Abstract

The system-wide attributes of ecological communities - such as community-level abundance and metabolic flux, and how these are distributed among species and organisms - emerge from a web of shifting environmental constraints, diverse species interactions, and ubiquitous mathematical rules. While this apparent complexity can present a challenge to synthesis in community ecology, a macroecological perspective embraces ecological complexity as a path towards general understanding. In this dissertation, I use a telescoping macroecological perspective to explore how these factors shape community properties and determine how they change over time, building from a granular focus on species interactions in a well-studied experimental system, to successively broader spatial and conceptual scales in pursuit of general insights. In chapter 1, I offer an overview of the macroecological approach as it applies to community ecology and the specific vignettes in this dissertation. In chapter 2, I use a long-term experiment on desert rodents to disentangle how shifting environmental conditions and species interactions modulate the impact of species loss on community function. In chapter 3, I leverage modern computational approaches to show how changes in community structure modulate nuanced relationships between the long-term trends in size- and individuals- based currencies of community function. In chapter 4, I borrow tools and conceptual frameworks from statistical mechanics to explore what common ecological patterns stand to teach us about ecological, as opposed to statistical, processes. Finally, in chapter 5, I offer concluding reflections on the current landscape of prospects and challenges associated with a macroecological lens on community structure and function.

Acknowledgements

Chapter 1: Introduction

The interplay between system-specific natural history narratives and ubiquitous ecological or even mathematical rules that combine to determine how abundance, biomass, and resource use are distributed among species and organisms and across different levels of organization in ecological communities lies at the core of community ecology (**???**, Lawton 1994, Brown 1995). A macroecological approach integrates classic modalities of ecological inquiry with conceptual frameworks drawn from across the scope of complex systems studies and computationally-intensive methodologies in order to disentangle the specific processes and dynamics at play in particular systems from general processes that operate across diverse taxonomic, geographic, or temporal contexts (**???**, @mcgill_whathowandwhy). In this dissertation, I adopt a macroecological perspective to understanding the structure and function of ecological communities, working from a narrow, system-specific focus on a well-studied long-term experiment, to a broad taxonomic and conceptual perspective on the interplay between combinatorics, statistical mechanics, and community ecology.

In chapter 2, I use 30 years' of accumulated data and natural history knowledge to explore the effects of species loss on community function in an experimentally manipulated desert rodent community. Understanding how community function responds to species loss, and how the effects of species loss interact with shifting environmental conditions, is a key problem for biodiversity science in the current era of unprecedented ecological change. In this system, I find that compensation due to functional redundancy temporarily buffered community function against species loss. However, because similar, but non-identical, rodent species have responded differently to changes in environmental conditions over time, this compensatory effect has broken down, leaving community function highly sensitive to the loss of keystone species.

In chapter 3, I undertake a continental-scale comparison across communities to explore how shifts in community-wide body size modulate the long-term dynamics of total abundance, biomass, and energy use in North American breeding birds. Although total abundance and size and energy-based currencies are intrinsically linked, they capture different dimensions of community function, and shifts in community size structure can decouple the dynamics of different currencies. I find that, in nearly 1/3 of communities, changes in the community size structure result in qualitatively different trajectories for biomass and total abundance over the past 30 years. In nearly 80% of instances, total abundance has decreased, but community-wide mean body size has increased, partially offsetting the decline in biomass expected due to declines in individual abundance.

In chapter 4, I step further back to examine how fundamental mathematical constraints inform our understanding of ecological "laws". Common patterns in community ecology, such as the "hollow-curve" or J-shaped species abundance distribution (SAD), emerge from a combination of biological processes and ubiquitous mathematical constraints on the emergent properties of complex systems. Disentangling the signal of ecological processes from these mathematical constraints can provide new sources of inferential power linking pattern to process in community ecology. I use combinatorics to characterize the mathematical constraint on the SAD, and compare the SADs of 22,000 empirically-observed communities to these "statistical baselines". This reveals that, while empirical SADs often match their statistical baselines, a substantial minority of real SADs deviate from these baselines - leaving an important role for ecological processes in shaping these distributions.

Takeaways para.

Chapter 2: Maintenance of community function through compensation breaks down over time in a desert rodent community.

In press at Ecology as Diaz and Ernest (2022): Maintenance of community function through compensation breaks down over time in a desert rodent community

Background

Determining the extent to which community-level properties are affected by species loss, and how and why this changes over time, is key for understanding how communities are structured and how community function may respond to future perturbations (Gonzalez and Loreau 2009). When species are lost from a community, their contributions to community function (e.g. total productivity or resource use) are also directly lost. Community function may be maintained, however, if in the new community context, species that remain perform similar functions to the species that were lost, and compensate for the decline in function directly caused by species loss - i.e., functional redundancy (Walker 1992, 1995; Ernest and Brown 2001; Rosenfeld 2002; Gonzalez and Loreau 2009). When compensation via functional redundancy occurs among consumers with a common resource base, it is consistent with a zero-sum competitive dynamic, in which resources not used by one species are readily absorbed by competitors, and any increases in the abundance of one species must come at the direct expense of others (Van Valen 1973; Ernest et al. 2008).

Because the response of system-level function to species loss is partially determined by the degree of functional redundancy in a community, processes that cause functional redundancy to change over time can have important consequences for the long-term maintenance of ecosystem function. Colonization events may buffer community function against species loss, if a community gains species that perform similar functions to the species that were lost (Ernest and Brown 2001; Leibold et al. 2017). The ability of colonization to supply functionally redundant species depends on the species (and traits) present in the broader metacommunity, and on the rate of dispersal supplying appropriate species to local communities (Leibold et al. 2017).

Even without the addition of new species and traits, however, functional redundancy within a consistent set of coexisting species may fluctuate over time. While, in theory, functional redundancy may occur via the special case of complete niche neutrality (e.g. Hubbell 2001), it may also occur in niche-structured systems that contain species that share some traits but differ along other niche axes (Thibault et al. 2010). In these systems, if similar, but non-identical, species respond to environmental change in similar ways, functional overlap can be maintained or even strengthened. However, if niche differences cause species to respond differently to changing conditions, the degree of functional overlap between those species may decline, resulting in a breakdown in compensation (Loreau 2004; Fetzer et al. 2015). Over time, as metacommunity dynamics and changing environmental conditions modulate functional redundancy within a community, the extent to which community function is robust to species loss - and the strength of zero-sum competition - may also be dynamic and context-dependent.

Despite logical conceptual support, and evidence from experimental microcosms (Fetzer et al. 2015), there is little empirical documentation of how, and through which mechanisms, temporal changes in functional redundancy modulate the effect of species loss on ecosystem function in natural assemblages. Although relatively plentiful, observational data cannot unambiguously detect compensation through functional redundancy, and even short-term experiments may not be sufficient to capture temporal variation in compensation (Ernest and Brown 2001; Houlahan et al. 2007). In contrast, long-term manipulative experiments are uniquely suited to address this question. In long-term experiments in which key species are removed from a community over an extended period of time, the impact of species loss on community function can be directly quantified by comparing community function between complete and manipulated assemblages. As metacommunity dynamics and environmental conditions shift over time, long-term monitoring can reveal how these processes contribute to changes in functional redundancy and ecosystem function across different time periods. Due to the financial and logistical resources required to maintain and monitor whole-community manipulations over long timescales, these experiments are rare in natural systems representative of realistic evolutionary, geographic, and environmental constraints (Hughes et al. 2017).

Here, we use a 30-year experiment on desert rodents to investigate how shifts in functional redundancy alter the effect of species loss on community function over time. In this study, kangaroo rats (Dipodomys spp.), the largest and competitively dominant species in the rodent community, have been removed from a subset of experimental plots to explore how the loss of key species affects community function, measured as community-level metabolic flux ("total energy use", or *Etot*) or total biomass (Ernest et al. 2019). For systems of consumers with a shared resource base, such as this community of granivorous rodents, *Etot* reflects the total amount of resources being processed by an assemblage, and total biomass directly reflects standing biomass. Both are important metrics of community function (Lawton 1994; Ernest and Brown 2001). Long-term monitoring of this experiment has documented repeated shifts in the habitat and species composition of this system, resulting in distinct time periods characterized by different habitat conditions and configurations of the rodent community (Christensen et al. 2018). Abrupt reorganization events in community composition occurred in 1997 and in 2010, associated with the establishment and subsequent decline of the pocket mouse Chaetodipus baileyi. *C. baileyi* is similar in size, and presumably other traits, to kangaroo rats, and its establishment in 1996-97 drove a pronounced increase in compensation due to functional redundancy between *C. baileyi* and kangaroo rats (Ernest and Brown 2001; Thibault et al. 2010). Over the course of this experiment, shifting environmental conditions have caused the habitat at the study site to transition from desert grassland to scrub, driving a shift in baseline rodent community composition away from kangaroo rats and favoring other, smaller, granivores (Brown et al. 1997; Ernest et al. 2008). By making comparisons across these time periods, we explored how shifts in community composition and functional overlap among the same species have contributed to long-term changes in the effect of species loss on community function.

Methods

The Portal Project

The Portal Project consists of a set of 24 fenced experimental plots located approximately 7 miles east of Portal, AZ, USA, on unceded land of the Chiricahua Apache. Beginning in 1977, kangaroo rats (*Dipodomys spectabilis*, *D. merriami*, and *D. ordii*) have been experimentally excluded from a subset of these plots (exclosures), while all other rodents are allowed access through small holes cut in the plot fencing. Control plots, with larger holes, are accessible to all rodents, including kangaroo rats. Rodents on all plots are censused via monthly bouts of live-trapping. Each individual captured is identified to species and weighed. For additional details on the site and methodology of the Portal Project, see Ernest et al. (2019).

Data

We used data for control and exclosure plots from February 1988 until January 2020. The experimental treatments for some plots have changed over time, and we used the subset of plots that have had the same treatments for the longest period of time (Ernest et al. 2019). Four control plots, and five exclosure plots, met these criteria. In order to achieve a balanced sample, we randomly selected four exclosure plots for analysis. We divided the timeseries into three time periods defined by major transitions in the rodent community surrounding the establishment and decline of *C. baileyi* (Ernest and Brown 2001; Christensen et al. 2018). The first time period (February 1988-June 1997) precedes *C. baileyi*'s establishment at site. We defined *C. baileyi*'s establishment date as the first census period in which *C. baileyi* was captured on all exclosure plots (following Bledsoe and Ernest, 2019). During the second time period (July 1997-January 2010), *C. baileyi* was abundant on both exclosure and control plots. This time period ended with a reorganization event in which C. balieyi became scarce sitewide. We used January 2010, the midpoint of the 95% credible interval for the date of this reorganization event as estimated in Christensen et al. (2018), as the end date for this time period. The last time period spans from Feburary 2010-January 2020. For each individual rodent captured, we estimated the individual-level metabolic rate using the scaling relationship between individual body mass and metabolic rate b = 5.69 * (m0.75), where m is body mass in grams and b is metabolic rate (for details, see White et al. 2004). We calculated treatment and species-level energy use as the sum of the appropriate individuals' metabolic rates, and total biomass as the sum of individuals' body mass measurements.

