

Improving plant biomass estimation in the field using partial least squares regression and ridge regression

Brian M. Ohsowski, Kari E. Dunfield, John N. Klironomos, and Miranda M. Hart

Abstract: Estimating primary productivity over time is challenging for plant ecologists. The most accurate biomass measurements require destructive sampling and weighing. This is often not possible for manipulative studies that involve repeated measures over time, or for studies in protected areas. Estimates of aboveground plant biomass using allometric equations or linear regression on single plant traits have been used, but tend to have poor predictability both within and across systems, or are limited to specific plant taxa. Here we estimate aboveground plant biomass using multiple collinear plant traits to generate a standard curve specific to the site of interest. Partial least squares regression (PLS) and ridge regression (RR), addressing predictor collinearity, are robust, highly accurate statistical methods to estimate plant biomass across gross differences in plant morphology and growth habit.

Key words: nondestructive sampling, plant biomass, partial least squares, grassland restoration, ridge regression.

Résumé : L'estimation de la productivité primaire au cours du temps est un défi pour les phytoécologistes. Les mesures les plus précises de la biomasse nécessitent un échantillonnage et un pesage destructifs. Cela n'est souvent pas possible pour la réalisation d'études expérimentales qui impliquent des mesures répétées au cours du temps, ou d'études en zones protégées. Les estimés de la biomasse végétale hors-sol à l'aide d'équations allométriques ou par régression linéaire sur les caractères d'une plante isolée ont été utilisés mais ils tendent à avoir une faible prévisibilité tant à l'intérieur qu'entre les systèmes, ou ils sont limités à un taxon végétal spécifique. Les auteurs ont estimé ici la biomasse végétale hors-sol à l'aide de caractères végétaux multiples colinéaires afin de générer une courbe standard spécifique au site d'intérêt. La régression par la méthode des moindres carrés partiels et la régression d'arête, explorant la colinéarité du prédicteur, constituent des méthodes statistiques robustes et précises pour estimer la biomasse végétale malgré des différences importantes sur le plan de la morphologie végétale et du type de croissance. [Traduit par la Rédaction]

Mots-clés : échantillonnage non-destructif, biomasse végétale, moindres carrés partiels, réhabilitation des prairies, régression d'arête.

Introduction

Plant biomass is considered a good approximation of primary productivity, especially in grassland communities (Hector et al. 1999). Directly measuring plant biomass, however, often requires fully destructive sampling techniques, which prevents repeated sampling in a community and impacts information on long-term growth dynamics. Direct biomass measurements are also labour intensive and impractical when experimental designs require a high level of replication. Minimally destructive estimation methods are useful when large-scale plant harvesting is impractical, and sampling design requires repeated measures (Catchpole and Wheeler 1992).

Historically, allometric equations have been used to estimate plant biomass with minimal destruction in the field. These equations are constructed using predictor variables such as diameter at breast height and total plant height to estimate biomass (Picard et al. 2012). The advantage of allometric equations is that they do not require calibration through destructive sampling. While used successfully to estimate woody plant mass (Sah et al. 2004), allometric equations developed to measure herbaceous plants are relatively scarce and can be highly variable in prediction accuracy because they have been shown to differ with plant morphology and environmental conditions (Niklas and Enquist 2002; Chave et al. 2004).

Received 13 January 2016. Accepted 15 April 2016.

B.M. Ohsowski. Institute of Environmental Sustainability, Loyola University Chicago – Lakeshore Campus, Chicago, IL 60660, USA.

K.E. Dunfield. School of Environmental Sciences, University of Guelph, Guelph, ON N1G 2W1, Canada.

J.N. Klironomos* and M.M. Hart.* Biology, University of British Columbia - Okanagan, Kelowna, BC V1V 1V7, Canada.

Corresponding author: Miranda M. Hart (email: Miranda.hart@ubc.ca).

*These authors contributed equally to this work.

Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from RightsLink.

An alternative to allometric equations is the use of linear regression (LR) models to create standard curves specific to the study's species of interest. This approach has advantages over allometric equations because a predictor variable from the system in question is regressed against actual biomass measurements collected on site (Catchpole and Wheeler 1992). This approach both improves accuracy over published allometric equations, and reduces the labour of harvesting and weighing biomass. Once a standard curve is established, rapid measurement of similar vegetation in the field is straightforward. Plant height is often assumed to be correlated with plant biomass when creating predictive univariate LR models, and has been used as an easily measured surrogate for plant biomass (Singh et al. 1975; Catchpole and Wheeler 1992). In the field, however, the predictive power of univariate LR models has been shown to vary, depending on ecosystem type and vegetative structure (Catchpole and Wheeler 1992). Even within a plant population, a single plant trait may not be necessarily well-correlated with plant biomass across all individuals (Gholz et al. 1979). Thus, while generating a standard curve improves accuracy over strict equations, in practice, these models may have weak predictive power.

