**Response to Previous Reviews**

The previous full proposal panel was enthusiastic that the research would advance the field of community ecology, but concerned about methodological details and whether the scope justified the budget. We addressed these issues by being more explicit about how we will evaluate environment-richness relationships, elevating the examination of spatial scale to a full Aim, eliminating analyses of phylogenetic community assembly, and by reducing the budget to that of a Small Grant. The response of the pre-proposal panel to these changes was extremely positive (all Excellents). The only challenge highlighted by the panel was compiling sufficient data to generalize our analyses across taxonomic groups. We have provided a detailed list of the thousands of communities we will use to accomplish this generalization in Table 1 and the Data Sources section of Aim 2.

**I. Conceptual Framework and Specific Aims**

Determining the processes governing community assembly and diversity is crucial to understanding and managing ecological systems. However, most studies investigating patterns of species richness fail to recognize a critical insight made 30 years ago by Shmida and Wilson (1985): species at a site typically fall into two distinct groups. C*ore species* maintain self-sustaining populations at a site, while *transient species* are poorly suited to the site but are present due to immigration from neighboring source areas. Because the dynamics and diversity of these two groups are expected to be driven by different processes, accurately inferring the importance of those processes requires the explicit consideration of each group separately (Coyle *et al.* 2013). Furthermore, since the ratio of transient to core species within a community may vary geographically and across taxonomic groups, this distinction may provide the key to reconciling conflicting evidence from previous studies regarding the relative importance of local versus regional, and niche versus neutral processes. Comparative, data-intensive studies of the impact of the core-transient distinction are critical for advancing our understanding of diversity and community assembly.

Here, we develop a unique compendium of community time series datasets with which we will distinguish core and transient species and analyze patterns of diversity. Our primary aims are:

**Aim 1**. **Evaluate the generality of the distinction between core and transient species across taxa and ecosystems.** We will use data from a wide range of taxonomic groups and ecosystems to determine how core-transient patterns vary across taxa, dispersal mode, and landscape context.

**Aim 2**. **Advance models of species richness by treating core and transient species separately.** We will develop models of species richness that allow local and regional variables to exert differential effects on core and transient species. We will compare the resulting parameter values and predictive abilities of these models to traditional approaches that ignore differences between the two groups.

**Aim 3. Determine the spatial-scaling of core and transient designations.** Understanding differences in the proportions of core and transient species between taxonomic groups and ecosystems will hinge on understanding how the prevalence of these groups varies with spatial scale.

By addressing the important differences in the processes driving core and transient species patterns, this research will produce a better understanding of the linkages between local and regional scale processes in driving patterns of species richness, and how the relative importance of those processes varies across taxonomic groups, ecosystems, and spatial scales.

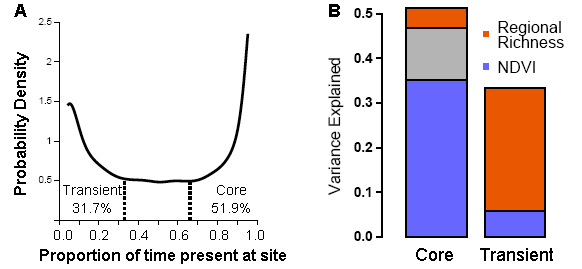
**II. Background and Significance**

For decades, the most important determinants of species richness were thought to be purely local processes such as competition, predation, and disturbance (Paine 1966; MacArthur 1972; Connell 1978). Even in recent years, many studies have been conducted seeking to characterize the extent to which local biotic or abiotic conditions dictate an equilibrial level or limit to the number of species that might coexist (Srivastava & Lawton 1998; Brown *et al.* 2001; Hurlbert & Haskell 2003; Sanders *et al.* 2007). However, an alternative perspective emerging over the past twenty-five years highlights the importance of regional processes where the richness of local communities primarily reflects the availability of colonists from the regional species pool and the regional-scale variables that affect speciation, colonization, and extinction dynamics (Ricklefs 1987, 2007; Karlson *et al.* 2004). While it is increasingly well accepted that processes operating at both local and regional scales contribute to local richness patterns (Freestone & Harrison 2006; Harrison & Cornell 2008; Hortal *et al.* 2008; White & Hurlbert 2010), exactly how these two scales of processes combine to structure communities and determine species number remains poorly understood. We highlight an important distinction between species that yields novel insights into how local and regional processes combine to structure ecological communities.

*Core versus transient species*

Species within a community can be divided into core and transient species based on their temporal persistence and ability to maintain viable populations. These two groups are expected to differ substantially in the processes influencing their presence and abundance in ecological systems (Shmida & Wilson 1985; Grime 1998; Novotný & Basset 2000; Magurran & Henderson 2003). In order to maintain a viable population, core species must be able to successfully compete in the local abiotic and biotic environment, and are therefore expected to be strongly influenced by deterministic processes such as environmental filtering and competition. On the other hand, transient species, which do not maintain viable populations, are not expected to exhibit a strong ecological match to their environment. As such, their presence should be primarily influenced by regional processes governing the prevalence of species available to disperse to the site from a broader species pool.

Since the processes influencing core and transient species are expected to differ so strongly, we hypothesize that when these groups can be distinguished, modeling them separately will improve both our understanding of ecological processes and the predictive power of related ecological models (Magurran 2007). This hypothesis is supported by empirical research showing that core and transient species exhibit distinct forms of the species abundance distribution, and modeling the two groups separately leads to better fits than a single fit to the community as a whole (Magurran & Henderson 2003; Ulrich & Ollik 2004). In addition, regional diversity has a stronger effect on transient species richness than on core species richness (Belmaker 2009). Recently, we conducted the first study of the core-transient influence on richness-environment relationships (Coyle *et al.* 2013). We found that across 467 North American bird communities, most species could be reliably assigned to core and transient categories based on how frequently they occur in the community (Fig. 1A). More importantly, we found that core species richness and transient species richness were best predicted by completely different variables: the number of core species was most strongly related to a local measure of primary productivity in the breeding season, while the number of transient species was most strongly related to regional measures of habitat heterogeneity and the richness of the regional species pool (Fig. 1B). These contrasting patterns are consistent with the predicted differences in the way local and regional processes should affect species richness. If these results apply broadly across ecological systems, then distinguishing between core and transient species has the potential to: 1) change the way we model diversity patterns; and 2) reconcile contradictory views regarding the relative importance of local versus regional influences (Ricklefs 1987; Harrison & Cornell 2008) and niche versus neutral processes (Holyoak & Loreau 2006; Vergnon *et al.* 2009).



**Figure 1.** (A) The distribution of temporal occupancy of 467 North American bird communities exhibits 2 distinct groups of species: core species that occurred in more than 2/3 of samples thru time, and transients that occurred in less than 1/3. (B) Fractions of variance in core and transient richness explained by a local vegetation index (NDVI, blue), or by the richness of the regional pool (*from Coyle et al. 2013*).

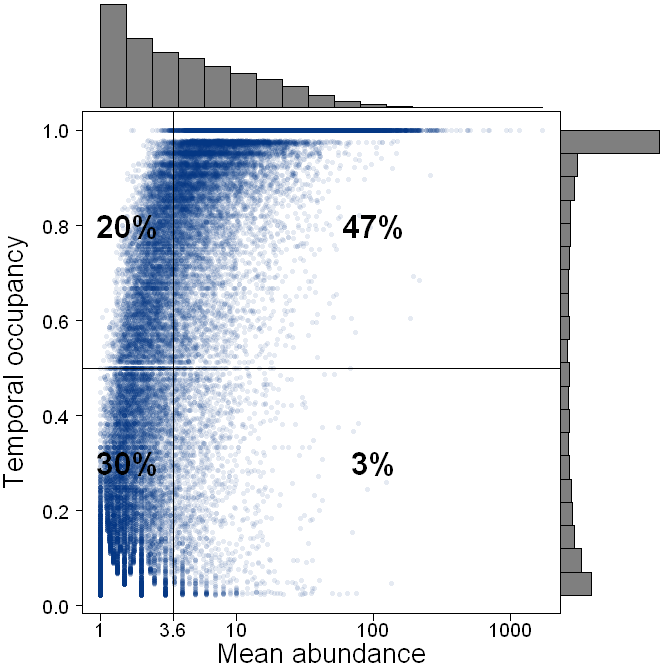
*Identifying core and transient species*

The most common approach for identifying core and transient species uses information on temporal occupancy. The frequency with which a species is observed at a sampling location through time is quantified and then species are divided into groups with high values of occupancy (core species) and low values of occupancy (transient species) (Costello & Myers 1996; Magurran & Henderson 2003; Ulrich & Ollik 2004; Vergnon *et al.* 2009). The distribution of occupancies for North American birds is clearly bimodal (Fig. 1) leading to a natural division of species into two groups. The simplicity of this empirical approach makes the distinction between core and transient species possible for any dataset with a sufficiently long community time-series. This general approach has been validated using independently derived habitat association data in other systems (Magurran & Henderson 2003) and analyses of richness using non-occupancy based designations yield similar results to those using occupancy (Belmaker 2009). We have further validated this approach by comparing the statuses of bird species based on occupancy distributions to independently collected breeding status data from a statewide breeding bird atlas project. Comparing 30 Breeding Bird Survey routes (BBS; see Aim 2) from New York state to high quality blocks in the New York Breeding Bird Atlas (Andrle & Carroll 1988) shows that on average 85% of both core and transient species identified using the BBS occupancy distributions were similarly classified by the atlas data (i.e., as probable or confirmed breeders as opposed to transients).

