Beyond bivariate correlations: three-block partial least squares illustrated with vegetation, soil, and topography

Daehyun Kim,[1],† Thomas J. DeWitt,[2] César S. B. Costa,[3] John A. Kupfer,[4]
Ryan W. McEwan,[5] and J. Anthony Stallins[1]

[1]*Department of Geography, University of Kentucky, Lexington, Kentucky 40506 USA*
[2]*Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, Texas 77843 USA*
[3]*Laboratório de Biotecnologia de Halófitas, Instituto de Oceanografia, Universidade Federal do Rio Grande-FURG,
Av. Itália, km 08, 96203-900 Rio Grande, Rio Grande do Sul, Brazil*
[4]*Department of Geography, University of South Carolina, Columbia, South Carolina 29208 USA*
[5]*Department of Biology, University of Dayton, Dayton, Ohio 45469 USA*

Abstract. Ecologists, particularly those engaged in biogeomorphic studies, often seek to connect data from three or more domains. Using three-block partial least squares regression, we present a procedure to quantify and define bi-variance and tri-variance of data blocks related to plant communities, their soil parameters, and topography. Bi-variance indicates the total amount of covariation between these three domains taken in pairs, whereas tri-variance refers to the common variance shared by all domains. We characterized relationships among three domains (plant communities, soil properties, topography) for a salt marsh, four coastal dunes, and two temperate forests spanning several regions in the world. We defined the specific bi- and tri-variances for the ecological systems we included in this study and addressed larger questions about how these variances scale with each other looking at generalities across systems. We show that a system tends to exhibit high bi-variance and tri-variance (tight coupling among domains) when subjected to the effects of frequent and widespread (i.e., broadly acting) hydrogeomorphic disturbance. When major disturbance events are uncommon, bi-variance and tri-variance decrease, because the formation of vegetation, soil, and topographic patterns is primarily localized, and the couplings of these properties diverge over space, contingent upon site-specific disturbance history and/or fine-grained environmental heterogeneity. We also demonstrate that the bi-variance and tri-variance of a whole system are not consistently either greater or smaller than those of the associated sub-zones. This point implies that the overall correlation structure among vegetation, soil, and topography is conserved across spatial scales. This paper addresses a critical aspect of ecology: the conceptual and analytical integration of data across multiple domains. By example, we show that bi-variances and tri-variances provide useful insight into how the strength of couplings among vegetation, soil, and topography data blocks varies across scales and disturbance regimes. Though we describe the simplest case of multi-variance beyond the usual two-block linear statistical model, this approach can be extended to any number of data domains, making integration tractable and more supportive of holistic inferences.

Key words: biogeomorphology; disturbance; ecosystem engineer; feedback; historical contingency; multicollinearity; scale invariance; three-block partial least squares; variance.

† **E-mail:** biogeokim@uky.edu

Introduction

The spatial distribution and temporal dynamics of vegetation are often correlated with a set of known environmental conditions, including those associated with soil and topography (e.g., Cowles 1899, Boerner 2006, Kim and Yu 2009). In floodplains, for example, the role of topography is often manifested through its effects on flood regime, resulting in communities that are distributed along gradients of flooding, geomorphology, and soils (e.g., Sharitz and Mitsch 1993). Conditions related to soils and topography are then used as independent factors to predict the behavior of individual plant species, populations, and communities measured as changing abundance, biomass, or composition. Over decades, this approach facilitated considerable progress in understanding the mode of vegetation–soil and vegetation–topography couplings in upland areas and vegetation–flooding couplings in floodplain settings. The present study aims to stretch the discipline's perspective beyond such bivariate relationships by estimating the variance simultaneously shared by vegetation, soil, and topography in various ecological systems.

Quantifying relationships between variables has a long history in scientific research (Galton 1888, Pearson 1920, Wright 1921, Kendall 1938, Fisher 1954). Conventional multivariate regression approaches, while appearing complex, are essentially based on quantifying the covariation between a response (i.e., dependent) variable of interest and each of several predictor (i.e., independent) variables. The amount of this covariation is hereafter termed bi-variance. The concept of bi-variance is widespread and deep-rooted in modern science; to wit, researchers often contort their analyses into bi-variance problems even when the intuitive structure or fundamental nature of the data involves three or more fundamental domains.

This research focuses on the simplest case beyond bi-variance in which the relationships among three data blocks are considered; thus, the amount of the common variance shared by all of them will be called tri-variance (i.e., δ in Fig. 1). Uni-variance will indicate the variance in a data block that is unrelated to any other data block under consideration (i.e., the "error" in standard parametric statistics; α, ϵ, and η in Fig. 1).
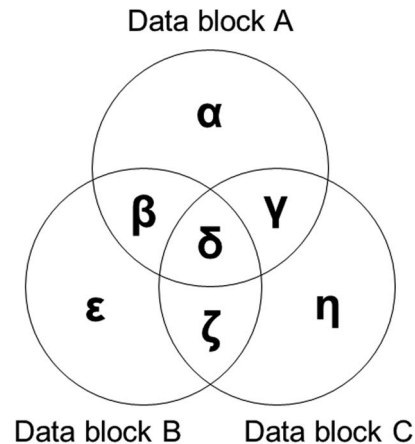

Fig. 1. A generic form of Venn diagram indicating what is meant by bi-variance ($\beta + \gamma + \delta + \zeta$) and tri-variance ($\delta$). The other symbols ($\alpha$, ϵ, and η) represent uni-variances, or unexplained variances.

Conceivably, covariance relationships among blocks may be structured in many ways. At one extreme, any two of the three factors are related to each other to some degree, but there is no variance shared by all of them simultaneously (Fig. 2A). Put another way, the interaction between Factors A and B occurs exclusively, regardless of the other interactions between A and C or between B and C. Therefore, knowing how A and B covary does not provide any insight into the A–C or B–C correlation (and vice versa). In the second case, covariance between any two factors and among all factors exists (Fig. 2B). Here, from a certain type of interaction, e.g., between Factors A and B, one can infer to some extent the pattern of the other interactions between A and C or between B and C. The reliance of such inference should improve as tri-variance relative to bi-variance increases, and the reliance should become maximized when the magnitude of bi-variance and tri-variance is the same, another extreme case (Fig. 2C). Overall, these three situations can be commonly summarized as
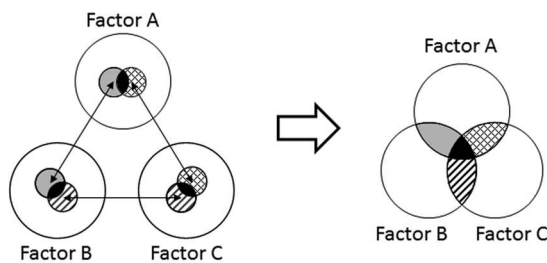
$$\text{Factor A} = f(\text{Factor B, Factor C, other factors})$$
$$\text{Factor B} = f(\text{Factor A, Factor C, other factors})$$
$$\text{Factor C} = f(\text{Factor A, Factor B, other factors})$$

(1)

where "other factors" refer to chance and latent (unmeasured causal) influences. As these equa-

A Bi-variance > zero, but tri-variance = zero
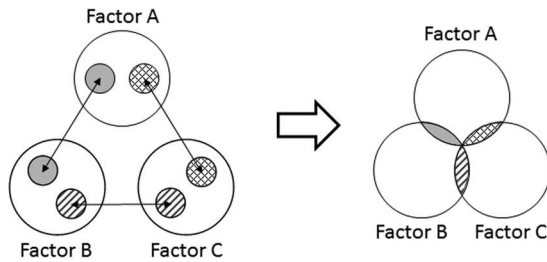
B Both bi-variance and tri-variance > zero

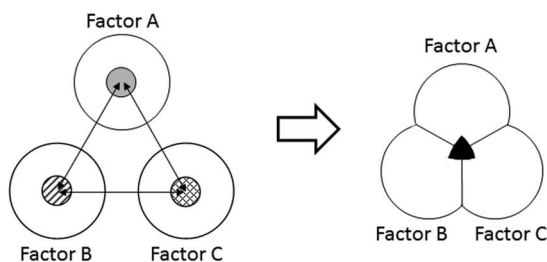**C** Bi-variance = tri-variance > zero

Fig. 2. Conceptual models of possible relationships among three distinct factors (left), and a representation of the relationships using Venn diagrams (right).

tions imply, any one of Factors A, B, and C is apparently related to the other two; however, the equations deliver no more than such bivariate information, as if the causal/correlative scenario were structured only as in Fig. 2A. Whether and how much the covariation among the three exists cannot be determined until one actually estimates the tri-variance.

As bi-variance indicates how closely two data blocks are related to each other, tri-variance is a proxy for the level of covariance among all three blocks. For example, many coastal dunes are located where the effects of overwash disturbance are not manifested (i.e., tide-dominated coasts; Stallins and Parker 2003). Under the absence of widespread control over the dunefield, the formation of vegetation, soil, and topographic patterns is primarily localized; therefore, their relationships may diverge over space. As a consequence, locations having similar topographic conditions may exhibit greatly different plant species composition or successional pathways, contingent upon site-specific variations in disturbance history or soil microsites. In such a case, the amount of the variance shared by any two factors and, especially, by all three is likely low over larger areas. In contrast, other dunes are highly subjected to periodic or episodic overwash events that influence the mortality and establishment of plant species, soil development, and surface morphodynamics (i.e., wave-dominated coasts; Miller et al. 2010, Young et al. 2011). Under such widespread hydrogeomorphic control across the dunefield, vegetation, soil, and topography likely behave in unity over space and time, namely, their relationships may be more consistent and less *noisy* than the case above. Accordingly, the associated tri-variance as well as the bi-variance is expected to be high over larger spatial areas.

The multi-block data methods we describe below allow inquiry into the specific nature and scaling pattern of bi-variance and tri-variance across systems with different biogeomorphic conditions. In this paper, we develop three-block partial least squares regression (3-B PLS) for field data on vegetation, soil, and topography acquired from widely divergent ecological systems (coastal dunes, salt marshes, regenerating floodplain forests, and upland temperate forests) in various regions of the world. Each of these systems encompasses a gradient of topo-edaphic conditions, and is considered to be spatially heterogeneous in terms of ecological and physical processes. Therefore, we will adopt a two-step approach in which we first characterize the vegetation–soil–topography relationships throughout a whole system, and then decompose it into sub-units. In the case of coastal marshes, for example, we will examine patterns for the entire marsh first, and then subdivide the system into creek bars, bank levees, and interior marsh platforms to perform separate investigations on the same relationships within each of these sub-zones. Specifically, we intend to evaluate the

A Analysis of a whole system

Whole system

B Division of a whole system

Sub-zone A   Sub-zone B   Sub-zone C

$\text{Bi-variance}_{Whole}$
$\text{Tri-variance}_{Whole}$

Bi-variance_A Bi-variance_B Bi-variance_C
Tri-variance_A Tri-variance_B Tri-variance_C

C Possible cases

$\text{Bi (Tri)-variance}_{Whole}$

Consistently greater than
Consistently smaller than
Approximately the average of

$\text{Bi (Tri)-variance}_A$
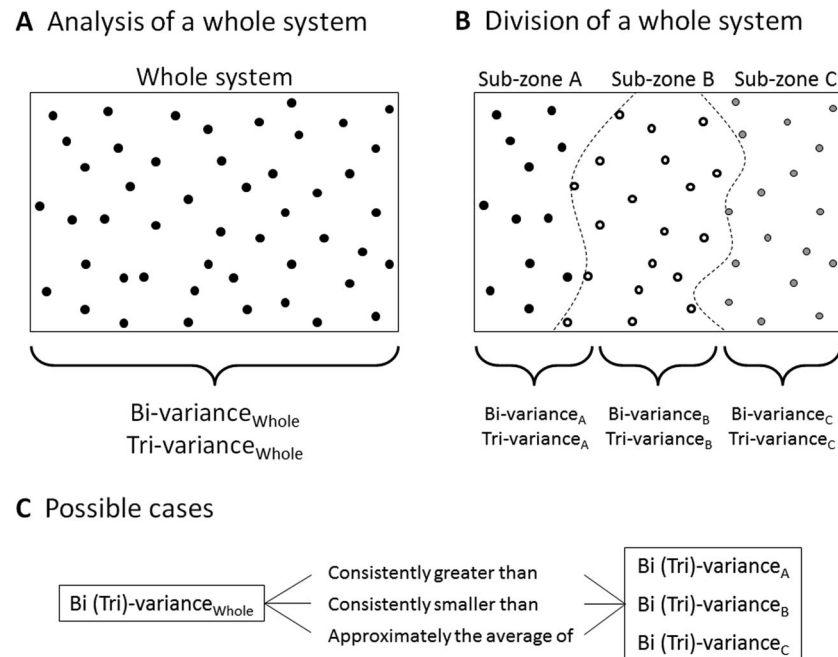$\text{Bi (Tri)-variance}_B$
$\text{Bi (Tri)-variance}_C$

Fig. 3. Comparison of analyses for a whole system and for its sub-zones. The factors of interest are vegetation, soil, and topography. The whole system encompasses a range of environmental gradients (e.g., landforms, disturbance regimes, etc.), so it is considered spatially heterogeneous with regard to ecological and physical processes. The sub-zones are relatively homogeneous, however. Each dot represents the location where vegetation–soil–topography relationships are examined.

following three hypotheses:

Hypothesis 1 (H_1): *Systems under widespread (i.e., broadly acting) environmental controls have higher bi-variance and tri-variance than those characterized by localized interactive biotic–abiotic processes.* This hypothesis is directly related to the example of two contrasting coastal dune types mentioned above. We will evaluate this idea in the context of vegetation–soil–topography couplings.

Hypothesis 2 (H_2): *Separate analyses of sub-zones produce consistently smaller or greater bi-variance and tri-variance than analysis of a whole system.* Scale-dependence is a fundamental notion in ecology (Levin 1992, Stohlgren et al. 1998, Turner 2010). It is widely known that distinctive ecological patterns and processes are dominant at different spatial scales. Thus, modifying scale, be it the size of a unit plot or the size of a whole study area, would involve a change in the correlation structure among biotic and abiotic components (Kim et al. 2012). In sum, the results of ecological analyses and the associated inter-

pretations are strongly dependent on the spatial scales in which a study is focused. For instance, small-scale studies often demonstrate a negative relationship between native and exotic plant species diversity (Tilman 1997, Stachowicz et al. 1999, Brown and Peet 2003), whereas at large scales the opposite relationship has been discovered (Lonsdale 1999, Stohlgren et al. 2003, Davies et al. 2005). In the same context, the strength of vegetation–soil–topography relationships should vary depending on the scales of data collection. However, there is no consensus as to whether such strength will increase or decrease after dividing a heterogeneous whole system into a number of homogeneous sub-zones. Alternatively, there might be no consistently increasing or decreasing strength after the division; that is, the relationships may become weaker in some sub-zones and stronger in the others. In this case, the bi-variance and tri-variance of the whole system represent more or less the average strength of the biotic–abiotic couplings observed in the sub-zones (see Fig. 3).

Hypothesis 3 (H₃): *Bi-variance is a good indicator of tri-variance.* This hypothesis is somewhat related to why there has been no explicit estimation and use of tri-variance in related literature. Is this because bi-variance would be a sufficient measure to provide enough detailed insight into vegetation–soil–topography relationships in any systems (i.e., do we ecologists not even have to bother with the concept of tri-variance)? We expect that, in general, there will be a positive correlation between the values of bi-variance and tri-variance: the greater the amount of the total variance shared by any two of vegetation, soil, and topography in pairs, the greater the amount of the variance shared by all of them simultaneously. However, the significance (or variability) of the correlation is a different issue; one should not ignore a wide range of possible patterns in vegetation–soil–topography relationships as briefly described in Fig. 2. It is plausible that two distinct ecological systems, possessing the same magnitude of bi-variance, may have different tri-variances. Also, one can imagine even a case of negative relation: the bi-variance of system A is smaller than that of system B, but A has a higher tri-variance than B.

