Grass and Forage Science
The Journal of the British Grassland Society | The Official Journal of the European Grassland Federation

Comparison of near infrared reflectance spectroscopy calibration strategies for the botanical composition of grass-clover mixtures

M. Cougnon*, C. Van Waes†, P. Dardenne‡, J. Baert§ and D. Reheul*

*Department of Plant Production, Gent University, Melle, Belgium, †ILVO Plant Crop Husbandry and Environment, Merelbeke, Belgium, ‡Valorisation of Agricultural Products Department, Walloon Agricultural Research Centre (CRA-W), Gembloux, Belgium, §ILVO Plant Applied Genetics and Breeding, Melle, Belgium

Abstract

Predicting species composition in mixed swards by near infrared reflectance spectroscopy (NIRS) can save labour in grassland research, provided equations are available. This study compares calibration strategies to predict species composition in swards with tall fescue (*Festuca arundinacea*), perennial ryegrass (*Lolium perenne*) and white clover (*Trifolium repens*). The compared calibration strategies were based on either real or artificial samples. Real samples are samples taken from multispecies swards; the species composition is known by hand separating; after separating, the original samples are recomposed. Artificial samples are samples obtained by mixing single species grown in pure stands in known proportions. The performance of the equation based on real samples was significantly better than the performance of the equation based on artificial samples. We hypothesized that the weak performance of the equation based on artificial samples was due to a lack of environmental variation in the spectra of the artificial samples. The hypothesis was supported by the good performance of a novel calibration strategy based on the spectra of the artificial samples with added variation. According to the obtained results, a calibration strategy based on few but diverse calibration samples is discussed.

Keywords: near infrared reflectance spectroscopy, botanical composition, *Festuca arundinacea* Schreb., *Trifolium repens* L., calibration

Correspondence to: M. Cougnon, Department of Plant Production, Gent University, Proefhoevestraat 22, 9090 Melle, Belgium.
E-mail: mathias.cougnon@ugent.be

Received 21 March 2012; revised 18 October 2012

Introduction

In research programmes dealing with mixtures of different forage species, the contribution of the different species in the yield is often of interest. Visual estimation of the botanical composition is a fast method, but is very subjective. Separation of a cut sample into the different constituents is very precise but is often too labour intensive for large field experiments. Near infrared reflectance spectroscopy (NIRS) can predict the botanical composition of a sample both quickly and precisely, provided an equation is available.

Several equations to predict the botanical composition of multi-species swards by NIRS have been developed in the past. These calibrations can be classified according to (i) the species and the number of different species for which the calibration was built, (ii) the magnitude of the calibration population and (iii) the calibration strategy used.

1 Mostly, equations were developed to discriminate between grass and one or more leguminous species. Discriminations have been made between a specific grass and a specific legume, e.g. tall fescue vs. white clover in Petersen *et al.* (1987); between non-specified grasses and either white or red clover in Wachendorf *et al.* (1999) or between functional groups e.g. grasses vs. leguminous species in Locher *et al.* (2005a). Equations which aim to discriminate between different grass species were built by Coleman *et al.* (1985) and Chataigner *et al.* (2010).
2 The objective of the calibration can be to predict the composition of a 'closed population', e.g. a well-defined trial with species mixtures (Surault *et al.*, 2006) or a 'open population', e.g. clover content in swards varying in phenological growth stages, in grass species and in nitrogen fertilization and climatic conditions (Wachendorf *et al.*, 1999).

doi: 10.1111/gfs.12031

3 Different strategies can be used to build a calibration according to the way in which the composition of the samples is determined. The most straightforward way to build a calibration is by separating samples (harvested from multispecies swards) into their different components, weighing the different species, recomposing the original samples, drying and grinding the samples and finally recording the spectra. Calibration samples obtained in this way are called 'real samples' (Shaffer *et al.*, 1990; Pitman *et al.*, 1991; Wachendorf *et al.*, 1999; Surault *et al.*, 2006). An alternative is the mixing of dried and ground components to provide a whole new set of samples differing in the proportions of their components. Samples obtained in this way are called 'artificial samples' (Coleman *et al.*, 1985; Pitman *et al.*, 1991; Locher *et al.*, 2005a; Surault *et al.*, 2006). The plant material used in this strategy is harvested in swards sown with a single species, which eliminates the labour-intensive hand separation.

