Towards an integration of scale and complexity in marine ecology

JON D. WITMAN,1,3 ROBERT W. LAMB,1 AND JARRETT E. K. BYRNES2

1Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912 USA
2Department of Biology, University of Massachusetts Boston, Boston, Massachusetts 02125 USA

Abstract. Manipulative field experiments provide a window into the complexity of nature. Yet there is concern that we lack resolution by conducting experiments on a scale that is too small and short to include the relevant complexity of the study system. We addressed this issue by asking how and why the scale (local and global spatial extent, spatial grain, duration) and complexity (number of species, factors, treatment combinations) of experiments performed on marine hard substrata (rocky intertidal, RI; coral reef, CR; rocky subtidal, RS; mangrove root, MR) has changed by assessing 311 total experiments published since 1961 in Ecology and Ecological Monographs and since 1967 in Journal of Experimental Marine Biology and Ecology. We show that the local spatial extent and all metrics of complexity increased as a positive, log-linear function of time. In contrast, the size of experimental units (spatial grain) decreased with time. Quantile regression analysis revealed that these trends are largely driven by changes in the upper bounds of experimental scale and complexity; most studies are still relatively simple in design and conducted over small areas. A structural equation model (SEM) incorporated the direct and indirect effects of six metrics indicating that the complexity of field experiments has increased both as a direct effect of time and because experiments have become smaller in spatial grain. The SEM also showed longer experiments tended to be more complex. We show striking habitat differences, as subtidal experiments (CR, RS) involved more species and were carried out on the largest global spatial scales. RI experiments were the longest.

Future prospects to incorporate more of the complexity of nature into field experiments include site replication, as only 34.7% of all experiments were conducted at more than one site, open experimental designs monitored by technology, and integrating experimental manipulations with long-term monitoring to achieve mechanistic insight across scales relevant to human alteration of the biosphere. The increasing resolution of remote sensing also creates opportunities to track experiment-driven changes in community structure across large scales. We suggest applying these methods to a wider class of problems to enhance our understanding of marine communities and ecosystems.

Key words: complexity; coral reef; experimental design; field experiments; human impacts; long-term monitoring; mangroves; marine community ecology; rocky intertidal; rocky subtidal; scale.

INTRODUCTION

Ecological experiments, by necessity, contain a subset of the biodiversity, species interactions, and processes present in natural communities. Given this constraint, how can ecologists develop a more comprehensive understanding of nature from manipulative field experiments? In this era of climate change (Field et al. 2014), how can manipulative experiments and comparative field research efforts be designed to expand the knowledge of climate and human impacts on species interactions, biodiversity, and the functioning of ecosystems?

While the need to increase the spatial and temporal scale of community, ecosystem-level research has been emphasized repeatedly (Dayton and Tegner 1984, Levin 1992, Lawton 2000, Witman and Roy 2009), it is not readily apparent how mechanistic insight can be achieved about complex ecological systems at large scales.

We address these general questions and related issues for communities and ecosystems living on hard substrate habitats of the ocean, including rocky intertidal and subtidal zones, coral reefs, and mangrove roots. Spatial and temporal scaling of ecological processes in marine soft substrate habitats has received wide attention (Schneider et al. 1997, Thrush et al. 1997, 2012, 2014). The scale and complexity of experiments designed to untangle effects of different interacting abiotic and
A) Scale

- **Duration**: length of experiment in days, range 0.08–4380 d.
- **Spatial grain**: grain size (area in m²) of an experimental unit; basal area used for 3D structures (cages), range 0.0001–1173 m².
- **Spatial extent**: maximum distance (m) between experimental units within a site, range 1–2000 m.
- **Global spatial extent**: maximum distance (km) between study sites, range 0.015–15130 km.

B) Complexity

- **No. species potentially interacting with response variables**: compiled from tables of species lists or description of experiment, range 1–399 species.
- **No. factors (synonymous with treatment)**: from described experimental design or statistical tables; time was considered a factor, as was block in blocked designs, range 1–13 factors.
- **No. treatment combinations**: total no. unique combinations assigned to an experimental unit, typically the product of the number of treatment levels between factors in a fully crossed design, range 1–60 combinations.

Table 1. Definition of the metrics used in the review of 311 total experiments published since 1961 in *Ecology* and *Ecological Monographs* and since 1967 in *Journal of Experimental Marine Biology and Ecology.*

Biotic drivers, less so. Thus, we focused on the scale and complexity of experiments conducted on hard substrate marine habitats, which have been instrumental to the development of marine community ecology (Bertness et al. 2014).

The general connotation of biological complexity is one of systems composed of multiple parts at different levels of organization interacting in multiple ways. It is readily apparent from a diagram of feeding interactions in a community, energy flow through an ecosystem or from a trophic cascade containing consumptive and nonconsumptive effects that complexity is an inherent property of natural systems. The idea of nature as a complex system that can be reduced to a simpler set of laws or properties traces back to Aristotle (Boylan 1984) and is a cornerstone of the philosophy of science (Radder 2003).

In conceptual frameworks of biological complexity, Cadenasso et al. (2006) identified three primary dimensions (spatial heterogeneity, connectivity, and history), while Loehle (2004) maintained that that there were at least six dimensions, including spatial, temporal, structural, process, behavioral, and geometric complexity. Space, time, and an aspect of structure or organization are central elements in most frameworks. Spatial and temporal scaling of biological properties provides a fruitful way to identify thresholds and emergent properties and to extract general patterns from complex systems (Wiens 1989, Levin 1992, Schneider 1994, Allen 2001, Benedetti-Cecchi et al. 2012, Stier et al. 2014).

We assess trends in the scale and complexity of manipulative experiments published in *Ecology* and *Ecological Monographs* from 1961 to 2014 and the *Journal of Experimental Marine Biology and Ecology* from 1967 to 2014 in order to set the stage for recommendations about future research. We ask how and why field marine experiments have changed over time. As shown in Table 1, we define the scale of experiments by traditional units of space and time (Wiens 1989) and the complexity of experiments by the number of species involved, the numbers of factors, and the number of treatment combinations in the experimental design. Thus, unlike Loehle (2004), we separate scale from our analysis of complexity in the review of published experiments, but discuss ways that they can be integrated in future experimental designs. To make the review tractable, we constrained the search to all papers published in consecutive four-year periods per decade beginning with the first issues through December 2014 (see Methods) from experiments conducted in rocky intertidal (RI), rocky subtidal (RS), coral reef (CR), and mangrove root (MR) habitats. To our knowledge, this is the first time that the scale–complexity space of experiments in hard substrate marine communities has been measured and analyzed.

We then highlight future prospects to reduce the mismatch between the scales and complexity of nature and experiments through innovative designs of field experiments, by combining experiments with long-term field observations, and by the application of new technologies. A main objective of these approaches is to enable more species interactions and processes to influence the response variables (i.e., more complexity) at larger scales of space and time, ultimately to obtain a broader understanding of marine communities and ecosystems.