*Related ideas*

Core-satellite. Temporal site occupancy shares conceptual overlap with spatial occupancy across a region, an idea which has been investigated extensively in the literature (Raunkiaer 1934; Hanski 1982; Collins & Glenn 1990; McGeoch & Gaston 2002). Frequently, the spatial occupancy distri-bution is observed to be bimodal as well, prompting the development of the well-known core-satellite hypothesis (Hanski 1982) and other explanations (see review in McGeoch & Gaston 2002). How-ever, spatial and temporal occupancy differ in important respects making these very different areas of research. Under the spatial framework a species is designated as core or satellite over an entire region or continent and this designation is a characteristic of the species. In contrast, the designation of a species as core or transient occurs at the local scale, may vary from site to site, and is a characteristic of a species at a particular location. Narrowly distributed satellite species can still be ‘core’ community members at the sites where they occur, and widely distributed species may be transient visitors to some of the sites at which they are observed. As such, the core-transient distinction is more directly tied to the dynamics of a local population and its ability to persist in a particular environment.

Abundant-rare. A second distinction related to the core-transient dichotomy is that between abundance and rarity. Temporal occupancy is correlated positively with mean abundance in the North American Breeding Bird Survey (BBS), but imperfectly so (Fig. 2). Preliminary analysis shows that 40% of the variance in occupancy in our data is not associated with abundance (Fig. 2). While many core species are abundant and many transient species are rare, a substantial proportion of species are numerically rare but persistently occurring species, and a noticeable fraction of points also indicate transient species that occur infrequently but are abundant when they do occur (Fig. 2). The most obvious difference between the abundant-rare and core-transient distinctions is that abundance distributions tend to be unimodal (Fig. 2, top histogram), and thus any categorical distinction between abundant and rare relies on choosing a somewhat arbitrary abundance threshold. In contrast, occupancy distributions are often strongly bimodal (Fig. 1; Fig 2, right histogram) lending greater confidence in the existence of two biologically distinct groups.



**Figure 2**. Relationship between average abundance of a species when it is observed and the proportion of years in which the species was observed for 345 bird species across 467 Breeding Bird Surveys in North America. Horizontal line at occupancy = 0.5 partitions data into core and transient, while vertical line at the median level of abundance across the entire dataset (abundance = 3.6) partitions data into abundant and rare. Marginal histograms depict the density of points along each axis.

**III. Research Aims and Approach**

**Aim 1**. **Evaluating the generality of the distinction between core and transient species across taxa and ecosystems**

Preliminary data suggests that the bimodal form of the temporal occupancy distribution is consistent across a variety of systems (Fig. 3), and that core and transient species are generally identifiable as distinct groups. However, the exact shape of this distribution and the relative representation of core and transient species varies. We will compile an extensive database of temporal occupancy distributions across a wide range of taxonomic groups and ecosystems to assess general patterns in the shape of the occupancy distribution and refine its use for designating core and transient species.

*1a. Does the shape of the occupancy distribution vary predictably with taxonomic group, dispersal ability, or landscape context?*

We have already identified approximately 60 datasets with community time series spanning over 5,000 communities (Table 1). These datasets are either publicly available or the PIs have established



**Figure 3.** Distributions of temporal occupancy for species in 6 taxonomically distinct communities. Data sources: fish, Magurran & Henderson 2003; butterflies, Dapporto 2009; plants, Guo et al. 2000; plankton, Dolan et al. 2009; rodents, Ernest et al. 2009; amphipods, Costello & Myers 1996. Two metrics of distribution shape are included: a measure of bimodality, *b*, and the distribution mean, μ.

relationships with the data holders and a history of access to the data via memoranda of understand-ing. We will add to this set through targeted literature searches. We will use all communities which have been sampled in at least 10 time intervals (usually years, but potentially months or weeks depending on the taxon). These data, or pointers to them, will be made available through the Knowledge Network for Biocomplexity and the Ecological Data Wiki (see Data Management Plan).

We will measure two critical aspects of the shape of temporal occupancy distributions with the goal of gaining insight into the generality of the core-transient distinction. The first is a measure of bimodality, *b*, using the variance in occupancy scaled by the maximum possible variance. When all species are evenly split between the highest and lowest occupancy bins, the scaled variance will be 1, while if all species have identical occupancy, the value will be 0. Most of the distributions in Figure 3 are strongly bimodal (*b* > 0.5). The significance of this bimodality will be assessed using methods by Silverman (1981) and Tokeshi (1992). A second measure is whether the weight of the distribution falls towards the core or the transient end, which is reflected in the overall mean occupancy. The fish and rodent datasets in Figure 3 provide examples of two distributions with a similar degree of bimodality but with a notable difference in the side to which occupancy values are skewed.

Our aim here is necessarily descriptive, as we will be the first to explore generalities in the shape of temporal occupancy distributions. We will employ a mixed modeling approach (as in Soininen *et al.* 2007) that examines how the above shape parameters vary as a function of taxonomic group, geographic location, dispersal mode, and environmental variables. We expect the proportion of core species in a community to be related to species traits and environments that reduce the importance of dispersal. Specifically the proportion of core species should be higher for: 1) taxonomic groups that are active rather than passive dispersers; 2) communities in homogeneous rather than heterogeneous environments; 3) for taxonomic groups with lower reproductive output; and 4) for groups in which local community richness is a greater proportion of the regional species pool.

**Table 1.** Available datasets for investigating temporal occupancy and the core-transient dichotomy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Taxonomic Group** | **Number of communities** | **Maximum**  **time span** | **Aims** | **Data sources** |
| Terrestrial birds | ~1,700 | 47 years | 1, 2, 3 | Beven 1976, Diamond & May 1977,  Kendeigh 1982, Williamson 1983,  Hall 1984, Svensson et al. 1984, Vickery & Nudds 1984, Knapp et al. 1998, Holmes et al. 2009, Sauer et al. 2011, National Audubon Society 2012 |
| Plants | ~2,200 | 41 years | 1, 2, 3 | See Table 2 |
| Rocky Intertidal | ~1,000 | 25 years | 1, 2, 3 | Petraitis & Vidargas 2006, Petraitis et al. 2008, Raimondi et al. 2009 |
| Fish | ~120 | 26 years | 1, 2, 3 | Grossman 1982, Grossman et al. 1982, Magnuson & Bowser 1990, Henderson & Holmes 1994, Pigg et al. 1998, Sosebee & Cadrin 2006, Sweatman et al. 2008 |
| Terrestrial invertebrates | ~130 | 25 years | 1, 2, 3 | Novak 1983, Taylor et al. 1990, Pollard et al. 1986, Pollard 1991, Bloch et al. 2007, Ernest et al. 2009, NABA 2009 |
| Small mammals | ~30 | 32 years | 1, 3 | Ernest et al. 2000, Kaufman et al. 2000, Ernest et al. 2009, Merritt 2009, Stapp 2009, Thibault et al. 2011, Kelt et al. 2013 |
| Plankton | ~70 | 60 years | 1, 2, 3 | Magnuson & Bowser 1990, Hampton et al. 2008, Yan et al. 2008 |

*1b. Refining core-transient species designations.*

Currently we define transient species operationally as those occurring in the bottom third of the occupancy spectrum (≤0.33), and core species as those occurring in the top third (≥0.67). The small fraction of species in the middle of the occupancy distribution for which there is equivocal evidence and for which classification errors may be greatest are excluded from analyses. This general approach to distinguishing between core and transient species performs well (see Background & Significance) and yields richness model results that are robust to specific occupancy thresholds or exclusion of intermediate species (Coyle et al. 2013). However, no effort has been made to further improve this approach since it was first described over a decade ago. We will further refine the methods for determining the richness of core and occasional species using simulation modeling.

