



ECOLOGICAL SOCIETY OF AMERICA

Ecology/Ecological Monographs/Ecological Applications

PREPRINT

This preprint is a PDF of a manuscript that has been accepted for publication in an ESA journal. It is the final version that was uploaded and approved by the author(s). While the paper has been through the usual rigorous peer review process of ESA journals, it has not been copy-edited, nor have the graphics and tables been modified for final publication. Also note that the paper may refer to online Appendices and/or Supplements that are not yet available. We have posted this preliminary version of the manuscript online in the interest of making the scientific findings available for distribution and citation as quickly as possible following acceptance. However, readers should be aware that the final, published version will look different from this version and may also have some differences in content.

The doi for this manuscript and the correct format for citing the paper are given at the top of the online (html) abstract.

Once the final published version of this paper is posted online, it will replace the preliminary version at the specified doi.

628 Running head: α and β -diversity in a temperate forest

629

630 Title: **Species associations structured by environment and land-use history promote beta-**
631 **diversity in a temperate forest**

632

633 Authors: Stephen J. Murphy*¹, Lívia D. Audino², James Whitacre³, Jenalle L. Eck¹, John W.
634 Wenzel³, Simon Queenborough⁴, Liza S. Comita^{4,5}

635

636 ¹Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W.
637 12th Avenue, Columbus, OH 43210-1293, USA

638 ²Departamento de Entomologia, Universidade Federal de Lavras, Lavras, Minas Gerais 37200-
639 000, Brazil

640 ³Powdermill Nature Reserve, Carnegie Museum of Natural History, 1847 Route 381, Rector, PA
641 15677, USA

642 ⁴School of Forestry and Environmental Studies, Yale University, 195 Prospect Street, New
643 Haven, CT 06511

644 ⁵Smithsonian Tropical Research Institute, Box 0843-03092, Balboa, Ancón, Panama

645

646 *Author for correspondence

647

648 Email: murphy.1132@osu.edu; livia.audino@gmail.com; jvwhit@illinois.edu; eck.34@osu.edu;

649 WenzelJ@CarnegieMNH.Org; simon.queenborough@yale.edu; liza.comita@yale.edu

650 **Abstract**

651 Patterns of diversity and community composition in forests are controlled by a combination of
652 environmental factors, historical events, and stochastic or neutral mechanisms. Each of these
653 processes has been linked to forest community assembly, but their combined contributions to
654 alpha and beta-diversity in forests has not been well explored. Here we use variance partitioning
655 to analyze ~40,000 individual trees of 49 species, collected within 137 ha of sampling area
656 spread across a 900 ha temperate deciduous forest reserve in Pennsylvania to ask: 1) To what
657 extent is site-to-site variation in species richness and community composition of a temperate
658 forest explained by measured environmental gradients and by spatial descriptors (used here to
659 estimate dispersal-assembly or unmeasured, spatially-structured processes)? 2) How does the
660 incorporation of land-use history information increase the importance attributed to deterministic
661 community assembly? And 3) How do the distributions and abundances of individual species
662 within the community correlate with these factors? Environmental variables (i.e., topography,
663 soils, and distance to stream), spatial descriptors (i.e. spatial eigenvectors derived from Cartesian
664 coordinates), and land-use history variables (i.e., land-use type and intensity, forest age, and
665 distance to road), explained about half of the variation in both species richness and community
666 composition. Spatial descriptors explained the most variation, followed by measured
667 environmental variables and then by land-use history. Individual species revealed variable
668 responses to each of these sets of predictor variables. Several species were associated with
669 stream habitats, and others were strictly delimited across opposing north and south-facing slopes.
670 Several species were also associated with areas that experienced recent (i.e. < 100 years) human
671 land-use impacts. These results indicate that deterministic factors, including environmental and
672 land-use history variables, are important drivers of community response. The large amount of
673 'unexplained' variation seen here (about 50%) is commonly observed in other such studies

674 attempting to explain distribution and abundance patterns of plant communities. Determining
675 whether such large fractions of unaccounted for variation are caused by lack of sufficient data, or
676 are an indication of stochastic features of forest communities globally, will remain an important
677 challenge for ecologists in the future.

678

679 **Key words:** community assembly, environmental filtering, forest diversity, habitat association,
680 multivariate regression tree analysis, spatial dependency, spatial eigenvector, variance
681 partitioning.

preprint

682 **Introduction**

683 Determining the underlying mechanisms that govern patterns of species diversity and community
684 composition continues to pose a major challenge to ecologists, despite the fact that this theme
685 has a rich history in ecological research (Whittaker 1956, Bray and Curtis 1957, Whittaker 1960,
686 Baldeck et al. 2013b). While such studies have made clear that topographic and edaphic
687 gradients can drive spatial variation in species richness and turnover in community composition
688 in plants, their relatively low explanatory power and lack of consistency among communities
689 makes a cohesive framework for community assembly elusive (Lawton 1999, Hubbell 2001).
690 This lack of generality is highlighted, in part, by the popular niche vs. neutral debate (i.e.
691 deterministic vs. stochastic assembly) spurred by Hubbell's unified neutral theory (2001, Adler
692 et al. 2007, Clark 2012), as well as by the notable dearth of predictive models in community
693 ecology (Lawton 1999, Hubbell 2001, Simberloff 2004).

694 Researchers have, however, begun to make substantial progress in understanding spatial
695 patterns of diversity and composition by returning to Whittaker's concept of beta-diversity,
696 defined as the variation or turnover of species across space and/or time (Whittaker 1960,
697 Legendre et al. 2005, Anderson et al. 2011). The analysis of beta-diversity provides a
698 quantitative means of linking local (α -scale) and regional (γ -scale) processes, and can be
699 potentially useful for determining the relative importance of niche- and neutral-based
700 mechanisms in generating observed patterns of plant diversity and composition (Condit et al.
701 2002, Legendre et al. 2009, De Cáceres et al. 2012, Myers et al. 2013). For example, many
702 authors have attempted to isolate the roles of environmental variables and spatial descriptors
703 derived from geographic coordinates into individual and combined components, and then
704 determine how these data are correlated with variation in species composition or diversity (i.e.,
705 variance partitioning; e.g. Borcard et al. 1992, Legendre et al. 2009, Myers et al. 2013, De

706 Cáceres et al. 2012, Baldeck et al. 2013b). This dichotomy between environmental and spatial
707 mechanisms has largely been driven by the aforementioned niche vs. neutral debate, as spatial
708 autocorrelation left unexplained by the environment has commonly been attributed to ‘neutral’
709 mechanisms, specifically dispersal limitation (Legendre et al. 2009, Siefert et al. 2012, Hu et al.
710 2013). However, authors have recently warned that simply attributing the unique variation
711 explained by spatial descriptors to neutral mechanisms is unjustified, arguing that this fraction
712 cannot be readily distinguished from unmeasured, spatially structured, environmental variables
713 (Anderson et al. 2011). For example, Baldeck et al. (2013b) and Chang et al. (2013) showed that
714 the inclusion of soils data dramatically lowered the amount of variation explained by space, thus
715 reducing the importance accredited to neutral community assembly.

716 In addition to the turnover in communities, researchers have also evaluated the role that
717 environmental variables play in shaping the distributions of individual species. Specifically, a
718 number of studies have examined the proportion of species showing significant associations with
719 particular habitat types based on environmental variables such as topography or soils (Harms et
720 al. 2001, Baldeck et al. 2013a). Such species-specific studies support the idea that niche
721 partitioning plays an important role in structuring plant communities, and are often invoked as
722 evidence against neutral theory (Harms et al. 2001, Baldeck et al. 2013a).

