Full title: Multivariate methods for testing hypotheses of temporal community dynamics

Running header: Testing hypotheses of temporal community dynamics

Authors and affiliations:

Hannah L. Buckley[1]*, Nicola J. Day[2], Bradley S. Case[1], Gavin Lear[3], Aaron M. Ellison[4]

[1] School of Science, Auckland University of Technology, Auckland, New Zealand

[2] Biology Department, Wilfrid Laurier University, Waterloo, Ontario, Canada

[3] School of Biological Sciences, University of Auckland, Auckland, New Zealand

[4] Harvard Forest, Harvard University, Petersham, Massachusetts, USA

* Corresponding author: Hannah.Buckley@aut.ac.nz

ABSTRACT

For ecological research to make important contributions towards understanding and managing temporally-variable global change processes, such as responses to land-use and climatic change, we must have effective and comparable ways to quantify and analyse compositional change over time in biological communities. These changes are the sum of local colonisation and extinction events, or changes in the biomass and relative abundance of taxa within and among samples. We conducted a quantitative review of currently available methods for the analysis of multivariate datasets collected at temporal intervals. This review identified the need for the application of quantitative, hypothesis-based approaches to understand temporal change in community composition, particularly for small datasets with less than 15 temporal replicates. To address this gap, we: (1) conceptually present how temporal patterns in community dynamics can be framed as specific, testable hypotheses; (2) provide three fully-worked case-studies, complete with R code, demonstrating multivariate analysis methods for temporal hypothesis testing and pattern visualisation; and (3) present a road map for testing specific, quantitative hypotheses relating to the underlying mechanisms of temporal community dynamics.

KEYWORDS

Beta diversity, biodiversity, community variation, compositional change, multivariate analysis, species turnover, temporal change, temporal variability, time series, zeta diversity

INTRODUCTION

A key task for ecologists is to monitor, quantify, analyse, and predict temporal changes in the structure of ecological communities as a function of ongoing changes in climate, land use, and other environmental drivers. Ecological communities comprise populations of multiple taxa. The structure of communities—the number of taxa (‘richness’) and their relative abundance (‘composition’)—changes through time because of fluctuations of individual

38 populations and local colonisations and extinctions of taxa (Micheli et al. 1999, Legendre and
39 Gauthier 2014). These temporal dynamics are governed by both intrinsic factors, including
40 intraspecific and interspecific interaction, colonisation, and extinction, and extrinsic factors,
41 which are processes that relate to periodic disturbances and directionally changing
42 environmental conditions (Connell and Slatyer 1977, Brown and Lawson 2010).
43 Understanding the temporal dynamics of community structure can illuminate fundamental
44 ecological processes, including effects of individual life histories on rates of ecosystem
45 change, the relative importance of biotic and abiotic factors in determining community
46 structure, or how taxa and the networks in which they are embedded respond to
47 environmental change (Tilman 1999, Hobbs et al. 2007, Kampichler et al. 2012). Temporal
48 changes in taxon richness and composition also can provide both simple summaries of
49 ecosystem change and information that can be translated rapidly into action, either to protect
50 and restore declining taxa, or to staunch unsustainable growth of unwanted organisms.

51 Representations of community structure as multivariate datasets that include
52 information on multiple taxa can be more sensitive to changes in explanatory or predictor
53 variables (e.g., pH, temperature, soil type), than univariate summary statistics, such as taxon
54 richness, diversity, or biomass (Micheli et al. 1999, Chao et al. 2014), and hence more
55 informative about causal processes. However, quantitative methods for investigating and
56 testing hypotheses about temporal dynamics of multivariate measures of community structure
57 (henceforth 'community dynamics') also are more complex than methods used for univariate
58 data (Schaefer et al. 2005, Legendre and Legendre 2012). To date, methods for the analysis
59 of spatial multivariate community structure data (e.g., Whittaker 1960, Legendre and
60 Legendre 2012) have received far more attention than methods for analysis of temporal
61 multivariate data. Well-known concepts used to describe spatial changes in community
62 structure—grain, extent, nestedness, beta diversity, taxonomic turnover—have been

transferred to the temporal domain (Azeria and Kolasa 2008, Anderson et al. 2011, Legendre and Gauthier 2014), and some well-developed multivariate methods for large spatial datasets have been used to detect temporal changes in community structure (Legendre and Gauthier 2014). However, although temporal studies of compositional change are commonly undertaken, it is not always clear which analytical methods are appropriate or available for any given research question, data type, or community structure; indeed, there may not be one ideal method for all situations and a range of methods may be required to answer a given question (Whittaker 1960). Comparisons of multivariate ecological data through time pose unique, non-trivial challenges because these analyses must also account for spatial variation in these data, (Philippi et al. 1998, Micheli et al. 1999, Collins et al. 2000, Schaefer et al. 2005).

In this paper, we: (1) review empirical studies that have undertaken analyses of multivariate temporal community datasets, to gain insight into how researchers have been applying analysis methods over the past several decades, (2) provide recommendations for the greater use of quantitative methods for testing specific temporal hypotheses, and (3) provide guidance for researchers, including those new to community dynamics research, in analysis decisions, particularly for small temporal datasets (fewer than 15 temporal replicates).

How to use this paper

This paper does not necessarily need to be read from start to finish; we suggest readers identify the section(s) most appropriate for their interests. The remainder of this paper is structured into five main sections: (A) a review of the recent literature; (B) a conceptual description of the types of hypotheses that can be addressed in analyses of temporal community dynamics data; (C) three worked case study examples illustrating hypothesis-driven analyses of temporal community data, including code for the software programme ‘R’

88 (R Core Development Team 2017) to allow readers to easily implement methods described in
89 this paper; (D) A "road map" to analysing temporal community data, including descriptions,
90 and how to identify and present, analyses appropriate for testing particular hypotheses; (E) A
91 brief discussion of recent advances in multivariate analysis of community dynamics.

92 A. LITERATURE REVIEW

93 DATA COLLECTION AND ANALYSIS

94 We reviewed and described the multivariate analyses of community dynamics in studies
95 published up until September 2017 to provide an overview of available methods, their
96 benefits and drawbacks, and what they can reveal about community dynamics. Specifically,
97 we asked: Are data collected from different taxonomic groups, habitats, spatial extents, and
98 time-scales analysed in particular (or different) ways, i.e., are there consistencies in analysis
99 approaches due to discipline-specific preferences?

100 We searched the literature for studies investigating changes in ecological
101 communities over time. We searched the ISI Web of Science Core Collection on 5 September
102 2017 for articles using the following search statement: ["temporal" AND "composition*"
103 AND "communit*" AND ("community dynamic*" OR "temporal dynamic*" OR "temporal
104 community varia*" OR "community change" OR "compositional change")]. We refined this
105 search to include only document types listed as 'articles' and published in 'English'. This
106 search yielded 692 articles for review. Of these, we excluded 241 papers that lacked
107 community data (i.e. those that included fewer than three taxa), had no temporal replication
108 or pooled all temporal data, used only artificial communities or simulated datasets, or used
109 space-for-time substitution (e.g., using chronosequences) to infer community ages. Although
110 we did not constrain the temporal search window, our search terms only returned papers after
111 1990, indicating that the specific terms we used were not part of the compositional change
112 literature prior to the 1990s.

We extracted and collated key attributes relating to each study in the remaining 451 papers from the ISI database (e.g., publication year, number of citations). Specific study attributes were determined from the body text of each manuscript, including information on (i) habitat type (e.g., estuarine); (ii) habitat location (e.g., U.S.A.); (iii) taxonomic identity (e.g., vertebrate; fish); (iv) manipulation (e.g., natural variation, experimental resource manipulation); (v) spatial extent (e.g., local, global); (vi) temporal grain and extent (e.g., minutes, months); (vii) how composition was measured (e.g., presence-absence, relative abundances of taxa); (viii) how time was included in each analysis (i.e., explicitly as a predictor variable, or where other variables represent time, e.g., climatic factors recorded at different times); and (ix) the key perceived research question (e.g., exploration of yearly dynamics) (Data S1). We categorised organisms being studied into the following taxonomic groups: plants (Kingdom Plantae); vertebrates (Phylum Chordata); invertebrates (Phyla Annelida, Arthropoda, Cnidaria, Ctenophora, Echinodermata, Mollusca, Platyhelminthes, or Porifera); microeukaryotes (algae and Kingdoms Fungi and Protista); prokaryotes (Domains Bacteria and Archaea); and viruses (non-cellular pathogenic organisms). The data types collated and included papers are provided in the supplementary material (Data S1).

LITERATURE REVIEW RESULTS AND DISCUSSION

The 451 papers identified by our search, when compared by year to the number of papers in community ecology, showed that every year since 1990, publications on temporal community dynamics consistently have comprised approximately 24% of studies within community ecology. Thus, studies on temporal dynamics have increased from less than five publications per year in the early 1990s to more than 30 per year in the last five years (Appendix S1: Fig. S1). Few of the 451 reviewed publications (20 papers; 4% of those reviewed) contained more than one independent dataset; those instances were treated as multiple studies that were analysed as separate data points (increasing our total sample size *n* to 475). The majority of

138 studies (70%) were conducted within North America and Europe (Fig. 1), likely reflecting the
139 relative availability of research funding and resources.

140 Reviewed studies generally focused on organisms, scales, and habitats that were
141 conducive to studying temporal community dynamics. For instance, prokaryotes and
142 invertebrates were studied most frequently (Fig. 1), being relatively dynamic and easy to
143 study over smaller temporal and spatial scales; by contrast, research on plants and vertebrate
144 fauna together accounted for less than one-third of all studies of community dynamics.
145 Although studies spanned a broad range of habitat types, taxa, spatial, and temporal scales,
146 there were two-to-three times more studies at small spatial extents (micro, point sample, and
147 local scales) across the 14 habitat categories identified from the articles (Appendix S1: Fig.
148 S2A). Similarly, the most common temporal extent (i.e., the time between the collection of
149 the first and last samples) of individual studies was years, and only a minority of studies on
150 plant and vertebrate communities extended for a decade or more (Appendix S1: Fig. S2B).
151 Smaller organisms were often measured at finer temporal resolutions (grain) than larger
152 organisms, and many studies on short-lived prokaryote and viral communities lasted more
153 than one year. The temporal grain of sampling generally was smaller for communities of
154 microorganisms than for communities of vertebrate animals and plants, for which the most
155 common grain for repeated sample collection was years rather than months or seasons
156 (Appendix S1: Fig. S2B).

157 Researchers frequently used a range of different analytical methods, even within one
158 study, to enable them to understand the complexities of observed temporal changes. Up to
159 eight different analytical methods were used (Fig. 2A), with about two-thirds of the studies
160 applying ≤ 3 methods, and most using two. Although the number of studies addressing
161 community dynamics have increased dramatically since the 1990s and have begun to
162 incorporate more computationally-complex methods over time (Appendix S1: Fig. S3A),

163 descriptive methods (i.e., visual, non-statistical comparisons) and ordination have remained
164 the most commonly-used analyses, comprising 33% and 22% of all applied methods,
165 respectively (Fig. 2B). In fact, 134 out of 451 studies (30%) used only descriptive approaches
166 to "analyse" their datasets, while another 35% of studies used descriptive methods in addition
167 to at least one other method. We suspect that the overall ease of use of descriptive and
168 ordination methods, their accessibility via readily-available software programs, and their
169 focus on visually-presenting complex multivariate data, together have driven this trend. The
170 use of the raw dissimilarity values (18%), time-lag analysis (9%), and cluster analysis (7%)
171 comprised the next most common types of methods; the remaining 12 methods were used
172 infrequently (Fig. 2B). Several methods captured by this review were not multivariate (see
173 Appendix 2), including univariate linear methods, temporal stability (*a.k.a.* coefficient of
174 variation), frequency change (testing for significant changes in the frequency of individual
175 taxa over time), and multiplicative change (percent cover change statistic for a single taxon or
176 taxon group used in subsequent regression analyses). Since 2005, the use of raw dissimilarity
177 methods has increased considerably. In the same decade, a steady stream of alternative
178 methods that incorporate time explicitly have been applied, albeit with poor uptake across the
179 wider literature in this area. These methods often are 'one-off' analyses that have been
180 applied only a handful of times; we refer to these as 'minor' methods (Appendix S1: Fig. S3).
181 Preferred ordination methods changed from detrended correspondence analysis (DCA) and
182 principal components analysis (PCA) in the early 1990s, to canonical correspondence
183 analysis (CCA) in the late 1990s, and, since the early 2000s, to non-metric multidimensional
184 scaling (NMDS) (Appendix S1: Fig. S3B).

185 Three-quarters of all analyses compiled were from observational studies aimed at
186 describing patterns of community change or related to monitoring community composition,
187 rather than exploring environmental effects or studying the effects of biotic processes or

188 interactions (Appendix S1: Fig. S4). In fewer cases, researchers were interested in the effect
189 of environmental change, such as seasonality or climatic cycles over longer time periods.

190 Our review also shows how common it is for researchers to analyse datasets with low
191 temporal replication. Across research aims and for most methods, most time series had fewer
192 than 15 observations (mean = 14.9, median = 7, mode = 4; Fig. 3); the few studies that used
193 network analyses and machine-learning methods used 50 or more observations taken through
194 time. Overall, neither the number of temporal replicates, nor any other study attributes we
195 extracted during our review (e.g., habitat, taxonomic group), appeared to explain well the
196 differences in how researchers analysed their data or their motivations for doing so.

197 Typically, researchers will have some notion of how their communities of interest
198 might be changing, which motivated them to undertake temporal sampling in the first place
199 (Bestelmeyer et al. 2011). In some cases, long-term datasets exist because previous
200 researchers began data collection with a rationale and method that is different from the
201 current research questions, and so we can sometimes be limited by past sampling designs.
202 Nonetheless, our review highlights that most studies describe temporal community dynamics
203 within their studied communities without explicit, quantitative hypotheses and specific,
204 testable predictions about the expected drivers of observed community changes (Appendix
205 S1: Fig. S4). Clearly, there are many benefits to entering the analysis pathway with a strong *a*
206 *priori* prediction of the compositional pattern(s) expected and the biological hypotheses that
207 underpin those predictions.

208 **B. HYPOTHESES OF TEMPORAL COMMUNITY DYNAMICS**

209 Quantitative and specific predictions of the nature and variability in temporal community
210 dynamics allows deeper understanding of the patterns and processes involved. However, to
211 achieve this, *a priori* expectations of change need to be developed and paired with the
212 appropriate analytical method(s) to achieve rigorous tests. Here, we describe in detail the

possible observed changes in community composition (taxon presence or abundance) in one sampling unit (site, host, etc.) over time (Fig. 4) or in a set of sampling units over time (Figs. 5-7) and what some of the ecological research questions are that may underpin such hypotheses. In each case, we give predictions for how pairwise dissimilarity values, the basis for most of the commonly-used quantitative measures of compositional change identified in the literature review, would change. Many of these hypotheses are not mutually exclusive although some could be used as multiple alternatives to be tested simultaneously, depending on the research question(s) of interest. Note that 'dissimilarity measures', also called 'distance measures' or 'beta-diversity', are the inverse of 'similarity measures'. These metrics quantify the compositional distance between two samples (pairwise dissimilarity).

Simple predictions for a single sample unit: Persistence and directional change

Aside from the 'null' prediction of no change in composition ('persistence'; Fig. 4b), one of the simplest predictions of compositional dynamics is that of 'directional' change, whereby certain taxa either increase or decrease so that the dissimilarity of a sampling unit to itself at previous sampling occasions increases (or decreases) over time (Fig. 4c); when composition turns over as taxa are replaced, the magnitude of change in dissimilarity can be high. Such a hypothesis could arise from many ecological processes, e.g., has the community undergone 'succession', whereby there is a predictable replacement of individual or groups of taxa over time; has it undergone 'degradation', whereby rare taxa are lost from the system over time; has it had an 'invasion' event, where the relative dominance of one taxon or functional group increases; is it changing in composition in response to a changing biotic or abiotic variable, such as predation or temperature?

Simple patterns for multiple sample units: Directional change, divergence and convergence

Where a set of sample units are all undergoing the same directional changes in composition over time, pairwise 'spatial' dissimilarity among that set of sampling units, would not change

238 significantly over time, because all sampling units are moving in the same direction at the
239 same rate in compositional space because the same taxa are changing in the same ways (Fig.
240 5b). We also might ask whether sample units are becoming more dissimilar to each other, due
241 to interactions between the changing factors described above and characteristics of the
242 community, such as initial composition or differences in habitat conditions. In this case
243 'divergence' represents the hypothesis where a set of sampling units are all becoming more
244 dissimilar in their composition, so that their pairwise dissimilarities increase over time; this
245 term therefore does not apply to change within one sampling unit (Fig. 5c).
246 Two very similar patterns predict decreasing pairwise dissimilarity values over time.
247 In contrast to 'divergence', under 'convergence' taxonomic composition of two or more
248 sampling units becomes more similar over time (Fig. 5e) in a predictable way because of
249 increases or decreases in taxon presence or abundance; this is also called 'homogenisation'
250 (Fig. 5e), where there is a concurrent decrease in diversity and dissimilarity becomes
251 relatively low. In contrast, convergence simply implies that between sample-unit dissimilarity
252 declines over time, but not necessarily to a very low level. In the cases where taxa are lost,
253 the predicted overall decline in dissimilarly is large. These hypotheses of change would be
254 relevant if we are interested in the effects of an experimental treatment designed to increase
255 the growth of a particular set of taxa, or under invasion, such as the biotic homogenisation of
256 cities due to urbanisation, or loss of native taxa due to competition with the invading taxa.
257 Note that the predicted effect of invasion on dissimilarity values depends on whether we are
258 considering the trajectory of one sample unit (divergence) or many sample units
259 (convergence) and the among-sample unit variation in other interacting factors.
260 *More complex temporal patterns: Perturbations, resilience, and cyclical change*
261 More complex patterns in temporal dynamics involve compositional shifts that change in
262 their magnitude or direction over time. Natural or manipulative experiments asking questions

263 about the effects of disturbance on community composition can lead to complex predictions
264 of change. 'Perturbation' is a pattern for which the change in composition is initially high but
265 lessens over time and can be detected by comparison to undisturbed treatments or areas (Fig.
266 6a). The perturbation can occur as either a 'press perturbation' (Fig. 6b), such as the removal
267 of a predator or a continuous nutrient addition treatment, or a 'pulse perturbation' (Fig. 6c),
268 such as a disturbance event or a population density reduction; these two perturbations result
269 in different rates of change in pairwise temporal dissimilarities. Post-perturbation, a
270 community either could remain in a state that is dissimilar to un-perturbed sample units
271 ('tipping point'), or could return to its previous state over time ('resilience'). The prediction
272 of a 'pivot point' or 'threshold' can be used to test for a significant compositional shift for
273 one or more sampling units over a relatively short period of time, meaning that temporal
274 pairwise dissimilarity among times for each sampling unit is low before, and also after, the
275 pivot point, but is high for temporal pairs across the pivot point (Fig. 6c).
276 Questions about the scale of temporal dynamics in a community, such as the presence
277 of seasonal cycles, lead to predictions of 'cyclical' change or 'periodicity', which represent
278 shifts in composition that repeatedly diverge and re-converge back to a previous state;
279 dissimilarity increases at short temporal scales but decreases at larger temporal scales, and
280 will alternate between high and low at increasing temporal scales that correspond to some
281 cyclical process, e.g., warming events (Fig. 4d).
282 *Predictability of change: Instability and stabilisation*
283 Questions about the predictability of compositional changes over time resulting from
284 interactions between causal-change processes, such as temperature shifts throughout a year,
285 and biotic or abiotic differences among sampling units, could also lead to hypotheses
286 regarding the variability in relative compositional change. For instance, 'instability' predicts
287 significantly increasing temporal variation in dissimilarity for a set of sampling units because

288 relative increases and decreases among taxa are not correlated or consistent; change is not
289 directional although mean dissimilarity would also likely increase with the variance (Fig. 5d).
290 In contrast, 'stabilisation' predicts the opposite: significantly decreasing temporal variation in
291 dissimilarity as initially significant compositional change lessens over time; pairwise
292 dissimilarity is initially high at short temporal scales (the opposite of Fig. 5d).