The current approach ignores the fact that some species which maintain viable local populations may be designated as transient because they are rare or otherwise difficult to detect (Coyle et al. 2013). For this reason, estimates of transient richness are expected to be biased high while estimates of core richness are expected to be biased low. We will quantify the degree of misclassification using simulation models and use these results to adjust the estimates of core and transient species at a site. A simulation model for this problem requires 1) distinct suites of species with particular habitat affiliations, 2) long-term persistence at a site dependent on the match between a species and its habitat, and 3) the dispersal of propagules across the landscape. We will use the individual-based model of Smith & Lundholm (2010; available as part of the “neutral.vp” R package) in which individuals reproduce, disperse, and die within a heterogeneous landscape. Individual grid cells will be coded as being one of two habitat types, and species will be assigned "trait values" *a priori* such that they are more suited to (i.e., core species in) one habitat type over the other. Birth and death rates are functions of the match of an individual to its environment, and dispersal operates as either a random walk (Smith & Lundholm 2010) or a random walk biased toward suitable habitat type. Each grid cell supports a fixed carrying capacity, and if after reproduction and dispersal the total number of individuals exceeds that level, excess individuals are removed at random (Smith & Lundholm 2010). The simulated landscape will be 100 x 100 cells and implicitly lies within a broader metacommunity from which individuals disperse at a fixed migration rate. Each cell will be initialized with an equal number of individuals of each species. After 10,000 time steps, we will examine the temporal occupancy of species within local community samples over a 25-time step period. We will use standard parameter values from Smith and Lundholm (2010), which produce suites of realistic ecological patterns.

We will simulate 10,000 sets of communities and compare core-transient assignments based on the occupancy distribution to their known state. This comparison will yield quantitative estimates of the magnitudes by which transient species richness is overestimated and core richness is underestimated using traditional approaches, and will indicate the range of potential bias over which to conduct sensitivity analyses in Aim 2.

**Aim 2**. **Advance models of species richness by treating core and transient species separately**

The status of a species as core or transient is a reflection of the extent to which that species is adapted to the local environment, and also an indication of whether that species is a predictable member of the community and subject to local stressors and resource constraints (Shmida & Wilson 1985; Magurran & Henderson 2003; Ulrich & Ollik 2004). This biological distinction between core and transient species yields five specific predictions about drivers of species richness of the two groups. Our initial work on birds was supportive of the first four predictions (Fig. 1b, Coyle et al. 2013).

**H1a)** The number of core species at a site will be best predicted by local scale environmental variables such as temperature, productivity, or soil nutrients.

**H1b)** The number of core species will *not* be directly influenced by regional variables reflecting broad-scale habitat heterogeneity and the diversity of the species pool.

**H2a)** The number of transient species at a site will be best predicted by regional variables reflecting habitat heterogeneity and the diversity of the species pool.

**H2b)** The number of transient species will *not* be directly influenced by local environmental variables.

**H3)** Modeling species richness as the sum of two sets of processes operating distinctly on core versus transient species will yield greater predictive power than approaches that ignore this distinction.

***Data sources***

In order to assess patterns of species richness, we focus on the subset of studies listed in Table 1 that include data for at least 30 locations sampled in an equivalent manner across an extent of at least 1,000 km. For each dataset, we will use information from previous studies to select variables that best characterize the local environment and regional heterogeneity for that taxon from a suite of remotely sensed environmental data (Table 2), and where appropriate, site-specific meteorological stations. We will also use taxon specific sources for data on the richness of the regional species pool. A brief summary of four terrestrial and four marine systems that will be investigated is presented below.

**Table 2.** Remotely sensed environmental data for Aim 2. Data will be averaged over the relevant temporal (e.g., seasonal or annual) and spatial scales to match the scale of community data.

|  |  |  |
| --- | --- | --- |
| Variable | Resolution | Source |
| *Terrestrial* |  |  |
| Temperature | 1 km | WorldClim (Hijmans et al. 2005) |
| Annual precipitation | 1 km | WorldClim (Hijmans et al. 2005) |
| Normalized difference vegetation index | 250 m | MODIS/Terra, MOD13 |
| Actual evapotranspiration | 1 km | MODIS/Terra, MOD16 |
| Elevation | 1 km | USGS GTOPO30 |
| Land cover | 30 m | National Land Cover Database (Fry et al. 2011) |
| Soil type | 10 m | Gridded Soil Survey Geographic Database |
|  |  |  |
| *Marine* |  |  |
| Sea surface temperature | 1 km | MODIS/Aqua, MOD28 |
| Chlorophyll-*a* | 1 km | MODIS/Aqua |
| Organic pollutants | 1 km | Halpern et al. (2008) |
| Inorganic pollutants | 1 km | Halpern et al. (2008) |
| Reef area | 1 km | Global shallow bathymetry from SeaWiFS |
| Ocean currents | 1/3 degree | NOAA OSCAR, Bonjean & Lagerloef (2002) |

Birds. The North American Breeding Bird Survey (BBS; Bystrak 1981) and Audubon Christmas Bird Counts (CBC; Bock & Root 1981) are long-term, large-scale monitoring programs that provide an unparalleled resource for examining spatial and temporal variation in avian populations and communities during the breeding and wintering seasons, respectively. 467 breeding surveys (25 km2 roadside routes) and ~1,200 Christmas counts (452 km2 circles) meeting *a priori* quality control criteria have been surveyed continuously between 1996 and 2010 (Fig. 4a). *Local environmental variables* include long-term means for seasonal and annual temperature, precipitation, and vegetation indices as described in Coyle et al. (2013). *Regional heterogeneity* is evaluated by calculating the spatial variance of these local variables over a 100 km radius circle centered on the community of interest. This scale is based on the typical dispersal distances of 100 bird species from two continents (Paradis *et al.* 1998; Tittler *et al.* 2009). The *regional species pool* of a site is determined independently of the survey data using the number of species with range maps that overlap this region (data from NatureServe, Ridgely *et al.* 2007).

Butterflies. The North American Butterfly Association's Fourth of July Butterfly Count (North American Butterfly Association 2009) is a continent-wide monitoring program with ~100 count circles (452 km2 each) with at least 10 years of continuous data meeting minimum survey effort criteria (Fig. 4a). *Local environmental variables* expected to be most important for explaining butterfly richness include growing season temperature (White & Kerr 2006) and actual evapotranspiration (Hawkins & Porter 2003). *Regional heterogeneity* measures include measures of land cover diversity (Kerr *et al.* 2001) and spatial variance of local variables over the 12-km radius count area which spans typical dispersal distances (Stevens *et al.* 2010). The *regional species pool* for each site is derived from county level range maps (Opler *et al.* 2013).

Plants. We will compile an extensive set of grassland and shrubland plant community quadrat time series (Fig. 4a, Table 3). When community data are collected as part of an experiment, we will use unmanipulated control plots only. *Local environmental variables* include seasonal and annual measures of precipitation and temperature (available from site-specific weather stations or from sources in Table 2), and for many of the datasets, soil variables such as pH, conductivity, % organic

****

**Figure 4**. Locations of long-term community monitoring datasets from (A) terrestrial and (B) marine environments to be used in this study.

matter, and mineral content. *Regional heterogeneity* measures include land cover diversity, topographic diversity, and soil diversity from the Natural Resources Conservation Service soils database (Soil Survey Staff 2013), and spatial variance in local environmental variables within a 5 km radius of each focal site. Five km is greater than the dispersal distance of >95% of terrestrial plants surveyed in a recent meta-analysis (Kinlan & Gaines 2003). Finally, the *regional species pool* of a site will be compiled from county-level lists from the USDA Plants Database (USDA, NRCS 2012).