723 While the current focus on environmental and spatial descriptors may be justified for
724 undisturbed old-growth stands, it ignores the impacts of past and current human land-use on
725 forest communities. For example, in the eastern United States, massive clear-cutting of forests
726 for commercial timber in the 19th century, and rapid changes in human land-use more recently,
727 have left the region a patchwork of stands with a history of differing disturbance types,
728 intensities, and ages (Whitney 1994). Ecologists have only recently begun to investigate the
729 impacts that such legacies have on closed canopy forest communities, and have concluded that

730 these factors can be just as important as the environment in determining patterns of plant
731 community composition and diversity (Motzin et al. 1996, Foster et al. 1998, Bürgi et al. 2000,
732 Foster et al. 2003). To date, many such studies have revealed convincing evidence for the effects
733 of human impacts on both forest diversity (Hermy 1994, Vellend 2004, Vellend et al. 2007,
734 Baeten et al. 2010) and on species composition (Foster 1992, Motzin et al. 1996, Bürgi et al.
735 2000, Thompson et al. 2013), and it has been shown that the legacies from such effects can be
736 long-lasting (Peterken and Game 1984, Dupouey et al. 2002, Foster et al. 2003, Koerner et al.
737 2007). Nonetheless, few studies have looked at the relative importance and interactive effects of
738 land-use history, environment, and spatial descriptors together at explaining patterns of alpha and
739 beta-diversity in forests (Svenning et al. 2004, López-Martínez et al. 2013). Likewise, the role
740 that past land-use history plays in shaping individual tree species distributions has been relatively
741 unexplored.

742 To help fill this gap, we combined land-use history data with environmental and tree
743 distribution data to determine the relative importance of each of these factors in explaining
744 spatial variation in composition and diversity in a mixed species temperate forest community in
745 southwestern Pennsylvania. To accomplish this, we analyzed data on ~40,000 individual adult
746 trees of 49 species, collected within a 137 ha sampling area spread across a 900 ha forest reserve.
747 The environment was defined using topographic and soils data, and information on land-use
748 history and stand age was obtained from historical aerial images. Spatial descriptors were
749 derived from principal coordinates of neighbor matrices (PCNM) (Borcard and Legendre 2002).
750 Specifically, we asked: 1) To what extent is site-to-site variation in species richness and
751 community composition of a temperate forest explained by measured environmental gradients
752 and by spatial predictors (used to estimate stochastic assembly processes and/or unmeasured
753 environmental variables)? 2) How does the incorporation of past land-use history information

754 and forest age increase the level of importance attributed to deterministic community assembly
755 in an anthropogenically impacted system? And 3) How do the distributions and abundances of
756 individual species within the community correlate with these factors?

757 **Materials and Methods**

758 *Field Site*

759 The study was conducted at Powdermill Nature Reserve (PNR), an approximately 900 ha
760 temperate deciduous forest located in southwestern Pennsylvania (Westmoreland County;
761 40°09'N, 79°16'W). The area lies within the mixed mesophytic vegetation region of the
762 Allegheny plateau at the base of the Appalachian Mountains, and is composed primarily of
763 maple (*Acer* spp.), tulip poplar (*Liriodendron tulipifera*), and other mesophytic species, as well
764 as a variety of oaks (*Quercus* spp.) and hickories (*Carya* spp.). The area receives approximately
765 1100 mm of precipitation annually, and temperatures can range from a low of -20°C in January
766 to a high of 33°C in July (NCDC 2013, based on years 2001 - 2009). Elevation ranges from 392
767 to 647 meters, and slope steepness ranges from 0 to 24 degrees. PNR is predominately forested,
768 but also contains a mix of maintained fields, human development, and other managed lands. The
769 entire region was likely logged in the 19th century (Whitney 1994), and much of PNR was used
770 for agriculture until the early to mid 20th century. Surface mining operations were also conducted
771 in areas of the reserve in the 1930's and 1940's. The area was designated a nature reserve in
772 1956 by the Carnegie Museum of Natural History and has since experienced only minor
773 anthropogenic disturbances. Today, the landscape directly to the west of PNR is moderately
774 developed, while the area to the east is state forestland.

775 *Data Collection*

776 In 2006, the entire 900 ha reserve was systematically mapped onto a grid using a 120 × 120 m
777 cell size (referred to hereafter as “blocks”). A total of 647 blocks encompassed the total area of

778 PNR. In the centers of each block, nine adjacent 10-m radius ($\sim 314 \text{ m}^2$) circular plots were
779 established for vegetation sampling, for a total of 5823 plots (Appendix A). During the summers
780 of 2007 and 2008, field crews identified and measured all live and dead stems ≥ 8 cm diameter at
781 breast height (DBH; 1.37 m above ground). Thus, the entire PNR dataset encompasses a total
782 sampling area of ~ 183 ha, with a total stem count of nearly 45,000 individual trees. Because not
783 all plots sampled were forested, only blocks containing nine plots with at least one stem ≥ 8 cm
784 DBH were used in the analysis, resulting in a total of 485 blocks, which we used as our statistical
785 sampling unit. Furthermore, 5224 dead stems and 328 individuals identified only to the genus
786 level (other than *Crataegus* spp. and *Malus* spp.) were excluded from the analysis, resulting in a
787 final stem count of 37,258 adult trees within 137 ha of sampling area.

788 Environmental variables used in our analysis included topography and soil type. GIS-
789 derived topographic variables included the continuous variables elevation, slope, and convexity
790 (measured as the center block's elevation minus the mean elevation of the surrounding eight
791 blocks; Legendre et al. 2009). Slope aspect, originally measured in degrees, was converted to six
792 categorical units (i.e., Northeast, East, Southeast, Southwest, West, and Northwest) to avoid
793 problems associated with a circular statistic being treated as linear. Each of these variables was
794 derived in GIS using a digital elevation model (DEM) based on Landsat data using a 5-meter
795 scale resolution (PAMAP 2008; See Appendix B for a detailed hill-shade relief map of PNR).
796 Because topography itself is unlikely to influence forest trees, these variables were primarily
797 used as proxy variables for soil moisture and microclimatic variables. We also measured nearest
798 distance to Powdermill Run, a 3rd-order stream running through the center of the reserve
799 (Raikow et al. 1995). Finally, data on soil type were derived from soil survey maps obtained
800 from publicly available records (Appendix C)(NRCS 2013).

801 Land-use history and forest age were quantified using historical aerial imagery.
802 Photographs were available for years 1939, 1957, 1967, 1993, and 2006. For each photograph,
803 land-use type was visually classified at each of the 5823 plots using a modification of the
804 standard national land cover database classification system (Anderson et al. 1976). This
805 classification system was used to identify plots as forest, shrub land, cultivated, developed, or
806 mine land (used here to denote areas where surface mining activity occurred). Cultivated areas
807 included both crop and pastureland, and were easily identified based on spatial configuration and
808 the presence of plow-lines. Developed areas were identified as those where human structures
809 such as roads, building, and parking lots were found (See Anderson et al. 1976 for further
810 details). This classification was conducted for each of the five years for which image data were
811 available. These data were then used to assign estimates of forest age (an endogenous
812 characteristic due to succession) and land-use type to each block (see Appendix D). For the few
813 areas of the reserve that experienced more than one land-use impact, the most recent land-use
814 type and age was used for the analysis. Furthermore, for the few blocks in which the nine
815 replicate plots did not all have the same land cover category, the most dominant type was used.
816 Additionally, the total proportion of non-forested plots since 1939 out of nine within each block
817 was used as an estimate of the intensity of disturbance. Finally, distance to nearest paved road
818 was calculated in GIS and included in our suite of land-use variables.

819 *Statistical Analyses*

820 To visualize the spatial variation in species richness, we mapped the total number of species per
821 block across the reserve (Fig. 1A). Likewise, a 'beta-diversity map' of PNR was generated to
822 summarize variation in community composition. To do this, we performed non-metric
823 multidimensional scaling on the community composition data using Bray-Curtis distances, and
824 then used the scores from the first three axes to assign R.G.B. color values to individual blocks

825 (Thessler et al. 2005, Baldeck et al. 2013b). The resulting image can thus be viewed as a color-
826 map, in which blocks of more similar color have a more similar species composition (Fig. 1B).

827 Next, we used variance partitioning to isolate the roles of environment, land-use history,
828 and spatial descriptors in explaining heterogeneity in the richness and composition of PNR
829 (Borcard et al. 1992, Peres-Neto et al. 2006, Legendre 2008). Based on partial redundancy
830 analysis for multivariate response data (e.g. species composition) and on partial multiple linear
831 regression for univariate response data (e.g. species richness), variance partitioning is ideal for
832 isolating the pure and combined explanatory power of suites of predictor variables (Borcard et al.
833 1992; Legendre, 2008). We were interested in how much variation in community composition
834 and species richness can be explained by the environment, spatial descriptors, and by land-use
835 history alone, and the degree to which these three predictor groups are redundant. Community
836 data were Hellinger-transformed prior to the analysis (Legendre and Gallagher 2001).

