2021

Novel Insights to Be Gained From Applying Metacommunity Theory to Long-Term, Spatially Replicated Biodiversity Data

Sydne Record
Bryn Mawr College, srecord@brynmawr.edu

Nicole M. Voelker
University of Maryland - Baltimore

Phoebe L. Zarnetske
Michigan State University

Nathan I. Wisnoski
Indiana University - Bloomington

Jonathan D. Tonkin
Oregon State University

Follow this and additional works at: https://repository.brynmawr.edu/bio_pubs

See next page for additional authors

This work is licensed under a Creative Commons Attribution 4.0 License.
Let us know how access to this document benefits you.

Custom Citation

This paper is posted at Scholarship, Research, and Creative Work at Bryn Mawr College. https://repository.brynmawr.edu/bio_pubs/36

For more information, please contact repository@brynmawr.edu.
Novel Insights to Be Gained From Applying Metacommunity Theory to Long-Term, Spatially Replicated Biodiversity Data

Sydne Record1*, Nicole M. Voelker2, Phoebe L. Zarnetske3,4, Nathan I. Wisnoski5, Jonathan D. Tonkin6,7, Christopher Swan2, Luca Marazzi8, Nina Lany3,4, Thomas Lamy9, Aldo Compagnoni10, Max C. N. Castorani11, Riley Andrade12 and Eric R. Sokol13,14

1 Department of Biology, Bryn Mawr College, Bryn Mawr, PA, United States, 2 Department of Geography and Environmental Systems, University of Maryland, Baltimore County, Baltimore, MD, United States, 3 Department of Integrative Biology, Michigan State University, East Lansing, MI, United States, 4 Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, United States, 5 Department of Biology, Indiana University, Bloomington, IN, United States, 6 Department of Integrative Biology, Oregon State University, Corvallis, OR, United States, 7 School of Biological Sciences, University of Canterbury, Christchurch, New Zealand, 8 Department of Biological Sciences, Southeast Environmental Research Center, Florida International University, Miami, FL, United States, 9 Marine Science Institute, University of California, Santa Barbara, Santa Barbara, CA, United States, 10 German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Martin Luther University Halle-Wittenberg, Halle, Germany, 11 Department of Environmental Sciences, University of Virginia, Charlottesville, VA, United States, 12 School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, United States, 13 National Ecological Observatory Network (NEON), Battelle, Boulder, CO, United States, 14 Institute of Arctic and Alpine Research (INSTAAR), University of Colorado Boulder, Boulder, CO, United States

Global loss of biodiversity and its associated ecosystem services is occurring at an alarming rate and is predicted to accelerate in the future. Metacommunity theory provides a framework to investigate multi-scale processes that drive change in biodiversity across space and time. Short-term ecological studies across space have progressed our understanding of biodiversity through a metacommunity lens, however, such snapshots in time have been limited in their ability to explain which processes, at which scales, generate observed spatial patterns. Temporal dynamics of metacommunities have been understudied, and large gaps in theory and empirical data have hindered progress in our understanding of underlying metacommunity processes that give rise to biodiversity patterns. Fortunately, we are at an important point in the history of ecology, where long-term studies with cross-scale spatial replication provide a means to gain a deeper understanding of the multiscale processes driving biodiversity patterns in time and space to inform metacommunity theory. The maturation of coordinated research and observation networks, such as the United States Long Term Ecological Research (LTER) program, provides an opportunity to advance explanation and prediction of biodiversity change with observational and experimental data at spatial and temporal scales greater than any single research group could accomplish. Synthesis of LTER network community datasets illustrates that long-term studies with spatial replication present an under-utilized resource for advancing spatio-temporal metacommunity research. We identify challenges towards synthesizing these data and
present recommendations for addressing these challenges. We conclude with insights about how future monitoring efforts by coordinated research and observation networks could further the development of metacommunity theory and its applications aimed at improving conservation efforts.