293 *Taxon dynamics: Synchronicity*

294 In many cases, we are interested in asking if different components of the community are
295 responding in similar or different ways over time. Where individual taxa, or particular groups
296 of taxa (e.g., functional groups), change in similar ways at the same time ('synchronicity'),
297 we can detect this because among-taxon-group dissimilarity increases, but within-taxon-
298 group dissimilarity does not change (this involves comparing the pairwise sampling unit
299 dissimilarities separately for different taxonomic subsets) (Fig. 7a). Synchronicity may result
300 in directional or other types of compositional dynamics patterns within one sampling unit or a
301 set of sampling units (Fig. 7b).

302 Hypotheses of biological processes causing dynamics may be very specific, such as a
303 decrease in certain prey taxa in response to a predator, or more general, such as a shift
304 towards a grassland community in response to fire disturbance of a woodland or an increase
305 in annual plant taxa in spring compared to winter. More specific and directional (increases or
306 decreases) hypotheses lead to more powerful tests of compositional change; we strongly
307 encourage researchers to generate such hypotheses *a priori* and to pair these with predictions
308 that are as quantitative as possible, e.g., define the predicted timescales of variation.

309 **C. WORKED CASE STUDY EXAMPLES**

310 Here, we provide three worked examples to demonstrate the hypothesis-testing approach we
311 advocate for different types of community dynamics. We used open-access datasets to test
312 for: (1) significant divergence in temporal community dynamics among disturbance

313 treatments imposed upon plant communities in experimental grasslands, (2) neutral
314 community dynamics for a fully-censused forest plot, and (3) seasonal (cyclical) temporal
315 dynamics in marine aquatic bacterial communities. We present a brief description of each
316 dataset and describe the results and interpretation for the relevant exploratory data analysis
317 (EDA; see section 'D. Road map for analysing temporal community dynamics') and
318 hypothesis-testing analyses. More detailed descriptions, explanations and R code for each
319 case study are provided as supporting information (Data S2-4).

320 **TESTING FOR DIVERGENT COMMUNITY DYNAMICS WITHIN A CONTROLLED EXPERIMENT**

321 Porensky et al. (2016a, 2016b) measured plant species composition annually from 2007 to
322 2014 in quadrats on 24 transects to which six grazing treatments were applied since 1982
323 (four transects in each treatment): none (exclosure), light, moderate, heavy, heavy-to-light
324 (lightly grazed since 2007; 'new light') and heavy-to-none (new exclosure since 2007). They
325 calculated the mean cover in 25 quadrats for each species on transects. We tested the
326 hypothesis that increasing grazing intensity causes increasing divergence (directional change;
327 as detailed in section B) in species composition relative to the control (no-grazing) treatment.

328 EDA (as described in section D)) showed that the 24 transects were dominated across
329 all measurements by ten common species out of a total of 92 species (Data S2: row and
330 column summaries, relative abundance plots and rank abundance curves). Species turnover
331 (section D 4.0) through time was high; mostly 30% or greater for time period-treatment
332 combinations (Fig. 8). Principal coordinates analysis (PCoA; section D 1.4 Ordination)
333 suggested that the main compositional differences corresponded to the grazing treatments,
334 and that transect composition changed over time, but not markedly (Data S2: PCoA analysis
335 and figure). PERMANOVA (section D 1.4a) showed that transects were significantly
336 compositionally distinct and changed significantly over time in different ways (Data S2:
337 PERMANOVA outputs – all three coefficients for a two-way PERMANOVA with transect,

338 year and their interaction were significant $P < 0.05$). Partitioning spatial and temporal beta
339 diversity (section D 1.2) showed that two of the control (exclosure) treatments differed
340 significantly (Fig. 9). Principal Response curves (section D 1.4b) showed that the grazed
341 treatments differed from the control and that the relative abundances of four species in
342 particular were influential in driving these differences among treatments, but did not show
343 increasing divergence over time (Fig. 10). This result is consistent with the analysis of
344 temporal coherence (icc = -0.04; see section D 1.4a and Data S2 for detailed description of
345 this statistic), which showed that transects were not synchronous in their temporal changes
346 (Data S2: temporal coherence analysis). Thus, we conclude that treatment differences in
347 turnover and shifts in composition were not consistent or directional and so we reject our
348 hypothesis that increasingly grazed transects diverged more over time relative to the control.

349 **TESTING FOR NEUTRAL COMMUNITY DYNAMICS ACROSS SPATIAL SCALES**

350 The 50-ha forest census plot on Barro Colorado Island, Panama, has been censused eight
351 times between 1982 and 2015 (Condit 1998, Hubbell et al. 1999,
352 http://ctfs.si.edu/webatlas/datasets/bci/, 2005). This plot was established in part to test for
353 neutral community dynamics (Hubbell 1979) by fully censusing all woody stems greater than
354 1 cm diameter at breast height every five years. Under this hypothesis, we expected to see no
355 directional or other patterns in temporal community dynamics ('persistence'; as described in
356 section B.). EDA showed that the dataset consisted of 324 species generating a highly right-
357 skewed abundance distribution (most species were rare) (Data S3: EDA analyses and section
358 D). A line graph plot of species' abundance values over time showed very little change in
359 total abundance over time (Data S3: line graph of abundance; Appendix S2: Univariate linear
360 methods). Two analyses that allowed us to test for stability in community composition in
361 time, i.e., change in species composition is lower than expected if individual trees were
362 randomly distributed in time (keeping their spatial positions fixed), showed that change in

363 species composition was lower than expected: (1) Zeta diversity analysis (Data S3: zeta
364 diversity and section D 2.0) showed that the presence of species across census times was
365 more consistent than expected (Fig. 11) and (2) total turnover in species abundances (Data
366 S3: total turnover rates analyses and section D 4.0) at each of the seven time periods was
367 lower than expected (Fig. 12), particularly when the plot data were at larger spatial scales.
368 Thus, we conclude that the observed lack of pattern is consistent with our hypothesis of
369 neutral community dynamics.

370 **TESTING FOR SEASONAL COMMUNITY DYNAMICS**

371 Gilbert et al. (2012) measured bacterial taxonomic composition monthly over six years using
372 pyrosequencing of the 16S rRNA gene for water samples from a single site off the coast of
373 Plymouth, United Kingdom (source data can be obtained from the journal's online
374 supplementary file). We tested the hypothesis that subsets of the community would show
375 cyclical patterns of temporal change due to seasonality. EDA (section D) showed that the
376 dataset consisted of 74 operational taxonomic units (OTUs) across 62 temporal samples.
377 Most taxa were common through time. Proportional turnover between temporal samples
378 varied from 4% to 48% (Data S4: turnover calculation and section D 4.0), which is relatively
379 low, and was consistent with the detrended correspondence analysis (DCA) result showing a
380 short gradient length of 1.8 for axis 1 (Data S4: DCA and section D. EDA). Similarly, PCoA
381 graphs (section D. EDA) showed little consistent change within years (Fig. 13). However, 13
382 of the OTUs were deemed to be highly variable in that they changed by more than 500 reads
383 across the sampling period (mean number of reads = 296 ± 829 S.D..; Data S4: row and
384 column summaries, relative abundance distributions and abundance plots and section D.
385 EDA).

386 Time-lag analysis (section D 1.3) showed that although there was a significant
387 increase in dissimilarity within increasing time lag, variability was high, showing that

388 temporal changes did not occur in a simple divergence over time pattern (Fig. 14). Temporal
389 synchrony analysis (sensu Loreau and de Mazancourt 2008) showed that some subsets of
390 OTUs in the community were significantly changing in similar ways (Synchrony = 0.27; $P <$
391 0.05; Data S4: temporal synchrony analysis and section D 3.0). When patterns in these
392 variable taxa were examined through time using zeta diversity analysis (Hui and McGeoch
393 2014) (Data S4: zeta diversity analysis; section D 2.0), which is based on the presence of
394 species in samples, winter samples shared significantly more species across sets of temporal
395 samples than autumn, spring and summer (Fig. 15; ANOVA, $P < 0.01$; Appendix S2:
396 Univariate linear methods). Further analysis of the most variable taxa using the raw
397 dissimilarity values based on abundance data calculated for temporal samples within seasons
398 showed clear cyclical patterns (Fig. 16; Data S4; section D 1.1). Thus, we conclude that the
399 patterns are consistent with our hypothesis of seasonal dynamics in community composition,
400 largely driven by a small subset of OTUs that varied in their abundance, rather than presence,
401 over time.

402 **D. ROAD MAP FOR ANALYSING TEMPORAL COMMUNITY DYNAMICS**

403 The two most common sets of methods seen in our review of the temporal community
404 dynamics literature were descriptive (mostly tables and figures that show changes in taxon
405 presence or abundance) and ordination-based methods. Although these techniques are useful
406 for illustrating temporal trends and qualitatively exploring patterns in the data, they are used
407 less often for testing specific quantitative hypotheses of patterns and processes in community
408 dynamics. We recommend that researchers augment such descriptions with methods that
409 quantify temporal dynamics in taxonomic composition in more powerful ways, thus allowing
410 more quantitative and potentially comparable descriptions of spatiotemporal patterns and
411 explicit hypothesis-testing of mechanisms underlying observed patterns in community
412 dynamics.

413 Once the hypotheses to be tested have been generated, we suggest that researchers
414 undertake five steps: (Step 1) perform exploratory data analysis (EDA), (Step 2) choose
415 multivariate methods that can test the *a priori* temporal hypotheses, (Step 3) test for statistical
416 significance, (Step 4) report on the analysis workflow, and (Step 5) consider how to present
417 the results of the selected methods in the text and in figures.

418 **STEP 1: EXPLORATORY DATA ANALYSIS (EDA)**

419 As with any data analysis, the first step in the analysis of temporal community dynamics is an
420 exploratory data analysis (EDA) phase (Zuur et al. 2010). This generates basic information
421 that can be used to both error-check the data and to gain an understanding of key aspects of
422 the dataset (Table 1). R code for the relevant EDA steps is provided in three worked
423 examples, described previously (Data S2-4).

424 ***a. Count the number of taxa, the number of samples and the number of empty***
425 ***samples and missing data points***

426 The ratio of the number of taxa and samples is an important consideration because some
427 multivariate methods are more susceptible than others to the influence of rare taxa (Lepš and
428 Šmilauer 2003) and these may be of more or less interest, depending on the research
429 hypotheses. Whilst long-term data and monitoring are highly valued (e.g., Hobbs et al. 2007),
430 it is unfortunately common for long-term datasets to contain missing values; these values
431 should be denoted as missing (i.e., 'NA' in R) and not given zero or other numerical values.
432 The importance of missing values and how to deal with them during data analysis depend on
433 the number of missing data points and how they are distributed across the dataset.
434 Multivariate analyses based on compositional dissimilarities, e.g., some ordination methods,
435 cannot be applied to datasets containing samples where no taxa are present (McCune and
436 Grace 2002). However, empty samples may be important for analyses where we are

437 interested in understanding responses to disturbance and local extinctions. The ratio of the
438 number of explanatory variables to the number of independent replicates is also important for
439 temporal analyses because it affects the degrees of freedom available for statistical testing
440 (Johnstone and Titterington 2009).

441 ***b. Examine taxon frequency distributions and response curves***

442 Some multivariate analyses assume specific frequency distributions of taxa (a histogram of
443 the number of individuals or their relative abundance across samples for each taxon), e.g.,
444 CCA assumes that taxa have optima along environmental gradients ('unimodal' responses).
445 Examining the shapes of taxon frequency distributions is necessary for determining whether
446 taxa are responding strongly to environmental gradients in the dataset and whether their
447 distributions conform to analysis assumptions. Where many taxa are involved, the common
448 ones should be examined.

449 ***c. Determine the ratio of temporal to spatial variation in composition***

450 The degree of spatial and temporal replication is important in determining the method of
451 analysis because few sites with many measurements over time would be analysed differently
452 from many sites measured only a few times. If the data have a nested structure, e.g., samples
453 are nested within sites, this needs to be accounted for in the analysis, particularly if one wants
454 to report significance values. For example, if plant species composition was recorded in
455 quadrats that were nested within transects that were nested within sites, then quadrats are not
456 fully independent replicates. Similarly, information from technical replicates (such as PCR
457 products) should be kept separate at this stage because they provide useful information about
458 variability at different scales in the dataset. It is inadvisable to average or pool data until the
459 significant sources of variation are understood. For instance, if we want to test hypotheses of
460 temporal dynamics, but there is a high degree of spatial variation among samples in the

461 dataset, this can alter analysis choices or how results are presented and interpreted (Thrush et
462 al. 2008). The relative amount of change in taxonomic composition can be estimated from the
463 differences in sample scores on DCA ordination axes, e.g., the 'gradient length' statistic
464 quantifies the turnover in taxonomic composition (Lepš and Šmilauer 2003); this relatively
465 descriptive analysis would ideally be conducted as part of the EDA process prior to selecting
466 an analysis pathway appropriate for testing a quantitative hypothesis of temporal dynamics,
467 such as divergence or convergence.

468 ***d. Determine the optimal dissimilarity measure***

469 A community dataset may be an abundance matrix (number of individuals), relative
470 abundance matrix (e.g., percent cover, sequence abundances), rank-abundance matrix
471 (ordinal abundance ranks), or a taxon-occurrence (presence-absence) matrix. These different
472 response variables can influence the choice of analysis methods, including which distance
473 measure to use to calculate pairwise community dissimilarity values for samples (see
474 Magurran 2004, or Legendre and Legendre 2012 for a comprehensive description). To ensure
475 that abundance data from different taxa are comparable using dissimilarity, data should be
476 standardised prior to analysis if they are measured on different scales (Legendre and
477 Legendre 2012). For example, taxon abundance values within each sample can be divided by
478 the total abundance within the sample, or can be standardised across all samples so that each
479 taxon has a mean of zero and standard deviation of one (Borcard et al. 2011). Alternatively, a
480 distance measure that incorporates a standardisation can be used, e.g., Hellinger distance
481 (Legendre and Gallagher 2001, Anderson et al. 2011, Legendre and Legendre 2012). The
482 commonly-used Euclidean distance measure is not usually appropriate for taxon abundance
483 data, which are often characterised by non-linear relationships among taxa, abundance values
484 measured on different scales, or by the 'double-zero problem', which is the idea that samples
485 that are missing the same taxa and therefore, have lower dissimilarity when using this

486 distance measure (Borcard et al. 2011). Comparing results based on presence-absence data
487 with those generated using relative abundance data allows us to evaluate the role of taxon
488 abundances in generating temporal patterns (e.g., Angeler and Johnson 2012). Estimating
489 abundances of each population within the community has its own important considerations,
490 i.e., variation in taxon detection probabilities, which have been addressed elsewhere (e.g.,
491 animals: Gaston and McArdle 1994, metagenomics data: Amend et al. 2010). Many distance
492 measures cannot be computed if there are empty samples or missing values. Deleting entire
493 rows or estimating missing data so that the dataset can be analysed may be an option in some
494 cases (McCune and Grace 2002). However, other measures, such as the Gower coefficient,
495 can be calculated while ignoring comparisons among samples with missing data (Legendre
496 and Legendre 2012; coefficient S15: 278).

497 **STEP 2: CHOOSE ANALYSES FOR TESTING HYPOTHESES OF TEMPORAL DYNAMICS**

498 In this section, we describe a range of methods for testing hypotheses of community
499 dynamics, including their outputs, highlighting in each case, the relevant ecological questions
500 and the kinds of hypotheses that those analyses can be used to address: pairwise dissimilarity-
501 based methods (section 1, below), zeta diversity (section 2), synchrony (section 3), turnover
502 rates (section 4) and multivariate regression modelling (section 5). Further, each of the
503 described methods has been cross-referenced with (i) which hypotheses it is best suited to
504 test; (ii) whether or not it allows the identification of individual taxon effects on, or responses
505 to, temporal change; (iii) whether or not explanatory and spatial variables can be explicitly
506 incorporated; (iv) the type of statistical testing normally used; and (v) any requirements for
507 the length of the time series (number of temporal samples) (Fig. 17). The selection of an
508 appropriate analytical method also can be influenced by the kind of output desired (Table 2).
509 For instance, some analyses result in outputs that require more effort to interpret than others,
510 whilst others show differences in taxon responses or generate outputs that can be added to a

511 geographic map. Further, as our literature review highlighted, in many cases several different
512 analytical methods often are appropriate for asking questions about temporal dynamics and
513 presenting the results at the desired level of complexity; it is also the case that many analyses
514 can be used to test a wide range of hypotheses, so our focus is on describing the strengths of
515 each analysis method and the minimum requirements for use, rather than on making explicit
516 recommendations. Thus, we recommend that, at this point, readers use Tables 2 and 3 to help
517 identify the methods of interest and then refer to the relevant sub-sections of section C that
518 describe those particular methods.

519 Our literature review revealed that in numerous cases, researchers performed and
520 compared separate analyses on their community matrices from each sample time, e.g.,
521 separate nestedness analyses at each sample time (Azeria et al. 2006); we do not specifically
522 address these applications here in any depth because these methods do not explicitly evaluate
523 community dynamics through time. We also focus on methods that are truly multivariate
524 rather than methods that are applied separately to abundances of individual taxa; however, if
525 less than ten taxa are of interest, it may be that hypothesis testing via univariate approaches
526 (e.g., separate linear regressions of taxon abundances) will be adequate for drawing synthetic
527 conclusions about community change. In the supplementary material (Appendix S2), we
528 summarise univariate methods that were commonly used in the literature review articles:
529 univariate linear methods, temporal stability (*a.k.a.* coefficient of variation), frequency
530 change, multiplicative change, and Markov chain modelling. A few studies used alternative
531 methods such as cluster analysis and network analysis; however, we excluded those because
532 we think the methods presented here provide equally good, or better, alternatives, but we
533 nonetheless provide an explanation of them (Appendix S2) because they appeared in our
534 literature review. Some of the analytical methods described below can be implemented in
535 several different software packages (e.g., PC-ORD, PRIMER, CANOCO), but, for simplicity

536 and widest accessibility, we illustrate their implementation in R (Data S2-4), except for
537 Moran's Eigenvector Maps (MEMs), which is given in Legendre & Gauthier (2014), and
538 Multivariate Autoregressive State Space modelling (MARSS), which is presented in the
539 documentation associated with the R package 'MARSS' (Holmes et al. 2012).