837 As described above, environmental variables included topography (elevation, slope,
838 aspect, and convexity), soil type (Appendix C), and distance to stream; and land-use history
839 variables included land-use type (i.e., cultivated, developed, mine land, shrub land, and forested
840 since 1939), forest age, proportion of plots disturbed, and distance to nearest road (Appendix E).
841 Spatial predictors were computed using principle coordinates of neighbor matrices (PCNM)
842 (Borcard and Legendre 2002). By extracting the positive eigenvectors from a truncated distance
843 matrix using principle coordinates analysis, PCNM provides highly detailed estimates of spatial
844 structure at a variety of scales. Therefore, individual eigenvectors can be viewed as spatial map-
845 patterns, and can easily be used as predictor variables in an appropriate linear model or
846 multivariate procedure (Griffith and Peres-Neto 2006). Because a large number of eigenvectors
847 will result when large sample sizes are used, a permutational forward selection procedure was

848 used to identify those eigenvectors that are significantly correlated with the richness and
849 community composition data (Peres-Neto et al. 2006).

850 Next, we tested whether individual species distributions at PNR were associated with the
851 environmental, spatial, and land-use history variables described above. Environmental variables
852 were used to measure ‘true’ habitat associations, while land-use history variables were used to
853 measure whether individual species were associated with a particular historical land-use type.
854 The spatial PCNM eigenvectors, while not based on any measured environmental variables, were
855 used to model the structure in the community composition data (sensu Baldeck et al. 2013a). In
856 this case, significant species associations represent either correlations with spatially-structured
857 environmental factors (i.e. true habitat associations), or spatial aggregation due to dispersal
858 limitation (Anderson et al. 2011, Baldeck et al. 2013a, 2013b). For the environmental and spatial
859 variables, habitats were delimited using multivariate regression tree analysis (De’Ath 2002). This
860 method uses an iterative clustering algorithm that partitions the composition data into maximally
861 dissimilar groups based on break points in the environmental variables. For land-use history, the
862 land-use types were used. Species associations with each environmental, spatial, and land-use
863 category were then assessed using indicator species analysis (Dufrêne and Legendre 1997).

864 All GIS data were derived using ArcGIS v. 10.1 (ESRI 2013) and QGIS v. 2.0 (QGIS
865 Development Team 2013), and all statistical analyses were performed using R v. 3.01 (R Core
866 Development Team 2013). All statistical tests were performed at a significance level of $\alpha = 0.05$.
867 The PCNM and variance partitioning analyses were implemented using the functions ‘pcnm’ and
868 ‘varpart’ in the vegan package, respectively (Oksanen et al. 2013). Forward selection using
869 permutation was carried out using the packfor package (Dray et al. 2011). Finally, multivariate
870 regression tree analysis was performed using the mvpart package (De’ath 2006), and indicator
871 species analysis using the labdsv package (Roberts 2013).

872 **Results**873 *Tree diversity and distribution patterns*

874 A total of 49 tree species from 18 families were recorded at PNR. *Acer rubrum* was the most
875 abundant species, encompassing more than 25% of all sampled individuals and appearing in
876 ~98% of the 485 forested blocks used in the analysis. Other mesophytic species such as *Acer*
877 *saccharum*, *Liriodendron tulipifera*, *Prunus serotina*, and *Betula lenta* were also abundant,
878 making up 13.3%, 11.1%, 8.0%, and 7.0% of individuals, respectively (Appendix F). Xeric-
879 tolerant species such as oaks (*Quercus* spp.) and hickories (*Carya* spp.) were much less common,
880 typically representing less than 5% of all individuals (Appendix F).

881 Species richness was generally greatest along Powdermill Run stream (Fig. 1A).
882 Distribution and abundance patterns varied widely amongst the 49 species sampled (Appendix
883 G). This heterogeneity in species' distributions resulted in substantial variation in community
884 structure, as evidenced by the dissimilar community types revealed on the beta-diversity map
885 (Fig. 1B). For example, the dark green areas are associated with high abundances of *A.*
886 *saccharum* and *L. tulipifera*. The pink area to the northeast of the stream can be distinguished
887 primarily by the high abundances of *A. rubrum*, *B. lenta*, and *Nyssa sylvatica*, and to a lesser
888 extent by *Magnolia acuminata* and *Sassafras albidum*.

889 *Variance Partitioning*

890 Environment, space, and land-use history together explained 52% of the variation in
891 species richness at PNR (Fig. 2A). Environmental variables alone explained a total of 30% of the
892 variation, and land-use history explained a total of 9%. However, the spatial descriptors
893 explained more variation than either environment or land-use history (45%), based on 44
894 eigenvectors that were retained through forward selection (Appendix H). In addition to
895 accounting for the largest amount of total variation, spatial descriptors also explained the largest

896 amount of residual variation after accounting for land-use history and environment (20%). In
897 contrast, environment and land-use history each accounted for 4% and 1% of unique variation
898 explained, respectively. This low percentage largely reflects the fact that environmental
899 variables, as well as patterns of land-use history, tend to be spatially structured, such that we
900 might expect most of the variation explained by these factors to overlap with variation explained
901 by space.

902 A total of 50% of the variation in community composition was accounted for in our
903 model (Fig. 2B). Again, environment explained more variation (29%) than land-use history
904 (5%). Eighty-five spatial eigenvectors were selected, and these again explained the most total
905 variation by far (45%), as well as the most variation left unexplained by any of the other
906 variables (19%). Environment and land-use history variables accounted for just 3% and 1% of
907 unique variation, respectively, which again reflects the strong spatial signal in these variables.

908 When looking at the individual environmental variables alone, soil type accounted for the
909 largest percentage of variation in both species richness (21%) and composition (19%) (Appendix
910 E). Topography accounted for slightly less, with elevation being the most important, explaining
911 nearly 10% of the variation in both cases. The three remaining topographic variables (i.e., aspect,
912 slope, convexity) accounted for a total of about 8% of the variation in species richness, and 10%
913 in community composition. Distance to stream was also important, accounting for 9% and 6% of
914 the variation in species richness and composition, respectively.

915 Land-use type and proportion of plots disturbed since 1939 were the most important land-
916 use history variables for both species richness and species composition, accounting for 3-6% of
917 the variation in both cases (Appendix E). Forest age was relatively unimportant in both cases
918 (non-significant for species richness, ~2% for composition). Similarly, distance to nearest paved
919 road was relatively unimportant (< 2%).

920 *Species habitat associations*

921 Multivariate regression tree (MVRT) analysis using environmental variables (i.e.,
922 topography and soils) and spatial variables (i.e., PCNMs) resulted in seven and six habitat
923 categories, respectively (Table 1, Appendix I). The most important partition in the MVRT
924 analysis using measured environmental variables was for stands above and below 500 m
925 elevation, followed by a soils split, an aspect split, and a slope split (Table 1; Appendix I). For
926 MVRT based on PCNMs, categories were delimited primarily through use of the smallest
927 eigenvectors (particularly PCNM4), indicating that large-scale gradients (> 1000 meters) were
928 most important at delimiting habitat types (Table 1, Appendix H). PCNM15 was also included in
929 the model, suggesting that some intermediate scale (i.e., < 500 meters) patchiness is also
930 important (Appendix H). In terms of the sizes and locations of habitat delimitations, results were
931 highly similar (based on visual inspection) when using actual measured environmental properties
932 versus PCNM eigenvectors (see Appendix I for habitat maps).

933 Thirty-four of the 49 (69%) tree species at PNR were significantly associated with at least
934 one ‘habitat’ type defined by one of the three predictor groups (i.e., environment, spatial
935 descriptors, or land-use history). Indicator species analysis resulted in 22 (45%) significant
936 associations with the environment (i.e., true habitat associations), 26 (53%) with spatial
937 ‘habitats’ (i.e., groupings based on PCNMs), and 14 (29%) with at least one of the five land-use
938 categories (Table 2). Most species were significantly associated with more than one of these
939 habitat groupings (i.e., multiple columns in Table 2). However, three species were only
940 associated with environment-defined habitats, five to spatial-defined habitats, and four to land-
941 use history-defined habitats (Table 2).