Keywords: LTER, NCO, synthesis, metacommunity, biodiversity, spatio-temporal, long-term

INTRODUCTION

Biodiversity is a key asset for environmental sustainability via its role in maintaining ecosystem functions and services (Cardinale et al., 2012; Hooper et al., 2012; Díaz et al., 2019). However, over the past half-century, human activities have increased the rate of biodiversity loss more than at any other time in history (World Health Organization, 2005; United Nations, 2017). The term “biodiversity” has been used to describe variation at many different levels of biological organization, but here we focus on species diversity (i.e., data sets from which a researcher can calculate measures of species richness or evenness) (Cleland, 2011). In association with declining biodiversity, researchers have documented changes in species interactions (Tylianakis et al., 2008), distributions, and phenology across ecosystems (Parmesan and Yohe, 2003; Dudgeon et al., 2006; Parmesan, 2006; Martay et al., 2016). These changes, which are not ubiquitous across taxa and landscapes, are expected to accelerate in the future (Brook et al., 2008; Bellard et al., 2012; Urban, 2015), further affecting ecosystem structure and function (Seto et al., 2012; Grimm et al., 2013).

A key goal of biodiversity science in the Anthropocene is to anticipate threshold shifts in the distributions and abundances of organisms and the ecosystem services that they afford to society (United Nations, 2017). However, without explicit consideration of multiple spatial scales and the within- and among-species pool connections, conclusions about the processes driving patterns of biodiversity are incomplete. Community structure and processes are still very common research themes in ecology; however, the greatest increase in number of publications has been observed on themes such as scale, anthropogenic impacts, and climate change (McCallen et al., 2019). Metacommunity theory and its applications are essential to addressing scaling in terms of space (e.g., local vs. regional) and time (i.e., linked to organisms’ generation times), especially in today’s rapidly changing communities and ecosystems (Mouquet and Loreau, 2002; Fahrig, 2003; Fischer and Lindenmayer, 2007).

In the past two decades, short-term field studies have considerably advanced our understanding of metacommunities across space by demonstrating how dispersal traits and landscape heterogeneity can affect community assembly in predictable ways (Logue et al., 2011; Leibold and Chase, 2017; Wilcox et al., 2017). Such information can provide insight into the types of metacommunity dynamics that organize biodiversity in each ecosystem (e.g., Pulliam, 1988; De Bie et al., 2012). However, spatially explicit snapshot data are limited in their utility to identify the processes that underlie observed patterns (Leibold et al., 2004; Brown et al., 2017; Sokol et al., 2017).Because ecological studies are often limited to the short-term (Hughes et al., 2017), considerable variability often remains unexplained in examinations of observational data. Assemblage composition and habitat availability may vary through time because of disturbance, seasonality, multi-year climatic variation (e.g., El Niño Southern Oscillation), and shifts of propagules in and out of dormancy (Holyoak et al., 2020). Many investigators acknowledge the limitation of not having temporal data (Frishkoff et al., 2014), particularly in dynamic systems such as streams (Datry et al., 2016; Sarremège et al., 2017; Tonkin et al., 2018). Not only can the environment change through time (e.g., habitat availability, environmental heterogeneity, and connectivity), but so too can the intrinsic processes regulating local biotic interactions (e.g., priority effects, intransitive competition) and dispersal from the relevant regional species pool (Chase, 2003; Fukami, 2015; Zarnetske et al., 2017). That is, temporal changes in biodiversity can occur regardless of environmental change.

With short-term spatially replicated metacommunity data, it can be difficult to distinguish between (1) exogenous environmental drivers of compositional changes and (2) endogenous community dynamics that may result from deterministic processes (e.g., multiple stable equilibria, endpoint assembly cycles, frequency-dependent coexistence in continuous space) on metacommunity composition. For instance, intransitive competition (i.e., “rock-paper-scissors” competition scenario) can result in endpoint assembly cycles (EACs) in which communities are decoupled from the influence of local environmental factors (Law and Morton, 1993, 1996; Steiner and Leibold, 2004). However, repeated observations in both space and time are necessary to demonstrate EAC dynamics. Specifically, the data would need to demonstrate that (1) all species involved in intransitive competitive dynamics persist regionally, (2) that the species involved cannot coexist locally, and (3) that low to moderate levels of dispersal among patches in the metacommunity allow for the EACs to play out. Therefore, both temporal and spatial replication are needed to test multiple hypotheses that explain how biodiversity is changing in such spatially complex landscapes (Leibold and Chase, 2017).