Statistical analysis of rodent community energy use and biomass

Here, we describe analyses for energy use. For biomass, we repeated these analyses substituting biomass values for energy use throughout. For all variables, we combined data for all plots within a treatment in each monthly census period and calculated treatment-level means. This is necessary to calculate compensation, and we treated other variables in the same way to maintain consistency. A provisional plot-level analysis yielded qualitatively equivalent results (Appendix S1). To measure the overall impact of kangaroo rat removal on $Etot$, we calculated a "total energy ratio" as the ratio of treatment-level $Etot$ for kangaroo-rat exclosure plots relative to unmanipulated control plots, i.e. $Etot_E/Etot_C$ where $Etot_E$ and $Etot_C$ are total energy use on exclosures and controls, respectively (Thibault et al 2010; Bledsoe and Ernest 2019). This ratio is distinct from compensation, which we defined as the proportion of the energy made available by kangaroo rat removal taken up via compensatory increases in energy use by small granivores (all granivores other than kangaroo rats; Baiomys taylori, *C. baileyi*, Chaetodipus hispidus, Chaetodipus intermedius, Chaetodipus penicillatus, Perognathus flavus, Peromyscus eremicus, Peromyscus leucopus, Peromyscus maniculatus, Reithrodontomys fulvescens, Reithrodontomys megalotis, and Reithrodontomys montanus). We calculated this as $(SGE - SGC)/KRC$, where SGE and SGC are the amount of energy used by small granivores (SG) on exclosure and control plots, respectively, and KRC is the amount of energy used by kangaroo rats (KR) on control plots (Ernest and Brown 2001). To compare these variables across time periods, we used generalized least squares models (GLS; the R package nlme; Pinheiro et al. 2020) of the form `(SGE - SGC)/KRC ~ time period`, for compensation, and `EtotE/EtotC ~ time period`, for the total energy ratio. We included a continuous-time autoregressive temporal autocorrelation term to account for temporal autocorrelation between values from monthly census periods within each multi-year time period (for details of model selection, see Appendix S2). To evaluate change in baseline community composition over time, we calculated the proportion of treatment-level energy use accounted for by kangaroo rats on control plots in each census period ($KRC/EtotC$). Proportional energy use is bounded 0-1 and is therefore not appropriate for GLS, so we compared values across time periods using a binomial generalized linear model (GLM) of the form `KRC/EtotC ~ time period`. Finally, we calculated the proportional energy use accounted for by *C. baileyi* (CB) on exclosure and control plots in each census period ($CBE/EtotE$ and $CBC/EtotC$, respectively). *C. baileyi* was not present at the site prior to 1996, and we restricted the analysis of *C. baileyi* proportional energy use to the second two time periods. We compared *C. baileyi* proportional energy use over time and across treatments using a binomial GLM of the form `CBE/EtotE ~ time period + treatment`. For all models, we calculated estimated means and 95% confidence or credible intervals for time-period (and, for *C. baileyi*, treatment) level values, and contrasts between time periods (and, for *C. baileyi*, treatments), using the R package emmeans (Lenth 2021). Analyses were conducted in R 4.0.3 (R Core Team 2020). Data and code are archived at https://doi.org/10.5281/zenodo.5544361 and https://doi.org/10.5281/zenodo.5539880.

Results

The impact of kangaroo rat removal on community function has changed repeatedly over time, through a combination of abrupt shifts in compensation associated with *C. baileyi*, and long-term changes in baseline community composition sitewide (Figure 1). These dynamics are qualitatively identical whether function is measured as total energy use (Figure 1; Appendix S2) or total biomass (Appendix S3). The first shift coincided with *C. baileyi*'s establishment in the community beginning in 1996-97 (Figure 1D). *C. baileyi* rapidly became dominant on exclosure plots and dramatically increased compensation (Figure 1B). From 1997-2010, small granivores compensated for an average of 58% of kangaroo rat energy use on control plots (95% interval 48-67%), an increase from an average of 18% from 1988-1997 (95% interval 8-29%; contrast p < 0.001; for complete results of all models, see Appendix S2) from 1997-2010. With *C. baileyi*'s addition to the community, the total energy ratio (on exclosures relative to controls; Figure 1A) increased from 30% (20-40%) to 71% (62-79%, contrast p < 0.014). In the second shift, beginning around 2010, *C. baileyi*'s abundance sitewide dropped precipitously (Figure 1D). *C. baileyi*'s proportional energy use dropped from an average of 72% (65-80%) to 26% (18-35%, contrast p < 0.001) on exclosure plots, and from 11% (6-16%) to essentially 0 on control plots (contrast p < 0.001). Other species of small granivore did not make compensatory gains to offset the decline in *C. baileyi* (Figure 1B). As a result, compensation declined from an

average of 58% (48-67%) to 28% (17-38%, contrast p = 0.002), a level not significantly different from the 18% (8-29%, contrast p = .44) observed prior to *C. baileyi*'s establishment at the site. Somewhat paradoxically, while the total energy ratio also dropped following *C. baileyi*'s decline, from an average of 71% (62-79%) from 1997-2010 to 50% (40-60%, contrast p = 0.0056) from 2010-2020, it remained higher than its average of 30% (20-40%, contrast p = 0.0144) from 1988-1997 (Figure 1A). Over the course of the experiment, community composition shifted sitewide. In later years, kangaroo rats accounted for a lower proportion of baseline Etot than they did at the beginning of the study (Figure 1C). From 1988-1997, kangaroo rats accounted for 92% (87-97%) of Etot on controls; after 1997, this dropped to an average of approximately 70% (1988-1997 compared to later time periods, both p = .0004; 1997-2010 and 2020-2020 not significantly different, p = .976). Because the proportion of Etot directly lost to kangaroo rat removal was smaller from 2010-2020 than from 1988-1997, the total energy ratio was higher from 2010-2020 than it was from 1988-1997 - even though there was not a detectable difference between the two time periods in the proportion of lost energy being offset through compensation.

Discussion

The dynamics of rodent community energy use at Portal illustrate that the role of functional redundancy in buffering community function against species loss fluctuates over time, due to changes in both species composition and in the degree of functional overlap among the same species. The 1997 increase in compensation, driven by *C. baileyi*'s establishment at the site, was a clear and compelling instance of colonization from the regional species pool overcoming limitations on functional redundancy (Ernest and Brown 2001; Leibold et al 2017). Although the small granivore species originally present in the community did not possess the traits necessary to compensate for kangaroo rats, *C. baileyi* supplied those traits and substantially, but incompletely, restored community function. In contrast, following the community reorganization event in 2010, *C. baileyi* remained present in the community, but ceased to operate as a partial functional replacement for kangaroo rats. This is consistent with fluctuating conditions modulating functional redundancy between similar, but non-identical, competitors. Kangaroo rats and *C. baileyi* are relatively similar in size and are demonstrably capable of using similar resources. However, *C. baileyi* prefers different, shrubbier microhabitats than kangaroo rats, and the two groups have been observed to replace each other in adjacent habitats (Ernest and Brown 2001). We suggest that this study site, which has historically been dominated by kangaroo rats, constitutes marginal habitat for *C. baileyi*, and that, while conditions from 1997-2010 aligned sufficiently with *C. baileyi*'s requirements to create appreciable functional redundancy between kangaroo rats and *C. baileyi*, conditions since have caused this redundancy to break down. *C. baileyi*'s decline occurred immediately following a period of low plant productivity and low rodent abundance community-wide, and in the decade following, the site experienced two long and severe droughts (Appendix S4; Christensen et al. 2018). These extreme conditions may themselves have limited *C. baileyi*'s fitness at the site, or the community-wide low abundance event may have temporarily overcome incumbency effects and triggered a community shift tracking longer-term habitat trends (Thibault and Brown 2008; Christensen et al. 2018). Regardless of the proximate cause of *C. baileyi*'s decline, the fact that *C. baileyi* remains in the community, but no longer compensates for kangaroo rats, illustrates that changing conditions can have profound effects on community function by modulating the degree of functional redundancy within a consistent set of species.

While changes in compensation have contributed to changes in community function in this system, changes in compensation alone do not fully account for the long-term changes in the overall impact of kangaroo rat removal on Etot. Since 2010, although the ratio of Etot on exclosure plots relative to control plots declined coinciding with the breakdown in compensation associated with *C. baileyi*, it remained higher than the levels observed prior to 1997 (Figure 1A). This difference between the first and last time periods cannot be explained by an increase in compensation, as compensation from 2010-2020 was not greater than pre-1997 levels (Figure 1B). Rather, the increase in Etot on exclosure plots relative to control plots was the result of a long-term decrease in the contribution of kangaroo rats to Etot sitewide. Because kangaroo rats accounted for a smaller proportion of Etot on control plots from 2010-2020 than they did prior to 1997, their removal had a smaller impact on community function – even though there was not an increase in the degree to which small granivores compensated for their absence. In fact, the comparable levels of compensation achieved in the decades preceding and following *C. baileyi*'s dominance at the site suggest a relatively stable, and

limited, degree of functional overlap between kangaroo rats and the original small granivores (i.e., excluding C. bailyei). Niche complementarity, combined with changing habitat conditions, may partially explain this phenomenon. It is well-documented that, while kangaroo rats readily forage in open microhabitats where predation risk can be relatively high, smaller granivores preferentially forage in sheltered microhabitats as an antipredator tactic (Kelt 2011). Over the course of this experiment, the habitat at this study site has transitioned from an arid grassland to a shrubland (Brown et al. 1997). As sheltered microhabitats became more widespread, small granivores may have gained access to a larger proportion of resources and increased their share of Etot sitewide. However, kangaroo rats may have continued to use resources in open areas, which would have remained inaccessible to smaller granivores even on exclosure plots. The long-term reduction in the impact of kangaroo rat removal on community function, driven by niche complementarity and consistent niche partitioning, contrasts with the temporary compensatory dynamic driven by functional redundancy with *C. baileyi*. Although changes in the overall effect of species loss are sometimes treated interchangeably with compensation (e.g. Ernest and Brown 2001 compared to Thibault et al. 2010), it is important to recognize that multiple distinct pathways modulate the long-term impacts of species loss on community function. Particularly in strongly niche-structured systems, complementarity effects and fluctuations in functional redundancy may occur simultaneously, with complex and counterintuitive impacts on community function.

Overall, the decadal-scale changes in energy use among the Portal rodents underscore the importance of long-term metacommunity dynamics to the maintenance of community function following species loss (see Leibold et al. 2017). Although a single colonization event may allow for temporary compensation via functional redundancy, as conditions shift, species that once compensated may no longer perform that function (see also Isbell et al. 2011). Particularly if limiting similarity prevents similar competitors from specializing on precisely the same habitats (Rosenfeld 2002), temporary, context-dependent compensation may be common. To maintain compensation over time, multiple colonization events, supplying species that are functionally redundant under different conditions, may be required. Depending on dispersal rates, and the diversity and composition of regional species pools, this may be unlikely or even impossible. At Portal, dispersal limitation introduced a 20-year delay in the compensatory response driven by *C. baileyi*. Theoretically, a new species capable of compensating for kangaroo rats, and better-suited to conditions at the site since 2010, could restore compensation under present conditions – but it is unclear whether this species exists or if it can disperse to this site. As ecosystems globally undergo reductions in habitat connectivity and regional beta diversity, and enter novel climatic spaces, maintenance of community function via functional redundancy may grow increasingly rare and fragile (Dornelas et al. 2014; Williams and Jackson 2007).