In so doing, we do not simply advocate that more complex experimental designs are inherently better than simple ones. Theoretically, increased complexity in experimental design comes with trade-offs imposed by statistics and logistics. Complex experiments with many treatments may have shortcomings over simpler ones, including lower statistical power if replication is reduced in order to increase the number of treatments, non-independence (Hurlbert 1984, Underwood 1997, Gotelli and Ellison 2013), and reduced effect size (Leuzinger et al. 2011). In addition, complex factorial designs can be difficult to analyze, although options are available including pooling nonsignificant factors (Winer et al. 1999, Underwood 1997, Quinn and Keough 2002).
Clearly, one cannot easily manipulate six factors in a field experiment successfully at many places and times without a large research budget and a Herculean effort. The advantages, however, of expanding the scale and complexity of carefully designed experiments and observational programs are that the scale of generalization may be increased. More complex experiments also allow for more detailed exploration of mechanism. Elegant yet complex designs make it less likely that important direct or indirect effects, interactions, or feedbacks will be missed compared to an over-simplified research effort. By reviewing trends in the scale of past experimentation here, we hope to foster a better appreciation for the need to conduct research at spatiotemporal scales appropriate to the hypotheses being tested (Dayton and Tegner 1984, Underwood and Petraitis 1993, Benedetti-Cecechi et al. 2012).

Ecological experiments in field marine ecology are analogous to small black boxes. The experimentalist removes or adds a factor from the box and, after a period of time, measures the response of the community or ecosystem. Typically, we cannot see how the manipulated factor translates to a measured response. Further, if a factor isn’t included in the box, its influence on the response variable or its interaction with the factors we have added cannot be measured. Our efforts are meant to explore how we can control more factors in the box, to augment the measurement of other variables with potentially causal effects on the outcome, and to increase the size of the box. Ultimately, we wish to understand the inner workings of a biological system, and hence must render the experimental box transparent to the tools of science.

METHODS

Historical analysis of experimental designs
We searched the literature of studies published in *Ecology* and *Ecological Monographs* and *Journal of Experimental Marine Biology and Ecology* for manipulative field experiments conducted in the four hard substrate habitats, the rocky intertidal (RI), rocky subtidal (RS), coral reefs (CR), and mangrove root (MR), from 1961 to 2014. To qualify for inclusion in the survey, the experiment had to be conducted at the community or ecosystem level, defined as involving two or more interacting species. The search was constrained to four-year blocks per decade, which encompassed the years 1961–1964, 1971–1974, 1981–1984, 1991–1994, 2001–2004, and 2011–2014 for *Ecology* and *Ecological Monographs*. The search of publications in *Journal of Experimental Marine Biology and Ecology* began with the start of the journal in 1967, so the four-year blocks consisted of 1967–1970, 1971–1974, 1981–1984, 1991–1994, 2001–2004, 2011–2014. No papers were published in *Journal of Experimental Marine Biology and Ecology* on experimental hard substrate community ecology from 1967–1970. The search yielded 311 experiments from 229 papers (Appendix A). The greatest number of experiments came from the rocky intertidal (126), representing 40.5% of the total, followed by the rocky subtidal (92 experiments, 29.6%), coral reefs (72, 23.1%), and mangrove root habitats (16, 5.1%). Three studies contained experiments spanning both rocky subtidal and coral reef habitats, while two studies transitioned rocky intertidal-subtidal habitats; these were excluded from habitat-based comparisons. We placed two deep-sea studies in the rocky subtidal category (Micheli et al. 2002, Mullineaux et al. 2003) because they were conducted on rocky substrates. Only two papers were published from manipulative experiments performed in subtidal habitats (CR, RS) during the 1961–1964, 1967–1970, and 1971–1974 time periods, reflecting the later development of ecological research requiring SCUBA (Witman et al. 2013).

Four metrics of scale (experiment duration, spatial grain, spatial extent, global spatial extent) and three metrics of complexity (number of species potentially interacting with the response variable, number of factors, treatment combinations) were determined as accurately as possible from each experiment (Table 1). The specific questions we attempted to answer with these analyses were as follows. (1) Does the scale and complexity of experiments vary over time from 1961–2014? (2) If so, are these relationships linear or nonlinear? (3) If they are nonlinear, can thresholds of scaling or complexity be identified? (4) What is the nature of the relationship between scale and complexity in marine field experiments? Possibly the most obvious alternative to the null hypothesis of no relationship is that the complexity of experiments is inversely related to spatial or temporal scale, but this has never been quantitatively tested. (5) Are there indirect effects of spatial grain, extent, and duration on the complexity of experiments, in addition to the effects of time? (6) Are there trade-offs in experimental design suggested by inverse relationships among the seven metrics? (7) Are there habitat-based (coral reefs, rocky intertidal, rocky subtidal, mangroves) differences in the complexity and scale of manipulative field experiments? (8) Do any of the general trends identified depend on the journal that the papers were published in (i.e., Ecological Society of America publications *Ecology* and *Ecological Monographs* vs. *Journal of Experimental Marine Biology and Ecology*?).

To determine how experiments have changed in complexity over time, we used linear regression analysis (generalized linear models). We began with a generalized linear model analysis (McCullagh and Nelder 1989, O’Hara 2009) seeking to capture simple temporal trends over time. We fit models with log-link functions and gamma, quasi-poisson, or negative binomial error terms to accommodate overdispersed errors. Inspecting results, we noted that while trends were often present, they appeared to be driven by shifts in the upper limits of experimental scale or complexity, rather than changes in
experimental design across all studies. To characterize whether there were general shifts across all experiments or only shifts in the upper echelons of experimental complexity, we reanalyzed the temporal trends in the data using quantile regression (Cade and Noon 2003, Koenker 2013). For these models, we looked at the relationship between year and log(x+1)-transformed response variables at a wide variety of quantiles (tau). Finally, we used two-way factorial generalized linear models (GLM) with each variable as a response to test for differences in mean values between decades used as a factor. Likelihood ratio chi-square tests were run on these models using type II model comparisons and the same error distribution families and link functions as established in the temporal trends analysis. When significant differences between decades were established, post hoc comparisons of all treatment combinations with a false-discovery-rate-corrected alpha were used to make pairwise comparisons (Hothorn et al. 2008).

To explore why experimental complexity may have changed over time, we fit a structural equation model (SEM; Bollen 1989, Grace 2006) to several variables in our data set based on an a priori hypothetical model of how the complexity of experimental designs might have evolved over time. In the SEM analysis, we sought to ask how different design choices affect experimental complexity. As this is a concept for which we have some measurements (more complex experiments have more species, more factors, and more treatment combinations) we represented it as a latent variable indicated by those three metrics. We examined the direct effect of time on complexity, as well as its indirect effect via shifts in the log of spatial grain, extent, and duration of experiments. We also accounted for experiments with a larger spatial grain naturally having a larger spatial extent and the mathematical correlation between number of factors and number of treatment levels. We fit the model to the covariance of the data using likelihood-based SEM using the lavaan package in R (Rosseel 2012). We used log(x+1)-transformed data, as SEM is a fundamentally linear-based technique. We could not use piecewise SEM with generalized linear models (Grace et al. 2012) due to the inclusion of latent variables. As all relationships were log-linear, this transformation provided a reasonable solution to the inclusion of nonlinear relationships. Results matched GLM analyses largely, and residuals, with the exception of spatial grain and treatment combinations, matched assumptions of normality. To accommodate this violation of assumptions of multivariate normality when we assessed whether our multivariate hypothesis was consistent with the data, we tested the model for fit and corrected standard errors of coefficients using the Satorra-Bentler correction (Satorra and Bentler 1988, Bentler and Dudgeon 1996).

Last, we looked at the differences between different habitats and journals using two-way factorial generalized linear models as before. We used chi-square likelihood ratio tests, with habitat type (CR, MR, RI, RS) and journal (ESA, JEMBE) as fixed effects. We evaluated whether these factors drove response variables and conducted post hoc tests on factors as before. In the case of significant main effects, we made all pairwise comparisons using a false-discovery-rate-adjusted alpha to determine which groups differed. When interaction terms were present, we only analyzed differences between habitats within journals.