540 **1. Methods based on pairwise dissimilarity values**

541 Pairwise dissimilarity methods can be used to test nearly the full range of questions and
542 hypotheses in community dynamics (Anderson et al. 2011; Figs. 4-7). Pairwise dissimilarity
543 methods are the most frequent way ecologists quantify the difference in the identities of taxa
544 occurring in two samples. The result is a single number that measures the distance in
545 compositional space between the two samples. Dissimilarity can also be represented as
546 similarity, i.e., 1 − dissimilarity (Whittaker 1960, Anderson et al. 2011, Magurran 2013).
547 When applied to multiple samples, it results in a matrix of all possible comparisons (a
548 triangular dissimilarity matrix). For example, Collins (1992) measured the between-year
549 percent change in plant species composition for grassland plots that had been repeatedly
550 measured over a nine-year period. Dissimilarity methods are more desirable than univariate
551 methods, which analyse the abundance of each taxon separately, because they consider the
552 entire community at once and directly quantify change in community structure.

553 The development of dissimilarity indices is an active area of research. Magurran
554 (2013) is a useful general discussion of dissimilarity measures, Anderson et al. (2011)
555 provides a practical guide to selecting an appropriate index and issues of importance, such as
556 whether to include joint absences in the calculation. Different authors often use different
557 names for the same dissimilarity measure, so we recommend using a comprehensive and
558 consistent guide, such as Legendre and Legendre (2012; Table 7.4: 324). Legendre and De
559 Cáceres (2013) also provide comprehensive descriptions of most dissimilarity measures and

560 rationales for choosing one. Several recent advances are specifically relevant to temporal data
561 (e.g., Podani and Schmera 2011, Baselga 2012, Shimadzu et al. 2015) and some contributions
562 aid in improving dissimilarity measures for use in certain circumstances, such as imperfect
563 detection of taxa (e.g., Chao et al. 2006) or which indices are most appropriate for calculating
564 dissimilarity (Legendre and De Cáceres 2013). For example, the unique fraction (UniFrac)
565 dissimilarity metric was designed for microbial communities where phylogenetic information
566 is available from sequencing that has been used to determine taxa in samples (Lozupone and
567 Knight 2005, Lozupone et al. 2007). This metric measures phylogenetic distance between
568 samples as the fraction of branch lengths leading to descendants from each sample. UniFrac
569 can be used with abundance data, which weights the branch lengths prior to calculating
570 distances (weighted UniFrac), or with presence-absence data (unweighted UniFrac Lozupone
571 et al. 2007).

572 In the context of temporal community dynamics, dissimilarity is calculated in two
573 different ways: (1) separately among spatial samples at each of a series of time points or (2)
574 among samples taken at different times at the same location(s). Dissimilarity matrices are the
575 basis for several methods recommended in our guide, including ordination analyses and
576 Moran's Eigenvector Maps (MEMs). It is important to realise that it is not easy to use such
577 methods to simultaneously assess both temporal and spatial variation in the same analysis;
578 however, temporal and spatial dissimilarity can be computed and assessed in parallel to test
579 hypotheses and make inferences about spatiotemporal community dynamics (e.g., Legendre
580 and Gauthier 2014, Lamy et al. 2015).

581 *1.1. Using the raw dissimilarity values*

582 There are two ways in which raw dissimilarity values for samples are typically used.
583 Either of these approaches could be used for different taxonomic subsets of the community to
584 test hypotheses of taxonomic variation in dynamics. First, compositional dissimilarity can be

585 calculated among samples taken at different locations to obtain a measure of spatial
586 heterogeneity in community composition. To use these values to analyse temporal
587 community dynamics, pairwise dissimilarity (spatial dissimilarity) must be calculated
588 separately for each sampling time, and the means of these spatial dissimilarity values can
589 then be compared across time visually in a table (e.g., Collins 1992), graph (e.g., Barros et al.
590 2014), or using a correlation or regression analysis where time is the predictor variable and
591 amount of spatial variability at each site is the response (e.g., Barros et al. 2014). The mean
592 dissimilarity can be divided by the number of taxa to generate an independent measure of
593 compositional change (Cleland et al. 2013). Such comparisons over time could be done
594 separately for different groups of samples, such as different experimental treatments, to test
595 for predictors of differences in temporal change in composition. Bagchi et al. (2017) illustrate
596 predictions of how temporal differences in dissimilarity can include non-linear patterns and
597 thus how applying linear and non-linear regression modelling to pairwise dissimilarity values
598 can be used to test more complex hypotheses of temporal dynamics. Regression approaches
599 can also include mixed-effects models that account for nesting and other types of non-
600 independence in the data, or non-linear models (e.g., generalised additive models) that may
601 be useful to predict cyclical or threshold patterns in community dynamics (e.g., Shenhav et
602 al. 2017). However, these modelling methods will apply well only in situations with longer
603 time series (i.e., greater than 25 temporal replicates) and the more complex the prediction of
604 temporal change, the greater replication is likely required. Alternatively, a Mantel test can be
605 used to compare pairs of dissimilarity matrices computed at different times and to test for
606 significant difference over time (see section in the main text 'Test for statistical
607 significance'). This approach is only useful for a small number of spatiotemporal
608 observations (i.e., ten or fewer) because the comparisons are pairwise.

609 Second, if a set of samples were all taken at the same site at different times, 'temporal
610 heterogeneity' can be calculated using the mean dissimilarity of the entire sequence of
611 temporal samples to obtain a single measure of within-site temporal variability (Collins and
612 Smith 2006). If multiple sites are sampled through time, this method can be applied at
613 different spatial scales (Collins 1992, Collins and Xia 2015) to examine spatial scale-
614 dependent temporal dynamics (Collins and Smith 2006). Regression approaches that relate
615 mean temporal dissimilarity values to other explanatory variables for sites or samples could
616 be used to explore spatial variability in temporal change. Alternatively, the 'dissimilarity-
617 overlap' method, first described by Bashan et al. (2016) and further explained by Kalyuzhny
618 and Shnerb (2017), can be applied to raw dissimilarity values to test for similarity among
619 samples in community dynamics. This method uses the relationship between the pairwise
620 dissimilarity values and the proportion of individuals in pairs of samples that overlap
621 taxonomically to detect if changes in taxonomic composition (degree of overlap) alters
622 relative abundances in a consistent way across samples (dissimilarities). It can be used for
623 very large numbers of samples and taxa and could be applied to different groups of taxa to
624 test for synchronicity.

625 *1.2. Beta diversity decomposition*

626 'Beta diversity' is term used for several different types of measures in ecology, e.g.,
627 (Whittaker 1960, Baselga 2010, Legendre and De Cáceres 2013), including as a synonym for
628 compositional dissimilarity. It always describes some measure of how taxonomic
629 composition changes across a set of samples. A number of recent advances in the calculation
630 of beta diversity have illustrated that, if calculated in particular ways, it can be 'partitioned'
631 into different components that relate to different predictions of how taxa or their abundances
632 change in space or time (Baselga 2010, Ellison 2010, Legendre and De Cáceres 2013). Beta
633 diversity decomposition methods can be applied to testing almost any hypothesis of

634 compositional change (Fig. 17). The outputs are focussed on species replacement and
635 whether samples are subsets of each other (Table 2), so detecting patterns such as
636 convergence or divergence is very easy. However, depending on what subsets of the data are
637 used in separate calculations, beta diversity decomposition is equally able to detect rapid
638 shifts in composition, e.g., perturbation, tipping point, or increases or decreases in temporal
639 variability, e.g., 'instability' or 'stability'. Two approaches so far have been usefully applied
640 to quantifying community dynamics measured in the field (Baselga 2010, Baselga and Orme
641 2012, Legendre and De Cáceres 2013, Lamy et al. 2015).

642 First, when among-sample dissimilarity is measured as Sørensen's dissimilarity [(b +
643 c)/2a + b + c), where a is the number of shared presences and b and c are the number of
644 unshared presences in each of the two samples], the Simpson's dissimilarity index (also
645 known as 'species turnover': b / (b + a)) can be subtracted from this value, resulting in a
646 measure of 'nestedness' (Baselga 2010, 2012, Baselga and Orme 2012). This measure of
647 nestedness reflects the degree to which the species composition of smaller samples (lower
648 species richness) is encompassed by the species composition of larger samples (higher
649 richness); samples may differ in space or time. In contrast, 'species turnover' measures the
650 degree to which samples do not share species: how the identities of species differ among
651 samples. It is useful to separate these two components (nestedness and turnover) to obtain a
652 more informative quantification of beta diversity that reflects differences in the underlying
653 processes. Such separation can be done either pairwise or for multiple sets of samples
654 (Baselga 2010, 2017), and has been used to decompose temporal beta diversity in tests of
655 hypotheses unique to temporal data (e.g., Baselga and Orme 2012).

656 Second, and in contrast, Legendre & De Cáceres (2013) defined beta diversity as the
657 sums of squares in the variation of taxon abundance (or presence) corrected for the number of
658 samples. They partitioned this measure of beta diversity into two different components:

659 ‘Species contribution to beta diversity’ (SCBD) and ‘local contribution to beta diversity’
660 (LCBD), which describe, respectively, the relative contributions of different taxa and sites to
661 overall beta diversity (Table 2; Legendre and Gauthier 2014). This method allows for tests of
662 predictions of patterns in temporal community dynamics related to different taxa in the
663 dataset. For example, Lamy et al. (2015) decomposed beta diversity (based on Bray-Curtis
664 dissimilarity) of fish communities at each of 13 reef sites into contributions from biomass
665 replacement and biomass differences over time. They demonstrated that total biomass in the
666 community changed little, but that the observed significant temporal changes in fish
667 composition resulted from differential changes in the relative biomass of species.
668 Spatiotemporal patterns in the two components of fish community beta diversity were
669 visualised using principal coordinates analysis (PCoA).

670 *1.3. Time-lag analysis*

671 A time-lag analysis involves relating the amount change in composition to the amount of
672 change in time across increasing temporal distances, called ‘lags’ (Table 2). For example,
673 Collins & Xia (Collins and Xia 2015) compared plant species composition of repeatedly-
674 measured grassland transects across all time lags between 1989-2008 (from a 1- to 19-year
675 difference). There are two ways of assessing the time-lag effect on community dissimilarity:
676 graphically (‘time-decay curve’) and statistically (time-lag regression analysis). For
677 meaningful inference, time-lag graphs require at least five samples and statistical time-lag
678 analysis requires at least ten samples for each temporal sequence. For both approaches, the
679 pairwise compositional dissimilarity (temporal beta diversity) among temporal samples is
680 calculated first.

681 To generate a time-decay curve, pairwise temporal dissimilarity values are plotted
682 against temporal distance between samples (time lag; pairwise Euclidean distances in sample
683 times) to show how community dissimilarity changes as samples become further apart in time

684 (Collins 2000, Collins et al. 2000, Hallett et al. 2016). This graph can be then used to visually
685 assess whether community composition is converging (becoming more similar) or diverging
686 (becoming less similar) over time (e.g., Collins 2000). For example, Bar-Massada & Hadar
687 (2017) plotted temporal beta diversity (as Bray-Curtis dissimilarity) of plants in grazed and
688 ungrazed treatments in Mediterranean herb communities, revealing that composition diverged
689 over time, regardless of grazing treatment. Alternatively, temporal community dissimilarity
690 can be graphed against dissimilarity in environmental variables (rather than time), if the
691 primary research focus is on response of the community to environmental change rather than
692 time *per se*.

693 Time-lag regression analysis was developed specifically for data with greater than ten
694 time points to assess directional, cyclic or stochastic changes within communities (Collins
695 2000, Collins et al. 2000) and has been evaluated against other temporal analysis methods
696 (Angeler et al. 2009). For time-lag regression analysis, the pairwise compositional distances
697 are plotted against the square-root of the time lag, which reduces the bias that results from
698 having fewer data points at larger time lags (Collins 2000, Collins et al. 2000). A regression
699 line is then fitted to determine the pattern of compositional change. A significantly positive
700 relationship indicates that compositional changes in communities is directional, because
701 sampling points that are further apart in time are more dissimilar than those that are closer
702 together in time (i.e., Fig. 2 in Collins 2000). A negative slope shows that some convergence
703 towards the taxonomic composition at one of the earlier time points has occurred, because
704 measurements close together in time are more dissimilar in composition compared to those
705 further apart in time. No slope, or a non-significant regression, implies stochastic changes in
706 the community over time. Non-linear regressions also can be used to describe more complex
707 temporal patterns (Magalhães et al. 2007).

A robust measure of significance in time-lag regression analysis can be determined by running Mantel tests on the dissimilarity matrix and the matrix of pairwise distances among samples in time, with Monte Carlo permutations (Bêche and Resh 2007, Day et al. 2017). Time-lag regression analysis is appealing because it is intuitive, produces one statistic (the regression coefficient), and can be easily visualised (e.g., freshwater macroinvertebrates: Bêche and Resh 2007, gut bacteria: Dimitriu et al. 2013, and plants: Collins and Xia 2015). However, time-lag analysis only provides one value for each sampling unit, so further analyses may be required to investigate changes within treatments or groups.

1.4. Ordination

Ordination summarises multivariate community data by optimising relationships between high-dimensional samples and taxa in low-dimensional space to detect important ecological gradients in communities (Table 2). This dimensionality reduction means that they are not especially useful for directly testing, or visualising, complex changes in community composition over time, such as tipping points, cyclical patterns or synchronicity. However, more simple patterns, such as divergence or convergence, can be more easily detected. Ordinations are either unconstrained, meaning that they do not incorporate explanatory variables, or constrained, which assess the effect of explanatory variables, including time (e.g., Kent 2012, Legendre and Legendre 2012, Palmer 2019). Ordination methods, including constrained ordination and using the ordination scores from an unconstrained ordination in a regression model (e.g., general linear model, mixed-effects model) are only appropriate for testing hypotheses of community dynamics in cases where EDA shows that the ordination axes used strongly represent temporal compositional change and therefore explain a substantial proportion of the variation. This is because the amount of temporal compositional variation represented by the ordination axis scores will be low if there is not a strong temporal gradient along the axes. Ordination methods allow only weak inference of the

733 relationship between taxa and community dynamics if the temporal pattern is complex
734 relative to the ordination axes because some ordinations, e.g., DCA, only provide an
735 ordination axis score for taxa, and others do not provide 'species scores' at all, e.g., PCoA
736 and NMDS. PCoA can be used to calculate the amount of variance explained by the
737 ordination axes, which is important if a time-series analysis is to be performed using the axis
738 scores. Where species scores are not provided, weighted averaging can be used to plot points
739 for each taxon on the ordination diagrams to relate differences in site composition to
740 differences in taxon presence or relative abundance if you wish to examine the associations
741 between samples and taxa (McCune and Grace 2002). One advantage of PCoA is that the
742 first PCoA axis represents the gradient explaining the greatest amount of variation in
743 composition, followed by the second axis that explains the next greatest amount, and so on;
744 this is not the case for NMDS (McCune and Grace 2002). Therefore, it is recommended that
745 NMDS scores are not used in community dynamics analyses unless the amount of variation
746 explained by each axis is clearly stated in order to validate their use in further analyses. 'Co-
747 inertia analysis' can be coupled with ordination in a 'partial triadic analysis' to investigate
748 community dynamics in relation to environmental variables. This method may be particularly
749 useful when testing for temporal stability (Kidé et al. 2015).

750 a. Using ordination site scores to assess compositional change

751 Ordination axis scores for samples taken at different times can be used to quantify and test
752 hypotheses of compositional change. Ordination diagrams are visually appealing, but for
753 illustrating compositional change through time, it is difficult to interpret the magnitude and
754 significance of temporal dynamics without statistical testing, such as permutational analysis
755 of variance (PERMANOVA: Anderson 2001), unless temporal trends in the data are very
756 strong along the ordination gradients in composition. This difficulty is amplified if there are
757 many samples or if the sampling design includes a large number of sites, environmental

factors, or temporal replicates. Note that PERMANOVA requires the assumption of homogeneity of dispersion to be met (this can be tested using the "betadisper" function in the R package vegan; Oksanen et al. 2017). Further, ordination diagrams and their statistics computed for different matrices (different sets of taxa or sites) are not comparable; results of individual ordinations cannot be compared among them. The partial exception to this is Procrustes analysis, which can measure the distortion, or 'lack of fit' between two ordinations, or dissimilarity matrices (Jackson 1995, Peres-Neto and Jackson 2001). It can be used to identify which taxa are most important in differentiating between the two ordination patterns.

A more quantitative, and often more flexible, approach is to use the ordination axis 'site' scores in linear models with time and other explanatory variables of interest entering as predictor variables (e.g., Berthon et al. 2014, Broadway et al. 2015). For example, Cotton et al. (2015) used DCA scores as a response variable in a linear mixed model to investigate the importance of carbon dioxide, ozone, and time on arbuscular mycorrhizal fungal community composition. This is like doing a constrained ordination, such as RDA, dbRDA, or CCA, where time is used as a constraining explanatory variable in the ordination; these methods will test the significance of time and any other explanatory variables. However, it is more difficult within a constrained ordination to test for non-linear, cyclical or other complex temporal patterns, and in this case, using the ordination axis scores in a more complex model may be more useful.

Ordination axis scores for samples can be used to quantify temporal changes in other ways. For example, Quinn et al. (1991) developed a 'malleability index' by regressing the pairwise Euclidean distances between the first three dimensions of ordination space for temporal samples, against environmental variables. One potential problem with this method is caused by the tendency of ordinations to distort the relative distances (dissimilarity)

783 between samples, depending on, for example, differences in the number of rare taxa, and so
784 values from distance-based indices may not be comparable, even in different parts of the
785 same ordination.

786 Temporal coherence measures the degree to which a response variable is similarly
787 variable in time across a set of sites. In the case of measuring temporal coherence in
788 community composition, the multivariate response variable is taxonomic composition in the
789 form of ordination axis site scores (Angeler and Johnson 2012). Temporal coherence
790 compares the amount of temporal variation in the ordination axis site scores across a set of
791 sites by using statistics generated by an ANOVA of the site scores (using year and site as
792 factors) to calculate the 'intra-class correlation coefficient', which ranges from $-1/(n-1)$ to
793 $+1$, where n is the number of sites sampled, given that $n > 1$ (for calculation details see Zar
794 1996 pp 398, Rusak et al. 1999). High positive values of this metric indicate increased
795 coherence of the different sites in composition over time. The measure can be recalculated
796 while dropping each time in sequence (an ordered jackknife) to determine the relative effect
797 of each time on the overall value (Rusak et al. 1999). Different sets of sites or treatments can
798 be compared using a Fisher's Z test (Angeler and Johnson 2012) or randomisation test in the
799 case of low sample sizes (Rusak et al. 1999).