942 Most true habitat associations based on measured environmental variables were for those
943 species associated with higher elevation areas (> 500 m) and soil types shown in table 1 (Groups

944 E1-E5). This partition likely reflects habitat differences between ridge top and valley slope
945 positions. This included three species each of *Acer* spp. and *Quercus* spp., two species of *Betula*
946 spp., as well as *Castanea dentata*, *Fagus grandifolia*, *L. tulipifera*, *N. sylvatica*, *P. serotina*, *S.*
947 *albidum*, *Tilia americana*, and *Tsuga canadensis* (Table 2). Species associated with groups E1
948 versus E2-E5 were segregated based on soil types that seemed to be distinguished by depth to
949 water table and by soil bulk density (Table 2, Appendix C). Soils associated with group E1
950 typically had larger depth to water table and soil bulk density values than soils associated with
951 groups E2-E5. Next, species were segregated based on slope steepness and then by aspect. *Acer*
952 *pennsylvanicum*, *A. saccharum*, *F. grandifolia*, *L. tulipifera*, and *T. canadensis* were all
953 associated with steeper north-facing slopes (Table 1 and 2, group E5). In contrast, all three
954 species of *Quercus* spp. were associated with south-facing slopes (group E3). Finally, the few
955 species that were found in greater abundances at lower elevation areas (groups E6 and E7)
956 included *Carya glabra*, *Carya ovata*, *Cornus florida*, and *Fraxinus americana*.

957 Results based on the PCNM eigenvectors were roughly similar to those found when using
958 measured environmental variables, although a greater total number of significant associations
959 were discovered (26 vs. 22). Fourteen species were significantly associated with land-use history
960 categories. Previously developed areas had a significantly higher proportion of *Acer*
961 *saccharinum*, *A. saccharum*, *Juglans nigra*, *L. tulipifera*, and *T. canadensis*. Previously mined
962 areas were associated with *C. glabra* and *C. ovata*, as well as with *Platanus occidentalis*. Lastly,
963 agricultural areas had a higher abundance of *Crataegus* spp., *Malus* spp., and *Robinia* spp.

964 Discussion

965 With the advent of new statistical tools, an increasing number of studies have quantified
966 the role of environment and spatial dependency in structuring plant species assemblages.
967 However, few of these studies have included the role of history, particularly the impact of

968 anthropogenic disturbances. Overall, our model was able to explain a relatively large percentage
969 of the variation in both species richness and community composition (about 50%). This was
970 surprising given the fact that the community at PNR is relatively homogenous given the very
971 large abundance of *A. rubrum*, which occurred in 98% of our plots. Our results demonstrate the
972 importance of deterministic processes, including land-use history, in shaping temperate mixed-
973 species forest communities at a local- to landscape-scale, but also provide evidence that spatial
974 patterns of tree diversity and community composition at the site might be influenced by dispersal
975 limitation.

976 *Role of environmental variables*

977 We found that tree species at PNR were sorted primarily along topographic and edaphic
978 gradients. Elevation was most important: *Carya* spp., *Cornus* spp., and *Fraxinus americana* were
979 all associated with sites below 500 meters elevation (i.e. valley positions), and *Acer* spp.,
980 *Quercus* spp., and *Betula* spp. were associated with areas of higher elevation (i.e. ridge-tops).
981 Furthermore, species such as *A. saccharum* and *T. canadensis* were strongly associated with
982 areas along Powdermill Run stream, likely due to increased soil moisture conditions in these
983 areas. There is also a striking contrast in community composition along the northeast and
984 southwest sides of Powdermill Run stream in the lower half of the reserve (i.e., green vs. pink
985 colors in Fig. 1B). This pattern likely arises from microclimatic variation between northerly- and
986 southerly-facing slopes (Boerner 2006). For example, *A. rubrum* is highly abundant on the more
987 southerly-facing slopes on the northeast side of the stream, but completely absent on the
988 southwest side (Appendix G). In contrast, *A. saccharum* showed the opposite pattern, being very
989 dominant on the northeast-facing slopes south of Powdermill Run, but was almost never found
990 on the southerly-facing slopes on the northeast side of the stream (Appendix G). Cooler and
991 moister conditions on these northeast-facing slopes are likely promoting the dominance of the

992 more mesophytic *A. saccharum* over *A. rubrum*. Similarly, *L. tulipifera* and *F. grandifolia* were
993 associated with the cooler and moister conditions, while all three *Quercus* spp. were found more
994 often on the drier south-facing slopes (Table 1 & 2, Appendix G).

995 We found soil type to be an important determinant of richness and composition at PNR.
996 This result is similar to those of Baldeck et al. (2013b) for eight tropical forests, as well as of
997 Chang et al. (2013) in a sub-tropical forest in Taiwan, although their soils data were from field
998 collected samples. In particular, the species in our study seemed to sort most strongly according
999 to soil types with high and low soil bulk densities and depths to water table (Appendix C). Both
1000 of these variables are associated with soil moisture capacity, indicating that water availability is
1001 probably playing a key role in governing local patterns of species distributions at PNR. However,
1002 because the soils data were derived from survey maps, direct chemical and physical
1003 measurements of field samples at the exact scale of our sampling design were not available. It
1004 should also be noted that these data represent an amalgamation of different properties (i.e.,
1005 chemical, physical, and topographic), and are not a univariate gradient like the other
1006 environmental variables used in this analysis. Therefore, some care should be taken when
1007 comparing the relative contributions of soil type with the topographic variables. It is likely that
1008 finer-scale soils data (e.g. field collected within each block) would help to paint a more nuanced
1009 picture of how edaphic properties influence tree species abundance distributions at PNR.
1010 Likewise, more detailed environmental information on solar irradiance, water availability,
1011 microclimate, etc., would likely further increase the fraction of variation attributed to niche-
1012 assembly (Baldeck et al. 2013b).

1013 *Role of space in shaping species richness and composition*

1014 The spatial descriptors explained a large fraction of the variation in both the species
1015 richness and community composition data. Almost all of the variation explained by measured

1016 environmental data was also co-explained by the spatial descriptors, indicating that the
1017 environmental variables are strongly spatially-structured, and that the spatial descriptors
1018 themselves are good approximators of real gradients. The spatial descriptors also explained the
1019 largest fraction of unique variation by far, suggesting that other processes besides those explicitly
1020 measured are important drivers of site-to-site variation in richness and community composition
1021 at PNR. For example, dispersal limitation could be driving aggregation in species patterning, or
1022 there may be additional spatially structured variables yet missing from the model.

1023 Wide variations in the relative contributions of spatial descriptors have been reported in
1024 the literature (Legendre et al. 2009, De Cáceres et al. 2012, Siefert et al. 2013, Myers et al. 2013,
1025 Baldeck et al. 2013b). The total amount of variation explained by the spatial descriptors can vary
1026 with plot size (Legendre et al. 2009, De Cáceres et al. 2012), the spatial configuration of the
1027 sampling design (Borcard and Legendre 2002), and with the number of eigenvectors used in the
1028 model (Gilbert and Bennett 2010). Likewise, the amount of unique variation explained by space
1029 (i.e. not co-explained by the environment) varies with environmental heterogeneity (De Cáceres
1030 et al. 2012), the quality of environmental variables used (Baldeck et al. 2013b, Chang et al.
1031 2013), and the spatial distribution of the environmental gradients (Smith and Ludlom 2010).
1032 Myers et al. (2013) suggested that latitude is a driver of the importance attributed to spatial
1033 predictors, with a greater proportion of variation explained by space in tropical over temperate
1034 forests. However, their lack of replication within temperate and tropical regions makes this
1035 relationship somewhat tenuous, and the fraction of variation that they attribute to dispersal-
1036 assembly could, again, be due to unmeasured environmental variables. For example, De Cáceres
1037 et al. (2012) found that the total amount of topographic relief within a plot was a stronger
1038 determinant of beta-diversity, and that the total fraction explained by their spatial descriptors was
1039 not strongly correlated with latitude. Ultimately, determining the true underlying mechanisms

1040 generating spatially autocorrelated diversity and composition patterns that are uncorrelated with
1041 measured variables will remain an important, albeit challenging, goal in future efforts.