RESULTS

Temporal patterns

Scale.—There were significant positive relationships between time and two of the four measures of the scale of experimental manipulations (Fig. 1B, C). For example, the spatial extent, indicating the size of the area that experimental units were spread across at a local site, increased significantly with time from the 1960s to the present (Fig. 1C). The maximum spatial extent was from coral reef (CR) habitats with Edmunds et al.’s (2014) study of refuge effects on coral recruitment encompassing 2500 m². Similarly, the largest global spatial extent (15310 km) was from Lindquist et al.’s (1992) study of chemical defenses against fish predation in CR habitats across the Pacific (Appendix A). However, there was no significant temporal trend in the global spatial extent of experiments (Fig. 1D). As indicated in Table 2, 65.3% of the total number of field experiments assessed were conducted at only one site, yielding a zero value for global spatial extent. Site replication appeared in the database for the first time in the 1971–1974 period and increased to represent 59.4% of all experiments in 2011–2014.

The second significant scale relationship was that the grain size of experiments decreased from larger, coarse-grained to smaller, fine-grained experimental units from 1961–2014 (Fig. 1B). This was in contrast to all the other significant temporal patterns in the data, which increased over time. The decrease in grain size is likely due to the difference between experiments performed during 1971–1974 and 2011–2014, as experimental units in 1971–1974 often consisted of large areas (reefs) of intertidal shore where a consumer species was removed or added (Paine 1971, Menge 1974), while contemporary (2011–2014) experimental units were generally less than 1.5 m². For example, the average grain size in 1971–1974 was 120.5 m² (96.3 SE) vs. 1.1 m² (0.4 SE) in 2011–2014 (Appendix B). There was no significant relationship between the duration of experimental manipulations and year (Fig. 1A; Appendix C). Ecologists were performing multiyear experiments in the 1960–1980s. The longest experiment was a 12-year Pisaster removal experiment (Paine 1984; Appendix A). Apparently, experiments performed in the last four years are no longer on average than those in the early 1970s (Fig. 1A; Appendix B).

Complexity.—All three of the complexity metrics (the number of species, factors, and treatment combinations) increased significantly with time (Fig. 1E–G), indicating that the complexity of manipulative experiments in
marine community ecology has increased since the early years of the field in the 1960s. As indicated by the coefficients of determination in the significant regression analyses, time, represented as the year of publication, explained a very small amount (0.05–2.5% of the variation in scale and complexity (Appendix C) due to extremely large variation among the data within year groups. The data in the plot of year against the number of factors (treatments) showed a large increase in the number of factors between 1961–1964 and 1971–1974 (Fig. 1F). The most complex experiment during this time period was Dayton’s (1971) caging and dog dish experimental design with six factors testing for biological disturbance and competition. Todd and Keough’s (1994) larval settlement experiments in epifaunal communities provided the highest overall value with 13 factors (Appendix A).

Similarly, the positive relationship between time and the number of species involved in the experiments was characterized by an exceptionally large range in the number of species in the 2011–2014 time period (Fig. 1E). For example, several experiments involved only three species, while Freestone and Osman’s (2011) test of latitudinal predation gradients with fouling plates involved up to 399 species, and Messmer’s (2011) experiment on fish community assembly involved 225 species.

Another sign of the increasing complexity of manipulative experiments was indicated by the growing number of treatment combinations (Fig. 1G), indicating that more levels of manipulated factors, such as different densities of consumers or predators were incorporated into field experiments. In multiple experiments from the 2001–2004 and 2011–2014 groups, the increased species richness represented biodiversity–ecosystem functioning experiments (Appendix A). The increase in complexity involved up to 399 species, and Messmer’s (2011) experiment on fish community assembly involved 225 species.

Table 2. Site replication in marine ecological experiments by year block.

<table>
<thead>
<tr>
<th>Year block</th>
<th>No. replicated studies</th>
<th>Total no. studies</th>
<th>Site replication (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1961–1964</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>1971–1974</td>
<td>6</td>
<td>12</td>
<td>50</td>
</tr>
<tr>
<td>1981–1984</td>
<td>11</td>
<td>45</td>
<td>24.44</td>
</tr>
<tr>
<td>1991–1994</td>
<td>19</td>
<td>57</td>
<td>33.33</td>
</tr>
<tr>
<td>2001–2004</td>
<td>35</td>
<td>99</td>
<td>35.35</td>
</tr>
<tr>
<td>2011–2014</td>
<td>37</td>
<td>92</td>
<td>40.22</td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
<td>311</td>
<td>34.73</td>
</tr>
</tbody>
</table>
might also reflect the growth in regression-based designs for experiments (Cottingham et al. 2005).

Scale and complexity

Analysis of mean values by decade for each variable reinforced the general trends apparent in the generalized linear models as significant differences in mean values between decades were observed for species, factors, and spatial grain (Fig. 2; Appendix D). The mean number of species incorporated into field experiments was consistently higher during the 1990–2010s than in the previous three decades (Fig. 2; Appendix D). In contrast, the mean spatial grain of experimental units in the 2010 group was significantly smaller than it has been in any of the past four decades. The mean number of factors was significantly greater in the 2000s than in the 1980s.

Still, the low percentage of variation explained by generalized linear models (Appendix C) raises questions of whether marine ecology as a discipline is changing its practices or whether as time goes on, a handful of investigators are executing progressively more complex designs. Quantile regression results indicate that is the case for some variables, but not all (Fig. 3; Appendix E). Global spatial extent, treatment combinations, and to some extent the number of species have all gone up at upper, but not always at lower, quantiles. In contrast,
the spatial extent of experiments has increased for almost all quantiles, indicating that as a discipline, we are performing experiments at progressively larger spatial scales. The number of factors appears to have only increased at intermediate quantiles, indicating a general trend, but one not reflected at the extremes. Curiously, spatial grain appears to have decreased at upper quantiles, likely driven by a handful of early
coarse-grained experiments, such as Menge (1974). Experimental duration appears to have not changed at all over time.

As shown in Fig. 4A, the number of treatment combinations increased simultaneously with the local spatial extent (m) of the seascape over which the experiments were conducted since 1961–1964. The experiments published in 1961 by Castenholz (1961) and Connell (1961) contained few treatment combinations and small local spatial extents (Fig. 4A; Appendix
A). There was a large increase in spatial scale (extent) and complexity (treatment combinations) from the early 1960s to the 1970s time period. The average spatial extent of some of the early 1970s experiments was a 12-fold increase over a decade earlier (Appendix F; Figs. 2 and 4A).

Similarly, the number of species potentially interacting with the response variable expanded over time and with the local spatial extent of the experiments (Fig. 4B). The most notable feature of the most recent (2011–2014) experiments was the large number of species in manipulations performed at large spatial extents (Fig. 4B).

Examining three variables simultaneously revealed an interaction between the grain size of the experimental units, the complexity of the experiment (number of species affecting response variable), and the global spatial extent (Fig. 5). From the early 1970s to the present, the grain size of the experiments decreased as the experiments increased in complexity while more and more of them were replicated across multiple sites, increasing the global extent. This suggests a potential trade-off of experimental macroecology (Witman and Roy 2009) between the size, complexity, and broad-scale replication of manipulative experiments.