An even simpler strategy is to collect the NIR spectra of dried and ground material of 'botanically pure samples' (i.e. samples taken from swards with a single species) and to use these spectra to estimate the composition of mixtures. This strategy is based on the assumption that spectra of mixtures can be obtained by linearly combining the spectra of the botanically pure samples. This strategy is called 'end point calibration' (Coleman *et al.*, 1990; Locher *et al.*, 2005a,b).

Pitman *et al.* (1991) compared the three strategies mentioned above: equations based on real, artificial and pure samples were built and validated with the same, independent set of actual pasture samples (which are real samples). Despite the excellent calibration statistics of the equations based on the artificial samples, the validation statistics were very poor. Using a calibration based on pure samples improved the results but validation statistics remained poor. Calibration based on real samples gave the best results. The comparison between the methods was not entirely justified, as the numbers of calibration samples and the origins of the material were not equal in the tested strategies. To date, there is no explanation as to why calibrations based on artificial samples and/or pure samples mostly fail to predict the composition of real pasture samples.

To disentangle this problem, we repeated the comparison made by Pitman *et al.* (1991) using fully comparable plant material and using equal numbers of samples with a fully comparable botanical composition. The equations were validated with the same set of real samples. In addition, we developed a calibration strategy, based on adding variation to the spectra of artificial samples. The rationale for this fourth strategy was based on our own observation that spectra of artificial samples contain less variation than spectra of real samples (Cougnon *et al.*, 2012)

The aim of our research was

1 To compare the performance of equations based on a same number of either real or artificial samples with the same botanical composition;
2 To explain the difference in performances in different calibration strategies;
3 To develop and assess a novel calibration strategy based on adding spectral variation to the spectra of artificial samples and/or botanical pure samples.

Materials and methods

Plant material

The plant material was collected from two trials comparing the yield and botanical composition of perennial ryegrass (*Lolium perenne* L.; Lp) or tall fescue (*Festuca arundinacea* Schreb.; Fa) with white clover (*Trifolium repens* L.; Tr) or mixtures of both grass species with white clover. Both trials were established in the spring of 2009 on a sandy-loam soil in Merelbeke, Belgium. The mixtures sown in both trials differed in the ploidy level of the ryegrass component: diploid (Lp2) or tetraploid (Lp4) and in the proportion of Lp in the sown mixture as indicated in Table 1. The two trials differed in the management, the number of replicates and in the plot size. The first trial was managed under a cutting-only management with five cuts per year; individual plot size was 7·8 m² and there were three replicates. In the second trial, apart from the spring growth which was cut, a grazing-only manage-

Table 1 Species composition of the mixtures sown in the trials.

	Sowing density (viable seeds per m^2)			
	Fa	**Lp2**	**Lp4**	**Tr**
2 grass species and clover				
1/4Lp2 + 3/4Fa	750	250	0	700
1/4Lp4 + 3/4Fa	750	0	250	700
1/8Lp2 + 3/8Fa	875	125	0	700
1/8Lp4 + 3/8Fa	875	0	125	700
1 grass species and clover				
Fa	1000	0	0	700
Lp2	0	1000	0	700
Lp4	0	0	1000	700

Fa, tall fescue; Lp2, diploid perennial ryegrass; Lp4, tetraploid perennial ryegrass; Tr, white clover.

ment was applied. Plot size in the second trial was 49 m² and there were four replicates.

At each cut (five cuts per year for the first trial; one cut per year for the second trial), a representative sample (150–300 g of fresh material) of the harvested material of each plot was separated into the different species and the fresh weights of the species were recorded. After the separation, mixtures containing two grass species (both Fa and Lp) were treated differently from mixtures containing only one grass species. The separated material from the plots with Fa + Lp + Tr was recomposed. The separated material from the plots with Fa + Tr or Lp + Tr was not recomposed, but pooled per species over the replications resulting in one pure sample for each of the four species (Fa, Lp2, Lp4 and Tr) per cut and per trial. Finally, all samples were dried (16 h, 75°C) and ground (Brabender shear mill) to pass a 1-mm sieve.

Results presented in this study are based on data of 2011 only. The origin and the construction of the samples are presented in Table 2. Each cut in both trials delivered one pure sample for each of the four single species. In addition, we collected 83 separated and recomposed samples with three constituent species: 60 samples came from the cutting trial (4 mixtures Fa + Lp + Tr * 3 replicates * 5 cuts); the spring cut in the grazing trial resulted in 23 samples. As the botanical composition within the plots was expected to be less homogenous due to the grazing management, two samples were taken per plot on two of the four blocks. As one sample was misplaced, we ended with 23 samples: 4 mixtures * 4 reps + (4 mixtures * 2 reps) − 1.