1042 *Impacts of human disturbance*

1043 A major goal of this study was to determine how historical land-use impacts influence
1044 diversity and composition of an eastern deciduous forest. While focusing on land-use history is
1045 not new in plant ecological research, it has received minimal attention in a beta-diversity or
1046 community assembly framework (Vellend 2004, Velland et al. 2007, Arroyo-Rodríguez et al.
1047 2013). This is despite the fact that the concept of beta-diversity is often deemed important for
1048 conservation issues (Condit et al. 2002, Myers et al. 2013). Our results indicate that land-use
1049 history had a small but significant influence on the distribution of trees at PNR, explaining nine
1050 and five percent of the total variation in species richness and composition, respectively. Much of
1051 this variation was co-explained by both environmental and spatial predictors (Fig. 2). This is not
1052 surprising, due to the fact that historical land-use impacts are often correlated with environmental
1053 variables (Motzin et al. 1996, Foster et al. 2003). For example, flatter areas are easier to log and
1054 farm than steep slopes. The low amount of variation explained by land-use history may be an
1055 indication of successional convergence, where abandoned stands progress to a particular species
1056 composition based on time and environment, but is independent of the particular land-use
1057 impact. However, the relatively weak relationship with stand age found for both richness and
1058 composition (Appendix E) does not seem to strongly support this hypothesis. Furthermore,
1059 blocks that experienced some form of land-use since 1939 made up only 14% of the 485 blocks
1060 used in the analysis (Appendix D), making it unsurprising that these variables explained a
1061 relatively small percentage of the total variation. Overall, we expect that given a larger
1062 proportion of disturbed area at the site, or historical data prior to 1939, the importance attributed
1063 to land-use history and forest age would be even greater. Furthermore, we also looked only at

1064 anthropogenic impacts and not at natural disturbances such as wind or fire, which likely also
1065 have a strong influence on community assembly.

1066 A total of 14 species (29%) were significantly associated with one of the four land-use
1067 history categories based on indicator species analysis (Table 2). The majority of these were
1068 species typically associated with early successional or disturbed habitats, such as *Crataegus* spp.,
1069 *Malus* spp., and *Robinia pseudoacacia* (Burns and Honkala 1990), all of which were associated
1070 with previously cultivated areas (Table 2). Interestingly, however, later successional species such
1071 as *A. saccharum*, *T. canadensis*, *Carya* spp. and *Quercus rubra* also showed positive associations
1072 with land-use history variables (Table 2). For example, *C. glabra* and *C. ovata* were associated
1073 with previously mined lands, *A. saccharum* and *T. canadensis* with previously developed areas,
1074 and *Q. rubra* with recent shrub land areas. However, the positive association of *T. canadensis*
1075 and *A. saccharum* with previously developed areas was due to just a single block in both cases.
1076 Each of these blocks were located near an old structure where these species may have been
1077 spared from cutting and/or released from neighborhood competition. No evidence of intentional
1078 planting was observed in either case, although this remains a possibility. When the blocks were
1079 removed from the analysis, neither species remained an indicator of developed areas.

1080 *Residual variation in species richness and composition*

1081 Lastly, about half of the variation in both species richness and community composition
1082 was unaccounted for in our model (Fig. 2). This large fraction of unexplained variation is
1083 commonly observed in plant community ecology studies, and was actually lower in the present
1084 study than in many others (Young and Peacock 1992, He et al. 1996, Lawton 1999, De Cáceres
1085 et al. 2012, Siefert et al. 2013, Myers et al. 2013). If relevant, non-redundant variables (e.g. soils
1086 data, land-use history) are missing from analyses of species diversity and composition, their
1087 inclusion would likely help to reduce this unexplained fraction (Baldeck et al. 2013b, Chang et

1088 al. 2013). Furthermore, it is also likely that smaller individuals (i.e. seedlings and saplings)
1089 partition environmental niches in more unique ways than adult trees, information that would be
1090 unaccounted for in studies looking at only larger individuals (e.g. > 8 cm DBH). In contrast, if
1091 stochastic processes or ecological drift are important at governing community assembly, then
1092 this unexplained variation is an important indication for their occurrence. Likewise, if
1093 communities are not at competitive equilibrium because of natural disturbances or climate
1094 change, then past conditions or events will leave a spatial signature that will be uncorrelated with
1095 present day measurements if those conditions have changed over time (Grubb 1977, Pickett
1096 1980, He et al. 1996, Chesson 2000). Therefore, a better understanding of the role of past events
1097 in structuring present day ecological communities may help to alleviate this issue (Ricklefs
1098 1987). Ultimately, determining whether such large fractions of unaccounted for variation are
1099 caused by a lack of sufficient data, or are simply an implicit feature of forest communities
1100 globally remains an important challenge for ecologists in the future.

1101 **Acknowledgements**

1102 The vegetation survey was made possible through a grant from the Heinz Endowments. We
1103 thank the field crews who conducted the survey, and Matthew Silveira for helping with
1104 Geographic Information Systems work. LDA thanks Capes for a doctorate scholarship (sandwich
1105 program - PDSE 11219/12-2).

1106 **Literature Cited**

- 1107 Adler, P. B., J. HilleRisLambers, and J. M. Levine. 2007. A niche for neutrality. *Ecology Letters*
1108 10:95–104.
- 1109 Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer. 1976. A land use and land cover
1110 classification system for use with remote sensor data. U.S. Geological Survey,
1111 Washington DC, USA.
- 1112 Anderson, M. J., T. O. Crist, J. M. Chase, M. Vellend, B. D. Inouye, A. L. Freestone, N. J.
1113 Sanders, H. V. Cornell, L. S. Comita, K. F. Davies, S. P. Harrison, N. J. B. Kraft, J. C.
1114 Stegen, and N. G. Swenson. 2011. Navigating the multiple meanings of β diversity: a
1115 roadmap for the practicing ecologist. *Ecology Letters* 14:19–28.
- 1116 Arroyo-Rodríguez, V., M. Rös, F. Escobar, F. P. L. Melo, B. A. Santos, M. Tabarelli, and R.
1117 Chazdon. 2013. Plant β -diversity in fragmented rain forests: testing floristic
1118 homogenization and differentiation hypotheses. *Journal of Ecology* 101:1449–1458.
- 1119 Baeten, L., M. Hermy, S. Van Daele, K. Verheyen. 2010. Unexpected understorey community
1120 development after 30 years in ancient and post-agricultural forests. *Journal of Ecology*
1121 98:1447–1453.
- 1122 Baldeck, C. A., K. E. Harms, J. B. Yavitt, R. John, B. L. Turner, R. Valencia, H. Navarrete, S.
1123 Bunyavejchewin, S. Kiratiprayoon, A. Yaacob, M. N. N. Supardi, S. J. Davies, S. P.
1124 Hubbell, G. B. Chuyong, D. Kenfack, D. W. Thomas, and J. W. Dalling. 2013a. Habitat
1125 filtering across tree life stages in tropical forest communities. *Proceedings of the Royal*
1126 *Society B* 280:20130548.
- 1127 Baldeck, C. A., K. E. Harms, J. B. Yavitt, R. John, B. L. Turner, H. Navarrete, S. J. Davies, G.
1128 B. Chuyong, D. Kenfack, W. Duncan, S. Madawala, N. Gunatilleke, S. Gunatilleke, S.
1129 Bunyavejchewin, S. Kiratiprayoon, A. Yaacob, M. N. N. Supardi, J. W. Dalling, R.