Our SEM (Fig. 6; see Appendix G for coefficients) shows that as ecology has progressed, experiments have grown in complexity. They have grown in complexity both on their own (direct effect of time) and because as the discipline has progressed, experiments have grown smaller in spatial grain size. Smaller experiments tend to be more intricate and have higher overall complexity. Experiments that are longer in duration also tend to be more complex, although we have not been running longer experiments as we have grown as a discipline, echoing results from the quantile regression analyses. While larger grain-sized experiments tended to be larger in spatial extent, this did not lead to additional experimental complexity. This model fit the data ($\chi^2_{SB} = 15.928$, df = 9, $P = 0.068$; $\chi^2_{SB}$ refers to the Satorra-Bentler corrected chi-square test statistic) and was no different from a model allowing variables to influence measurements of complexity directly ($\chi^2_{SB} = 12.05$, df = 8, $P = 0.100$). The latent variable of complexity was itself well explained with $R^2 = 0.387$.

**Habitat and journal effects**

Chi-square tests analyzing habitat type and journal (ESA, JEMBE) indicated significant main effects of habitat type on local spatial extent, study duration, and number of species. Values were significantly different between journals for spatial grain and global spatial extent (Fig. 7; Appendix H). A significant interaction between habitat and journal occurred for local spatial extent, global extent, number of species potentially affecting the response variable, and treatment combinations. This last variable was not tested further due to the

---

**FIG. 5.** Three-dimensional scatter plot showing trends in two aspects of spatial extent (spatial grain, global spatial extent) and one measure of complexity (number of species potentially interacting with the response variable) with the time period of publication color-coded. The time series starts in 1971–1974; none of the experiments in 1961–1964 were site-replicated. Values were $\log(x + 1)$-transformed. The trend illustrated is that the complexity and global spatial scale of the experiments increased as spatial grain (measured as m$^2$) of the experiments decreased (from coarse-grained 1970s, blue dots to fine-grained 2010s, red dots) visualized from the upper right to lower left of the plot. Axes labeled “log” have been $\log(x + 1)$-transformed; spatial grain was originally measured in m$^2$.
DISCUSSION

Historical analysis of experimental designs

We began our analysis of trends in the scale and complexity of manipulative experiments conducted in hard substrate marine communities without prior knowledge of any general relationships or how they have changed over the past six decades.

Multiple lines of evidence indicated that field experiments have been growing in spatial extent and complexity over time. For example, our analysis indicated that local spatial extent increased as a positive function of year of publication since 1961 (Fig. 1; Appendix C). Similarly, the analysis indicated that all measures of complexity, the numbers of species, factors, and treatment combinations, involved in the experiments increased with time (Fig. 1; Appendix C). A common feature of these models is that they are significant, however, time (year of publication) explained a very low percent of the variation in the response variable, ranging from 0.85% to 2.57% (Appendix C). This reflects the enormous variation in each of these metrics within year groups, i.e., there was up to a 2500-fold range in spatial extent within a particular year group (Fig. 1C).

At the same time, there are unmistakable trends in the upper ranges of the data over time that were revealed by quantile regression models (Fig. 3). These models indicate that the upper bounds of the spatial extent and number of treatment combinations in the manipulations have been increasing over time (Fig. 3). Our interpretation of this pattern is that researchers have been conducting simple experiments with few treatment combinations over small spatial extents at the same time that more complex experiments with larger number of treatment combinations have been performed across larger spatial extents. Both simple and complex experiments hold places of distinct utility within marine ecology. Early simple experiments allowed ecologists to dissect single mechanisms. The growing appreciation for the complexity of real-world systems, however, has led ecologists to confront this problem head on. We are now seeing a better mix of experiments that can tease out a single mechanism at a single time and place and those that can address several mechanisms at once over larger areas. We are encouraged for the future of the field and hope this trend continues with more equal weight to both the simple and complex.

The SEM identified nonintuitive ways that grain size affected complexity, as experiments with larger (coarser) grain tended to be less complex, yet coarser-grained experiments tended to be larger in spatial extent. Surprisingly, the spatial grain of the experiments decreased with year of publication (Figs. 1–3). The SEM integrated the individual measures of scale and complexity, showing overall that experiment complexity has increased over time, partly driven by this decrease in grain. This growth is both due to ecologists designing increasingly complex experiments, as well as ecologists performing smaller-grained, more intricate experiments over time. This corresponds with an increase in spatial scale due to a decrease in grain size. Put another way, experiments with small plots run at small scales were simple. But experiments with small plots run over large spatial extents (areas), such as the scale of a whole reef, tended to incorporate far more complexity. But it was ultimately the grain size of plots that determined the complexity of the manipulation. Scale and complexity covary merely due to grain size.

Analysis of the mean values of each of the scale and complexity metrics broke the continuous temporal patterns revealed by regression into discrete comparisons by year group. There were peaks in the data averaged within decades (Fig. 2) that likely reflected the type of questions being asked. For example, the maximum area of experimental units (spatial grain) occurred in the 1970s, (Figs. 1B and 2B; Appendix B) as researchers began to apply the knowledge gained from initial manipulative studies in the 1960s, which were small-scale species exclusions and removals across large rocky intertidal slopes and platforms (reefs). While spatial grain has declined, average global extent peaked in the 1990s, displaying an 8.8-fold increase between the 1980s and the 1990s groups (Fig. 2D; Appendix B). One metric of complexity, average number of species,
doubled during the same period (Fig. 2E; Appendix B). We suggest that the early 1990s period represents a "step change" in ecological experimentation, reflecting a concerted effort in the 1990s to expand the spatial scales of ecological inference. This step change follows an emphasis on supply-sided ecology (Gaines and Roughgarden 1985, Roughgarden et al. 1988) and oceanographic influences in benthic ecology (Dayton and Tegner 1984, Menge 1992, Witman et al. 1993). This change is also coincident with the development of macroecology (Brown 1985, Witman and Roy 2009). A caveat about trends in one aspect of complexity, the number of species potentially influencing the response variable, is that it could reflect a geographic bias in the location of studies. For example, complexity could increase if more studies were conducted in the species-rich tropics or decrease if carried out in lower diversity regions.

Another possible explanation for the increase in the complexity of experiments in marine ecology over time is that it simply reflects the maturation of a scientific field (Carmel et al. 2013), which could lead to a greater number of factors considered in experiments on larger spatiotemporal scales. This dispersion is thus driven by the growth of a number of subdisciplines, each of which focused on maximizing a different aspect of experimental complexity or scale. We cannot discount this explanation without further investigation. Since the initial emphasis on competitive interactions in the 1960s group (Connell 1961a), the importance of predation (Paine 1966), disturbance (Dayton 1971), environmental stress (Menge and Sutherland 1976), recruitment (Gaines and Roughgarden 1985), oceanographic processes affecting marine benthos (Dayton and Tegner 1984, Menge 1992, Witman et al. 1993), facilitation (Bertness and Callaway 1994, Bruno et al. 2003), chemical ecology (Hay 1996), macroecology (Witman and Roy 2009), human exploitation (Estes and Palisano 1974, Witman and Sebens 1992, Jackson et al. 2001, Coverdale et al. 2013), biodiversity and ecosystem functioning (Emmerson et al. 2001, Duffy et al. 2002, Stachowicz et al. 2002), and trait-mediated indirect interactions (Trussell et al. 2003) all have come to the attention of marine ecologists (see Bertness et al. [2014] for a review of the field).

We believe that is also likely that Hurlbert’s (1984) influential critique of experimental design contributed to the increase in the local spatial extent of experiments over time by pointing out that experimental units need to be spatially interspersed within a study area to avoid pseudoreplication. It is also possible that Underwood’s (1997, 1999) advocacy of procedural controls contributed to the rise in the number of factors and treatment combinations over time. These more intricate manipulations are often difficult to carry out in large plot sizes, leading to the links between year, grain size, and complexity.