Finally, the long-term variability in functional redundancy documented here adds important nuance to our understanding of how zero-sum dynamics operate in natural assemblages. Theories invoking zero-sum dynamics, and tests for compensatory dynamics in empirical data, often treat a zero-sum dynamic as a strong and temporally consistent constraint (Hubbell 2001; Houlahan et al. 2007). In this framing, any resources made available via species loss should immediately be taken up by other species. This is not consistent with the dynamics that occur at Portal, which has seen extended periods of time when resources are available on exclosure plots but are not used. Rather, these results are more consistent with a zero-sum constraint operating at metacommunity or evolutionary scales (Van Valen 1973; Terry and Rowe 2015; Leibold et al. 2017). Over short timescales, or within a closed local assemblage, niche differences may weaken zero-sum effects, especially under fluctuating conditions. However, over larger temporal and spatial scales, dispersal or evolution may supply new species equipped to use available resources - via either functional redundancy, or niche complementarity allowing them to exploit novel niches. A long-term, metacommunity, and even macroevolutionary approach may be necessary to fully understand how zero-sum constraints, functional redundancy, and niche complementarity contribute to the maintenance of community-level function in the face of species extinctions and changing conditions over time.

Chapter 3: Shifts in the individual size distribution decouple long-term trends in abundance, biomass, and energy use of North American breeding birds

Background

Understanding the interrelated dynamics of size- and -abundance based dimensions of biodiversity is key to understanding biodiversity change in the Anthropocene. Total abundance - i.e. the total number of individual organisms present in a system - and size-based currencies, such as the total biomass or total metabolic flux ("energy use") of a system, are intertwined, but nonequivalent, measures of biological function. Abundance is more closely tied to species-level population dynamics, while size-based metrics more directly reflect assemblage-level resource use and contributions to materials fluxes at the broader ecosystem scale (Morlon et al. 2009, Dornelas et al. 2011, Connolly et al. 2005, White et al. 2007). While these currencies are naturally linked (Morlon et al. 2009, Henderson and Magurran 2010), changes in size composition can decouple the dynamics of one currency from another (Ernest et al. 2009, Dornelas et al. 2011, White et al. 2004, 2007, Yen et al. 2017). This can mean that intuition from one currency may be misleading about others; a trend in numerical abundance might mask alternative dynamics occurring with respect to biomass or total energy use (White et al. 2004). Changes in size composition strong enough to decouple currencies may be symptomatic of important changes in ecosystem status- e.g. abundance-biomass comparison curves (Petchey and Belgrano 2010); size-biased extinctions (Young et al. 2016, Smith et al. 2018). This underscores the need to understand how these dynamics are playing out in the Anthropocene (Fisher et al. 2010).

At the community scale, changes in the relationship between size and abundance can signal important shifts in community structure and functional composition. To the extent that size is a proxy for other functional traits, changes or consistency in the community-level size structure (individiual size distribution, ISD) over time may reflect processes related to niche structure (White et al. 2007, Petchey and Belgrano 2010). Strong size shifts can decouple the relationship between abundance and biomass. In aquatic systems, such changes in the scaling between abundance and biomass often signal ecosystem degradation (**???**, Warwick and Clarke 1994, Petchey and Belgrano 2010). Compensatory shifts in the size structure can buffer community function (in terms of biomass or energy use) against changes in abundance (Ernest et al. 2009, White et al. 2004, Terry and Rowe 2015). Or, consistency in the size structure may maintain the relationship between size- and -abundance based currencies, even as species composition, total abundance, and total biomass/total energy use fluctuate over time, which can reflect consistency in the niche structure over time (Holling 1992).

It is important to improve our understanding of these dynamics for terrestrial animal communities in particular. In contrast to terrestrial trees and aquatic systems (**???**, White et al. 2007), how the relationship between size and abundance changes over time, and the consequences of these changes for ecosystem-level properties, remain relatively unknown for terrestrial animals (but see White et al. (2004)). Terrestrial animal communities exhibit size structure (Thibault et al. 2011, Ernest 2005), and case studies have demonstrated that size shifts can decouple the dynamics of abundance, biomass, and energy use for terrestrial animals (White et al. 2004, Yen et al. 2017), but do not always do so (Hernández et al. 2011). Establishing generalities in these dynamics is especially pertinent in the Anthropocene, as these communities are experiencing extensive and potentially size-structured change, with implications at community, ecosystem, and global scales (Young et al. 2016, Schmitz et al. 2018).

Macroecological-scale synthesis on the interrelated dynamics of the ISD, total abundance, and community function for terrestrial animals has been constrained by 1) a lack of community-level size and abundance timeseries data for these systems (Thibault et al. 2011, White et al. 2007), and 2) appropriate statistical methods for relating change in the size structure to changes in abundance and function (Thibault et al. 2011, Yen et al. 2017). In contrast to aquatic and forest systems, most long-term surveys of animal communities do not collect data on individuals' *sizes* across a full community (with the exception of small mammal studies, which have made major contributions to our understanding of the dynamics of size, abundance, and function for these systems; (White et al. 2004, Ernest 2005, Hernández et al. 2011, Kelt et al. 2015)). Global, continental, or population-wide studies capture different phenomena (White et al. 2007, McGill et

al. 2015). The ISDs for terrestrial animals, and specifically for determinate growing taxa (e.g. mammals, birds), are often complex, multimodal distributions strongly determined by community *species* composition (Holling 1992, Thibault et al. 2011, Ernest 2005, Yen et al. 2017) and less statistically tractable than the power-law ISDs found in aquatic and tree systems (**???**, White et al. 2007). Quantifying change in the size structure, and relating this to change in community-wide abundance and function, is not as straightforward as computing and comparing slopes. As a result, we do not have a general understanding of either 1) how the size structures for these systems behave over time or 2) the extent to which changes in community size structure decouple the community-level dynamics of abundance, biomass, and energy use in these systems.

Here, we begin to address this gap by exploring how temporal changes in the size structure modulate the relationship between total abundance, energy, and biomass for communities of North American breeding birds. We used allometric scaling to estimate community size and abundance data for the North American Breeding Bird Survey, and evaluated how changes in total abundance, biomass, and energy use have co-varied from 1988-2018. Specifically, we examined: 1) How often do these currencies change together vs. have decoupled dynamics?; 2) What are the dominant directions and magnitudes of the overall change over time and degree of decoupling between the currencies?; 3) To what extent do changes in species composition and community size structure translate into decoupling in the temporal trends of different currencies at the community scale?

Methods

Bird abundance data

We used data from the Breeding Bird Survey (Pardieck et al. 2019) to evaluate trends in abundance, biomass, and energy use. The Breeding Bird Survey consists of roughly 40km-long survey routes distributed throughout the United States and Canada. Routes are surveyed annually during the breeding season (predominately May-June), via 50 3-minute point counts during which all birds seen or heard are identified to species (Pardieck et al. 2019). Sampling began in 1966, and routes have been added over time to a current total of roughly 3000 routes (Pardieck et al. (2019)) We explored trends in abundance, biomass, and energy use over the 30-year time period from 1989-2018. We selected these years to provide a temporal window sufficient to detect trends (Cusser et al. (2020)), while allowing for a substantial number of routes. To avoid irregularities caused by missing time steps, we restricted the main analysis to routes that had been sampled in at least 27 of 30 years in this window (n = 739), and compared these results to a more strict selection of routes that were sampled in every year (n = 199). We take the route to be the "community" scale (Thibault et al. 2011). We filtered the data to remove taxa that are poorly sampled through these methods, following Harris et al. (2018). We accessed the data, and performed this preliminary cleaning and filtering, using the R package `MATSS` (Ye et al. 2020).

Estimated size data

BBS contains abundances for all species along a route in each year, but does not include measurements of individual body size. We generated body size estimates for individual birds assuming that intraspecific size distributions are normally distributed around a species' mean body size (following Thibault et al. (2011)). Using records of species' mean and standard deviation body sizes from Dunning (2008), we drew individuals' body sizes from the appropriate normal distributions. For species for which there was not a standard deviation recorded in Dunning (2008) (185 species affected, of 421 total), we estimated the standard deviation using an allometric scaling relationship between mean and standard deviation in body mass constructed by fitting a linear model to the records in our dataset that did have mean and standard deviation measurements (resulting in the scaling relationship $log(variance) = -5.273 + (log(mass) * 1.995)$); model R2 .86; see also Thibault et al. (2011)). For species with multiple records in Dunning (2008), we used the mean mean and standard deviation body sizes across all records (averaging across sexes, subspecies, and records from different locations). We performed this averaging after estimating any missing standard deviation measurements. For each individual bird observed, we estimated metabolic rate as $10.5 * (mass^{.713})$

(Fristoe 2015, Nagy 2005, McNab 2009). For each route in a given year, we compute total energy use, total biomass, and total abundance by summing over all individuals observed on that route in that year. This method does not incorporate intraspecific variation in body size across geographies or over time (Dunning 2008, Gardner et al. 2011). However, it makes it possible to conduct macroecological studies of avian size distributions at a spatial and temporal scale that would otherwise be impossible (Thibault et al. 2011).

Comparing abundance- and size- based currencies

Comparing trends across different currencies is a nontrivial statistical problem. Because different currencies vary widely in their units of measure (e.g. abundance in the hundreds of individuals; total biomass in the thousands of grams) and , it is challenging to interpret differences in magnitude of slope across different currencies. Transformation and scaling using common approaches (such as a square-root transformation, or rescaling each currency to a mean of 0 and a standard deviation of 1) destroys information about the degree of variability within each currency that is necessary in order to make comparisons *between* currencies for the same timeseries.

Rather than attempting to compare slopes across currencies or to transform different currencies to a common scale, we used a simple null model to compare the observed dynamics for biomass and energy use to the dynamics that would occur in a scenario in which the species composition (and therefore, in this context, size structure) of the community was consistent throughout the timeseries, but in which total abundance varied over time consistent with the observed dynamics. For each route, we characterized the "observed" timeseries of total biomass and total energy use by simulating size measurements for all individuals observed in each time step and summing across individuals, using the method described above. We then simulated timeseries for "abundance-driven" dynamics of biomass and energy use incorporating observed changes in community-wide abundance over time, but under a scenario of consistent species (and therefore approximate size) composition over time. For each community, we characterized the timeseries-wide probability of an individual drawn at random from the community belonging to a particular species ($P(s_i)$) as each species' mean relative abundance taken across all timesteps:

$$P(s_i) = \frac{\sum_t^T \frac{n_{i,t}}{N_t}}{T}$$

where $n_{i,t}$ is the abundance of species i in timestep t, N_t is the total abundance of all species in timestep t, and T is the total number of timesteps. For each timestep t, we randomly assigned species' identities to the total number of individuals of all species observed in that time step (N_t) by drawing with replacement from a multinomial distribution with probabilities weighted according to $P(s)$ for all species. We then simulated body size measurements for individuals, and calculated total energy use and total biomass, following the same procedure as for the observed community. This characterizes the dynamics for size-based currencies expected if the species (and size) composition of the community does not change over time, but incorporating observed fluctuations in total abundance. We refer to these dynamics as "abundance-driven" dynamics.

Long-term trends

For each route, we evaluated the 30-year trend in biomass (or energy use) and compared this to the trend derived from the "abundance-driven" null model using generalized linear models with a Gamma family and log link (appropriate for strictly-positive response variables such as biomass or total metabolic flux). We fit four models to characterize 1) the trend in biomass (or energy use) over time and 2) whether this trend deviates from the trend expected given only changes in abundance:

1. `biomass ~ year * dynamics`, in which "dynamics" refers to being either the "observed" or "abundance-driven" (null model) dynamics. This model fits a slope and intercept for the observed trend in biomass (or energy use) over time, and a separate slope and intercept for the trend drawn from the abundance-driven, or null model, dynamics.

2. `biomass ~ year + dynamics`. This model fits a separate intercept, but not slope, for the abundance-driven and observed dynamics. These models were never selected as the best-performing descriptions of community dynamics.
3. `biomass ~ year`. This model fits a temporal trend, but does not fit separate trends for the observed and abundance-driven dynamics.
4. `biomass ~ 1`. The intercept-only model describes no directional change over time for either the observed or abundance-driven dynamics.