Incorporating multiple plant traits has been suggested as a way to improve the accuracy of regression models (Ahmed and Bonham 1982). The use of classic multiple predictors should increase biomass estimation accuracy by taking into account several correlated plant traits. As plant traits are often collinear, this violates an assumption of classical multiple linear regression and can lead to erratic predictions of a response variable (Graham 2003).

Partial least squares regression (PLS) is a statistical method commonly used in computational chemistry that predicts a response variable from multiple, collinear predictor variables (Wold et al. 2001). PLS is becoming increasingly popular in ecological data analysis (Carrascal et al. 2009) as it address two problems common in ecological studies: (i) datasets with a high number of predictor variables relative to the number of samples, and (ii) high collinearity amongst predictor variables, such as plant morphological traits. Additionally, continuous and categorical data can be used simultaneously in PLS, which is an essential feature when measuring morphological aspects of plants. An alternative method such as ridge regressions (RR) is another viable statistical technique that analyzes collinear predictor variables. RR is a more robust version of LR that places constraints on regression coefficients and reduces the potential of overfitting data in a model.

Here, we apply two multivariate statistical approaches for estimating plant biomass in the field, PLS regression and RR. Compared with destructive harvesting, minimally destructive PLS plant biomass estimation will increase sampling volume, reduce data collection time,

and minimize labour. We show that both RR and PLS increase plant biomass prediction accuracy in three morphologically distinct plant growth forms compared with simple linear regression models.

Materials and methods

Species selection and data collection

We selected three plant species representing differences in morphology/growth habit to determine the robustness of the PLS approach: a small shrub (*Cornus racemosa* Lam., grey dogwood), a tussock grass (*Sporobolus cryptandrus* Torr., sand dropseed), and a fern with radial rhizomes (*Osmunda claytoniana* L., interrupted fern). In September 2012, 41 individuals from each plant species were sampled from southern Ontario's hardwood forest near Simcoe, Ontario, Canada (42°40'15.25"N, 80°28'44.96"W). We established a 50 m transect through a representative plant population for the target species and selected the closest individual at each 1 m increment. The target species selected along each transect captured the range of size and shape morphology present in the plant populations. Before harvesting each plant, a suite of plant trait measurements was collected (see below) before the plant was clipped at the soil surface and dried in a forced-air oven at 60 °C. The plant biomass was considered dry with the plant mass stabilized within ±0.3 g.

