

SCIENTIFIC REPORTS

OPEN

Fine root lignin content is well predictable with near-infrared spectroscopy

Oliver Elle^{1,4}, Ronny Richter^{1,2,3}, Michael Vohland^{2,3} & Alexandra Weigelt^{1,3}

1. Root lignin is a key driver of root decomposition, which in turn is a fundamental component of the terrestrial carbon cycle and increasingly in the focus of ecologists and global climate change research. However, measuring lignin content is labor-intensive and therefore not well-suited to handle the large sample sizes of most ecological studies. To overcome this bottleneck, we explored the applicability of high-throughput near infrared spectroscopy (NIRS) measurements to predict fine root lignin content. **2.** We measured fine root lignin content in 73 plots of a field biodiversity experiment containing a pool of 60 grassland species using the Acetylbromid (AcBr) method. To predict lignin content, we established NIRS calibration and prediction models based on partial least square regression (PLSR) resulting in moderate prediction accuracies (RPD = 1.96, $R^2 = 0.74$, RMSE = 3.79). **3.** In a second step, we combined PLSR with spectral variable selection. This considerably improved model performance (RPD = 2.67, $R^2 = 0.86$, RMSE = 2.78) and enabled us to identify chemically meaningful wavelength regions for lignin prediction. **4.** We identified 38 case studies in a literature survey and quantified median model performance parameters from these studies as a benchmark for our results. Our results show that the combination Acetylbromid extracted lignin and NIR spectroscopy is well suited for the rapid analysis of root lignin contents in herbaceous plant species even if the amount of sample is limited.

Soils are a major sink of terrestrial carbon (C)¹ primarily via plant litter. Litter decomposition thus contributes a substantial amount of C to the fluxes in soils and from soil to atmosphere and represents an important link between plant productivity and C stocks in the soil². For grassland ecosystems, up to 70% of plant biomass is stored in the roots^{3,4}, producing litter with closer contact to soil particles and longer residence times in soils than other plant tissues⁵. Thus, especially in grasslands, root decomposition constitutes a major C source^{2,6}. Root lignin is one important driver of root decomposition⁷ but lignin quantification is labor-intensive especially for large sample sizes as often used in ecological studies. Here, we present a method to predict fine root lignin content from high-throughput near infrared spectroscopy (NIRS) measurements to overcome this bottleneck.

Root decomposition is driven by three major processes: chemical litter composition, soil biota and soil abiotic environment^{8,9}. In grasslands, root chemical composition and specifically lignin and the carbon to nitrogen ratio are the major drivers for root decomposition^{10–12}. High contents of holocellulose and nitrogen are positively correlated whereas a high lignin content is negatively correlated with root decomposition^{13–16}. Lignin is the second most abundant biopolymer on Earth¹⁷ and represents a class of aromatic macromolecules characterized by complex, three-dimensional structures, which increase the stability of cell walls¹⁸. Lignification of plant tissues therefore increases their recalcitrance against chemical, biotic and abiotic degradation.

There is a multitude of methods for lignin determination¹⁹. The most common method of gravimetric determination of an acid insoluble residue (Klason lignin²⁰) or the related acid detergent lignin method (van Soest method²¹) both require sample volumes of 500 mg to 5 g for accurate results²². For smaller sample volumes such as herbaceous samples or fine roots (usually ≤ 10 mg), the acetyl bromide (AcBr) extraction offers an alternative approach to make lignin soluble for quantitative UV absorption measurement^{23–25}. However, both methods are

¹Systematic Botany and Functional Biodiversity, Institute of Biology, Leipzig University, Johannisallee 21–23, D-04103, Leipzig, Germany. ²Geoinformatics and Remote Sensing, Institute for Geography, Leipzig University, Johannisallee 19a, D-04103, Leipzig, Germany. ³German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, D-04103, Leipzig, Germany. ⁴Macromolecular Chemistry and Paper Chemistry, Ernst-Berl-Institute, Technische Universität Darmstadt, Alarich-Weiss-Straße 8, D-64287, Darmstadt, Germany. Oliver Elle and Ronny Richter contributed equally. Correspondence and requests for materials should be addressed to R.R. (email: ronny.richter@idiv.de)

labor-intensive and not well-suited to handle large numbers of samples as commonly examined in ecological studies.

Ecological models, by their very nature, have to consider a high biological variation between species, taxa, sites or experimental treatments. Near infrared spectroscopy (NIRS) is a fast and economic approach to predict chemical compounds in such large sample sizes. NIRS has been used successfully for a variety of compounds and materials i.e. chemical and biological properties of soils²⁶, non-structural carbohydrates in plant tissues²⁷, cellulose in paper based materials²⁸ as well as for lignin in plant materials²⁹. NIRS is based on the absorption of photon energy ranging from 800 nm–2500 nm and the excitation of molecular overtone and combined vibrations from mainly hydrogen containing chemical groups^{30,31}. NIR spectra of lignocellulosic materials reveal complex absorption patterns due to the high chemical diversity of plant material including that of lignin³⁰.

Chemometric methods are essential to model the relationship between NIR spectra and measured analytical constituents³⁰. The most widely used chemometric method is partial least squares regression (PLSR)^{32,33} which uses the complete NIR spectra to define latent variables as combinations of the original spectral variables. However, this regression technique does not eliminate the problem of uninformative or even noisy spectral data that may affect the modelling approach. Therefore, combining methods which select the most informative spectral variables with the PLS regression often provides models with a greater predictive power³⁴. There are several comprehensive overviews of such variable selection techniques for PLS models^{34–36}. In principle, these methods can be subdivided into two groups: (1) methods considering only separate effects of the introduced statistical features and (2) methods considering the interaction of variables in the search space (which is often linked to a more exhaustive search compared to the approaches of the first group)³⁶. The competitive adapted reweighted sampling (CARS) algorithm³⁷ belongs to the first group of methods. It often provides more accurate estimates than full spectrum-PLSR³⁸. CARS algorithm was successfully benchmarked against other state of the art variable selection techniques in regression tasks such as genetic algorithm (GA)³⁹, successive projections algorithm (SPA) and iteratively retaining informative variables (IRIV)³⁸. CARS algorithm works in a fast and computationally efficient way even for a great number of spectral bands^{36,38}. Furthermore, CARS-PLS-DA models were significantly superior in classifying hyperspectral images as compared to non-parametric support vector machines (SVM) and random forest (RF)⁴⁰. Moreover, the identity of wavelengths selected with CARS algorithm in a final model might inform the physical interpretation of results due to the link to specific chemical groups or bindings^{37,41,42}. In addition to spectral variable selection, the pre-processing of NIR spectra is an integral part of chemometric modeling⁴³. Rinnan *et al.*⁴³ systematically compared a wide range of different pre-processing methods for moisture and sugar contents in marzipan samples. According to their results, the choice of the best pre-processing technique could improve model prediction up to 25% relative to unprocessed NIR spectra.

To put our results in a wider methodological and statistical perspective we performed a literature survey for studies predicting lignin content via spectroscopic measurements. We listed studies (1) measuring chemical content of lignin and using spectroscopy on plant tissue samples, (2) reporting the set of species used for lignin extraction, (3) using PLSR or other multivariate calibration methods to predict lignin contents spectrally and (4) validating the model on independent test data. We identified 30 relevant studies, which predicted lignin content via spectroscopic measurement following our four prerequisites. These studies reveal that NIRS has been used to predict Klason lignin or comparable methods of lignin analysis in wood^{44,45} or leaves^{46,47} of mainly silvicultural species such as pine, birch or eucalyptus. In addition, a number of studies measured lignin in agriculturally relevant species such as rice, corn, elephant grass or bamboo^{29,48–50}. However, we did not find any study evaluating the prediction of acetyl bromide extracted lignin with NIRS nor a study trying to predict lignin in root material.

