Appendix S1 Why species tell more about traits than vice versa: Predictive analysis to connect species, traits, and environment

JAMES S. CLARK

In this supplement I provide mathematics for trait norms and co-dispersion, data used in the example, and modeling/computation background.

Trait norms and co-dispersion

To facilitate connections to joint trait distributions I provide norms and codispersion in matrix notation. In the following t_{ms} is trait m for species s, organized in a M by S matrix of traits by species matrix **T**. For a single trait the *trait norm* is

$$\bar{\mathbf{u}}_m = S^{-1} \mathbf{t}'_m \mathbf{1}_S$$

where \mathbf{t}_m is the length-*S* vector of trait values, the *m*th row of **T**, and $\mathbf{1}_S$ is the length-*S* vector of ones. For *M* traits the trait norm (TN) is a vector

$$TN = \overline{\mathbf{u}} = S^{-1} \mathbf{T} \mathbf{1}_S$$
 S2

This is the LQ matrix of RLQ analysis (Doledec et al. 1996, Brown et al. 2014, Dray et al. 2014).

The CWM value is a *weighted trait norm* taken for a vector of observed species weights at location *i*

$$WTN_i = Tw_i$$
 S3

Taken over a sample of *n* observations the weighted trait norm is WTN = n^{-1} **TW**'**1**_{*n*}, where **W** is the *n* by *S* species composition sample. WTN varies between minimum and maximum trait values; these extremes obtain when the species having those values dominate a sample. Exploratory methods, including randomization techniques, are often used to evaluate WTN.

Although traits are fixed in CWM analysis they do have dispersion that can be represented in ways that permit analysis of species and environment effects. Trait co-dispersion independent of species diversity quantifies tendency for traits to be shared across species, the columns in trait-by-species matrix **T**,

$$TC = S^{-1}(T - \overline{U})(T - \overline{U})'$$
 S4

where

$$\overline{\mathbf{U}} = \overline{\mathbf{u}}\mathbf{1}'_{S}$$

1

is a *M* by *S* matrix, each column being the length-*M* vector of trait norms $\overline{\mathbf{u}}$ (eqn S1). I refer to TC as co-*dispersion* rather than covariance, because traits are treated as fixed in CWM analysis. TC quantifies dispersion of individual traits and codispersion between traits in the same manner as would be described by a covariance for a random variable. Just as TN is a baseline for effects of species diversity on trait norms, TC is a baseline for evaluating effects of species diversity on trait dispersion. The abundance-weighted trait dispersion for sample *i* requires *M* by *S* matrices of weights and traits. The weight matrix for observation *i* is

 $\mathbf{W}_i = \mathbf{1}_m \mathbf{w}'_i$

The weighted trait co-dispersion is

$$WTC_i = [(\mathbf{T} - \overline{\mathbf{U}}) \circ \mathbf{W}_i](\mathbf{T} - \overline{\mathbf{U}})'$$

where \circ is the Hadamard product. WTC_i is a *M* by *M* matrix. Analogous to a coefficient of variation, the square root of the diagonal of the WTC_i matrix can be divided by TN as an index of dispersion for a location (Fig. 1B). Trait co-dispersion over *n* observations requires the *n* by *S* matrix **W**,

WTC =
$$n^{-1} [(\mathbf{T} - \overline{\mathbf{U}}) \circ \mathbf{1}_m \mathbf{1}'_n \mathbf{W}] (\mathbf{T} - \overline{\mathbf{U}})'$$
 S5

WTC differs from Laliberté and Legendre's (2010) functional trait dispersion (FDis) in that WTC is defined on a variance (t^2) scale rather than a linear scale. I adopt WTC for two reasons. First, it is consistent with the weighted mean for trait norm, with WTN being the corresponding weighted norm. Second, it allows direct comparison with trait covariance that emerges from a model-based approach to species and trait diversity, discussed below. Still, WTC is expected to show patterns similar to FDis. As with the WTN this dispersion is induced by the sample **W**; there is not yet a model.