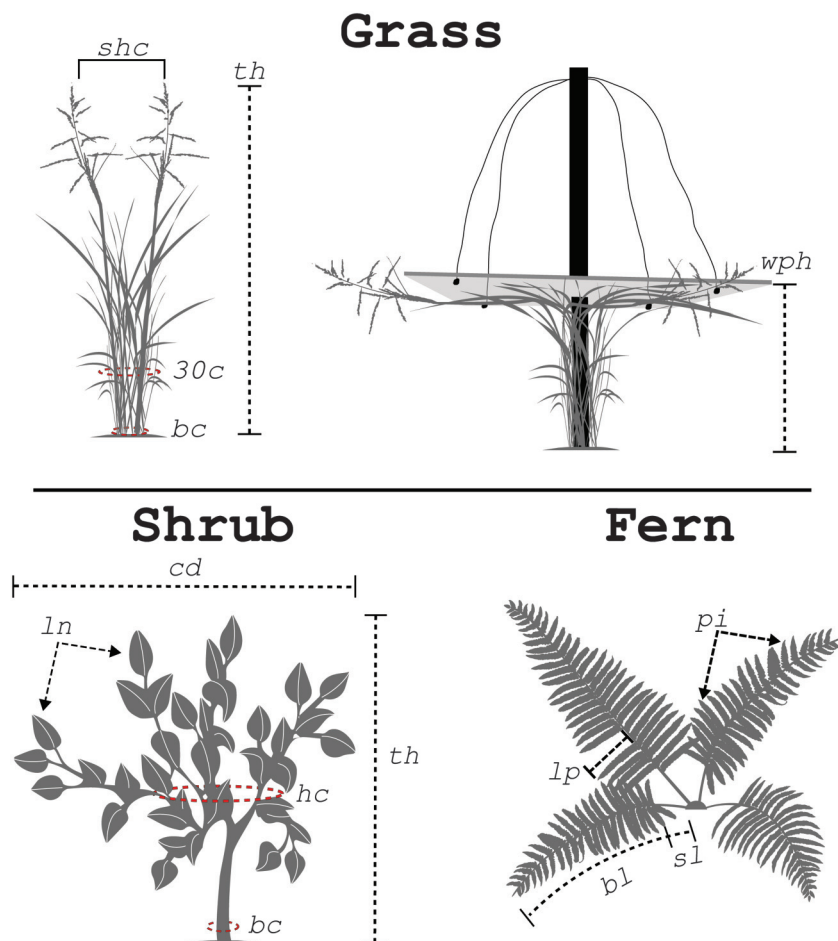
Measured plant traits

Plant height and density were determined with a standardized plate meter following Rayburn and Rayburn (1998). To accomplish this, an acrylic plexiglass sheet (40.0 cm × 40.0 cm × 3.2 mm) with a dowel (122.0 cm × 1.9 cm) inserted into the centre was lowered onto the plant until four leaves touched the plate (see Fig. 1). Plant height was recorded by measuring height of the plate from the soil surface. After measuring plant height, the weighted plate was then lowered to rest on an individual, compressing the plant tissue. The resting plate height gives an approximate estimate of the density of an individual plant (Rayburn and Rayburn 1998).

Circumference was obtained by measuring the circumference of all the plant material from one individual using a cloth measuring tape at three locations: half of the plant height, at 30 cm from the soil surface, and at the soil surface. Measuring plant circumference at a height of 30 cm allowed for the tight bundling of plant biomass when gathered. A list of measurements collected for each species is given in Fig. 1.

In the case of the *O. claytoniana*, plant morphology was distinctly different from the *S. cryptandrus* and *C. racemosa*. *Osmunda claytoniana* has radial rhizomes with multiple fronds growing in a circular cluster. Plant measurements were taken per frond and subsequently averaged and summed per individual. The average and sum measurements were the variables used to create the statistical models. See Fig. 1 for details on measurements.

Fig. 1. Workflow for predicting plant biomass with partial least squares regression. 30c, circumference at height of 30 cm; bc, basal circumference; bl, fern blade length; cd, maximum canopy diameter; fl, frond length (blade length + stipe length); hc, circumference at half plant height; ln, leaf number; lp, longest pinna per blade; pi, pinnae number per blade; shc, seed head count; sl, stipe length; th, total plant height; wph, resting height of falling plate meter.



Model creation

A training dataset and a test dataset were created from the 41 individuals from each species. The training dataset was built by randomly selecting 35 individuals, with the remaining six plants serving as external data points (test dataset). The same training and test datasets were used when creating the PLS, RR, and LR prediction models. Allometric equations were not compared because published equations do not exist for the three plant species in this study.

Data transformation, auto-scaling, and polynomial terms

Each variable was transformed to approximate normality (Table 1). Data were normalized (mean centered and auto-scaling) because PLS and RR are sensitive to fluctuations in scale and variance among predictor variables. Diagnostic plots indicated potential curvilinear relationships after autoscaling between predictor and response variables. Polynomial terms (2nd and 3rd orders) were calculated for each response variable after autoscaling and included in the variable selection calculations (Schielzeth 2010).

Variable reduction and model averaging

Using the training dataset ($n = 35$), all possible combinations of transformed variables and associated 2nd and 3rd order polynomials were scored using Bayesian Information Criterion (BIC) model selection in the MuMin package (Barton 2015) in R (R Core Team 2013). MuMin's dredge function was used to force the condition that polynomial regression coefficients must be evaluated in conjunction with 1st order regression coefficients to ensure proper fitting of the model (Schielzeth 2010; Symonds and Pither 2012). Models with $BIC \leq 2$ are considered to be equivalent. Therefore, BIC models ≤ 2 were averaged and variable importance values for the predictor variables were extracted. The plant measurements selected by BIC model selection represent the optimized variables that will best predict plant mass in the dataset (Johnson and Omland 2004; Grueber et al. 2011).

Partial least squares regression and linear regression models

PLS regression models were created with the pls package (Mevik and Wehrens 2007) in R. The optimized pre-

Table 1. Measured plant traits included in the optimized partial least squares regression, ridge regression, and linear regression models.