Therefore, the goals of this study are (i) to test, for the first time, if NIR based models can predict AcBr extracted fine root lignin content, and (ii) to explore whether it is possible to predict root lignin content in mixtures of up to 60 different herbaceous grassland species. Our chemometric approach to link lignin content to NIR spectra uses two steps. First, we evaluate the benefits of different spectral pre-processing techniques and second we compare the performance of model prediction using either full spectra or spectral variable selection (with the CARS algorithm). As plant material, we used biomass of fine roots (i.e. <2 mm in diameter) sampled from different plant communities of the Jena Experiment, a large grassland biodiversity experiment in Thuringia, Germany.

Results

AcBr based fine root lignin content. We found a lignin content of $20.52 \pm 7.7\%$ (mean \pm SD) in the fine roots of mixed grassland communities of the Jena Experiment. The 74 samples spanned a species gradient of 1–16 species from a pool of 61 species with a minimum lignin content of 7.7% and a maximum of 42.8%.

The RSEL of $\pm 3.61\%$ is within the range of other studies reporting RSEL for AcBr based lignin determinations^{51,52}.

Spectral Pre-processing. We evaluated the effect of six different methods of spectral pre-processing on the accuracies of full spectrum PLSR in 100 repeated independent model validations (Fig. 1). We based our model evaluation on root mean square error of prediction (RMSEP) and residual predictive deviation (RPD) (Table 1). Standard normal variate (SNV) pre-processing achieved the overall highest model accuracies with highest RPD and lowest RMSEP values. This pre-processing was therefore selected for further modelling in block II (Fig. 1). Overall accuracies were lowest for multiplicative scatter correction (MSC) yet the high BIAS and RMSEP values indicate a strong offset between predicted and observed lignin content. Irrespective of spectral pre-processing, all full spectrum PLSR-models showed RPD values below 2 and were therefore not promising to accurately predict lignin in quantitative analyses.

Variable selection. The variable selection procedure based on CARS-PLSR significantly improved the accuracies in independent model validation to RPD values above 2.5 (Table 2). Starting from the full spectrum,

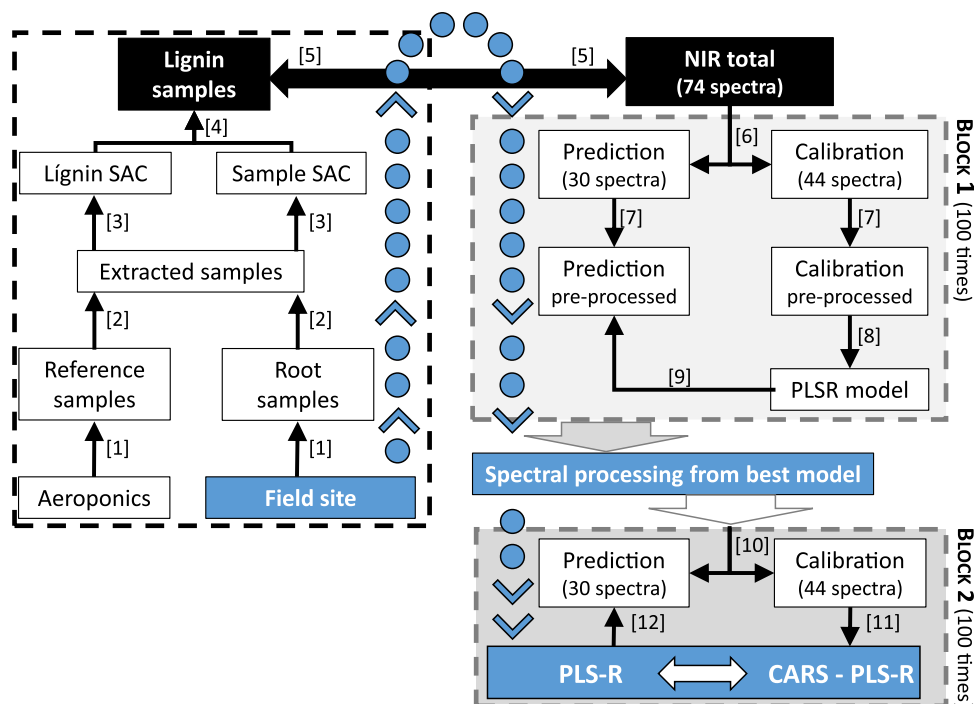


Figure 1. Schematic depiction of the workflow of chemical analysis (left box with steps 1–4 indicated in square brackets), spectral analysis (outlined in black boxes and step 5) and statistical analysis. The latter is subdivided in spectral pre-processing (block 1 with steps 6–9) and final analysis (block 2 with steps 10–12). See text for further details.

| Method | nLV | RPD | R_p^2 | RMSEP [%] | SEP [%] | BIAS [%] |
|------------|--------------------|--------------------|--------------------|--------------------|--------------------|---------------------|
| Raw | 8.17 ± 0.75 | 1.83 ± 0.32 | 0.71 ± 0.08 | 4.11 ± 0.62 | 4.05 ± 0.59 | -0.43 ± 0.95 |
| B.als | 6.62 ± 1.48 | 1.73 ± 0.27 | 0.69 ± 0.07 | 4.34 ± 0.64 | 4.27 ± 0.59 | -0.44 ± 1.05 |
| B.irls | 7.96 ± 1.40 | 1.69 ± 0.30 | 0.68 ± 0.09 | 4.47 ± 0.69 | 4.40 ± 0.65 | -0.49 ± 1.04 |
| MSC | 7.34 ± 1.61 | 1.95 ± 0.37 | 0.74 ± 0.09 | 11.64 ± 8.74 | 3.84 ± 0.64 | -1.83 ± 13.97 |
| SNV | 7.37 ± 1.28 | 1.96 ± 0.35 | 0.74 ± 0.08 | 3.83 ± 0.57 | 3.79 ± 0.56 | -0.22 ± 0.86 |
| D1 | 7.95 ± 2.08 | 1.84 ± 0.39 | 0.71 ± 0.10 | 4.14 ± 0.73 | 4.09 ± 0.70 | -0.35 ± 0.98 |
| D2 | 5.04 ± 2.31 | 1.56 ± 0.17 | 0.58 ± 0.09 | 4.77 ± 0.64 | 4.71 ± 0.59 | -0.02 ± 1.18 |

Table 1. Accuracies for different pre-processing methods achieved in model validation (mean and standard deviation from 100 runs). Given are measures of accuracies as number of latent variables used in PLSR (nLV), residual predictive deviation (RPD), measure of determination between predicted and observed lignin contents (R_p^2), root mean squared error of prediction (RMSEP), standard error of prediction (SEP) and deviation from the line of equality of linear regression between predicted and observed values (BIAS). We used raw data (Raw) and six different pre-processing methods: asymmetric least squares baseline offset correction (B.als), iterative restricted least squares baseline offset correction (B.irls), multiplicative scatter correction (MSC), standard normal variate (SNV), first derivative (D1) and second derivative (D2). The best model is highlighted in bold.

