a high potential for a broad use. However, it cannot be deduced from these data that the spectral range specifically describes a legume. Finally, it described the difference between grasses and legumes, which does quite likely refer to the same components (e.g., starch, protein, lignin), which are present in different relations and cellular structures.

Compared with the literature, the models already represent a broad variability (growing site, species composition, plant age), but the data can still be described as a closed population because the samples were collected on one farm. For the overall aim, the prediction of legume content in any multispecies legume–grass mixture grown under temperate climate conditions, further validation has to prove the ability of the models to accurately predict legume content. Then, it should be again questioned which calibration strategy is superior. In future samples, new sources of variability may affect the linearity stated so far and therefore demand a calibration design with intermediate samples. This important step of validation will be reported in a future paper.

CONCLUSIONS

Based on samples of multispecies clover–grass from real fields, a NIRS application was developed that predicted legume content at a high accuracy over a broad variation of species composition, plant age, and site conditions. The different calibration strategies compared did not lead to relevant differences. From these data, it can be concluded that less than 200 calibration samples for end-points calibration will create a reliable model. It

was shown that sample preparation was less important for prediction accuracy than the precision of the reference data used. The intersection of the spectral regions used in all models obviously described legume content quite well. But it cannot be concluded that it describes the spectral characteristics of legumes. However, further validations are necessary to confirm the general applicability of the models developed. Then, this method will strongly simplify the determination of legume content, which is a key parameter for the determination of N_2 fixation of legume–grass mixtures.

ACKNOWLEDGMENTS

We are grateful to the technical staff from the Research Station of Dürnast for harvesting and sample processing. Many thanks to the associate editor and the unknown reviewers for their most valuable suggestions, which helped to improve the paper. This research was part of the Agro-Ecosystem Research Network Munich (FAM) and was supported by the Federal Ministry of Research and Technology (BMBF 0339370I), Berlin, Germany. Rent and operating expenses are paid by the Bavarian State Ministry for Education and Culture, Science and Art.

REFERENCES

Androsoff, G.L., C. van Kessel, and D.J. Pennock. 1995. Landscape-scale estimates of dinitrogen fixation by *Pisum sativum* by nitrogen-15 natural abundance and enriched isotope dilution. Biol. Fertil. Soils 20:33–40.

Boller, B.C. 1988. Biologische Stickstoff-Fixierung von Weiß- und Rotklee unter Feldbedingungen. Landwirtsch. Schweiz 1:251–253.

Brereton, R.G. 2000. Introduction to multivariate calibration in analytical chemistry. Analyst (Cambridge, UK) 125:2125–2154.

Coleman, S.W., F.E. Barton II, and R.D. Meyer. 1985. The use of near-infrared reflectance spectroscopy to predict species composition of forage mixtures. Crop Sci. 25:834–837.

Coleman, S.W., S. Christiansen, and J.S. Shenk. 1990. Prediction of botanical composition using NIRS calibrations developed from botanically pure samples. Crop Sci. 30:202–207.

Dardenne, P., and N. Féménias. 1999. Diode array near infrared instrument to analyse fresh forages on a harvest machine. p. 121–124. *In* A.M.C. Davies and R. Giangiacomo (ed.) Near infrared spectroscopy. Proc. Int. Conf., 9th, Verona, Italy. 14–19 June 1999. NIR Publ., Chichester, UK.

Deaville, E.R., and P.C. Flinn. 2000. Near infrared (NIR) spectroscopy: An alternative approach for the estimation of forage quality and voluntary intake. p. 301–320. *In* D.I. Givens, E. Owen, R.F.E. Axford, and H.M. Omed (ed.) Forage evaluation in ruminant nutrition. CAB Int., Wallingford, UK.

Hansen, J.P., and F.P. Vinther. 2001. Spatial variability of symbiotic N_2 fixation in grass–white clover pastures estimated by the ^{15}N isotope dilution method and the natural ^{15}N abundance method. Plant Soil 230:257–266.

LaRue, T.A., and T.G. Patterson. 1981. How much nitrogen do legumes fix? Adv. Agron. 34:15–38.

Martens, H., and T. Naes. 1987. Multivariate calibration. John Wiley & Sons, New York.

Moore, K.J., C.A. Roberts, and J.O. Fritz. 1990. Indirect estimation of botanical composition of alfalfa–smooth bromegrass mixtures. Agron. J. 82:287–290.

Naes, T., T. Isaksson, T. Fearn, and T. Davies. 2002. A userfriendly guide to multivariate calibration and classification. NIR Publ., Chichester, UK.

Peoples, M.B., J.K. Ladha, and D.F. Herridge. 1995. Enhancing legume N_2 fixation through plant and soil management. Plant Soil 174:83–101.

Petersen, J.C., F.E. Barton II, W.R. Windham, and C.S. Hoveland. 1987. Botanical composition definition of tall fescue–white clover mixtures by near infrared reflectance spectroscopy. Crop Sci. 27: 1077–1080.

Pitman, W.D., C.K. Piacitelli, G.E. Aiken, and F.E. Barton II. 1991. Botanical composition of tropical grass–legume pastures estimated with near infrared reflectance spectroscopy. Agron. J. 83:103–107.

Shaffer, J.A., G.A. Jung, J.S. Shenk, and S.M. Abrams. 1990. Estimation of botanical composition in alfalfa/ryegrass mixtures by near infrared reflectance spectroscopy. Agron. J. 82:669–673.

Shenk, J.S., and M.O. Westerhaus. 1991. Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. Crop Sci. 31:469–474.

Stevenson, F.C., J.D. Knight, and C. van Kessel. 1995. Dinitrogen fixation in pea: Controls at the landscape and micro-scale. Soil Sci. Soc. Am. J. 59:1603–1611.

Wachendorf, M., B. Ingwersen, and F. Taube. 1999. Prediction of the clover content of red clover– and white clover–grass mixtures by near-infrared reflectance spectroscopy. Grass Forage Sci. 54:87–90.

Weißbach, F. 1995. Über die Schätzung des Beitrages der symbiontischen N_2–Fixierung durch Weißklee zur Stickstoffbilanz von Grünlandflächen. Landbauforsch. Voelkenrode 45:67–74.

Whalley, R.D.B., and M.B. Hardy. 2000. Measuring the botanical composition of grasslands. p. 67–102. *In* L. 't Mannetje and R.M. Jones (ed.) Field and laboratory methods for grassland and animal production research. CABI Publ., Wallingford, UK.