800 b. Principal Response Curves (PRC)

801 The method of principal response curves (PRC) was first developed by van den Brink & ter
802 Braak (1999) for comparing compositional change in treatment sites relative to a control site.
803 It is similar to redundancy analysis (RDA) in that a multivariate regression is performed on
804 the sample × taxon matrix constrained by variables of interest, but PRC is explicitly
805 constrained by time (Lepš and Šmilauer 2003, ter Braak and Šmilauer 2015). It was
806 developed to eliminate the need for cluttered biplots that make it difficult to see temporal
807 trends within and between treatments. To overcome this problem, PRC uses a single axis,

808 plotted against time, to express trajectories of compositional change (Lepš and Šmilauer
809 2003: 145); further axes can be calculated if they are of interest. Taxon weights also can be
810 calculated and plotted for each PRC to show the change in the relative abundance (e.g.,
811 percent cover) of each taxon in the treatments relative to the control. The significance of the
812 PRC is tested using Monte Carlo simulations (Van den Brink and Ter Braak 1999). As with
813 RDA, PRC yields a measure of the percent variance explained for the overall taxonomic
814 composition and how much is attributable to each variable (treatment and time). For example,
815 Sugiura et al. (2013) used PRC to show that plant invasion in forests on a small island in the
816 north-western Pacific had little effect on the temporal dynamics of invertebrate communities.
817 The potential drawbacks of this method include its requirement for a 'baseline' or 'control'
818 treatment level to be set for comparison, an attribute that many observational datasets will not
819 have. Furthermore, PRC allows for only one 'treatment' variable and each analysis focusses
820 on the change at one site, relative to the control site. However, Auber et al. (2017) have
821 recently generalised this method so that it can be used to analyse compositional change at
822 multiple sites, relative to a baseline or control, but this application removes the ability to
823 consider an environmental or treatment effect. Thus, a principal response curve analysis
824 yields a line graph showing the relative change in taxonomic composition for each
825 'treatment' along one compositional gradient relative to a control (Table 2).

826 *1.5. Moran's eigenvector maps (MEMs)*

827 Moran's eigenvector maps (MEMs; also known as 'principal coordinates of neighbour
828 matrices', PCNM) is a method designed to characterise the spatial scales of variation in
829 univariate or multivariate data (Borcard and Legendre 2002, Dray et al. 2006, Legendre and
830 Gauthier 2014). However, MEMs have recently been applied to detect temporal scales of
831 variation in communities (Angeler et al. 2009, Caruso et al. 2013, Legendre and Gauthier
832 2014). For instance, if a cyclical pattern in compositional change, such as seasonality, was

833 detected as significant for a dataset, this can be illustrated using line graphs showing the
834 temporal pattern (Table 3; Legendre and Gauthier 2014). This method is particularly useful
835 for detecting very complex temporal patterns, such as seasonality (periodicity), cyclical
836 changes or tipping points, but requires relatively large temporal sample sizes (Fig. 17). For
837 example, Kampichler et al. (2014) used MEMs to detect temporal patterns in multiple
838 datasets that had at least 20 annual measurements of bird communities.

839 First, a distance matrix is generated. Then, to focus on neighbour relationships, a
840 threshold value is chosen (see Borcard et al. 2011 for advice). For values below this
841 threshold, the distance value is retained; values above this threshold are assigned an
842 arbitrarily large value. Legendre & Gauthier (2014) recommend using one time interval as
843 the threshold for regular time series or the length of the longest lag for irregular series. A
844 PCoA is then applied to this modified distance matrix to generate a large number of
845 eigenvectors that describe a wide range of temporal scales of variation.

846 The eigenvectors from the PCoA can be used as explanatory variables, most often in a
847 redundancy analysis (RDA) or multiple regression. Combined with RDA, MEMs can be used
848 to determine whether there is cyclic or synchronised temporal change in composition and
849 which taxa, or taxon groups, are the drivers of it (Fig. 17). MEMs-RDA scores also can be
850 plotted as a function of time to assess visually the intervals of fluctuating community
851 dynamics (Angeler et al. 2009, Kampichler et al. 2014). In a comparison of different methods
852 to assess community dynamics, Angeler et al. (2009) observed that MEMs-RDA consistently
853 explained a greater proportion of variation in data with low taxonomic richness and linear
854 changes in time relative to time-lag regression analysis, Mantel tests, or RDA. However,
855 power was higher for taxon-rich communities because there was a greater probability of more
856 taxa fitting the temporal models. As a scale-detection method, MEMs could be applied to
857 determining the relevant time intervals for community monitoring. This method can be

858 combined with variance partitioning to show the relative variance in composition explained
859 by temporal, spatial and other explanatory variables (Legendre and Gauthier 2014).

860 *1.6. Compositional pivot days*

861 Compositional pivot days is a dissimilarity-based method developed by, and unique to,
862 Lellouch et al. (2014). This method is appropriate for situations where we want to test for the
863 occurrence of fast changes in taxonomic composition across one point in time, such as the
864 application of an experimental treatment or the effects of a rapid change in environmental
865 conditions (Table 2). It works by comparing pairwise distance values for pairs of temporal
866 observations taken at the same sample location to identify the time points where rapid shifts
867 in composition occurred. A 'pivot day' is a sampled time point that is identified by pairwise
868 similarities being low before and after that time point, but relatively higher across it.
869 Compositional pivot days are identified either visually from the pairwise distance matrix plot,
870 or using clustering (see Lellouch et al. 2014 for details). This method may be a good option
871 for a preliminary identification of thresholds or rapid shifts in community structure, but does
872 not provide a statistical test for the significance of the threshold relative to background. The
873 power of this analysis will increase as the number of sampled time points increases relative to
874 the number of pivot days.

875 *1.7. Community trajectory analysis*

876 Community trajectory analysis quantitatively compares the temporal change in multivariate
877 compositional dissimilarity space among samples using geometric calculations (de Caceres et
878 al. (2019). These comparisons are based on the relative amounts, rates, and directions of
879 sample movement in compositional space, thus, the method allows flexible testing of a wide
880 range of very specific, quantitative hypotheses of change, such as the divergence or
881 convergence of samples over time. This method is especially useful for a small number of

spatial replicates because the output can be graphical; plots use arrows on an ordination diagram to show how samples have moved in compositional space over time; however, larger numbers of samples could be graphed separately. We do not illustrate this method here because comprehensive R code and explanations are presented in the R package 'vegeclust' (De Cáceres et al. 2010) and demonstrative examples are given in de Caceres et al. (2019).

2. Zeta diversity

Zeta diversity, a measure of similarity, analyses change in community composition among samples in the dataset by calculating the average number of taxa that are shared between samples that are grouped in pairs, triples, quadruples, etc. (Table 3). The method was developed and introduced by Hui & McGeoch (2014), extended to temporal data by McGeoch et al. (2017), and comes with its own R package, 'zetadiv' (Latombe et al. 2017b). Zeta diversity estimates compositional turnover from multiple ($n \geq 2$) samples, rather than the pairwise comparisons used to estimate beta diversity. What is obtained by the method is a series of zeta values (similarity) for different orders of zeta: zeta_1 to zeta_n, where n is the total number of time points. Each of these values represents the average of the number of taxa shared by each set of times in the dataset. For example, zeta_2 (zeta order 2) represents the average number of taxa shared by all possible pairs of samples in the dataset (equivalent to 1 − beta diversity or mean similarity), zeta_3 represents the average number of taxa shared by all possible subsets of three samples, and so on. Zeta_1 (zeta order 1) represents the average alpha diversity (taxon richness) of all the samples in the dataset.

If there is complete turnover in community composition in time, at some point, zeta_i (where $i \neq 1$) will equal zero when there are no taxa shared by a given sized subset of samples. In contrast, if community composition never changes through time, $\text{zeta}_n = \text{zeta}_1$. Therefore, the rate of decline in zeta with increasing 'zeta order' tells us about the rate of

change in community composition in time (‘zeta decline’). ‘Zeta decay’ can be considered analogous to time-lag graphs and time-lag regression analyses, which we described earlier for pairwise dissimilarity measures (Table 3). This means it experiences the same issue of few time points at higher lags, and thus higher uncertainty. The temporal scale(s) of community dynamics can be illustrated by plotting the change in sample zeta diversity of a particular zeta order against the temporal lag of samples, meaning that this method is useful for detecting both directional changes and cyclical patterns. For instance, if a multi-year time series in composition is cyclical (e.g., shows seasonality), zeta would seasonally decrease and then increase. These analyses can be done for different sets of samples to make spatial or temporal comparisons and, if necessary, zeta values can be normalised using the gamma diversity (total number of taxa in the dataset) to account for differences in taxon richness across samples (McGeoch et al. 2017). Further, different methods of creating the subsets for calculating zeta can be used, such as nearest neighbours or relative to a time-zero sample, further facilitating hypothesis testing (McGeoch et al. 2017). Latombe et al. (2017a) used zeta diversity to extend generalised dissimilarity modelling (Ferrier et al. 2007). This enables the user to assess taxonomic turnover at different orders of zeta and understand drivers of turnover in rare and common taxa (Latombe et al. 2017a). This method could be extended to assess temporal turnover at different orders of zeta by adding time as an environmental variable.

3. Synchrony

Although other methods can be performed on taxonomic subsets to look for similarities and differences in the temporal dynamics among groups of taxa (“synchronicity”), a useful measure of the ‘temporal synchrony’ in multivariate taxon abundances was developed by Loreau & Mazancourt (2008) and comes with its own R package, ‘synchrony’ (Gouhier and Guichard 2014). This standardised value measures how closely changes in taxon abundances

931 within a community track one another in time. It varies between zero (complete asynchrony
932 of taxon abundances in the community) and one (perfect synchrony; all taxon abundances
933 changing in the same way) and, by exploring different sized 'time windows', can be used to
934 detect scales of temporal variation. Overall significance tests of synchronicity in community
935 change can be obtained by performing null model analysis (e.g., Pedersen et al. 2017).

936 A similar method (Houlahan et al. 2007) uses the sum of the temporal variances and
937 pairwise covariances of all taxa in the community to distinguish if the community shows
938 'compensatory dynamics' over time. This method results in either a positive, negative or zero
939 community-level covariance value, indicating that taxa are increasing or decreasing in
940 synchrony (positive community covariance), some taxa are increasing, but this is
941 compensated for by decreases in other taxa (negative community covariance), or that no
942 change is evidence (zero community covariance) (Houlahan et al. 2007, Hallett et al. 2016).

943 **4. Turnover Rates**

944 The degree of temporal variation in community composition can be assessed by calculating
945 the turnover rate of the community (a.k.a. 'temporal turnover' or 'species turnover'), which is
946 a measure of the rate of change in taxonomic composition for a 'site' over time (e.g., Aguirre
947 et al. 2003). Turnover rates can be computed to measure the magnitude of compositional
948 change, but in a way that does not result in an understanding of the direction of change or the
949 taxa involved in compositional shifts. This method can detect when, and how rapidly,
950 temporal shifts occur in a temporal sequence to detect complex temporal patterns such as
951 perturbations, tipping points and pivot points (Fig. 17). Turnover rates are calculated using
952 combinations of colonisation, immigration, extinction, mortality, recruitment, and survival.
953 One of the commonly used methods is that of Diamond (1969) where the percentage turnover
954 rate, T, is calculated as $T = 100 \times (E + C) / (S1 + S2)$, where E is the number of taxa that went

955 extinct between two time points, *C* is the number of taxa that colonised between the same two
956 time points, *S*1 is the number of taxa present at time 1 and *S*2 is the number of taxa present at
957 time 2. *T* gives the percent change in taxon identities over the time period and varies from 0,
958 representing no turnover, to 1, indicating complete turnover (Diamond 1969). This measure
959 can be calculated only for a single location at two time points; pairwise turnovers must be
960 calculated separately for datasets with multiple time periods or multiple sites. *T* values can be
961 compared through time, at different time intervals, or at different spatial scales (e.g.,
962 Boulinier et al. 2001, Tworek 2004). They also can be used in a regression-style analysis to
963 assess the effects of a treatment variable or other explanatory factor on compositional
964 turnover (e.g., Aguirre et al. 2003, Tworek 2004). These methods are only useful for
965 communities that have relatively high rates of temporal turnover in taxa. The degree of
966 compositional turnover also is tied to the temporal extent of the study for the given system
967 and the average lifespan of taxa. For example, complete turnover in prokaryotes can be
968 observed over a period of days (e.g., Kara and Shade 2009), whilst many decades may be
969 required to detect any species turnover in some plant communities (e.g., Kardol et al. 2010).
970 However, because it can be standardised by the number of taxa in the community, *T* is a
971 potentially comparable measure of compositional change among different datasets, if they are
972 on similar timescales and encompass communities with a similar distribution of life-histories.

973 **5. Multivariate regression modelling**

974 Multivariate autoregressive state-space models (MARSS) are multivariate linear regression
975 methods that model the stochastic temporal dynamics in matrices of species' abundances in
976 relation to abiotic or biotic covariates, while accounting for potential observation
977 (measurement) error (Hampton et al. 2013). The models are autoregressive in that previous
978 measurements in time are assumed to be correlated to future measurement, and these

autocorrelations are formalised by modelling the current values of the response matrix as a weighted linear sum of previous values at a specified temporal lag (Penny and Harrison 2006). MARSS models can be used to disentangle temporal variation in species abundances into the relative contributions of inter- and intraspecific interactions, abiotic variations, and of demographic stochasticity (Mutshinda et al. 2009). Effect sizes of these different components are estimated from the multivariate model structure through the use of conditional least-squares, maximum-likelihood, or Bayesian estimation methods (Hampton et al. 2013). Observation errors, and other complexities such as spatially-replicated time series and non-uniform sampling, can be formally modelled within the state-space framework (Holmes et al. 2012). Further, MARSS models can be used to compute several metrics that indicate the degree to which communities are possibly becoming less stable over time (Ives et al. 2003). This level of flexibility means that they are able to test for even the most complex temporal patterns in communities, and can simultaneously reveal similarities and differences among taxa, and account for environmental factors or other covariates. However, they do require relatively long time series (ideally more than 25 temporal replicates), and, like any statistical model, more independent replication is required for more complex models. There are several examples of the use of MARSS models in freshwater and marine ecology, particularly in terms of estimating aquatic food web dynamics (Beisner et al. 2003, Hampton et al. 2008, Thibaut et al. 2012); there have been relatively few applications for terrestrial communities (e.g., Yamamura et al. 2006, Mutshinda et al. 2009). The 'MARSS' R package (Holmes et al. 2012) enables the fitting of MARSS models and includes a range of ecological examples, so we have not included example code in this paper.

Multispecies N mixture models and joint species distribution models are other regression methods that can be used to model explicit, quantitative hypotheses of community change, with a focus on interactions among taxa (e.g., Dorazio et al. 2015, Ovaskainen et al.

2017). These methods can allow for useful additions, such as accounting for uncertainty in the detection of taxa; however, due to their complexity and high computational requirements, they have yet to be implemented for datasets with large numbers of species.

STEP 3: TEST FOR STATISTICAL SIGNIFICANCE

Whichever multivariate analysis has been selected, most ecologists interested in community dynamics seek a way to determine if the changes that they see over time in the results of those analyses are greater than, or less than, they would expect by chance alone. Some of the methods presented above have their own built-in significance tests, often based on randomisations of the community matrix. Commonly-used randomisation tests include the Mantel test (Legendre and Legendre 2012) and permutational multivariate analysis of variance (PERMANOVA; Anderson 2001, Legendre and Legendre 2012). Null models, whereby community matrices are randomised under sensible criteria (Gotelli and Graves 1996, Gotelli 2000, 2001, Gotelli and Entsminger 2001, Ulrich and Gotelli 2007, 2010), can be used to test the significance of any observed change in compositional dissimilarity, and can easily be incorporated into analyses described above including zeta diversity. The primary consideration in applying any randomisation test is what assumptions are being made about taxon niches and whether they are valid. If the interest is in testing the effects of interactions among taxa on their co-occurrence, but the randomisation procedure (null model) allows taxa to occur in samples where they are unable to occur for other reasons, such as habitat differences, a difference among samples may be falsely detected (an inflated Type I error rate). Thus, if strong inferences are to be made about compositional differences, we must be sure that the differences among the groups of sites that we are testing are the only difference. Similarly, the sample size needs to be high enough to allow the detection of differences among sets of sites for the given level of compositional variability (Type II error rate).

1029 Schaefer, et al. (2005) introduced a novel method for detecting significant shifts in
1030 community composition using randomisations under null models of taxon mean abundances
1031 through time. They demonstrate that this approach, which uses both compositional distance
1032 (pairwise dissimilarities) and the coefficient of variation (CV), could detect significant
1033 dynamics in community composition, even where only small amounts of change had
1034 occurred. In this method, temporal changes in communities are assessed by using a
1035 compositional distance measure that is selected by the user. This quantifies the multivariate
1036 distance between the randomly generated communities (null model randomisations) and
1037 observed 'reference' samples (e.g., 'Time 0' samples, compared to randomisations of 'Time
1038 1' samples). If the observed distance is greater than under randomisation, then we can infer
1039 statistically significant temporal change. Because this can be done in an ordination context,
1040 increases and decreases of particular taxa driving the change can be determined from the
1041 ordination plots. Limitations of this method are that it requires a relatively large degree of
1042 temporal change in the data compared to spatial variation (i.e., more temporal, than spatial,
1043 variation) and it is relatively computationally intensive; however, we provide the R code for a
1044 worked example in the supplementary material (Data S2).

1045 **STEP 4: REPORT ON THE ANALYSIS WORKFLOW**

1046 Our literature review suggested a lack of consistent reporting of critical aspects of the data
1047 analysis workflow. Such reporting of ensures that the results are robust, represent the
1048 variation deemed important for the research aim and are reproducible. The number of taxa,
1049 samples, and independent replicates in space and time all should be clearly reported in the
1050 Methods or Results section of the paper (see Hurlbert 1984 for guidance on statistical
1051 independence). The dissimilarity measure selected, and the rationale and procedures used to
1052 arrive at that decision, should be reported in the Methods section of the paper. We encourage
1053 researchers to refer to a specific source, such as Legendre & Legendre (2012), or to give the

formula applied to maximise the repeatability and transparency of their research; these indices can be inherent in particular methods or built-in as defaults in software, but nevertheless should still be reported. Different methods of dissimilarity or ordination (or other analysis types) that were compared, e.g., in the EDA phase, should be listed in the Methods section of the paper stating that their results were consistent with each other. The choice of dissimilarity or ordination method is subjective, but rules of thumb can be used for assessing different methods, so that the method selected optimises the reduction of dimensionality in the data and thus maximises the variation accounted for by the analysis (see McCune and Grace 2002 for comprehensive advice).