Four calibration sets

Four different sets of calibration samples were built with the available plant material (Figure 1). The first set consisted of the 83 hand-separated and recomposed samples; these samples are 'real samples' (Figure 1a). As hand separating is the reference method to determine the botanical composition, equations should always be validated with real samples. Therefore, the set of 83 real samples was split into a calibration set and a validation set. Half of the samples of each cut were assigned to the calibration set (42 samples) and the remaining 41 samples were assigned to the validation set.

A second calibration set consisted of artificial samples (Figure 1b): for each of the 42 real samples in the calibration set, an artificial sample with the same botanical composition was made, by physically mixing the ground material of pure samples in the proportions found in the real sample. As a result the number of samples and the botanical composition were identical in both the real and artificial calibration set.

A third calibration set (again with the same number of samples and identical botanical composition as the real calibration set) was built by making linear combinations of the spectra of the pure samples. This third calibration set therefore had no physical samples but was a set of spectra. The spectra were calculated as $(a_i X_{Fa,j} + b_i X_{Lp2,j} + c_i X_{Lp4,j} + d_i X_{Tr,j})$; a_i, b_i, c_i, d_i, being the proportion of Fa, Lp2, Lp4 and Tr respectively in the i-th sample of the 42 real samples and $X_{Fa,j}$ $X_{Lp2,j}$ $X_{Lp4,j}$ $X_{Tr,j}$ being the spectra of the four pure species in the j-th cut (Figure 1c). We called these linear combinations of spectra 'artificial spectra'.

Table 2 Number of samples used and their origin.

		Trial 1					Trial 2	
	Post-harvest treatment	**C1**	**C2**	**C3**	**C4**	**C5**	**C1**	**Total**
2 grass species and clover								
1/4Lp2 + 3/4Fa	Separated, recomposed,	3	3	3	3	3	6	21
1/4Lp4 + 3/4Fa	dried, ground	3	3	3	3	3	6	21
1/8Lp2 + 3/8Fa		3	3	3	3	3	5*	20
1/8Lp4 + 3/8Fa		3	3	3	3	3	6	21
Total		12	12	12	12	12	23	83
1 grass species and clover								
Fa	Separated, pooled per species	1	1	1	1	1	1	6
Lp2	over replicates, dried, ground	1	1	1	1	1	1	6
Lp4		1	1	1	1	1	1	6
Tr		1	1	1	1	1	1	6
Total		4	4	4	4	4	4	24

*One lost sample.C1–C5, first to fifth cut.