Fish. Fish communities will be examined using two datasets at different spatial scales. The Australian Institute of Marine Science has conducted 20 years of visual fish censuses on 46 reefs spanning ~1,200 km of the Great Barrier Reef (Sweatman *et al.* 2008; Fig. 4b). Each reef is characterized by fifteen 50 m transects at 6-9 m depth. In addition, the Northeast Fisheries Science Center (NEFSC) Bottom Trawl Survey has monitored the species composition and abundance of fish in the Northwest Atlantic from Cape Hatteras to the Gulf of Maine at hundreds of sampling stations since 1948 (NEFSC 1988; Fig. 4b). Data are aggregated into regional strata (~103 km2) as defined by the NEFSC (Sosebee & Cadrin 2006), with a focus on annual autumnal surveys in offshore areas. *Local*

**Table 3**. Preliminary list of plant community time series datasets.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Site** | **Citation** | **No. plots** | **Plot size** | **Years of data** | **Soil data?** |
| Arctic LTER, AK | Shaver and Chapin 1991 | 20 | 0.04 m2 | 20 | Y |
| Carpinteria salt marsh, CA | Cleland et al. 2008 | 25 | 0.25 m2 | 7 | N |
| Cedar Creek LTER, MN | Tilman 1987 | 36 | 16 m2 | 30 | Y |
| Central Plains Exptl Range, CO | Chu et al. 2013 | 24 | 1 m2 | 14 | N |
| Channel Islands, CA | Johnson & Rodriguez 2001 | 163 | 30 m transect | 17 | N |
| Fort Hays, KS | Adler et al. 2007 | 51 | 1 m2 | 41 | N |
| Jasper Ridge, CA | Zavaleta et al. 2003 | 96 | 0.8 m2 | 8 | Y |
| Jornada LTER, NM | Nelson 1934, Huenneke et al. 2001 | 751 | 1 m2 | 32 | N |
| Kellogg Biological Station, MI | Huberty et al. 1998 | 12 | 1 m2 | 23 | Y |
| Konza Prairie LTER, KS | Gibson & Hulbert 1987 | 120 | 10 m2 | 25 | Y |
| Lac Croche, Quebec | Paquette et al. 2007 | 43 | 400 m2 | 9 | Y |
| McLaughlin Reserve, CA | Elmendorf & Harrison 2011 | 355 | 1 m2 | 10 | Y |
| Niwot Ridge LTER, CO | Theodose & Bowman 1997 | 40 | 4 m2 | 11 | Y |
| Portal, AZ | Ernest et al. 2009 | 384 | 0.25 m2 | 30 | N |
| Sagebrush Steppe, ID | Zachmann et al. 2010 | 26 | 1 m2 | 35 | N |
| Sevilleta LTER, NM | Baez et al. 2006 | 36 | 1 m2 | 10 | Y |
| Sonoran Desert Lab, AZ | Rodriguez-Buritica et al. 2013 | 30 | 100 m2 | 20 | N |
| Tall Grass Prairie Preserve, OK | McGlinn et al. 2010 | 80 | 1 m2 | 12 | Y |

*environmental variables* for both datasets include annual mean estimates of sea surface temperature, salinity, nitrate and ocean productivity (estimated from chlorophyll-*a*) which have all been found to be important drivers of fish richness (Mellin *et al.* 2010). In addition, for the coral reef dataset, percent coral cover recorded along the fish transects will be used as a local variable. *Regional heterogeneity* measures will be assessed over a circle of radius 30 km in the Great Barrier Reef, and over a circle of radius 150 km in the North Atlantic. These differing regional scales account for extended larval duration at colder temperatures, and hence greater distances that larvae may be passively dispersed (O’Connor *et al.* 2007), and are also in accordance with published estimates of larval dispersal in the two regions (Kinlan & Gaines 2003; Planes *et al.* 2009). Heterogeneity measures include the spatial variance of local variables over these regions, as well as total regional reef area for the Australian sites. The *regional species pool* will be estimated for each location by overlaying distribution maps of the Australian ichthyofauna provided by FishMap (<http://fish.ala.org.au/>), and by expected species richness based on latitudinal extents for fish in the Northwest Atlantic from FishBase (Froese & Pauly 2012).

Plankton. The Sir Alistair Hardy Foundation for Ocean Science manages the Continuous Plankton Recorder which has collected data on both phyto- and zooplankton from sample tows at ~10 m depth from throughout the North Atlantic for over 60 years (Richardson *et al.* 2006; Fig. 4b). We will characterize plankton assemblages in the sixty-six 0.5 degree latitude-longitude cells that have at least 10 consecutive years of data with at least 10 sample tows in each year. Each sample tow spans 10 nautical miles. *Local environmental variables* include annual mean estimates of sea surface temperature, chlorophyll-*a*, salinity, and measure of organic and inorganic pollutants from the atlas of human impacts on marine systems (Halpern *et al.* 2008). *Regional heterogeneity* measures include spatial variance in local variables and variance in the source of incoming current directions within a 2-degree block centered on the focal cell. The *regional species pool* for each cell will be defined as the set of species encountered within the regional 2-degree block over the previous 25 years.

Rocky Intertidal. We will use data on the abundance of marine invertebrates from a set of over 1,000 long-term 50 x 75 cm photoplots from 70 sites along the Pacific coast of North America, monitored by the Marine Rocky Intertidal Network and its partners for up to 18 years (Raimondi *et al.* 2009; Fig. 4b). We will conduct analyses separately for invertebrates and plants, as well as an analysis that considers them together. We will use *local environmental variables* following Cruz-Motta et al. (2010), including sea surface temperature, chlorophyll-*a*, and organic and inorganic pollution. *Regional heterogeneity* measures include spatial variance in these local variables within 100 km of each site based on published estimates of dispersal distances (Kinlan & Gaines 2003). The *regional species pool* for each site will be based on the total number of taxa expected based on known latitudinal extents of each species as well as habitat preferences (Ricketts 1985).

***Statistical methods***

We will fit a series of regression models to predict core, transient, and total richness in each dataset based on local environmental variables, regional heterogeneity variables, and all variables combined. We will use random forest modeling (Olden *et al.* 2008) to identify non-linearities in the response of richness to environmental variables as well as important variable interactions. These results will inform the structure (i.e., inclusion of interaction and quadratic terms) of conditional autoregressive models (Wall 2004) that explain spatial patterns of species richness while accounting for the spatial autocorrelation inherent to these data. For datasets that are hierarchically structured in space (e.g., the plant data that has large numbers of quadrats at a single site) we will use hierarchical models to properly account for within versus among site variation (Gelman & Hill 2007). All independent variables will be normalized to *z*-scores (subtracting the mean and dividing by the standard deviation) to facilitate the comparison of effect sizes for variables with disparate units and values. We will use cross-validation to avoid overfitting in the conditional autoregressive models by evaluating model performance on data that have not been part of the model fitting process (Geisser 1993).

*Assessing the relative importance of local and regional variables (Hypotheses 1-2):* We will evaluate the importance of different processes for determining core and transient richness using regression models based on only local variables, only regional variables, and all variables combined. All three models will be fit to core and transient richness separately and compared in two ways. We will use the coefficients of determination for the local, regional, and combined models as inputs for variance partitioning analyses (Legendre & Legendre 1998) to determine the relative explanatory power of local and regional variables (White & Hurlbert 2010, Coyle et al. 2013). This provides information on the relative importance of the different sets of processes. We will also conduct a stronger test using information theoretic-based model selection, by selecting the best fitting model for each dataset using AIC (Burnham & Anderson 2002). This will determine whether variables that are predicted not to influence a particular group’s species richness have no meaningful effect or are simply less important than other variables. In cases where more than one model provides a good fit to the data we will also use model averaging to assess the relative importance of local and regional variables from an AIC perspective (Burnham & Anderson 2002). We will use the results of *Aim 1b* to re-evaluate model performance and variance partitioning using core and transient richness estimates that have been adjusted for sampling bias.

*Does the core-transient distinction improve predictive power? (Hypothesis 3):* We will test the hypothesis that modeling the two groups separately yields better predictions for overall species richness by modeling total species richness using a combination of local and regional variables and comparing the predictions of this model to those obtained by modeling core and transient species richness separately and summing the predicted richness values. Comparisons will be done using leave-one-out cross validation (LOOCV; Arlot & Celisse 2010). Cross validation is used to assess how accurately a predictive model will perform in practice, and reduces the likelihood of model overfitting leading to superior performance. In LOOCV, the value of species richness for each data point is validated against the prediction from a model based on all data excluding that data point. Models of overall species richness and core + transient species richness will be compared using the mean squared error of cross validation analyses, and analyses will be conducted with and without adjustments for sampling bias.

**Aim 3. Determine the spatial-scaling of core-transient designations.**

Patterns of diversity change with the spatial scale of analysis due to shifts in the processes that dominate at different scales (Chase & Leibold 2002; Hurlbert & Haskell 2003). The identification of core and transient species will also depend on spatial scale. At continental scales, nearly every species can be considered to have a sustainable population somewhere, and hence all will be core species. Once spatial scales fall below the average home range size of an individual, then species will be inadequately censused in any given year and will all appear to be transients. This calls for a deeper examination of how the shape of the temporal occupancy distribution varies with spatial scale, which we will pursue using both empirical and simulation-based analyses.

A number of the datasets in Table 1 were collected in a spatially hierarchical manner allowing for the natural aggregation of community data at different spatial scales. For example, plant quadrats are frequently arrayed in grids, plankton samples can be aggregated at coarser and coarser resolutions, and coral reef fish can be examined at the scale of a single transect, a group of transects at a site, and the set of sites sampled within a reef. Figure 5 illustrates how occupancy distributions of bird communities vary with spatial scale using the BBS dataset. A single BBS survey route describes the species seen over 50 point count stops, so we calculated occupancy at the individual stop scale, over 10 adjacent stops, the entire survey route, and over the 27 survey routes within the state of Maryland. At each scale, the distribution is bimodal, but the frequency of core species increases with spatial scale as expected. Within datasets we will characterize this shift quantitatively by using AIC based model selection to identify the best fitting spatial scaling function for both the mean occupancy and bimodality metrics described in Aim 1.