3-B PLS

Vegetation, soil, and topography are multivariate in nature: most vegetation data sets are composed of multiple *l* plant species, the abundances or biomass of which are measured in a certain number of plots. In these same plots, one can collect and analyze a set of soil data consisting of *m* physical and chemical properties, as well as topographic data of *n* landform attributes. For each data domain, there are relationships among variables both within and across data blocks. In this paper, use of community, soil, and topographic data results in *three data blocks*. Taken together, these blocks produce *l* × *m* × *n* possible bivariate cases of vegetation–soil–topography relationships. An attempt to understand each of these relationships will be considered inefficient if several variables within each block are significantly correlated with one another—the problem of multicollinearity (Farrar and Glauber 1967, Graham 2003).

In such a situation, the most widely employed solution has been the use of ordination, a family

of multivariate data reduction techniques that extract one or a few axes from the original variables. The data points (e.g., plots, species) are then ordered to aid interpretations of ecological and environmental patterns (Pielou 1984, ter Braak and Prentice 1988, Palmer 1993, Austin 2013). For example, canonical correspondence analysis (CCA) can directly and immediately relate plant species composition to measured environmental variables, along a few major axes (ter Braak 1986). Although CCA has been widely used in ecology (Lepš and Šmilauer 2003), it cannot actually help to achieve the goal of the present research: understanding the tri-variance among vegetation, soil, and topography. This is because, in CCA (and in similar methods), all soil and topography variables are lumped together as environmental variables (i.e., predictors) and the community data are treated as dependent variables; but of course, many taxa also change the environment around them, making covariance a very fluid concept that can propagate in multiple directions among data domains.

As a possible approach (Fig. 4), one may perform a principal component analysis (PCA) based on the correlation matrix produced from a single block. Among the resulting eigenvectors composed of axis scores for each plot, those axes with eigenvalues greater than 1 (assuming standardized data) are considered significant and selected as the principal components of that block. The same procedure is repeated for the remaining other blocks. Then, the components extracted from and representing each block are correlated to understand the patterns of vegetation–soil–topography relationships (e.g., Kim et al. 2012).

This PCA-based approach has one clear limitation, however: the extraction of components in a block is conducted without taking into account the covariation structure of the other blocks (see Carrascal et al. 2009 for more details). In other words, the extracted components maximize the correlation among variables within a block, but these components do not reflect the maximization of the correlation across blocks. As a consequence, in many cases, the principal components from an environmental block and those from a vegetation block exhibit low and non-significant correlations, hindering ecologically meaningful interpretation (Carrascal et al.
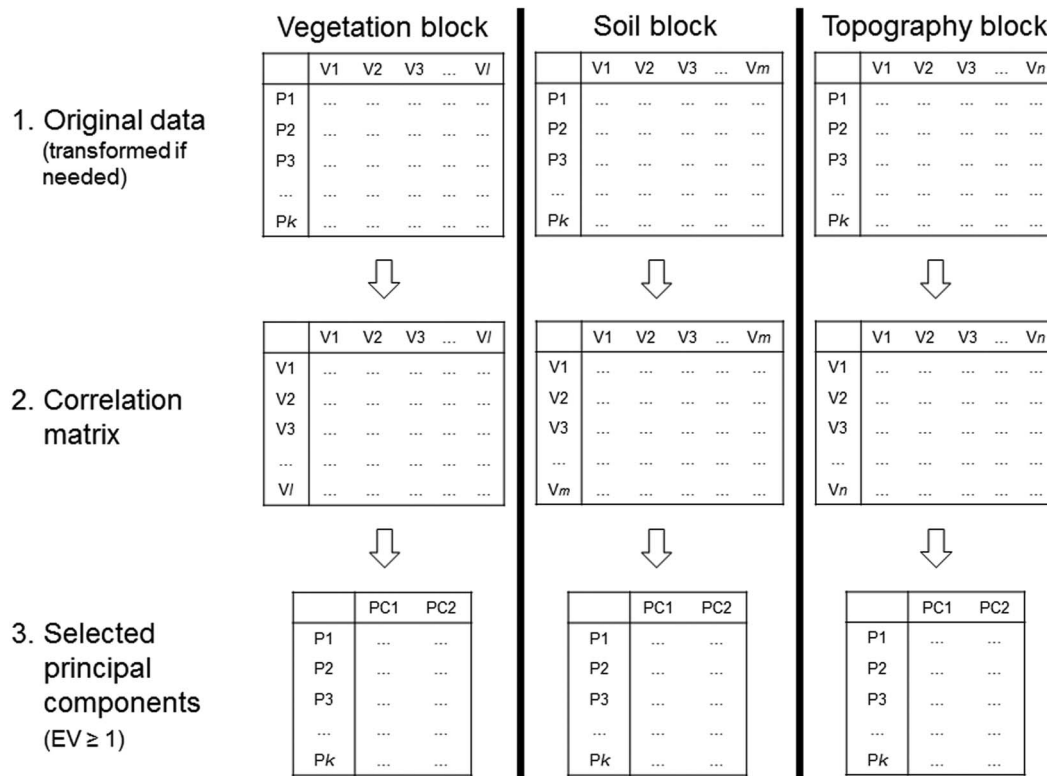
Fig. 4. A possible ordination procedure followed by correlating primary axes of different data blocks. In this particular case, we present an example of principal component analysis, but this analytic flow is also applicable to the other modes of ordination such as correspondence analysis and principal coordinate analysis. In this conceptual example, two principal components (PC) are extracted in each block, but the number can change depending on the magnitude of eigenvalues. The thick, solid, lines imply that the extraction of PCs is conducted exclusively to each block, that is, without taking into account the covariation structure of the other blocks (V: variable; P: plot; EV: eigenvalue).

2009).

Various multivariate techniques have been developed to address this issue, dealing with multi-block data sets in a symmetrical perspective (for a more thorough discussion, see Borcard et al. 2011, Legendre and Legendre 2012): co-inertia analysis (Dray et al. 2003), multiple factor analysis (Escofier and Pagès 1994, Bécue-Bertaut and Pagès 2008, Abdi et al. 2013), and Procrustes analysis (Gower 1975, Gower and Dijksterhuis 2004). While these methods have been employed by a number of ecologists in recent decades, the present work intends to introduce PLS as an alternative powerful approach that has nonetheless been underrepresented in ecological litera-

ture (Wold et al. 1984, Rosipal and Krämer 2006, Abdi 2007, Esposito Vinzi et al. 2010). It is an extension of multiple regression analysis that, when extracting major components from a certain block, seeks simultaneously to maximize the explained variance of the other blocks. Thus, components from different blocks would have the greatest mutual linear predictive power (Næs and Martens 1985, Geladi and Kowalski 1986, de Jong 1993). PLS can be a useful approach, especially in ecological studies, because it is designed to overcome typical situations, where (1) the number of predictor variables is similar to or greater than the number of observations and/ or (2) environmental variables, measured with an

aim to predict the spatial and temporal dynamics of organisms, are significantly correlated with one another. Despite such advantages of using PLS, there is a surprising paucity of research that employs the method in ecological literature (Carrascal et al. 2009, but see ter Braak and Juggins 1993, Birks 1996). Furthermore, to the best of our knowledge, PLS has not been conducted for three-block biotic and abiotic data, which is the subject of this paper. It should be noted that an exhaustive comparison of 3-B PLS and the several multi-block methods mentioned above is beyond the scope of this paper. There should be relative pros and cons of all these approaches depending on the type of ecological data—e.g., number of samples, degree of skewness, coefficient of variation, and direction of causality among data blocks. We illustrate the potential of 3B-PLS for addressing questions about reciprocal interactions between the environment and biota that go beyond bivariate relationships but do not become obscured by the dynamic instantaneousness of the feedbacks among them.

Materials and Methods

Overall, this paper analyzes vegetation, soil, and topographic data collected by five groups of ecologists from seven different systems at widely varied locations in the world. Brief explanations of each study area and field survey are presented below (for more detailed information, see Appendix).

Costa's study area and data collection

Costa et al. (1996) collected vegetation, soil, and topography data sets at coastal dunes along the southernmost 217 km long stretch of Brazil. The study region encompassed northern and southern sites: the northern dunes exhibited relatively high foredune ridges (~7 m) and the southern sites showed relatively large seashore deflation plains. At each site, Costa et al. (1996) established five 100 m long and 1 m wide strip transects perpendicular to the coastline, starting at the leading edge of the vegetation. Each transect was divided into 10 zones each with an equal length of 10 m. Along each zone, the presence of plant species was recorded at 2 m intervals using 1 m^2 square quadrats. Then, all

floristic data from these zones were pooled to represent the overall frequency of each species. In each zone, the researchers also examined the pH of superficial sand (10–20 cm depth) and the proportion (%) of coarse (>0.50 mm), medium (0.25–0.50 mm), fine (0.062–0.25 mm), and clay/silt (<0.062 mm) sediments. Two topographic attributes of each zone were included in the present work: surface elevation (m) and distance from the coastline (m), which corresponded to the gradients in sedimentation rates and salinity.

Kim's study area and data collection

Kim's research sites included the Sindu coastal dunefield in South Korea and the Skallingen salt marsh in Denmark. The dunefield ranged for about 4 km along the coastline and extended 0.5–1.5 km inland. This system exhibited well-developed topographic features, such as foredune, second dune, and interdune depression (or swale), as well as distinct vegetation zonation in each of these habitats (Kim et al. 2008, Kim and Yu 2009). In the summer of 2003, Kim established a rectangular plot (180 × 280 m) that fully encompassed these three geomorphic zones. The plot was divided into 126 grids of 20 × 20 m. In each grid, he surveyed the abundance of 21 main plant species, the mean cover of which exceeded 0.5% across the plot. Also, a total of 11 physical and chemical properties were analyzed for the 126 soil samples collected from the grid centers. Finally, for these centers, eight terrain parameters were extracted from a digital elevation model of the dunefield.

The Skallingen marsh was characterized by a complex tidal creek system that actively migrated through bar deposition and cutbank erosion across the marsh's platform (Kim 2012, Kim et al. 2012). In the summer 2006, Kim established 11 line transects along four tidal creeks. These transects were, on average, about 25 m long and perpendicular to the streamline. Each transect fully encompassed three distinct topographic zones, such as newly deposited point bar, bank levee (or transition), and inner marsh field (or marsh interior). Along these transects, a total of 9–10 locations were selected for soil and topographic surveys. Kim identified 10 soil physical and chemical properties and two topographic variables (surface elevation and distance to the channel). At each location, two replicate square

quadrats (1 × 1 m) were identified for vegetation survey.

Kupfer's study area and data collection

The Kupfer data set comes from a study of post-logging succession on the floodplain of the Bates Fork tract within Congaree National Park (NP), near Gadsden, South Carolina, USA (Kupfer et al. 2010). Sample transects ($N = 63$; 30 m in length) were established in forest clearcuts that were located in three different geomorphic settings (a backswamp, a remnant natural levee associated with a former meander of the Congaree River, and the active natural levee adjacent to the current river channel) and had been undergoing succession for 4–5 years. Vegetation composition was measured in three circular subplots (2.82 m radius) spaced evenly along the length of the transects; these transects were oriented parallel to the forest edge and located 5–50 m into the clearcut. Soil cores were taken from the top 15 cm of soil on each subplot, composited by transect, dried, and then analyzed for pH, extractable phosphorus, potassium, calcium, magnesium, zinc, copper and sodium, organic matter, cation exchange capacity (CEC), exchangeable acidity (the amount of total CEC occupied by H^+ and Al^{3+}), and total percent base saturation (the percent of exchange sites occupied by base cations). Two variables related to topography, and by extension, flood regime, were determined for each subplot and averaged by transect: (1) elevation, which was extracted from a high resolution, LiDAR-derived digital terrain model of the study area, and (2) the depth of inundation during a 98,000 cfs flood, which was estimated using a flood inundation model (HEC-RAS; Kupfer et al. 2010).

McEwan's study area and data collection

The McEwan data set was generated in a study of a 52-ha old-growth forest in southeastern Kentucky, USA (McEwan and Muller 2011). Sampling took place within four 1-m^2 circular plots ($N = 320$) nested within 80 larger plots established by Muller (1982). Muller (1982) used ordination techniques to classify these 80 plots into three ecological communities: sugar maple (*Acer saccharum*; 17 plots), beech (*Fagus grandifolia*; 31 plots), and chestnut oak (*Quercus montana*; 32 plots; McEwan and Muller 2006).

An inventory of herbaceous plants, with an associated visual estimate of cover for individual species, was conducted on all herb-layer subplots beginning the first week of April, June, and August of 2000. Topographic variables were collected by Muller (1982) who measured elevation and slope aspect on each of the plots. Soil samples ($N = 4$/subplot) were collected within each herb subplot in June of 2001. Samples were homogenized and analyzed for plant-available nitrogen (NH_4^+, NO_3^-), Mehlich III-extractable P, K, Ca, Mg and Mn, cation exchange capacity, and base saturation. Hemispherical digital canopy images were collected by McEwan at each permanent plot center and analyzed using Gap Light Analyzer image processing software (Frazer et al. 1999, McEwan and Muller 2011).

Stallins's study area and data collection

The two U.S. barrier islands in the studies conducted by Stallins were Sapelo Island, Georgia and South Core Banks, North Carolina, USA. Dune vegetation and soils were sampled in the summer of 1999. At each of the five sites along a 5-km stretch of Holocene dunes on Sapelo, three transects were aligned perpendicular to the shoreline. Vegetation was sampled at 1-m intervals with a point frame. One hundred and forty soil samples were systematically drawn from representative habitat types (dune ridges, flats, and swales) among these transects, with the additional requirement that the distance between successive soil samples was no greater than 10 m. A similar sampling design (five sites of three transects each) was employed on South Core Banks. However, island length here was greater (35 km), as was the distance between successive soil samples (20 m) given longer transect lengths. A total of 136 soils samples were collected. Full details are provided in Stallins (2001) and Stallins and Parker (2003).

Statistical procedure for 3-B PLS

Before the 3-B PLS was conducted, all variables were transformed using appropriate mathematical functions so that the data met the normality assumption in statistics. Because these variables were measured in distinct units, they were standardized and unitless (i.e., their mean and standard deviation were 0 and 1, respectively). Not surprisingly, vegetation, soil, and topogra-
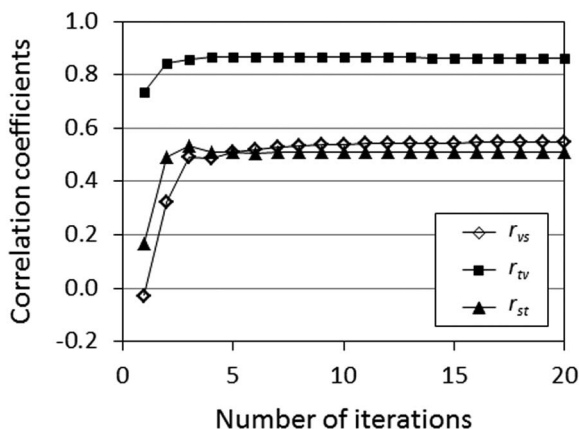
Fig. 5. An example of converging correlation coefficients among the unit vectors of vegetation, soil, and topography data blocks in pairs, acquired by Kim at the Sindu coastal dunefield, South Korea. Correlation coefficients r_{vs}, r_{tv}, and r_{st} indicate vegetation–soil, topography–vegetation, and soil–topography couplings, respectively. In this particular case, the three coefficients started to become stable (i.e., convergence) at about the 8th iteration.

phy blocks were of different size in each data set; that is, the number of variables in each block was different. For example, Kim's data from the Sindu coastal dunefield consisted of 24 plant species, 11 physical and chemical soil attributes, and eight landform parameters. To scale these varying block sizes (BS) into 1, we applied the following modification for each data value (DV) within each block:

$$DV \times \sqrt{1/BS}. \tag{2}$$

This scaling equation (adapted from Ginter et al. 2012) can also be used if desired to scale data blocks to any size, such as the number of independent dimensions, or "true rank" of data, by replacing the numeral 1 with the desired block size.