Trait prediction for environmental covariance

The effect of environmental variables is available using a predictive approach. Based on the fitted model in eqn 6 we can predict a realization having the same structure as observed data, starting from the joint distribution for species weights and predictors

 $[\mathbf{w}_i, \mathbf{x'}_i] = [\mathbf{w}_i | \mathbf{x'}_i] [\mathbf{x'}_i]$

The first factor on the right-hand side is eqn 6. The second factor is a distribution for predictors with mean and covariance,

 $[\mathbf{x}'_i] = MVN(\mathbf{m}, \mathbf{Q})$

Marginalizing \mathbf{x}'_i followed by a variable change from \mathbf{w}_i to \mathbf{u}_i gives a predictive trait distribution

 $\mathbf{u}_{i}^{\prime} \sim MVN(PM, PTC)$ S6

having predictive mean and predictive trait covariance

$$PM = T\beta'm$$

$$PTC = \mathbf{T}(\Sigma + \boldsymbol{\beta}' \mathbf{Q} \boldsymbol{\beta}) \mathbf{T}'$$
 S8

PTC is the model-based equivalent of WTC in eqn S5. The two terms in eqn S8 are i) the covariance in trait diversity unrelated to environmental predictors, and ii) the covariance that is related to the environment. For the PTM this second term is the *environmental trait covariance*,

$$ETC = T\beta'Q\beta T'$$
 S9a

As a relative contribution of the environmental predictors to the variance we have the ratio

$$RETC = \frac{\operatorname{diag}(\mathbf{T}\boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta}\mathbf{T}')}{[\operatorname{diag}(\mathbf{T}\boldsymbol{\Sigma}\mathbf{T}') + \operatorname{diag}(\mathbf{T}\boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta}\mathbf{T}')]}$$
S9b

Because the PTM is fitted to species before translating to traits we can evaluate the same relationships for species, $\text{ESC} = \beta' Q \beta$. The environmental contribution to variance in traits can be compared to its contribution to species

$$RESC = \frac{diag(\beta' Q\beta)}{[diag(\Sigma) + diag(\beta' Q\beta)]}$$
 S10

Full covariance matrices can be compared using standard tools such as cluster analysis.

Species and trait diversity can be summarized by volume of the hyper-ellipsoid in trait space, taken here at a probability value of 0.05,

$$V(\mathbf{C}) = \left[\frac{2(\pi \chi_M^2(0.05))^{M/2} det(\mathbf{C})^{1/2}}{M\Gamma(M/2)}\right]^{1/M}$$

where **C** is a *M*-dimensional covariance matrix, det() is the determinant, $\chi_M^2(0.05)$ is the chi-square distribution, and $\Gamma()$ is the gamma function. Volumes are multiplicative in dimension *M*; the exponent 1/M places this volume on a per-trait basis.

Example Data

The example uses FIA data extracted from 23,916 forested non-plantation plots in 31 eastern states from FIADB version 5.1 on May 5, 2012 (<u>http://fia.fs.fed.us/</u>). All trees with a diameter at breast height (DBH) of at least 12.7 cm were measured on four 7.2 m radius subplots.

The predictors for the FIA data include climate variables, soils, and local drainage. Environmental predictors include a moisture index from the FIA and climate data extracted from the 800 m resolution Parameter-elevation Regressions on Independent Slopes Model (PRISM) data set (http://prism.nacse.org/) (Daly *et al.* 2009). I used climate data corresponding to each FIA plot to evaluate energy and moisture availability and stress. From mean monthly temperature (°C) and monthly precipitation (mm) I computed the hydrothermal deficit HTD (degree hours with a negative water balance, P < PET), and winter temperature (Clark et al. 2015). To evaluate collinearity I computed variance inflation factors (VIFs) and correlations presented in Table S1.

Stand age is recorded as part of the FIA census for each plot.

Slope and aspect are captured by three variables,

 $\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}' = \begin{bmatrix} sin(slope_i) \\ sin(slope_i)sin(aspect_i) \\ sin(slope_i)cos(aspect_i) \end{bmatrix}$

There is a main effect of slope (first element), but there can be no main effect of aspect independent of slope. If the first coefficient is not different from zero, then slope does not affect the response. If both of the last two coefficients are not different from zero then aspect does not affect the response (Clark 1990).

		temp	u1	u2	u3	stand age	moisture	deficit	UltKan	SpodHist	Mol	EntVert
	VIF	1.7	1.3	1.0	1.0	1.10	1.40	1.5	1.2	1.4	1.5	1.1
	factor								1	1	1	1
Correlation	u1	0.13										
	u2	0.01	0.01									
	u3	-0.04	-0.13	0.00								
	stand age	-0.19	0.23	0.00	-0.03							
	moisture	-0.03	-0.04	-0.01	0.01	0.02						
	deficit	0.49	-0.21	0.01	0.04	-0.24	0.03					
	UltKan	0.19	0.02	0.00	-0.03	-0.10	0.06	0.13				
	SpodHist	-0.48	-0.19	0.00	0.04	0.11	-0.04	-0.30	-0.10			
	Mol	-0.04	-0.03	-0.01	0.03	-0.02	0.10	0.13	-0.03	-0.08		
	EntVert	0.06	-0.16	0.00	0.05	-0.06	-0.01	0.13	-0.05	-0.14	0.04	

Table S1. Variance inflation and variable correlation in the design matrix.