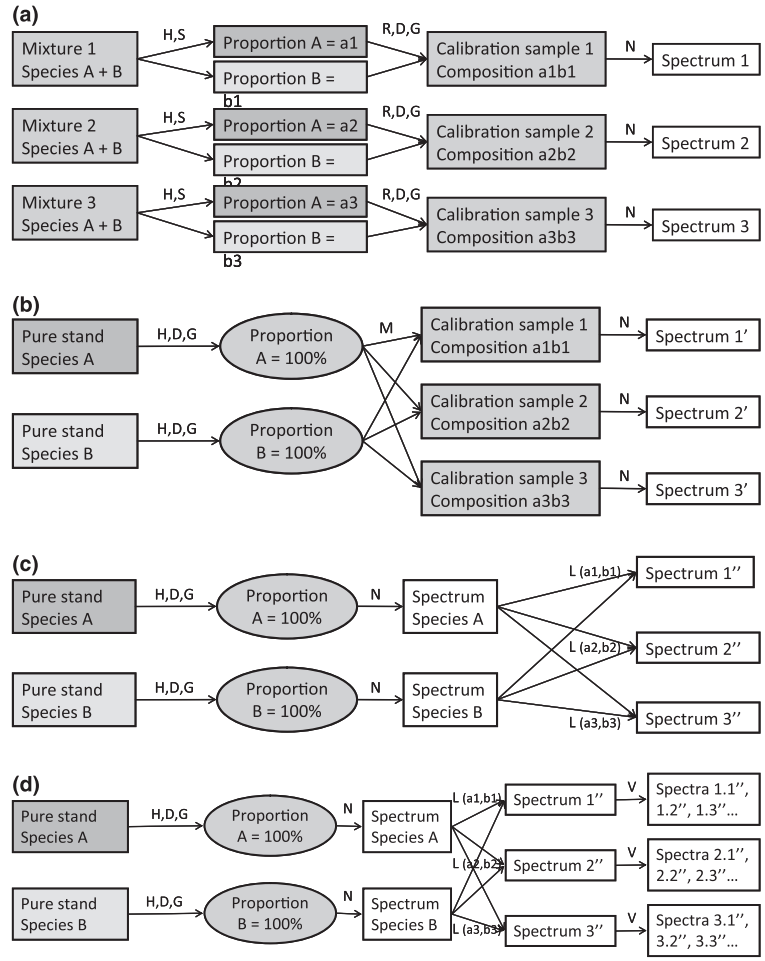

Figure 1 Calibration strategies for distinction of two species A and B based on: (a) real samples, (b) artificial samples, (c) artificial spectra, (d) artificial spectra with added variation. H, harvest; S, separate; R, recompose; D, dry; G, grind; N, collection of NIR spectrum; M, mixing powders in known proportions; L, making linear combinations of spectra with scalars (a, b); V, adding variation.

A fourth calibration set was obtained by adding variation to the spectra of the artificial samples (Figure 1d). The added variation was the difference found between spectra of real and artificial samples with the same botanical composition. Among the spectra of the real calibration samples, twelve spectra (two from each cut of the first trial and two from the single cut in the second trial) were chosen randomly and corrected by subtracting the spectra of the 12 corresponding artificial samples resulting in 12 spectra differences. These 12 spectra differences were added to each of the spectra of the 42 different artificial calibration samples, resulting in 504 (12 * 42) new spectra. This set of spectra was called 'artificial spectra with added variation'.

Collection of spectra and calibration

The spectra of the pure species and of the real and artificial samples were collected on the same day with a Foss NIRSystems 5000 (Foss, Hillerod, Denmark) and ISIscan 2.85.1 (Infrasoft, Port Mathilda, PA, USA) software. One quartercup was scanned per sample. The inverse reflectance [log(1/R)] was measured from 1100 to 2500 nm in steps of 2 nm, resulting in 700 datapoints per spectrum. All equations were calculated with WinISI II 1.5 (Infrasoft, Port Mathilda, PA, USA) using the partial least squares regression method. A data pre-treatment was performed before regression: the first derivative of the spectra was taken (with gap 4 and smoothing 4) and scatter correction using stan-

dard normal variate (SNV) and detrend was applied. Principal component analysis and calculation of Mahalanobis distance were performed in WinISI (Shenk and Westerhaus, 1991).

Linear combinations and adding variation

The linear combinations of the spectra of the botanically pure samples, needed to create the artificial samples, were calculated in R (R Development Core Team, 2011). The 24 (6 cuts * 4 species) spectra of botanically pure species were exported from WiniISI as a 24*700 matrix, which was imported in R using the *read.table()* function. This matrix was split into six 700*4 matrices, each containing the spectra of Fa, Lp2, Lp4 and Tr for one cut). The 42*4 matrix with the proportions of Fa, Lp2, Lp4 and Tr for each of the 42 real calibration samples was also imported. Using a loop, each row of the matrix with the proportions was multiplied with the matrix containing the pure spectra of the appropriate cut, resulting in a 42*700 matrix. This matrix was exported using the *write.table()* function and imported in WinISI.

The addition of spectral variation to each of the artificial samples, creating the artificial spectra with added variation, was also performed in R. A 24*700 matrix including the spectra of the 12 selected real samples and the 12 corresponding artificial samples and the 42*700 matrix containing the spectra of the 42 artificial samples were imported in R using the *read.table()* function. The 12 spectra of the artificial samples were subtracted from the 12 corresponding real samples, resulting in a 12*700 matrix containing the spectral differences between real and artificial samples. Using a loop, these 12 spectral differences were added to each of the 42 spectra of the artificial samples, resulting in a 504 (12*42) new spectra of artificial samples with added variation. This 504*700 matrix was exported using the *write.table()* function.