Following Streissguth et al. (1993) and Bookstein et al. (2003), the first stage of 3-B PLS defined three unit vectors, one each for vegetation, soil, and topography blocks (U_v, U_s, and U_t) and then iteratively optimized them. Bookstein et al. (2003) suggested using arbitrary starting values for the starting vectors. However, we performed PCA for each block to select the first principal components as the starting unit vectors.

In our preliminary analysis, we found that these two different approaches produced almost identical results in the end. Interpreting these unit vectors as linear combinations, we calculated

$$\begin{aligned} s_v &= X_v \times U_v \\ s_s &= X_s \times U_s \\ s_t &= X_t \times U_t \end{aligned} \tag{3}$$

where X_v, X_s, and X_t represented the whole vegetation, soil, and topography blocks, respectively. X was a data matrix with N rows (samples) and k columns (variables; i.e., $X = N \times k$). Each s was normalized to sample variance 1, and we then computed the correlations among s_v, s_s, and s_t in pairs

$$\begin{aligned} r_{vs} &= \text{correlation}(s_v, s_s) \\ r_{st} &= \text{correlation}(s_s, s_t) \\ r_{tv} &= \text{correlation}(s_t, s_v). \end{aligned} \tag{4}$$

This was the end of the first iteration in our approach of 3-B PLS. The second iteration began with developing updated unit vectors, U_v, U_s, and U_t, based on the inter-block correlation structure as follows:

$$\begin{aligned} U_v &\leftarrow X_v(r_{vs}s_s + r_{tv}s_t) \\ U_s &\leftarrow X_s(r_{st}s_t + r_{vs}s_v) \\ U_t &\leftarrow X_t(r_{tv}s_v + r_{st}s_s). \end{aligned} \tag{5}$$

In other words, these new unit vectors were generated by summing "predictions of the individual variables using combinations of the scores s with the correlations r as weights" (Bookstein et al. 2003:185). We then returned to Eq. 3 to repeat the same iteration until arriving at a convergence of stable correlation coefficients, r_{vs}, r_{st}, and r_{tv}, that did not vary between iterations (e.g., see Fig. 5).

At convergence, the final vectors, U_v, U_s, and U_t, were labeled as the first singular axes representing the blocks of vegetation, soil, and topography, respectively. A clear distinction exists between these singular axes and principal components: principal components are extracted from the internal correlation structure within each block of data, with no regard to covariation across data domains (vegetation, soil, topography). Thus, each principal component maximally represents the overall pattern of only a given data block. In contrast, a singular axis from 3-B PLS is predicted by the weighted combination of the other two blocks as Eq. 5 indicates. Therefore,

Table 1. A generic form of analysis of variance table developed to estimate extra sums of squares.

Source	df	Sum of squares	Mean squares
X1	1	SSR (X1)	MSR (X1)
X2	1	SSR (X2|X1)†	MSR (X2|X1)
Residuals	$N - 3$	SSE (X1, X2)	MSE (X1, X2)
Total	$N - 1$	TSS‡	

† The additional variation explained by X2 when added to a regression model already containing X1 (i.e., the pure effect of X2 on the response variable, excluding that of X1).

‡ In this present paper, because all of the first singular axes from vegetation, soil, and topography data blocks have been standardized before regression modeling, TSS = $N - 1$.

each singular axis represents the correlations of the elements of a given data block with other blocks.

While the first singular axes express the patterns that have the greatest correlation among vegetation, soil, and topography blocks, one may also be interested in extracting the second (and higher) axes. By construction, the second singular axes—analogous to the second principal component in PCA—are uncorrelated with the first vectors (Bookstein et al. 2003). Hence, the second (and third or more) axes can be easily extracted by applying the same algorithms, presented above, to the residuals after the first vectors have been regressed out of the original data blocks. In this paper, however, our analysis and results are concerned primarily with the first singular axes.

Estimating bi-variance and tri-variance

The first three singular vectors (each from vegetation, soil, and topography data blocks) were standardized (i.e., scaled to unit length) to make them directly comparable to one another. Therefore, in the Venn diagram shown in Fig. 1 (Factor A = vegetation; Factor B = soil; Factor C = topography), the circles of these three factors were identical in size. In the context of regression modeling, the size of each circle was equivalent to the total sum of squares (TSS = $N - 1$).

We then employed the concept of extra sums of squares (ESS), tentatively treating vegetation as the response variable, with soil and topography as the predictor variables (indeed, it did not matter which factor was treated as the response variable). ESS measured the marginal reduction in the residual sum of squares that occurred due to the addition of a (group of) predictor variable(s) to the existing regression model (i.e.,

SSR (X2|X1) in Table 1). In this research, the original regression model was

$$\text{Vegetation} = f(\text{topography, soil}) \qquad (6)$$

where it was assumed that topography was the preexisting predictor variable and soil was the newly added predictor variable. The α value in the Venn diagram (Fig. 1) equaled vegetation's variance unexplained by soil and topography, i.e., residual sum of squares (SSE in Table 1). β indicated soil's sum of squares (i.e., SSR (X2|X1)), namely, the pure influence of soil on vegetation, excluding topography's effect.

To estimate γ, we changed the role of soil and topography, such that topography became the newly added predictor variable, whereas soil became the preexisting predictor variable

$$\text{Vegetation} = f(\text{soil, topography}). \qquad (7)$$

In the subsequent regression, the γ value was the same as topography's sum of squares, or the pure influence of topography on vegetation, excluding soil's effect. Finally, tri-variance (δ) was calculated by subtracting α, β, and γ values from TSS.

Now, we switched Eq. 6 into

$$\text{Soil} = f(\text{vegetation, topography}) \qquad (8)$$

where soil was the response variable, and topography and vegetation were the predictor variables. In particular, vegetation was the preexisting predictor and topography was the newly added predictor. In this regression, the ε value equaled soil's variance unexplained by vegetation and topography, i.e., the residual sum of squares. ζ could be estimated in two ways, either as TSS $- β - δ - ε$ or as topography's sum of squares (i.e., topography's pure effect on soil excluding that of vegetation). η was equivalent to TSS $- γ - δ - ζ$. As a result, bi-variance was the sum of β, γ, δ, and ζ. Finally, we expressed these raw bi-variance and tri-variance values in proportion (%) to TSS

$$\text{Final bi-variance} = 100 \times (\text{Raw bi-variance/TSS}) \qquad (9)$$

$$\text{Final tri-variance} = 100 \times (\text{Raw tri-variance/TSS}). \qquad (10)$$

All of these procedures were conducted using

R 2.14.2 (R Development Core Team 2012). An example R-code for this procedure is given in the supplementary material (see Supplement).

Comparison of PCA-based and PLS-based approaches

Before evaluating the empirical hypotheses set out in the Introduction, we first examined the structure of inferences from a traditional ordination approach and 3-B PLS. For the traditional part, we followed the steps presented in Fig. 4 so that the first principal components—each derived from vegetation, soil, and topography data blocks using PCA—were directly correlated in pairs. In the same way, the first singular axes—each extracted from vegetation, soil, and topography data blocks using 3-B PLS—were directly correlated.

Evaluation of the hypotheses

Within each system, an examination of the specific structure of bi-variances and tri-variances would be a useful implementation of the 3-B PLS, but since detailed examination of the systems has appeared in our published accounts, albeit not using 3-B PLS, we elect here to illustrate use of the method in order to address higher-level questions that transcend detailed results from specific systems. In particular, system function will be explored only in relation to evaluating the hypotheses.

H_1: *Systems under widespread (i.e., broadly acting) environmental controls have higher bi-variance and tri-variance than those characterized by localized interactive biotic–abiotic processes*. This hypothesis was evaluated using Stallins's barrier islands data because his two systems, Sapelo and South Core Banks, were well known in terms of their contrasting hydrogeomorphic disturbance regimes and the consequent ecological processes of vegetation recovery and soil formation (see Stallins and Parker 2003).

With increasing latitude along the southeastern U.S. Atlantic coast, there is a climatological gradient of increasing exposure to extra-tropical cyclones (Mather et al. 1964, Davis et al. 1993) and the consequent gradient of increasing shore disturbance exposure (Williams and Leatherman 1993). In the wave-dominated microtidal barrier islands of South Core Banks in North Carolina, relatively frequent and intensive overwash events mediate biotic–abiotic interactions: disturbance-dependent plant species, which are burial tolerant, recolonize overwash sediments using their horizontally extensive rhizome networks. Due to the stabilized sediment surface, the formation of protective foredune ridges is often inhibited and the landforms are generally flat (Godfrey and Godfrey 1976). This low-relief topography, in turn, reinforces exposure to future overwash events, which will further facilitate the establishment of the disturbance-dependent vegetation (Godfrey et al. 1979, Stallins and Parker 2003). This clearly explains a positive feedback between vegetation and geomorphology.

By contrast, the mixed-energy mesotidal barrier islands of Sapelo in Georgia, USA, exhibit a conspicuous ridge-and-swale topography. Such terrain ruggedness is maintained by dune-building taxa that enhance sedimentation rates and by burial-intolerant woody shrubs and grass species that stabilize intervening low swales. These ridge-and-swale plant species reinforce their presence by contributing to the high-relief topography that acts as a barrier to the spread of wave energy into backshore dune habitats. In both island types, these landform-modifying plants are considered ecosystem engineers, in that they regulate the distribution of soil resources and species coexistence (Jones et al. 1997, Corenblit et al. 2007, Zarnetske et al. 2012).

H_2: *Separate analyses of sub-zones produce consistently smaller or greater bi-variance and tri-variance than analysis of a whole system*. First, we estimated the bi-variance and tri-variance values for all of the seven whole systems introduced above. These whole systems encompassed a range of landforms and disturbance regimes, and each of them was apparently heterogeneous in terms of physical conditions and ecological processes. Therefore, we divided each whole system into sub-zones as follows: Costa's whole dune into northern and southern sites; Kim's whole dune into foredune, swale, and second dune; Kim's whole marsh into point bar, transition, and interior (or platform); Kupfer's whole forest into remnant levee, backswamp, and current levee; McEwan's whole forest into beech, maple, and oak communities; Stallins's whole dune in Sapelo into flat, ridge, and swale; and Stallins's whole dune in South Core Banks into flat, ridge, and swale. Then, the bi-variance and tri-variance

values for all of these 20 sub-zones were also estimated. Finally, we calculated the average difference (D_{ave}) between the bi-variances of the i whole systems and the bi-variances of their j sub-zones

$$D_{\text{ave}} = \frac{1}{n} \sum \left(x_{\text{whole}(i)} - x_{\text{sub-zone}(i)} \right) \qquad (11)$$

where n was 20 (the number of sub-zones) and x_{whole} and $x_{\text{sub-zone}}$ indicated the bi-variance values of each whole system and their sub-zones, respectively. Then, we verified if D_{ave} was significantly different from zero

$$\frac{D_{\text{ave}} - (\mu_{\text{whole}} - \mu_{\text{sub}})}{SD/\sqrt{n}} = \frac{D_{\text{ave}}}{SD/\sqrt{n}} \qquad (12)$$

where μ_{whole} and μ_{sub} indicated the population mean of the whole systems and the sub-zones, respectively. Statistically, these mean values are assumed to be same because the null hypothesis of this test is that there is no difference in the mean variances between the whole systems and the sub-zones. SD was the standard deviation of the differences. After conducting Eq. 12, we estimated the corresponding P-value assuming a t distribution (Fisher 1954). We repeated the same procedure to compare the tri-variances of the whole systems and their sub-zones.

H_3: *Bi-variance is a good indicator of tri-variance.* This hypothesis was tested by conducting a simple linear regression, in which bi-variance and tri-variance were treated as predictor and response variables, respectively.

RESULTS

Comparison of PCA-based and PLS-based approaches

In most cases, 3-B PLS revealed much stronger correlations of vegetation–soil, soil–topography, and topography–vegetation couplings than PCA (Table 2). The coefficients produced by PCA indicated the correlations of the major principal components for each pair of domains. There were only three exceptions in which the magnitude of the correlations estimated by PLS was slightly lower than that by PCA, all of which came from Stallins's Sapelo data (see the values in boldface in Table 2). These results were directly related to the difference between the two approaches with regard to the bi-variance and tri-variance values

they produced (Table 3). With only one exception (the tri-variance of Stallins's 'Ridge' data from Sapelo), 3-B PLS produced much greater bi-variance and tri-variance than PCA. In 10 cases, the PCA-based approach indicated an absence of variance simultaneously shared by vegetation, soil, and topography (i.e., tri-variance = 0).

Evaluation of the hypotheses

H_1: *Systems under widespread (i.e., broadly acting) environmental controls have higher bi-variance and tri-variance than those characterized by localized interactive biotic–abiotic processes.* This hypothesis was clearly supported: we found that both the bi-variance and tri-variance of South Core Banks were greater than those of Sapelo Island (Fig. 6 and Table 3). Notably, the difference in the tri-variance (44.8 vs. 30.0) between the two regions was greater than the difference in the bi-variance (72.9 vs. 65.0).

Table 4 demonstrates the detailed contribution of each variable or plant species to the first singular axes of topography, soil, and vegetation. In both regions, as the distance to the sea increased, surface elevation declined, while the content of silt, clay, and organic matter increased. However, each of pH and fine sand content exhibited contrasting spatial gradients between Sapelo Island and South Core Banks. Concomitantly, away from the coastline, there were decreases in the cover of dune grasses (e.g., *Panicum amarum*) and increases in the cover of swale-dwelling species like *Andropogon* spp. and *Phyla nodiflora* in Sapelo Island. In the inland part of South Core Banks, frequent overwash effects seemed to favor the increased abundance of *Spartina patens* and a dune-stabilizing moss species *Ditrichum*.

H_2: *Separate analyses of sub-zones produce consistently smaller or greater bi-variance and tri-variance than analysis of a whole system.* In Fig. 7, the bi-variance and tri-variance plots of the seven whole systems were generally located around the middle of those plots estimated from the 20 sub-zones. The average values of the bi-variance and tri-variance from these sub-zones were slightly smaller than those from the whole systems, but a quantitative comparison revealed that such differences were not statistically significant at the 0.05 significance level (see the P values (0.45 and 0.60) at the bottom of Table 3). In

Table 2. Pearson's correlation coefficients among the first principal components of vegetation (V), soil (S), and topography (T) in pairs, and among the first singular axes of vegetation, soil, and topography in pairs.