In the context of metacommunity assembly, given the problem of inferring spatio-temporal processes from spatial patterns alone, there have been multiple calls for long-term, spatial data collection to advance metacommunity research (Adler et al., 2005; Cottenie, 2005; Leibold and Chase, 2017; Holyoak et al., 2020). Financial and logistical constraints in field studies often hamper a comprehensive exploration of biodiversity trends across temporal and spatial scales. However, the various types of coordinated research networks [e.g., the United States Long Term Ecological Research (LTER) program, the International-LTER...
(ILTER), the global Nutrient Network, NutNet¹, the Smithsonian ForestGEO Network], and observatory networks [e.g., the United States Global Lake Ecological Observatory Network (GLEON), the National Critical Zone Observatory (CZO), and the United States National Ecological Observatory Network (NEON; Bourgeron et al., 2018)], provide a growing resource of long-term data that can be leveraged in synthesis science. Most of these coordinated research and observation networks, hereafter referred to as observation networks, have been working for over 10 years (e.g., the United States LTER will be 40 years old in 2020, the ILTER is 27 years old, GLEON is 16 years old, CZO is 12 years old, and the Long Term Agricultural Research network is 9 years old), while NEON recently completed construction of all 81 field sites and is slated to collect data for the next 30 years (Table 1).

With the maturation of long term data sets from such observation networks, the field of ecology is approaching an exciting point where there is the opportunity to empirically explore spatial and temporal representativeness of species within and among sites across ecosystems. Ultimately, a deeper understanding of the representativeness of species in space and time will lead to a better understanding of how scale influences metacommunity organization and biodiversity dynamics, which is key to making more general theoretical insights that transcend beyond the nuances of individual study systems. Here we provide evidence from a synthesis effort of LTER data showing that some LTER sites have the spatial replication within sites over time needed to capture asymptotic species-time-area-relationships (sensu Adler et al., 2005). The results of this synthesis suggest that these datasets could provide an untapped resource for metacommunity studies. When long-term studies include spatial contexts through multiple sites (within the regional species pool), they enable a more complete assessment of biodiversity change relative to long-term studies at a single site and allow for researchers to explore how sampling design might influence insights into metacommunities (Box 1). Spatio-temporal biodiversity patterns can take decades or longer to be described and explained (Magurran et al., 2010), partly owing to high temporal variability in community dynamics or lags in the processes, such as extinction debt, that structure biodiversity (Tilman et al., 1994; Kuussaari et al., 2009). More synthesis of data from long-term observation networks has the potential to uncover additional long-term, spatially replicated data that will aid in exploring the problem in identifying the spatio-temporal mechanisms underlying metacommunity assembly manifested in spatial biodiversity patterns.

Another key aspect of some observation networks (e.g., LTER, LTAR) is the existence of manipulative experiments that may also aid in teasing apart pattern from process in metacommunities. Observational studies considering metacommunity dynamics across space and time can still result in limited inferences about the dynamics of a metacommunity in the future. Most ecological systems may exist in different states wherein species compositions and abundances vary considerably, for example, with gradual shifts between states during succession or abrupt transitions when tipping points are reached (Suding and Hobbs, 2005).