Comparison of the models

For each of the calibration sets, an equation for the prediction of the proportion Fa, Lp and Tr in the dry matter was built (Equations 1 to 4) (Table 3). Standard errors of calibration (SEC), standard errors of cross-validation (SECV) and the determination coefficient of the linear regression between the predicted and the reference values (RSQ) were calculated for the four equations. The four equations were validated with the validation set of 41 real samples. Root mean square error of prediction (RMSEP), bias and the ratio of prediction to deviation (RPD), which is the ratio of the standard deviation of the reference values to the RMSEP, were calculated (Shenk and Westerhaus, 1993). For each of the spectra of the real calibration samples, artificial calibration samples and artificial calibration spectra, the Neighbourhood H (NH) value, which is the Mahalanobis distance of a spectrum to its closest neighbour, was calculated (Shenk and Westerhaus, 1993).

Table 3 Calibration sets on which different equations were based.

Equation	Calibration set
1	42 real samples
2	42 artificial samples
3	42 artificial spectra
4	504 artificial spectra with added variation

The different calibration strategies were compared statistically by calculating confidence intervals for the ratios of the standard errors of the prediction errors (SEP), taking into account the paired nature of the prediction errors (Fearn, 1996; Naes *et al.*, 2002). Confidence intervals for the ratios of the SEPs obtained with the equation based on real samples (Equation 1) and each of the three other equations were calculated. If a confidence interval excluded 1, the SEPs were different at the 5% significance level, and the results obtained with the two methods were considered as different.

Results and discussion

As the calibration strategies used to build Equations 1, 2 and 3 were based on an equal amount of calibration points with the same composition, a direct comparison of these strategies was appropriate. Both the calibration sets and the validation set contained samples with a very diverse composition (Table 4). The validation set used to compare the different calibration strategies was a subset of our real samples. Hence, the validation set was not completely independent. The results shown here are only valuable in the closed population of samples of the concerned trials. This is not problematic as it was not the aim of this experiment to present the performance of the calibrations, but merely to compare different strategies to develop calibrations. The same principles would apply to a larger data set, validated with a real, independent data set (e.g. plant material harvested on a different site).

Real samples vs. artificial samples and artificial spectra

The SEC was higher for the equation based on the real samples (Equation 1) compared to those for the calibrations based on artificial samples (Equation 2) or artificial spectra (Equation 3) (Table 5). The RMSEP,

Table 4 Distribution of the botanical composition (% by dry weight) of the samples in the calibration sets and in the common validation set.

	Calibration set				Validation set			
	Mean	**Stdev**	**Min**	**Max**	**Mean**	**Stdev**	**Min**	**Max**
Fa	49·3	16·9	18·4	89·0	53·5	15·2	26·5	93·9
Lp	32·4	11·8	11·0	60·1	28·2	13·8	5·7	66·9
Tr	18·2	11·7	0·0	41·3	18·3	10·9	0·4	41·1

Table 5 Calibration and validation statistics of four equations for the prediction of the botanical composition of forage samples.

	Calibration			Validation		
	SEC	**RSQ**	**SECV**	**RMSEP**	**Bias**	**RPD**
Equation 1						
Fa	3·7	0·92	6·1	5·4	−1·2	2·8
Lp	5·0	0·87	7·8	6·1	1·4	2·3
Tr	2·5	0·95	3·9	2·9	0·4	3·8
Equation 2						
Fa	3·6	0·94	4·9	15·6	−9·1	1·0
Lp	3·3	0·94	5·0	15·9	7·4	0·9
Tr	1·4	0·98	1·9	6·8	0·7	1·6
Equation 3						
Fa	1·8	0·98	2·3	14·3	0·6	1·1
Lp	1·8	0·98	2·2	21·7	−6·8	0·6
Tr	0·7	0·99	0·9	13·9	8·0	0·8
Equation 4						
Fa	1·9	0·98	2·4	6·0	0·3	2·5
Lp	1·9	0·98	2·4	7·7	0·5	1·8
Tr	0·6	0·99	0·8	3·5	−0·7	3·1

SEC, standard error of calibration (%); SECV, standard error of cross-validation (%); RSQ, R squared value; RMSEP root mean square error of prediction (%); Bias, average difference between the predicted values and the reference value of the validation samples; RPD, ratio of standard deviation of validation samples and RMSEP; Fa, tall fescue; Lp, perennial ryegrass; Tr, white clover.

on the other hand, was lower for Equation 1 compared to Equations 2 and 3 for the three species involved. For the three equations, Tr was the species which was best predicted (the lowest RMSEP, highest RPD) (Table 5), which was no surprise as the difference in chemical composition between clover and grass is higher than between two grass species.