Name of ecologist and systems studied	Based on principal component analysis			Based on 3-B PLSR		
	V–S	S–T	T–V	V–S	S–T	T–V
Costa						
Whole dune	0.57**	0.09	0.33**	0.74**	−0.65**	−0.78**
Northern sites	0.25	0.14	−0.62**	−0.85**	0.87**	−0.84**
Southern sites	0.06	0.00	0.72**	−0.73**	0.62**	0.89**
Kim						
Whole dune	−0.19*	0.14	0.21*	0.55**	0.51**	0.86**
Foredune	0.65**	0.31	−0.30	0.86**	−0.79**	−0.86**
Swale	−0.14	−0.11	−0.19	−0.72**	−0.74**	0.68**
Second dune	−0.16	0.18	0.04	−0.77**	0.42**	−0.43**
Whole marsh	−0.42**	0.13	0.38*	−0.56**	0.65**	−0.77**
Bar	0.24	−0.13	−0.53**	−0.34	−0.29	0.76**
Transition	0.30	0.64**	0.30	−0.78**	−0.79**	0.88**
Interior	0.22	0.61**	0.29	0.50**	0.61**	0.44**
Kupfer						
Whole forest	−0.26*	0.13	−0.13	−0.67**	−0.59**	0.72**
Remnant levee	0.54*	0.35	−0.84**	0.77**	−0.60**	−0.87**
Backswamp	0.09	−0.41	−0.19	0.67**	0.47	0.59*
Current levee	−0.15	−0.24	0.41*	−0.58**	−0.37*	0.68**
McEwan						
Whole forest	0.67**	0.30**	0.33**	0.78**	0.46**	0.71**
Beech community	0.50**	0.32	−0.14	0.68**	−0.73**	−0.86**
Maple community	0.66**	−0.11	0.18	−0.87**	−0.47	0.60*
Oak community	−0.07	−0.11	−0.47**	−0.73**	−0.57**	0.64**
Stallins						
Whole dune (Sapelo)	−0.21*	0.56**	0.23**	0.78**	**0.55**†**	0.58**
Flat	−0.61**	−0.06	−0.21	−0.82**	0.36**	−0.42**
Ridge	−0.16	−0.24	0.27	0.58**	0.43**	**0.14**
Swale	0.02	0.52**	0.28	−0.62**	**0.51****	−0.63**
Whole dune (South Core)	0.31**	0.77**	0.37**	−0.71**	−0.78**	0.73**
Flat	0.37**	−0.72**	−0.39**	−0.72**	−0.73**	0.75**
Ridge	−0.32	0.28	−0.66**	0.50*	0.64**	0.78**
Swale	−0.19	−0.65**	0.29	−0.41	0.65**	−0.54*
Average magnitude						
Of whole systems	0.38	0.30	0.28	0.68	0.60	0.74
Of separate sub-zones	0.29	0.31	0.37	0.68	0.58	0.66
P values‡	0.18	0.91	0.10	0.86	0.74	0.10

** Significant at the 0.01 level; * Significant at the 0.05 level.

† Boldface values indicate that the magnitude of the correlation coefficient resulting from 3-B PLSR approach is slightly smaller than that from principal component analysis.

‡ The values were produced by following Eq. 12 of the main text to test whether, in each column, there was a significant difference between the seven whole systems and the 20 separate sub-zones in terms of their mean magnitude of correlations. None of the six cases proved significant.

a similar context, we compared the whole systems and the sub-zones in terms of the average magnitude of their correlations for vegetation–soil, soil–topography, and topography–vegetation pairs. In all of these relationships, the differences were not statistically significant at the 0.05 significance level (see the P values at the bottom of Table 2).

Looking at the seven whole systems individually, we found that the bi-variance and tri-variance values of each whole system were generally somewhere around the middle of those values from the associated sub-zones, with only three exceptions (see the values in boldface in

Table 3): in Costa's data, both bi-variance and tri-variance of the whole dune were smaller than those of both the northern and southern sites. In Stallins's data from Sapelo Island, the tri-variance of the whole system was consistently greater than that of the three sub-zones. Overall, these results failed to support the second hypothesis.

H_3: *Bi-variance is a good indicator of tri-variance.* There was a general trend in which tri-variance increased with increasing bi-variance, and this positive relationship was statistically significant ($R^2 = 0.61$, $P < 0.01$; Fig. 7).

Table 3. Percent variance values (bi-variance and tri-variance), identified by principal component analysis (PCA) and 3-B PLS.

Name of ecologist and systems studied	Based on PCA		Based on 3-B PLS	
	Bi-variance	Tri-variance	Bi-variance	Tri-variance
Costa				
Whole dune	41.2	3.2	**71.6†**	**45.3**
Northern sites	41.4	3.6	**84.7**	**66.1**
Southern sites	52.8	0.0	**84.2**	**47.8**
Kim				
Whole dune	13.8	0.0	76.8	28.2
Foredune	44.3	8.0	83.7	65.2
Swale	9.0	0.0	68.2	41.4
Second dune	6.7	0.0	61.0	17.0
Whole marsh	52.0	0.0	72.0	30.9
Bar	30.5	2.7	59.6	9.9
Transition	43.9	6.8	83.8	58.3
Interior	41.3	4.5	46.5	16.3
Kupfer				
Whole forest	8.6	0.7	63.4	35.9
Remnant levee	80.6	22.3	91.0	50.9
Backswamp	19.9	0.8	57.5	26.4
Current levee	20.6	1.9	58.2	21.5
McEwan				
Whole forest	47.0	8.9	77.2	34.5
Beech community	59.1	0.0	81.5	46.0
Maple community	57.1	0.0	81.8	31.4
Oak community	26.6	0.0	62.5	33.3
Stallins				
Whole dune (Sapelo)	33.8	3.3	65.0	**30.0**
Flat	48.7	0.0	69.2	**15.5**
Ridge	12.5	1.6	47.8	**0.4**
Swale	43.9	0.0	53.9	**26.2**
Whole dune (South Core)	64.0	8.9	72.9	44.8
Flat	56.3	11.9	70.6	45.2
Ridge	45.9	8.4	77.9	24.4
Swale	47.2	3.6	56.0	15.9
Average of whole systems	37.2	3.6	71.3	35.7
Average of separate sub-zones	39.4	3.8	69.0	33.0
P values‡	0.63	0.89	0.45	0.60

† Boldface values indicate that separate analyses of sub-zones produced consistently lower bi- and tri-variance (Costa's data) or higher tri-variance (Stallins's Sapelo data), than the analysis of the associated whole system. In other words, in these three cases, the bi-variance and tri-variance of the whole system did not represent the average strength of the biotic–abiotic couplings in each sub-zone. These are the only cases that support the second hypothesis (H_2). All of the other cases do not support H_2.

‡ The values were produced by following Eq. 12 of the main text to test whether, in each column, there was a significant difference between the seven whole systems and the 20 separate sub-zones with regard to their bi- or tri-variances. None of these four cases proved significant.

Discussion

Comparison of PCA-based and PLS-based approaches

We found that PCA-based approaches explained much less of the relationships between vegetation, soil, and topography in pairs than did PLS methods (Table 2). Furthermore, in many cases, the PCA-based approach statistically implied that the variance simultaneously shared by vegetation, soil, and topography does not exist (i.e., tri-variance = 0; Table 3). This is an illogical conclusion which, according to the 3-B PLS, is demonstrably false. Those who work in ecolog-ical systems would almost certainly assume there are causal links among these three natural components so that tri-variance $\neq$ 0. In our analysis, PCA was an example of ordination methodologies and we hypothesize that other ordination techniques (e.g., principal coordinate analysis, correspondence analysis, multidimensional scaling, etc.) would also have yielded much weaker correlations than those found using the PLS method. By nature, these non-PLS methods extract major axes that maximally represent the patterns of a certain data block. However, these axes may not significantly correlate with the axes (or variables) belonging

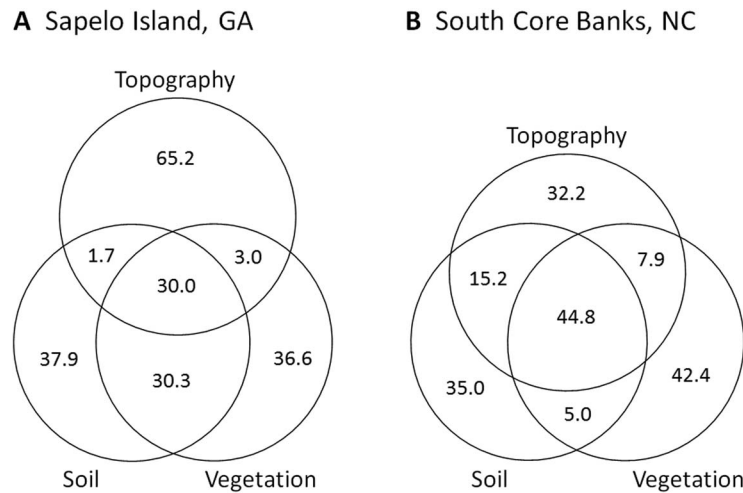**A** Sapelo Island, GA **B** South Core Banks, NC

Fig. 6. Proportion (%) of variances estimated from the relationships among the first singular axes of vegetation, soil, and topography, identified by 3-B PLS for Stallins's barrier islands data from Sapelo Island (A) and South Core Banks (B). These proportional Venn diagrams were generated using a statistical loss function and a minimization procedure (Wilkinson 2012). The three circles of each region are of the same size because the corresponding three singular axes had been standardized before estimating the proportions.

to the other blocks, thereby underestimating the amount of inter-block covariation. Our results demonstrate that this lack of correlation between ordination axes from different blocks may be a limitation when attempting to explain patterns observed in natural systems (see also Carrascal et al. 2009 for a simulation experiment).

Our point is not to suggest a blanket rejection of PCA and other conventional ordination approaches. Such approaches are useful, but, as our results indicate, there is additional or residual information that can be partitioned by PLS. Traditional methods can provide insights into the variance structure among vegetation, soils, and landforms. However, PLS is a powerful quantitative tool because it is constructed to directly estimate the covariance of the native variables of all data blocks simultaneously. We recommend the use of PLS for more integrative, holistic investigations that are explicitly focused on abiotic–biotic feedbacks and their geographic variability.

Evaluation of the hypotheses

H_1: *Systems under widespread (i.e., broadly acting) environmental controls have higher bi-variance and tri-variance than those characterized by localized interactive biotic–abiotic processes.* In ecological literature, it is widely agreed that disturbance is an environmental process that regulates plant species abundance and composition and that the effects of disturbance are interrelated with topographic conditions and soil resources (Swanson et al. 1988, Bendix 1998, Shafroth et al. 2002, Hupp and Rinaldi 2007). We interpret the high tri-variance among vegetation, soil, and topography on South Core Banks as a reflection of the profound influence of frequent exposure to extensive overwash across the relatively flat dunes in this wave-dominated coastal region. In this context, Stallins (2001) earlier found that pH and organic matter, even in the unstable, poorly sorted substrate of South Core Banks, were significantly correlated with the cover of burial-tolerant stabilizers such as *Spartina patens* and *Ditrichum* spp., which recolonized after an overwash to maintain the flat surface. These plants, soil conditions, and uniform topography reproduce one another by facilitating the spread of future overwash effects across space and the subsequent intrinsic recovery processes.

On Sapelo, the low tri-variance is attributable to the rarity of major hydrogeomorphic disturbance in this tide-dominated region, which in turn induces fine-scale habitat heterogeneity (Phillips 1999, 2000, 2004). Here, efficient dune-

Table 4. Contribution (correlation coefficient) of each variable or species to the first singular axes of topography, soil, and vegetation, identified by 3-B PLS for Stallins's data from Sapelo Island in Georgia and South Core Banks in North Carolina.

Variables	Sapelo Island	South Core Banks
Topographic variables		
Distance from the sea	−0.79**	−0.93**
Surface elevation	0.83**	0.91**
Soil variables		
pH	0.84**	−0.78**
Fine sand content	−0.60**	0.01
Silt and clay content	−0.82**	−0.86**
Carbonate	0.08	−0.16
Organic matter	−0.91**	−0.92**
Plant species†		
Ammophila brevigulata Fernald		0.18*
Andropogon species	−0.22**	
Arenaria serpyllifolia L.		−0.43**
Cenchrus species	0.03	
Chloris petrea Swartz.		−0.49**
Commelina erecta L.		−0.47**
Conyza canadensis L. Cronq.	−0.05	−0.15
Croton punctatus Jacquin	0.33**	
Cynanchum angustifolium Pers		−0.42**
Diodia teres Walter		−0.07
Ditrichum pallidum		−0.63**
Eragrostis species	−0.13	0.04
Eupatorium species	−0.39**	
Fimbristylis species	−0.26**	
Gaillardia pulchella Foug.		−0.04
Gnaphalium purpureum L.	0.00	
Heterotheca subaxillaris (Lam.) Britton & Rusby	0.11	−0.17*
Hydrocotyle bonariensis Lam.	0.06	−0.15
Juncus roemerianus Scheele	−0.26**	
Muhlenbergia filipes M.A. Curtis	−0.54**	−0.39**
Myrica cerifera L.	−0.44**	−0.21*
Oenothera humifosa Nuttall	−0.02	−0.15
Panicum amarum Ell.	0.40**	0.16
Phyla nodiflora (L.) Greene	−0.23**	
Physalis viscosa L. ssp maritima (Curtis)		−0.21*
Rubus species	−0.47**	
Rumex hastatulus Baldwin ex Ell.	−0.20*	
Smilax species	−0.49**	
Solidago species		−0.12
Spartina patens (Aiton) Muhl.		−0.19*
Strophostyles helvola (L.) Ell		0.26**
Sporobolus virginicus (L.) Kunth	0.14	
Triplasis purpurea (Walter) Chapman	0.05	
Uniola paniculata L.	0.57**	0.67**
Vulpina octoflora (Walt) Rydb.	−0.05	−0.23**

** Significant at the 0.01 level; * Significant at the 0.05 level.
† No data indicate that the associated species were not present in the corresponding study region.
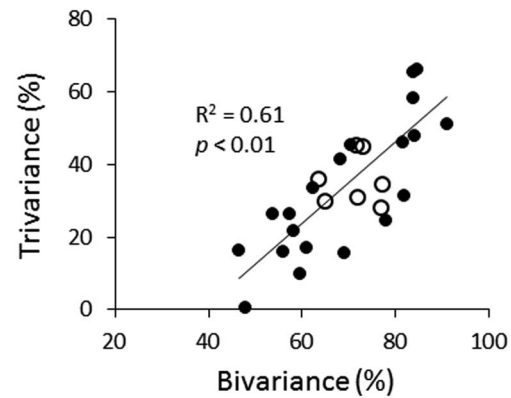

Fig. 7. The relationships between bi-variance and tri-variance estimated from the relationships among the first singular axes of vegetation, soil, and topography, identified by 3-B PLS. The seven open circles represent the seven whole systems examined, while the other data points are for the 20 separate sub-zones of those whole systems.

not necessarily correspond to the overall dune morphologies. Accordingly, the bi-variance of vegetation–topography and soil–topography pairs is very low on Sapelo. Conversely, such localized processes gradually strengthen over time the coupling between vegetation and soil on Sapelo, and the associated bi-variance is indeed greater than that on South Core Banks (i.e., 60.3% vs. 49.8%; see Fig. 6).

These discussions can further be related to the interval between disturbance events and the time it takes for an original biogeomorphic feedback among dune vegetation, soil, and landform to be restored after a disturbance event (Brunsden and Thornes 1979, Phillips 1995, Corenblit et al. 2011, Eichel et al. 2013). On Sapelo, due to the rarity of overriding disturbance, time is allowed for vegetation and soil to become closely inter-twined, and this vegetation–soil coupling likely diverges over space, contingent upon site-specific disturbance history and/or soil microbial compo-sition (Phillips 2013). In such a case of high stochasticity, the importance of dune morphologies to floristic and pedological processes should decline gradually. As a consequence, it is quite possible to observe variable plant species com-position or edaphic characteristics at different locations possessing very similar topographic conditions (Gleason 1926).

builders (e.g., *Uniola* spp.) trap suspended sediments and develop high-relief dune ridges that dampen the direct impact of future over-wash. Both rarity and dampening of disturbance effects eventually enhance the localization of vegetation and pedological processes that do

On South Core Banks, development time is too short to induce such spatial diversification of vegetation and soil factors over varying dune surface topography. Rather, frequent and spatially extensive overwash events lead to dynamic landform changes, which immediately influence spatial patterns of both vegetation and soil. Therefore, relatively tight vegetation–soil–topography couplings are likely maintained before spatial divergence (i.e., decoupling) of these relationships occurs across the dunefield.