Only one trade-off between scale and complexity was identified. The grain size of experimental units decreased as the number of species potentially affecting the response variable and number of factors and treatment combinations of the experiments increased from the 1970s to the 2010s group (Fig. 5; Appendix B). This suggests that ecologists are deploying smaller experimental units while incorporating a larger number of different treatments (factors) and simultaneously testing the responses of more species, leading to an increase in complexity. As researchers test more interacting variables and incorporate a greater proportion of the local species pool, a trade-off may arise between the size of the unit of replication and the need to include sufficient replicates for statistical robustness, given a finite resource base and logistical constraints. Frankly, we expected to identify more trade-offs between scale and complexity. One possible interpretation of the general increase in the spatial scale and complexity of field experiments over time (Figs. 1–3) is that when large-scale research is conducted in marine community ecology, effort is made to maintain complex experiments (e.g., many factors and species) by collaborative research.
groups such as the Partnership for the Interdisciplinary Study of Coastal Oceans (PISCO) on the U.S. west coast. However, many large-scale collaborative projects also utilize relatively simple experiments, allowing the large-scale biogeographic variation of different experiment sites to create variation in predictors (e.g., the Zostera Experimental Network; Reynolds et al. 2014). Both strategies produce results that far exceed previous limitations of experimental complexity (Borer et al. 2013).

There were striking habitat-specific differences in parameters of experiment scale and complexity (Fig. 7; Appendix H). Spatial scale was highest in coral reef studies, while the trends in rocky intertidal and subtidal habitats depended on the journal of publication. We suggest that this is due to experimentalists spreading out their treatments on a scale matching the spatial extent of their study habitat. Hard bottom habitats on coral and rocky reefs cover large areas (Fig. 8) that are typically larger than rocky intertidal zones constrained by upper

Fig. 7. Bar graph comparing seven metrics of scale and complexity (mean ± SE) across four habitats: coral reefs (CR), mangrove root (MR), rocky intertidal (RI), and rocky subtidal (RS), and by journals (ESA represents Ecology and Ecological Monographs; JEMBE is Journal of Experimental Marine Biology and Ecology). The effect of different habitats and journals was analyzed by two-way factorial generalized linear models (Appendix H). The main habitat differences were that in ESA publications, RI experiments were performed over significantly smaller local spatial extents (m) than in either CR or RS habitats and across both journals subtidal experiments (CR, RS) involved more species and were carried out on the largest global spatial extents. RI experiments were the longest. Experiments published in ESA journals were performed over global spatial scales that were an order of magnitude greater than those published in JEMBE, and they incorporated units of replication that were three times larger.
and lower tidal heights (Witman and Dayton 2001), so there is more space to place experimental units in subtidal than in intertidal habitats. Within ESA studies, global spatial extent of experimental manipulations was significantly larger in both coral reefs and rocky subtidal habitats than in the rocky intertidal. Given the rise of experimental macroecology in rocky intertidal ecosystems (Fig. 9; Menge et al. 2004, Navarrete et al. 2005, Wieters 2005, Bulleri et al. 2012, Tamburello et al. 2013), it was surprising that the global spatial extent of experiments published in ESA journals in coral reefs and the rocky subtidal exceeded that of rocky intertidal habitats. The fact that experiments performed in intertidal habitats ran twice as long as those in subtidal habitats likely reflects the greater accessibility of the intertidal zone to researchers. Likewise, the number of species interacting with the response variable was 1.5–2.0 times higher in CR and RS than in RI studies, which may be related to high species richness of subtidal habitats. To summarize the journal-based differences; experiments published in ESA journals were performed over global spatial scales that were an order of magnitude greater than those published in JEMBE, and they incorporated units of replication three times larger. The larger spatial scales of experiments published in ESA journals may be related to inclusion of Ecological Monographs in the analysis, which has the purview to publish comprehensive, integrative studies that demand greater length than those typically published in other journals.

The expansion of experimental scale emerging from our analysis is tempered by the fact that only 40% of the most recent (2011–2014) manipulative experiments were replicated at more than one site (Table 2). The importance of site replication seems obvious. The ecological contributions made possible when experiments are replicated at different sites include testing the generality and context dependency of ecological paradigms (Menge et al. 1994, Estes and Duggins 1995, Graham and Dayton 2002, Watson and Estes 2011), elucidating shifts in the range of native and invasive species (Sax et al. 2005), metacommunity dynamics (Leibold et al. 2004), and the scale of anthropogenic impacts (Jackson et al. 2001). Consequently, a major recommendation to enhance the understanding of ecological complexity is simply to replicate field experiments at multiple sites.

Future prospects to incorporate more of the complexity of nature into marine community ecology

As marine ecology continues to mature as a discipline, can we advance by incorporating more complexity (i.e., number of species, species interactions, factors, treatment combinations) into our experiments while not creating unsustainably complicated research programs? Below, we review several techniques that have been incorporated by investigators in the past to fold more complexity into their experimental designs. These techniques have been used to great effect, and all afford more ecological realism. They are not overly complicated to implement, but do require more care and thoughtfulness on the part of the investigator in order to derive a clean signal of ecological processes. We then
discuss novel ways to expand the scale and complexity of future experiments.

Experimental designs

Open experiments.—Open cage designs allowing the free movement of organisms through an experimental replicate provide a simple point of entry for adding experimental realism. It is well known that enclosing consumers in cages can result in unnaturally high or low rates of predation or herbivory (Hall et al. 1990) and altered consumer behavior (Englund 1997). Caging experiments also eliminate extrinsic influences, such as any direct or indirect effects of multiple consumer species, unless they are explicitly built into the experimental design. These extrinsic influences often represent desirable information about the natural complexity of the study system that may be informative.

Open experimental designs represent one option of alleviating this problem. Englund (1997) defined open experiments as experiments with treatments where prey can move in and out of experimental units. A control plot is open to these effects but nothing is technically manipulated, so they are not the focus of attention here. We extend the definition to experiments where consumers (predators, omnivores, herbivores) and/or prey can move in and out of treatments, such as the use of fences to allow predatory or herbivorous fish access to the substrate (Fig. 10; Hereu et al. 2008) or where diffusion of chemicals lowers the abundance of herbivores, such as amphipods (Whalen et al. 2012). In open experiments with mobile prey, the effects of prey mobility decrease with size of the experimental units, yet the effects of mortality from predation do not change with size, and consequently, the results of experiments with large experimental units are most likely to detect predator effects (Englund 1997, Pérez-Matus and Shima 2010).

Experiments where predators are removed or added to unenclosed areas of intertidal or subtidal habitats are open experiments (Paine 1966, Kitching and Ebling 1967, Paine and Vadas 1969, Menge 1972, Duggins 1983). For example, when the urchin Paracentrotus was removed from a 20 m long stretch of subtidal shore (Kitching and Ebling 1967) or Pisaster was removed from the lower edge of the Mytilus californianus zone (Paine 1966), other species were free to migrate into the experimental unit and affect the response variable. Such whole reef removals or additions of consumers have been valuable for identifying strongly interacting species (Paine 1980, 1984), and indeed, the keystone species concept emerged from this type of experiment. They are also significant because many of them were coarse-grained (Fig. 1B), that is, the experimental units were large (tens to hundreds of square meters) so that they could be expected to approximate the scale of the process under investigation, such as predation or herbivory from fronts of consumers (Witman et al. 2003, Silliman et al. 2005, 2013). In one sense, many open experimental designs represent a missed opportunity to gain further information about the complexity of ecological interactions unless they are monitored by technology to capture effects of the focal species and of interactions with other species (i.e., direct and behavioral effects of multiple consumers) when observers cannot be physically present in the field.