Equation 1 had an RPD higher than 2 for Fa, Lp and Tr; for Equations 2 and 3, the RPD was below 2 for the three species. As an RPD of 2 is generally regarded as the minimum for a suitable calibration, only Equation 1 was considered as suitable to predict the composition of the validation set (Table 5).

Confidence intervals for the ratio of the SEPs obtained with Equations 1 and 2 were [0·35, 0·60], [0·41, 0·68] and [0·26, 0·48] for Fa, Lp and Tr respectively. For Equations 1 and 3, the confidence intervals of the ratio of the SEPs were [0·33, 0·61], [0·26, 0·49] and [0·20, 0·36] for Fa, Lp and Tr respectively. This means that the standard deviations of the prediction errors obtained with the calibrations based on artificial samples or on artificial spectra were significantly higher for the three species. Hence, the equation based on real samples predicted the botanical compositions of the samples in the validation set significantly better.

The different performances of the equations based on artificial samples and artificial spectra on the one hand and the real samples on the other hand can be understood from a principal component analysis (Figure 2). Two clear trends can be observed. First, the scores of the artificial samples and the artificial spectra almost overlapped, indicating that the spectral information contained in both types of spectra was very similar. This was confirmed by the study of the NH distances, the Mahalanobis distance between a sample

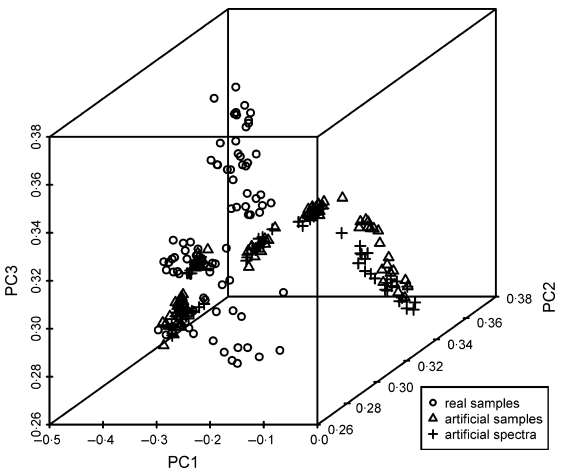

Figure 2 Scores of spectra of real and artificial samples and of artificial spectra plotted in their first three principal components.

Table 6 Neighbourhood H values (means and standard deviation) of real samples, artificial samples and artificial spectra calculated in their principal components.

	Trial 1					Trial 2	
	C1	C2	C3	C4	C5	C1	Mean
Real samples							
Mean	0·096	0·110	0·091	0·086	0·122	0·185	0·126
Stdev	0·061	0·040	0·048	0·041	0·050	0·120	0·086
Artificial samples							
Mean	0·013	0·011	0·024	0·028	0·074	0·025	0·028
Stdev	0·065	0·007	0·009	0·017	0·060	0·019	0·031
Artificial spectra							
Mean	0·014	0·010	0·025	0·023	0·072	0·027	0·028
Stdev	0·021	0·008	0·008	0·023	0·052	0·021	0·031

C1–C5, first to fifth cut.

and its closest neighbour. Averaged over all cuts, the NH values between the artificial samples and the artificial spectra were almost equal (Table 6).

Second, the real samples occur in larger, wider clouds than the artificial samples or artificial spectra. The mean NH distance was at least twice as high for the real samples compared to the artificial samples. The standard deviation on the NH distances between the points was also larger for the real samples, except for the fifth cut of the first trial. Averaged over all cuts, the mean NH distance between the real samples was four times higher than the distance between artificial samples; standard variation was twice as high.

We can conclude that less variation is present in the spectra of artificial samples and artificial spectra compared to real samples even though the botanical composition of both types of samples was the same. The variation present in the validation samples is not spanned by the cloud of artificial samples which explains the weak performance of Equation 2. To test this hypothesis, we tried to add relevant variation to the artificial samples and compared the performance of the resulting equation with the equation based on real samples.

Artificial spectra with added variation

Adding relevant spectral variation to the artificial samples was expected to enhance the prediction ability of the calibrations based on these samples. Indeed, the equation based on the artificial spectra with added variation (Equation 4) performed well: RMSEPs were on the same levels as these obtained with the real samples (Table 5) and the RPD's were higher than 2 for Fa and Tr. Confidence intervals for the ratio of the standard deviations of the prediction errors of Equations 1 and 4 were [0·82, 1·39], [0·76, 1·24] and [0·69, 1·16] for Fa, Lp and Tr respectively, which means that there was no significant difference in the performance of the equation based on real samples and the equation based on artificial spectra. The values of the artificial spectra with added variation spanned the whole space filled by the real (validation) samples in the first three principal components (Figure 3).