the RMSE steadily declined during the variable selection process such that the remaining wavelengths are more likely to allow better predictions. Moreover, CARS-PLSR identified wavelength regions, which code for the most important information within the regression procedure and thus potentially facilitate a mechanistic understanding of model results. The best CARS-PLSR model used 16 wavelengths, partly adjacent so that they formed nine clusters over the whole spectral region (as indicated by arrows and black bars in Fig. 2b). All wavelengths with PLS regression coefficients >0 are positively correlated with lignin content, whereas wavelengths with regression coefficients <0 indicate a negative relationship with lignin content and are most likely linkable to other chemical compounds (e.g. cellulose and polyoses). The selected 9 wavelength clusters from the best model can be assigned to local extrema of the regression coefficients obtained through PLSR modeling (Fig. 2a). The relative importance of selected wavelengths in the CARS-PLSR increases from lower to higher wavelengths with local maxima around 1428 nm, 1881 nm and 2274 nm (Fig. 2a) but increasing importance for the prediction of lignin content above the cluster of 1881 nm (Fig. 2b). However, our results also imply that wavelengths below 1881 nm likely have strong supplementary qualities since prediction accuracies strongly decreased if single wavelengths within that region were excluded from our model. Our best model indicated four wavelength clusters, which were positively

| Models (S = 100) | WL-Selection | Full spectrum PLSR | CARS-PLSR |
|------------------|-----------------------------|--------------------|--------------|
| Calibration | Variables | 662 | 16 |
| | nLV | 7.37 ± 1.28 | 6.99 ± 0.58 |
| | RMSE [%] | 3.58 ± 0.54 | 2.75 ± 0.30 |
| Validation | RPD | 1.96 ± 0.35 | 2.67 ± 0.46 |
| | R _p ² | 0.74 ± 0.08 | 0.86 ± 0.05 |
| | RMSEP [%] | 3.83 ± 0.57 | 2.82 ± 0.40 |
| | SEP [%] | 3.79 ± 0.56 | 2.78 ± 0.40 |
| | BIAS | -0.22 ± 0.86 | -0.19 ± 0.65 |

Table 2. Comparison of accuracies (mean ± SD) achieved for full spectrum PLSR versus best model from variable selection using CARS-PLS resulting from 100 repetitions for the calibration and validation. Abbreviations are as given in Table 1. In addition we use root mean square error of calibration (RMSE).

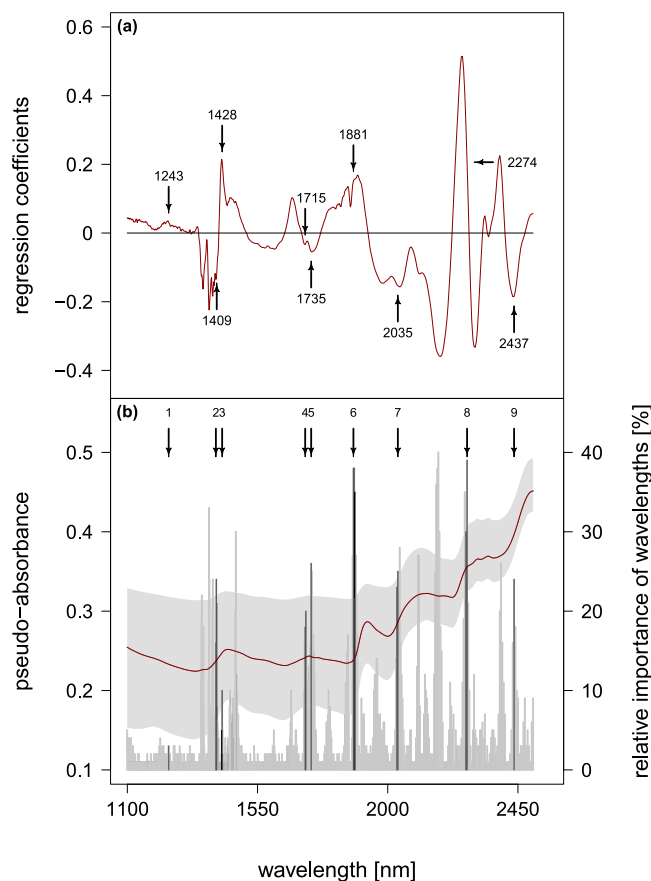


Figure 2. (a) Regression coefficients from full spectrum PLS (red line) with the selected wavelengths from the best CARS-PLS model as overlay (black arrows with wavelength indicated). (b) Mean absorbance spectrum of fine roots (red line) and its range (light grey) together with the relative importance (in %) of wavelength selection during 100 runs (dark grey bars) and the most important variables from the best model (black bars). Black arrows and number refer to the wavelength clusters indicated by the most important selected variables.

correlated with lignin content (clusters 1, 3, 6 and 8 corresponding to wavelength 1243 nm, 1428 nm, 1881 nm and 2274 nm, respectively). At the same time, we found five clusters which were negatively correlated with lignin content (clusters 2, 4, 5, 7, 9 corresponding to wavelength 1409 nm, 1715 nm, 1735 nm, 2035 nm and 2437 nm, respectively).

The gradient in species richness had no significant negative effect on model residuals in calibration or in validation (Fig. 3).

Literature survey. Our literature survey identified 30 papers, which predicted lignin content (extracted either based on Klason or the van Soest method) in plant tissue via spectroscopic measurements. These papers reported details on 38 case studies either as different species within one paper or different tissue (e.g. stem and

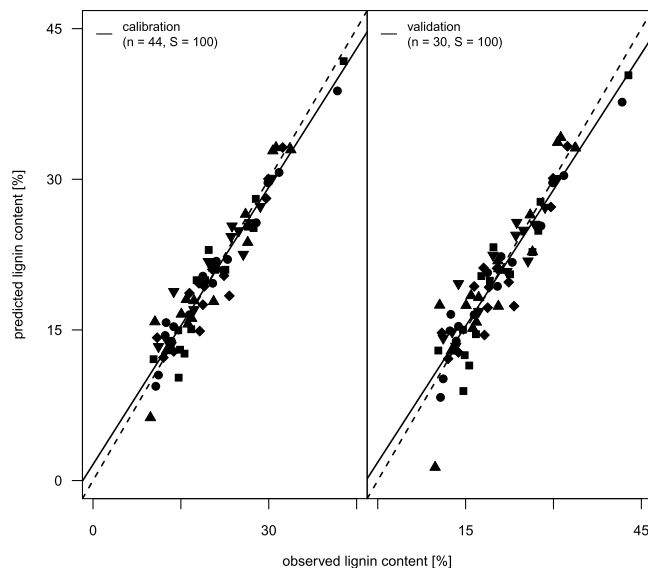


Figure 3. Predicted versus observed lignin contents (mean over 100 runs) from model calibration (left) and model validation (right) given as linear regression (dashed line) against the 1:1 line (solid line). Symbols represent the gradient of species richness (squares = 1, circles = 2, triangle = 4, rhombus = 8, triangle (upside down) = 16). n indicates the number of samples used in calibration and validation.

| | Species ³⁸ | Lignin [mg] ³⁶ | Val/Cal ³⁵ | RPD ²⁸ | R _p ²³³ | RMSEP [%] ¹⁹ | SEP [%] ²¹ |
|-------------------|-----------------------|---------------------------|-----------------------|-------------------|-------------------------------|-------------------------|-----------------------|
| Min | 1 | 100 | 0.12 | 0.5 | 0.05 | 0.19 | 0.28 |
| Q25 | 1 | 275 | 0.29 | 1.61 | 0.61 | 0.68 | 0.65 |
| Q75 | 1 | 500 | 0.50 | 3.56 | 0.90 | 1.88 | 2.27 |
| Max | 32 | 2500 | 3.55 | 9.5 | 0.99 | 3.48 | 5.00 |
| Med | 1 | 300 | 0.38 | 2.45 | 0.83 | 1.03 | 0.87 |
| This study | 61 | 10 | 0.68 | 2.67 | 0.86 | 2.82 | 2.78 |

Table 3. Summary statistics of a literature survey of studies predicting lignin contents from different plant tissues using spectroscopic methods. Given are minimum, maximum and interquartile ranges of the number of species examined per study (Species), sample volume used for chemical lignin extraction (Lignin) and proportion of samples used for model validation and calibration (Val/Cal). Additional abbreviations as in Table 1. Superscript numbers indicate the number of individual datasets that entered the analysis. See Table S1 for details on individual studies.

leaves) from the same species (see Supplementary Table S1 for full details of individual studies). The literature survey revealed three critical issues: First, 58% of the case studies analyzed lignin content in woody tissue, where a higher lignin content results in a more pronounced spectral signal. Second, 78% of the case studies were limited to one species and only 8% had more than five species. This minimizes the complexity of spectral signals reported in literature, as interspecific chemical variability is likely to be higher than intraspecific variability in chemical compounds^{53,54}. Third, the minimum amount of sample was 100 mg and the median amount 300 mg, which is one order of magnitude higher compared to this study (10 mg).