Although originally used to exclude invertebrate consumers (Castenholz 1961, Connell 1961b, Haven 1973), fenced treatments represent an effective way to increase complexity in an open experimental design. Two principal ways that fences can increase complexity is by enabling differential or simultaneous access to consumer groups. For example, (Dayton 1971) used dog dish fences to exclude limpets in and out of Pisaster removal areas, while fenced treatments in subtidal experiments enabled predatory or herbivorous fish to feed on epibionta of natural (Aylling 1981) or artificial substrata (Karlson 1978). Fences were also employed early on in tropical rocky intertidal habitats where consumer pressure is high (Menge et al. 1986). Simultaneous access to more than one type of consumer increases the number of species potentially affecting the response variable in a treatment, but to avoid confounding by multiple factors, it is only effective if the groups represent predator and prey, or if their individual effects on a common resource can be discerned either from direct observation or by species-specific predation or grazing damage. Common examples of fenced treatments where multiple consumers have simultaneous access involve fenced sea urchins or limpets with predatory or grazing fish entering the treatments from above (Fig. 10; Ayling 1981, Hereu et al. 2008).

Corridors allowing prey to move across ecosystem boundaries represent a classic type of open experiment (Englund 1977). They offer an effective way to quantify the importance of a major type of complexity, cross-ecosystem subsidies to food webs (Polis et al. 2004). For example, Ellis et al. (2007) erected large (5 × 6 m) roofs of construction mesh over the subtidal-intertidal ecotone one meter above the rock substrate in order to prevent gulls from preying on rock crabs migrating between ecosystems with the tide, to test the hypothesis that gull predation drove a trophic cascade on New England rocky shores (see Plate 1). Open experimental designs are commonly employed in hard substrate manipulative experiments, but they are rarely monitored by technology to provide new detailed information.

Last, we have begun to see the increased use of deterrents for some species. Simple applications such as chemical (i.e., for amphipods, Whalen et al. [2012], Reynolds et al. [2014]), visual (i.e., for birds, Morris and Keough [2003]), and tactile (i.e., for gulls, Ellis et al. [2007]) deterrents provide a unique method to exclude key taxa while maintaining the full complexity of the ecosystem. Provided these methods have minimal additional artifacts, they represent a powerful tool for ecologists. In particular, simple deterrents allow for ecologists to scale up the global extent of their experi-
ments with much lower effort, enabling global experiments with extreme potential for generality, such as those conducted by the Zostera Experimental Network.

Monitoring open experiments by technology.—In current manipulations of consumers and prey, unless we watch each individual predation event, we have no idea who is eating whom in a way that affects the outcome of our experiments. They are a black box, and our incomplete knowledge of food web structure can hamper our ability to draw generalizations. Paine (1980) pointed out the difference between connectedness webs, which simply show who eats whom, and functional food webs depicting strong controlling interactions elucidated by manipulative experiments. In emphasizing the need for more complexity in ecological research, we are not simply advocating for the construction of more complicated connectedness webs, which contain many weak interactions (Menge 1995). Rather, we suggest that greater understanding of the consequences of complexity could be achieved if we were more vigilant for missed trophic connections in our experimental black boxes.

Time-lapse imaging of treatments in open experiments provides a valuable opportunity to obtain information about direct and indirect effects of consumers and on temporal variation in processes driving the patterns observed. Underwater cameras and video systems have been used increasingly in mensurative experiments to study rates of grazing and predation (Steneck 1983, Witman and Sebens 1992, Hoey and Bellwood 2009, Rasher et al. 2013), but time-lapse imaging has been used much less frequently to document changes in open treatments of experimental manipulations. One of the early applications in RS habitats was Witman’s (1985) use of a time-lapse camera to test the hypothesis that subtidal mussel bed treatments provided a refuge from predation for invertebrates associated with the beds. More recently, we have used time-lapse photography and videography from compact sports cameras (GoPro; GoPro, San Mateo, California, USA) to test for trophic cascades and the effects of consumer diversity on interaction strength in Galápagos RS habitats, where upper trophic level predators are abundant (J. Witman and F. Smith, unpublished data; Fig. 11). The trophic cascade experimental designs were open, as fences were used to contain sea urchins on substrates covered with edible benthic algae, while the urchins were exposed to predatory fish from the overlying water column. The experiments were photographed throughout the day for 7–8 days, enabling predation by triggerfish and trait-mediated indirect interactions to be discovered (J. Witman and F. Smith, unpublished data). The complexity of this open experiment was high, as eight species of consumers potentially affected the response variables, since six species were observed affecting the two focal species of triggerfish (Fig. 11).
**Biodiversity ecosystem functioning experimental designs.**—Experimental designs in the field of biodiversity–ecosystem function (BEF) have long grappled with the problem of involving more complexity (i.e., greater numbers of species) in manipulations. The BEF field has begun to develop a wide set of tools for generating experimental designs that can yield a signal from a complicated system. It is time to move these designs outside of just the BEF field and find where they can be applied to other problems of ecological complexity. The field has primarily concentrated on manipulating the number of species to evaluate the relative contribution of different types of diversity effects (Loreau and Hector 2001, Duffy et al. 2003, Stachowicz et al. 2007, Fox and Harpole 2008, Kirwan et al. 2009, O’Connor and Byrnes 2014). As the field has progressed, it has furthered the exploration of the utility, benefits and limitations of additive, replacement, response surface and simplex designs (O’Connor and Byrnes 2013). While these designs often already existed for other fields of inquiry (e.g., competition), the BEF field has explored how they can be used to tease apart the influence of multiple interacting drivers within a system. The lessons learned by the discipline may well be applied to other areas, such as multiple stressor impacts. Further, the field has begun to examine how we can evaluate simultaneous changes in multiple response variables (ecosystem functions) in a way that collapses information into a biologically meaningful fashion rather than being guided solely by statistical concerns (Byrnes et al. 2014). This approach has widespread utility outside of BEF, increasing not just the complexity of experimental designs, but providing a way of acquiring and understanding increasingly complicated responses. While this traditionally has been dealt with by nonparametric clustering techniques, the rise of tools such as the many generalized linear model approaches (Warton et al. 2011, Wang et al. 2012) provide even greater flexibility to create parametric predictive models.

**Other designs: beyond ANOVA.**—We are nearly a decade away from Cottingham et al.’s (2005) exhortation to move beyond the ANOVA and begin to use regression-based designs to tease apart nonlinearities, tipping points, and complex interactions between variables. Climate change, land conversion, and overfishing are all not discrete processes. In a rejoinder to critiques of her proposal (Steury et al. 2005), Cottingham notes significant “historical biases in the field of ecology” that have slowed the recognition that regression is as much a hypothesis-testing framework as ANOVA. This idea is echoed by other critiques of her call to ecologists as actually not novel at all, as similar ideas can be found in many statistics textbooks (Steury et al. 2005). Still, many ecologists have yet to embrace continuous regression-based designs as a common practice in field experiments, and the idea remains controversial.