From the observations above, we concluded that the weak performance of Equations 2 and 3 was based on a lack of variation in the spectra of the artificial samples and artificial spectra. As the variation in botanical composition in both sets of samples was exactly the same, the difference has to be explained by environmental variation. However, the main

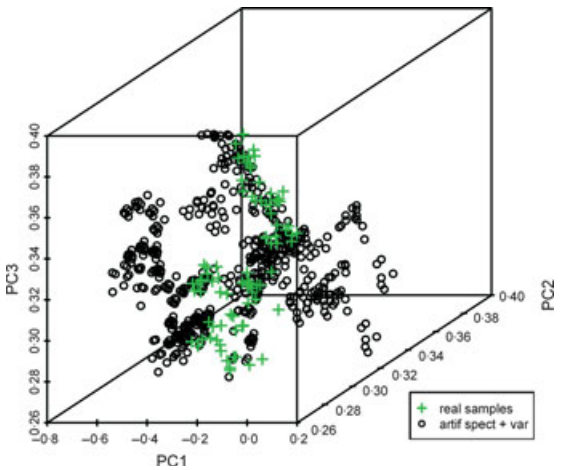

Figure 3 Scores of the artificial spectra with added variation (artif. spect. + var.) plotted in the first three principal components of the spectra of the real and the artificial samples.

sources of environmental variation (different cuts, different trials, ploidy of the perennial ryegrass, drying, grinding) were identical for both sets. The only source of environmental variation present in the real samples and absent in the artificial samples was the variation caused by different plots. As indicated above, the pure material from the plots sown with only one grass species and clover was pooled over the replicates of the same cut, resulting in one pure sample per cut for each species. By acting like this, we diminished the labour which was one of the prerequisites of our research. We did not expect this to have any effect on the calibration, as the trials were very uniform, and the variation coefficients for dry matter yield, crude protein content and digestibility of the organic material were very low (data not shown). The difference between the real and the artificial samples may find its origin in the post-harvest handling and behaviour of the plant material. Different authors (Coleman *et al.*, 1990; Locher *et al.*, 2005a) suggest that there is a difference in particle size distribution whether a specific species was ground solely or in a mixture, but these effects are believed to be largely eliminated by the mathematical treatment applied on the spectra. Another source of variation between the real and artificial samples might by due to the interaction of species growing together in a mixture compared to species growing in a pure stand. A similar suggestion was made by Surault *et al.* (2006), who explained the difference between real and artificial samples by stating that the spectral signatures of grasses growing in pure stands and mixtures are different. This interaction effect is especially true when one of the species is clearly dominant. Indeed, there is a clear morphological difference between a species growing in a mixture where it is dominated by another species, and the same species growing in a pure stand. Whether the spectra of species growing in pure stands or in mixtures are different remains to be proven.

Whatever the reason or origin may be, it is remarkable that apparently small differences in environmental variation in the calibration samples have a huge influence on the resulting equations.

One way to increase the environmental variation in the artificial samples is to harvest pure material from more plots as apparently a pure sample for each species in the real samples is needed to obtain the same amount of environmental variation in the artificial samples as in the real samples.

Surault *et al.* (2006) noticed that broadening a calibration based on artificial samples with some real samples greatly improved the performance of their calibration. The strategy we used to improve the performance of the artificial samples was to increase the variation in the spectra of the artificial samples by adding the variation present in the spectra between real and artificial samples with an identical botanical composition. No new reference values were added, just differences between spectra. The number of spectra in the fourth calibration strategy was inflated a lot, but the number of reference values did not change.

Implications for practise

Although good calibrations for botanical composition based on botanical pure samples were described (Locher *et al.*, 2005a,b), our findings suggest there is more advantage in taking time and effort to use real samples, which represent all the variables that would affect the NIRS spectra, rather than creating artificial samples. The labour required to separate samples can be reduced by following a strategy in which relevant environmental variation is added to the available reference values (Fernández Pierna *et al.*, 2010). By scanning a restricted set of samples for example on different days (with different temperature and humidity), the variation due to different scanning circumstances can be added to all the reference samples. The variation present between the spectra of a particular species grown as a single species in different locations or managed under different circumstances is another source of variation that may be added to make a calibration more robust for botanical composition.