<table>
<thead>
<tr>
<th>Network</th>
<th>Ecosystems</th>
<th>Organisms</th>
<th>Data publicly available? (if yes, link provided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long Term Ecological Research Network</td>
<td>Forest (tropical and temperate), grassland, desert, tundra, freshwater</td>
<td>Plants, animals, microbes</td>
<td>Yes <a href="https://portal.edirepository.org/nis/home.jsp">https://portal.edirepository.org/nis/home.jsp</a></td>
</tr>
<tr>
<td>National Ecological Observation Network</td>
<td>Forest (tropical and temperate), grassland, desert, tundra, freshwater</td>
<td>Plants, animals, microbes</td>
<td>Yes <a href="https://data.neonscience.org/">https://data.neonscience.org/</a></td>
</tr>
<tr>
<td>International Long Term Ecological Research Network</td>
<td>Forest (tropical and temperate), grassland, tundra</td>
<td>Plants, animals, microbes</td>
<td>Yes <a href="https://deims.org/">https://deims.org/</a></td>
</tr>
<tr>
<td>Nutrient Network</td>
<td>Grassland, tundra</td>
<td>Plants, animals, microbes</td>
<td>Yes <a href="https://nutnet.org/data">https://nutnet.org/data</a></td>
</tr>
<tr>
<td>Global Lake Ecological Observatory Network</td>
<td>Freshwater</td>
<td>Plants, animals, microbes</td>
<td>Yes <a href="https://gleon.org/data">https://gleon.org/data</a></td>
</tr>
<tr>
<td>Smithsonian ForestGEO</td>
<td>Forest (tropical and temperate)</td>
<td>Plants</td>
<td>Mixed <a href="https://forestgeo.si.edu/explore-data">https://forestgeo.si.edu/explore-data</a></td>
</tr>
</tbody>
</table>

¹http://www.nutnet.org/
One of the most challenging aspects of any metacommunity study is adequately characterizing the regional species pool of interest because it is context-dependent and the “region” is defined by the researcher. When samples are taken in multiple localities at a single point in time, we often assume that the species found across all localities capture the whole regional species pool. However, we know that not all species that could occur at study sites are observed in a single temporal snapshot because stochastic colonization/extinction events and/or transient dynamics might be occurring during the study (reviewed by Holyoak et al., 2020). For example, species experiencing diapause or having low detection probability might be missed or demographic stochasticity may eliminate one or more species in an isolated year (Figure 1). It takes time to sample the regional species pool, but spatially replicated sampling schemes can (1) increase the total number of species encountered and (2) decrease the time it takes to discover them. Additionally, succession and environmental change may add to the pool of species that can potentially occur in the study sites over time, as can the arrival of invasive species (Figure 2; Pickett, 1989; Li et al., 2020). Although the number of species observed at a site may be approximately the same across years, substantial compositional temporal turnover of species may take place (Forkin et al., 2017).

Long-term, spatially replicated data collection allows for the data to suggest the relevant spatial extent to study for characterizing the regional species pool for a given question. For example, researchers can see whether spatial or temporal samples saturate first and which sites continue to reveal more species diversity with greater sampling extent. Then, by looking at the temporal curves in a spatial context, researchers can better understand the most dynamic or undersampled regions of the landscape. This may help identify local sites that are on the periphery of the metacommunity that may be less influenced by spatio-temporal variation and/or dispersal (i.e., locally saturated sites whose composition is nested inside of the composition of other sites); and sites that are important integrators of spatio-temporal processes or important dispersal connections (i.e., sites with highly variable composition where repeated sampling continues to reveal new species and different composition). In a similar vein, Erős and Schmera (2010) combined field survey data on fish in a temporally and spatially dynamic stream system with simulation experiments to explore how spatial and temporal scales and their interaction influence species accumulation, which can lead to different inferences about metacommunity organization (e.g., the role of spatial versus environmental effects; Sály and Erős, 2016).

However, it is also important to note the limitations of species-time-area-relationships in evaluating sample representativeness over space and time within the context of understanding spatio-temporal metacommunity dynamics. For instance, species-time-area-relationship curves do not reveal insights about the temporal variability in changes in the abundance of organisms within and between sites, which is central to understanding mechanisms of metacommunity assembly (e.g., mass effects, species recovery from a historic disturbance; Scheffer, 2010). A new method for distinguishing changes in species richness across space due to separate effects of species abundance distribution, density, and the spatial configuration of individuals on a landscape presented by McGlinn et al. (2020) presents a promising path forward for more informed assessments of species-area-relationships, but incorporation of the temporal component into such new methods remains lacking. Regardless, we recommend that monitoring programs include the longitude and latitude of each sample plot if they want to be amenable to this new method, which requires plot locality information.