- 1130 Valencia, and D. W. Thomas. 2013b. Soil resources and topography shape local tree
1131 community structure in tropical forests. *Proceedings of the Royal Society B*
1132 280:20122532.
- 1133 Boerner, R. E. J. 2006. Unraveling the Gordian knot: interactions among vegetation, topography,
1134 and soil properties in the central and southern Appalachians. *The Journal of the Torrey*
1135 *Botanical Society* 133:321–361.
- 1136 Borcard, D., and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of
1137 principal coordinates of neighbour matrices. *Ecological Modelling* 153:51–68.
- 1138 Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of
1139 ecological variation. *Ecology* 73:1045–1055.
- 1140 Bray, J., and J. Curtis. 1957. An ordination of the upland forest communities of southern
1141 Wisconsin. *Ecological Monographs* 27:325–349.
- 1142 Bürgi, M., E. Russell, G. Motzkin. 2000. Effects of postsettlement human activities on forest
1143 composition in the north- eastern United States: a comparative approach. *Journal of*
1144 *Biogeography* 27:1123–1138.
- 1145 De Cáceres, M., P. Legendre, R. Valencia, M. Cao, L. Chang, G. Chuyong, R. Condit, Z. Hao, C.
1146 Hsieh, S. P. Hubbell, D. Kenfack, K. Ma, X. Mi, M. N. Supardi Noor, A. R. Kassim, H.
1147 Ren, S. Su, I.-F. Sun, D. Thomas, W. Ye, and F. He. 2012. The variation of tree beta
1148 diversity across a global network of forest plots. *Global Ecology and Biogeography*
1149 21:1191–1202.
- 1150 Dupouey, J., E. Dambrine, J. Laffite, and C. Moares. 2002. Irreversible impact of past land use
1151 on forest soils and biodiversity. *Ecology* 83:2978–2984.
- 1152 Chang, L., D. Zeleny, C. Li, S. Chiu, and C. Hsieh. 2013. Better environmental data may reverse
1153 conclusions about niche- and dispersal-based processes in community assembly. *Ecology*

- 1154 94:2145–2151.
- 1155 Chesson, P. 2000. Mechanisms of maintenance of species diversity. *Annual Review of Ecology*
1156 *and Systematics* 31:343–366.
- 1157 Clark, J. S. 2012. The coherence problem with the unified neutral theory of biodiversity. *Trends*
1158 *in Ecology & Evolution* 27:199–203.
- 1159 Condit, R., N. Pitman, E. G. Leigh, J. Chave, J. Terborgh, R. B. Foster, P. Núñez, S. Aguilar, R.
1160 Valencia, G. Villa, H. C. Muller-Landau, E. Losos, and S. P. Hubbell. 2002. Beta-
1161 diversity in tropical forest trees. *Science* 295:666–669.
- 1162 De’Ath, G. 2002. Multivariate regression trees: a new technique for modeling species-
1163 environment relationships. *Ecology* 83:1105–1117.
- 1164 De’Ath, G. 2006. mvpart: multivariate partitioning. R Package Version 1.2–4.
1165 <http://cran.r-project.org/i>.
- 1166 Dray, S. 2011. packfor: forward selection with permutation (Canoco p.46). R package version
1167 0.0-8. <http://R-Forge.R-project.org/projects/sedar/>.
- 1168 Dufrière, M., and P. Legendre. 1997. Species assemblages and indicator species: the need for a
1169 flexible asymmetrical approach. *Ecological Monographs* 67:345–366.
- 1170 ESRI 2013. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research
1171 Institute.
- 1172 Foster, D., G. Motzkin, and B. Slater. 1998. Land-use history as long-term broad-scale
1173 disturbance: regional forest dynamics in central New England. *Ecosystems* 1:96–119.
- 1174 Foster, D., F. Swanson, J. Aber, and I. Burke. 2003. The importance of land-use legacies to
1175 ecology and conservation. *BioScience* 53:77–88.
- 1176 Gilbert, B. and J. R. Bennett. 2010. Partitioning variation in ecological communities: do the
1177 numbers add up? *Journal of Applied Ecology* 47:1071–1082.

- 1178 Griffith, D. A., and P. R. Peres-Neto. 2006. Spatial modeling in ecology: the flexibility of
1179 eigenfunction spatial analyses. *Ecology* 87:2603–2613.
- 1180 Grubb, P. J. 1977. The maintenance of species-richness in plant communities: the importance of
1181 the regeneration niche. *Biological Reviews* 52:107–145.
- 1182 Harms, K. E., R. Condit, S. P. Hubbell, and R. B. Foster. 2001. Habitat associations of trees and
1183 shrubs in a 50 ha neotropical forest plot. *Journal of Ecology* 89:947–959.
- 1184 He, F., P. Legendre, J. LaFrankie. 1996. Spatial pattern of diversity in a tropical rain forest in
1185 Malaysia. *Journal of Biogeography* 23:57–74.
- 1186 Hermy, M. 1994. Effects of former land use on plant species diversity and pattern in European
1187 deciduous woodlands. *Biodiversity, Temperate Ecosystems, and Global Change* 20:123–
1188 144.
- 1189 Hu, Y.-H., D.-Y. Sheng, Y.-Z. Xiang, Z.-J. Yang, D.-P. Xu, N.-N. Zhang, and L.-L. Shi. 2013.
1190 The environment, not space, dominantly structures the landscape patterns of the richness
1191 and composition of the tropical understory vegetation. *PLoS One* 8:e81308.
- 1192 Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton
1193 University Press, Princeton, NJ.
- 1194 Koerner, W., J. L. Dupouey, E. Dambrine, and M. Benoît. 1997. Influence of past land use on the
1195 vegetation and soils of present day forest in the Vosges mountains, France. *Journal of*
1196 *Ecology* 85:351–358.
- 1197 Lawton, J. 1999. Are there general laws in ecology? *Oikos* 84:177–192
- 1198 Legendre, P. 2008. Studying beta diversity: ecological variation partitioning by multiple
1199 regression and canonical analysis. *Journal of Plant Ecology* 1:3–8.
- 1200 Legendre, P., D. Borcard, and P. Peres-Neto. 2005. Analyzing beta diversity : partitioning the
1201 spatial variation of community composition data. *Ecological Monographs* 75:435–450.

- 1202 Legendre, P., X. Mi, H. Ren, K. Ma, M. Yu, I.-F. Sun, and F. He. 2009. Partitioning beta
1203 diversity in a subtropical broad-leaved forest of China. *Ecology* 90:663–674.
- 1204 Legendre, P. and E. Gallagher. 2001. Ecologically meaningful transformations for ordination of
1205 species data. *Oecologia* 129:271–280.
- 1206 López-Martínez, J. O., L. Sanaphre-Villanueva, J. M. Dupuy, J. L. Hernández-Stefanoni, J. A.
1207 Meave, and J. A. Gallardo-Cruz. 2013. β -Diversity of functional groups of woody plants
1208 in a tropical dry forest in Yucatan. *PLoS One* 8:e73660.
- 1209 Motzin, G., D. Foster, A. Allen, J. Harrod, and R. Boone. 1996. Controlling site to evaluate
1210 history: vegetation patterns of a New England sand plain. *Ecological Monographs*
1211 66:345–365.
- 1212 Myers, J. A., J. M. Chase, I. Jiménez, P. M. Jørgensen, A. Araujo-Murakami, N. Paniagua-
1213 Zambrana, and R. Seidel. 2013. Beta-diversity in temperate and tropical forests reflects
1214 dissimilar mechanisms of community assembly. *Ecology Letters* 16:151–157.
- 1215 National Climate Data Center (NCDC), National Atmospheric and Oceanic Administration.
1216 2013. <http://www.ncdc.noaa.gov/>.
- 1217 Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson,
1218 P. Solymos, M. H. M. Stevens, and H. Wagner. 2013. *vegan*: community ecology
1219 package. R package version 2.0-10. <http://CRAN.R-project.org/package=vegan>.
- 1220 PAMAP. 2008. PAMAP Program LAS Files (LiDAR Data of Pennsylvania). PAMAP Program,
1221 PA Department of Conservation and Natural Resources, Bureau of Topographic and
1222 Geologic Survey. <http://www.pasda.psu.edu/>.
- 1223 Peterken, G. F., and M. Game. 1984. Historical factors affecting the number and distribution of
1224 vascular plants species in Central Lincolnshire. *Journal of Ecology* 72:155–182.
- 1225 Peres-Neto, P. R., P. Legendre, S. Dray, and D. Borcard. 2006. Variation partitioning of species

- 1226 data matrices: estimation and comparison of fractions. *Ecology* 87:2614–25.
- 1227 Pickett, S. T. A. 1980. Non-equilibrium coexistence of plants. *Bulletin of the Torrey Botanical*
1228 *Club* 107:238–248.
- 1229 QGIS Development Team. 2013. QGIS Geographic Information System. Open Source
1230 Geospatial Foundation Project. <http://qgis.osgeo.org>.
- 1231 R Core Development Team. 2013. R: A language environment for statistical computing. Vienna,
1232 Austria. <http://www.r-project.org/>.
- 1233 Raikow, D. F., S. A. Grubbs, and K. W. Cummins. 1995. Debris dam dynamics and coarse
1234 particulate organic matter retention in an Appalachian mountain stream. *Journal of the*
1235 *North American Benthological Society* 14:535–546.
- 1236 Ricklefs, R. E. 1987. Community diversity: relative roles of local and regional processes.
1237 *Science* 235:167–171.
- 1238 Roberts, D. W. 2013. labdsv: ordination and multivariate analysis for ecology. R package version
1239 1.6-1. <http://CRAN.R-project.org/package=labdsv>.
- 1240 Siefert, A., C. Ravenscroft, M. D. Weiser, N. G. Swenson. 2013. Functional beta-diversity
1241 patterns reveal deterministic community assembly processes in eastern North American
1242 trees. *Global Ecology and Biogeography* 6:682–691.
- 1243 Simberloff, D. 2004. Community ecology: is it time to move on? *The American Naturalist*
1244 163:787–99.
- 1245 Smith, T. W., and J. T. Lundholm. 2010. Variation partitioning as a tool to distinguish between
1246 niche and neutral processes. *Ecography* 33:648–655.
- 1247 Soil Survey Staff, Natural Resources Conservation Service (NRCS), Web Soil Survey, U.S.
1248 Department of Agriculture. 2014. <http://websoilsurvey.nrcs.usda.gov/>.
- 1249 Svenning, J.-C., D. A. Kinner, R. F. Stallard, B. M. J. Engelbrecht, and S. J. Wright. 2004.