Table 3 summarizes the results of this literature survey as quantile ranges of common measures of accuracy and predictive power (RMSEP, SEP, R² and RPD). The median accuracy over these 38 case studies indicates an acceptable quantitative prediction of lignin content in plant tissue with a RPD of 2.45 and R² of 0.83 in validation. In comparison, our best model had a better quantitative prediction (RPD of 2.67) and a higher predictive R² of 0.86.

Discussion

Our study demonstrates that fine root lignin content is well predictable using near infrared spectroscopy. This is true for field samples from monocultures up to 16 species mixtures assembled from a pool of 60 different herbaceous grassland species. Further, we find a higher accuracy in model prediction when combining PLSR with CARS variable selection (RMSEP = 2.82) as compared to models using the full spectrum PLSR (RMSEP = 3.83). All measures of accuracy for lignin prediction in our study were in the range of results from 38 relevant case studies extracted from a literature survey, despite much lower sample volumes and higher sample variation in our study. Thus, we could show that the combination of Acetylbromid extracted lignin and FT-NIR spectroscopy is well suited for the rapid analysis of root lignin contents of herbaceous plant species even if the amount of sample is limited.

Fine root lignin content and range. The range of lignin content based on Acetylbromid extraction in our fine root samples (10–43%) exceeds that typically reported in the literature. However, most studies report lignin content in aboveground biomass based on Klason lignin extraction for woody gymnosperms (25–33%)⁵⁵, woody angiosperms (20–25%)⁵⁴ or herbaceous angiosperms (3–19%)^{51,56}. There are three potential reasons for the discrepancy in lignin ranges between our study and literature values. First, the extraction methods might produce different results. Though Acetylbromid is better correlated to Klason lignin than other extraction methods⁵¹, there is an offset between both methods for herbaceous species⁵⁶, which might account for part of the range differences we found. Second, roots often have higher lignin content than shoots⁵⁷. The limited data available on lignin content of grass species fine roots suggest a mean of $19.6\% \pm 1.7\%$, which is only slightly lower than our $20.5\% \pm 7.7\%$ for mixed herb species^{7,57}. Third, our samples most likely contained a larger proportion of partially degraded fine root debris for which lignin content is known to be higher⁴⁷. The field samples of mixed species were harder to sort for dead roots given the different color and texture of roots of different species. This problem might have been less pronounced for the data on individual species root lignin content published so far.

Spectral pre-processing. We tested six different types of spectral pre-processing and raw spectra to determine the most adequate method for our sample type. We found that standard normal variate (SNV) and multiplicative scatter correction (MSC) obtained the highest prediction accuracies according to residual predictive deviation (RPD) and R^2 in validation. Both types of spectral pretreatments have been specifically designed for correcting NIR spectra of scattered radiation⁵⁸ and have been widely applied in NIR spectroscopy³². Yet, error terms of prediction (RMSEP and BIAS) were higher in MSC compared to SNV in MSC compared to SNV. This difference can be traced back to the fact that we performed an independent spectral pre-processing for the validation and calibration set. Since MSC relies mathematically on the mean spectrum of the entire dataset⁵⁹, any pre-processing results diverge for calibration and validation datasets. Thus, prediction accuracies widely change for individual random draws of validation and calibration data. In contrast, SNV acts on the individual spectrum only, which makes it less dependable on pre-processing results⁶⁰. In addition, our results reveal that first and second derivative standardization, which have been found valuable in previous studies^{45,46,61} did not result in enhanced accuracies compared to using unprocessed spectra (raw spectra). This might indicate relatively minor contributions of spectral signal distortion in the data. While it remains difficult to a priori select the best pre-processing method⁴³, the case of MSC revealed that it is possible to choose an incapable one.

Selection of key wavelength-clusters based on CARS-PLS. We compared model accuracies between partial least square regression (PLSR) models and the combination of PLSR models with a predictor selection algorithm (competitive adaptive reweighted sampling, CARS). Our results indicate a significant increase in model validation accuracies through variable selection with CARS for all measures of model accuracy. Most likely, CARS predictor variable selection reduces the noise caused by non-informative variables for each individual model run. These findings are in line with other studies combining PLS based models with the CARS framework for variable selection^{27,37,38,40}, also reporting an enhanced model performance compared to using the full set of predictors from PLSR models alone.

In addition to increased model accuracies, the selection of 9 clusters of wavelengths from the best model might also enhance our mechanistic understanding of spectral lignin prediction. The finding that wavelength cluster 1, 3, 6 and 8 (1243 nm, 1428 nm, 1881 nm and 2274 nm) are positively correlated with lignin content (indicated by positive regression coefficients in PLS) is in line with previous studies where these wavelengths seemed indicative of lignin or lignified structures. Li *et al.*⁵⁰ also found the wavelength at 1243 nm to be relevant for predicting lignin content, but in contrast to our study, identified a larger set of important wavelengths (20), many of which located in the spectral region between 1450 nm and 1700 nm. The absorption at 1428 nm is often related to the first overtone of phenolic O-H stretching in lignin^{62–64}. The cluster at 1881 nm is ascribed to C-H⁶² and O-H stretching and deformation⁶⁵ and is associated with lignin in general⁶². The cluster around 2270 nm may also arise from O-H stretching overtones in lignin^{63,64}. Thus, all selected positively correlated wavelength clusters can be directly linked to lignin or chemical structures related to lignin, which further strengthens the applicability of near-infrared prediction of fine root lignin content.

Similarly, the wavelength clusters 2, 4, 5, 7 and 9 (1409 nm, 1715 nm, 1735 nm, 2035 nm and 2437 nm) which were negatively correlated with lignin content, have previously been linked to chemical groups with some ecological counterpart or trade-off towards fine root lignin content.

During lignification, the protein containing, holocellulosic fiber scaffolding of the compound middle lamella and the secondary cell wall are incrustated with lignin⁶⁶. This process increases the mechanical stability of lignified tissue against high compression and capillary forces. In addition, lignification enhances the recalcitrance of plant tissue against the alkaline soil solution, root exudates and exoenzymes from bacterial and fungal communities^{5,67}. According to Schwanninger *et al.*³⁰ wavelengths around 1715 nm and 1735 nm result from overtone C-H stretching vibrations in polyoses and cellulose. The wavelength cluster around 2035 nm could be connected to amide-I stretching in peptides and proteins^{63,64,68} or similarly to C=O stretching vibration from acetyl groups in polyoses⁵⁰. Bands at 2437 nm probably indicate α -helical protein structures⁶³. Thus, the wavelength clusters 4, 5 and 7, which are negatively correlated to root lignin in our study, are indicative of cellulose, polyoses and proteins, indicating a trade-off well known from literature⁶⁹. This trade-off may be explained by a substrate competition for the biosynthesis of cellulose and protein in a living cell, versus biosynthesis of lignin and the subsequent death of the cell⁷⁰.