Model	Comp	RMSECV (g)	Predictors (mass $\sim x_1 + x_2 + \dots x_n$)
<i>Cornus racemosa</i>			
PLS	3	47.7	Mass $\sim bc + bc^2 + cd + cd^2 + hc + \text{sqrt}(\ln) + \text{sqrt}(\ln)^2 + th$
RR	NA	48.0	Mass $\sim bc + bc^2 + cd + cd^2 + hc + \text{sqrt}(\ln) + \text{sqrt}(\ln)^2 + th$
LR	NA	65.0	Mass $\sim th$
<i>Osmunda claytoniana</i>			
PLS	4	18.2	Mass $\sim \text{sqrt}^{-1}(\Sigma(\text{sl})) + \log(\text{fc}) + \text{sqrt}^{-1}(\text{mean}(\text{bl})) + \log(\Sigma(\text{fl}))$
RR	NA	18.4	Mass $\sim \text{sqrt}^{-1}(\Sigma(\text{sl})) + \log(\text{fc}) + \text{sqrt}^{-1}(\text{mean}(\text{bl})) + \log(\Sigma(\text{fl}))$
LR	NA	22.5	Mass $\sim \log(\text{mean}(\text{fl}))$
<i>Sporobolus cryptandrus</i>			
PLS	4	38.1	Mass $\sim bc + bc^2 + \text{square}(30c) + \text{cube}(\text{th}) + \text{cube}(\text{th})^2 + \text{sqrt}(\text{wph})$
RR	NA	38.1	Mass $\sim bc + bc^2 + \text{square}(30c) + \text{cube}(\text{th}) + \text{cube}(\text{th})^2 + \text{sqrt}(\text{wph})$
LR	NA	47.2	Mass $\sim \text{cube}(\text{th})$

Note: Partial least squares (PLS) regression component selection based on lowest root mean squared error from cross-validation (RMSECV) using 10-fold cross-validation. RR, ridge regression; LR, linear regression; 30c, circumference at height of 30 cm; bc, basal circumference; cd, maximum canopy diameter; fc, frond count; fl, frond length (blade length + stipe length); hc, circumference at half plant height; ln, number of leaves; sl, stipe length; th, total plant height; wph, resting height of falling plate meter.

dicator variables and transformations used to calculate each plant's PLS model are given in Table 1. Ridge regression models were calculated with the MASS package in R (Venables and Ripley 2002). Simple linear regression (LR) models were calculated with the plant height predictor variable to establish a standard curve in the lm function found in R's base package (R Core Team 2013).

The number of terms retained in each model was determined using the root mean squared error of cross-validation (RMSECV) calculated from 10-fold cross-validation. RMSECV is a diagnostic metric used to test each component's contribution to the overall predictive fit of the model. The latent variable with the lowest average RMSECV indicates the number of components to retain in the PLS model, thus maximizing each model's predictive performance. The value of lambda, a vector of ridge constants, in the ridge regression model was optimized to the lowest generalised cross validation score for each plant species.

After generating standard curves from the PLS, RR, and LR models, plant mass was predicted and back transformed (P_{mass}) for the training and test datasets. P_{mass} was subtracted from the corresponding reference plant mass (R_{mass}) weighed in the laboratory to determine how well the model predicted each data point. A perfect model prediction for a sample is equal to zero (i.e., $P_{\text{mass}} - R_{\text{mass}} = 0$). Root mean squared error (RMSE) and R^2 estimates for the linear relationship between P_{mass} versus R_{mass} were calculated to determine the training and test dataset's actual predictive performance. The regression slope that was used to calculate RMSE and R^2 for the P_{mass} versus R_{mass} training dataset is equal to a slope = 1 with an intercept = 0. Mean and standard deviations of $P_{\text{mass}} - R_{\text{mass}}$ were calculated for the training and test datasets.

Results

Variable selection in PLS models

BIC model selection reduced the number of variables from the full model in the *O. claytoniana* (5 \rightarrow 4) and *S. cryptandrus* (5 \rightarrow 4) datasets. All *C. racemosa* models included the five variables in the optimized PLS models (Table 1). Field-measured variables retained in the PLS models for all three species included plant diameter, height, and structural counts. Weighted plate measurements were only relevant when predicting the biomass of *S. cryptandrus*. PLS models used for *C. racemosa* and *S. cryptandrus* corrected for curvilinear relationships between plant biomass and several field measured variables (i.e., 2nd order polynomial terms) (Table 1).