We selected the best-fitting model using AICc. In instances where multiple models had AICc scores within two AICc units of the best-fitting model, we selected the simplest model within two units of the best score.

For each route's selected model, we extracted the predicted values for the first (usually 1988) and last (usually 2018) year sampled, for both the observed and null trajectories. We calculated the magnitude of change over time as the ratio of the last (2018) to the first (1988) value, and characterized the direction of the long-term trend as increasing if this ratio was greater than one, and decreasing if it was less than one.

Relating change in community structure to decoupling between abundance and size-based dynamics

Community dissimilarity metrics are most readily interpretable when making pairwise comparisons (as opposed to repeated comparsions over a timeseries). We compared the first and last five-year intervals in each timeseries, resulting in a "begin" and "end" comparison separated by a relatively consistent window of time across routes (usually 19-20 years). The use of five-year periods corrects for sampling effects (White (2004)), smooths out interannual variability, and, by including a relatively large proportion (1/3) of the total timeseries, partially mitigates the impact of scenarios where the start and end values do not align with the long-term trend.

We calculated three metrics to explore how changes in community composition and size structure translate into decoupling between abundance-driven and observed dynamics for biomass and energy use. First, we evaluated the change in average community-wide body size, calculated as the log ratio of mean body size in the last five years relative to the mean body size in the first five years:

$\ln(\frac{\bar{m}_{last5}}{\bar{m}_{first5}})$

where $\bar{m}_{first5}$ and $\bar{m}_{last5}$ is the mean body size of all individuals observed in the first and last 5 years, respectively. Large changes in average body size are, by definition, expected to translate into decoupling between observed and abundance-driven dynamics.

Second, we calculated measures of turnover in the size structure and in species composition. We calculated turnover in the ISD using a measure inspired by an overlap measure that has previously been applied to species body size distributions in mammalian communities (Read et al. (2018)). We characterized each "begin" or "end" ISD as a smooth probability density function by fitting a Gaussian mixture model (with up to 15 Gaussians; following Thibault et al. (2011)) to the raw distribution of body masses, and extracting the fitted probability density at 1000 evaluation points corresponding to body masses extending beyond the range of body masses present in this dataset (specifically, from 0 to 15 kilograms; mean body masses in this dataset range from 2.65 grams, for the Calliope hummingbird *Selasphorus calliope*, to 8.45 kg, for the California condor *Gymnogyps californianus*). We rescaled each density function such that the total probability density summed to 1. To calculate the degree of turnover between two ISDs, we calculated the area of overlap between the two density smooths as $\sum \min(density1_i, density2_i)$ where $density1_i$ is the probability density from the density smooth for the first ISD at evaluation point i, and $density2_i$ is the probability density from the density smooth for the second ISD at that evaluation point. We subtracted this quantity from 1 to obtain a measure of turnover between two ISDs.

To evaluate turnover in species composition between the five-year time periods, we calculated Bray-Curtis dissimilarity between the two communities using the R package `vegan` (Pinheiro et al. 2020).

We tested whether routes whose dynamics were best-described using different syndromes of change (no trend, couple trends, or decoupled trends) differed in 1) the magnitude of change in mean body size; 2) turnover in the ISD over time; or 3) species compositional turnover over time. For change in mean body size, we fit an ordinary linear model of the form `abs(log ratio (mean body size)) ~ best fitting model type`. We used the absolute log ratio so as to focus on the magnitude, rather than the direction, of change in body size (see also Supp and Ernest (2014) for the use of the absolute log ratio to examine the magnitudes of differences between values). We compared this model to an intercept-only null model of the form `abs(log ratio(mean body size)) ~ 1`. Because our metrics for turnover in the ISD and species composition are bounded from 0-1, we analyzed these metrics using binomial generalized linear models of the form `ISD turnover ~ best fitting model type` and `dissimilarity ~ best fitting model type`, and again compared these models to intercept-only null models. In instances where the model fit with `best fitting model type` outperformed the intercept-only model, we calculated model estimates and contrasts using the R package `emmeans` (Lenth 2021).

Results

This analysis revealed qualitatively different continent-level patterns in the long-term trends for biomass, energy use, and total abundance (Figure 2). Of the 739 routes in this analysis, approximately 70% (500/739 for biomass, and 509/739 for energy use) were best-described using a model incorporating a temporal trend in abundance and/or biomass or energy use (Table 1). Trends driven by abundance, as reflected by the dynamics of a simple null model, were strongly dominated by declines (335 decreases and 165 increases for abundance-driven dynamics in biomass, and 355 decreases and 154 increases for abundance-driven dynamics in energy use; Figure 2; Table 2). In contrast, for biomass, the long-term temporal trends were evenly balanced between increases and decreases (256 decreasing trends, and 244 increasing trends; Figure 2; Table 2). For energy use, there was a greater representation of decreasing trends than for biomass, but still less so than for strictly abundance-driven dynamics (329 decreasing trends and 180 increasing trends; Figure 2; Table 2).

These divergent aggregate outcomes in abundance, energy use, and especially biomass occurred due to decoupling in the long-term trends for these different currencies. For a substantial minority of routes (20% of all routes for biomass, and 7% of all routes for energy use), the best-fitting model fit a different long-term trend for biomass or energy use than for the "null", abundance-driven, trend (Table 1). When this decoupling occurred, it was dominated by scenarios in which the slope for abundance-driven dynamics was more negative than that for biomass or energy use (Figure 3).

Decoupling between the long-term trajectories of abundance and energy use or biomass is, by definition, indicative of some degree of change in the ISD over time. Routes whose dynamics were best-described as decoupled trends over time had a higher absolute log ratio of mean mass (i.e. greater magnitude of change, either increasing or decreasing, in mean mass over time) than routes with coupled or no trends (Tables 3-5). However, there was not a detectable difference in the degree of temporal turnover in the ISD overall (Table 6), or in species composition (Table 7), compared between routes that exhibited different dynamics (i.e. no linear temporal trend, a consistent temporal trend for abundance and biomass or energy use, or differing trends for abundance and biomass or energy use) .

Discussion

Abundance, biomass, and energy use are nonequivalent currencies

Simultaneously examining multiple currencies of community-level abundance revealed qualitatively different continent-wide patterns in the long-term trends for abundance in terms of individuals, biomass, and energy use. While long-term trends in individual abundance were dominated by decreases, long-term trends in biomass were evenly split between increases and decreases, and trends in energy use were again dominated by declines (Figure 2). These different currencies, though intrinsically linked, describe nonequivalent dimensions of community function and reflect different classes of structuring processes (Morlon et al. 2009).

Abundance, in terms of individuals, is most directly linked to species-level population dynamics of the type often considered in classic, particularly theoretical, approaches to studying competition, compensation, and coexistence (e.g. Hubbell (2001); Chesson (2000)). Biomass most directly reflects the productivity of a community and its potential contributions to materials fluxes in the broader ecosystem context, whereas energy use - by taking into account the metabolic inefficiencies of organisms of different body size - characterizes the total *resource use* of a community and may come the closest to capturing signals of bottom-up constraints, "Red Queen" effects, or zero-sum competitive dynamics (Van Valen 1973, Ernest et al. 2009, 2008, Morlon et al. 2009, White et al. 2004). Our results underscore that, while trends in abundance, biomass, and energy use naturally co-vary to some extent, shifts in the community size structure can and do produce qualitatively different trends for these different currencies. These may reflect contrasting long-term changes in different types of community processes - for example, shifts in habitat structure that affect the optimal body sizes for organisms in a system, but do not result in overall changes in resource availability (e.g. White et al. (2004)). Moreover, extrapolating the long-term trend from one currency to another may elide underlying changes in the community that complicate these dynamics. To appropriately monitor different dimensions of biodiversity change, it is therefore important to focus on the specific currency most closely aligned with the types of processes and dynamics - e.g. population fluctuations, resource limitation, or materials fluxes - of interest in a particular context.

For North American breeding birds, biomass has declined less than abundance or energy use

For communities with a decoupling in the long-term trends of biomass, energy use, and abundance, this decoupling is indicative of a directional shift in the size structure of the community. For the communities of breeding birds across North America considered here, the long-term trends in total biomass are often less negative than trends in total abundance or total energy use (Figure 3). This consistent (but not ubiquitous) signal corresponds to community-level increases in average body size that partially or completely buffer changes in total against declines in abundance. This contrasts with general, global concerns that larger-bodied species are more vulnerable to extinction and population declines than smaller ones (**???**, Young et al. 2016, Dirzo et al. 2014, Smith et al. 2018). However, it is consistent with previous findings from the Breeding Bird Survey (Schipper et al. 2016). Increases in body size may reflect forests in recovery across North America over this timespan (Schipper et al. 2016), or the contributions of relatively few, large-bodied species that may in fact benefit from recent ecological changes (**???**). The long-term trends for communities of different taxonomic groups, geographies, or temporal spans may show different effects related to different facets of global change and biodiversity responses.

We note that these increases in body size do not generally appear great enough to decouple the long-term trends in *energy use* from total abundance (Figure 3). Energy use scales nonlinearly with body size with an exponent less than 1, which means that community-wide increases in mean body size result in smaller increases in total energy use than in total biomass.

Complex relationships between compositional change and community-level properties

The decoupling between the long-term trends for biomass, abundance, and energy use demonstrated in many of the communities studied here is symptomatic of a directional shift in the size structure - in these instances, generally favoring larger bodied species. However, examining the community-wide dynamics of turnover in species composition and the overall size structure reveals that the relationship between changes in community structure and changes in the scaling between different currencies of community-wide abundance is considerably more nuanced than simple directional shifts in mean size. Routes that exhibit a statistically detectable decoupling between total biomass and total abundance show large changes in average body size compared to routes for which biomass and abundance either change more nearly in concert with each other or do not show temporal trends (Figure 4). This aligns naturally with mathematical intuition given the intrinsic relationship between average body size, total abundance, and total biomass. However, these routes are *not* extraordinary in terms of their overall degree of temporal turnover in either the size structure or in species composition. Rather, the levels of turnover in overall community structure are comparable between

routes that show decoupling between abundance and biomass, statistically indistinguishable trends, or no temporal trends in either currency (Figure 4).

For many communities, therefore, there has been appreciable change in the species and size composition that simply does not manifest in a shift in the overall community-wide mean body size or mean metabolic rate sufficient to decouple the dynamics of biomass, abundance, and energy use. These changes may signal changes in functional composition equally important as the ones that manifest in directional shifts in community-wide average body size. For the complex, multimodal size distributions that are the norm for avian communities (Thibault et al. 2011), changes in the number and position of modes may be as important as changes in higher-level statistical moments such as the overall mean. At present, the field lacks the statistical tools and conceptual frameworks to quantify and interpret these nuanced changes, especially at the macroecological scale of the current study (Thibault et al. 2011, Yen et al. 2017). However, this is an excellent opportunity for more system-specific work, informed by natural history knowledge and process-driven expectation, to characterize more nuanced changes in the size structure of specific communities and identify the underlying drivers of these changes. To facilitate these efforts in the context of the Breeding Bird Survey, the R package we have developed to characterize the individual size distributions for avian communities based on species' identities and/or mean body sizes is freely available for re-use and wider applications (*cite dissertation version here - I am working on a more general-use version to publish more widely, e.g. in JOSS, but not by March!*).