H₂: *Separate analyses of sub-zones produce consistently smaller or greater bi-variance and tri-variance than analysis of a whole system.* In this hypothesis, we inquired what would happen to vegetation–soil–topography couplings if a whole system was divided into a number of sub-zones (Fig. 3). On Sapelo, the tri-variate couplings become weaker after division (Table 3), presumably due to the high spatial stochasticity driven by the gradual decoupling of local vegetation and soil processes from topography within each sub-zone, as discussed above. By contrast, the stronger couplings we observed after dividing Costa's whole system are related to highly different environmental settings between the northern and southern sites: these two regions are geographically separated by ~120 km, exhibiting distinct coastline orientations. Hence, they undergo differing approach angles and impacts of onshore winds, which, in turn, induce varying gradients in overwash frequency and sand transport (see Appendix). As a result, dissipative and fine-grained beaches are developed in the north, whereas reflective and coarse-grained beaches are found in the south (Calliari and Klein 1993, Calliari et al. 1996, Costa et al. 1996). To reveal significant relationships among vegetation, soil, and topography, it would be reasonable to perform separate analyses for the northern and southern regions, rather than combining these greatly different sub-zones into a single whole system.

Other than these exceptional cases observed just for the dune systems (Stallins's Sapelo and Costa's data), there is no prevalent pattern of either stronger or weaker vegetation–soil–topography couplings; rather, both occur simultaneously upon division. Thus, the bi-variance and tri-variance values of the whole systems are more or less the average of those values of their sub-zones; that is, the relationships become weaker in some sub-zones and stronger in the others. This suggests that, in many cases, the overall correlation structure among vegetation, soil, and topography is *conserved* across spatial scales. In other words, during the aggregation of sub-zones or the division of a whole landscape, there is no net loss or gain of the strength of vegetation–soil–topography couplings. Such potential scale-invariance confers some degree of freedom to ecologists who, due to their concern about spatial non-stationarity induced by landscape heterogeneity (Milne 1998), have been reluctant to predict large-scale biotic–abiotic correlative patterns by investigating and averaging the information of smaller units. Concerning this issue of scale-invariance, two caveats should be noted: (1) the sub-zones are not characterized by too high spatial stochasticity of vegetation and soil processes, as in the case of Stallin's Sapelo data; and (2) the whole system is not too spatially heterogeneous, that is, it does not aggregate physiographic types that are too different, as in the case of Costa's northern and southern sites. There is a danger that the local heterogeneity in variance structure cancels out when aggregated to the large scale.

H₃: *Bi-variance is a good indicator of tri-variance.* Despite the positive relationship we found between bi-variance and tri-variance values, it is still an open question whether bi-variance can be a good indicator of tri-variance, and further, whether there would be no need to struggle to estimate the tri-variance. In a statistical sense, bi-variance significantly explains tri-variance at the confidence level of 99% (i.e., $P < 0.01$); however, it should also be noted that 39% of the tri-variance still remains unexplained. The associated scatterplot (Fig. 7) even illustrates several cases of locally negative relations in which the bi-variance of system (or sub-zone) A is smaller than that of system (sub-zone) B, but A has a higher tri-variance than B. Given such uncertainties, we posit that, while the overall coefficient of determination looks generally encouraging ($R^2 = 0.61$), the need for estimating tri-variance as well as bi-variance still exists. In particular, the large tri-variance residuals could represent important three-way interactions among variables in the three data domains. Such insights are less accessible using traditional approaches.

*Reciprocal nature
of vegetation–soil–topography relationships*

Our 3-B Venn diagram approach (e.g., Fig. 6) closely accords with one of the current mainstream topics of ecology and, more generally, Earth sciences: feedbacks among biological and physical factors. Until recently, one-way relationships between vegetation, soil, and topography have been dominantly considered when studying the spatial pattern or evolution of landscapes. Such a unidirectional viewpoint inspired investigations that focused on the influence of one factor upon another. Accordingly, this cause-and-effect logic necessitated a division of abiotic and biotic components into groups of response and predictor variables, followed by the application of such traditional approaches as multiple regression and ordination. This causal approach has generally held true even in cases where researchers have sought to identify the unique and shared contributions of two sets of predictor variables (e.g, environmental and spatial variables) on a response variable (e.g., vegetation composition) through variance partitioning (e.g., Borcard et al. 1992, Kupfer et al. 1997). Today, Earth scientists are increasingly redirecting their attention from discrete one-way couplings to integrated two- or three-way interactions that address the importance of the reciprocal nature of vegetation–soil–topography linkages (Stallins 2006, Corenblit et al. 2007, Osterkamp and Hupp 2010). Key to this new paradigm is an understanding of the feedback mechanisms among natural phenomena that drive a perpetual, gradual modification or an abrupt collapse of the existing landscape complexity (Malanson 1999, Savage et al. 2000, Naylor et al. 2002, Bever et al. 2010, van der Putten et al. 2013).

Our results suggest that statistical approaches commonly used in contemporary ecology are not keeping pace with the conceptual advances in our understanding of multi-block and multi-directional ecological relationships. Within the constraint of conventional multivariate statistics, we face a challenge concerning how to divide our data into groups of response and predictor variables. Our results refute the idea of treating vegetation strictly as a response variable, given that there are clearly mutual influences between vegetation and environmental factors. This paper presents a method that avoids such a monodirec-

tional assumption structure, and estimates bi-variance and tri-variance by treating all involved components at the same level (i.e., a symmetrical approach). The resulting Venn diagrams fully reflect the idea of mutual interactions among vegetation, soil, and topography in a wide range of ecological systems from various regions of the world. Although the focus of this research is on 3-B data, the general procedure illustrated in Eqs. 3–5 can be extended to deal with 4-B data and beyond (Bookstein et al. 2003). In general, statistical models are static in nature; therefore, we do not claim that our 3B-PLS approach would immediately represent the dynamic nature of feedback mechanisms between environmental conditions and organisms. Rather, our point is that the symmetrical approach of, e.g., 3B-PLS, multiple factor analysis (Escofier and Pagès 1994), and co-inertia analysis (Dray et al. 2003) would be consistent with the ever-increasing perspective of reciprocal interactions between the environment and biota.

Suggestions for future research

In this article, we have introduced a new conceptual and methodological framework for understanding vegetation–soil–topography relationships. We do not posit that our approaches and perspectives are unquestionable or decisive; rather, more research should be conducted to further evaluate and improve their applicability to a variety of ecological and environmental contexts (e.g., Cope et al. 2014). We conclude this paper by presenting key research topics yet to be explored.

First, ecologists need to investigate vegetation–soil–topography couplings in more spatially explicit ways. Vegetation, soil, and geomorphic patterns and processes are essentially spatial phenomena. Hence, the associated variables and data blocks are likely to possess inherent spatial autocorrelation (SAC), or spatial structure, which is now a standard paradigm in ecology (Legendre 1993, Dray et al. 2006, Miller et al. 2007, Franklin 2009). To take into account the presence and effects of SAC, we should develop a methodology to partial out from each data block the pure variance explained by spatial filters only (Borcard and Legendre 2002, Griffith 2003, de Marco et al. 2008, Kim 2013). The resulting new vegetation, soil, and topography data blocks

could then be subjected to the same 3B-PLS procedure that we introduced in this paper.

Second, it is desirable to examine biotic–abiotic relationships over time, preferably over long timescales that encompass both periods of discrete disturbance events and biogeomorphic recovery. In the seven systems we investigated and similar to most of the previous studies, our analyses were performed based on one snapshot of patterns of vegetation, soil, and topography. In an attempt to model species distributions (Franklin 1998, Lehmann et al. 2003, Austin and van Niel 2011), we ecologists have often postulated the presence of tight couplings among community structure, edaphic conditions, and terrain settings. In contrast, there have been few attempts to examine the changing correlation structure among these factors, responding to several recurring disturbance events at the same landscape. Recently, Kim and Arthur (2014) demonstrated a gradual weakening of biotic–abiotic couplings over the course of their field surveys between 2002 and 2010, each occurring before and after six prescribed fire events in an oak-dominated temperate forest in eastern Kentucky, USA. These findings suggest that, in general, sharp breaks may exist, at which bi-variance and tri-variance transition as a result of gradual or abrupt changes in climate conditions or modifications of substrate.

Third, related to the temporal viewpoint above, there still is much uncertainty as to how bi-variance and tri-variance would change if a design for data collection worked across domains of scale. As mentioned, distinctive ecological patterns and processes are known to be dominant at different spatial scales (Ludwig et al. 2000). This scalar discrepancy has been conceptualized as transition between domains of scale (Wiens 1989) or discrete breaks (O'Neill and King 1998). Here, domain indicates "a portion of the scale spectrum within which process-pattern relationships are consistent regardless of scale" (Wiens 1989:392). An attempt to extrapolate the information of vegetation–soil–topography relationships acquired at a particular scale, to other scales within the same domain, may be successful; however, an extrapolation across the disjunction between domains will likely fail. As such, a key question in future research will be how to develop scaling theories that effectively link bi-variances and tri-variances estimated at different spatial scales (Wiens 1989, Ludwig et al. 2000).

Last, more disturbance types, other than just hydrogeomorphic disturbance in riparian and coastal systems, should receive attention. The exact trajectories of bi-variance and tri-variance through time and across space may differ among different biogeomorphic contexts, due to sediment mobility, landform stability, the type of plants, and the disturbance regime. From a rugged mountain forested landscape to a highly mobile marine sea floor sandy substrate, there could be some linkages between systems in terms of bi-variance and tri-variance. How variances change through time and in response to the type of disturbance forcings (historically prevalent versus novel) would not be expected to be the same, but each would exhibit patterns of convergence and divergence in variances as feedbacks between the biota and the geomorphic context developed (Phillips 1999, 2004). In this way, bi-variances and tri-variances potentially become ways of tracking the multicollinear relationships among vegetation, soil, and topography in various ecological systems. Eventually, bi-variances and tri-variances can be used to predict overall ecosystem responses to changing climate conditions and disturbance regimes.

Conclusions

This paper has addressed a critical, underdeveloped aspect of ecology: conceptual and analytical integration across multiple data domains. Often ecologists, especially as the era of big data progresses, deal with data from fundamentally, sometimes radically different domains. A given project may have a data block on community composition, soil (or water) chemistry, geography, genetic data (sequence or expression arrays) or anthropogenic impacts. Typically, we attempt integration of data domains by combining principal components analysis (a within-block approach) and two-block multivariate ordinations. But we miss a staggering amount of information by doing so, as illustrated in our paper using several case studies. Partial least squares regression, though an established tool in chemometrics and bioinformatics (e.g., Cope et al. 2014), is important to develop and popularize in ecology.

We developed three-block partial least squares, an intuitive graphical method for summarizing data relations and illustrated high-order questions that can be addressed using vegetation, soil, and topography examples from several ecological systems and their respective sub-zones. We demonstrated that bi-variances and tri-variances provide useful insight into how the strength of couplings among vegetation, soil, and topography data blocks varies across scales and disturbance regimes.

A perspective shift like we are describing is important as ecology grows even more integrative. Until actually measured, trivariate relationships cannot be fully understood with even the most detailed bivariate approaches. Though we describe the simplest case of multi-variance beyond the usual two-block linear statistical model, this approach can be extended seamlessly to any number of data domains, making true integration possible, so that the long-held ideal of ecology being a holistic science can be more fully realized.

Acknowledgments

Literature Cited

Aagaard, T., N. Nielsen, and J. Nielsen. 1995. Skallingen: origin and evolution of a barrier spit. Meddelelser fra Skalling-Laboratoriet, Bind 35. C.A. Reitzels, Denmark.

Abdi, H. 2007. Partial least square regression (PLS regression). Pages 792–795 in N. J. Salkind, editor. Encyclopedia of measurement and statistics. Sage, Thousand Oaks, California, USA.

Abdi, H., L. J. Williams, and D. Valentin. 2013. Multiple factor analysis: principal component analysis for multitable and multiblock data sets. WIREs Computational Statistics 5:149–179.

Austin, M. P. 2013. Inconsistencies between theory and methodology: a recurrent problem in ordination studies. Journal of Vegetation Science 24:251–268.

Austin, M. P., and K. P. van Niel. 2011. Improving species distribution models for climate change studies: variable selection and scale. Journal of Biogeography 38:1–8.

Bécue-Bertaut, M., and J. Pagès. 2008. Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. Computational Statistics and Data Analysis 52:3255–3268.

Beers, T. W., P. E. Dress, and L. C. Wensel. 1966. Aspect transformation in site productivity research. Journal of Forestry 64:691–692.

Bendix, J. 1998. Impact of a flood on southern California riparian vegetation. Physical Geography 19:162–174.

Bernardi, H., C. V. Cordazzo, and C. S. B. Costa. 1987. Efeito de ressacas sobre *Blutaparon portulacoides* (St. Hill) Mears, nas dunas costeiras do sul do Brasil. Ciência e Cultura 39:545–547.

Bever, J. D., I. A. Dickie, E. Facelli, J. M. Facelli, J. Klironomos, M. Moora, M. C. Rillig, W. D. Stock, M. Tibbett, and M. Zobel. 2010. Rooting theories of plant community ecology in microbial interactions. Trends in Ecology and Evolution 25:468–478.

Birks, H. J. B. 1996. Statistical approaches to interpreting diversity patterns in the Norwegian mountain flora. Ecography 19:332–340.

Boerner, R. E. J. 2006. Unraveling the Gordian Knot: interactions among vegetation, topography, and soil properties in the central and southern Appalachians. Journal of the Torrey Botanical Society 133:321–361.

Bookstein, F. L., P. Gunz, P. Mitterœcker, H. Prossinger, K. Schæfer, and H. Seidler. 2003. Cranial integration in *Homo*: singular warps analysis of the midsagittal plane in ontogeny and evolution. Journal of Human Evolution 44:167–187.

Borcard, D., F. Gillet, and P. Legendre. 2011. Numerical ecology with R. Springer, New York, New York, USA.

Borcard, D., and P. Legendre. 2002. All-scale analysis of ecological data by means of principal coordinates of neighbor matrices. Ecological Modelling 153:51–68.

Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. Ecology 73:1045–1055.

Brown, R. L., and R. K. Peet. 2003. Diversity and invasibility of southern Appalachian plant communities. Ecology 84:32–39.

Brunsden, D., and J. B. Thornes. 1979. Landscape sensitivity and change. Transactions of the Institute of British Geographers 4:463–484.

Calliari, L. J., and A. H. F. Klein. 1993. Características morfodinâmicas e sedimentológicas das praias Oceânicas entre Rio Grande e Chuí, RS. Pesquisas 20:48–56.

Calliari, L. J., A. H. F. Klein, and F. C. R. Barros. 1996. Beach differentiation along the Rio Grande do Sul

coastline (Southern Brazil). Revista Chilena de Historia Natural 69:485–493.

Carrascal, L. M., I. Galván, and O. Gordo. 2009. Partial least squares regression as an alternative to current regression methods used in ecology. Oikos 118:681–690.

Conrad, O. 2002. DiGeM terrain analysis software. http://downloads.sourceforge.net/saga-gis/digem_2.0_bin.zip

Conrads, P. A., T. D. Feaster, and L. G. Harrelson. 2008. The effects of the Saluda Dam on the surfacewater and groundwater hydrology of the Congaree National Park floodplain, South Carolina. U.S. Geological Survey Scientific Investigations Report 2008–5170. U.S. Department of the Interior, U.S. Geological Survey, Reston, Virginia, USA.

Cope, J. L., A. J. M. Hammett, E. A. Kolomiets, A. K. Forrest, K. W. Golub, E. B. Hollister, T. J. DeWitt, T. J. Gentry, M. T. Holtzapple, and H. H. Wilkinson. 2014. Evaluating the performance of carboxylate platform fermentations across diverse inocula originating as sediments from extreme environments. Bioresource Technology 155:388–394.

Cordazzo, C. V., and U. Seeliger. 1988. Phenological and biogeographical aspects of dune plant communities in southern Brazil. Plant Ecology 75:169–173.