Comparing models for predicting plant biomass in the training dataset

PLS and RR models consistently performed equally well in prediction diagnostics for all three plant species compared to LM models (Tables 2 and 3). The RMSECV model performance indicators translated to higher prediction accuracy when evaluating P_{mass} versus R_{mass} in each training dataset. R^2 values ranged from 0.990–0.995 (PLS) and 0.990–0.996 (RR) when comparing the predicted vs. reference biomass values. Predicted P_{mass} versus R_{mass} regression diagnostics in the training dataset were more variable in LR models with R^2 values ranging from 0.872–0.928. In all LR models, the lower R^2 values were due to reduced model performance when predicting heavier plants in the population (Fig. 2). In comparison, RR and PLS models accurately predicted training dataset plant mass across all plant weights resulting in higher R^2 values (Fig. 2). Model prediction performance in the training dataset, determined by RMSE values, indicated that all PLS and RR models consistently outperformed LR models when comparing $P_{\text{mass}} - R_{\text{mass}}$

Table 2. Summary of statistics for the optimized partial least squares regression, ridge regression, and linear regression model training datasets.

Species	Mass \pm 1 SD	Height \pm 1 SD	Model	RMSE (g)	R ²
<i>Cornus racemosa</i>	42.9 \pm 41.3 g	77.7 \pm 32.9 cm	PLS	4.4	0.995
			RR	3.9	0.996
			LR	20.5	0.872
<i>Osmunda claytoniana</i>	17.1 \pm 12.9 g	81.0 \pm 18.1 cm	PLS	2.1	0.992
			RR	2.2	0.992
			LR	5.6	0.928
<i>Sporobolus cryptandrus</i>	32.9 \pm 23.6 g	88.3 \pm 19.0 cm	PLS	4.4	0.990
			RR	4.2	0.990
			LR	13.4	0.880

Note: R² and root mean squared error (RMSE) values are based on P_{mass} versus R_{mass} estimates where slope = 1 and intercept = 0.

data (Table 2). RR model prediction performance compared with PLS was only slightly improved when analyzed by RMSE in *C. racemosa* and *S. cryptandrus* (Table 2).

Comparing models for predicting plant biomass for external data

Similar to models using the training set data, external data points predicted by PLS models consistently outperformed LR models (Table 3). R² of P_{mass} versus R_{mass} PLS test data points ranged between 0.991–0.995 (PLS) and 0.991–0.995 (RR), whereas R² data for LR ranged between (0.669–0.950). PLS and RR model RMSE of P_{mass} versus R_{mass} was consistently lower when using LR models to predict species biomass. Average $P_{\text{mass}} - R_{\text{mass}} \pm 1$ SD for externally predicted data using PLS was highest in *C. racemosa* (2.7 \pm 7.4 g) and lowest in *O. claytoniana* (1.4 \pm 2.2 g) (Table 3). In comparison, average $P_{\text{mass}} - R_{\text{mass}} \pm 1$ SD using LR had reduced predictive model performance with highest prediction variability in *C. racemosa* (–8.2 \pm 16.2 g) and lowest variability in *O. claytoniana* (–1.7 \pm 3.1 g).

Discussion

In our study, PLS and RR were far better predictors of plant biomass compared with LR for the three plant species in this study. This indicates a distinct advantage of using a multivariate approach to predict plant biomass in the field, since growth form did not strongly influence the predictive performance of PLS and RR. Despite the weak correlation of the individual variables to plant mass in the *O. claytoniana* population, both multivariate approaches resulted in an accurate model with good predictive performance. This highlights the usefulness of the regression techniques that account for collinearity when estimating plant mass in the field.

The success of the multivariate PLS and RR approaches show the importance of incorporating several predictor variables to estimate plant mass while accounting for variable collinearity. Incorporating variables that estimated plant density, circumference, and structural counts accounted for morphological variation. Approaches that employ only one predictor variable largely ignore the fact that two plants with identical heights may have distinctly different plant volumes in the field. This ulti-