Conclusion

This analysis demonstrates the current power, and limitations, of a data-driven macroecological perspective on the interrelated dynamics of community size structure and different dimensions of community-wide abundance for terrestrial animal communities. For breeding bird communities across North America, we find that changes in species and size composition produce qualitatively different aggregate patterns in the long-term trends of abundance, biomass, and energy use, highlighting the nuanced relationship between these related, but decidedly nonequivalent, currencies and reflecting widespread changes in community size structure that may signal substantive changes in functional composition. Simultaneously, the complex relationship between turnover in community species and size composition, and the scaling between different currencies of community-level abundance, highlights opportunities for synergies between recent computational and statistical advances, case studies grounded in empiricism and natural history, and future macroecological-scale synthesis to realize the full potential of this conceptual space.

Chapter 4: Empirical abundance distributions are more uneven than expected given their statistical baseline

Published in Ecology Letters as Diaz et al. (2021): Empirical abundance distributions are more uneven than expected given their statistical baseline

Background

Ecological communities are complex systems made of numerous interacting entities subject to a vast array of processes operating in different contexts and at different scales (Levin 1992; Lawton 1999; Maurer 1999; Brown et al 2002; Nekola and Brown 2007; McGill 2019). One strategy for making sense of this inherent complexity is to identify patterns that occur consistently across many communities, and use these common phenomena to develop and test theories regarding general mechanisms that shape community structure (Brown and Maurer 1989; Maurer 1999; Lawton 1999; Gaston and Blackburn 2000; McGill 2019). Some of these patterns, however, can have counterintuitive emergent statistical properties (Frank 2009; 2019). Left unexamined, these properties can confound the interpretation of the observed patterns: what we interpret to be the result of generative mechanism may be an artifact of statistical constraints. However, when these properties are properly understood and accounted for, they can provide leverage for detecting and identifying the processes at work in a system (Jaynes 1957, Harte and Newman 2014).

The species abundance distribution (SAD) – the distribution of how all of the individuals in a community are divided among the species in that community – is a prime example of an ecological pattern that is both commonly invoked in the search for general processes, and subject to statistical constraints that have thus far complicated efforts to use it in this way (Nekola and Brown 2007; McGill et al. 2007; Locey and White 2013). The shape of the SAD is so consistent that it is often considered an ecological law (Preston 1948, 1962a, 1962b, 1980; Lawton 1999, McGill 2003, McGill et al. 2007). Across varied ecosystems and taxa, the species abundance distribution is dominated by a few very abundant species and a larger number of increasingly rare species, generating a distinctive hollow- or J-shaped curve when plotted with species rank on the x-axis and abundance on the y-axis (Fisher et al. 1943; McGill et al 2007). Community ecologists have used the SAD to test numerous theories regarding which biological processes are most important for structuring assemblages of species, by comparing theoretical predictions for the SAD to observed SADs (McGill 2003; McGill et al. 2007). However, this approach has proven inconclusive because many theories predict similar shapes for the SAD (McGill 2003; McGill et al. 2007), and even experimental manipulations generate little variation in the shape of the SAD (Supp and Ernest 2014). Investigating and accounting for the statistical considerations that constrain the shape of the SAD may open up new avenues for ecological interpretations of the SAD.

In fact, the nearly ubiquitous shape of the SAD may transcend ecological processes and instead reflect mathematical properties inherent to abundance distributions. Complex systems across domains ranging from economics to information technology often exhibit empirical abundance distributions with hollow-curve forms similar to ecological SADs (Shockley 1957; Gaston et al. 1993; Nekola and Brown 2007, Blonder et al. 2014; Keil et al. 2018). This suggests that the hollow curve is a common feature of abundance distributions and not necessarily an ecological phenomenon. Because the hollow-curve is observed in diverse systems and many theoretical generative processes converge to power-law or log-series abundance distributions (i.e. hollow curves) (Preston 1950; McGill 2003; Nekola and Brown 2007; Frank 2009; Frank 2019), approaches from statistical mechanics and complexity science may best explain the expected emergent shape for the distribution (Preston 1950; McGill 2003; Nekola and Brown 2007; Dewar and Porté 2008). Indeed, frameworks grounded in both entropy maximization (e.g. the Maximum Entropy Theory of Ecology; Harte et al. 2008, Harte 2011) and combinatorics (i.e. 'the feasible set'; Locey and White 2013) generate realistic hollow curves via the random division of the total number of individuals in a community, N, into the total number of species present S. If the SAD is statistically inclined to be a hollow curve, the hollow-curve in itself may be of limited use for developing and testing ecological theories.

While SADs may be statistically constrained, this does not necessarily mean that they cannot be biologically

informative. Biological factors may introduce subtle, but meaningful, deviations between observed SADs and the shapes of the SADs expected due to the mathematical constraints imposed by S and N, which we hereafter refer to as the "statistical baseline" (Locey and White 2013, Harte and Newman 2014). If the vast majority of mathematically achievable SADs for a community share a similar shape, an empirically observed SAD that deviates even slightly from this statistical baseline is unlikely to have emerged at random (Locey and White 2013), and may be the signature of a non-random – i.e., biological – process operating on the relative abundances of species (Harte and Newman 2014). If, over many communities, there are consistent deviations between observed SADs and their statistical baselines, these deviations can help evaluate and refine ecological theories. For example, the high prevalence of rare species in ecological communities has attracted considerable empirical and theoretical attention (e.g. Nee et al. 1991; Magurran and Henderson 2003), but it is unclear to what extent this phenomenon may derive from mathematical constraints on the SAD rather than ecological processes. If the prevalence of rare species in observed distributions consistently exceeds what would be expected to emerge from the statistical baseline, we would be prompted to look for ecological mechanisms promoting rarity. Candidate theories could then be evaluated based on how well their predictions for the rare tail of the SAD matched observed distributions. Thus, the deviations from the statistical baseline may enable us to detect strong ecological processes or evaluate theories (Harte and Newman 2014, Xiao et al. 2016).

Successfully interpreting SADs in this fashion depends on our capacity to detect and quantify deviations between empirical observations and statistical baselines, which requires metrics and computational approaches that allow us to quantify and interpret whatever deviations may exist. Here, we build upon the combinatoric approach developed by Locey and White (2013) to define and explore statistical baselines for SADs. For a given N (total number of individuals) and S (total number of species), there exists a finite set of possible distributions of individuals into species. Collectively, this set of possible SADs is the feasible set, with each possible SAD constituting a single element of the set. If an observed SAD is drawn at random from the feasible set, it is likely to have a shape similar to the shapes most common in the feasible set. The feasible set therefore allows us to define statistical baselines for assessing deviations between observed SADs and what is likely to occur due to mathematical constraints (Locey and White 2013).

The feasible set can also be used to explore how the characteristics of the statistical baseline, and the presence and nature of any deviations that occur, vary over ranges of values for S and N. Although most feasible sets are dominated by the hollow-curve shape, variation in S, N, and the ratio of N to S modulate the detailed attributes of the SADs in a feasible set (Locey and White 2013). For example, if the ratio of N to S is close to 1, all possible SADs are mathematically constrained to be fairly even (Locey and White 2013). Although an SAD that is very even would be highly unusual in most cases, it would be expected in this situation. The feasible set therefore allows us to appropriately calibrate our expectations for what types of observations would be surprising for an SAD given the specific constraints imposed by its S and N. Additionally, accounting for variation in the specificity, or vagueness, of the expectations derived from the statistical baseline may be critically important for disentangling the aspects of the SAD that can be attributed to statistical constraints from those that result from other processes. If the vast majority of mathematically possible SADs are similar in shape – generating a very specific, narrowly defined statistical baseline – then even small deviations between an observed SAD and this baseline can signal the operation of ecological processes. However, if many different shapes occur with more even frequency in the feasible set, the statistical baseline is less specific and less well defined, and our sensitivity for distinguishing biological signal from statistical constraints is greatly reduced. This is more likely to occur when the size of the community, in terms of S and N, is small, because in such cases the feasible set may be too small for a particular shape to emerge as the most common shape. These statistical baselines with broad distributions may therefore impede our ability to assess whether observed deviations are ecologically generated or expected to emerge randomly (Jaynes 1957). This general concern has been acknowledged in efforts to compare ecological observations to statistical baselines (Harte 2011, White et al. 2012, Locey and White 2013) but there has not yet been a quantification of these effects for the SAD or an identification of the range of community sizes most strongly affected. Because ecologists study the SAD for communities varying in size from the very small – S and N < 5 – to the enormous – S and N » 1000 – identifying the community sizes for which we can and cannot confidently detect deviations from the statistical baseline is necessary to appropriately contextualize our interpretations.

Here we use the feasible set to define statistical baselines for empirical SADs for 22,000 communities of birds, mammals, trees, and miscellaneous other taxa. We then compare observed SADs to their corresponding statistical baselines and evaluate 1) if the shapes of observed SADs consistently deviate from their statistical baseline, 2) how the characteristics and specificity of the statistical baseline vary over ranges of S and N, and 3) whether this variation appears to be associated with variation in our capacity to detect deviations between observations and the corresponding baselines.

Methods

Data and code for all of our analyses can be accessed at http://doi.org/10.5281/zenodo.4711104.

Datasets

We used a compilation of community abundance data for trees, birds, mammals, and miscellaneous additional taxa (White et al. 2012, Baldridge 2015, Baldridge 2016, data from Baldridge 2016). This compilation consists of cleaned and summarized community abundance data for trees obtained from the Forest Inventory and Analysis (Woudenberg et al 2010) and Gentry transects (Phillipes and Miller 2002), birds from the North American Breeding Bird Survey (Sauer et al. 2013), mammals from the Mammal Community Abundance Database (Thibault et al. 2011), and a variety of less commonly sampled taxa from the Miscellaneous Abundance Database (Baldridge 2015). Because characterizing the random expectation of the SAD is computationally intractable for very large communities, we filtered our datasets to remove 4 communities that had more than 40714 individuals, which was the largest community we successfully analyzed. We further filtered the FIA database. Of the 103,343 communities in FIA, 92,988 have fewer than 10 species. Rather than analyze all these small communities, we randomly selected 10,000 small communities to include in the analysis. We also included all FIA communities with more than 10 species, which added 10,355 FIA communities to the analysis and resulted in a total of 20,355 FIA communities. Finally, for sites that had repeated sampling over time, we followed White et al. (2012) and Baldridge (2016) and analyzed only a single, randomly selected, year of data, because samples taken from a single community at different time points are likely to covary. It should be noted that our analyses include data from the Mammal Community Database and Miscellaneous Abundance Database that were collected over longer timescales and cannot be disaggregated into finer units of time. Our final dataset consisted of ~22,000 communities with S and N ranging from 2 to 250 and 4 to 40714, respectively (Figure 1). Details and code for the filtering process can be found in Appendix S1 in Supporting Information.

Accounting for empirical sampling error

Because it is logistically impossible to exhaustively census all individuals present in most empirical systems, SADs derived from field sampling will inevitably be subject to some degree of sampling error (Bonar et al. 2011). Therefore, in addition to analyzing the raw SADs in our database, we employed two resampling schemes to test if, and how, different forms of observation error affect our results. First, we explored the possibility that empirical sampling systematically undercounts the true number of rare species in a community (Preston 1948; Gotelli and Colwell 2011). Rare species are more likely to escape detection during sampling, leading to an underestimate of both the total species richness of a community and the proportion of species in the rare tail of the SAD (Preston 1948). We used a procedure based on species richness estimators to adjust for this possibility (see also Ulrich et al. 2010 for the use of richness estimators to distinguish between completely and incompletely censused communities). We computed the estimated richness for each community using the bias-corrected Chao and the ACE estimators (as implemented in the R package "vegan"; O'Hara 2005; Chiu et al 2014; Oksanen et al. 2020). To each of these richness estimates, we added one standard deviation of the estimate, and then took the mean of the two results. This yields a generous estimate of the true number of species in the system. If this estimate exceeded the observed species richness, we added the missing species each with abundance 1. These adjusted SADs allowed us to explore the consequences of undersampling rare species while making the smallest possible changes to S and N.