Corenblit, D., A. C. W. Baas, G. Bornette, J. Darrozes, S. Delmotte, R. A. Francis, A. M. Gurnell, F. Julien, R. J. Naiman, and J. Steiger. 2011. Feedbacks between geomorphology and biota controlling Earth surface processes and landforms: a review of foundation concepts and current understandings. Earth-Science Reviews 106:307–331.

Corenblit, D., E. Tabacchi, J. Steiger, and A. M. Gurnell. 2007. Reciprocal interactions and adjustments between fluvial landforms and vegetation dynamics in river corridors: a review of complementary approaches. Earth-Science Reviews 84:56–86.

Costa, C. S. B., C. V. Cordazzo, and U. Seeliger. 1996. Shore disturbance and dune plant distribution. Journal of Coastal Research 12:133–140.

Cowles, H. C. 1899. The ecological relations of the vegetation of the sand dunes of Lake Michigan. Botanical Gazette 27:95–117.

Davies, K. F., P. Chesson, S. Harrison, B. D. Inouye, B. A. Melbourne, and K. J. Rice. 2005. Spatial heterogeneity explains the scale dependence of the native–exotic diversity relationship. Ecology 86:1602–1610.

Davis, R. E., R. Dolan, and G. Demme. 1993. Synoptic climatology of Atlantic coast north-easters. International Journal of Climatology 13:171–189.

de Jong, S. 1993. SIMPLS: an alternative approach to partial least squares regression. Chemometrics and Intelligent Laboratory Systems 18:251–263.

de Marco, P., J. A. F. Diniz, and L. M. Bini. 2008. Spatial

analysis improves species distribution modelling during range expansion. Biology Letters 4:577–580.

Dray, S., D. Chessel, and J. Thioulouse. 2003. Co-inertia analysis and the linking of ecological data tables. Ecology 84:3078–3089.

Dray, S., P. Legendre, and P. R. Peres-Neto. 2006. Spatial modeling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). Ecological Modelling 196:483–493.

Eichel, J., M. Krautblatter, S. Schmidtlein, and R. Dikau. 2013. Biogeomorphic interactions in the Turtmann glacier forefield, Switzerland. Geomorphology 201:98–110.

Environmental Systems Research Institute. 2000. ESRI GIS and mapping software. ESRI, Redlands, California, USA.

Escofier, P., and J. Pagès. 1994. Multiple factor analysis (AFMULT package). Computational Statistics and Data Analysis 18:121–140.

Esposito Vinzi, V., W. W. Chin, J. Henseler, and H. Wang. 2010. Handbook of partial least squares: concepts, methods and applications. Springer, New York, New York, USA.

Farrar, D. E., and R. R. Glauber. 1967. Multicollinearity in regression analysis: the problem revisited. Review of Economics and Statistics 49:92–107.

Fisher, R. A. 1954. Statistical methods for research workers. Oliver and Boyd, Edinburgh, UK.

Franklin, J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. Journal of Vegetation Science 9:733–748.

Franklin, J. 2009. Mapping species distributions. Cambridge University Press, Cambridge, UK.

Frazer, G. W., C. D. Canham, and K. P. Lertzman. 1999. Gap Light Analyzer (GLA): imaging software to extract canopy structure and gap light transmission indices from true-color fisheye photographs. Simon Fraser University, Burnaby, British Columbia, Canada and the Cary Institute of Ecosystem Studies, Millbrook, New York, USA.

Galton, F. 1888. Co-relations and their measurement, chiefly from anthropometric data. Proceedings of the Royal Society of London 45:135–145.

Geladi, P., and B. R. Kowalski. 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta 185:1–17.

Ginter, C. C., T. J. DeWitt, F. E. Fish, and C. D. Marshall. 2012. Fused traditional and geometric morphometrics demonstrate pinniped whisker diversity. PLoS ONE 7:e34481.

Gleason, H. A. 1926. The individualistic concept of the plant association. Bulletin of the Torrey Botanical Club 53:7–26.

Godfrey, P. J., and M. M. Godfrey. 1976. Barrier island ecology of Cape Lookout National Seashore and vicinity, North Carolina. National Park Service

Monograph Series, publication number 9. U.S. Government Printing Office, Washington, D.C., USA.

Godfrey, P. J., S. P. Leatherman, and R. Zaremba. 1979. A geobotanical approach to classification of barrier beach systems. Pages 99–126 in S. P. Leatherman, editor. Barrier islands. Academic Press, New York, New York, USA.

Gower, J. C. 1975. Generalized Procrustes analysis. Psychometrika 40:33–51.

Gower, J. C., and G. B. Dijksterhuis. 2004. Procrustes problems. Oxford University Press, Oxford, UK.

Graham, M. H. 2003. Confronting multicollinearity in ecological multiple regression. Ecology 84:2809–2815.

Griffith, D. A. 2003. Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer-Verlag, Berlin, Germany.

Hupp, C. R., and M. Rinaldi. 2007. Riparian vegetation patterns in relation to fluvial landforms and channel evolution along selected rivers of Tuscany (Central Italy). Annals of the Association of American Geographers 97:12–30.

Jones, C. G., J. H. Lawton, and M. Shachak. 1997. Positive and negative effects of organisms as physical ecosystem engineers. Ecology 78:1946–1957.

Jones, R. L. 2005. Plant life of Kentucky: an illustrated guide to the vascular flora. University of Kentucky Press, Lexington, Kentucky, USA.

Kendall, M. 1938. A new measure of rank correlation. Biometrika 30:81–89.

Kim, D. 2012. Biogeomorphic feedbacks drive dynamics of vegetation–landform complex in a coastal riparian system. Ecosphere 3:art74.

Kim, D. 2013. Incorporation of multi-scale spatial autocorrelation in soil moisture–landscape modeling. Physical Geography 34:441–455.

Kim, D., and M. A. Arthur. 2014. Changes in community structure and species–landform relationship after repeated fire disturbance in an oak-dominated temperate forest. Ecological Research 29:661–671.

Kim, D., D. M. Cairns, J. Bartholdy, and C. L. S. Morgan. 2012. Scale-dependent correspondence of floristic and edaphic gradients across salt marsh creeks. Annals of the Association of American Geographers 102:276–294.

Kim, D., and K. B. Yu. 2009. A conceptual model of coastal dune ecology synthesizing spatial gradients of vegetation, soil, and geomorphology. Plant Ecology 202:135–148.

Kim, D., K. B. Yu, and S. J. Park. 2008. Identification and visualization of complex spatial pattern of coastal dune soil properties using GIS-based terrain analysis and geostatistics. Journal of Coastal Research 24(4C):50–60.

Kupfer, J. A., G. P. Malanson, and J. R. Runkle. 1997. Factors influencing species composition in canopy gaps: the importance of edge proximity in Hueston Woods, Ohio. Professional Geographer 49:165–178.

Kupfer, J. A., K. M. Meitzen, and A. R. Pipkin. 2010. Hydrogeomorphic controls of early post-logging successional pathways in a southern floodplain forest. Forest Ecology and Management 259:1880–1889.

Lee, W. T. 1996. Standard illustrations of Korean plants. Academy Publishing, Seoul, South Korea.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology 74:1659–1673.

Legendre, P., and L. Legendre. 2012. Numerical ecology. Third edition. Elsevier, Amsterdam, The Netherlands.

Lehmann, A., J. M. Overton, and J. R. Leathwick. 2003. GRASP: generalized regression analysis and spatial prediction. Ecological Modelling 157:189–207.

Lepš, J., and P. Šmilauer. 2003. Multivariate analysis of ecological data using CANOCO. Cambridge University Press, Cambridge, UK.

Levin, S. A. 1992. The problem of pattern and scale in ecology. Ecology 73:1943–1967.

Lonsdale, W. M. 1999. Global patterns of plant invasions and the concept of invasibility. Ecology 80:1522–1536.

Ludwig, J. A., J. A. Wiens, and D. J. Tongway. 2000. A scaling rule for landscape patches and how it applies to conserving soil resources in savannas. Ecosystems 3:84–97.

Malanson, G. P. 1999. Considering complexity. Annals of the Association of American Geographers 89:746–753.

Mather, J. R., H. A. Adams, and G. A. Yoshioka. 1964. Coastal storms of the eastern United States. Journal of Applied Meteorology 3:693–706.

McEwan, R. W., and R. N. Muller. 2006. Spatial and temporal dynamics in canopy dominance of an old-growth mixed mesophytic forest. Canadian Journal of Forest Research 36:1536–1550.

McEwan, R. W., and R. N. Muller. 2011. Dynamics, diversity and resource gradient relationships in the herbaceous layer of an old-growth Appalachian forest. Plant Ecology 212:1179–1191.

McEwan, R. W., R. N. Muller, M. A. Arthur, and H. H. Housman. 2000. Temporal and ecological patterns of flowering dogwood mortality in the mixed mesophytic forest of eastern Kentucky. Journal of the Torrey Botanical Society 127:221–229.

Miller, J., J. Franklin, and R. Aspinall. 2007. Incorporating spatial dependence in predictive vegetation models. Ecological Modelling 202:225–242.

Miller, T. E., E. S. Gornish, and H. L. Buckley. 2010. Climate and coastal dune vegetation: disturbance, recovery, and succession. Plant Ecology 206:97–104.

Milne, B. T. 1998. Motivation and benefits of complex systems approaches in ecology. Ecosystems 1:449–456.

Muller, R. N. 1982. Vegetation patterns in the mixed mesophytic forests of eastern Kentucky. Ecology 63:1901–1917.

Næs, T., and H. Martens. 1985. Comparison of prediction methods for multicollinear data. Communications in Statistics: Simulation and Computation 14:545–576.

Naylor, L. A., H. A. Viles, and N. E. A. Carter. 2002. Biogeomorphology revisited: looking towards the future. Geomorphology 47:3–14.

O'Neill, R. V., and A. W. King. 1998. Homage to St. Michael; or, Why are there so many books on scale?. Pages 3–15 in D. L. Patterson and T. Parker, editors. Ecological scale: theory and applications. Columbia University Press, New York, New York, USA.

Osterkamp, W. R., and C. R. Hupp. 2010. Fluvial processes and vegetation: glimpses of the past, the present, and perhaps the future. Geomorphology 116:274–285.

Palmer, M. W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. Ecology 74:2215–2230.

Pearson, K. 1920. Notes on the history of correlation. Biometrika 13:25–45.

Phillips, J. D. 1995. Biogeomorphology and landscape evolution: the problem of scale. Geomorphology 13:337–347.

Phillips, J. D. 1999. Divergence, convergence, and self-organization in landscapes. Annals of the Association of American Geographers 89:466–488.

Phillips, J. D. 2000. Signatures of divergence and self-organization in soils and weathering profiles. Journal of Geology 108:91–102.

Phillips, J. D. 2004. Divergence, sensitivity, and nonequilibrium in ecosystems. Geographical Analysis 36:369–383.

Phillips, J. D. 2013. Networks of historical contingency in Earth surface systems. Journal of Geology 121:1–16.

Pielou, E. C. 1984. The interpretation of ecological data: a primer on classification and ordination. John Wiley & Sons, New York, New York, USA.

Raunkiær, C. 1934. The life forms of plants and statistical plant geography: being the collected papers of C. Raunkiær. Clarendon Press, Oxford, UK.

R Development Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org

Rosipal, R., and N. Krämer. 2006. Overview and recent advances in partial least squares. Pages 34–51 in C. Saunders, M. Grobelnik, S. Gunn, and J. Shawe-Taylor, editors. Subspace, latent structure and feature selection. Springer, Berlin, Germany.

Savage, M., B. Sawhill, and M. Askenazi. 2000. Community dynamics: What happens when we rerun the tape? Journal of Theoretical Biology 205:515–526.

Shafroth, P. B., J. C. Stromberg, and D. T. Patten. 2002. Riparian vegetation response to altered disturbance and stress regimes. Ecological Applications 12:107–123.

Sharitz, R. R., and W. J. Mitsch. 1993. Southern floodplain forests. Pages 311–372 in W. H. Martin, S. G. Boyce, and A. C. Echternacht, editors. Biodiversity of the southeastern United States: upland terrestrial communities. John Wiley, New York, New York, USA.

Stachowicz, J. J., R. B. Whitlatch, and R. W. Osman. 1999. Species diversity and invasion resistance in a marine ecosystem. Science 286:1577–1579.

Stallins, J. A. 2001. Soil and vegetation patterns in barrier-island dune environment. Physical Geography 22:79–98.

Stallins, J. A. 2006. Geomorphology and ecology: unifying themes for complex systems in biogeomorphology. Geomorphology 77:207–216.

Stallins, J. A., and A. J. Parker. 2003. The influence of complex systems interactions on barrier island dune vegetation pattern and process. Annals of the Association of American Geographers 93:13–29.

Stohlgren, T. J., R. R. Bachand, Y. Onami, and D. Binkley. 1998. Species-environment relationships and vegetation patterns: effects of spatial scale and tree life-stage. Plant Ecology 135:215–228.

Stohlgren, T. J., D. T. Barnett, and J. Kartesz. 2003. The rich get richer: patterns of plant invasions in the United States. Frontiers in Ecology and the Environment 1:11–14.

Streissguth, A. P., P. Sampson, and H. Barr. 1993. The enduring effects of prenatal alcohol exposure on child development. University of Michigan Press, Ann Arbor, Michigan, USA.

Swanson, F. J., T. K. Kratz, N. Caine, and R. G. Woodmansee. 1988. Landform effects on ecosystem patterns and processes. BioScience 38:92–98.

ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67:1167–1179.

ter Braak, C. J. F., and S. Juggins. 1993. Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. Hydrobiologia 269/270:485–502.

ter Braak, C. J. F., and I. C. Prentice. 1988. A theory of gradient analysis. Advances in Ecological Research 18:271–317.

Tilman, D. 1997. Community invasibility, recruitment

limitation, and grassland biodiversity. Ecology 78:81–92.

Turner, M. G. 2010. Disturbance and landscape dynamics in a changing world. Ecology 91:2833–2849.

USACE [United Stated Army Corps of Engineers]. 2002. HEC-RAS, River Analysis System user's manual. U.S. Army Corps of Engineers, Hydrologic Engineering Center, Davis, California, USA.

USACE [United Stated Army Corps of Engineers]. 2005. HEC-GeoRAS: GIS Tools for support of HEC-RAS using ArcGIS. U.S. Army Corps of Engineers, Hydrologic Engineering Center, Davis, California, USA.

USDA NRCS [United States Department of Agriculture, Natural Resources Conservation Service]. 2008. Web soil survey. USDA Natural Resources Conservation Service, Washington, D.C., USA. http://websoilsurvey.nrcs.usda.gov

van der Putten, W. H., et al. 2013. Plant–soil feedbacks: the past, the present and future challenges. Journal of Ecology 101:265–276.

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology 3:385–397.

Wilkinson, L. 2012. Exact and approximate area-proportional circular Venn and Euler diagrams. IEEE Transactions on Visualization and Computer Graphics 18:321–331.

Williams, A. T., and S. P. Leatherman. 1993. Process-form relationships of USA East Coast barrier islands. Zeitschrift für Geomorphologie 37:179–197.

Wold, S., A. Ruhe, H. Wold, and W. J. Dunn III. 1984. The collinearity problem in linear regression: the partial least squares (PLS) approach to generalized inverses. SIAM Journal on Scientific and Statistical Computing 5:735–743.

Wright, S. 1921. Correlation and causation. Journal of Agricultural Research 20:557–585.

Young, D. R., S. T. Brantley, J. C. Zinnert, and J. K. Vick. 2011. Landscape position and habitat polygons in a dynamic coastal environment. Ecosphere 2:art71.

Zarnetske, P. L., S. D. Hacker, E. W. Seabloom, P. Ruggiero, J. R. Killian, T. B. Maddux, and D. Cox. 2012. Biophysical feedback mediates effects of invasive grasses on coastal dune shape. Ecology 93:1439–1450.