- 1250 Ecological determinism in plant community structure across a tropical forest landscape.
1251 Ecology 85:2526–2538.
- 1252 Thessler, S., K. Ruokolainen, H. Tuomisto, and E. Tomppo. 2005. Mapping gradual landscape-
1253 scale floristic changes in Amazonian primary rain forests by combining ordination and
1254 remote sensing 14:315–325.
- 1255 Thompson, J. R., D. N. Carpenter, C. V. Cogbill, and D. R. Foster. 2013. Four centuries of
1256 change in northeastern United States forests. PLoS One 8:e72540.
- 1257 Vellend, M. 2004. Parallel effects of land-use history and genetic diversity of forest herbs.
1258 Ecology 85:3043–3055.
- 1259 Vellend, M., K. Verheyen, K. M. Flinn, H. Jacquemyn, A. Kolb, H. V. Calster, G. Peterken, B. J.
1260 Graae, J. Bellemare, O. Honnay, J. Brunet, M. Wulf, F. Gerhardt, and M. Hermy. 2007.
1261 Homogenization of forest plant communities and weakening of species–environment
1262 relationships via agricultural land use. Journal of Ecology 95:565–573.
- 1263 Whitney, G. G. 1994. From coastal wilderness to fruited plain: a history of environmental change
1264 in temperate North America from 1500 to the present. Cambridge University Press, New
1265 York, NY.
- 1266 Whittaker, R. H. 1956. Vegetation of the Great Smoky Mountains. Ecological Monographs
1267 26:1–80.
- 1268 Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. Ecological
1269 Monographs 30:279–338.
- 1270 Young, T. and M. Peacock. 1992. Giant senecios and alpine vegetation of Mount Kenya. Journal
1271 of Ecology 80:141–148.

- 1272 Ecological Archives Appendix A: Sampling design of the vegetation survey
- 1273 Ecological Archives Appendix B: Hill-shade relief map of Powdermill Nature Reserve
- 1274 Ecological Archives Appendix C: Soils Survey Data
- 1275 Ecological Archives Appendix D: Numbers of land-use history categories and stand ages used
- 1276 Ecological Archives Appendix E: Significance of individual variables used in the variance
1277 partitioning model
- 1278 Ecological Archives Appendix F: Species list for 2007-2008 vegetation survey at Powdermill
1279 Nature Reserve
- 1280 Ecological Archives Appendix G: Distribution and abundance maps for the 32 most abundant
1281 species sampled
- 1282 Ecological Archives Appendix H: Spatial map patterns (PCNM eigenvectors)
- 1283 Ecological Archives Appendix I: Maps of habitat categories derived from multivariate regression
1284 tree analysis

1285 **Table 1.** Habitat groupings at Powdermill Nature Reserve based on multivariate regression tree analysis using both environmental and
 1286 spatial variables (PCNMs). For groups based on measured environmental variables (i.e. true habitats), descriptions based on each of the
 1287 environmental variables used are shown. Similarly, the space variables are represented by the eigenvectors obtained in the PCNM
 1288 analysis. Additionally, the number of blocks and the number of species significantly associated with each group based on indicator
 1289 species analysis is shown. See appendix I for a map of both groups of habitat types.

Variables	Group	Description	No. blocks	No. species
Environmental	E1	Elevation ≥ 500 ; Soil = BuB,MkF,WrB,WsB,WsD.	119	3
	E2	Elevation ≥ 500 ; Soil = BuD,CaB,CeB,ErB,GcB,GcC,LaD,LaE,MaF, RgB,RgD,WrC; Slope < 10.65 .	126	1
	E3	Elevation ≥ 500 ; Soil = BuD,CaB,CeB,ErB,GcB,GcC,LaD,LaE,MaF, RgB,RgD,WrC; Slope ≥ 10.65 ; Asp = SE, SW, W.	16	7
	E4	Elevation ≥ 500 ; Soil = BuD,CaB,CeB,ErB,GcB,GcC,LaD,LaE,MaF, RgB,RgD,WrC; Slope ≥ 10.65 ; Asp = E, NE, NW; Dist ≥ 256 .	28	0
	E5	Elevation ≥ 500 ; Soil = BuD,CaB,CeB,ErB,GcB,GcC,LaD,LaE,MaF, RgB,RgD,WrC; Slope ≥ 10.65 ; Asp = E, NE, NW; Dist < 256 .	16	7
	E6	Elevation < 500 ; Soil = BeD,BuB,BuD,CrB,GcB,GcD,MaF,RgD,	112	0

WrB,WrC.

	E7	Elevation < 500; Soil = ErB,LaD,LaE,RgB,WsD.	68	4
Spatial	S1	PCNM4 \geq -0.005365; PCNM3 < 0.05072557; PCNM5 \geq 0.00211196	89	4
	S2	PCNM4 \geq -0.005365; PCNM3 < 0.05072557; PCNM5 < 0.00211196	161	3
	S3	PCNM4 \geq -0.005365; PCNM3 \geq 0.05072557	23	5
	S4	PCNM4 < -0.005365; PCNM15 \geq 0.01949	60	0
	S5	PCNM4 < -0.005365; PCNM15 < 0.01949; PCNM1 < 0.0262459	108	5
	S6	PCNM4 < -0.005365; PCNM15 < 0.01949; PCNM1 \geq -0.0262459	44	9

1290

1291

1292

1293

1294

1295

1296

1297

1298

1299 **Table 2:** Statistically significant indicator species with corresponding indicator value (IV) and p-value for the different categories of
 1300 habitat types and spatial descriptors (PCNMs) produced by the multivariate regression tree analysis, as well as for the different
 1301 categories of land-use types. See table 1 for descriptions of habitat groups. Species are listed based on true habitat categories
 1302 (environment columns), followed by alphabetical order.

	Environment			Spatial descriptors			Land-use history		
	Category	I.V.	<i>p</i>	Category	I.V.	<i>p</i>	Category	I.V.	<i>p</i>
<i>Betula alleghaniensis</i>	E1	26.4	0.001	S1	31.7	0.001			
<i>Betula lenta</i>	E1	31.1	0.001	S1	34.6	0.001			
<i>Magnolia acuminata</i>	E1	31.5	0.001	S1	31.1	0.001			
<i>Prunus serotina</i>	E2	21.4	0.03	S5	23.1	0.01			
<i>Acer rubrum</i>	E3	26.7	0.001	S1	27.6	0.001			
<i>Castanea dentata</i>	E3	22.2	0.001						
<i>Nyssa sylvatica</i>	E3	33.3	0.001	S2	22.8	0.01			
<i>Quercus prinus</i>	E3	39.5	0.001	S2	19.9	0.01			
<i>Quercus rubra</i>	E3	18.3	0.04	S2	20.2	0.01	Shrub land	34.8	0.008
<i>Quercus velutina</i>	E3	7.5	0.03						
<i>Sassafras albidum</i>	E3	25.9	0.001						
<i>Acer saccharum</i>	E5	36.7	0.001	S6	38.7	0.001	Developed	23.2	0.009
<i>Fagus grandifolia</i>	E5	21.4	0.01	S5	22.6	0.006			