Further, statistics is an evolving science, and advances in model structure, how we deal with issues of non-independence, and the rich number of techniques for dealing with nonnormal, and nonlinear hierarchical data (Bolker et al. 2009) provide a vast new toolbox. These
tools provide opportunities to start with models of how ecological interactions occur and then shape our designs to test our hypotheses about these models. This approach provides many advantages, not the least of which is the possibility for a better dialogue between empiricists and theoreticians. They do require some more statistical savvy on the part of ecologists (Ellison and Dennis 2009), but the potential gains are enormous. However, we recognize that complex analyses are often not necessary, and many is the case where simple analytic strategies to straightforward questions can produce the most powerful results (Murtaugh 2007).

Experiments designed with continuous variables in mind have the potential to gain fundamentally new insights about the complex function of ecosystems. In order to tease apart how communities and ecosystems work in the face of ecological complexity, the design must fit the question. In our data set, one of the factors leading to the growth in number of treatment levels was the use of regression-based designs (e.g., Byrnes et al. 2013). These experiments, despite manipulating comparable numbers of factors, in some cases had far more treatment levels. Would the inferences the investigators who ran these experiments drew be the same or as nuanced if they had used a more classical ANOVA? We suggest that the increased use of regression or other continuous treatment-level designs might provide for fascinating growth in the types of questions asked in the future of marine community ecology.

Furthermore, if our aim as ecologists is to tease apart complicated webs of ecological drivers and responses, careful modeling of this observational data is essential. Complicated webs of interactions can now be modeled using the straightforward tools of structural equation modeling (Bollen 1989, Grace et al. 2010) that can accommodate a wide variety of nonlinear forms and nonnormal error structures (Shipley 2009). These techniques allow tests of causal models of how systems

**Fig. 11.** Species co-occurrences during a trophic cascade experiment in the Galápagos, Ecuador rocky subtidal zone using an open experimental design where consumer species interactions and co-occurrences were recorded by time-lapse photography at 1-s intervals in the rocky subtidal at 10 m depth in July 2012 (J. Witman and F. Smith, unpublished data). Data plotted are from one day (12 daylight h) of an 8-day experiment. Urchins (*L. semituberculatus*) were contained in fences on substrates of consumable benthic algae, which were open to known sea urchin predators: hogfish (*B. diplotaenia*) and triggerfish (*Balistes polylepis, Pseudobalistes naufragium*). The data plotted represents the first appearance (occurrence) of eight different interacting consumer species; thus the experiment was characterized by high complexity in terms of the number of species potentially affecting the response variable, the number of sea urchins eaten. This level of natural complexity with highly mobile consumers entering and leaving the experimental arena would be impossible to replicate in a consumer enclosure treatment. Species images and labels along the x-axis are A, adult hogfish, *B. diplotaenia*; BTR, black triggerfish, *Melichthys niger*; FT, finescale triggerfish, *Balistes polylepis*; SE, sea lion, *Zalophus wollebaeki*; TR, unknown triggerfish species; TU, green turtle, *Chelonia agassizii*; WT, whitetip shark, *Trianedon obesus*; YBT, yellow bellied triggerfish, *Sufflamen verres*. N = 41 239 images analyzed.
work and can provide a jumping off point for exploring the implications of causal models (Pearl 2009). These techniques can provide validation of experimental results in a broader context if used after experiments are conducted. If models are fit on observations before experiments, then they may suggest several testable hypotheses for dissection by experiments.

**Pseudo experimental approaches to study effects of human removal of predators.**—Human interventions in ecosystems provide an opportunity for prepared investigators to examine a response to ecosystem change while all other factors of natural complexity are still at play. They require careful selection of control sites for comparison, but once these and other BACI issues are dealt with (Underwood 1991), they can provide an important opportunity to understand how the world works. In this era of extreme weather events, human-driven species invasions, and environmental habitat degradation, we suggest that they provide a way to learn about systems under stress so that we can understand how to plan for the future.

Community–ecosystem impacts of predator release from overfishing and hunting of apex and mesopredators are large and pervasive (Jackson et al. 2001, Baum and Worm 2009, Estes et al. 2011), presenting one of the greatest imperatives for marine conservation. Bascompte et al. (2005) pointed out that exploitation removes strongly interacting species (sensu Paine 1980), as well as entire trophic levels. As mentioned in the previous section, typical manipulative experiments to test for trophic cascades involving predator enclosures are logistically challenging, if not infeasible, due to the large spatiotemporal scales of foraging by top predators. In fact, our review shows that the average grain size of experimental units has decreased over past decades as large predators have been overexploited, suggesting that the spatial scaling of contemporary experimental designs is trending in the wrong direction for evaluating the impacts of overexploitation, even if an experimental approach was advocated.

Several authors have suggested employing a “pseudo experimental” research strategy (Kerr et al. 2007) involving areas where predators have been removed by fishing as predator removal treatments, and places where they have been protected from fishing or hunting as controls (Baum and Worm 2009). While this approach can never prove causation as in a controlled manipulation, it can provide a correlative test of the underlying hypothesis at spatiotemporal scales matching the scale of human exploitation of marine consumers (Baum and Worm 2009). Perhaps the most well-known example of using spatial gradients in human exploitation of top consumers as a pseudo-experiment is Estes and Palmisano’s (1974) description of a sea otter–sea urchin–kelp trophic cascade where benthic communities in otter dominated areas were compared to those historically depleted of otters by hunting. Diamond and Case (1986) referred to this type of comparison as a natural experiment. The application of a simple mensurative experiment, such as quantifying rates of predation on tethered prey across large spatial gradients of historical consumer removal, greatly improves inferences based solely on correlations because it provides some process-based information to test hypotheses (Witman and Sebens 1992). Myers et al. (2007) went one step further and used historical fisheries data on shark declines to motivate a manipulative experiment testing the community consequences of predator release, effectively linking large spatiotemporal scales to trophic complexity.

**Embedding experimental manipulations in a background of long-term monitoring.**—While strictly observational studies have long earned a bad name for being correlative, observational studies designed with a causal model in mind can provide a wealth of information about natural processes. Observational studies can explore a far wider range of variation in ecological drivers under natural conditions than can typically be tested in a field or lab experiment. While there is always the old quote, “Essentially, all models are wrong, but some are useful,” (Box and Draper 1987), an essential corollary is “All experiments are right. Some are useful.” (J. Stachowicz, personal communication). Match or mismatch between observations and experiments under key information about the relative importance and functional form of relationships evaluated in the lab.

The Myers et al. (2007) study is an example of a general approach that we advocate for increasing the understanding of complexity on large scales. It is the practice of conducting manipulative experiments in a system that is also monitored for long time periods. This approach has the potential to yield insight about mechanism, scale, and complexity because it can capture the effects of episodic events that ripple through ecosystems, such as storm disturbances (Woodley et al. 1981), El Niño Southern Oscillations and coral bleaching (Glynn 1988, Bruno et al. 2001), massive prey recruitment events (Witman et al. 2003), pathogens (Scheibling et al. 1999), or gradual changes associated with climate, such as temperature and ocean acidification, all of which may create legacy effects. If mechanistic insight is available from experimental manipulations performed in the same system for which long term data are available, then the consequences of altered species abundances, distributions, and diversities for the resilience of the system may be predicted. Although the early focus was on monitoring, ecologists working at sites in the United States Long Term Ecological Network (LTER) are using this combined approach by conducting experimental manipulations (Byrnes et al. 2011, Peters and Yao 2012).