**FIGURE 1** | Species accumulation curves over time (years) and space with increasing numbers of study sites, indicated with different colors. Species accumulation curves through time (A,C) and space (B,D) are shown for birds from the Central Arizona – Phoenix United States LTER (2001–2016, Bate et al., 2017) (A,B) and for sessile invertebrates from the Santa Barbara Channel United States LTER (2016–2016, Reed, 2019) (C,D). The total number of species increases with the number of study sites, and the steepness of the curve also tends to increase with the number of study sites. Curves were fit according to the Arrhenius, Lomolino, and Michaelis-Menten models described in Dengler (2009) using the R package vegan (Oksanen et al., 2017). The most parsimonious model (lowest AIC) is shown for each subset of sites. Data and R code for generating this figure can be found in the Supplementary Data Sheets 2–4.
FIGURE 2 | Changes in a regional species pool over time. The species accumulation curve for a collection of sites may not level off over time due to the successional nature of communities, species invasions, or environmental change. (A) shows the species accumulation curve over 21 years for plants growing in the pumice plain habitat on Mt. St. Helens, United States (1989–2009; del Moral, 2010), a community undergoing succession following volcanic eruption. (B) shows the species accumulation curve over space for the same sites. Data and R code for generating this figure can be found in the Supplementary Data Sheets 1, 4.

2009). Even with extensive time series, it may be difficult to know where organisms within communities and metacommunities lie on a given temporal trajectory relative to a previous disturbance or an impending state change (Bestelmeyer et al., 2011). By contrast, field experiments with manipulations that influence aspects of metacommunity dynamics present the opportunity to test theoretical metacommunity frameworks to learn where theory aligns with observations (Logue et al., 2011), thus allowing for more robust predictions.

Long-term field manipulations are a key component of most United States LTER Network and other long-term ecological research programs because they can experimentally test how biodiversity responds to ecological processes occurring over extended periods (Turner et al., 2003). For example, by simulating the loss of eastern hemlock (Tsuga canadensis) due to the invasive wooly adelgid (Adelges tsugae) and measuring ant biodiversity for 13 years, researchers at the Harvard Forest LTER found that this experimental disturbance reduced the importance of species sorting for community composition (Sackett et al., 2011; Record et al., 2018). Long-term manipulations can also reveal how biodiversity responds to repeated environmental fluctuations, yielding knowledge that cannot be obtained with short-term experiments. By simulating annual kelp forest loss from ocean storms for 9 years, researchers at the Santa Barbara Coastal LTER demonstrated strong shifts in marine biodiversity that were contrary to findings from an earlier 2-year study (Castorani et al., 2018).

The Cedar Creek Biodiversity Experiment is a long-term manipulation of soil nitrogen availability and plant diversity that critically demonstrated the importance of biodiversity for ecosystem functioning (Tilman et al., 2012). This experiment has also allowed United States LTER researchers to answer additional questions related to community assembly and metacommunity ecology. For example, to assess the extent to which plant communities were dispersal-limited, seed mixtures were added into plots of a native grassland (Tilman, 1997). Increased seed additions led to greater local species richness, providing evidence that some species are dispersal limited (Tilman, 1997). Seed addition experiments have also shown that local species interactions, such as resource competition, are important for structuring local plant communities (Fargione et al., 2003). These results, when placed in the broader context of metacommunity ecology, show how both local and regional processes can influence biodiversity.