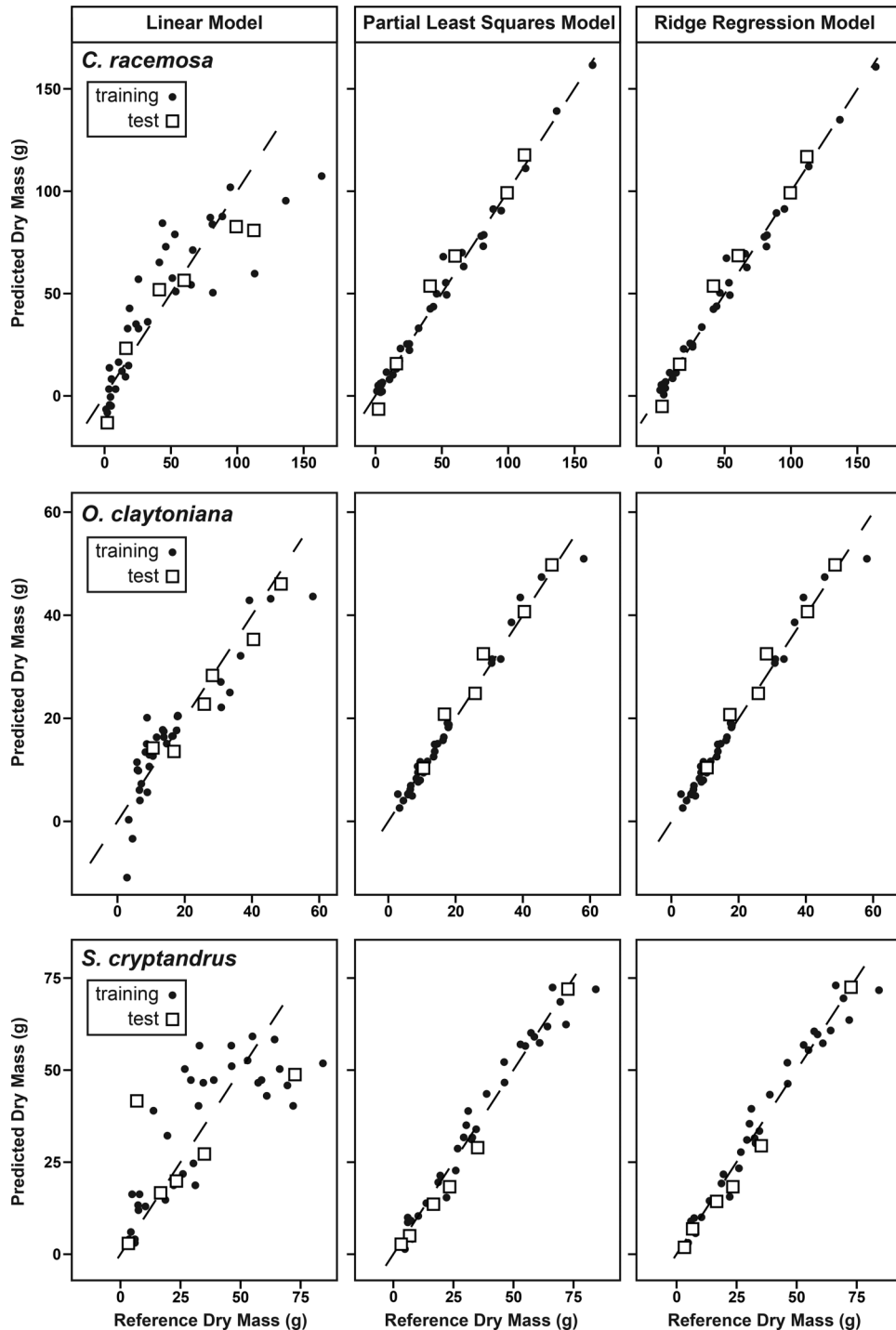
mately leads to higher variance in prediction accuracy and less reliable results (Catchpole and Wheeler 1992). In our study, predicting the biomass of larger individuals with LR reduced measurement accuracy and introduced uncertainty into the model. Because PLS and RR models have no statistical restrictions when variables exhibit multicollinearity (Wold et al. 2001), this allows for the incorporation of all variables that adequately describe aboveground plant architecture and morphological variation, resulting in excellent predictive performance in the field compared to all tested LR models.

In terms of variation within plant species, abiotic conditions in the physical environment has been shown to influence plant morphology (Westoby et al. 2002). In our study, variation in both externally and internally predicted data tended to increase with increased average plant mass for all plant species. Larger prediction error can reasonably be expected, owing to the higher variability associated with plant growth rate response during competition for water, light, and nutrients (Poorter and Nagel 2000). In this study, LR was more sensitive to plants with higher average biomass and standard deviations, resulting in less robust biomass prediction models using plant height as the sole predictor variable. This effect was most pronounced when evaluating predicted mass versus reference mass in the *S. cryptandrus* population, but was present in all evaluated plant species. PLS and RR models, on the other hand, exhibited high predictive performance across all plants, regardless of plant biomass ranges in the field. The robustness of the multivariate approach accounts for more morphological variability even within species, thus increasing the reliability of the model.

PLS and RR in practice

Compared with published allometric equations and LR using plant height, these multivariate models are customizable to specific study systems. Our multivariate models included both continuous (height, diameter) and categorical variables (structural). This approach incorporates the flexibility to choose not only the number but the types of variables used.