STEP 5: PRESENT TEMPORAL PATTERNS AND RESULTS

Some of the analysis methods that we describe in this paper do not produce simple visual outputs that clearly show temporal trends in community datasets and visualising temporal dynamics can be difficult for species-rich datasets or complex sampling designs. However, there are a variety of useful methods that can be used to complement a hypothesis-driven analysis (Table 3). Many descriptive methods work best where a small group of taxa is measured only a few times. However, there are examples of descriptive methods being used effectively for larger, more complex examples. Pereira et al. (2017) used a heat map to show compositional changes of 808 bacterial genera, at a single location, sampled 24 times over two years. Certain descriptive methods are useful for only specific types of temporal patterns, e.g., ‘time cores’ is a method that identifies ‘core’ groups of taxa that remain present over time (Vallès et al. 2014). Others, such as Venn diagrams and measures of overlap in composition, are useful only for small numbers of temporal comparisons. We recommend and provide R code (Data S2) for time cores and Venn diagrams because these methods are suitable for assessing hypotheses of taxon replacement while specifically identifying the taxa driving compositional shifts (Tables 2 and 3). Relative abundance distributions (RADs, *a.k.a.*

1079 ‘rank abundance curves’), can provide relatively sophisticated visualisations of community
1080 dynamics, and rank abundances can be used to quantitatively measure the shift in taxon ranks
1081 within samples taken at different time points (Hallett et al. 2016). RADs are plots of the
1082 relative abundances (on the *y*-axis) of all species in a community ranked (on the *x*-axis) from
1083 highest to lowest. They are closely related to *k*-dominance curves, in which species ranks are
1084 shown on the *x*-axis, but cumulative proportional abundance (dominance) is plotted on the *y*-
1085 axis (Lambshead et al. 1983). If samples taken at different times are plotted separately, the
1086 set of RADs can illustrate changes in community structure over time. RADs also can be used
1087 to explore community dynamics if a dissimilarity measure, such as Bray-Curtis, is calculated
1088 for the species’ ranks at each time point. The resulting temporal dissimilarity of the RADs
1089 can be explored by plotting rank dissimilarity against time lag or by ordinating rank
1090 dissimilarities in the same way that spatial compositional dissimilarities are examined. For
1091 example, temporal changes in RAD dissimilarity can illustrate changes in species evenness
1092 over time as might be expected if compositional homogenisation and a decline in evenness
1093 occurs because of invasion by an increasingly dominant species (Ruhl 2008).

1094 **E. DISCUSSION**

1095 There is an ever-expanding set of analyses that use a wide variety of multivariate and
1096 univariate methods to quantitatively address temporal questions about communities that
1097 continue to be developed (e.g., De Cáceres et al. 2019). One of the key reasons for the
1098 proliferation of methods appears to be variability in the length of the time series. For
1099 example, if the time series is long enough, methods such as temporal MEMs could be used;
1100 however, consistent with other reviews (e.g., Vaughn and Young 2010) and databases
1101 (Dornelas et al. 2018), most datasets captured by our literature had relatively few temporal
1102 measurements (mean = 14.9, median = 7, mode = 4), requiring very different methods. In
1103 other cases, the ability to implement some methods in temporal community ecology still

1104 requires development, such as the application of time-ordered and time-aggregated network
1105 analysis for spatiotemporal community datasets (Blonder et al. 2012).

1106 Some analysis types are well established for spatial data, but can be applied easily to
1107 temporal data, potentially yielding useful and comparable results for explicit temporal
1108 analyses. Further research is required to generate standard extensions and approaches using
1109 such methods for temporal data. For example, nestedness (*sensu* Patterson and Atmar 1986)
1110 is usually applied to spatial data and the analysis is repeated for different time points;
1111 however, there is no reason that we cannot apply this method across times to look at whether
1112 sites are becoming more or less nested over time. For instance, if younger sites were nested
1113 within older sites, this may indicate that sites were becoming more homogeneous over time.

1114 Our understanding of temporal dynamics in biological communities, and the methods
1115 we use to investigate them, have improved greatly over the past 25 years, probably because
1116 of increases in computational power and in dataset size. We can continue to improve our
1117 understanding by moving away from merely describing community changes, to linking these
1118 changes to understanding and testing the processes driving community change and
1119 biodiversity; that is, moving from hypothesis-generating methods towards hypothesis-testing
1120 methods enabled by well-defined research questions with specific, testable hypotheses that
1121 result in clear predictions for observed temporal patterns (e.g., Figs 4-7). Even for
1122 exploratory studies of compositional variation through time, we emphasise the benefits of
1123 making *a priori* predictions for how we think composition should change through time.

1124 Interactions between community composition and environmental or other factors
1125 across space and time are complicated. For example, Brown (2003) showed that temporal
1126 variability in species composition of stream invertebrates was reduced by high spatial
1127 heterogeneity, which may be because there are increased niches or refugia in changing
1128 environmental conditions. Manipulative experiments are useful, but not always possible,

1129 especially for long-lived taxa. Therefore, combining current knowledge of intrinsic and
1130 extrinsic factors controlling community change into null models and simulations may be the
1131 way forward (Brown and Lawson 2010). Carefully considering sample sizes and appropriate
1132 replication of particular factors are also important if we are to properly test hypotheses of
1133 species coexistence over time (Philippi et al. 1998). There is a need for us to make useful
1134 predictions about how communities may continue to respond to ecosystem disturbances, such
1135 as climate change, invasive species, and habitat fragmentation (Hobbs et al. 2007,
1136 Kampichler et al. 2012). Many methods commonly used to analyse temporal community
1137 dynamics are not quantitative, meaning conclusions can only be narrative descriptions, rather
1138 than clear tests of hypotheses. Many of the methods presented here push us to quantitatively
1139 assess dynamics, which naturally leads to clearer hypothesis testing.

1140 There is a wide variety of methods available for analysing temporal dynamics of
1141 biological communities; however, many studies in our literature review using these methods
1142 are descriptive, showing that a large body of literature in community ecology is not focussed
1143 on testing quantitative hypotheses of change. Surprisingly few studies in our review
1144 presented hypotheses that were more complex than simply predicting that composition would
1145 have changed in some way over time, instead of specifying predictions of quantitative and
1146 directional patterns of change in community composition. Even additional searching beyond
1147 our original search terms that included more specific terminology for compositional change
1148 (succession, turnover, synchron*, resilience, homogenization, divergence, convergence)
1149 revealed few empirical papers that tested specific, quantitative hypotheses. This search
1150 resulted in an additional 1,514 papers of which we read 10% (151 papers). Of these, 60
1151 papers met the inclusion criteria of our original review and only 13 papers tested hypotheses
1152 that were more sophisticated than the simple prediction that composition would have changed

in some way over time, e.g., directional change or that the abundances of some taxa would change more than others.

This lack of specific, quantitative hypothesis testing is not surprising because to predict change in composition in a quantitative way, we require an in-depth understanding of a system and how different components of the diversity differ in their ecology or other factors that could cause directional, cyclical, or other differential change in time. In many cases, particularly in the observational, rather than experimental, community ecology literature, this is not the situation. This means that, although such studies may state hypotheses, these hypotheses are not very specific or quantitative. In most cases, they predict some change in response to time or another factor that is correlated with time (such as disturbance or pollution), but the types of temporal change (divergence, convergence, synchrony, etc.) are not specified. We strongly recommend that researchers generate *a priori* hypotheses of compositional change (e.g., Figs. 4-7; Bagchi et al. 2017). These predictions should be used to select appropriate quantitative data analysis methods (Fig. 17). The comprehensive description and hypothesis testing of temporal community patterns requires the use of multiple data analysis methods, but by using standard exploratory data analysis methods followed by testing *a priori* hypotheses, the methods selected can be explicitly focussed on the most appropriate tests for the situation. We hope that this synthesis will allow researchers to make informed decisions about their choice of data analysis for their research question and pave the way for more uniformity in reporting of these analyses and/or innovation in comparability. Our understanding of patterns and drivers of changes in species communities would benefit by having more longer-term and fine-scale studies.

ACKNOWLEDGEMENTS

We thank Audrey Lustig for initial database and R code development. This paper is a contribution from the Harvard Forest LTER program. We thank members of the Forest

Ecology Research Group at Wilfrid Laurier University, Glenda Mendieta-Leiva, Brian Inouye, Miguel De Cáceres, and an anonymous reviewer for comments on the manuscript. We thank the authors of Porensky et al. (2016a, 2016b) and Gilbert et al. (2012) for making their data publicly available. The BCI forest dynamics research project was founded by S.P. Hubbell and R.B. Foster and is now managed by R. Condit, S. Lao, and R. Perez under the Center for Tropical Forest Science and the Smithsonian Tropical Research in Panama. Numerous organizations have provided funding, principally the U.S. National Science Foundation, and hundreds of field workers have contributed.

FUNDING DISCLOSURE

H. L. Buckley and B. S. Case were supported by a Charles Bullard Fellowship at Harvard Forest. A. M. Ellison's work on this project was supported by the Harvard Forest.

SUPPORTING INFORMATION

Data S1. Description of attributes, and associated data, extracted from papers in the literature review.

Data S2. Example R code for plant community dataset

Data S3. Example R code for BCI dataset

Data S4. Example R code for marine bacterial community dataset (from high throughput sequencing)

Appendix S1. Supplementary figures from the literature review.

Appendix S2. Descriptions of methods encountered in the literature review that are not included in the main text.

REFERENCES CITED

Aguirre, L. F., L. Lens, R. van Damme, and E. Matthysen. 2003. Consistency and variation in the bat assemblages inhabiting two forest islands within a neotropical savanna in Bolivia. Journal of Tropical Ecology 19:367–374.

Amend, A. S., K. A. Seifert, and T. D. Bruns. 2010. Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Molecular Ecology 19:5555–5565.

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology 26:32–46.

Anderson, M. J., T. O. Crist, J. M. Chase, M. Vellend, B. D. Inouye, A. L. Freestone, N. J. Sanders, H. V. Cornell, L. S. Comita, K. F. Davies, S. P. Harrison, N. J. B. Kraft, J. C. Stegen, and N. G. Swenson. 2011. Navigating the multiple meanings of β diversity: A roadmap for the practicing ecologist. Ecology Letters 14:19–28.

Anderson, M. J., and T. J. Willis. 2003. Canonical analysis of principal coordinates: A useful method of constrained ordination for ecology. Ecology 84:511–525.

Angeler, D. G., and R. K. Johnson. 2012. Patterns of temporal community turnover are spatially synchronous across boreal lakes: Temporal coherence in invertebrate community turnover. Freshwater Biology 57:1782–1793.

Angeler, D. G., O. Viedma, and J. M. Moreno. 2009. Statistical performance and information content of time lag analysis and redundancy analysis in time series modeling. Ecology 90:3245–3257.

Auber, A., M. Travers-Trolet, M. C. Villanueva, and B. Ernande. 2017. A new application of principal response curves for summarizing abrupt and cyclic shifts of communities over space. Ecosphere 8:e02023.

Azeria, E. T., A. Carlson, T. Pärt, and C. G. Wiklund. 2006. Temporal dynamics and nestedness of an oceanic island bird fauna. Global Ecology and Biogeography 15:328–338.

Azeria, E. T., and J. Kolasa. 2008. Nestedness, niche metrics and temporal dynamics of a metacommunity in a dynamic natural model system. Oikos 117:1006–1019.

Bagchi, S., N. J. Singh, D. D. Briske, B. T. Bestelmeyer, M. P. McClaran, and K. Murthy. 2017. Quantifying long-term plant community dynamics with movement models: implications for ecological resilience. Ecological Applications 27:1514–1528.

Bar-Massada, A., and L. Hadar. 2017. Grazing and temporal turnover in herbaceous communities in a Mediterranean landscape. Journal of Vegetation Science 28:270–280.

Barros, F., H. Blanchet, K. Hammerstrom, P.-G. Sauriau, and J. Oliver. 2014. A framework for investigating general patterns of benthic β-diversity along estuaries. Estuarine, Coastal and Shelf Science 149:223–231.

Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity: Partitioning beta diversity. Global Ecology and Biogeography 19:134–143.

Baselga, A. 2012. The relationship between species replacement, dissimilarity derived from nestedness, and nestedness: Species replacement and nestedness. Global Ecology and Biogeography 21:1223–1232.

Baselga, A. 2017. Partitioning abundance-based multiple-site dissimilarity into components: Balanced variation in abundance and abundance gradients. Methods in Ecology and Evolution 8:799–808.

Baselga, A., and C. D. L. Orme. 2012. betapart : an R package for the study of beta diversity: Betapart package. Methods in Ecology and Evolution 3:808–812.

Bashan, A., T. E. Gibson, J. Friedman, V. J. Carey, S. T. Weiss, E. L. Hohmann, and Y.-Y. Liu. 2016. Universality of human microbial dynamics. Nature 534:259–262.

Bêche, L. A., and V. H. Resh. 2007. Short-term climatic trends affect the temporal variability of macroinvertebrates in California 'Mediterranean' streams. Freshwater Biology 52:2317–2339.

Beisner, B. E., A. R. Ives, and S. R. Carpenter. 2003. The effects of an exotic fish invasion on the prey communities of two lakes. Journal of Animal Ecology 72:331–342.

Berthon, V., B. Alric, F. Rimet, and M.-E. Perga. 2014. Sensitivity and responses of diatoms to climate warming in lakes heavily influenced by humans. Freshwater Biology 59:1755–1767.

Bestelmeyer, B. T., A. M. Ellison, W. R. Fraser, K. B. Gorman, S. J. Holbrook, C. M. Laney, M. D. Ohman, D. P. C. Peters, F. C. Pillsbury, A. Rassweiler, R. J. Schmitt, and S. Sharma. 2011. Analysis of abrupt transitions in ecological systems. Ecosphere 2:art129.

Blonder, B., T. W. Wey, A. Dornhaus, R. James, and A. Sih. 2012. Temporal dynamics and network analysis. Methods in Ecology and Evolution 3:958–972.

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

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

Boulinier, T., J. D. Nichols, J. E. Hines, J. R. Sauer, C. H. Flather, and K. H. Pollock. 2001. Forest fragmentation and bird community dynamics: Inference at regional scales. Ecology 82:1159–1169.

ter Braak, C. J. F., and P. Šmilauer. 2015. Topics in constrained and unconstrained ordination. Plant Ecology 216:683–696.

Broadway, K. J., M. Pyron, J. R. Gammon, and B. A. Murry. 2015. Shift in a large river fish assemblage: Body-size and trophic structure dynamics. PLOS ONE 10:e0124954.

Brown, B. L. 2003. Spatial heterogeneity reduces temporal variability in stream insect communities. Ecology Letters 6:316–325.

Brown, B. L., and R. L. Lawson. 2010. Habitat heterogeneity and activity of an omnivorous ecosystem engineer control stream community dynamics. Ecology 91:1799–1810.

Caruso, T., V. Trokhymets, R. Bargagli, and P. Convey. 2013. Biotic interactions as a structuring force in soil communities: Evidence from the micro-arthropods of an Antarctic moss model system. Oecologia 172:495–503.

Chao, A., R. L. Chazdon, R. K. Colwell, and T.-J. Shen. 2006. Abundance-based similarity indices and their estimation when there are unseen species in samples. Biometrics 62:361–371.

Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs 84:45–67.

Chapman, J. I., and R. W. McEwan. 2016. Thirty years of compositional change in an old-growth temperate forest: The role of topographic gradients in oak-maple dynamics. PLOS ONE 11:e0160238.

Clavero, M., and E. Garcia-Berthou. 2006. Homogenization dynamics and introduction routes of invasive freshwater fish in the Iberian Peninsula. Ecological Applications 16:2313–2324.

Cleland, E. E., S. L. Collins, T. L. Dickson, E. C. Farrer, K. L. Gross, L. A. Gherardi, L. M. Hallett, R. J. Hobbs, J. S. Hsu, L. Turnbull, and K. N. Suding. 2013. Sensitivity of grassland plant community composition to spatial vs. temporal variation in precipitation. Ecology 94:1687–1696.

Collins, S. L. 1992. Fire frequency and community heterogeneity in tallgrass prairie vegetation. Ecology 73:2001–2006.

Collins, S. L. 2000. Disturbance frequency and community stability in native tallgrass prairie. The American Naturalist 155:311–325.

Collins, S. L., F. Micheli, and L. Hartt. 2000. A method to determine rates and patterns of variability in ecological communities. Oikos 91:285–293.

Collins, S. L., and M. D. Smith. 2006. Scale-dependent interaction of fire and grazing on community heterogeneity in tallgrass prairie. Ecology 87:2058–2067.

Collins, S. L., and Y. Xia. 2015. Long-term dynamics and hotspots of change in a desert grassland plant community. The American Naturalist 185:E30–E43.

Condit, R. 1998. Tropical Forest Census Plots. Springer-Verlag and R. G. Landes Company, Berlin, Germany, and Georgetown, Texas.

Connell, J. H., and R. O. Slatyer. 1977. Mechanisms of succession in natural communities and their role in community stability and organization. The American Naturalist 111:1119–1144.

Cotton, T. E. A., A. H. Fitter, R. M. Miller, A. J. Dumbrell, and T. Helgason. 2015. Fungi in the future: interannual variation and effects of atmospheric change on arbuscular mycorrhizal fungal communities. New Phytologist 205:1598–1607.

Day, N. J., and H. L. Buckley. 2013. Twenty-five years of plant community dynamics and invasion in New Zealand tussock grasslands. Austral Ecology 38:688–699.

Day, N. J., S. Carrière, and J. L. Baltzer. 2017. Annual dynamics and resilience in post-fire boreal understory vascular plant communities. Forest Ecology and Management 401:264–272.

De Cáceres, M., L. Coll, P. Legendre, R. B. Allen, S. K. Wiser, M.-J. Fortin, R. Condit, and S. Hubbell. 2019. Trajectory analysis in community ecology. Ecological Monographs:e01350.

De Cáceres, M., X. Font, and F. Oliva. 2010. The management of vegetation classifications with fuzzy clustering: Fuzzy clustering in vegetation classifications. Journal of Vegetation Science 21:1138–1151.

Diamond, J. M. 1969. Avifaunal equilibria and species turnover rates on the Channel Islands of California. Proceedings of the National Academy of Sciences 64:57–63.

Dimitriu, P. A., G. Boyce, A. Samarakoon, M. Hartmann, P. Johnson, and W. W. Mohn. 2013. Temporal stability of the mouse gut microbiota in relation to innate and adaptive immunity: Mouse gut microbiota dynamics. Environmental Microbiology Reports 5:200–210.

Dorazio, R. M., E. F. Connor, and R. A. Askins. 2015. Estimating the Effects of Habitat and Biological Interactions in an Avian Community. PLOS ONE 10:e0135987.

Dornelas, M., L. H. Antão, F. Moyes, A. E. Bates, A. E. Magurran, D. Adam, A. A. Akhmetzhanova, W. Appeltans, J. M. Arcos, H. Arnold, N. Ayyappan, G. Badihi, A. H. Baird, M. Barbosa, T. E. Barreto, C. Bässler, A. Bellgrove, J. Belmaker, L. Benedetti-Cecchi, B. J. Bett, A. D. Bjorkman, M. Błażewicz, S. A. Blowes, C. P. Bloch, T. C. Bonebrake, S. Boyd, M. Bradford, A. J. Brooks, J. H. Brown, H. Bruelheide, P. Budy, F. Carvalho, E. Castañeda-Moya, C. A. Chen, J. F. Chamblee, T. J. Chase, L. Siegwart Collier, S. K. Collinge, R. Condit, E. J. Cooper, J. H. C. Cornelissen, U. Cotano, S. Kyle Crow, G. Damasceno, C. H. Davies, R. A. Davis, F. P. Day, S. Degraer, T. S. Doherty, T. E. Dunn, G. Durigan, J. E. Duffy, D. Edelist, G. J. Edgar, R. Elahi, S. C. Elmendorf, A. Enemar, S. K. M. Ernest, R. Escribano, M. Estiarte, B. S. Evans, T.-Y. Fan, F. Turini Farah, L. Loureiro Fernandes, F. Z.