Long-term ecological research programs situated within a network of sites are critically important for understanding trends in biodiversity, especially when historical contexts are known and infrastructure enables long-term field experiments. Data sets published by observation networks often contain abundant and co-located biotic and abiotic data to investigate patterns of biodiversity through the lens of metacommunity theory. For example, spatio-temporal metacommunity dynamics can be investigated across numerous ecosystems and taxa by means of NEON’s frequent and coordinated biotic and abiotic sampling within plots or stream reaches nested within sites across 20 climatic domains (Keller et al., 2008). Such networks provide economies of scale for infrastructure for long-term and networked sites, dedicated human resources for data collection, and taxonomic expertise (Bourgeron et al., 2018). Based on insights we made while curating LTER data for metacommunity analyses, we identify challenges towards synthesizing these data and present recommendations for addressing these challenges. We also provide guidance for long-term monitoring programs based on our insights.

CHALLENGES TO ADVANCING METACOMMUNITY SCIENCE AND PATHS FORWARD

Challenge One – Scale Mismatch Among Data Sets in Synthesis Efforts

To address broad questions about the generality of the metacommunity framework in ecology, it is necessary to broadly...
test the theory across ecosystems and organisms. This presents a challenge with respect to mismatches in temporal and spatial scaling across data sets (Lamy et al., 2018). Observatory networks can employ various scales of temporal resolution – days, weeks, seasons, years, or even decades and plot sizes (e.g., Keller et al., 2008). Cross-ecosystem syntheses of biodiversity often compare data that has been aggregated or standardized to a common temporal or spatial grain (e.g., annual observation frequencies; Collins et al., 2018). However, trends or shifts in biodiversity can be affected by species' generation times (Kuussaari et al., 2009), phenological patterns, and the frequency and duration of dispersal events (e.g., Tilman, 1997) and environmental fluctuations (e.g., Free et al., 2013), which do not always align with common ways of standardizing space and time across data sets (e.g., annual sampling schedules).

In light of these challenges, research questions focused on understanding the underlying processes that structure metacommunity assembly (i.e., species interactions, environmental filtering, dispersal limitation) must be aware of heterogeneity in sampling effort and spatial grain (i.e., plot size) across studies, which biodiversity estimates and variability among samples are sensitive to [Chase and Knight (2013), Spake et al. (2020)]. Spake et al. (2020) suggest that in formal meta-analyses scale dependence in effect sizes may be assessed using meta-regressions exploring relationships between either spatial (i.e., plot size) or temporal (i.e., sampling interval) grain and effect sizes across studies. They also illustrate with simulated community data how effect sizes calculated with the log response ratio metric applied to biodiversity estimates (i.e., species richness) were more accurate than those calculated with the common Hedge's g metric. In instances, when effect sizes applied to biodiversity estimates are highly scale dependent, the use of a scale-independent metric (e.g., Hurlbert's Probability of Interspecific Encounter) is preferred (Chase and Knight, 2013). Given the challenges of synthesizing biodiversity data with varying grains of sampling in space and time, we have two recommendations for monitoring programs. First, ensure that raw data are published with ample metadata, so that synthesis researchers can extract relevant information on grains of sampling (Spake et al., 2020). Second, we recommend that programs coordinate efforts to agree upon standardized sampling protocols for particular taxa to promote synthesis (see more specifics on such coordination in the Challenge Three subsection below).

**Challenge Two – Rare Species**

In a non-stationary world, rare species will be crucial for predicting future states of novel ecosystems (Lyons et al., 2005; Jain et al., 2014). However, trends and patterns observed for rare taxa can be challenging to interpret because they arise from a combination of observation error and stochastic colonization dynamics (Hanski et al., 2004; McGill et al., 2007). Capturing rare species dynamics is an important step in quantifying the regional species pool, which is an essential component of studies that embrace spatial dimensions and/or dispersal. Only with long-term temporal data from multiple sites can one understand the presence or absence of rare taxa because transient or local fluctuations (e.g., Free et al., 2013), which do not always align with common ways of standardizing space and time across data sets (e.g., annual sampling schedules).