Second, we tested the sensitivity of our results to sampling variability across all species in the SAD – not just rare species - using subsampling. For each observed community, we constructed subsamples by randomly drawing 60% of the observed number of individuals from the total pool of individuals in the community, without regard to species and without replacement. The precise proportion of individuals drawn in each subsample should not dramatically affect the qualitative outcome. We selected 60% so as to introduce appreciable room for sampling error between the raw and subsampled SADs, but to produce subsampled SADs with N (and presumably S) in a comparable size range to the raw ones. Extremely small subsamples (e.g. 10%) could introduce complications related to small N and S that could obscure the effects of sampling error, while very large subsamples (e.g. 90%) could recapture the raw distributions too closely to be informative. We generated 10 resampled communities for each observed community. We ran our computational pipeline using all raw SADs and all SADs adjusted for undersampling of rare species. Because the subsampling approach increased computational effort approximately tenfold, we analyzed all subsampled communities for the Mammal Community, Miscellaneous Abundance, and Gentry databases, but only a random subset of 300 (of 2773) communities from the Breeding Bird Survey and 2000 (of 20179) from the FIA – 1,000 with S < 10, and 1,000 with S >= 10.

Generating the statistical baseline

We use the concept of the "feasible set" to establish a statistical baseline for the SAD (Locey and White 2013). For a given number of individuals N, there are a finite number of unique ways to partition those individuals into S species. The complete set of these unique partitions is the feasible set. In this approach, neither species nor individuals are distinguishable from each other; thus partitions are unique if and only if they differ in the number of species that have a particular abundance (Locey and White, 2013). Operationally, this means that for S = 3 and N = 9, the SADs (1, 3, 5) and (2, 2, 5) count as distinct partitions, but (1, 3, 5) and (3, 1, 5) do not, because they each contain one species with an abundance 1, 3, and 5, respectively, and differ only in the order of the numbers. In the absence of justification for additional assumptions regarding the distinguishability of species and/or individuals, we adopted this simple set of assumptions that has previously been shown to generate realistic statistical baselines (Locey and White 2013).

While it is possible to list all possible partitions in the feasible set for small S and N, the size of the feasible set increases rapidly with S and N. An exhaustive characterization of the statistical properties of the feasible set for large S and N quickly becomes computationally intractable. This renders it necessary to draw samples from the feasible set, rather than enumerating all of its elements. Previous efforts in this vein (Locey and White 2013) have been constrained by the problem of unbiased sampling of large feasible sets. We developed an algorithm to efficiently and uniformly sample feasible sets even for large values of S and N. In brief, the algorithm takes a generative approach to sample the feasible set for a given combination of S and N, based on recurrence relations used to calculate the size of the feasible set. Let f(S, N) be the number of possible partitions of N individuals into exactly S species, i.e. the size of the feasible set for given values of S and N. Computation of f(S, N) can be achieved without enumerating the entire feasible set through the recurrence relation f(S, N) = f(S-1, N-1) + f(S, N-S) (originally documented in a 1742 letter from Euler to Bernoulli; 1862). For example, consider the feasible set with S = 3 and N = 7. For all possible partitions, either (a) at least one species has an abundance equal to 1, or (b) all of the species have abundance greater than 1. In the case of (a), removing one species with abundance equal to 1 must result in a partition of 6 individuals into 2 species. In fact, all of the unique partitions in (a) must have a corresponding unique partition in the feasible set for S = 2 and N = 6, and vice versa. In the case of (b), removing 1 individual from each species must result in a partition from the feasible set with S = 3 and N = 4. Here, all the partitions in (b) must have a corresponding unique partition in the feasible set with S = 3 and N = 4, and vice-versa. Therefore, f(3,7) = f(2,6) + f(3,4). By storing the values in a lookup table, f(S, N) can be calculated for increasing values of S and N through straightforward summation.

This recurrence relation also makes it possible to draw random samples from the feasible set without enumerating all possible partitions of N into S. For the example of S = 3 and N = 7, there are a total of 4 possible partitions (i.e. f(S, N) = 4). Because f(2, 6) = 3 and f(3, 4) = 1, we know that (a) 3 of the 4 partitions must correspond to a partition of the feasible set with S = 2 and N = 6 (but with a species of abundance equal

to 1 removed), and (b) 1 of the 4 partitions must correspond to a partition of the feasible set with S = 3 and N = 4 (but with 1 individual removed from each species). Thus, we can determine the probability that a partition drawn at random from the feasible set for S = 3 and N = 4 is in case (a) – probability ¾ - or case (b) – probability ¼. To generate a partition in case (a), we sample a partition for S = 2 and N = 6 and then add a species with abundance equal to 1; for case (b), we sample a partition for S = 3 and N = 4 and then add 1 individual to each species. In this way, we use the recurrence relation to transform the problem of sampling from a large feasible set into the problem of sampling from a smaller, different feasible set. This procedure continues until a partition is uniquely determined, after which some back-transformation yields a unique partition for the feasible set of interest. A detailed description of the algorithm we use, based on a slightly different recurrence relation, is available in Appendix S2 and is implemented in the R package feasiblesads available at http://doi.org/10.5281/zenodo.4710750.

For every community in our database, we drew 4000 samples from the feasible set to characterize the distribution of statistically probable shapes for the SAD. We filtered the 4000 samples to unique elements. For small values of S and N, it can be impossible or highly improbable for the 4000 samples from the feasible set to all be unique, but for large communities, all 4000 are usually unique. We refer to this as the sampled feasible set.

Comparing observed SADs to their statistical baselines

We compared SADs to their statistical baselines using several metrics, including a general measure of dissimilarity, as well as skewness, Simpson's evenness, Shannon's index, and the proportion of rare species (species with abundance = 1). These metrics represent just a few of the vast array of possible summary metrics to describe the shape of the SAD, each of which emphasize different aspects of the distribution. In this first effort to compare empirical distributions to a statistical baseline, we selected a suite of complementary metrics and explored whether our overall results were consistent between metrics. By calculating these metrics for each the community's sampled feasible set (see Generating the statistical baseline, above), we generated a portfolio of measures describing the shapes expected from randomly sampled SADs.

First, as a general characterization of whether observed SADs have rare or common shapes relative to their feasible sets, we computed a dissimilarity score comparing SADs to the central tendencies of their feasible sets (following Locey and White, 2013). We defined the degree of dissimilarity between two SADs with the same S and N as the proportion of individuals allocated to species with different abundances between the two SADs, calculated as:

where n1i is the abundance at rank i for one SAD and n2i is the abundance at rank i for the other SAD. This value ranges from 0 to 1, with 1 being high dissimilarity. To find the central tendency of a given sampled feasible set, we identified the sampled SAD with the lowest mean dissimilarity compared to the rest of the SADs in the feasible set. We calculated the dissimilarity between every sample drawn from the feasible set and a random set of 500 other samples, using a subset of samples for comparisons because it is computationally impractical to make all pairwise comparisons between large numbers of samples. To assess whether an observed SAD was highly dissimilar to its central tendency, we calculated the degree of dissimilarity between the central tendency of the corresponding feasible set and all other samples from that feasible set, and between the central tendency and the observed SAD. Although the dissimilarity score is scaled from 0 to 1, the distributions of dissimilarity scores for samples from the feasible set can vary over broad ranges in S and N. We therefore used the percentile rank of the observed dissimilarity scores, relative to the distribution of dissimilarity scores from the corresponding sampled feasible sets, to quantify how likely or unlikely observed dissimilarity scores are across the range of S and N in our datasets. For a single community, an observed percentile score of 95 indicates that there is a 5% chance of drawing a value greater than the observed value from the distribution of values from the sampled feasible set. Aggregating across communities, if observed SADs reflect random draws from their feasible sets, their percentile rank values should be uniformly distributed from 0 to 100. However, if observed SADs are consistently more dissimilar to their feasible sets that expected at random, the percentile values will be disproportionately concentrated at high values. We used a one-tailed 95 confidence interval and tested whether the percentile values for the dissimilarity scores of observed SADs fell above 95 more than 5% of the time. We note that it is impossible for an observation fall

above the 95th percentile if there are fewer than 20 values in the sampled distribution. We therefore excluded from this analysis communities with fewer than 20 unique SADs in their feasible sets, yielding a total of 22,490 communities. Finally, note that, if the observed dissimilarity scores for individual communities are not systematically higher than the distributions of dissimilarity scores from the corresponding feasible sets, increasing the number of communities in the analysis will not increase the frequency of extreme percentile scores. While the degree of dissimilarity between SADs and the central tendency of the feasible set provides an overall sense of deviations among possible SADs, it does not describe how observed SADs may differ from their feasible set. We therefore used a set of more targeted, ecologically interpretable metrics to explore how observed SADs compare to their feasible sets in their shape and proportion of rare species. We examined three metrics for the shape of the SAD - skewness, Simpson's evenness (1-D), and Shannon's index. Skewness measures the asymmetry of a distribution around its mean. The Simpson and Shannon indices are commonly used metrics for assessing how equitably abundance is distributed across species (Maurer and McGill 2011). We also calculated the proportion of rare species (species with abundance = 1) in each SAD, because the proportion of rare species in a community is comparable across different community sizes and is of special interest to ecologists.

As with the degree of dissimilarity score, to assess whether the shape of an observed SAD was statistically unlikely, we used percentile ranks to compare the observed values of the summary metrics to the distributions of values for those metrics obtained from each community's sampled feasible set. The actual ranges and values of summary metrics vary widely over the ranges of S and N in our data and thus cannot directly compared, but percentile ranks are comparable across different community sizes and allow assessment across our entire dataset. We used two-tailed 95% intervals to test whether observed communities' percentile values for each metric were disproportionately concentrated below 2.5 or above 97.5. In all cases, in testing for unusually high percentile scores, we defined the percentile score as the proportion of values in the sampled distribution strictly less than the observed value, while in testing for low values, we defined it as the proportion of sampled values less than or equal to the observed value. This ensured a conservative estimate of how extreme the observed values were relative to the sampled distribution. Because it is impossible for an observed percentile score to be above or below the 97.5th or 2.5th percentile if there are fewer than 40 values in the sample distribution, we excluded from these analyses communities with fewer than 40 SADs in their feasible sets. Finally, note that skewness, as implemented in the R package "e1071" (Meyer et al. 2019), always evaluates to 0 for distributions with only two species, and we therefore excluded those cases from analyses of skewness. Our final analysis included 21,395 communities for skewness and 21,403 communities for all other shape metrics.

The narrowness of the expectation

We also used the distributions of dissimilarity scores and shape metrics to quantify the relative specificity of the statistical baseline, in order to assess when there could be challenges in determining whether observed communities differ from their statistical baselines. For an overall sense of how tightly elements of the feasible set were clustered around its central tendency, we calculated the 95th percentile of dissimilarity scores for all samples from a feasible set compared to the central tendency of that feasible set. This value shows how dissimilar to the central tendency an observed SAD would need to be to count as "extreme" relative to the feasible set. For the shape metrics, we calculated a breadth index defined as the ratio of the range of values encompassed within a two-sided 95% density interval relative to the full range of values in the distribution (Figure 2). This breadth index for the statistical baseline ranges from 0 (a very narrow distribution and well-resolved baseline) to 1 (a very broad distribution), and is comparable across feasible sets for varying combinations of S and N. These approaches correspond qualitatively to more computationally-intensive approaches to measuring the self-similarity of the elements of feasible sets (see Appendix S3). We explored how the narrowness of the statistical baseline varies with the size of the feasible set and the ratio of N to S.