**Figure 5.** (a) Mean temporal occupancy distributions of bird communities in Maryland at four spatial scales from the North American Breeding bird survey data. The relative magnitude of core and transient modes shifts with scale. (b) Mean temporal occupancy as a function of community size, the total number of individuals in a sample, for the four scales of bird data in (a) and for the datasets included in Figure 3.

Understanding how occupancy is influenced by scale is particularly important because different taxonomic groups are sampled at different absolute scales, as well as at different scales relative to average body sizes and dispersal distances. We will explore potential differences among groups by comparing the form of the spatial scaling relationships among different taxonomic groups and datasets. We will use these results to help control for scale in comparisons of core and transient designations across taxa. Preliminary analyses suggest that average community size (i.e., number of individuals) serves as an indicator of the taxon relevant spatial scale. We will plot occupancy as a function of the total number of individuals encompassed by the sample (Fig. 5b). In this way, a community of 1,000 plankton individuals in a liter of seawater can potentially be compared to a community of 1,000 birds over 25 km2. We will validate this approach (and fine tune it if necessary) using simulations. We will use the simulation model described in Aim 1b to establish a null expectation for the relationship between scale and the proportions of core and transient species in a community. After simulating spatially explicit communities we will examine the temporal occupancy of species within focal regions spanning a 100-fold range of area over a 25 time-step period. We will also conduct simulations in which the spatial scale of analysis is held constant but organism density is varied. In so doing, we will be able to assess the independent contributions of spatial scale and community size in driving occupancy scaling relationships.

Understanding the impacts of spatial scale on occupancy distributions will improve our ability to evaluate differences in these distributions among ecosystems, taxonomic groups, and datasets as described in Aim 1. Characterizing the amount of between dataset variation in occupancy that results from absolute and relative spatial scale differences allows us to focus more directly on environmental and biological drivers of variation in occupancy distributions. To assess the benefits of explicitly considering spatial scale in occupancy distribution based modeling we will rerun analyses conducted in Aim 1a after accounting for spatial scale and evaluate how this influences conclusions about differences among groups and underlying processes.

**IV. Project Management and Milestones**

Hurlbert and White have a long history of successful collaboration on macroecological research involving large datasets (e.g., Hurlbert & White 2005, 2007; White & Hurlbert 2010; Coyle *et al.* 2013). As such, the PIs will be able to accomplish this substantial ecoinformatics undertaking quickly and productively within the proposed two-year award period. As the lead PI, Hurlbert will oversee the training and supervision of the graduate student, data collection and analysis, and writing of manuscripts and annual reports. Co-PI White will oversee database development, ecoinformatics, and statistical analysis. White will also run a scientific programming workshop each year. The PIs and project graduate student will hold monthly skype conferences to discuss project status and to troubleshoot any obstacles. In addition, White will travel to UNC once per year for more in depth collaboration on analysis and manuscript writing. The graduate RA will be in charge of compiling the data in Tables 1 and 2, searching for additional community datasets, and ensuring all datasets are registered and accessible, where permitted, in standard repositories and data registries. In the final year of the project, the RA will develop and plan the Pre-College Saturday Academy (see Broader Impacts), and ensure that the developed curriculum is disseminated through the Environmental Educators of North Carolina (www.eenc.org) and the National Environmental Education Foundation (www.neefusa.org). A summary of the expected project timeline is below.

**Table 4.** Project timeline.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Project task** | **F 2014** | **S 2015** | **F 2015** | **S 2016** |
| Cross-taxon database compilation | x | x |  |  |
| Aim 1 - generality of core/transient | x | x |  |  |
| Aim 2 - species richness modeling |  | x | x |  |
| Aim 3 - effect of spatial scale |  |  | x | x |
| Software Carpentry workshops |  | x |  | x |
| Pre-College Saturday Academy |  |  |  | x |

**V. Broader Impacts**

*Previous activities*

PI Hurlbert has successfully mentored four undergraduates (3 of them females from underrepresented groups) in independent research, one of whom recently published results from her project in *PLoS One.* In addition to his graduate mentoring (3 students, 2 female), he co-leads a graduate seminar on using R for data manipulation and analysis in ecology each year. Hurlbert has also demonstrated a commitment to public outreach through Science Day events at local middle schools, and public talks at venues such as the North Carolina Museum of Natural Sciences, local Audubon Society chapters, and the Chapel Hill Bird Club. Co-PI White has been actively involved in providing computational training to biologists. He teaches courses in computational skills to both undergraduate and graduate students at Utah State University and is both an instructor and member of the steering committee for the Software Carpentry project (<http://software-carpentry.org>). As such the researchers involved in this proposal are well suited to training students how to work with large datasets effectively.

*Planned activities*

Recruitment and training. The work will involve the training of one graduate student in biodiversity science and the use of big data in ecological research. We are committed to recruiting and training young scientists from underrepresented groups for these positions. To insure that a diverse pool is reached, advertisements for the graduate position will be placed with groups focusing on diversity in science, including: ESA SEEDS, the Society for the Advancement of Chicanos and Native Americans in Science (SACNAS), and American Women in Science. The graduate student will receive both formal and informal professional development training at UNC, including a semester-long seminar on Professional Development Skills for Ecologists offered by the department in addition to weekly meetings with PI Hurlbert.

K-12 outreach. We will also engage in outreach activities targeted toward younger students. To facilitate this outreach, in collaboration with the UNC Center for Mathematics and Science Education, we will run a special session of the Pre-College Program's Saturday Academy. The Pre-College Program is an inquiry-based program of enrichment and encouragement in science and mathematics for women and minorities underrepresented in scientific and mathematical careers. Our Saturday Academy session will bring in 200 middle school students from the Chapel Hill-Carrboro City Schools, Durham Public Schools, and Orange County Schools to experience hands-on learning in math and science. In this session we will engage middle school students with their natural surroundings, introduce them to the idea that ecological systems are inherently dynamic, and identify the forces, such as global climate change and human disturbance, which might contribute to those dynamics. In conjunction with Hurlbert, the project graduate student will develop activities that engage middle school students with their natural surroundings at the North Carolina Botanical Gardens (NCBG), introducing them to the concepts of biodiversity, biotic versus abiotic interactions, and global change. Funds are included in the budget to support these activities.

Computational skills for working with big data. Working with the scale of data in this proposal requires computational and data management skills that most ecologists lack. White is a member of the Software Carpentry team (<http://software-carpentry.org>) that trains scientists in core computational skills and tools, and also teaches computational classes targeted at biologists. This proposal will provide funding (including travel for instructors and some participant support costs) for White to deliver one workshop each year, one at UNC and one at USU.

The project also involves the compilation of a large number of community and trait datasets. To facilitate the use and discovery of these data by other scientists, we will catalog them on both the Knowledge Network for Biocomplexity (<http://knb.ecoinformatics.org/>) and the Ecological Data Wiki (http://ecologicaldata.org), a wiki site developed by White for the identification of datasets and the sharing of best practices in their analysis (see Data Management Plan). Public datasets that are not already part of the EcoData Retriever (software development by White’s lab for simplifying the use of large datasets; <http://ecodataretriever.org/>) will be added to allow other researchers to quickly download and analyze the data.

**VI. Results of prior support**

**Ethan P. White:** NSF CAREER: Advancing Macroecology Using Informatics and Entropy Maximization (0953694; $657,499, 2010-2015). **Intellectual Merit** – This research develops and evaluates general theories of macroecology to establish linkages between patterns and simplify efforts to make predictions about ecological systems at large scales. Using one of the largest and most diverse datasets ever assembled in community ecology this research has: 1) shown that entropy maximization theories can predict commonness and rarity across ecosystems and taxonomic groups; 2) identified weaknesses with the current theories using strong tests evaluating numerous theoretical predictions simultaneously; and 3) developed new methods for contextualizing the agreement between macroecological patterns and theoretical predictions. **Broader Impacts** – This grant provides training and computational tools to allow ecologists to take advantage of the large amounts of ecological data that are available for analysis and include: 1) software to automatically download, cleanup, and install most of the major macroecological databases (this software is Highly Recommended, Highly Discussed, and Highly Cited based on Impact Story metrics); 2) an advanced wiki-based system to allow crowd-sourced identification and discussion of ecological datasets; and 3) courses taught at Utah State University and nationally through Software Carpentry to train ecologists in cutting edge computational skills. The grant has also supported the training of 2 postdoctoral researchers, 5 graduate students, and an undergraduate researcher. **Publications** – 8 publications (already cited over 30 times) including papers in *PLOS Biology*, *Ecology Letters*, *Ecology*, *American Naturalist*, and *Philosophical Transactions of the Royal Society B*.