In light of these challenges, research questions focused on understanding the underlying processes that structure metacommunity assembly (i.e., species interactions, environmental filtering, dispersal limitation) must be aware of heterogeneity in sampling effort and spatial grain (i.e., plot size) across studies, which biodiversity estimates and variability among samples are sensitive to [Chase and Knight (2013), Spake et al. (2020)]. Spake et al. (2020) suggest that in formal meta-analyses scale dependence in effect sizes may be assessed using meta-regressions exploring relationships between either spatial (i.e., plot size) or temporal (i.e., sampling interval) grain and effect sizes across studies. They also illustrate with simulated community data how effect sizes calculated with the log response ratio metric applied to biodiversity estimates (i.e., species richness) were more accurate than those calculated with the common Hedge's g metric. In instances, when effect sizes applied to biodiversity estimates are highly scale dependent, the use of a scale-independent metric (e.g., Hurlbert's Probability of Interspecific Encounter) is preferred (Chase and Knight, 2013). Given the challenges of synthesizing biodiversity data with varying grains of sampling in space and time, we have two recommendations for monitoring programs. First, ensure that raw data are published with ample metadata, so that synthesis researchers can extract relevant information on grains of sampling (Spake et al., 2020). Second, we recommend that programs coordinate efforts to agree upon standardized sampling protocols for particular taxa to promote synthesis (see more specifics on such coordination in the Challenge Three subsection below).

**Challenge Three – Economies of Scale**

It can often be difficult to assess whether long-term data are collected at optimal temporal resolutions and spatial extents to capture all relevant community assembly dynamics (e.g., dispersal kernel shape, demographic rates) or structural characteristics (e.g., spatial heterogeneity of suitable habitat, species occurrence, biomass). Increasing the spatial extent of data collection around existing long-term study sites will enable researchers and managers to compile the information needed for understanding trends in biodiversity and will allow for better characterization of regional species pools informed by data. Resources are often not available in any individual research program to capture both the necessary spatial and temporal resolution. There is a growing need for both spatial and temporal replication in biodiversity data, which requires a plan for coordination among single-PI projects and multiple long-term observatories in a network of networks to provide economies of scale within the research community.

We recommend that scientific societies provide a hub for coordination among and between single-PI and large-scale observation networks to help identify opportunities where single-PI and observation network projects can fill complementary knowledge gaps. For instance, the National Science Foundation's Macrosystems Biology program and NEON-Enabled Science solicitations provide an opportunity for short duration (3 years) studies that leverage NEON...
Data collection priorities could inform monitoring frequency and spatial replication based on organismal life histories (e.g., time until reproduction, dispersal abilities) and current knowledge gaps (Wolfe et al., 1987). Adopting pre-existing, standardized sampling and data archiving protocols that are consistent among sites will also enable researchers and policy makers to scale-up local studies to global scale research initiatives (e.g., Group on Earth Observations – Biodiversity Observation Network’s essential biodiversity variables; Haase et al., 2018). Recent proposals for integrating measures of biodiversity and ecosystem integrity across observatory networks may advance synergy within and among these networks (Haase et al., 2018). In addition to observation networks, spatially replicated studies in long-term databases, such as BioTime (Dornelas et al., 2018), offer additional data sources.

Data standards can also aid synthesis efforts. Furthermore, data to be used for multi-site analyses are best archived in harmonized datasets (with consistent structure and format). Examples include the GLEON DataONE Repository for synchronized hydrological sensor data2 and the ecocomDP standard data pattern for community data3 that members of our group have developed with the Environmental Data Initiative (EDI)4 for implementation in their data portal, which publishes data products from the United States LTER and NSF Macrosystems Biology programs. A key outcome of the LTER synthesis group that we are a part of has been to harmonize LTER community ecology data sets into the ecocomDP standard data pattern to promote future use of LTER data in metacommunity studies. A key first step in using LTER time series to address hypotheses in metacommunity ecology is the identification of appropriate data sets. The analysis ready data provided by ecocomDP provides additional metadata for improved discovery with information on taxonomic resolutions, and nesting of sampling designs over space and time – key pieces of information for identifying the suitability of a data set for a metacommunity study. Currently, seventy community ecology data sets generated by the LTER network have been formatted into the ecocomDP standard data pattern and there are ~100 more data sets in the queue for processing. Completed data sets are discoverable by going to the EDI Data Repository5 and searching for the term “ecocomDP.” Ultimately, these types of harmonized data sets will allow for greater advances in metacommunity studies because efforts to clean and format data leading up to analyses do not have to be repeatedly performed by individual researchers and there is the additional benefit that results from studies can then also be more reproducible (Reichman et al., 2011).