Results

Comparing observed SADs to their statistical baselines

For four of the five datasets we analyzed – BBS, Gentry, Mammal Communities, and Miscellaneous Abundance – observed SADs are more dissimilar to their statistical baselines than would be expected by chance (Figure 3). Combined over these four datasets, 29% of observed SADs are more dissimilar to the central tendency than are 95% of samples from the corresponding feasible sets (Table 1). If observed SADs reflected random draws from the feasible set, we would expect only 5% to be that dissimilar. These highly unlikely SADs have dissimilarity scores from 1.007 to 6.3 times the 95th percentile of dissimilarity scores comparing the central tendency to samples from the feasible set, an absolute increase ranging from .001 to .54 on a scale from 0-1 (Figure S4). These datasets also contain highly unlikely observed SADs in terms of their shape metrics. At random, roughly 2.5% of observed percentile scores for these metrics should be very high (>97.5) or very low (<2.5). Compared to their feasible sets, these four datasets contain a disproportionate number of communities with very low values for Simpson's evenness and Shannon diversity, and very high skewness, relative to their feasible sets (Table 1). The Mammal Community and Miscellaneous Abundance databases also have high proportions of rare species, but this tendency is weaker for BBS and nonexistent for Gentry – in fact, the Gentry dataset has a high representation of sites with low proportions of rare species (20% of sites; Table S5). The Gentry dataset also has a disproportionate number of communities with the opposite tendencies to the other datasets for the other shape metrics– i.e., an overrepresentation of communities with high Simpson's evenness and Shannon diversity, and low skewness.

In contrast to the other datasets, percentile scores for sites from the FIA dataset are more uniformly distributed, and the proportions of extreme values are closer to what would be expected by chance (Figure 3, Table 1). Only 7% of FIA communities are highly dissimilar to their feasible sets (compared to a random expectation of 5%). Among the shape metrics, only 2.8% (compared to 2.5% at random) of sites have high values for skewness, 1.3% have high proportions of rare species, 5.7% have low Simpson's evenness, and 5.4% have low Shannon diversity.

The narrowness of the expectation

The ability to detect deviations from the statistical baseline depends in part on the distribution of SADs in the feasible set. Overall, as the size of the feasible set increases, the SADs in a feasible set become more narrowly clustered around the central tendency of that feasible set, and the sampled distributions for shape metrics generally become less variable (Figure 4). In small communities, the breadth indices are highly variable and often very large – approaching 1, meaning that a 95% density interval of the values in the distribution spans nearly the entire range of values – while the breadth indices for larger communities rarely exceed ~.7 for skewness, Simpson evenness, and Shannon diversity, and ~.8 for the proportion of rare species. Among our datasets, the FIA and Mammal Community databases have the smallest communities, in terms of S and N, and tend to have the largest proportions of feasible sets with high breadth indices (Figure S6).

Sensitivity to sampling variability

In almost all cases, SADs adjusted for the under-observation of rare species are even more extreme relative to their feasible sets than unadjusted SADs (Figure 5; see Appendix A7 for complete results of resampling). For all datasets, adjusted SADs show more high values for skewness and the proportion of rare species, and low values for Simpson's evenness and Shannon diversity, than unadjusted SADs. Subsampling consistently reduces the proportion of extreme observations across all datasets and metrics (Figure 5; Appendix A7). In most instances, the proportion of extreme observations still exceeds the proportion that would be expected by chance. However, the proportion of sites with high numbers of rare species observed for the BBS and Mammal Community databases drop from 4.5% to 1% and ~13% to 3.5% with resampling. For FIA, the proportions of sites with high dissimilarity, low evenness and Shannon diversity all drop from 6-8% to 2-3%. Note that, for FIA, neither the raw nor the resampled SADs have a disproportionate representation of extreme values for the remaining metrics.

Discussion

We found widespread evidence that SADs for a range of real ecological communities deviate from the forms expected given the distribution of shapes within their feasible sets. Overall, these deviations may signal that ecological processes operate on top of statistical constraints, thereby driving the SAD away from shapes generated by purely statistical processes. We also found that the magnitude and form of deviation varied among the datasets we considered. This variability may reflect statistical phenomena related to the size of S and N and their ratio, or it may reflect different biological processes dominating in different contexts. Finally, although a disproportionate number of communities deviated statistically from their feasible sets, there were also many communities for which we did not detect deviations. This does not imply the absence of ecological processes operating on these SADs. Rather, one possible explanation is that multiple ecological processes are operating simultaneously and with countervailing effects, resulting in no dominating net impact on the shape of the distribution beyond that imposed by fundamental constraints (Harte 2011; Harte and Newman 2014). Going forward, testing whether ecological theories or common functional approximations (e.g. the log-normal distribution) accurately predict the deviations between observed SADs and their statistical baselines may be much more fruitful than focusing only on the general form of the SAD (McGill et al. 2007; Locey and White 2013; Harte and Newman 2014).

In most cases, and most pronouncedly for the Breeding Bird Survey, Mammal Community, and Miscellaneous Abundance databases, our results suggest that the prevailing processes cause abundance distributions to be highly uneven, rather than those that produce more even abundances across species. For these communities, observed SADs tended to be unusually skewed and uneven, and to have a high proportion of rare species, compared to their feasible sets. Accounting for undersampling of rare species strengthened these effects, while subsampling weakened them. Perhaps unsurprisingly, the effect of these two resampling approaches was especially noticeable for the proportion of rare species; enriching the SAD directly adds rare species, while subsampling is likely to drop rare species even if it otherwise recaptures the general shape of a distribution. The long tail of rare species in the SAD has been a consistent focus in SAD research, and our results highlight that the rare tails of observed SADs are extraordinary, even among the hollow-curve shapes that dominate the feasible set. Ecological processes may lengthen the rare tail and decrease the evenness of the SAD, for example by promoting the persistence of rare species at very low abundances (Yenni et al. 2012). Or, they could drive abundant species to have larger populations than would be statistically expected, without also driving other species entirely to extinction (Chesson 2000). While the Gentry database also exhibits deviations tending towards high unevenness, an even greater proportion of its communities are more even, and have a lower proportion of rare species, than would be expected given their feasible sets. This could indicate that there are biological differences between the systems in the Gentry and other datasets that result in different forms for the SAD. Alternatively, the statistical characteristics of the feasible sets for these communities could modulate the detected deviations. Communities in the Gentry database have high species richness and low average abundance (Figure 1). Among these, many of the communities exhibiting high evenness and low proportions of rare species are those with very high species richness and low average abundance (N/S < ~3) (see Appendix A8). As a result, these communities have unusual statistical baselines: for example, the corresponding feasible sets have the highest proportions of rare species of any of the feasible sets in our analysis. Although observed SADs for these communities also have high proportions of rare species, taking the statistical baseline into account would suggest that the extraordinary thing about these SADs is that they do not have even more rare species. Simultaneously, there may be biological reasons why the species-rich but relatively low-abundance tropical tree communities of the Gentry database differ from those in other datasets. The same mechanisms that promote high diversity may manifest in high evenness, and/or ecological features particular to these forests may produce unusual shapes for the SAD. Because no communities from our other datasets are comparable in S and N, we cannot disentangle these statistical and biological explanations. This is an excellent opportunity to develop additional theoretical and empirical approaches to predict and explain variation in the deviations between SADs and their feasible sets, in particular for species-rich communities across ecosystems.

Unlike the other four datasets, communities in the FIA dataset showed weak or no evidence of deviations from their feasible sets. We entertained two general classes of explanation for why the FIA dataset differs from the others in our analysis: first, that biological attributes of the FIA communities cause the SADs for these

communities to differ from the others in our database, and second, that statistical phenomena related to S and N may modulate the capacity to detect deviations for these communities. To distinguish between possible biological drivers causing FIA to differ from the other datasets, and factors intrinsic to S and N, we compared a subset of ~300 FIA communities to communities from other datasets with directly matching S and N. We did not find differences in the distribution of percentile scores for any metrics between communities from FIA and communities from other datasets, confirmed via Kolmogorov-Smirnov tests (Appendix A9). Although 300 communities constitute a small sample relative to the 20,355 FIA communities we analyzed, these results point to statistical phenomena, and not biological attributes unique to FIA, as the likely explanation for the differences.

A second possibility is that these differences reflect statistical phenomena related to community size in terms of S, N, and as a result, the number of possible SADs in a community's feasible set. The FIA communities are the smallest across our datasets (Figure 1), and communities with small values of S and N have smaller feasible sets. When there are relatively few possible SADs in the feasible set, they may be less tightly clustered around their central tendencies, and the distributions for their shape metrics may be less narrowly peaked, than when there are very large numbers of possible SADs. High variability within the feasible set weakens the statistical distinction between "common" and "extreme" shapes (Figure 2). Under these circumstances, any deviations – or lack thereof – will be less informative than for communities with more strongly defined statistical baselines (Jaynes 1957). The dissimilarity to the central tendency, and the distributions of breath indices for specific metrics, broadly align with this principle. Across the range of community sizes represented in our datasets, small feasible sets have highly variable, and often very broad, feasible sets (Figure 4). More specifically, very small communities – for example, those with fewer than 2000 possible SADs in their feasible sets, or S ~20 and N ~40 – exhibit more highly variable feasible sets than large communities, and these small communities also show less consistent deviations (Figure 6; Appendix A10). Of our datasets, FIA is most dominated by small communities (68% of communities have fewer than 2000 possible SADs), and these small-community phenomena may therefore have the greatest impact on results aggregated over the FIA dataset.

If it is true that the highly variable feasible sets associated with small communities contribute to the weak evidence of deviations observed for the FIA dataset, such considerations affect our capacity to use this approach to distinguish signal from noise for a substantial contingent of ecological communities. Because the combinations of S and N represented in our analyses are irregularly distributed among different datasets (Figure 1), and because there is a great deal of variation in our breadth indices not accounted for by the size of the feasible set (Figure 4), we do not interpret these results as showing a threshold for defining problematically small communities. A more systematic exploration of the S and N state space, combined with more nuanced metrics for characterizing the variability of the feasible set, could clarify the relationship between S and N, the size of the feasible set, and statistical power. However, FIA and other small, highly variable communities have on the order of 10-20 species and 30-60 individuals, suggesting a general range of values below which we have diminished power to detect deviations from the statistical baseline represented by the feasible set. Communities with on the order of 5 species, or 100s to 1000s of individuals, have previously been suggested as "small" in this context (Preston 1948; McGill et al. 2007). To meaningfully draw inferences using deviations in these small communities, we will need more sensitive metrics than those used here, and/or theories that generate more specific predictions for the SAD. In the absence of such, we may stand to learn the most by focusing on SADs from relatively large communities.