Fig. 2. Predicted (P_{mass}) vs. reference (R_{mass}) plant biomass using the optimized partial least squares regression, ridge regression, and linear regression model for each plant species. The solid black points represent internally predicted data used to train each model ($n = 35$). Open squares represent external data predictions from the test dataset using only predictor variables ($n = 6$). Each broken line indicates a perfect prediction ($P_{\text{mass}} = R_{\text{mass}}$) with a slope = 1 and intercept = 0.



Labor cost and time necessary to create a multivariate model is similar to collecting a single response variable using LR. This approach is easily adapted to a variety of field and greenhouse situations, thus increasing sample replication, work efficiency, and prediction accuracy. While the destruction of a small subset of plants is inevitably required for all nondestructive biomass prediction

analyses, this can be accounted for when designing an experiment. Increasing sample replication with the intent of destructive harvesting, creating a preliminary experiment under the same environmental conditions and harvesting its biomass, or choosing a representative population in the field similar to the population of interest are all viable options to account for initial biomass destruction.

Table 3. Summary of statistics for the optimized partial least squares regression, ridge regression, and linear regression model externally predicted data.

Species	Model	Mean	RMSE	
		$(P_{\text{mass}} - R_{\text{mass}}) \pm 1 \text{ SD}$ (g)	(g)	R^2
<i>Cornus racemosa</i>	PLS	2.7±7.4 g	7.0	0.991
	RR	2.8±7.2 g	7.0	0.991
	LR	-8.2±16.2 g	14.8	0.950
<i>Osmunda claytoniana</i>	PLS	1.4±2.2 g	2.4	0.995
	RR	1.4±2.2 g	2.4	0.995
	LR	-1.7±3.1 g	2.8	0.926
<i>Sporobolus cryptandrus</i>	PLS	-2.9±2.3 g	6.0	0.991
	RR	-2.4±2.5 g	5.7	0.992
	LR	-0.1±19.3 g	17.1	0.699

Note: R^2 and root mean squared error (RMSE) values are based on P_{mass} compared with R_{mass} estimates where slope = 1 and intercept = 0.

Alternatives to PLS and RR have been suggested, such as principal component regression (PCR) and artificial neural networks (ANN) (Hastie et al. 2001). Compared with PLS, PCR does not account for variance associated with response variables and resulting models tend to be less parsimonious with higher variability. Consistent with our results, studies evaluating PLS performance compared with RR show similar predictive performance on a simulated dataset (Frank and Friedman 1993) or marginally better predictive performance using observations in economic data investigating response variables to GDP per capita in Turkey (Yeniay and Goktas 2002). In both cases, RR and PLS outperform PCR. To the best of our knowledge, this is the first direct comparison of RR and PLS. In general, both PLS and RR models are equally parsimonious, easy to interpret, and user friendly thus affording no distinct statistical advantage over the other. Advances in computational statistics and machine learning suggest that ANN may someday create better predictive models than the preceding linear regression techniques. To date, ANN methodologies are not widely used in ecology. ANN computations have a steep learning curve, as the underlying statistics do not use common statistical methods. Thus, ANN methods are less accessible and more difficult to implement.

In conclusion, PLS and RR are robust statistical techniques that accurately predicts plant biomass in field and greenhouse experiments where destructive sampling is not possible. The customizable nature of this technique makes the usage of these models a powerful statistical tool for researchers in ecological and environmental science.

Acknowledgements

Thank you to the Nature Conservancy of Canada for the use of their land when conducting this field research and The Ontario Aggregate Research Corporation (TOARC) for funding and resource support. I would also like to thank my field assistant, Sarah Krus, for her invaluable help in the field.