Farneda, A. Fidelis, R. Fitt, A. M. Fosaa, G. A. Daher Correa Franco, G. E. Frank, W. R. Fraser, H. García, R. Cazzolla Gatti, O. Givan, E. Gorgone-Barbosa, W. A. Gould, C. Gries, G. D. Grossman, J. R. Gutierréz, S. Hale, M. E. Harmon, J. Harte, G. Haskins, D. L. Henshaw, L. Hermanutz, P. Hidalgo, P. Higuchi, A. Hoey, G. Van Hoey, A. Hofgaard, K. Holeck, R. D. Hollister, R. Holmes, M. Hoogenboom, C. Hsieh, S. P. Hubbell, F. Huettmann, C. L. Huffard, A. H. Hurlbert, N. Macedo Ivanauskas, D. Janík, U. Jandt, A. Jażdżewska, T. Johannessen, J. Johnstone, J. Jones, F. A. M. Jones, J. Kang, T. Kartawijaya, E. C. Keeley, D. A. Kelt, R. Kinnear, K. Klanderud, H. Knutsen, C. C. Koenig, A. R. Kortz, K. Král, L. A. Kuhnz, C.-Y. Kuo, D. J. Kushner, C. Laguionie-Marchais, L. T. Lancaster, C. Min Lee, J. S. Lefcheck, E. Lévesque, D. Lightfoot, F. Lloret, J. D. Lloyd, A. López-Baucells, M. Louzao, J. S. Madin, B. Magnússon, S. Malamud, I. Matthews, K. P. McFarland, B. McGill, D. McKnight, W. O. McLarney, J. Meador, P. L. Meserve, D. J. Metcalfe, C. F. J. Meyer, A. Michelsen, N. Milchakova, T. Moens, E. Moland, J. Moore, C. Mathias Moreira, J. Müller, G. Murphy, I. H. Myers-Smith, R. W. Myster, A. Naumov, F. Neat, J. A. Nelson, M. Paul Nelson, S. F. Newton, N. Norden, J. C. Oliver, E. M. Olsen, V. G. Onipchenko, K. Pabis, R. J. Pabst, A. Paquette, S. Pardede, D. M. Paterson, R. Pélissier, J. Peñuelas, A. Pérez-Matus, O. Pizarro, F. Pomati, E. Post, H. H. T. Prins, J. C. Priscu, P. Provoost, K. L. Prudic, E. Pulliainen, B. R. Ramesh, O. Mendivil Ramos, A. Rassweiler, J. E. Rebelo, D. C. Reed, P. B. Reich, S. M. Remillard, A. J. Richardson, J. P. Richardson, I. van Rijn, R. Rocha, V. H. Rivera-Monroy, C. Rixen, K. P. Robinson, R. Ribeiro Rodrigues, D. de Cerqueira Rossa-Feres, L. Rudstam, H. Ruhl, C. S. Ruz, E. M. Sampaio, N. Rybicki, A. Rypel, S. Sal, B. Salgado, F. A. M. Santos, A. P. Savassi-Coutinho, S. Scanga, J. Schmidt, R. Schooley, F. Setiawan, K.-T. Shao, G. R. Shaver, S. Sherman, T. W. Sherry, J.

Siciński, C. Sievers, A. C. da Silva, F. Rodrigues da Silva, F. L. Silveira, J. Slingsby, T. Smart, S. J. Snell, N. A. Soudzilovskaia, G. B. G. Souza, F. Maluf Souza, V. Castro Souza, C. D. Stallings, R. Stanforth, E. H. Stanley, J. Mauro Sterza, M. Stevens, R. Stuart-Smith, Y. Rondon Suarez, S. Supp, J. Yoshio Tamashiro, S. Tarigan, G. P. Thiede, S. Thorn, A. Tolvanen, M. Teresa Zugliani Toniato, Ø. Totland, R. R. Twilley, G. Vaitkus, N. Valdivia, M. I. Vallejo, T. J. Valone, C. Van Colen, J. Vanaverbeke, F. Venturoli, H. M. Verheye, M. Vianna, R. P. Vieira, T. Vrška, C. Quang Vu, L. Van Vu, R. B. Waide, C. Waldock, D. Watts, S. Webb, T. Wesołowski, E. P. White, C. E. Widdicombe, D. Wilgers, R. Williams, S. B. Williams, M. Williamson, M. R. Willig, T. J. Willis, S. Wipf, K. D. Woods, E. J. Woehler, K. Zawada, and M. L. Zettler. 2018. BioTIME: A database of biodiversity time series for the Anthropocene. Global Ecology and Biogeography 27:760–786.

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

Ellison, A. 2010. Partitioning diversity. Ecology 91:1962–1963.

Ferrier, S., G. Manion, J. Elith, and K. Richardson. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Diversity and Distributions 13:252–264.

Franca, A., S. Caredda, F. Sanna, F. Fava, and G. Seddaiu. 2016. Early plant community dynamics following overseeding for the rehabilitation of a Mediterranean silvopastoral system. Grassland Science 62:81–91.

Gaston, K. J., and B. H. McArdle. 1994. The temporal variability of animal abundances: measures, methods and patterns. Philosophical Transactions of the Royal Society B: Biological Sciences 345:335–358.

Gilbert, J. A., J. A. Steele, J. G. Caporaso, L. Steinbrück, J. Reeder, B. Temperton, S. Huse, A. C. McHardy, R. Knight, and I. Joint. 2012. Defining seasonal marine microbial community dynamics. The ISME journal 6:298.

Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. Ecology 81:2606–2621.

Gotelli, N. J. 2001. Research frontiers in null model analysis. Global Ecology and Biogeography 10:337–343.

Gotelli, N. J., and G. L. Entsminger. 2001. Swap and fill algorithms in null model analysis: rethinking the knight's tour. Oecologia 129:281–291.

Gotelli, N. J., and G. R. Graves. 1996. Null Models in Ecology. Smithsonian Institution Press, Washington and London.

Gouhier, T. C., and F. Guichard. 2014. Synchrony: Quantifying variability in space and time. Methods in Ecology and Evolution 5:524–533.

Hallett, L. M., S. K. Jones, A. A. M. MacDonald, M. B. Jones, D. F. B. Flynn, J. Ripplinger, P. Slaughter, C. Gries, and S. L. Collins. 2016. codyn: An R package of community dynamics metrics. Methods in Ecology and Evolution 7:1146–1151.

Hampton, S. E., E. E. Holmes, L. P. Scheef, M. D. Scheuerell, S. L. Katz, D. E. Pendleton, and E. J. Ward. 2013. Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. Ecology 94:2663–2669.

Hampton, S. E., L. R. Izmest'Eva, M. V. Moore, S. L. Katz, B. Dennis, and E. A. Silow. 2008. Sixty years of environmental change in the world's largest freshwater lake - Lake Baikal, Siberia. Global Change Biology 14:1947–1958.

Hobbs, R. J., S. Yates, and H. A. Mooney. 2007. Long-term data reveal complex dynamics in grassland in relation to climate and disturbance. Ecological Monographs 77:545–568.

Holmes, E. E., E. J. Ward, and K. Wills. 2012. MARSS: Multivariate Autoregressive State-space models for analyzing time-series data. R Journal 4:11–19.

Houlahan, J. E., D. J. Currie, K. Cottenie, G. S. Cumming, S. K. M. Ernest, C. S. Findlay, S. D. Fuhlendorf, U. Gaedke, P. Legendre, J. J. Magnuson, B. H. McArdle, E. H. Muldavin, D. Noble, R. Russell, R. D. Stevens, T. J. Willis, I. P. Woiwod, and S. M. Wondzell. 2007. Compensatory dynamics are rare in natural ecological communities. Proceedings of the National Academy of Sciences 104:3273–3277.

Hubbell, S. P. 1979. Tree dispersion, abundance, and diversity in a tropical dry forest. Science 203:1299–1309.

Hubbell, S. P., R. Condit, and R. B. Foster. 2005. Barro Colorado Forest Census Plot Data.

Hubbell, S. P., R. B. Foster, S. T. O'Brien, K. E. Harms, R. Condit, B. Wechsler, S. J. Wright, and S. Loo de Lao. 1999. Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. Science 283:554–557.

Hui, C., and M. A. McGeoch. 2014. Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns. The American Naturalist 184:684–694.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187–211.

Ives, A. R., B. Dennis, K. L. Cottingham, and S. R. Carpenter. 2003. Estimating community stability and ecological interactions from time-series data. Ecological monographs 73:301–330.

Jackson, D. A. 1995. PROTEST: A PROcrustean Randomization TEST of community environment concordance. Écoscience 2:297–303.

Johnstone, I. M., and D. M. Titterington. 2009. Statistical challenges of high-dimensional data. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 367:4237–4253.

Kalyuzhny, M., and N. M. Shnerb. 2017. Dissimilarity-overlap analysis of community dynamics: Opportunities and pitfalls. Methods in Ecology and Evolution 8:1764–1773.

Kampichler, C., D. G. Angeler, R. T. Holmes, A. Leito, S. Svensson, H. P. van der Jeugd, and T. Wesołowski. 2014. Temporal dynamics of bird community composition: An analysis of baseline conditions from long-term data. Oecologia 175:1301–1313.

Kampichler, C., C. A. M. van Turnhout, V. Devictor, and H. P. van der Jeugd. 2012. Large-scale changes in community composition: Determining land use and climate change signals. PLoS ONE 7:e35272.

Kara, E., and A. Shade. 2009. Temporal dynamics of South End Tidal Creek (Sapelo Island, Georgia) bacterial communities. Applied and Environmental Microbiology 75:1058–1064.

Kardol, P., D. E. Todd, P. J. Hanson, and P. J. Mulholland. 2010. Long-term successional forest dynamics: species and community responses to climatic variability. Journal of Vegetation Science.

Kent, M. 2012. Vegetation Description and Data Analysis: A practical approach. Second Edition. Wiley-Blackwell, Chichester, UK.

Kidé, S. O., C. Manté, L. Dubroca, H. Demarcq, and B. Mérigot. 2015. Spatio-temporal dynamics of exploited groundfish species assemblages faced to environmental and fishing forcings: insights from the Mauritanian exclusive economic zone. PloS one 10:e0141566.

Lambshead, P. J. D., H. M. Platt, and K. M. Shaw. 1983. The detection of differences among assemblages of marine benthic species based on an assessment of dominance and diversity. Journal of Natural History 17:859–874.

Lamy, T., P. Legendre, Y. Chancerelle, G. Siu, and J. Claudet. 2015. Understanding the spatio-temporal response of coral reef fish communities to natural disturbances: Insights from beta-diversity decomposition. PLOS ONE 10:e0138696.

Latombe, G., C. Hui, and M. A. McGeoch. 2017a. Multi-site generalised dissimilarity modelling: using zeta diversity to differentiate drivers of turnover in rare and widespread species. Methods in Ecology and Evolution 8:431–442.

Latombe, G., M. A. McGeoch, D. A. Nipperess, and C. Hui. 2017b. zetadiv: Functions to Compute Compositional Turnover Using Zeta Diversity. R package version 1.0. https://CRAN.R-project.org/package=zetadiv.

Legendre, P., and M. De Cáceres. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. Ecology Letters 16:951–963.

Legendre, P., and E. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129:271–280.

Legendre, P., and O. Gauthier. 2014. Statistical methods for temporal and space-time analysis of community composition data. Proceedings of the Royal Society B: Biological Sciences 281:20132728–20132728.

Legendre, P., and L. F. J. Legendre. 2012. Numerical Ecology. Third Edition. Elsevier, Radarweg, The Netherlands.

Lellouch, L., S. Pavoine, F. Jiguet, H. Glotin, and J. Sueur. 2014. Monitoring temporal change of bird communities with dissimilarity acoustic indices. Methods in Ecology and Evolution 5:495–505.

Lepš, J., and P. Šmilauer. 2003. Multivariate Analysis of Ecological Data Using CANOCO. Cambridge University Press, Cambridge.

Lewis, S. L., and M. A. Maslin. 2015. Defining the Anthropocene. Nature 519:171–180.

Loreau, M., and C. de Mazancourt. 2008. Species synchrony and its drivers: neutral and nonneutral community dynamics in fluctuating environments. The American Naturalist 172:E48–E66.

Lozupone, C. A., M. Hamady, S. T. Kelley, and R. Knight. 2007. Quantitative and qualitative diversity measures lead to different insights into factors that structure microbial communities. Applied and Environmental Microbiology 73:1576–1585.

Lozupone, C., and R. Knight. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology 71:8228–8235.

Magalhães, M. F., P. Beja, I. J. Schlosser, and M. J. Collares-Pereira. 2007. Effects of multi-year droughts on fish assemblages of seasonally drying Mediterranean streams. Freshwater Biology 52:1494–1510.

Magurran, A. E. 2004. Measuring Biological Diversity. Blackwell, Malden.

Magurran, A. E. 2013. Measuring Biological Diversity. John Wiley and Sons, Oxford.

McCune, B., and J. B. Grace. 2002. Analysis of Ecological Communities. MjM Software Design, Gleneden Beach.

McGeoch, M. A., G. Latombe, N. R. Andrew, S. Nakagawa, D. A. Nipperess, M. Roige, E. . Marzinelli, A. H. Campbell, A. Vergés, T. Thomas, P. D. Steinberg, K. E. Selwood, and C. Hui. 2017. The application of zeta diversity as a continuous measure of compositional change in ecology. bioRxiv 216580.

Micheli, F., K. L. Cottingham, J. Bascompte, O. N. Bjornstad, G. L. Eckert, J. M. Fischer, T. H. Keitt, B. E. Kendall, J. L. Klug, and J. A. Rusak. 1999. The dual nature of community variability. Oikos 85:161–169.

Mutshinda, C. M., R. B. O'Hara, and I. P. Woiwod. 2009. What drives community dynamics? Proceedings of the Royal Society B: Biological Sciences 276:2923–2929.

Noll, M., D. Matthies, P. Frenzel, M. Derakshani, and W. Liesack. 2005. Succession of bacterial community structure and diversity in a paddy soil oxygen gradient. Environmental Microbiology 7:382–395.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2017. vegan: Community Ecology Package. R package version 2.4-2. http://CRAN.R-project.org/package=vegan.

O'Sullivan, D. J., P. D. Cotter, O. O'Sullivan, L. Giblin, P. L. H. McSweeney, and J. J. Sheehan. 2015. Temporal and spatial differences in microbial composition during the manufacture of a Continental-type cheese. Applied and Environmental Microbiology 81:2525–2533.

Ovaskainen, O., G. Tikhonov, D. Dunson, V. Grøtan, S. Engen, B.-E. Sæther, and N. Abrego. 2017. How are species interactions structured in species-rich communities? A new method for analysing time-series data. Proceedings of the Royal Society B: Biological Sciences 284:20170768.

Palmer, M. 2019. Ordination Methods for Ecologists. http://ordination.okstate.edu/.

Patterson, B. D., and W. Atmar. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. Biological journal of the Linnean society 28:65–82.

Pedersen, E. J., P. L. Thompson, R. A. Ball, M.-J. Fortin, T. C. Gouhier, H. Link, C. Moritz, H. Nenzen, R. R. E. Stanley, Z. E. Taranu, A. Gonzalez, F. Guichard, and P. Pepin. 2017. Signatures of the collapse and incipient recovery of an overexploited marine ecosystem. Royal Society Open Science 4:170215.

Penny, W., and L. Harrison. 2006. Multivariate autoregressive models. Statistical Parametric Mapping: The Analysis of Functional Brain Images, eds K. J. Friston, JT Ashburner, SJ Kiebel, TE Nichols, and WD Penny (London: Elsevier):534–540.

Pereira, R. P. A., J. Peplies, M. G. Höfle, and I. Brettar. 2017. Bacterial community dynamics in a cooling tower with emphasis on pathogenic bacteria and Legionella species using universal and genus-specific deep sequencing. Water Research 122:363–376.

Peres-Neto, P. R., and D. A. Jackson. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia 129:169–178.

Philippi, T. E., P. M. Dixon, and B. E. Taylor. 1998. Detecting trends in species composition. Ecological Applications 8:300–308.

Podani, J., and D. Schmera. 2011. A new conceptual and methodological framework for exploring and explaining pattern in presence - absence data. Oikos 120:1625–1638.

Porensky, L. M., K. E. Mueller, D. J. Augustine, and J. D. Derner. 2016a. Thresholds and gradients in a semi-arid grassland: long-term grazing treatments induce slow, continuous and reversible vegetation change. Journal of Applied Ecology 53:1013–1022.

Porensky, L. M., K. E. Mueller, D. J. Augustine, and J. D. Derner. 2016b. Data from: Thresholds and gradients in a semi-arid grassland: long-term grazing treatments induce slow, continuous and reversible vegetation change. https://doi.org/10.5061/dryad.3m8p1.

Quinn, M. A., S. E. Halbert, and L. Williams. 1991. Spatial and temporal changes in aphid (Homoptera: Aphididae) species assemblages collected with suction traps in Idaho. Journal of Economic Entomology 84:1710–1716.

R Core Development Team. 2017. R: A language and environment for statistical computing v. 3.3.3. R Foundation for Statistical Computing, Vienna.

Rayner, T. S., B. J. Pusey, and R. G. Pearson. 2009. Spatio-temporal dynamics of fish feeding in the lower Mulgrave River, north-eastern Queensland: The influence of

seasonal flooding, instream productivity and invertebrate abundance. Marine and Freshwater Research 60:97.

Rojas-Sandoval, J., E. Meléndez-Ackerman, and D. S. Fernández. 2012. Plant community dynamics of a tropical semi-arid system following experimental removals of an exotic grass. Applied Vegetation Science 15:513–524.

Ruhl, H. A. 2008. Community change in the variable resource habitat of the abyssal Northeast Pacific. Ecology 89:991–1000.

Rusak, J. A., N. D. Yan, K. M. Somers, and D. J. McQueen. 1999. The temporal coherence of zooplankton population abundances in neighboring north-temperate lakes. The American Naturalist 153:46–58.

Schaefer, J., K. Gido, and M. Smith. 2005. A test for community change using a null model approach. Ecological Applications 15:1761–1771.

Shenhav, L., O. Furman, I. Mizrahi, and E. Halperin. 2017. Modeling the temporal dynamics of the gut microbial community in adults and infants. bioRxiv:212993.

Shertzer, K. W., E. H. Williams, and J. C. Taylor. 2009. Spatial structure and temporal patterns in a large marine ecosystem: Exploited reef fishes of the southeast United States. Fisheries Research 100:126–133.

Shi, S., D. Zou, Q. Wang, X. Xia, T. Zheng, C. Wu, and M. Gao. 2016. Responses of ammonia-oxidizing bacteria community composition to temporal changes in physicochemical parameters during food waste composting. RSC Adv. 6:9541–9548.

Shimadzu, H., M. Dornelas, and A. E. Magurran. 2015. Measuring temporal turnover in ecological communities. Methods in Ecology and Evolution 6:1384–1394.