It is also important to recognize that there are multiple plausible approaches to defining a statistical baseline for the SAD, of which we have taken only one (Haegeman and Loreau 2008, Locey and White 2013). Our approach follows Locey and White (2013) and reflects the random partitioning of individuals into species, with the resulting distributions considered unique if the species' abundance values are unique, regardless of the order in which the values occur. This philosophy reflects a longstanding approach in the study of abundance distributions: to focus on the shape of the distribution without regard to species' identities (McGill et al 2007). Other assumptions regarding the statistical baseline may be equally valid and generate different statistical expectations, which may alter if, and in what ways, empirical distributions appear unusual. For example, incorporating differences in species order into the statistical baseline – which would imply that identifying which species contain the most or least individuals is important – might reduce the representation of long-tailed, highly uneven SADs within the feasible set, and make the rare tail observed for real SADs

appear more unlikely than it does here. Under our assumptions, the SADs (1,2,3,4) and (1, 1, 1, 7) each count as only one unique SAD. Taking species order into account would mean that (1,2,3,4) would count as 24 (4!) unique SADs, because there are 4! ways to assign the abundances to each species. However, an SAD containing species with equal abundances, such as (1, 1, 1, 7), would only count as 4 unique SADs. For SADs, equal abundances are likely most prevalent among rare species. If this is true, then this set of assumptions would generate feasible sets where rare-tailed SADs are relatively scarce, making observed SADs with rare tails seem even more extraordinary. Additional formulations for the statistical baseline exist, including those that approximate exponential, Poisson, or log-series distributions in the limit (Harte et al. 2008, Favretti 2018). Investigating and comparing the results that emerge from different baselines will be an important next step towards reinvigorating the use of the SAD as a diagnostic tool.

Our study demonstrates the utility, and the potential challenges, of applying tools from the study of complex systems and statistical mechanics to the study of ecological communities (Haegeman and Loreau 2008, Harte 2011, White et al. 2012, Harte and Newman 2014). While concepts such as maximum entropy and the feasible set are promising horizons for macroecology, the small size of some ecological communities may present difficulties that are rare in the domains for which these tools were originally developed (Jaynes 1957, Haegeman and Loreau 2008). When the observed numbers of species and individuals are too small to generate highly resolved statistical baselines, these approaches will be less informative than we might hope – as appears to be the case for the smallest communities in our analysis. In larger communities, where mathematical constraints have more resolved effects on the form of the SAD, our results show that these constraints alone do not fully account for the extremely uneven SADs we often observe in nature – leaving an important role for ecological processes. This ability to detect and diagnose the specific ways in which empirical SADs deviate from randomness can generate new avenues for understanding how and when biological drivers affect the SAD. There are, of course, still many elements to be improved in our ability to distinguish biological signal from randomness, including assessing alternative statistical baselines and calibrating our expected power to detect deviations, especially for small communities. Indeed, more sensitive metrics could also enable identification of processes that operate through time. Continuing to explore and account for the interplay between statistical constraint and biological process constitutes a promising and profound new approach to our understanding of this familiar, yet surprisingly mysterious, ecological pattern.

Chapter 5: Concluding remarks

Biosketch

Renata Diaz grew up surrounded by the fields of eastern Colorado and the forests of New England, in a family of enthusiastic puzzle solvers and natural historians. During her undergraduate work in Ecology and Evolutionary Biology at Princeton University (class of 2015), she developed a keen interest in the interplay between theoretical, computational, and empirical modalities in ecology. For her senior thesis (advised by Rob Pringle), she combined field experiments with a mathematical model to predict how the joint effects of large mammalian herbivores, small mammalian seed predators, and historical legacies of human-wildlife interactions modulate the spatial distribution of trees in East African savannas. From 2015 to 2017, she continued to explore dimensions of ecology as a research assistant in the Staver Lab at Yale University (2015-2016), a field intern with the Ecology of Bird Loss Project (2016), and a spatial data analysis intern at the Missouri Botanic Garden (2017). In 2017, she joined the Ernest lab at the University of Florida to pursue a PhD in macroecology, combining field work with the Portal Project with data- and computationally-intensive synthesis, software development projects, and intellectually adventurous theoretical pursuits.

References

Brown, J. H. 1995. Macroecology. University of Chicago Press.

Chesson, P. 2000. Mechanisms of Maintenance of Species Diversity. Annual Review of Ecology and Systematics 31:343–366.

Connolly, S. R., T. P. Hughes, D. R. Bellwood, and R. H. Karlson. 2005. Community Structure of Corals and Reef Fishes at Multiple Scales. Science 309:1363–1365.

Cusser, S., C. Bahlai, S. M. Swinton, G. P. Robertson, and N. M. Haddad. 2020. Long-term research avoids spurious and misleading trends in sustainability attributes of no-till. Global Change Biology 26:3715–3725.

Dirzo, R., H. S. Young, M. Galetti, G. Ceballos, N. J. B. Isaac, and B. Collen. 2014. Defaunation in the Anthropocene. Science 345:401–406.

Dornelas, M., D. A. T. Phillip, and A. E. Magurran. 2011. Abundance and dominance become less predictable as species richness decreases. Global Ecology and Biogeography 20:832–841.

Dunning, J. B. 2008. CRC handbook of avian body masses. CRC handbook of avian body masses. 2nd ed. CRC Press, Boca Raton.

Ernest, S. K. M. 2005. Body size, energy use, and community structure of small mammals. Ecology 86:1407–1413.

Ernest, S. K. M., J. H. Brown, K. M. Thibault, E. P. White, and J. R. Goheen. 2008. Zero Sum, the Niche, and Metacommunities: Long-Term Dynamics of Community Assembly. The American Naturalist 172:E257–E269.

Ernest, S. K. M., E. P. White, and J. H. Brown. 2009. Changes in a tropical forest support metabolic zero-sum dynamics. Ecology Letters 12:507–515.

Fisher, J. A. D., K. T. Frank, and W. C. Leggett. 2010. Dynamic macroecology on ecological time-scales. Global Ecology and Biogeography 19:1–15.

Fristoe, T. S. 2015. Energy use by migrants and residents in North American breeding bird communities. Global Ecology and Biogeography 24:406–415.

Gardner, J. L., A. Peters, M. R. Kearney, L. Joseph, and R. Heinsohn. 2011. Declining body size: A third universal response to warming? Trends in Ecology & Evolution 26:285–291.

Harris, D. J., S. D. Taylor, and E. P. White. 2018. Forecasting biodiversity in breeding birds using best practices. PeerJ 6:e4278.

Henderson, P. A., and A. E. Magurran. 2010. Linking species abundance distributions in numerical abundance and biomass through simple assumptions about community structure. Proceedings of the Royal Society B: Biological Sciences 277:1561–1570.

Hernández, L., J. W. Laundré, A. González-Romero, J. López-Portillo, and K. M. Grajales. 2011. Tale of two metrics: Density and biomass in a desert rodent community. Journal of Mammalogy 92:840–851.

Holling, C. S. 1992. Cross-Scale Morphology, Geometry, and Dynamics of Ecosystems. Ecological Monographs 62:447–502.

Hubbell, S. P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32). Princeton University Press.

Kelt, D. A., J. R. Aliperti, P. L. Meserve, W. B. Milstead, M. A. Previtali, and J. R. Gutierrez. 2015. Energetic compensation is historically contingent and not supported for small mammals in South American or Asian deserts. Ecology 96:1702–1712.

Lawton, J. H. 1994. What Do Species Do in Ecosystems? Oikos 71:367–374.

Lenth, R. V. 2021. Emmeans: Estimated Marginal Means, aka Least-Squares Means.

McGill, B. J., M. Dornelas, N. J. Gotelli, and A. E. Magurran. 2015. Fifteen forms of biodiversity trend in the Anthropocene. Trends in Ecology & Evolution 30:104–113.

McNab, B. K. 2009. Ecological factors affect the level and scaling of avian BMR. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 152:22–45.

Morlon, H., E. P. White, R. S. Etienne, J. L. Green, A. Ostling, D. Alonso, B. J. Enquist, F. He, A. Hurlbert, A. E. Magurran, B. A. Maurer, B. J. McGill, H. Olff, D. Storch, and T. Zillio. 2009. Taking species abundance distributions beyond individuals. Ecology Letters 12:488–501.

Nagy, K. A. 2005. Field metabolic rate and body size. Journal of Experimental Biology 208:1621–1625.

Pardieck, K. L., D. J. Ziolkowski, M. Lutmerding, V. Aponte, and M.-A. Hudson. 2019. North American Breeding Bird Survey Dataset 1966 - 2018, version 2018.0. U.S. Geological Survey.

Petchey, O. L., and A. Belgrano. 2010. Body-size distributions and size-spectra: Universal indicators of ecological status? Biology Letters 6:434–437.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team. 2020. Nlme: Linear and Nonlinear Mixed Effects Models.

Read, Q. D., J. M. Grady, P. L. Zarnetske, S. Record, B. Baiser, J. Belmaker, M.-N. Tuanmu, A. Strecker, L. Beaudrot, and K. M. Thibault. 2018. Among-species overlap in rodent body size distributions predicts species richness along a temperature gradient. Ecography 41:1718–1727.

Schipper, A. M., J. Belmaker, M. D. de Miranda, L. M. Navarro, K. Böhning-Gaese, M. J. Costello, M. Dornelas, R. Foppen, J. Hortal, M. A. J. Huijbregts, B. Martín-López, N. Pettorelli, C. Queiroz, A. G. Rossberg, L. Santini, K. Schiffers, Z. J. N. Steinmann, P. Visconti, C. Rondinini, and H. M. Pereira. 2016. Contrasting changes in the abundance and diversity of North American bird assemblages from 1971 to 2010. Global Change Biology 22:3948–3959.

Schmitz, O. J., C. C. Wilmers, S. J. Leroux, C. E. Doughty, T. B. Atwood, M. Galetti, A. B. Davies, and S. J. Goetz. 2018. Animals and the zoogeochemistry of the carbon cycle. Science.

Smith, F. A., R. E. Elliott Smith, S. K. Lyons, and J. L. Payne. 2018. Body size downgrading of mammals over the late Quaternary. Science 360:310–313.

Supp, S. R., and S. K. M. Ernest. 2014. Species-level and community-level responses to disturbance: A cross-community analysis. Ecology 95:1717–1723.

Terry, R. C., and R. J. Rowe. 2015. Energy flow and functional compensation in Great Basin small mammals under natural and anthropogenic environmental change. Proceedings of the National Academy of Sciences 112:9656–9661.

Thibault, K. M., E. P. White, A. H. Hurlbert, and S. K. M. Ernest. 2011. Multimodality in the individual size distributions of bird communities. Global Ecology and Biogeography 20:145–153.

Van Valen, L. 1973. A new evolutionary law. Evolutionary Theory 1:1–30.

Warwick, R. M., and K. R. Clarke. 1994. Relearning the ABC: Taxonomic changes and abundance/biomass relationships in disturbed benthic communities. Marine Biology 118:739–744.

White, E. P. 2004. Two-phase speciesTime relationships in North American land birds. Ecology Letters 7:329–336.

White, E. P., S. K. M. Ernest, A. J. Kerkhoff, and B. J. Enquist. 2007. Relationships between body size and abundance in ecology. Trends in Ecology & Evolution 22:323–330.

White, E. P., S. K. M. Ernest, and K. M. Thibault. 2004. Trade-offs in Community Properties through Time in a Desert Rodent Community. The American Naturalist 164:670–676.

Ye, H., E. K. Bledsoe, R. Diaz, S. K. M. Ernest, J. L. Simonis, E. P. White, and G. M. Yenni. 2020, May. Macroecological Analyses of Time Series Structure. Zenodo.

Yen, J. D. L., J. R. Thomson, J. M. Keith, D. M. Paganin, E. Fleishman, D. S. Dobkin, J. M. Bennett, and R. Mac Nally. 2017. Balancing generality and specificity in ecological gradient analysis with species abundance distributions and individual size distributions: Community distributions along environmental gradients. Global Ecology and Biogeography 26:318–332.

Young, H. S., D. J. McCauley, M. Galetti, and R. Dirzo. 2016. Patterns, Causes, and Consequences of Anthropocene Defaunation. Annual Review of Ecology, Evolution, and Systematics 47:333–358.