SUPPLEMENTAL MATERIAL

APPENDIX

Costa's study area and data collection

Costa collected vegetation, soil, and topography data sets at coastal dunes along the southernmost 217 km-long stretch of Brazil (32°55′ S, 52°52′ W; Fig. A1). Although Costa et al. (1996) originally selected a total of 10 sites for their research, in this paper, we restricted our analysis to the six representative locations as in the map: four northern sites and two southern sites. The northern dunes exhibit relatively high foredune ridges (~7 m) and the southern sites show relatively large seashore deflation plains. These differences between dunefields are associated with shoreline orientation and local hydrodynamics, which determinate dissipative beaches with fine quartz sands at the north and characteristic reflective beaches with high proportion of shell fragments at the south (Calliari and Klein 1993, Calliari et al. 1996). These regions belong to a warm temperate transition zone (Cordazzo and Seeliger 1988) and to a micro-tidal zone with mean range being less than 50 cm (Bernardi et al. 1987).

At each site, Costa et al. (1996) established five 100 m long and 1 m wide strip transects perpendicular to the coastline, starting at the leading edge of the vegetation. Each transect was divided into 10 zones with the equal length of 10 m. Along each zone, the presence of plant species was recorded at 2-m intervals using 1-m^2 square quadrats. Then, all floristic data from these zones were pooled to represent the overall frequency of each species. In each zone, they also examined pH of superficial sand (10–20 cm depth), and the proportion (%) of coarse (>0.50 mm), medium (0.25–0.50 mm), fine (0.062–0.25 mm), and clay/silt (<0.062 mm) sediments. Two topographic attributes of each zone were included in the

Fig. A1. Location of Costa's study area in southern Brazil. The four white circles at the top indicate northern sites, whereas the two black circles at the bottom represent southern sites (see Tables 2 and 3 of the main article).

present work: surface elevation (m) and distance from the coastline (m), which corresponded to the gradients in sedimentation rates and salinity.

Kim's study area and data collection

The Sindu coastal dunefield is located in west Korean Peninsula (36°50′33.81″ N, 126°11′37.85″ E; Fig. A2A). The foredune ridge is adjacent to the beach and is dominated by *Elymus mollis*, a typical grassy dune species in northeastern Asia, and a shrub species *Vitex rotundifolia*. The second dune is situated in the landward part of Sindu and is characterized by the presence of woody individuals of *Robinia pseudoaccacia* and an invasive plant *Phytolacca americana*. The inter-dune depression is prone to the concentration of groundwater flows from the surrounding dune ridges; hence, this swale experiences frequent submergence in freshwater during wet summer seasons.

In each of the 126 grids (20 × 20 m; Fig. A2B),

Kim systematically placed five square quadrats (Fig. A2C) and recorded the percent cover of each plant species present in each quadrat. The size of quadrats was 1 × 1 m when there were no woody individuals and 2 × 2 m when any of them was present. Cover data from the five quadrats were then averaged to represent the grid as a whole. Species nomenclature followed Lee (1996).

Kim also collected subsurface soil samples (∼1 kg at 15–20 cm depth) in the middle of each grid. Soil properties analyzed included soil moisture (%), find sand (%), organic matter (%), soil pH, electrical conductivity (mS/cm), nitrate (NO_3^-) (mg/kg), total phosphorus (mg/kg), and exchangeable Na^+, Mg^{2+}, Ca^{2+}, K^+ (mg/kg). See Kim et al. (2008) for more detailed information of analytic procedures.

Using a total station (Sokkia SET5FS), Kim acquired horizontal (x, y) and vertical (z) coordinates of 1383 surface points across the study plot.
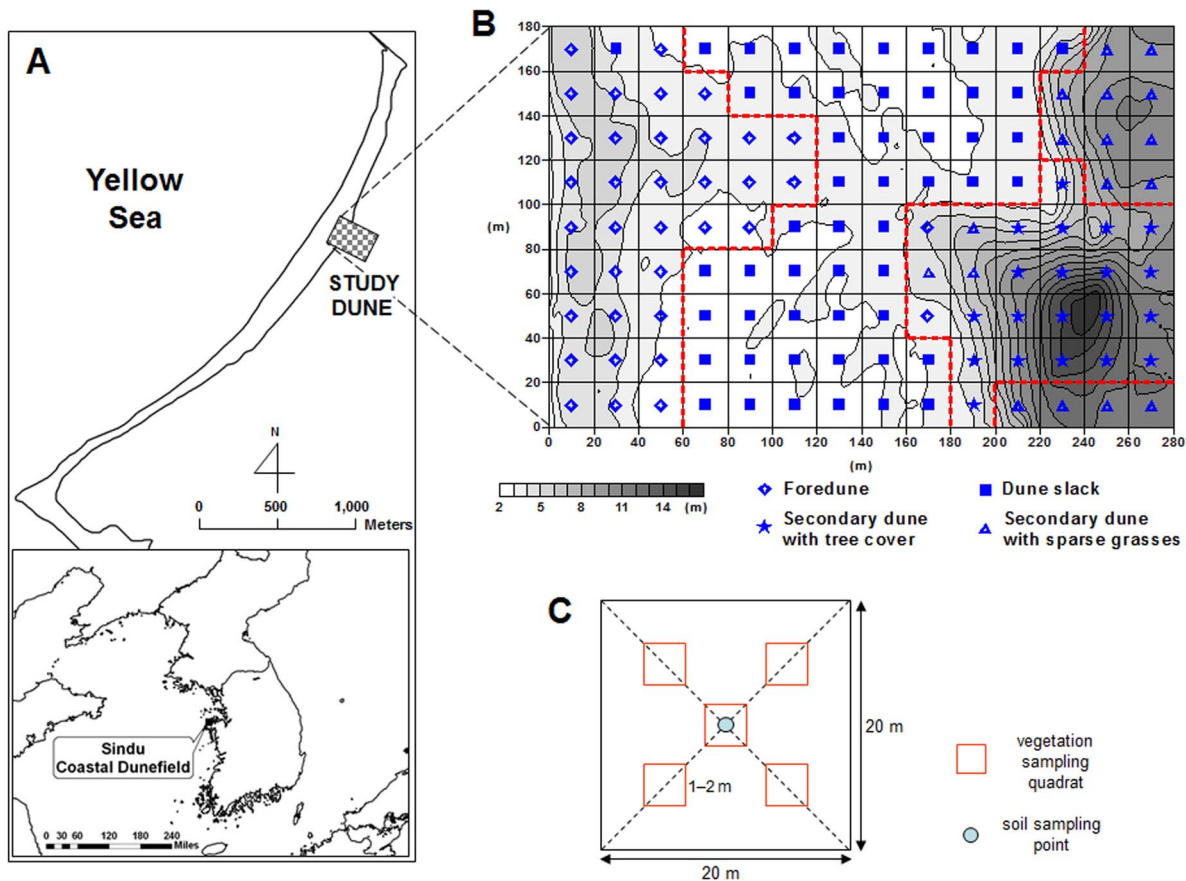
Fig. A2. The location of the Sindu coastal dune in South Korea (A) with spatial zonation of habitat types (B). The zonation was based on the result of correspondence analysis on floristic data. Each symbol represents soil sampling point as well as the habitat types. Panel C shows the location of five quadrats within each grid of 20 × 20 m. The size of quadrats was exaggerated to make them more visible.

He then used Surfer 7.0 (Golden Software) to perform semivariogram analysis and to create a digital elevation model (DEM). DiGeM 2.0 (Conrad 2002) was utilized to acquire values for terrain descriptors for each grid of the DEM in tandem with the Getgridvalue extension of Arc-View GIS 3.3 (Environmental Systems Research Institute 2000). Finally, eight topographic variables were calculated: elevation (m), slope gradient (radians), aspect (radians), plan curvature (radians/m), profile curvature (radians/m), upslope area (m^2), wetness index (dimensionless), and distance from the coastline (m).

The Skallingen salt marsh lies at the northern end of the Wadden Sea (55°30′29.85″ N, 8°15′03.08″ E; Fig. A3A). The marsh has a micro-tidal condition with the mean range being approximately 150 cm. The range becomes about 170 cm at spring and 130 cm at neap tides. The marsh has formed since the beginning of the 20th century when the accretion of silt and clay started on top of the peninsula's centuries old extensive sand flat (Aagaard et al. 1995).

Along each transect, Kim put 1 × 1 m square quadrats, each of which was subdivided into 25 cells of 0.2 × 0.2 m (Fig. A3B). The number of quadrats and the distance between them increased as he moved from marsh edges toward the interiors: (1) two or three quadrats were separated by 0.5 m in the zone of point bar, (2) two or three quadrats were separated by 1 m in transitional zones, and (3) four or five quadrats were separated by 2–4 m in marsh interiors. Two replicate quadrats were established at every site
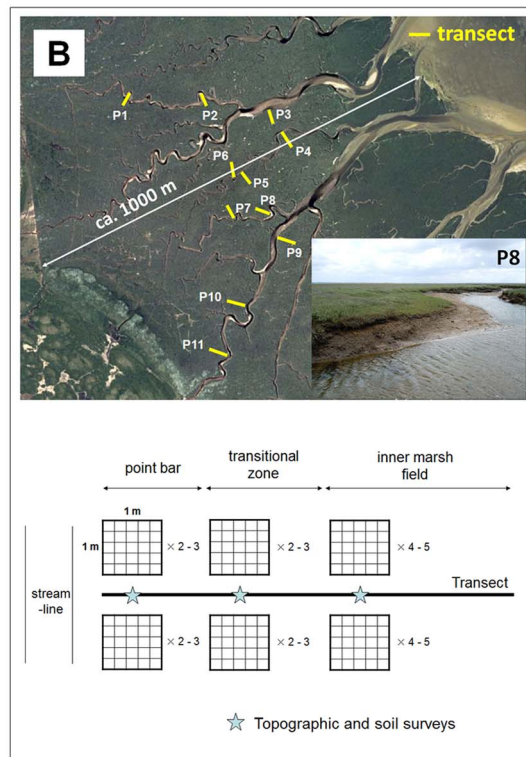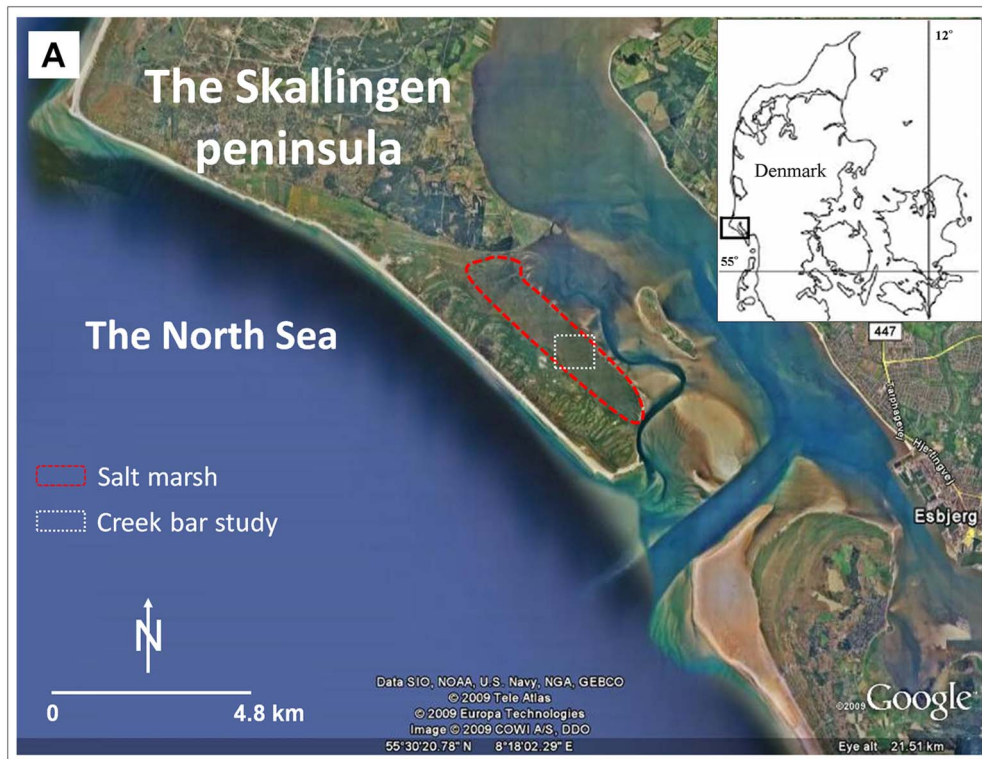
Fig. A3. Location of the study area in southwestern Denmark (A). The establishment of permanent plots across tidal creeks and the associated sampling designs (B).
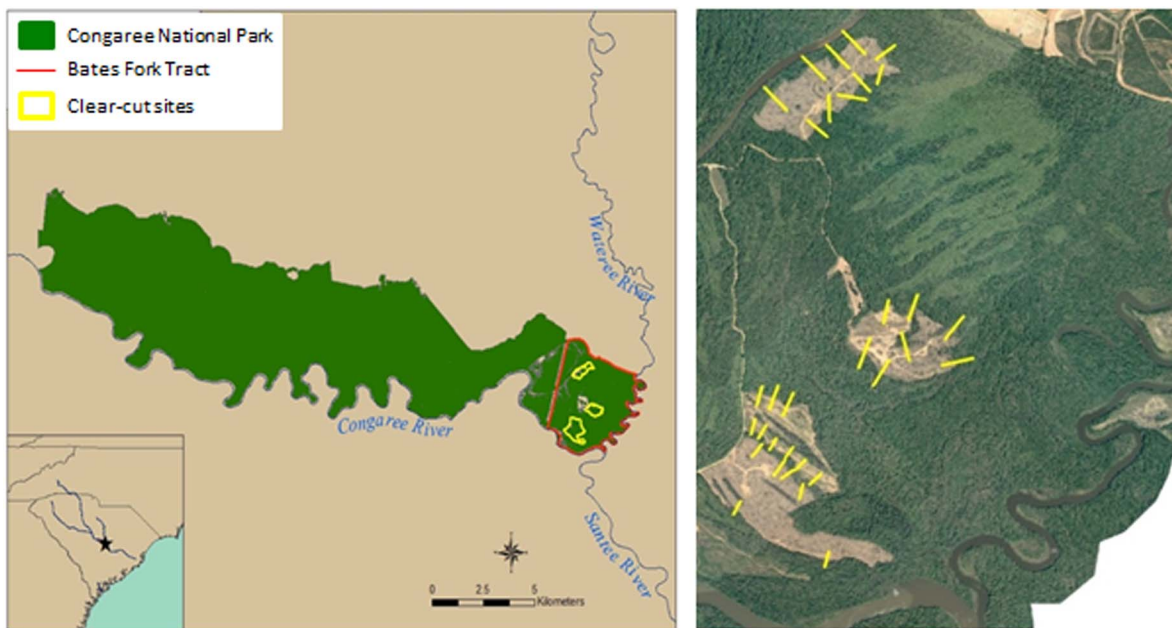
Fig. A4. Study area: the Bates Fork tract of the Congaree National Park.

surveyed. Kim recorded the presence of vascular plant species in each cell (0.2 × 0.2 m); and hence, the frequency of each species in one quadrat varied between 0 and 25. Then, the frequencies from two replicate quadrats were averaged.

At the middle between the two replicate quadrats, soil and topographic surveys were conducted (Fig. A3B). For the soil sampling, Kim used a cylindrical core with diameter and depth being 4.5 cm and 10 cm, respectively. A total of 10 soil physical and chemical analyses were then performed on each soil sample including bulk density, soil pH, electrical conductivity (umhos/cm), phosphorus (mg/kg), sulfur (mg/kg), nitrate (NO_3^-; mg/kg), Na^+ (mg/kg), K^+ (mg/kg), Ca^{2+} (mg/kg), and Mg^{2+} (mg/kg). For the topographic survey, a differential Global Positioning System (Trimble R4 GPS Receiver and Trimble Recon Controller) was used. The horizontal and vertical precision was 1 cm and 2 cm, respectively. Distance (m) was measured from each point to the streamline. See Kim et al. (2012) for more detailed information of analytic procedures.