A potentially competing chemical group, providing similar functionality to the cell as lignin, are suberins. Suberins and associated waxes are important hydrocarbon-containing biopolymers that also increase the mechanical stability of plant tissue yet without impairing the water absorption capacity and plasticity of the root

tissue. Our wavelength cluster 2 (around 1409 nm) has been connected to C-H stretching and deformation in aliphatic hydrocarbons^{64,65} and hydrocarbon-containing plant waxes⁶⁸ as well as to O-H deformations in extractable alcohols^{30,64,65}. Again, there might be a trade-off in allocating carbon to enhance tissue stability via either lignin or suberin and associated waxes. Suberins occur in the secondary walls of endodermal and hypodermal cells of primary roots, while the incorporation of waxes into the suberin polymer is limited to root tissues with secondary growth⁷¹. Yet, monocotyledonous roots, such as grasses, have no secondary growth and should invest into higher lignin concentration at the expense of wax contents. Indeed, Zeier and Schreiber^{72,73} could show an inverse relationship between endodermal lignin and suberin content in five monocotyledonous species. In addition, Chen *et al.*¹¹ could show that grass containing plots from the same experiment had significantly higher lignin contents and lower nitrogen contents in fine roots than legume containing plots.

Overall, our results show that wavelength assignments in NIRS have the potential to reveal insightful information regarding the underlying chemistry. Wavelengths with high explanatory power in the regression model provide us with background knowledge on the chemical groups or bindings, which absorb at distinct wavelengths⁷⁴. However, this is not a trivial task in complex lignocellulosic materials³⁰ since exact band positions depend on different equilibrium moisture contents⁷⁵ and thus intrinsically on the chemical composition of constituents.

Model design and best model outcome. The aim of our study was to develop a model to predict fine root lignin content, which is well applicable to a larger range of herbaceous species. For this reason, we put a strong focus on evaluating the predictive power of our models. Consequently, we assigned a much larger than usual proportion of samples to model validation. In addition, we used repeated outer validation instead of single cross-validation, as this allows for a more realistic determination of uncertainty when predicting unknown samples⁷⁶. Both procedures increase the reliability of our models yet at the cost of enhanced error terms especially given the overall limited number of samples.

Despite this conservative procedure our best model has a better quantitative prediction (higher RPD value of 2.67) and a higher predictive R^2 of 0.86 compared to the median of 38 case studies extracted from our literature survey (RPD = 2.45, $R^2 = 0.83$). In contrast, our measures of error of prediction (RMSE and SEP) are higher than the 75% quantile range of literature studies. This discrepancy is due to the fact that our data set comprised a much larger sample variability due to the higher number of species. We are aware of only one other study⁴⁷ with comparable sample complexity using leaves, leaf litter and organic residue of 31 tree species to predict lignin content via near infrared spectroscopy (SEP = 5, $R^2 = 0.76$ and RPD = 2.1). Similar results (RMSEP = 5.5 and $R^2 = 0.5$) were achieved by Kelly *et al.*⁷⁷ when predicting lignin content for 14 agricultural species used for fiber production, indicating that in fact sample heterogeneity might well determine the error of model prediction for lignin. This phenomenon is well known for model prediction via spectral analysis for biological applications⁷⁸. However, Petisco *et al.*⁴⁶ found that lignin content in tree leaves from 17 species was well predictable (SEP < 1), but in that study lignin content was spanning a smaller range and standard deviation than realized in our study. Overall, our results reveal that root lignin in a large number of herbaceous species is at least as well predictable as the median of studies predicting lignin content in one or few mostly woody or crop species.

Conclusion

Our study explored - for the first time - the applicability of NIR spectroscopy to determine AcBr extracted root lignin content in mixtures of up to 60 grassland species. We used CARS-PLS to select the most relevant wavelengths for root lignin prediction and concurrently increased prediction accuracy compared to PLS models based on full spectrum information. Despite the large number of species and the limited amount of sample available in our study, we found higher prediction accuracies than the median of 38 case studies extracted from a literature survey. Thus, we conclude that predicting AcBr extracted root lignin from NIRS spectroscopy shows great potential to overcome the limitation of large sample sizes as commonly examined in ecological studies.

Material and Methods

Sampling. The Jena Experiment is a biodiversity experiment studying the effects of plant species and functional group richness on ecosystem functions. 60 mesophilic central European grassland species were classified into four functional groups: grasses (16 species), legumes (12 species), small herbs (12 species) and tall herbs (20 species). Detailed information about the design and main results of 15 years Jena Experiment is given in Weisser *et al.*⁷⁹. For the current study, we collected roots of plant communities from 76 experimental plots, spanning the existing gradient of species richness (1, 2, 4, 8, 16) and functional group richness (1, 2, 3, 4). Community root collection is explained in detail in Chen *et al.*¹¹. In brief, we excavated soil volumes with a surface area ranging from 20 × 10 cm to 40 × 15 cm and a depth of 20 cm depending on the spatial extent of the respective root biomass. Roots were soaked in tap water prior to washing them over a 630 μm-mesh sieve and removing coarse roots (>2 mm), debris and highly decomposed roots before drying (65 °C). Three samples did not contain enough fine root material and had to be excluded from further analysis, leaving 73 samples for this study. In addition, we included one sample of standardized *Lolium perenne* L. roots. We used this as a reference sample of young and fully clean grass roots contrasting our field root samples with low diameter but potentially higher age (and lignification) as well as partly decaying roots or soil residue. *L. perenne* plants were cultivated in aeroponics in the greenhouse of the Botanical Garden of Leipzig University over 4 months. Newly grown roots and shoots were cut to 3 cm length every 4 weeks and roots were dried (65 °C) and stored. Dry *L. perenne* roots from all harvests were shredded and thoroughly mixed before grinding multiple smaller portions of the large sample volume with a vibratory ball mill (MM 400, Retsch Technology GmbH, Germany). After grinding, the powder was again carefully mixed and oven dried again (70 °C, 48 hours). Our overall sample number was therefore 74 (73 field samples plus *L. perenne*). Moreover, we used homogenized root biomass as internal reference standard to test replicability and recovery rates of chemical analysis. This reference sample was comprised of a mixed field community root

sample bulked from the 50 largest field samples described above. This sample contained a large variety of grass-land species and was extracted and measured with each analytical run.

Chemical analysis. The left box of Fig. 1 depicts the chemical analysis where individual steps are indicated with numbers in square brackets. We refer to these numbers in the following text. Dried root samples were ground with a vibratory ball mill (MM 400, Retsch Technology GmbH, Germany) and oven dried again (70 °C, 48 hours) before further processing ([1], Fig. 1). For liquid-solid pre-extraction we extracted up to 50 mg of sample with 12 mL solvent (acetone: ethanol: water; 5:3:2 volume) at 70 °C for 150 minutes, turning the tubes regularly. The extracted samples were centrifuged and washed three times before drying (70 °C, 48 hours). For lignin extraction, we used the acetyl bromide (AcBr) extraction as described by Iiyama & Wallis²⁴, but avoided to use 70% perchloric acid that causes the formation of hydrobromic acid and unwanted acid catalyzed, chromophor-forming oxidation of polysaccharides [2]. In brief, we extracted 10 mg sample with 5 mL 25% (vol:vol) solution of AcBr in glacial acetic acid and heated the vials in an oil bath (70 °C, 60 min) with regular shaking to promote sample digestion. We chilled samples on ice (15 min), equilibrated to room temperature (30 min) and centrifuged. To mask strongly absorbing polybromide anions, 1 mL of the supernatant was diluted in 1 mL of 2 N NaOH and 8 mL glacial acetic acid. We included microcrystalline cellulose (Sigma-Aldrich, USA) as control⁸⁰. Finally, we measured 3 mL of the sample solution at 280 nm in a spectrophotometer (Jasco V730, Jasco Labor- u. Datentechnik GmbH, Germany) to determine the specific absorption coefficients (SAC) [3]:

$$SAC = \frac{(OD_S - OD_B) * F}{W_d} * ml * cm^{-1} * mg^{-1} \quad (1)$$

where OD_S = optical density of the sample, OD_B = optical density of the blank, W_d = weight of the sample and F [$ml\ mg^{-1}$] = dilution factor (=50) and d [cm] = diameter of the quartz cuvette. We calculated the SAC per sample as mean over all replicates per sample ($2 \leq n \leq 4$). Each replicate of a sample was measured on a different day.