(Nelson 1934; Beven 1976; Diamond & May 1977; Grossman 1982; Grossman *et al.* 1982; Kendeigh 1982; Novak 1983; Williamson 1983; Hall 1984; Svensson *et al.* 1984; Vickery & Nudds 1984; Pollard *et al.* 1986; Gibson & Hulbert 1987; Tilman 1987; Magnuson & Bowser 1990; Taylor *et al.* 1990; Henderson & Holmes 1991; Pollard 1991; Shaver & III 1991; Costello & Myers 1996; Theodose & Bowman 1997; Huberty *et al.* 1998; Knapp *et al.* 1998; Pigg *et al.* 1998; Ernest *et al.* 2000, 2009; Guo *et al.* 2000; Kaufman *et al.* 2000; Huenneke *et al.* 2001; Johnson & Rodriguez 2001; Bonjean & Lagerloef 2002; Zavaleta *et al.* 2003; Hijmans *et al.* 2005; Báez *et al.* 2006; Petraitis & Vidargas 2006; Sosebee & Cadrin 2006; Adler *et al.* 2007; Bloch *et al.* 2007; Paquette *et al.* 2007; Hampton *et al.* 2008; Petraitis *et al.* 2008; Sweatman *et al.* 2008; Yan *et al.* 2008; Cleland *et al.* 2008; Dapporto 2008; Dolan *et al.* 2009; Holmes *et al.* 2009; Merritt 2009; North American Butterfly Association 2009; Raimondi *et al.* 2009; Stapp 2009; McGlinn *et al.* 2010; Zachmann *et al.* 2010; Elmendorf & Harrison 2011; Fry *et al.* 2011; Sauer *et al.* 2011; Thibault *et al.* 2011; National Audubon Society 2012; Kelt *et al.* 2013; Rodriguez-Buritica *et al.* 2013; Chu *et al.* 2013)

**References**

1.  
Adler, P.B., Tyburczy, W.R. & Lauenroth, W.K. (2007). Long-term mapped quadrats from Kansas prairie: A unique source of demographic information for herbaceous plants. *Ecology*, 88, 2673.

2.  
Andrle, R.F. & Carroll, J.R. (Eds.). (1988). *The Atlas of Breeding Birds in New York State*. 1st Edition. Cornell Univ Pr.

3.  
Arlot, S. & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statist. Surv.*, 4, 40–79.

4.  
Báez, S., Collins, S.L., Lightfoot, D. & Koontz, T.L. (2006). Bottom-up regulation of plant community structure in an aridland ecosystem. *Ecology*, 87, 2746–2754.

5.  
Belmaker, J. (2009). Species richness of resident and transient coral-dwelling fish responds differentially to regional diversity. *Global Ecology and Biogeography*, 18, 426–436.

6.  
Beven, G. (1976). Changes in breeding bird populations of an oak-wood on Bookham Common, Surrey, over twenty-seven years. *London Naturalist*, 55, 23–42.

7.  
Bloch, C.P., Higgins, C.L. & Willig, M.R. (2007). Effects of large-scale disturbance on metacommunity structure of terrestrial gastropods: temporal trends in nestedness. *Oikos*, 116, 395–406.

8.  
Bock, C.E. & Root, T.L. (1981). The Christmas Bird Count and avian ecology. *Studies in Avian Biology*, 6, 17–23.

9.  
Bonjean, F. & Lagerloef, G.S.E. (2002). Diagnostic Model and Analysis of the Surface Currents in the Tropical Pacific Ocean. *Journal of Physical Oceanography*, 32, 2938–2954.

10.  
Brown, J.H., Ernest, S.K.M., Parody, J.M. & Haskell, J.P. (2001). Regulation of diversity: maintenance of species richness in changing environments. *Oecologia*, 126, 321–332.

11.  
Burnham, K.P. & Anderson, D.R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach*. Springer, New York.

12.  
Bystrak, D. (1981). The North American Breeding Bird Survey. *Studies in Avian Biology*, 6, 34–41.

13.  
Chase, J.M. & Leibold, M.A. (2002). Spatial scale dictates the productivity-biodiversity relationship. *Nature*, 416, 427–430.

14.  
Chu, C., Norman, J., Flynn, R., Kaplan, N., Lauenroth, W.K. & Adler, P.B. (2013). Cover, density, and demographics of shortgrass steppe plants mapped 1997–2010 in permanent grazed and ungrazed quadrats. *Ecology*, 94, 1435–1435.

15.  
Cleland, E.E., Clark, C.M., Collins, S.L., Fargione, J.E., Gough, L., Gross, K.L., *et al.* (2008). Species responses to nitrogen fertilization in herbaceous plant communities, and associated species traits. *Ecology*, 89, 1175–1175.

16.  
Collins, S.L. & Glenn, S.M. (1990). A hierarchical analysis of species’ abundance patterns in grassland vegetation. *American Naturalist*, 135, 633–648.

17.  
Connell, J.H. (1978). Diversity in Tropical Rain Forests and Coral Reefs. *Science*, 199, 1302–1310.

18.  
Costello, M.J. & Myers, A.A. (1996). Turnover of transient species as a contributor to the richness of a stable amphipod (Crustacea) fauna in a sea inlet. *Journal of Experimental Marine Biology and Ecology*, 202, 49–62.

19.  
Coyle, J.R., Hurlbert, A.H. & White, E.P. (2013). Opposing mechanisms drive richness patterns of core and transient bird species. *American Naturalist*, in press.

20.  
Cruz-Motta, J.J., Miloslavich, P., Palomo, G., Iken, K., Konar, B., Pohle, G., *et al.* (2010). Patterns of Spatial Variation of Assemblages Associated with Intertidal Rocky Shores: A Global Perspective. *PLoS ONE*, 5, e14354.

21.  
Dapporto, L. (2008). Core and satellite butterfly species on Elba island (Tuscan Archipelago, Italy). A study on persistence based on 120 years of collection data. *Journal of Insect Conservation*, 13, 421–428.

22.  
Diamond, J.M. & May, R.M. (1977). Species turnover rates on islands: dependence on census interval. *Science*, 197, 266–270.

23.  
Dolan, J.R., Ritchie, M.E., Tunin‐Ley, A. & Pizay, M. (2009). Dynamics of core and occasional species in the marine plankton: tintinnid ciliates in the north‐west Mediterranean Sea. *Journal of Biogeography*, 36, 887–895.

24.  
Elmendorf, S.C. & Harrison, S.P. (2011). Is plant community richness regulated over time? Contrasting results from experiments and long-term observations. *Ecology*, 92, 602–609.

25.  
Ernest, S.K., Brown, J.H. & Parmenter, R.R. (2000). Rodents, plants, and precipitation: spatial and temporal dynamics of consumers and resources. *Oikos*, 88, 470–482.

26.  
Ernest, S.K.M., Valone, T.J. & Brown, J. H. (2009). Long-term monitoring and experimental manipulation of a Chihuahuan Desert ecosystem near Portal, Arizona, USA. *Ecology*, 90, 1708.

27.  
Freestone, A.L. & Harrison, S. (2006). Regional enrichment of local assemblages is robust to variation in local productivity, abiotic gradients, and heterogeneity. *Ecology Letters*, 9, 95–102.

28.  
Froese, R. & Pauly, D. (2012). FishBase. World Wide Web electronic publication. www.fishbase.org.

29.  
Fry, J.A., Xian, G., Jin, S., Dewitz, J.A., Homer, C.G., Yang, L., *et al.* (2011). Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogrammetric engineering and remote sensing*, 77, 858–864.

30.  
Geisser, S. (1993). *Predictive interference: an introduction*. CRC Press.

31.  
Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

32.  
Gibson, D.J. & Hulbert, L.C. (1987). Effects of Fire, Topography and Year-to-Year Climatic Variation on Species Composition in Tallgrass Prairie. *Vegetatio*, 72, 175–185.

33.  
Grime, J.P. (1998). Benefits of Plant Diversity to Ecosystems: Immediate, Filter and Founder Effects. *Journal of Ecology*, 86, 902–910.

34.  
Grossman, G.D. (1982). Dynamics and Organization of a Rocky Intertidal Fish Assemblage: The Persistence and Resilience of Taxocene Structure. *The American Naturalist*, 119, 611–637.