Kupfer's study area and data collection

The Kupfer data were collected on the floodplain of the 969-ha Bates Fork tract within Congaree National Park (NP). Originally set aside as a National Monument in 1976 and redesignated as a National Park in 2003, Congaree NP protects the largest intact expanse of old-growth bottomland hardwood floodplain forest in the southeastern USA and has been designated as an International Biosphere Reserve and a Ramsar wetland site. The Bates Fork tract is located at the easternmost extent of the park at the confluence of the Congaree and Wateree Rivers (33°46′ N, 80°37′ W; Fig. A4), which have upstream drainage areas of ~21500 km² and 14500 km², respectively. Discharges of these rivers are generally highest from January to April, when evapotranspiration is low, but they are also regulated to some degree by dams on the Saluda River, one of the two major tributaries to the Congaree River, and the main stem of the Wateree River (Conrads et al. 2008). Floodplain geomorphology at the study area reflects hydrologic influences from both rivers and is characterized by levee complexes, ridge and swale systems, abandoned channel meanders, and low gradient backswamps.

The Bates Fork tract was not within the original boundary of Congaree NP, but was acquired and incorporated into the park in 2005. The data for this study come from three

forested stands that were clear-cut shortly before the tract's acquisition. All of these stands are located on the active floodplain surface, but they encompass a gradient of conditions associated with differences in elevation, hydroperiod, geomorphic setting, and soil conditions. The Congaree River site (clearcut area = 51.7 ha; elevation, mean = 77.86 m asl; range = 76.67–79.57 m asl) is most affected by flood disturbance as it sits on an old meander scroll complex immediately below the active levee of the current river. The Bates Old River site (clearcut area = 24.0 ha; elevation, mean = 80.26 m asl; range = 78.82–82.27 m asl) is located on the natural levee of a permanently flooded, channel meander of the Congaree River that was abandoned in the 1850s. The levee is still building from overbank sediment deposition during high floods, but these sites flood less often and for shorter durations than those at the Congaree River site. The Sampson Island site (total clearcut area = 24.0 ha; elevation, mean = 79.50 m asl; range = 78.62–80.41 m asl) is located in the interior of the Bates Fork floodplain and is influenced by a sandsplay deposit resulting from the denudation of a Late-Pleistocene age aeolian sand dune.

As part of a long-term study on post-logging succession at the Bates Fork tract, vegetation data were collected from May 2006 through August 2007 using a replicated sample design with sample plots located along each of 32 transects spanning the boundary between logged and unlogged forest (Kupfer et al. 2010). Along each transect, two plot corners were randomly located: (1) 5–10 m into the clearcut (field edge) and (2) 30–50 m into the clearcut (field interior). Beginning at each corner, a 30 × 10 m sample plot was laid out, with the long axis aligned perpendicular to the transect direction. Three circular subplots (2.82 m radius) were then established at even intervals along a 30-m line running lengthwise down the middle of the sample plot. Measurements included diameter at breast height (dbh; 1.37 m) of all individuals >1 cm in dbh (saplings) and counts by species of all individuals >0.25 m in height but <1 cm in dbh (seedlings). Individuals >10 cm dbh, of which there were few, were measured and recorded within a larger radius of the subplot center (3.99 m). Sapling and seedling values were integrated into a single measure of abundance for the entire plot by: (1) transforming sapling diameters to basal areas and summing the values by species for each plot, and (2) assigning each seedling a nominal basal area of 0.1 cm^2 and adding these values to the sapling basal area. This resulted in data on woody plant regeneration for 64 plots (two plots each at 32 transects).

Soils are predominantly classified as moderately well-drained, silty clay loams on the Bates Old River and Congaree River sites and as poorly drained loams on the Sampson Island site (USDA NRCS 2008). However, variations in microtopography and flood regime create considerable within-stand differences in soil characteristics. Samples from the top 10–15 cm of soil were taken from each of the three subplots and composited to provide data on plot-level soil conditions (Kupfer et al. 2010). All soil samples were collected over two weeks during a period of uniform weather with no substantial precipitation. They were then analyzed for pH, extractable phosphorus, potassium, calcium, magnesium, zinc, copper and sodium, organic matter, cation exchange capacity(CEC), exchangeable acidity (the amount of total CEC occupied by H$^+$ and Al^{3+}), and total percent base saturation (the percent of exchange sites occupied by base cations) at the Clemson Agricultural Laboratory.

Two measures of topography and flood regime were calculated. Elevation (m asl) was extracted for each of the sample subplots from a digital terrain model constructed using high-resolution Lidar data with 3-m horizontal accuracy and 10-cm vertical accuracy. However, because elevation is an imperfect indicator of site hydrology, flood regime was examined using an interactive GIS model of floodplain inundation by coupling the U.S. Army Corps of Engineers Hydrologic Engineering Center-River Analysis System (HEC-RAS v.3.1.1: USACE 2002) with a Geographic Information System (ArcGIS 9.2), using a specialized set of ArcGIS extension tools (HEC-GeoRAS v.4.1.9.2: USACE 2005). Details of this modeling process can be found in Kupfer et al. (2010), but the end result is a seamless map of projected flood inundation depths that are specific to a user-defined river discharge. Data used in this study were the extracted inundation depths for each subplot during a 98000 cfs flood, a discharge that represents a flood which inundates all of the Bates Fork floodplain. These
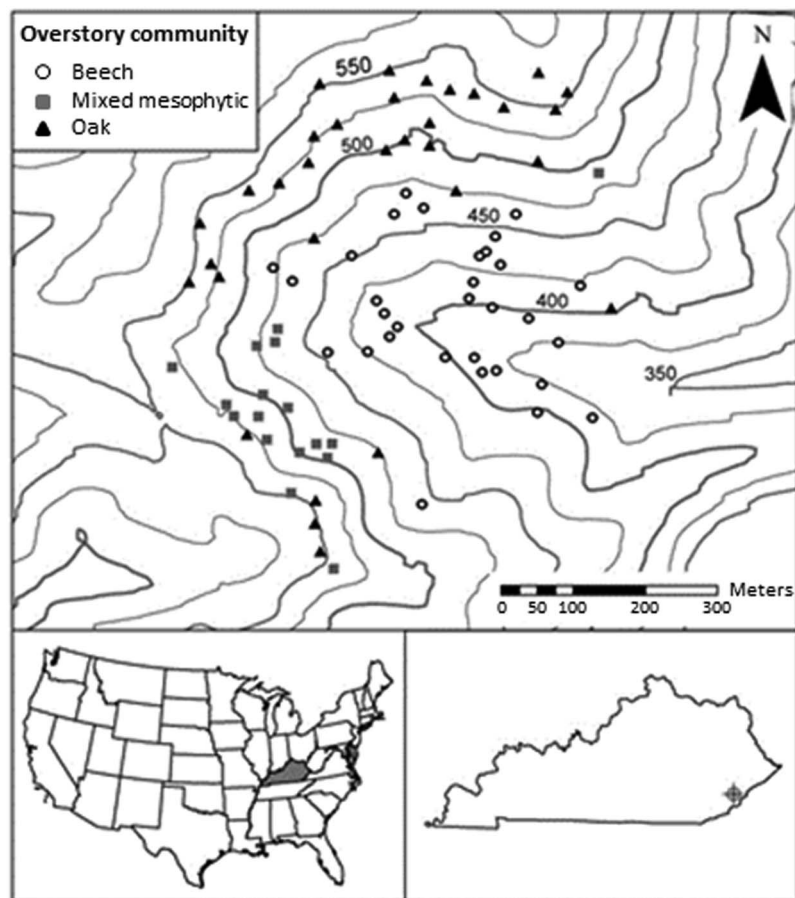
Fig. A5. Location of the McEwan study area in southeastern Kentucky, USA.

values were then averaged to get plot-level inundation values that were consistent with the vegetation and soils data.

McEwan's study area and data collection

The McEwan data set was generated in a study of a Big Everidge Hollow (BEH), a 52-ha watershed within the Lilley Cornett Woods Appalachian Research Station (37°05″ N, 83°00″ W, Roxana Quadrangle; Fig. A4) on the Cumberland Plateau, in southeastern Kentucky (McEwan and Muller 2011). Muller (1982) used a random sampling scheme stratified by elevation to establish 80 permanent circular plots, 0.04 ha in size on the horizontal plane throughout the BEH watershed, which is U-shaped and opens to the east. Woody species inventories of stems ≥2.5cm diameter at breast height were conducted in 1979, 1989, 1999, and 2010 (Muller 1982, McEwan

and Muller 2006). McEwan et al. (2000) established four 1-m^2 subplots within each of these permanent plots ($N = 320$) to assess ground-layer vegetation. Subplot centers were established at 90° intervals, around the overstory center stake at a distance that was equal to the radius of a circle bisecting the area of the overstory plot. The first subplot was located 45° to the right of uphill, and the rest were equally spaced around the plot center. The combined area of the four herb-layer subplots measured 1/100th of the overstory plot area (0.0004 ha/plot or 1 m^2 per subplot). For the purposes of this study, only herbaceous species (i.e., cryptophytes; Raunkiær 1934) were considered. Botanical nomenclature follows Jones (2005) and voucher specimens were deposited into the University of Kentucky Herbarium.

Muller (1982) measured elevation and slope aspect on each of the plots. Elevation has since

(a) South Core Banks, North Carolina
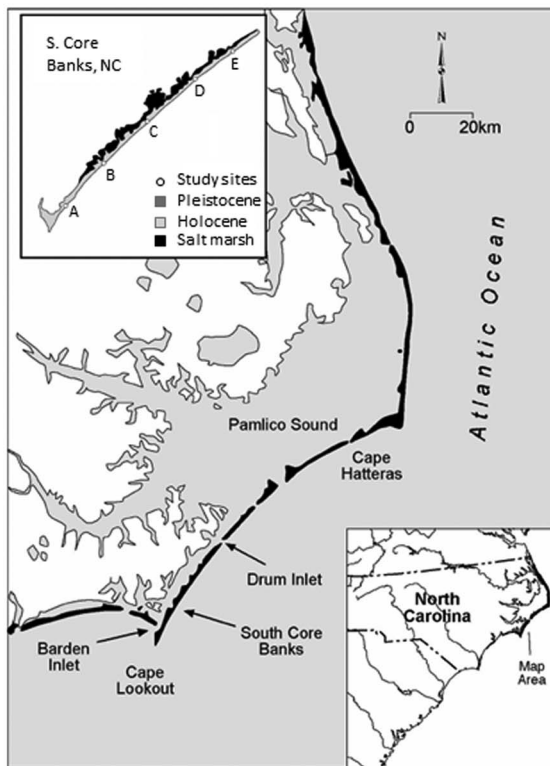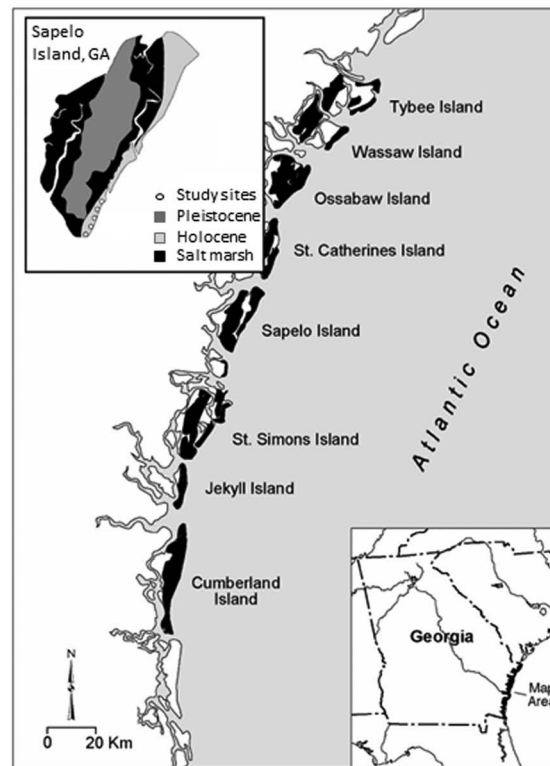
(b) Sapelo Island, Georgia

Fig. A6. Location of the Stallins study areas along the Atlantic coast, USA.

been verified using GPS technology. For the purposes of analysis, slope aspect data were converted using the Beers transformation (Beers et al. 1966), which rescaled the original 360° observations to a scale representing dryness where a value of 2 is for northeast facing slopes that are assumed to receive the least sunlight (are more mesic) and a value of 0 to southwest facing slopes (assumed to be the most xeric plots). All other aspects were linearly interpolated between these two (Beers et al. 1966). We also assessed soil chemistry through composite soil samples consisting of 16 pooled samples per overstory plot (80 total samples; collected in June of 2001). Samples were collected in the mineral soil with a 2.5-cm tube sampler to a depth of 2.5 cm, starting just below the A_o layer. Samples were homogenized and then a subsample was analyzed for plant available nitrogen (NH_4^+, NO_3^-) on a Bran-Luebbe autoanalyzer following KCl extraction. Mehlich III-extractable P, K, Ca, Mg, and Mn were analyzed on an inductively coupled plasma

spectrometer. Cation exchange capacity and base saturation were determined via ammonium saturation followed by analysis with an ammonium ion-selective electrode. To evaluate the understory light environment, hemispherical digital canopy images were collected at each permanent plot center. Digital images were collected with a Nikon Cool-Pix 950 digital camera equipped with a Nikon FC-E8 183° fisheye converter, with the lens positioned 1 meter above the ground. We used percent canopy openness in the Gap Light Analyzer image processing software (Frazer et al. 1999), as the measure of understory light conditions because it incorporates influences on light availability from both canopy vegetation and topography.

Stallins's study area and data collection

The Georgia Bight, South Core Banks, North Carolina (34°41 N, 76°28 W), and Sapelo Island, Georgia (31°23 N, 81°15 W) differ strongly in

their wave and tidal regime (Fig. A5). They also exhibit different island and dunefield morphologies, which necessitated an adapted sampling design. A 2-m point-frame sampler, aligned perpendicular to the centerline of a transect, was used to measure species presence. Point-frame observations were spaced at 1-m intervals along transects on Sapelo. On South Core, point frame observations were made every 2 m due to greater width of the dune zone and an overall longer length for transects. Each point-frame sample along transects consisted of 20 observations of species presence, with 10-cm intervals between observations. Presence was summed for individual species and expressed as percent absolute cover for each point-frame sample. A total of 1082 point-frame samples were made on Sapelo, and 1139 on South Core.

For each point-frame sample along a transect, a TOPCON total station was used to determine the elevation and distance from the seaward base of each foredune. Transects began on the seaward base of the foredune, a position approximating the mean high water mark, and extended inland until the first occurrence of extensive woody shrub cover. Soils were characterized for pH, particle-size distributions, and percent content by weight for organic matter and carbonate. Soil pH was measured with a handheld, electronic pH-meter. Wet and dry sieving was used to determine the percent contribution of particle sizes in four classes: silt and clay (<0.0625 mm), very fine to fine sand (<0.25 mm and >0.0625 mm), very coarse to medium sand (<2 mm and >0.25 mm), and granules plus coarser-grained inorganic material (>2 mm). Soil organic matter, as a proxy for soil nutrient status, was measured using loss on ignition. A HCl digest was employed to determine percent content by weight of calcium carbonate.

Supplement

R scripts for estimating the extra sums of squares in the main text (*Ecological Archives*, http://dx.doi.org/10.1890/ES15-00074.1.sm).