To translate SAC into lignin content we used the reference sample from mixed field roots described above. Two subsamples individually underwent the same procedure as described above. In addition, we purified and isolated these reference samples as described in detail by Fukushima & Dehority⁸¹. For the calibration curve we diluted 10–750 μ L extracted lignin aliquots in 8 mL masking solution, made up to 10 mL with blank solution (25% AcBr in acetic acid) and measured at 280 nm as detailed above. The lignin content (L) of all samples was calculated using regression eq. (2) [4].

$$L = \frac{(SAC - 0.05) * 100}{13.06} * \% \quad (2)$$

Spectral measurements. The ground and dried root samples were measured between $12489\ cm^{-1}$ – $3594\ cm^{-1}$ (800 nm – 2782 nm) with $8\ cm^{-1}$ spectral resolution in transmission (Multi-Purpose FT-NIR-Analyzer, Bruker Cooperation, USA). Transmission (T) was converted to absorbance via $\log_{10}(1/T)$. Each sample was measured 5 times; spectra were subsequently averaged. Samples were shaken between each replicate to ensure a spectral recording of the sample-inherent variability [5].

Statistical analysis. We used partial least squares regression (PLSR) to relate measured spectra of 74 collected samples (each with a total of 662 spectral bands as predictor variables) to the measured lignin contents of these samples (response variable) [5]. We used PLSR as the most widespread multivariate calibration method. PLSR is capable of handling a high degree of collinearity in the predictor variables⁸² as well as a small number of response samples in relation to the number of predictors³³. In addition, we combined full spectrum-PLSR with the CARS algorithm, introduced and described in detail by Li *et al.*³⁷. In short, CARS, aims at selecting key wavelengths in a computationally efficient way. The selection procedure in CARS is largely based on the PLS regression coefficients and consists of two major steps. In the first step an exponential decreasing function (EDF) defines the number of selected spectral variables. As a result, the number of kept variables decreases exponentially from one sampling run to the next and wavelengths linked to small absolute PLS regression coefficients are removed before the second step in each sampling run. This second step, also referred to as adaptive reweighted sampling (ARS), uses the absolute regression coefficients to define the probability of a single variable to be drawn in random sampling procedure. As a result, variables with low probabilities might not be drawn and are thus also removed from potential variable space. After a defined number of sampling runs the algorithm chooses the overall best model.

In this study, PLSR modelling was divided into two blocks [block I and II, Fig. 1]. In the first block, we used random sampling to divide the 74 spectra into 44 spectra for model building (calibration data) and 30 spectra for an outer model validation [6]. We sampled randomly 100 times, saved these data splits and reused them in all following steps. To mirror the operational application of NIR spectroscopy for predicting fine-root lignin from unknown field samples, using an already established prediction model, calibration and prediction data sets were pre-processed separately [6]. To test the effects of different pre-processing procedures, we used raw data as well as six pre-processing methods: asymmetric least squares baseline offset correction (B.als)⁸³, iterative restricted least squares baseline offset correction (B.irls)⁸⁴, multiplicative scatter correction (MSC)⁵⁸, standard normal variate (SNV)⁸⁵, first derivative (D1) and second derivative (D2)⁴³ [7]. Based on the differently pre-processed calibration data set the PLS regression models were computed and cross validated (ten-fold) to avoid model overfitting and to determine the optimal number of latent variables (nLV)³³ [8]. Then all final models were applied to predict the respective prediction data [9]. From Block I we analyzed the obtained prediction accuracies averaged from 100 data splits and selected the pre-processing method with the highest accuracy for further analysis in Block II [10].

In the second block of model building, we used the CARS-PLSR approach as described above. We used the same data splits as in Block I to allow a direct comparison with the results of full spectrum-PLSR [10].

We describe model quality using the following indices: coefficient of determination for prediction (R_p^2), the root mean square error of cross validation (RMSECV), the root mean square error of prediction (RMSEP), the standard error of prediction (SEP) and the residual predictive deviation (RPD), calculated from standard deviation of the prediction data set and SEP. We consider RPD values >1.5 – 2 as good for preliminary screenings and initial predictions⁸⁶, RPD values >2 – 2.5 as acceptable for quantitative predictions, values >2.5 – 3 as good and values >3 as excellent for predictions⁵⁰. We conducted all statistical analyses in the statistics software R (version 3.3.1)⁸⁷ using the packages: baseline for B.als/B.irls corrections⁸⁴, pls⁸⁸, carspls³⁷ (downloadable at: <https://code.google.com/archive/p/carspls/downloads>) and prospectr for SNV and MSC⁸⁹.

Data Availability

The data used in this article is accessible via <https://doi.pangaea.de/10.1594/PANGAEA.895501>.