35.  
Grossman, G.D., Moyle, P.B. & Whitaker, J.O. (1982). Stochasticity in Structural and Functional Characteristics of an Indiana Stream Fish Assemblage: A Test of Community Theory. *The American Naturalist*, 120, 423–454.

36.  
Guo, Q., Brown, J.H. & Valone, T.J. (2000). Abundance and Distribution of Desert Annuals: Are Spatial and Temporal Patterns Related? *Journal of Ecology*, 88, 551–560.

37.  
Hall, G.A. (1984). A Long-Term Bird Population Study in an Appalachian Spruce Forest. *The Wilson Bulletin*, 96, 228–240.

38.  
Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D’Agrosa, C., *et al.* (2008). A Global Map of Human Impact on Marine Ecosystems. *Science*, 319, 948–952.

39.  
Hampton, S.E., Izmest’eva, L.R., Moore, M.V., Katz, S.L., Dennis, B. & Silow, E.A. (2008). Sixty years of environmental change in the world’s largest freshwater lake – Lake Baikal, Siberia. *Global Change Biology*, 14, 1947–1958.

40.  
Hanski, I. (1982). Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos*, 38, 210–221.

41.  
Harrison, S. & Cornell, H.V. (2008). Toward a better understanding of the regional causes of local community richness. *Ecology Letters*, 11, 969–979.

42.  
Hawkins, B.A. & Porter, E.E. (2003). Water-energy balance and the geographic pattern of species richness of western Palearctic butterflies. *Ecological Entomology*, 28, 678–686.

43.  
Henderson, P.A. & Holmes, R.H.A. (1991). On the population-dynamics of dab, sole and flounder within Bridgwater bay in the lower Severn Estuary, England. *Netherlands Journal of Sea Research*, 27, 337–344.

44.  
Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.

45.  
Holmes, R.T., Sturges, F.W. & Sherry, T.W. (2009). Bird abundances at Hubbard Brook (1969-2004) and on three replicate plots (1986-2000) in the White Mountain National Forest.

46.  
Holyoak, M. & Loreau, M. (2006). Reconciling empirical ecology with neutral community models. *Ecology*, 87, 1370–1377.

47.  
Hortal, J., Rodríguez, J., Nieto-Díaz, M. & Lobo, J.M. (2008). Regional and environmental effects on the species richness of mammal assemblages. *Journal of Biogeography*, 35, 1202–1214.

48.  
Huberty, L.E., Gross, K.L. & Miller, C.J. (1998). Effects of Nitrogen Addition on Successional Dynamics and Species Diversity in Michigan Old-Fields. *Journal of Ecology*, 86, 794–803.

49.  
Huenneke, L.F., Clason, D. & Muldavin, E. (2001). Spatial heterogeneity in Chihuahuan Desert vegetation: implications for sampling methods in semi-arid ecosystems. *Journal of Arid Environments*, 47, 257–270.

50.  
Hurlbert, A.H. & Haskell, J.P. (2003). The effect of energy and seasonality on avian species richness and community composition. *American Naturalist*, 161, 83–97.

51.  
Hurlbert, A.H. & White, E.P. (2005). Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. *Ecology Letters*, 8, 319–327.

52.  
Hurlbert, A.H. & White, E.P. (2007). Ecological correlates of geographic range occupancy in North American birds. *Global Ecology and Biogeography*, 16, 764–773.

53.  
Johnson, L. & Rodriguez, D. (2001). *Terrestrial vegetation monitoring, Channel Islands National Park, 1996-2000 Report* ( No. Technical Report 01-06). National Park Service, Channel Islands National Park.

54.  
Karlson, R.H., Cornell, H.V. & Hughes, T.P. (2004). Coral communities are regionally enriched along an oceanic biodiversity gradient. *Nature*, 429, 867–870.

55.  
Kaufman, D.M., Kaufman, G.A. & Kaufman, D.W. (2000). Faunal structure of small mammals in tallgrass prairie: an evaluation of richness and spatiotemporal nestedness. In: *Reflections of a naturalist: papers honoring Professor Eugene D. Fleharty*, Fort Hays Studies, Special Issue 1. pp. 47–70.

56.  
Kelt, D.A., Meserve, P.L., Gutiérrez, J.R., Milstead, W.B. & Previtali, M.A. (2013). Long-term monitoring of mammals in the face of biotic and abiotic influences at a semiarid site in north-central Chile. *Ecology*, 94, 977–977.

57.  
Kendeigh, S.C. (1982). *Bird populations in east central Illinois: fluctuations, variations, and development over a half-century.* Illinois Biological Monographs 52. University of Illinois Press, Champaign, USA.

58.  
Kerr, J.T., Southwood, T.R.E. & Cihlar, J. (2001). Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. *PNAS*, 98, 11365–11370.

59.  
Kinlan, B.P. & Gaines, S.D. (2003). Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology*, 84, 2007–2020.

60.  
Knapp, A.K., Briggs, J.M., Hartnett, D.C. & Collins, S.L. (1998). *Grassland dynamics: long-term ecological research in tallgrass prairie*. Oxford University Press, Oxford, UK.

61.  
Legendre, P. & Legendre, L. (1998). *Numerical ecology*. Elsevier Science, Amsterdam.

62.  
MacArthur, R.H. (1972). *Geographical ecology*. Princeton University Press, Princeton, NJ.

63.  
Magnuson, J.J. & Bowser, C.J. (1990). A network for long-term ecological research in the United States. *Freshwater Biology*, 23, 137–143.

64.  
Magurran, A.E. (2007). Species abundance distributions over time. *Ecology Letters*, 10, 347–354.

65.  
Magurran, A.E. & Henderson, P.A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, 422, 714–716.

66.  
McGeoch, M.A. & Gaston, K.J. (2002). Occupancy frequency distributions: patterns, artefacts and mechanisms. *Biological Reviews*, 77, 311–331.

67.  
McGlinn, D.J., Earls, P.G. & Palmer, M.W. (2010). A 12-year study on the scaling of vascular plant composition in an Oklahoma tallgrass prairie. *Ecology*, 91, 1872.

68.  
Mellin, C., Bradshaw, C.J.A., Meekan, M.G. & Caley, M.J. (2010). Environmental and spatial predictors of species richness and abundance in coral reef fishes. *Global Ecology and Biogeography*, 19, 212–222.

69.  
Merritt, J.F. (2009). Long term mammal data from Powdermill Biological Station. Dataset ID VCR99062.

70.  
National Audubon Society. (2012). *The Christmas Bird Count Historical Results. Available http://www.christmasbirdcount.org*.

71.  
NEFSC. (1988). *An evaluation of the bottom trawl survey program of the Northeast Fisheries Center* ( No. NMFS-F/NEC-52). U.S. Dep. Commer., NOAA Tech. Memo.

72.  
Nelson, E.W. (1934). *The influence of precipitation and grazing upon black grama grass range*. U.S. Dept. of Agriculture.

73.  
North American Butterfly Association. (2009). NABA Butterfly Counts: 2009 Report.

74.  
Novak, I. (1983). An efficient light-trap for catching insects. *Acta Entomologica Bohemoslovaca*, 80, 29–34.

75.  
Novotný, V. & Basset, Y. (2000). Rare Species in Communities of Tropical Insect Herbivores: Pondering the Mystery of Singletons. *Oikos*, 89, 564–572.

76.  
O’Connor, M.I., Bruno, J.F., Gaines, S.D., Halpern, B.S., Lester, S.E., Kinlan, B.P., *et al.* (2007). Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation. *PNAS*, 104, 1266–1271.

77.  
Olden, J.D., Lawler, J.J. & Poff, N.L. (2008). Machine learning methods without tears: a primer for ecologists. *Q Rev Biol*, 83, 171–193.

78.  
Opler, P.A., Lotts, K. & Naberhaus, T. (2013). Butterflies and Moths of North America. http://www.buterfliesandmoths.org/.

79.  
Paine, R.T. (1966). Food Web Complexity and Species Diversity. *The American Naturalist*, 100, 65–75.

80.  
Paquette, A., Laliberte, E., Bouchard, A., de Blois, S., Legendre, P. & Brisson, J. (2007). Lac Croche understory vegetation data set (1998–2006). *Ecology*, 88, 3209.

81.  
Paradis, E., Baillie, S.R., Sutherland, W.J. & Gregory, R.D. (1998). Patterns of natal and breeding dispersal in birds. *Journal of Animal Ecology*, 67, 518–536.

82.  
Petraitis, P.S., Liu, H. & Rhile, E.C. (2008). Densities and cover data for intertidal organisms in the Gulf of Maine, USA, from 2003 to 2007. *Ecology*, 89, 588.