This approach of combining experimental manipulations with data on long-term and episodic changes in the ecosystem is superficially similar to the general observational-experimental approach advocated by Menge (2004), but differs in that the observations are
composed of biological and environmental data collected over long periods of time rather than observations of patterns of community structure used to motivate experiments. There are a number of examples of this type of approach from RI and RS habitats. In the subtidal, Dayton et al.'s (1984) study contributed to the understanding of oceanographic climate shifts on kelp forests because kelp removal experiments testing hypotheses about competition and patch dynamics were performed over a 10-yr period, which was long enough to encompass several monitored changes in the upwelling regime. Ling et al. (2009) combined long-term data on urchin range shifts, trends in urchin predators (lobsters), and temperature with mensurative urchin-tethering experiments to demonstrate a trophic cascade driven by a climate associated range shift. In the Galápagos rocky subtidal, we are combining results from our manipulative experiments (Irving and Witman 2009, Witman et al. 2010, Brandt et al. 2012, Altieri and Witman 2014) with results from ongoing, biannual monitoring at 12 sites over 13–16 yrs to understand the broader negative and positive impacts of ENSO (Witman et al., unpublished manuscript).

Some of the most well-known applications of this observational-experimental approach come from the RI zone where Menge et al. (2011) conducted recruitment limitation manipulations in intertidal communities across a latitudinal gradient on the U.S. west coast to understand the consequences of climate-induced variation in upwelling, where they have been monitoring relevant climate drivers for over a decade in the PISCO program. By combining manipulative experiments with long-term monitoring of recruitment and oceanographic conditions at 18 sites spanning 900 km of coast, Navarrete et al. (2005) documented geographically discontinuous oceanographic regimes that predictably modify the outcome of species interactions and the regulation of benthic communities. Ecologists in the EuroRock program performed limpet removals from Norway to Portugal to examine climate-related differences in grazing as they monitored latitudinal gradients in temperature (Bulleri et al. 2012).

With few exceptions, the scale of manipulative experiments in marine ecology will always be orders of magnitude smaller than the spatiotemporal scales of climate change and other anthropogenic disturbances, which operate on the largest regional/global scales. Even examples from the most recent literature on marine ecological experiments (2011–2014) have covered relatively small spatial scales, ranging from lows in the rocky subtidal to peak values in coral reef habitats, with average local spatial extents from 112 to 183 m. The size of the unit of replication within these experiments (spatial grain) is also lower now than it has been in any of the previous decades, with average grain sizes ranging from 0.34 m² in the rocky intertidal to 2.9 m² in mangrove root habitats.

Can we use the results from small-scale experiments to learn about impacts of larger scale processes? This central question preoccupies contemporary ecology; indeed, biological scaling is a subdiscipline that has been extensively reviewed (Schneider 1994, Gardner et al. 2001). While there is no single solution, several fruitful avenues involve building quantitative information about spatial structures of the focal species or community from previous surveys into the experimental design (Legendre 1993, Dray et al. 2012) and applying scale transition theory (Benedetti-Cecchi et al. 2012). Schneider (in Schneider et al. 1997 and Schneider 2001) proposed a mathematical and graphical method of using spatial allometry to examine the relationship between the scope of the experiment and the scope of the process being tested. Plotting the spatial scope (the ratio of spatial range, i.e., extent, and the resolution, i.e., grain) and the temporal scope (ratio of the experiment duration and minimum time between measurements) of both the process under investigation and the experiment is useful for identifying the limitations of inference.
Remote sensing of change in experiments, communities, and ecosystems.—Placing even simple experiments into a larger experimental context is key. But how can we acquire that data when our boots can only tread across so much ground? The use of unmanned aerial vehicles (UAVs) and autonomous underwater vehicles (AUVs) now provides a novel opportunity to integrate scale and complexity in marine ecosystems. Conducting experiments, simple experiments, at large spatial scales and using remote sensing data from UAVs and AUVs to create a rich quilt of biotic and abiotic covariates provides a tremendous opportunity to discover more about the shape and context dependence of relationships within our experimental arenas. In addition, UAVs will likely revolutionize spatial ecology due to their ability to map large areas of habitats from above (Anderson and Gaston 2013). Similarly, AUVs have been used to map the spatial distribution of coral reefs (Patterson and Relles 2008).

Many marine consumers (whales, seabirds, sea lions, seals, walruses, manatees, some sharks) and foundation species (kelp, corals, oysters, mangroves) can be remotely sensed on large space and time scales either when they are exposed at the surface, submerged in clear water, or exposed on intertidal shores (Cavanaugh et al. 2010, Turner 2014). For example, Fretwell et al. (2012) used remote sensing to monitor emperor penguins and their production of guano in Antarctica, which could serve as a basis for evaluating bottom-up (productivity) and cross-ecosystem effects on large scales. In addition, dynamic changes in resources such as algal productivity can be measured by remote sensing. Sensing ocean color from space is commonly used to investigate of the role of oceanographic processes and bottom-up effects in marine communities (Dayton and Tegner 1984, Menge et al. 1997, Navarrete et al. 2005, Wieters 2005, Witman et al. 2008). However, remotely sensed data on the abundance of top consumers and their prey would shed new light on the potential for top-down control over very large areas. As the spatial resolution of the airborne sensors increases, more community variables and even response variables in intertidal experiments may be measured over large areas. It is now feasible to conduct species removal experiments over large areas, such as the whole reef removals of the 1970s, and monitor them with drones (UAVs) for the effects of additional factors or transitory species interactions that may have a key influence on community structure. Although current models of UAV sensors are limited in terms of their ability to penetrate deep into the ocean, new developments in bathymetric LIDAR and depth-penetrating lasers hold promise for future investigation of subtidal communities (Turner 2014). We foresee a significant contribution from remote sensing in the near future in terms of our ability to understand more of the complexity of marine communities across large scales by frequent monitoring of field experiments, particularly in rocky intertidal habitats. The precedent already exists in terrestrial ecosystems where competition for space among trees has been measured (Kellner and Asner 2014) and a manipulative experiment (ungulate removal) was monitored on large spatiotemporal scales (9 years, 4.3 km²) by UAVs (Kellner et al. 2011).

Conclusion

We have learned much over the past several decades of experimental marine ecology. As we mature as a science, we necessarily seek to ask questions that explain progressively more about how the natural world works. As we move beyond initial heuristic depictions of nature, we must embrace both simple and more ambitious complex experimental designs equally. We have already begun to rise to the challenge. We suggest that we have further still to go and need a robust dialogue about the best techniques to develop a better understanding of communities and ecosystems in a changing world.

Acknowledgments

We thank Aaron Ellison for the opportunity to write this review and for his insight about experimental design. This research was supported by a grant from the U.S. National Science Foundation (Biological Oceanography Program, grant number OCE 1061475) to J. Witman and Rufford Small Grant for Conservation to R. Lamb. J. Byrnes was supported by SeaGrant 2014 R/RCI4-36. The manuscript was improved by comments from A. Ellison, J. Kellner, P. Petraitis, and an anonymous reviewer. Thanks to F. Smith, S. Castello y Tickell, and N. Low for help in the field, and E. Moore, J. Ang, and S. Castello y Tickell for helping with the analysis of time-lapse images. E. Gross and E. Moore assisted with aspects of the literature review. A. Adams illustrated Fig. 10. Discussions with F. Smith, K. Roy, R. Etter, M. Novak, S. Navarrete, and E. Wieters helped shape our thinking about scale and complexity issues. Our field work in the Galápagos Islands was facilitated by the Galápagos National Park and the Charles Darwin Foundation. To all we are grateful.
LITERATURE CITED


November 2015 499


SUPPLEMENTAL MATERIAL

Ecological Archives

Appendices A–I and the Supplement are available online: http://dx.doi.org/10.1890/14-2265.1.sm