References

- Ahmed, J., and Bonham, C.D. 1982. Optimum allocation in multivariate double sampling for biomass estimation. *J. Range Manage.* **35**: 777–779. doi:10.2307/3898263.
- Barton, K. 2015. MuMIn: multi-model inference. R package version 1.15.1.
- Carrascal, L.M., Galván, I., and Gordo, O. 2009. Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, **118**: 681–690. doi:10.1111/j.1600-0706.2008.16881.x.
- Catchpole, W.R., and Wheeler, C.J. 1992. Estimating plant biomass: a review of techniques. *Aust. J. Ecol.* **17**: 121–131. doi:10.1111/j.1442-9993.1992.tb00790.x.
- Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., and Perez, R. 2004. Error propagation and scaling for tropical forest biomass estimates. *Philos. Trans. R. Soc. Ser. B Biol. Sci.* **359**: 409–420. doi:10.1098/rstb.2003.1425.
- Frank, L.E., and Friedman, J.H. 1993. A statistical view of some chemometrics regression tools. *Technometrics*, **35**: 109–135. doi:10.1080/00401706.1993.10485033.
- Gholz, H.L., Grier, C.C., Campbell, A.G., and Brown, A.T. 1979. Equations for estimating biomass and leaf area of plants in the Pacific Northwest. Forest Research Lab, Oregon State University, Corvallis, Ore. pp. 1–43.
- Graham, M.H. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology*, **84**: 2809–2815. doi:10.1890/02-3114.
- Grueber, C.E., Nakagawa, S., Laws, R.J., and Jamieson, I.G. 2011. Multimodel inference in ecology and evolution: challenges and solutions. *J. Evol. Biol.* **24**: 699–711. doi:10.1111/j.1420-9101.2010.02210.x.
- Hastie, T., Tibshirani, R., and Friedman, J.J.H. 2001. The elements of statistical learning: prediction, inference and data mining. Springer, New York.
- Hector, A., Schmid, B., Beierkuhnlein, C., Caldeira, M.C., Diemer, M., Dimitrakopoulos, P.G., Finn, J.A., et al. 1999. Plant diversity and productivity experiments in European grasslands. *Science*, **286**: 1123–1127. doi:10.1126/science.286.5442.1123. PMID:10550043.
- Johnson, J.B., and Omland, K.S. 2004. Model selection in ecology and evolution. *Trends Ecol. Evol.* **19**: 101–108. doi:10.1016/j.tree.2003.10.013. PMID:16701236.
- Mevik, B.H., and Wehrens, R. 2007. The pls package: principal component and partial least squares regression in R. *J. Stat. Softw.* **18**: 1–24. doi:10.18637/jss.v018.i02.
- Niklas, K.J., and Enquist, B.J. 2002. On the vegetative biomass partitioning of seed plant leaves, stems, and roots. *Am. Nat.* **159**: 482–497. doi:10.1086/339459. PMID:18707431.
- Picard, N., Saint-André, L., and Henry, M. 2012. Manual for building tree volume and biomass allometric equations: from field measurement to prediction. Food and Agricultural Organization of the United Nations, Rome, Italy, and Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, France.
- Poorter, H., and Nagel, O. 2000. The role of biomass allocation in the growth response of plants to different levels of light, CO₂, nutrients and water: a quantitative review. *Aust. J. Plant Physiol.* **27**: 595–607. doi:10.1071/PP99173.
- R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rayburn, E.B., and Rayburn, S.B. 1998. A standardized plate meter for estimating pasture mass in on-farm research trials. *Agron. J.* **90**: 238–241. doi:10.2134/agronj1998.0002196200900020022x.
- Sah, J.P., Ross, M.S., Koptur, S., and Snyder, J.R. 2004. Estimating aboveground biomass of broadleaved woody plants in the

- understory of Florida Keys pine forests. *For. Ecol. Manage.* **203**: 319–329. doi:10.1016/j.foreco.2004.07.059.
- Schielzeth, H. 2010. Simple means to improve the interpretability of regression coefficients. *Methods Ecol. Evol.* **1**: 103–113. doi:10.1111/j.2041-210X.2010.00012.x.
- Singh, J.S., Lauenroth, W.K., and Steinhorst, R.K. 1975. Review and assessment of various techniques for estimating net aerial primary production in grasslands from harvest data. *Bot. Rev.* **41**: 181–232. doi:10.1007/BF02860829.
- Symonds, J.E., and Pither, J. 2012. Multiscale analyses of exotic and native plant species diversity within Canada's endangered antelope brush ecosystem. *Plant Ecol.* **213**(8): 1263–1275. doi:10.1007/s11258-012-0084-7.
- Venables, W.N., and Ripley, B.D. 2002. *Modern applied statistics with S*. Springer, New York.
- Westoby, M., Falster, D.S., Moles, A.T., Vesk, P.A., and Wright, I.J. 2002. Plant ecological strategies: some leading dimensions of variation between species. *Annu. Rev. Ecol. Syst.* **33**: 125–159. doi:10.1146/annurev.ecolsys.33.010802.150452.
- Wold, S., Sjöström, M., and Eriksson, L. 2001. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**: 109–130. doi:10.1016/S0169-7439(01)00155-1.
- Yeniay, O., and Goktas, A. 2002. A comparison of partial least squares regression with other prediction methods. *Hacettepe J. Math. Stat.* **31**: 99–111.