83.  
Petraitis, P.S. & Vidargas, N. (2006). Marine intertidal organisms found in experimental clearings on sheltered shores in the Gulf of Maine, USA. *Ecology*, 87, 796.

84.  
Pigg, J., Coleman, M.S., Wright, J. & Gibbs, R. (1998). An ecological investigation of the ichthyofauna in the Deep Fork River, central Oklahoma: 1976 to 1996. *Proceedings of the Oklahoma Academy of Sciences*, 78, 67–110.

85.  
Planes, S., Jones, G.P. & Thorrold, S.R. (2009). Larval dispersal connects fish populations in a network of marine protected areas. *PNAS*, 106, 5693–5697.

86.  
Pollard, E. (1991). Monitoring butterfly numbers. In: *Monitoring for Conservation and Ecology*. Chapman and Hall, London.

87.  
Pollard, E., Hall, M.L. & Bibby, T.J. (1986). *Monitoring the abundance of butterflies 1976-1985*. Research and survey in nature conservation. Nature Conservancy Council, UK.

88.  
Raimondi, P.R., Ambrose, R., Blanchette, C., Brancato, M.S., Engle, J., Fong, D., *et al.* (2009). *PISCO: Intertidal: MARINe Core Surveys: Photo Plots*. Multi-Agency Rocky Intertidal Network.

89.  
Raunkiaer, C. (1934). *The life forms of plants and statistical plant geography*. Clarendon Press, Oxford, UK.

90.  
Richardson, A.J., Walne, A.W., John, A.W.G., Jonas, T.D., Lindley, J.A., Sims, D.W., *et al.* (2006). Using continuous plankton recorder data. *Progress in Oceanography*, 68, 27–74.

91.  
Ricketts, E.F. (1985). *Between Pacific Tides*. Stanford University Press, Stanford, CA.

92.  
Ricklefs, R.E. (1987). Community Diversity: Relative Roles of Local and Regional Processes. *Science*, 235, 167 –171.

93.  
Ricklefs, R.E. (2007). History and Diversity: Explorations at the Intersection of Ecology and Evolution. *The American Naturalist*, 170, S56–S70.

94.  
Ridgely, R.S., Allnutt, T.F., Brooks, T., McNicol, D.K., Mehlman, D.W., Young, B.E., *et al.* (2007). *Digital Distribution Maps of the Birds of the Western Hemisphere, version 3.0*. NatureServe, Arlington, Virginia, USA.

95.  
Rodriguez-Buritica, S., Raichle, H., Webb, R.H., Turner, R.M. & Venable, L. (2013). One hundred and six years of population and community dynamics of Sonoran Desert Laboratory perennials. *Ecology*, 94, 976–976.

96.  
Sanders, N.J., Lessard, J.-P., Fitzpatrick, M.C. & Dunn, R.R. (2007). Temperature, but not productivity or geometry, predicts elevational diversity gradients in ants across spatial grains. *Global Ecology and Biogeography*, 16, 640–649.

97.  
Sauer, J.R., Hines, J.E., Fallon, J.E., Pardieck, K.L., Ziolkowski, D.J. & Link, W.A. (2011). The North American Breeding Bird Survey, Results and Analysis 1966 - 2010. Version 12.07.2011. USGS Patuxent Wildlife Research Center, Laurel, MD.

98.  
Shaver, G.R. & III, F.S.C. (1991). Production: Biomass Relationships and Element Cycling in Contrasting Arctic Vegetation Types. *Ecological Monographs*, 61, 1–31.

99.  
Shmida, A. & Wilson, M.V. (1985). Biological determinants of species diversity. *Journal of Biogeography*, 12, 1–20.

100.  
Silverman, B.W. (1981). Using Kernel Density Estimates to Investigate Multimodality. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43, 97–99.

101.  
Smith, T.W. & Lundholm, J.T. (2010). Variation partitioning as a tool to distinguish between niche and neutral processes. *Ecography*, 33, 648–655.

102.  
Soil Survey Staff. (2013). Gridded Soil Survey Geographic (gSSURGO) Database for the Continental United States.

103.  
Soininen, J., Lennon, J.J. & Hillebrand, H. (2007). A multivariate analysis of beta diversity across organisms and environments. *Ecology*, 88, 2830–2838.

104.  
Sosebee, K.A. & Cadrin, S.X. (2006). A historical perspective on the abundance and biomass of Northeast complex stocks from NMFS and Massachusetts inshore bottom trawl surveys, 1963-2002. US Dep Commer, Northeast Fish Sci Cent Ref Doc. 06-05; 200 p.

105.  
Srivastava, D.S. & Lawton, J.H. (1998). Why more productive sites have more species: An experimental test of theory using tree-hole communities. *American Naturalist*, 152, 510–529.

106.  
Stapp, P. (2009). SGS-LTER Long-Term Monitoring Project: Small Mammals on Trapping Webs.

107.  
Stevens, V.M., Turlure, C. & Baguette, M. (2010). A meta-analysis of dispersal in butterflies. *Biological Reviews*, 85, 625–642.

108.  
Svensson, S., Carlsson, U.T. & Liljedahl, G. (1984). Structure and dynamics of an alpine bird community, a 20-year study. *Annales Zoologici Fennici*, 21, 339–350.

109.  
Sweatman, H., Cheal, A., Coleman, G., Emslie, M., Johns, K., Jonker, M., *et al.* (2008). *Long-term Monitoring of the Great Barrier Reef. Status Report Number 8*. Australian Institute of Marine Science, Townsville, Queensland, Australia.

110.  
Taylor, L.R., Woiwod, I.P., MacCaulay, E.D.M., Dupuch, M.J. & Nicklen, J. (1990). Rothamsted Insect Survey Annual Report.

111.  
Theodose, T.A. & Bowman, W.D. (1997). Nutrient availability, plant abundance, and species diversity in two alpine tundra communities. *Ecology*, 78, 1861–1872.

112.  
Thibault, K., Supp, S.R., Giffin, M., White, E.P. & Ernest, S.K.M. (2011). Species composition and abundance of mammalian communities. *Ecology*, 92, 2316.

113.  
Tilman, D. (1987). Secondary Succession and the Pattern of Plant Dominance Along Experimental Nitrogen Gradients. *Ecological Monographs*, 57, 189–214.

114.  
Tittler, R., Villard, M.-A. & Fahrig, L. (2009). How far do songbirds disperse? *Ecography*, 32, 1051–1061.

115.  
Tokeshi, M. (1992). Dynamics and distribution in animal communities; theory and analysis. *Researches in Population Ecology*, 34, 249–273.

116.  
Ulrich, W. & Ollik, M. (2004). Frequent and occasional species and the shape of relative-abundance distributions. *Diversity and Distributions*, 10, 263–269.

117.  
USDA, NRCS. (2012). *The PLANTS Database (http://plants.usda.gov, 20 July 2012)*. National Plant Data Team, Greensboro, NC 27401-4901 USA.

118.  
Vergnon, R., Dulvy, N.K. & Freckleton, R.P. (2009). Niches versus neutrality: uncovering the drivers of diversity in a species‐rich community. *Ecology Letters*, 12, 1079–1090.

119.  
Vickery, W.L. & Nudds, T.D. (1984). Detection of density-dependent effects in annual duck censuses. *Ecology*, 65, 96–104.

120.  
Wall, M.M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121, 311–324.

121.  
White, E.P. & Hurlbert, A.H. (2010). The combined influence of the local environment and regional enrichment on bird species richness. *The American Naturalist*, 175, E35–E43.

122.  
White, P. & Kerr, J.T. (2006). Contrasting spatial and temporal global change impacts on butterfly species richness during the 20th century. *Ecography*, 29, 908–918.

123.  
Williamson, M. (1983). The Land-Bird Community of Skokholm: Ordination and Turnover. *Oikos*, 41, 378–384.

124.  
Yan, N.D., Somers, K.M., Girard, R.E., Paterson, A.M., Keller, W. (Bill), Ramcharan, C.W., *et al.* (2008). Long-term trends in zooplankton of Dorset, Ontario, lakes: the probable interactive effects of changes in pH, total phosphorus, dissolved organic carbon, and predators. *Canadian Journal of Fisheries and Aquatic Sciences*, 65, 862–877.

125.  
Zachmann, L., Moffet, C. & Adler, P.B. (2010). Mapped quadrats in sagebrush steppe: long-term data for analyzing demographic rates and plant–plant interactions. *Ecology*, 91, 3427.

126.  
Zavaleta, E.S., Shaw, M.R., Chiariello, N.R., Thomas, B.D., Cleland, E.E., Field, C.B., *et al.* (2003). Grassland responses to three years of elevated temperature, CO2, precipitation, and N deposition. *Ecological Monographs*, 73, 585–604.