Challenges

**Challenge 4 – Statistical Integration of Long-Term, Spatially Replicated Data With Theory**

Although we are at a point where some observation networks have amassed long-term, spatially replicated community data sets with saturating species-time-area relationships, the statistical integration of these data with theoretical concepts remains a key challenge. There have long been calls for moving beyond the classic metacommunity conceptual archetypes (i.e., mass effects, patch dynamics, species sorting, neutral theory) to better account for temporal dynamics, but theoretical and statistical approaches remain incomplete (Leibold and Chase, 2017). For instance, ecosystem stability over large spatial scales can be addressed in a metacommunity framework with long-term, spatially replicated data (Wang and Loreau, 2014, 2016). However, such analyses often consider aggregate metrics of ecosystems (e.g., biomass) rather than species diversity and composition, which may be of greater interest to federal agencies or non-governmental organizations, as these often uphold policy based on species diversity rather than aggregate metrics (i.e., the Endangered Species Act).

Greater strides in metacommunity science will be made as long-term, spatially replicated observation network data enter a loop, wherein the data impart information into the development of models and theory, and models and theory inform future data collection (Dietze, 2017). A promising path forward involves incorporating novel approaches to quantify metacommunity dynamics [e.g., joint species distribution models (Ovaskainen et al., 2019), open-source simulation tools (Sokol et al., 2017), process-based models (Keyel et al., 2016; Thompson et al., 2020)] into such cyclical rapid data assimilation and model/theory refinement. For instance, recent work by Thompson et al. (2020) revisits the metacommunity concept with a process-based framework that integrates local and regional dynamics of ecological communities with three main underlying dimensions (i.e., density independent responses to abiotic conditions, density-dependent biotic interactions, and dispersal) that link to the classic metacommunity conceptual archetypes. Near-term forecasting of process-based metacommunity models based on this reconceived metacommunity framework could help to identify which species traits best capture variation within and between species that influence density dependence to inform future monitoring and data collection efforts.

**PROSPECTUS**

Given the alarming rate at which biodiversity and associated ecosystem services are being lost (Ceballos et al., 2015; Johnson et al., 2017), understanding changes in biodiversity in both space and time is fundamental for science-informed conservation. While the metacommunity framework has the potential to uncover mechanisms explaining biodiversity patterns to inform conservation, the lack of spatio-temporal data has hindered researchers’ ability to disentangle environmental drivers from biotic niche-based processes generated within the
Incorporating a long-term temporal dimension into field-based metacommunity research is key to understanding the mechanisms generating observed patterns in biodiversity. However, resources for field studies are limited and temporal replication often comes at the expense of spatial replication and taxonomic resolution (e.g., Keller et al., 2008). Large, collaborative observation networks provide the opportunity to inform metacommunity theory with empirical data at spatial and temporal scales greater than any single researcher could accomplish on their own. To better understand how metacommunity dynamics operate in reality, the infrastructure of manipulative field experiments at sites within these networks allows ecologists to test challenging questions posed by metacommunity theory at real-world scales. Looking forward, such large-scale efforts can be better leveraged (Bourgeron et al., 2018) to address issues of scale mismatches in data synthesis, rare species, economies of scale, and the integration of data with theory. By coordinating biodiversity research efforts, ecologists will better understand how and why species persist across space and time, and how biodiversity patterns emerge across a diverse range of ecosystems and over long temporal scales.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SR, NV, and ES led the writing and editing of the manuscript. ES, NW, and CS secured funding from the LTER to support the working group. All authors contributed to the conceptualization, review, data collection, and editing of the manuscript.

REFERENCES