Stead, T. K., J. M. Schmid-Araya, and A. G. Hildrew. 2003. All creatures great and small: patterns in the stream benthos across a wide range of metazoan body size. Freshwater Biology 48:532–547.

1592 Sugiura, S., T. Tsuru, and Y. Yamaura. 2013. Effects of an invasive alien tree on the diversity
1593 and temporal dynamics of an insect assemblage on an oceanic island. Biological
1594 Invasions 15:157–169.
1595 Thibaut, L. M., S. R. Connolly, and H. P. A. Sweatman. 2012. Diversity and stability of
1596 herbivorous fishes on coral reefs. Ecology 93:891–901.
1597 Thrush, S. F., J. Halliday, J. E. Hewitt, and A. M. Lohrer. 2008. The effects of habitat loss,
1598 fragmentation, and community homogenization on resilience in estuaries. Ecological
1599 Applications 18:12–21.
1600 Tilman, D. 1999. The ecological consequences of changes in biodiversity: A search for
1601 general principles. Ecology 80:1455–1474.
1602 Tworek, S. 2004. Factors Affecting Temporal Dynamics of Avian Assemblages in a
1603 Heterogeneous Landscape. Acta Ornithologica 39:155–163.
1604 Ulrich, W., and N. J. Gotelli. 2007. Null model analysis of species nestedness patterns.
1605 Ecology 88:1824–1831.
1606 Ulrich, W., and N. J. Gotelli. 2010. Null model analysis of species associations using
1607 abundance data. Ecology 91:3384–3397.
1608 Vallès, Y., A. Artacho, A. Pascual-García, M. L. Ferrús, M. J. Gosalbes, J. J. Abellán, and M.
1609 P. Francino. 2014. Microbial succession in the gut: Directional trends of taxonomic
1610 and functional change in a birth cohort of Spanish infants. PLoS Genetics
1611 10:e1004406.
1612 Van den Brink, P. J., and C. J. Ter Braak. 1999. Principal response curves: Analysis of time-
1613 dependent multivariate responses of biological community to stress. Environmental
1614 Toxicology and Chemistry 18:138–148.

Vaughn, K. J., and T. P. Young. 2010. Contingent conclusions: Year of initiation influences ecological field experiments, but temporal replication is rare. Restoration Ecology 18:59–64.

Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs 30:279.

Xiong, J., J. Zhu, K. Wang, X. Wang, X. Ye, L. Liu, Q. Zhao, M. Hou, L. Qiuqian, and D. Zhang. 2014. The temporal scaling of bacterioplankton composition: high turnover and predictability during shrimp cultivation. Microbial Ecology 67:256–264.

Yamamura, K., M. Yokozawa, M. Nishimori, Y. Ueda, and T. Yokosuka. 2006. How to analyze long-term insect population dynamics under climate change: 50-year data of three insect pests in paddy fields. Population Ecology 48:31–48.

Zar, J. H. 1996. Biostatistical Analysis. Third. Prentice Hall, New Jersey.

Zuur, A. F., E. N. Ieno, and C. S. Elphick. 2010. A protocol for data exploration to avoid common statistical problems: Data exploration. Methods in Ecology and Evolution 1:3–14.

Appendix S1

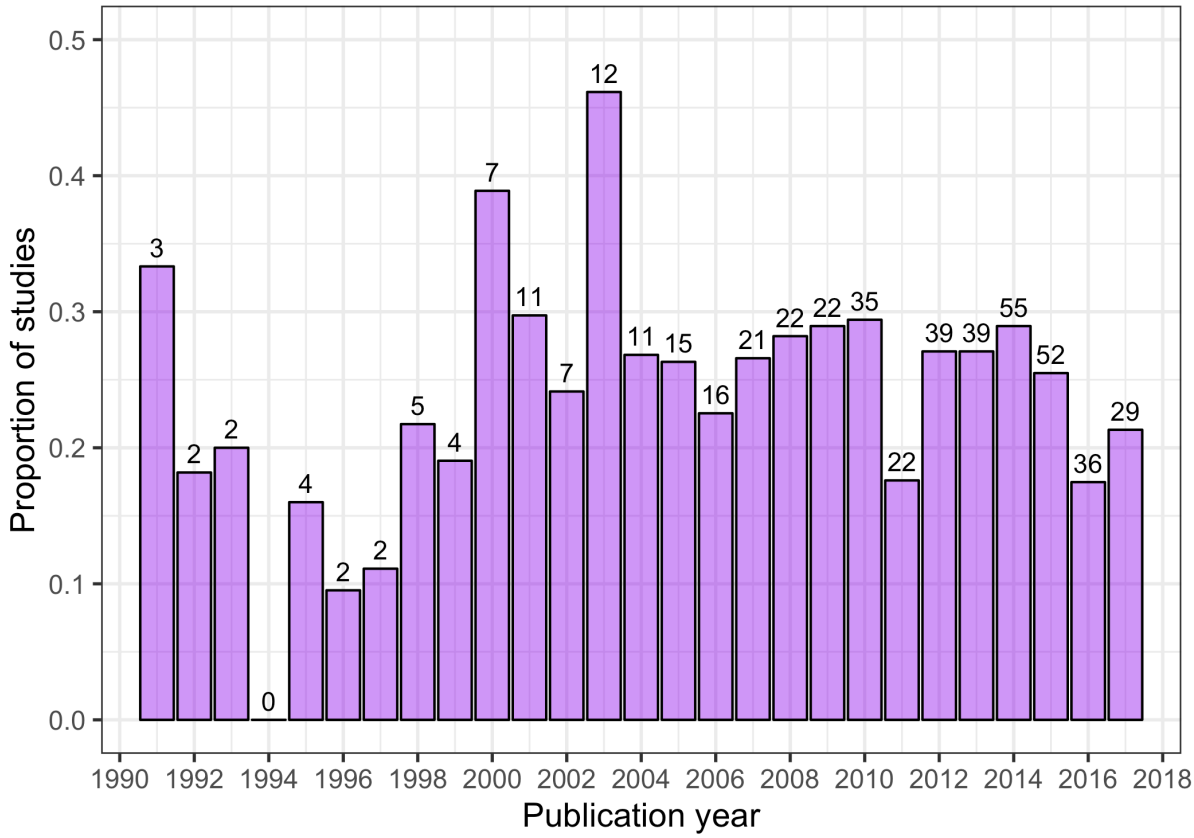
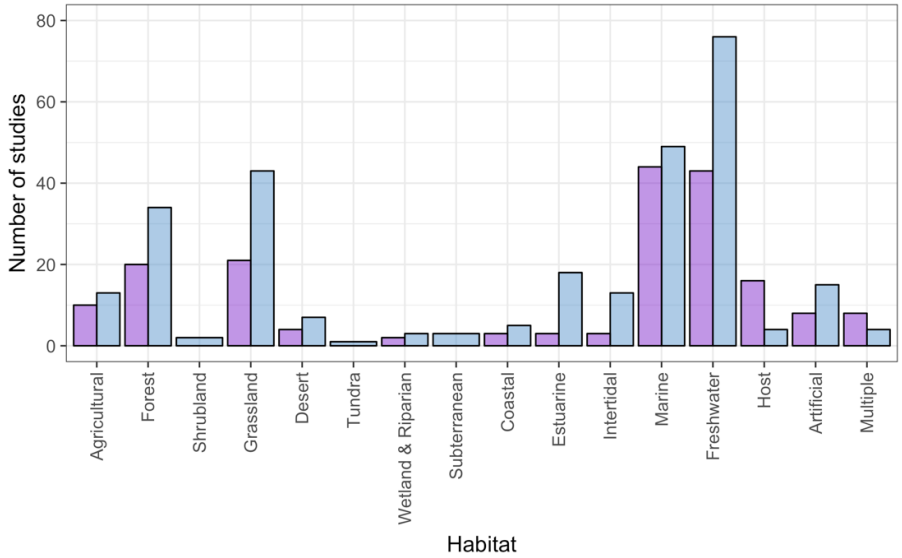


Figure S1. Proportion of studies identified using our search terms that were published each year until September 2017. No studies prior to 1991 were captured by our search terms. The total number of studies in each year is the value above the bars.

(A)



(B)

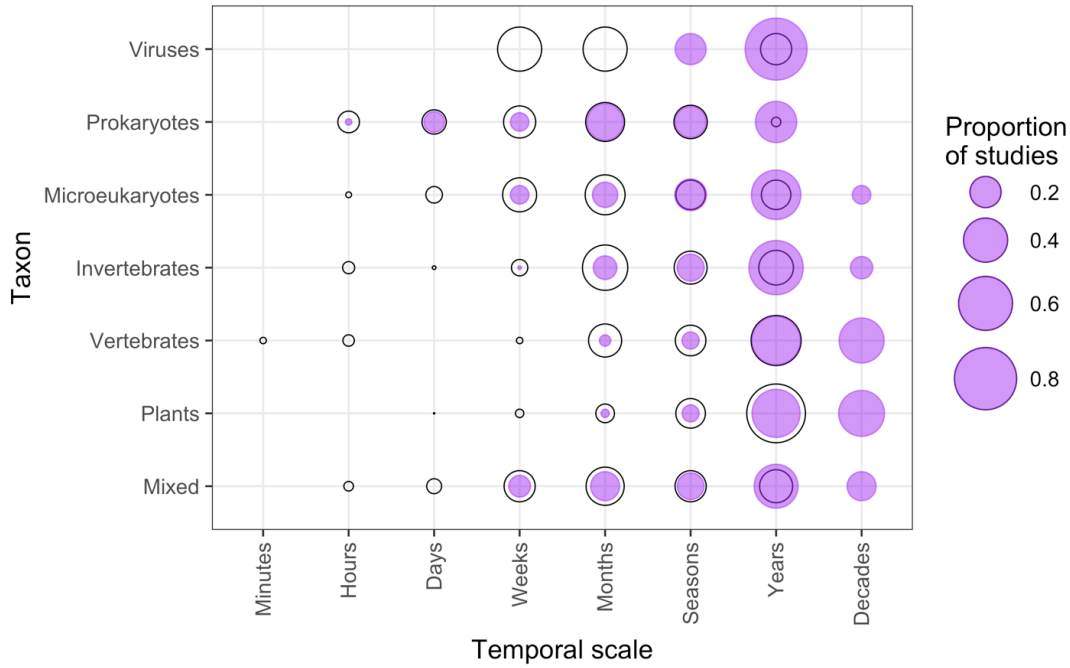


Figure S2. (A) The numbers of small and large spatial scale studies by habitat category. Small spatial scale (blue bars) studies were considered to be those carried out at micro, point sample, and local scales, while large spatial scale studies (purple bars) were those at regional, continental, or global scales. (B) The proportion of studies of different taxa within categories defined by the temporal grain (black circles) and extent (purple dots). Temporal grain is the minimum time between sampling events within the study and the temporal extent is the maximum time between sampling events within the study.

(A)

(B)

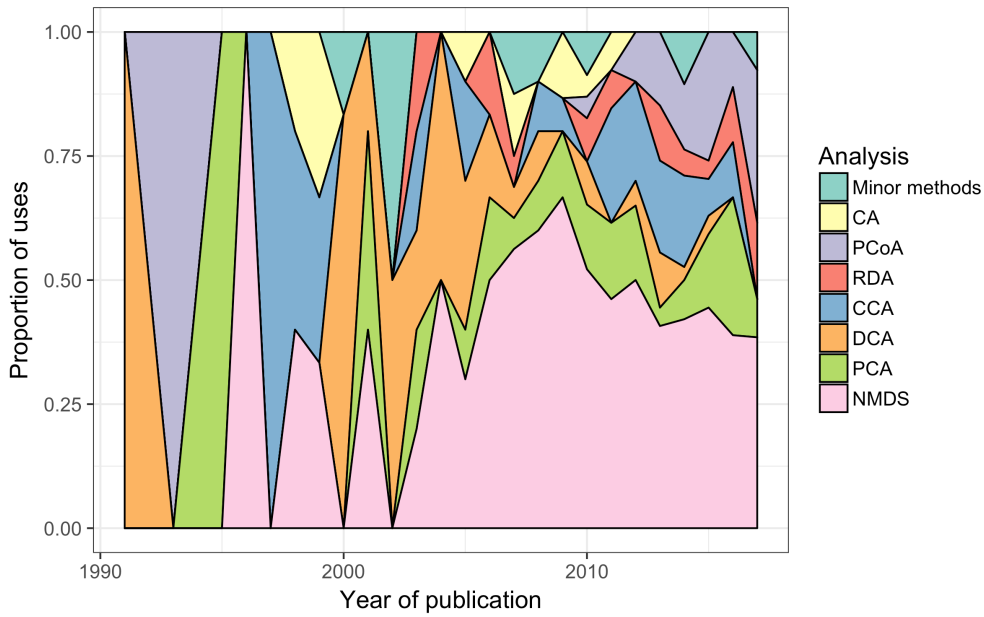


Figure S3. (A) The proportion of uses of each type of analysis across all sampled publications between 1990 and 2017. 'Minor methods' include compositional pivot days, frequency change, Markov chain modelling, multiplicative change, Moran's eigenvector maps (MEMs), Nestedness analysis, and synchrony. (B) The popularity of the different ordination methods used across the 27 years. 'Minor methods' are: dbRDA, DCCA, pCCA, pRDA, Procrustes, RA and multiple co-inertia analysis. Descriptions of analysis types are detailed the S1 Text supplementary material.

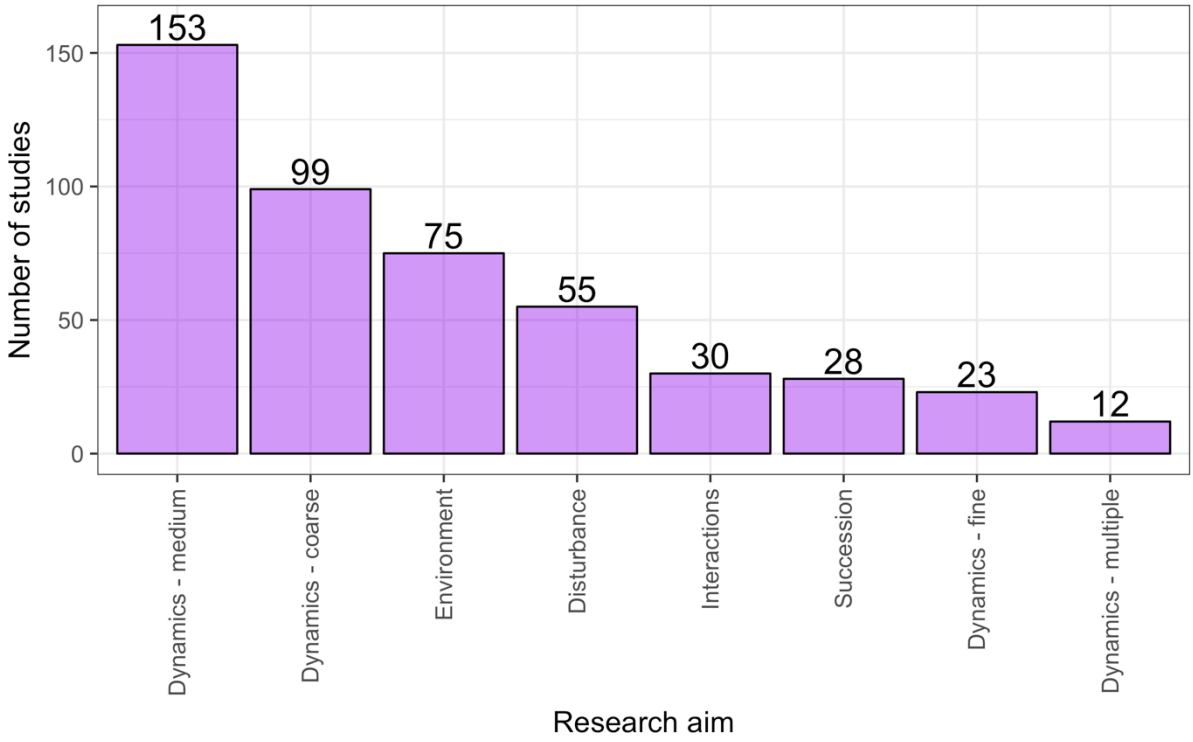


Figure S4. The distribution of studies across the different research aims. Research aims vary from understanding compositional dynamics at fine, medium and coarse temporal scales, to investigating the effect on species composition over time of species interactions, environmental conditions, and ecological disturbance, to understanding successional change in communities. Scales of temporal dynamics were classified as fine (days), medium (weeks, months, or seasons), and coarse (annual, years, or decades).

Appendix S2. Other methods encountered in the literature review, not explained in the text

UNIVARIATE LINEAR METHODS

Univariate linear methods, including analysis of variance (ANOVA), analysis of covariance (ANCOVA), repeated measures ANOVA (rmANOVA), simple correlation (e.g., Pearson's and Spearman's correlation coefficients), and regression (linear regression, generalised linear models, generalised linear mixed effects models, logistic regression, *etc.*), are used to examine temporal changes in a univariate response (dependent) variable. The independent variable is either time, or an environmental variable that changes over time, and the response variable usually is the (relative) abundance of a focal taxon or functional group.

Univariate linear methods are applied most frequently to community dynamics by constructing separate models for each species. These methods of assessing compositional dynamics are limited because multiple analyses, conducted in parallel on selected taxa, are required to make inferences regarding composition. For even modest numbers of taxa, these multiple analyses can become unwieldy and require *P*-value corrections for multiple tests (Benjamini and Hochberg 1995, Zar 1996). For example, Bugalho et al. (2013) used repeated-measures mixed-effects modelling to assess the effects of deer exclusion over time on the abundance of seedlings and saplings of forest tree species. In this case, the univariate modelling approach was usefully applied because fewer than ten species were analysed. Alternatively, abundance can be analysed for all species in the same model as long as there are enough degrees of freedom in the model to include species as a fixed or random effect term (e.g., Buckley and Freckleton 2010); see Harrell (2001) for a rule of thumb stating that we should have at least ten independent samples for each predictor variable in our model. Generalised linear models are more flexible in that they can handle designs with unbalanced data and nested sampling (e.g., plots within sites) (Bolker et al. 2009, Zuur et al. 2009). R code is provided in Appendix S4.

CLUSTER ANALYSIS (A DISSIMILARITY BASED METHOD)

Cluster analysis is a catch-all term applied to dissimilarity-based methods that either group samples together (agglomerative methods) or split all samples into sub-groups (divisive methods); the clustering is based on the distances between clusters (Legendre and Legendre

2012). There are two approaches for using cluster analysis to assess community dynamics: (1) identifications of transitions between clusters over time, and (2) examination of differences of clusters of repeated observations. In the first approach, a separate cluster analysis is computed for each temporal sample of community data (e.g., Day and Buckley 2013). In the second approach, all samples taken at different times are included in a single cluster analysis (e.g., Araujo et al. 2008). If compositional shifts are small relative to the total compositional variation in the dataset, the first approach will not be able to detect temporal differences. However, if compositional shifts are relatively large, it yields a clear way of showing shifts in 'community types': groups of samples of similar composition that can be characterised using a method such as indicator species analysis (e.g., Day and Buckley 2013). The first method can potentially be applied to a larger number of time points if all that is desired is to report the percent of sites that changed cluster groups each time. For example, cluster analysis of 34 phytoplankton species in Laoshan Bay, China revealed distinct groups of species occurred at different times over a six-month period (Xu et al. 2010).