References

- Jones, M. B. & Donnelly, A. Carbon sequestration in temperate grassland ecosystems and the influence of management, climate and elevated CO₂. *New Phytol.* **164**, 423–439 (2004).
- Berg, B. & McClaugherty, C. *Plant Litter. Decomposition, Humus Formation, Carbon Sequestration* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008).
- Jackson, R. B. *et al.* A global analysis of root distributions for terrestrial biomes. *Oecologia* **108**, 389–411 (1996).
- Poorter, H. *et al.* Biomass allocation to leaves, stems and roots. Meta-analyses of interspecific variation and environmental control. *New Phytol.* **193**, 30–50 (2012).
- Rasse, D. P., Rumpel, C. & Dignac, M.-F. Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation. *Plant Soil* **269**, 341–356 (2005).
- Mendez-Millan, M., Dignac, M.-F., Rumpel, C., Rasse, D. P. & Derenne, S. Molecular dynamics of shoot vs. root biomarkers in an agricultural soil estimated by natural abundance ¹³C labelling. *Soil Biol. Biochem.* **42**, 169–177 (2010).
- Silver, W. L. & Miya, R. K. Global patterns in root decomposition: comparisons of climate and litter quality effects. *Oecologia* **129**, 407–419 (2001).
- Hättenschwiler, S., Tiunov, A. V. & Scheu, S. Biodiversity and Litter Decomposition in Terrestrial Ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **36**, 191–218 (2005).
- Solly, E. F. *et al.* Factors controlling decomposition rates of fine root litter in temperate forests and grasslands. *Plant Soil* **382**, 203–218 (2014).
- Chen, H. *et al.* Plant species richness negatively affects root decomposition in grasslands. *J Ecol* **105**, 209–218 (2017).
- Chen, H. *et al.* Root chemistry and soil fauna, but not soil abiotic conditions explain the effects of plant diversity on root decomposition. *Oecologia* **185**, 499–511 (2017).
- Liang, X., Erickson, J. E., Silveira, M. L., Sollenberger, L. E. & Rowland, D. L. Tissue chemistry and morphology affect root decomposition of perennial bioenergy grasses on sandy soil in a sub-tropical environment. *GCB Bioenergy* **8**, 1015–1024 (2016).
- Goebel, M. *et al.* Decomposition of the finest root branching orders. Linking belowground dynamics to fine-root function and structure. *Ecol Monogr* **81**, 89–102 (2011).
- Prieto, I., Stokes, A. & Roumet, C. Root functional parameters predict fine root decomposability at the community level. *J. Ecol.* **104**, 725–733 (2016).
- Austin, A. T. & Ballaré, C. L. Dual role of lignin in plant litter decomposition in terrestrial ecosystems. *Proc. Natl. Acad. Sci. USA* **107**, 4618–4622 (2010).
- Hättenschwiler, S. & Jørgensen, H. B. Carbon quality rather than stoichiometry controls litter decomposition in a tropical rain forest. *J. Ecol.* **98**, 754–763 (2010).
- Fengel, D. & Wegener, G. *Wood. Chemistry, ultrastructure, reactions* (Kessel, Remagen, 2003).
- Boerjan, W., Ralph, J. & Baucher, M. Lignin biosynthesis. *Annu. Rev. Plant Biol.* **54**, 519–546 (2003).
- Calvo-Flores, F. G., Dobado Jiménez, J. A., Garcia, J. I. & Martín-Martínez, F. J. *Lignin and lignans as renewable raw materials. Chemistry, technology and applications* (Wiley, Chichester, West Sussex, 2015).
- Van Soest, P. J., Robertson, J. B. & Lewis, B. A. Symposium: Carbohydrate methodology, metabolism and nutritional implications in dairy cattle. Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *J. Dairy Sci.* **74**, 3583–3597 (1991).
- TAPPI T 222 om-02, Acid-insoluble lignin in wood and pulp (2002–2003 TAPPI Test Methods) (Tappi Press, Atlanta, GA, USA, 2002).
- Sluiter, A., *et al.* Determination of structural carbohydrates and lignin in biomass (Technical Report NREL/TP-510-42618) (National Renewable Energy Laboratory, Washington DC, USA, 2008).
- Johnson, D. B., Moore, W. E. & Zank, L. C. The spectrophotometric determination of lignin in small wood samples. *Tappi J* **44**, 793–798 (1961).
- Iiyama, K. & Wallis, A. F. A. An improved acetyl bromide procedure for determining lignin in woods and wood pulps. *Wood Sci. Technol.* **22**, 271–280 (1988).
- Chang, X. F., Chandra, R., Berleth, T. & Beatson, R. P. Rapid, microscale, acetyl bromide-based method for high-throughput determination of lignin content in *Arabidopsis thaliana*. *J. Agric. Food Chem.* **56**, 6825–6834 (2008).
- Cécillon, L. *et al.* Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS). *Eur J Soil Sci* **60**, 770–784 (2009).
- Ramirez, J. A. *et al.* Near-infrared spectroscopy (NIRS) predicts non-structural carbohydrate concentrations in different tissue types of a broad range of tree species. *Methods Ecol Evol* **6**, 1018–1025 (2015).
- Hayes, D. J. M., Hayes, M. H. B. & Leahy, J. J. Use of near infrared spectroscopy for the rapid low-cost analysis of waste papers and cardboards. *Faraday Discuss.* **202**, 465–482 (2017).
- Jin, X. *et al.* Determination of hemicellulose, cellulose and lignin content using visible and near infrared spectroscopy in *Miscanthus sinensis*. *Bioresour. Technol.* **241**, 603–609 (2017).
- Schwanninger, M., Rodrigues, J. C. & Fackler, K. A Review of Band Assignments in near Infrared Spectra of Wood and Wood Components. *J. Near Infrared Spectrosc.* **19**, 287–308 (2011).
- Bokobza, L. Near Infrared Spectroscopy. *J. Near Infrared Spectrosc.* **6**, 3–17 (1998).
- Roggo, Y. *et al.* A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J Pharm Biomed Anal* **44**, 683–700 (2007).
- Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**, 109–130 (2001).
- Xiaobo, Z., Jiewen, Z., Povey, M. J. W., Holmes, M. & Hanpin, M. Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* **667**, 14–32 (2010).