Murphy

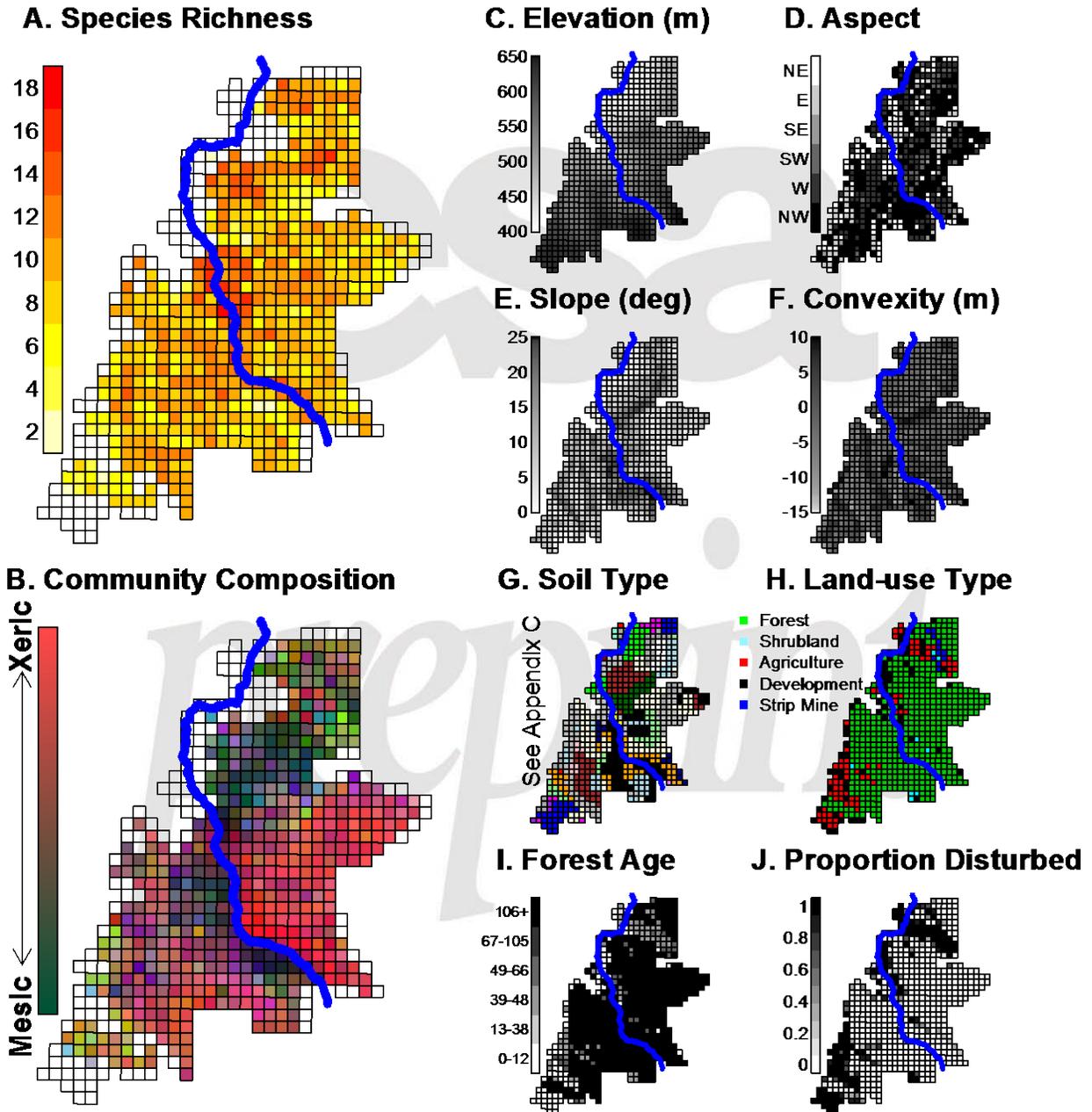
ECOLOGICAL SOCIETY OF AMERICA

<i>Fraxinus nigra</i>	E5	6.2	0.04						
<i>Liriodendron tulipifera</i>	E5	30.6	0.001	S3	30	0.001	Developed	34.5	0.02
<i>Tilia americana</i>	E5	14.3	0.01	S6	14.4	0.009			
<i>Tsuga canadensis</i>	E5	27.1	0.001	S3	28.4	0.001	Developed	19.5	0.01
<i>Carya glabra</i>	E7	14	0.01	S6	22	0.001	Mine land	20.1	0.04
<i>Carya ovata</i>	E7	26.3	0.001	S6	35.7	0.001	Mine land	28.9	0.02
<i>Cornus florida</i>	E7	9.5	0.03	S6	11.5	0.01			
<i>Fraxinus americana</i>	E7	18.8	0.02	S6	32.6	0.001	Agriculture	28.2	0.04
<i>Acer saccharinum</i>				S3	12.3	0.01	Developed	32.7	0.03
<i>Carya cordiformis</i>				S6	14.4	0.009			
<i>Crataegus spp.</i>							Agriculture	12.6	0.03
<i>Juglans nigra</i>							Developed	11.1	0.03
<i>Kalmia latifolia</i>				S3	4.3	0.04			
<i>Malus spp.</i>				S5	7.7	0.01	Agriculture	10.9	0.04
<i>Ostrya virginiana</i>				S6	11.5	0.005			
<i>Platanus occidentalis</i>							Mine land	13.5	0.02
<i>Quercus alba</i>				S5	15.9	0.02			
<i>Robinia pseudoacacia</i>				S6	11.1	0.01	Agriculture	28.1	0.01
<i>Ulmus rubra</i>				S5	5.1	0.04			

1304 **Figure 1:** Maps of block-scale variables used in the analysis **A.** Species richness, ranging from a
1305 low of 0 (white) to a high of 19 (bright red). **B.** Beta-diversity map. Similar colors indicate
1306 similar species composition based on Bray-Curtis dissimilarity. White areas are where no trees
1307 were present. **C.** Elevation. **D.** Slope aspect. **E.** Slope steepness. **F.** Slope convexity, measured as
1308 the elevation of each block minus the mean of the surrounding 8 blocks. **G.** Soil type based on
1309 soil survey maps (see Appendix C). **H.** Most recent land-use type. **I.** Average stand age, ranging
1310 from 0 to >100 years. **J.** Proportion of plots disturbed within each block. Blue lines indicate the
1311 location of Powdermill Run stream.

1312

1313 **Figure 2:** Fractions of explained variation for species richness (**A**) and community composition
1314 (**B**). Total fractions are listed in the bottom-right corner, and individual fractions are indicated
1315 within the Venn-diagrams.

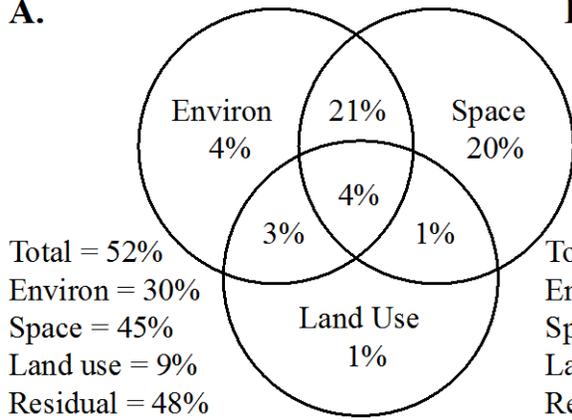


1316

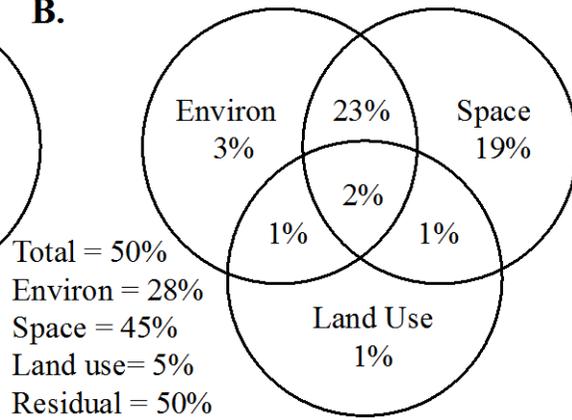
1317 **Figure 1:**

1318

A.



B.



1319

1320 **Figure 2:**

preprint