Soil types constitute a factor with five levels, Entisol/Vertisol, Mollisol, Spodosol/Histosol, Ultisol-Udults-Kanhapludults (Piedmont Plateau soils), and remaining Ultisols. Data were extracted from the State Soil Geographic (STATSGO) database. For each FIA plot, soil types are categorized based on the first to third soil taxonomy orders. To limit the number of levels in the model similar first order classes were merged into these categories:

Entisol: poor profile development Vertisol: plastic clays Mollisol: former prairie rich soils Spodosol-Histosol: moist, organic, nutrient-poor, low pH, Ultisol-Udults-Kanhapludults: clay rich, dominant on the southeastern Piedmont

Trait data discussed here include wood density, seed mass, maximum height, leaf [N], leaf [P], specific leaf area, three leaf habit variables, and three ordinal tolerance variables. They are included as a Supplement file. Foliar traits were compiled from Aber and Martin (1999), Alonso et al. (2010), Beaudet, and Messier (2008), Chapin and Kedrowski (1983), Finzi et al. (2001), Gilmore et al. 1995, Grotkopp, and Rejmanek. 2007, LeBauer et al. (2010), Ricklefs, and Matthew (1982), Reich et al. (1998), Rieske (2002), USDA (2015), Van Sambeek et al. (2008), Vergutz et al. (2012), Walker et al. (2014). Leaf habit variables are categorical (0 or 1), including broadleaf deciduous, broadleaf evergreen, and needleleaf evergreen. Drought, shade, and flood tolerances are ordinal data (Niinemets and Valladares 2006, Russell et al. 2014). All others are continuous variables and were centered and standardized for analysis. Seed mass was first transformed to a log scale. Ninety-eight dominant tree species were included with remaining species in FIA gathered in a remainder category. Weights are species abundances, represented as biomass for each sample obtained by allometric conversion of diameter data (Schliep et al., 2015).

Computation in GJAM

Posterior simulation is done by Gibbs sampling, as detailed for generalized joint attribute modeling (GJAM) in Clark et al. (2014, 2015a). GJAM is available for R on <u>CRAN</u> and fully described <u>here</u>. Sampling the GJAM algorithm for the data types included in this model is summarized here.

Regression coefficients for the predictive trait model (PTM) are sampled from

$$\prod_{i=1} MVN(\mathbf{w}_i | \boldsymbol{\beta} \mathbf{x}_i, \boldsymbol{\Sigma}) \ MVN(vec(\boldsymbol{\beta}) | \boldsymbol{0}, \boldsymbol{\Sigma} \otimes \mathbf{C})$$

where prior covariance **C** is assumed to have infinite variance. The covariance matrix is updated from

$$\prod_{i=1}^{n} MVN(\mathbf{w}_{i}|\boldsymbol{\beta}\mathbf{x}_{i},\boldsymbol{\Sigma}) |\boldsymbol{\Sigma}|^{-(S+1)/2}$$

Latent \mathbf{w}_i are sampled from the truncated normal distribution subject to a partition that differs for each data type. For ordinal tolerance classes the partition is also sampled. Each is described <u>here</u>. For the trait response model (TRM) substitute \mathbf{u}_i , $\boldsymbol{\alpha}$, $\boldsymbol{\Omega}$, and M for \mathbf{w}_i , $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, and S.