35. Mehmood, T., Liland, K. H., Snipen, L. & Sæbø, S. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **118**, 62–69 (2012).
36. Yun, Y.-H. *et al.* Using variable combination population analysis for variable selection in multivariate calibration. *Anal. Chim. Acta* **862**, 14–23 (2015).
37. Li, H., Liang, Y., Xu, Q. & Cao, D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal. Chim. Acta* **648**, 77–84 (2009).
38. Vohland, M., Ludwig, M., Harbich, M., Emmerling, C. & Thiele-Bruhn, S. Using Variable Selection and Wavelets to Exploit the Full Potential of Visible–Near Infrared Spectra for Predicting Soil Properties. *J. Near Infrared Spectrosc.* **24**, 255–269 (2016).
39. Sun, J. *et al.* Detection of internal qualities of hami melons using hyperspectral imaging technology based on variable selection algorithms. *J. Food Process Eng.* **40**(3) (2017).
40. Richter, R., Reu, B., Wirth, C., Doktor, D. & Vohland, M. The use of airborne hyperspectral data for tree species classification in a species-rich Central European forest area. *Int. J. Appl. Earth Obs. Geoinf.* **52**, 464–474 (2016).
41. Vohland, M., Ludwig, M., Thiele-Bruhn, S. & Ludwig, B. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma* **223–225**, 88–96 (2014).
42. Hutengs, C., Ludwig, B., Jung, A., Eisele, A. & Vohland, M. Comparison of Portable and Bench-Top Spectrometers for Mid-Infrared Diffuse Reflectance Measurements of Soils. *Sensors (Basel, Switzerland)* **18** (2018).
43. Rinnan, Å., van den Berg, F. & Engelsen, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *Trends Anal. Chem.* **28**, 1201–1222 (2009).
44. Castillo, R. *et al.* NIR spectroscopy applied to the characterization and selection of pre-treated materials from multiple lignocellulosic resources for bioethanol production. *J. Chil. Chem. Soc.* **59**, 2347–2352 (2014).
45. Jiang, W. *et al.* Rapid assessment of coniferous biomass lignin–carbohydrates with near-infrared spectroscopy. *Wood Sci Technol* **48**, 109–122 (2014).
46. Petisco, C. *et al.* Near-infrared reflectance spectroscopy as a fast and non-destructive tool to predict foliar organic constituents of several woody species. *Anal. Bioanal. Chem.* **386**, 1823–1833 (2006).
47. Ono, K., Hiraide, M. & Amari, M. Determination of lignin, holocellulose, and organic solvent extractives in fresh leaf, litterfall, and organic material on forest floor using near-infrared reflectance spectroscopy. *J. For. Res.* **8**, 191–198 (2003).
48. Ye, X. P. *et al.* Fast classification and compositional analysis of cornstover fractions using Fourier transform near-infrared techniques. *Bioresour. Technol.* **99**, 7323–7332 (2008).
49. Niu, W., Huang, G., Liu, X., Chen, L. & Han, L. Chemical Composition and Calorific Value Prediction of Wheat Straw at Different Maturity Stages Using Near-Infrared Reflectance Spectroscopy. *Energy Fuels* **28**, 7474–7482 (2014).
50. Li, X., Sun, C., Zhou, B. & He, Y. Determination of Hemicellulose, Cellulose and Lignin in Moso Bamboo by Near Infrared Spectroscopy. *Sci. Rep.* **5**, 17210 (2015).
51. Fukushima, R. S. & Hatfield, R. D. Comparison of the acetyl bromide spectrophotometric method with other analytical lignin methods for determining lignin concentration in forage samples. *J. Agric. Food Chem.* **52**, 3713–3720 (2004).
52. Moreira-Vilar, F. C. *et al.* The acetyl bromide method is faster, simpler and presents best recovery of lignin in different herbaceous tissues than Klason and thioglycolic acid methods. *PLoS one* **9**, e110000 (2014).
53. Jung, V., Violle, C., Mondy, C., Hoffmann, L. & Muller, S. Intraspecific variability and trait-based community assembly. *J. Ecol.* **98**, 1134–1140 (2010).
54. Schweiger, A. K. *et al.* Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. *Nat. Ecol. Evol.* **2**, 976–982 (2018).
55. Sjöström, E. & Alén, R. *Analytical Methods in Wood Chemistry, Pulping, and Papermaking* (Springer Berlin Heidelberg, Berlin, Heidelberg, 1999).
56. Fukushima, R. S. & Hatfield, R. D. Extraction and Isolation of Lignin for Utilization as a Standard to Determine Lignin Concentration Using the Acetyl Bromide Spectrophotometric Method. *J. Agric. Food Chem.* **49**, 3133–3139 (2001).
57. Brown, P. H., Graham, R. D. & Nicholas, D. J. D. The effects of manganese and nitrate supply on the levels of phenolics and lignin in young wheat plants. *Plant Soil* **81**, 437–440 (1984).
58. Pasquini, C. Near Infrared Spectroscopy. Fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.* **14**, 198–219 (2003).
59. Geladi, P., MacDougall, D. & Martens, H. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Appl. Spectrosc.* **39**, 491–500 (1985).
60. Dhanoa, M. S., Lister, S. J., Sanderson, R. & Barnes, R. J. The Link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformations of NIR Spectra. *J. Near Infrared Spectrosc.* **2**, 43–47 (1994).
61. Rijal, D., Walsh, K. B., Subedi, P. P. & Ashwath, N. Quality Estimation of Agave Tequilana Leaf for Bioethanol Production. *J. Near Infrared Spectrosc.* **24**, 453–465 (2016).
62. Barton, F. E. & Himmelsbach, D. S. Two-Dimensional Vibrational Spectroscopy II. Correlation of the Absorptions of Lignins in the Mid- and Near-Infrared. *Appl. Spectrosc.* **47**, 1920–1925 (1993).
63. Workman, J. & Weyer, L. *Practical guide and spectral atlas to interpretive near-infrared spectroscopy*. 2nd ed. (CRC Press, Boca Raton, London, New York, op. 2012).
64. Osborne, B. G. & Fearn, T. *Near infrared spectroscopy in food analysis*. 2nd ed. (Longman, Harlow, 1988).
65. He, W. & Hu, H. Prediction of hot-water-soluble extractive, pentosan and cellulose content of various wood species using FT-NIR spectroscopy. *Bioresour. Technol.* **140**, 299–305 (2013).
66. Fergus, B. J. & Goring, D. A. I. The location of guaiacyl and syringyl lignins in birch xylem tissue. *Holzforschung* **24**, 113–117 (1970).
67. Steinauer, K., Chatzinotas, A. & Eisenhauer, N. Root exudate cocktails: The link between plant diversity and soil microorganisms? *Ecol. Evol.* **6**, 7387–7396 (2016).
68. Barton, F. E., Himmelsbach, D. S., Duckworth, J. H. & Smith, M. J. Two-Dimensional Vibration Spectroscopy: Correlation of Mid- and Near-Infrared Regions. *Appl. Spectrosc.* **46**, 420–429 (1992).
69. Aulen, M., Shipley, B. & Bradley, R. Prediction of *in situ* root decomposition rates in an interspecific context from chemical and morphological traits. *Ann. Bot.* **109**, 287–297 (2012).
70. Novaes, E., Kirst, M., Chiang, V., Winter-Sederoff, H. & Sederoff, R. Lignin and biomass: A negative correlation for wood formation and lignin content in trees. *Plant Physiol.* **154**, 555–561 (2010).
71. Schreiber, L. Transport barriers made of cutin, suberin and associated waxes. *Trends Plant Sci.* **15**, 546–553 (2011).
72. Zeier, J. & Schreiber, L. Comparative investigation of primary and tertiary endodermal cell walls isolated from the roots of five monocotyledonous species: chemical composition in relation to fine structure. *Planta* **206**, 349–361 (1998).
73. Zeier, J. & Schreiber, L. Fourier transform infrared-spectroscopic characterisation of isolated endodermal cell walls from plant roots: chemical nature in relation to anatomical development. *Planta* **209**, 537–542 (1999).
74. Westad, F., Schmidt, A. & Kermit, M. Incorporating Chemical Band-Assignment in near Infrared Spectroscopy Regression Models. *J. Near Infrared Spectrosc.* **16**, 265–273 (2008).
75. Papadopoulos, A. N. & Hill, C. A. S. The sorption of water vapour by anhydride modified softwood. *Wood Sci. Technol.* **37**, 221–231 (2003).
76. Filzmoser, P., Liebmann, B. & Varmuza, K. Repeated double cross validation. *J. Chemometrics* **23**, 160–171 (2009).

77. Kelley, S. Rapid analysis of the chemical composition of agricultural fibers using near infrared spectroscopy and pyrolysis molecular beam mass spectrometry. *Biomass Bioenergy* **27**, 77–88 (2004).
78. Peirs, A., Schenk, A. & Nicolai, B. M. Effect of natural variability among apples on the accuracy of VIS-NIR calibration models for optimal harvest date predictions. *Postharvest Biol. Technol.* **35**, 1–13 (2005).
79. Weisser, W. W. *et al.* Biodiversity effects on ecosystem functioning in a 15-year grassland experiment: Patterns, mechanisms, and open questions. *Basic Appl. Ecol.* **23**, 1–73 (2017).
80. Hatfield, R. D., Grabber, J., Ralph, J. & Brei, K. Using the Acetyl Bromide Assay to Determine Lignin Concentrations in Herbaceous Plants: Some Cautionary Notes. *J. Agric. Food Chem.* **47**, 628–632 (1999).
81. Fukushima, R. S. & Dehority, B. A. Feasibility of using lignin isolated from forages by solubilization in acetyl bromide as a standard for lignin analyses. *J. Animal Sci.* **78**, 3135–3143 (2000).
82. Wold, S., Ruhe, A., Wold, H. & Dunn, I. W. J. I. I. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. and Stat. Comput.* **5**, 735–743 (1984).
83. Eilers, P. H. C., Boelens, H. F. M. *Baseline Correction with Asymmetric Least Squares Smoothing*. (Leiden University Medical Centre, Leiden, 2005).
84. Liland, K. H., Mevik, B.-H. baseline: Baseline Correction of Spectra. R package version 1.2-1. <https://CRAN.R-project.org/package=baseline> (2015).
85. Barnes, R. J., Dhanoa, M. S. & Lister, S. J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.* **43**, 772–777 (1989).
86. Jones, P. D., Schimleck, L. R., Daniels, R. F., Clark, A. & Purnell, R. C. Comparison of Pinus taeda L. whole-tree wood property calibrations using diffuse reflectance near infrared spectra obtained using a variety of sampling options. *Wood Sci. Technol.* **42**, 385–400 (2008).
87. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> (2016).
88. Mevik, B.-H., Wehrens, R., Liland, K. H. pls: Partial Least Squares and Principal Component Regression. R package version 2.6-0. <https://CRAN.R-project.org/package=pls> (2016).
89. Stevens, A., Ramirez-Lopez, L. An introduction to the prospectr package. R package Vignette, R package version 0.1.3. <https://CRAN.R-project.org/package=prospectr> (2013).

Acknowledgements

This study was performed on samples from the Jena Experiment which was funded by the German Science Foundation (DFG, FOR 1451) and was supported by the Friedrich-Schiller-University Jena and the Max Planck Society. We thank the gardeners of the Jena Experiment for maintaining the plots and the scientific manager and A. Ebeling for field work coordination. We also thank H. Chen for sampling and pre-processing fine roots and the gardeners of the Botanical Garden (Leipzig University) for their help with community root collection and *Lolium perenne* cultivation.

Author Contributions

A.W. designed the experiment. O.E. collected the data. R.R. and O.E. analyzed the data with input of M.V. O.E., R.R. and A.W. wrote the manuscript with input from M.V. and all authors provided input on the final written manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-42837-z>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019