TEMPORAL STABILITY (A.K.A. COEFFICIENT OF VARIATION)

The coefficient of variation (CV) is used to measure temporal stability of the abundance or biomass of an individual taxon, or group of taxa, across all times (not space); this is not a multivariate method (Grossman et al. 1990, Brown and Lawson 2010, Hector et al. 2010). The CV is the ratio of the standard deviation to the mean, and often is multiplied by 100 to obtain a percentage. The mean of the CV values for all taxa (e.g., abundance or biomass) is used as an aggregate measure of temporal stability for a whole community; smaller values imply greater stability. Because the CV is a ratio, changes in community composition can be compared even if there are differences in total abundances among taxa or in taxon richness among taxon groups. Such values are often presented in tables or the text as a measure of variation or are sometimes used as a response variable against other (spatial) variables of interest (e.g., Doxa et al. 2012). For example, CV can be calculated for disturbed and undisturbed communities to assess if the disturbance has resulted in a more temporally variable community. Grossman et al. (1990) categorised particular values of CV to assess stability of freshwater fish communities: a community was 'highly stable' if it had a CV < 25%, 'moderately stable' if 25% < CV < 50%, 'moderately fluctuating' if 50% < CV < 75% or 'highly fluctuating' if CV > 75%. The statistical significance of CV values could be obtained by comparing the observed values to those from communities generated using

randomisations under sensible null models. Note that temporal CV is necessarily be measured on a per site basis to exclude spatial variation.

FREQUENCY CHANGE

Changes in the frequency of individual taxa within a community over time can be assessed by subtracting their frequency at each sampling occasion from a prior sampling occasion; this is not a multivariate method. Permutation tests can be used to test the significance of these changes in frequencies (Kapfer et al. 2011). Community samples (with their given taxon frequencies) are randomised among sampling times and the change in taxon frequency recalculated with each permutation. A significant change in the observed taxon frequency is inferred if the observed change is greater than the 95th percentile change from the permutations. Permutations should be spatially restricted so that only samples from homogeneous habitat types are compared; otherwise the chance of spuriously detecting change is greatly increased. The outcome of these tests is a list of taxa that had significant increases or decreases in frequencies, and the amount, between sampling times. For example, Kapfer et al. (2011) used permutation tests to assess changes in frequencies of 85 plant species from two time points 50 years apart. This method could be used on data from multiple sampling occasions, but would quickly become cumbersome with many temporal observations. It can only be used to compare individual taxa over time, does not apply to rare taxa (taxa need to occur in enough sites to make the calculations tractable), and results in a (potentially) large table of taxa with significant (or not) changes, therefore, it has limited applicability to taxon-rich datasets where changes in rare taxa are of interest.

MULTIPLICATIVE CHANGE

The change in percent cover of a taxon, or group of taxa, at a single site can be used to obtain a ‘growth rate’ using a linear regression equation; this is not a multivariate method (Debinski et al. 2010, Collins and Xia 2015). This rate of change can then be used in a regression or ANOVA to assess predictors of change for each species (‘multiplicative change’). For example, different sites or sets of sites, can be compared in their relative change in the percent cover of different functional groups (Debinski et al. 2010), as long as communities consist of taxa with similar ecology and life-histories, e.g., grassland plant communities in different moisture regimes. This method appears to be most useful for studies where (1) there is little temporal replication, or at least where differences among temporal replicates are not

of interest, because the growth rate is calculated using only two temporal samples, and (2) few species, species groups, are of interest because a different analysis must be performed for each. However, it would be possible to calculate the change in percent cover for different time periods within the overall study, if one was interested in investigating finer-scale trends.

MARKOV CHAIN MODELLING

Markov chain modelling uses a series of transition probabilities (a 'Markov chain'), which describe how the occupancy of a taxon or functional group in a site changes from time t to time $t + 1$ (Hill et al. 2004). The user varies these probabilities in space and/or time to create a set of different hypothesis-driven models that predict the observed occupancy. The model set can then be evaluated against the observed data using log-linear modelling and AIC model selection to determine the influential and significant processes on taxon site occupancy through time (Hill et al. 2004). Community dynamics usually are examined by comparing predicted changes in individual taxa, making the method easier to apply to communities of few taxa (or groups of taxa). Because models often are parameterised using the modelled data, dependence of samples needs to be dealt with through sampling design or post-hoc statistical analyses (Hill et al. 2004). This method is most appropriate for sessile communities where patch occupancy is easy to replicate and define (e.g., Hill et al. 2004, Pueyo and Beguería 2007). It best describes directional dynamics, such as secondary succession (e.g., Baasch et al. 2010, Jiménez-Franco et al. 2011) and it is relatively computationally difficult. However, it makes detailed, taxon-level predictions that can be useful for the analysis of experimental studies where researchers want to explore specific predictions from multiple scenarios, such as taxon extinctions (e.g., Wootton 2004) or restoration manipulations (Baasch et al. 2010).

MACHINE LEARNING AND DATA MINING METHODS

Machine learning is branch of computer science that deals with the development of learning algorithms that are used to explore large, multivariate datasets, and has the primary aim of generating accurate, predictive models (Carbonell et al. 1983). Common machine-learning methods include artificial neural networks and decision-tree methods which can be used with either univariate or multivariate response variables, and Bayesian versions can incorporate prior knowledge (Cutler et al. 2007, Olden et al. 2008). Whilst the field of machine learning focusses on making predictions from data, many of its methods also are used in a data-mining

capacity to discern spatiotemporal patterns from complex and noisy datasets and can be used for classification, regression, clustering, and data reduction (Olden et al. 2008).

Machine-learning methods are attractive because they do not have many of the restrictive assumptions, such as linearity or data normality, that underpin many traditional statistical methods (Olden et al. 2008). Thus, machine-learning methods are becoming increasingly popular for exploratory and predictive analyses of community dynamics (e.g., Kampichler et al. 2010). For example, Vallès et al. (2014) used self-organising maps, an implementation of artificial neural networks (Kohonen 1982), to show that complex microbial communities sampled from infants' guts formed distinct clusters at each of five sample times (based on relative abundance data), suggesting that successional processes were occurring within the first year of these infants' lives. Sheaves et al. (2010) used a multivariate classification and regression tree analysis (De'ath 2002) to investigate spatial and temporal differences in fish assemblages; results showed that despite considerable spatial variation in community structure among different estuaries, there nonetheless were clear patterns of temporal change in species' abundances and assemblage composition. These methods are very flexible and can handle both univariate and multivariate input datasets with large numbers of predictor variables, including space and time, allowing simultaneous investigation of spatial and temporal community dynamics. Standardised approaches for characterising different types of temporal patterns and causal processes in community datasets will hopefully emerge from further research. R code is not provided, but R packages are available (the MultivariateRandomForest package: Rahman 2017, and rpart package: Therneau and Atkinson 2018).

NETWORK ANALYSIS

Ecological network analysis enables the investigation of the complex associations among species, and among species and their environment, through space and time (Newman 2010, Blonder et al. 2012). This method recognises that an ecological community is a complex biological system comprised of interconnected units whose associations can be modelled mathematically using constructs such as vertices (representing taxa) and edges, which are the connections between the vertices (representing ecological interactions) (Blonder et al. 2012). In the context of investigating community dynamics, patterns of species' associations and community memberships are most typically visualised as a topological network diagram created from taxon abundance data collected at multiple time points (*sensu* "time-aggregated

networks" (Blonder et al. 2012). From the network topology, a variety of statistics can be calculated that measure variability in community composition and interactions at particular times, which can then be compared. This method is most applicable for multi-trophic communities or where interaction networks can be clearly defined for the community; it is less useful where a single trophic level is of interest and interactions are not quantified, e.g., a plant community, but see Cram et al. (2015) for an example of spatiotemporal microbial networks. In R, the package 'timeordered' is the easiest way to implement a network analysis of community dynamics data. R code is not provided.

REFERENCES CITED

Araujo, H. M. P., D. A. Nascimento-Vieira, S. Neumann-Leitão, R. Schwamborn, A. P. O. Lucas, and J. P. H. Alves. 2008. Zooplankton community dynamics in relation to the seasonal cycle and nutrient inputs in an urban tropical estuary in Brazil. Brazilian Journal of Biology 68:751–762.

Baasch, A., S. Tischew, and H. Bruelheide. 2010. Twelve years of succession on sandy substrates in a post-mining landscape: a Markov chain analysis. Ecological Applications 20:1136–1147.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57:289–300.

Blonder, B., T. W. Wey, A. Dornhaus, R. James, and A. Sih. 2012. Temporal dynamics and network analysis. Methods in Ecology and Evolution 3:958–972.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24:127–135.

Brown, B. L., and R. L. Lawson. 2010. Habitat heterogeneity and activity of an omnivorous ecosystem engineer control stream community dynamics. Ecology 91:1799–1810.

Buckley, H. L., and R. P. Freckleton. 2010. Understanding the role of species dynamics in abundance-occupancy relationships. Journal of Ecology 98:645–658.

Bugalho, M. N., I. Ibáñez, and J. S. Clark. 2013. The effects of deer herbivory and forest type on tree recruitment vary with plant growth stage. Forest Ecology and Management 308:90–100.

Carbonell, J. G., R. S. Michalski, and T. M. Mitchell. 1983. An overview of machine learning. Pages 3–23 Machine Learning: An artificial intelligence approach. Tioga Publishing Company, Palo Alto.

Collins, S. L., and Y. Xia. 2015. Long-term dynamics and hotspots of change in a desert grassland plant community. The American Naturalist 185:E30–E43.

Cram, J. A., L. C. Xia, D. M. Needham, R. Sachdeva, F. Sun, and J. A. Fuhrman. 2015. Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. The ISME journal 9:2573.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783–2792.

Day, N. J., and H. L. Buckley. 2013. Twenty-five years of plant community dynamics and invasion in New Zealand tussock grasslands. Austral Ecology 38:688–699.

De'ath, G. 2002. Multivariate regression trees: A new technique for modeling species-environment relationships. Ecology 83:1105–1117.

Debinski, D. M., H. Wickham, K. Kindscher, J. C. Caruthers, and M. Germino. 2010. Montane meadow change during drought varies with background hydrologic regime and plant functional group. Ecology 91:1672–1681.

Doxa, A., M. L. Paracchini, P. Pointereau, V. Devictor, and F. Jiguet. 2012. Preventing biotic homogenization of farmland bird communities: The role of high nature value farmland. Agriculture, Ecosystems & Environment 148:83–88.

Grossman, G. D., J. F. Dowd, and M. Crawford. 1990. Assemblage stability in stream fishes: A review. Environmental Management 14:661–671.

Harrell, F. E. 2001. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer, New York.

Hector, A., Y. Hautier, P. Saner, L. Wacker, R. Bagchi, J. Joshi, M. Scherer-Lorenzen, E. M. Spehn, E. Bazeley-White, M. Weilenmann, M. C. Caldeira, P. G. Dimitrakopoulos, J. A. Finn, K. Huss-Danell, A. Jumpponen, C. P. H. Mulder, C. Palmborg, J. S. Pereira, A. S. . Siamantziouras, A. C. Terry, A. Y. Troumbis, B. Schmid, and M. Loreau. 2010. General stabilizing effects of plant diversity on grassland productivity through population asynchrony and overyielding. Ecology 91:2213–2220.

Hill, M. F., J. D. Witman, and H. Caswell. 2004. Markov chain analysis of succession in a rocky subtidal community. The American Naturalist 164:E46–E61.

Jiménez-Franco, M. V., J. E. Martínez, and J. F. Calvo. 2011. Territorial occupancy dynamics in a forest raptor community. Oecologia 166:507–516.

Kampichler, C., R. Wieland, S. Calmé, H. Weissenberger, and S. Arriaga-Weiss. 2010. Classification in conservation biology: A comparison of five machine-learning methods. Ecological Informatics 5:441–450.

Kapfer, J., J.-A. Grytnes, U. Gunnarsson, and H. J. B. Birks. 2011. Fine-scale changes in vegetation composition in a boreal mire over 50 years: Fine-scale changes in boreal mire. Journal of Ecology 99:1179–1189.

Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. Biological Cybernetics 43:59–69.

Legendre, P., and L. F. J. Legendre. 2012. Numerical Ecology. Third Edition. Elsevier, Radarweg, The Netherlands.

Newman, M. E. J. 2010. Networks: An introduction. Oxford University Press, New York.

Olden, J. D., J. J. Lawler, and N. L. Poff. 2008. Machine learning methods without tears: A primer for ecologists. The Quarterly Review of Biology 83:171–193.

Pueyo, Y., and S. Beguería. 2007. Modelling the rate of secondary succession after farmland abandonment in a Mediterranean mountain area. Landscape and Urban Planning 83:245–254.

Rahman, R. 2017. MultivariateRandomForest: Models Multivariate Cases Using Random Forests. R package version 1.1.5, https://CRAN.R-project.org/package=MultivariateRandomForest.

Sheaves, M., R. Johnston, and R. Connolly. 2010. Temporal dynamics of fish assemblages of natural and artificial tropical estuaries. Marine Ecology Progress Series 410:143–157.

Therneau, T., and B. Atkinson. 2018. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-12. https://CRAN.R-project.org/package=rpart.

Vallès, Y., A. Artacho, A. Pascual-García, M. L. Ferrús, M. J. Gosalbes, J. J. Abellán, and M. P. Francino. 2014. Microbial succession in the gut: Directional trends of taxonomic and functional change in a birth cohort of Spanish infants. PLoS Genetics 10:e1004406.

Wootton, T. J. 2004. Markov chain models predict the consequences of experimental extinctions. Ecology Letters 7:653–660.

Xu, H., G.-S. Min, J.-K. Choi, K. A. S. Al-Rasheid, X. Lin, and M. Zhu. 2010. Temporal dynamics of phytoplankton communities in a semi-enclosed mariculture pond and their responses to environmental factors. Chinese Journal of Oceanology and Limnology 28:295–303.

Zar, J. H. 1996. Biostatistical Analysis. Third. Prentice Hall, New Jersey.

Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith. 2009. Mixed Effects Models and Extensions in Ecology with R. Springer, New York.

Hannah L. Buckley, Nicola J. Day, Bradley S. Case, Gavin Lear, Aaron M. Ellison. 2019. Multivariate methods for testing hypotheses of temporal community dynamics. *Ecological Monographs.*

Data S1

Data extracted as part of the literature review (as .csv file)

Author(s) [of the material provided in DataS1.zip]

Hannah L. Buckley
School of Science
Auckland University of Technology
Auckland
New Zealand
Hannah.Buckley@aut.ac.nz

Nicola J. Day
Biology Department
Wilfrid Laurier University
Waterloo, Ontario
Canada

Bradley S. Case
School of Science
Auckland University of Technology
Auckland
New Zealand
Bradley.Case@aut.ac.nz

Gavin Lear
School of Biological Sciences
University of Auckland
Auckland
New Zealand

Aaron M. Ellison
Harvard Forest
Harvard University ,
Petersham, Massachusetts

USA

File list (files found within DataS1.zip)

```
Literature review data attributes.txt
Literature review data.csv
```

Description

Please write a prose description of the files, including variable descriptions, use, *etc.*

`Literature review data attributes.txt` – This text file contains attributes of the data extracted during the quantitative literature review, providing a description of the data found in the associated spreadsheet (Data S1.csv).

`Literature review data.csv` – This is a .csv spreadsheet of the data extracted as part of the quantitative literature review carried out in this study.

Hannah L. Buckley, Nicola J. Day, Bradley S. Case, Gavin Lear, Aaron M. Ellison. 2019. Multivariate methods for testing hypotheses of temporal community dynamics. *Ecological Monographs.*

Data S2

Worked case study example 1 R code files

Author(s) [of the material provided in DataS2.zip]

Hannah L. Buckley
School of Science
Auckland University of Technology
Auckland
New Zealand
Hannah.Buckley@aut.ac.nz

Nicola J. Day
Biology Department
Wilfrid Laurier University
Waterloo, Ontario
Canada
njday.ac@gmail.com

Bradley S. Case
School of Science
Auckland University of Technology
Auckland
New Zealand
Bradley.Case@aut.ac.nz

Gavin Lear
School of Biological Sciences
University of Auckland
Auckland
New Zealand
g.lear@auckland.ac.nz

Aaron M. Ellison
Harvard Forest
Harvard University

Petersham, Massachusetts
USA
aellison@fas.harvard.edu

File list (files found within DataS2.zip)

`Worked example 1.txt`

Description

`Worked example 1.txt` – R code for the worked case study example 1: Testing for significant temporal community dynamics within a controlled experiment. Further description and instruction can be found within the R script as annotation.

Hannah L. Buckley, Nicola J. Day, Bradley S. Case, Gavin Lear, Aaron M. Ellison. 2019. Multivariate methods for testing hypotheses of temporal community dynamics. *Ecological Monographs.*

Data S3

Worked case study example 2 R code file

Author(s) [of the material provided in DataS3.zip]

Hannah L. Buckley
School of Science
Auckland University of Technology
Auckland
New Zealand
Hannah.Buckley@aut.ac.nz

Nicola J. Day
Biology Department
Wilfrid Laurier University
Waterloo, Ontario
Canada
njday.ac@gmail.com

Bradley S. Case
School of Science
Auckland University of Technology
Auckland
New Zealand
Bradley.Case@aut.ac.nz

Gavin Lear
School of Biological Sciences
University of Auckland
Auckland
New Zealand
g.lear@auckland.ac.nz

Aaron M. Ellison
Harvard Forest
Harvard University

Petersham, Massachusetts
USA
aellison@fas.harvard.edu

File list (files found within DataS3.zip)

`Worked example 2.txt`

Description

`Worked example 2.txt` – R code for the worked case study example 2: Community dynamics analysis of the BCI dataset. Further description and instruction can be found within the R script as annotation.

Hannah L. Buckley, Nicola J. Day, Bradley S. Case, Gavin Lear, Aaron M. Ellison. 2019. Multivariate methods for testing hypotheses of temporal community dynamics. *Ecological Monographs.*

Data S4

Worked case study example 3 R code files

Author(s) [of the material provided in DataS4.zip]

Hannah L. Buckley
School of Science
Auckland University of Technology
Auckland
New Zealand
Hannah.Buckley@aut.ac.nz

Nicola J. Day
Biology Department
Wilfrid Laurier University
Waterloo, Ontario
Canada
njday.ac@gmail.com

Bradley S. Case
School of Science
Auckland University of Technology
Auckland
New Zealand
Bradley.Case@aut.ac.nz

Gavin Lear
School of Biological Sciences
University of Auckland
Auckland
New Zealand
g.lear@auckland.ac.nz

Aaron M. Ellison
Harvard Forest
Harvard University

Petersham, Massachusetts
USA
aellison@fas.harvard.edu

File list (files found within DataS4.zip)

`Worked example 3.txt`

Description

`Worked example 3.txt` – R code for the worked case study example 3: Cyclical dynamics example from Gilbert et al. (2012). Further description and instruction can be found within the R script as annotation.