Additional references

- Aber, J. D., and M. Martin. 1999. Leaf Chemistry, 1992-1993 (ACCP). [http://www.daac.ornl.gov] Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/421.
- Alonzo, A., González-Muñoz, N. & Castro-Díez, P. 2010. Comparison of leaf decomposition and macroinvertebrate colonization between exotic and native trees in a freshwater ecosystem. Ecological Research 25: 647–653.
- Augspurger, C. K., J. M. Cheeseman, and C. F. Salk. 2005. Light gains and physiological capacity of understory woody plants during phenological avoidance of canopy shade. Functional Ecology 19:537–546.
- Beaudet, M. and Messier, C. 2008. Beech regeneration of seed and root sucker origin: A comparison of morphology, growth, survival, and response to defoliation. Forest Ecology and Management, 255(10):3659-3666.
- Chapin III FS, and Kedrowski RA. 1983 . Seasonal changes in nitrogen and phosphorus fraction and autumn retranslocation in evergreen and deciduous Taiga trees. Ecology 64: 376–391.
- Clark, J.S. 1990. Landscape interactions among nitrogen mineralization, species composition, and long-term disturbance. Biogeochemisty 11:1-22.
- Clark, J.S., L. Iverson, C. W. Woodall, C. D. Allen, D. M. Bell, D. Bragg, A. D'Amato, F. W. Davis, M. Hersh, I. Ibanez, S. T. Jackson, S. Matthews, N. Pederson, M. Peters, M. W. Schwartz, K. Waring, and N. E. Zimmermann. 2015. The impacts of increasing drought on forest dynamics, structure, diversity, and management. In J. Vose, C. Luce, and J.S. Clark (eds). The National Assessment of Drought Impacts on Forests, in press.
- Daly C, Conklin DR, Unsworth MH (2009) Local atmospheric decoupling in complex topography alters climate change impacts. International Journal of Climatology, doi: 10.1002/joc.2007.
- Finzi, A.C., A.S. Allen, E.H. DeLucia, D.S. Ellsworth and W.H. Schlesinger. 2001. Aboveground litter production, chemistry, and decomposition following two years of free-air CO2 enrichment. Ecology 82(2):470-484.
- Gilmore WD, Seymour RS, Halteman WA, Greenwood MS. 1995. Canopy dynamics and the morphological development of Abies balsamea: effects of foliage age on specific leaf area and secondary vascular development. Tree Physiology 15: 47 -55.
- Grotkopp, E. and M. Rejmanek. 2007. High seedling relative growth rate and specific leaf area are traits of invasive species: phylogenetically independent contrasts of woody angiosperms. American Journal of Botany 94: 526-532.
- LeBauer, David; Dietze, Michael; Kooper, Rob; Long, Steven; Mulrooney, Patrick; Rohde, Gareth Scott; Wang, Dan; (2010): Biofuel Ecophysiological Traits and Yields Database (BETYdb); Energy Biosciences Institute, University of Illinois at Urbana-Champaign. http://dx.doi.org/10.13012/J8H41PB9

- Middleton, E. M., and J. H. Sullivan. 1998. BOREAS TE-10 Leaf Chemistry Data. Data set. Available on-line [http://www.daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/345.
- Niinemets, U., and Valladares, F., 2006. Tolerance to shade, drought, and waterlogging of temperate northern hemisphere trees and shrubs. Ecol. Monogr. 76, 521–547.
- Ricklefs, R. E., and K. K. Matthew. 1982. Chemical characteristics of the foliage of some deciduous trees in southeastern Ontario. Canadian Journal of Botany 60: 2037-2045.
- Reich, P.B., M.B. Walters, D.S. Ellsworth, J. Vose, J. Volin, C. Gresham, W. Bowman. 1998. Relationships of leaf dark respiration to leaf N, SLA, and life-span: a test across biomes and functional groups. Oecologia 114:471-482.
- Rieske, L.K. 2002. Wildfire alters oak growth, foliar chemistry, and herbivory. Forest Ecology and Management 168, 91-99.
- Russell, M.B., C.W. Woodall, A.W. D'Amato, G.M. Domke, and S.S. Saatchi. 2014. Beyond mean functional traits: influence of functional trait profiles on forest structure, production, and mortality across the eastern US. Forest Ecology and Management 328: 1-9.
- St-Jacques, C. and P. Bellefleur. 1991. Determining leaf nitrogen concentration of broadleaf tree seedlings by reflectance measurements Tree Physiol 1991 8: 391-398.
- Schliep, E.M., A.E. Gelfand, J.S. Clark, and K. Zhu. 2015. Modeling change in forest biomass across the eastern US. Environmental and Ecological Statistics, in press.
- Van Sambeek, J.W., Navarrete-Tindall, N. E., Hunt, K. L. 2008. Growth and foliar nitrogen concentrations of interplanted native woody legumes and pecan. In: Jacobs, Douglass F.; Michler, Charles H., eds. Proceedings, 16th Central Hardwood Forest Conference; 2008 April 8-9; West Lafayette, IN. Gen. Tech. Rep. NRS-P-24. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station: 580-588.
- Vergutz, L., S. Manzoni, A. Porporato, R.F. Novais, and R.B. Jackson. 2012. A Global Database of Carbon and Nutrient Concentrations of Green and Senesced Leaves. [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. http://dx.doi.org/10.3334/ORNLDAAC/1106

Walker, A.P., I. Aranda, A.P. Beckerman, H. Bown, L.A. Cernusak, Q.L. Dang, T.F. Domingues, L. Gu, S. Guo, Q. Han, J. Kattge, M. Kubiske, D. Manter, E. Merilo, G. Midgley, A. Porte, J.C. Scales, D. Tissue, T. Turnbull, C. Warren, G. Wohlfahrt, F.I. Woodward, and S.D. Wullschleger. 2014. A Global Data Set of Leaf Photosynthetic Rates, Leaf N and P, and Specific Leaf Area. Data set. [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1224.