In some cases, the use of a calibration strategy based on real samples is not possible (e.g. because the species for which the calibration is built are very difficult to separate). In that case, a calibration strategy based on botanically pure samples can be used keeping some important points in mind.

There is not much gain in mixing physically only a few botanically pure samples to obtain series of artificial samples with a whole range of compositions. Our results and results of Locher *et al.* (2005a) and Coleman *et al.* (1990) and Pitman *et al.* (1991) indicate that acting like this merely creates linear combinations of spectra of pure samples. It is the environmental variation in the pure samples that is of the greatest importance, and environmental variation can be added to the spectra using a strategy as presented in this article.

Conclusion

The performance of equations for the prediction of botanical composition based on artificial samples or artificial spectra was not satisfactory. This weak performance could be explained by a lack of environmental variation in the artificial samples relative to the real samples. Adding the variation among artificial and real samples to the artificial calibration samples resulted in

a calibration which performed nearly as well as the calibration based on real samples. Based on this experience, we recommend a calibration strategy based on diverse hand-sorted samples, rather than making a lot of artificial samples that contain relatively little spectral information. Adding environmental variation to the spectra of the calibration samples allows obtaining a robust calibration with a minimum of hand-separated samples.

References

Chataigner F., Surault F., Huyghe C. and Jullier B. (2010) Determination of botanical composition in multispecies forage mixtures by near infrared reflectance spectroscopy. In: Huyghe C. (ed.) *Sustainable use of genetic diversity in forage and turf breeding*, pp. 199–203. Dodrecht, The Netherlands: Springer.

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

Coleman S.W., Christiansen S. and Shenk J.S. (1990) Prediction of botanical composition using nirs calibrations developed from botanically pure samples. *Crop Science*, **30**, 202–207.

Cougnon M., Van Waes C., Baert J. and Reheul D. (2012) NIRS calibration strategies for the species composition of grass clover mixtures. In: Barth S. and Milbourne D. (eds.) *Breeding strategies for sustainable forage and turf grass improvement*, pp. 371–375. Dodrecht, The Netherlands: Springer.

Fearn T. (1996) Comparing standard deviations. *NIRnews*, **7**, 5–6.

Fernández Pierna J., Chauchard F., Preys S., Roger J., Galtier O., Baeten V. and Dardenne P. (2010) How to build a robust model against perturbation factors with only a few reference values: a chemometric challenge at 'Chimiométrie 2007'. *Chemometrics and Intelligent Laboratory Systems*, **106**, 152–159.

Locher F., Heuwinkel H., Gutser R. and Schmidhalter U. (2005a) Development of near infrared reflectance spectroscopy calibrations to estimate legume content of multispecies legume-grass mixtures. *Agronomy Journal*, **97**, 11–17.

Locher F., Heuwinkel H., Gutser R. and Schmidhalter U. (2005b) The legume content in multispecies mixtures as estimated with near infrared reflectance spectroscopy: method validation. *Agronomy Journal*, **97**, 18–25.

Naes T., Isaksson T., Fearn T. and Davies T. (2002) *A user-friendly guide to multivariate calibration and classification*. Chichester, UK: NIR Publications.

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

Pitman W.D., Piacitelli C.K., Aiken G.E. and Barton F.E. (1991) Botanical composition of tropical grass-legume pastures estimated with near-infrared reflectance spectroscopy. *Agronomy Journal*, **83**, 103–107.

R Development Core Team (2011) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Shaffer J.A., Jung G.A., Shenk J.S. and Abrams S.M. (1990) Estimation of botanical composition in alfalfa ryegrass mixtures by near-infrared spectroscopy. *Agronomy Journal*, **82**, 669–673.

Shenk J.S. and Westerhaus M.O. (1991) Population structuring of near-infrared spectra and modified partial least-squares regression. *Crop Science*, **31**, 1548–1555.

Shenk J. and Westerhaus M. (1993) *Analysis of agriculture and food products by near infrared reflectance spectroscopy*. Port Matilda, PA, USA: Infrasoft International.

Surault F., Briand M., Veron R. and Huyghe C. (2006) NIRS calibration equations to determine species contribution in grassland swards. In: Lloveras J. (ed.) *Grassland science in Europe*, Vol. **11**, pp. 598–600.

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