Ecological Indicators 125 (2021) 107522

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Marine ecosystem indicators are sensitive to ecosystem boundaries and spatial scale

Kurt C. Heim [a,*], Lesley H. Thorne [a], Joseph D. Warren [a], Jason S. Link [b], Janet A. Nye [a]

[a] School of Marine and Atmospheric Science, Stony Brook University, Stony Brook, NY, USA
[b] National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Office of the Assistant Administrator, Woods Hole, MA, USA

ARTICLE INFO

Keywords:
Ecosystem indicators
Problem of scale
Spatial scale
Ecosystem based management
Dynamic Ocean management
Marine ecosystems

ABSTRACT

Time series indicators are widely used in ecosystem-based management. A suite of indicators is typically calculated for a static region or multiple subregions and presented in an ecosystem assessment (EA). These are used to guide management decisions or determine environmental status. Yet, few studies have examined how the spatial scale of an EA influences indicator behavior. We explore this question using the Northwest Atlantic continental shelf ecosystem (USA). We systematically divided the ecosystem at six spatial scales (31 unique units), covering spatial extents from 250,000 km^2 to 20,000 km^2. The same 22 indicators were calculated for each unit, assessed for trends, and evaluated as 31 independent EAs. We found that the detected signals of indicator trends depended on the spatial scale at which the ecosystem was defined. A single EA for the whole region differed by 23% (in terms of the 22 indicator trend tests) relative to ones for spatially nested 120,000 km^2 subunits, and by up to 36% for EAs at smaller scales. Indicator trend disagreement occurred because (most common) a localized trend was perceived as widespread, (common) a local trend was obscured by aggregating data over a large region, or (least common) a local trend switched direction when examined at a broader scale. Yet, there was variation among indicators in their scale sensitivity related to trophic level. Indicators of temperature, chlorophyll-a, and zooplankton were spatially coherent: trends portrayed were similar regardless of scale. Mid-trophic level indicators (fish and invertebrates) showed more spatial variation in trends. We also compared trend magnitude and indicator values to spatial extent and found relationships consistent with scaling theory. Indicators at broad scales produced subdued trends and values relative to indicators developed at smaller spatial scales, which often portrayed 'hotspots' of local abundance or strong trend. Our results imply that subsequent uses of indicators (e.g., determining environmental status, risk assessments, management decisions) are also sensitive to ecosystem delineation and scale. We suggest that indicators and EAs should be done at multiple spatial scales and complimented with spatially explicit analysis to reflect the hierarchical structure of ecosystems. One scale is not best, but rather we gain a new level of understanding at each scale examined that can contribute to management decisions in a multiscale governance framework characterized by goals and objectives with relevance at different scales.

1. Introduction

Integrating information across disciplines in marine science is often facilitated by indicators, which we define here as quantitative measurements that represent key attributes of interest. When a vetted suite of indicators is assembled, it can be used to assess ecosystem status, drivers of change, and performance of management actions (Rice and Rochet, 2005; Shin and Shannon, 2010; Shin et al., 2010). Indicators have come to play a central role in implementing Ecosystem Based Management (EBM) and an Ecosystem Approach to Fisheries Management (EAFM); they are widely considered an important bridge between science and policy (Turnhout et al., 2007). Some governments require or strongly request indicators to support ecosystem and environmental impact assessments through formal policy (e.g., determining Good Environmental Status in the European Union under the Marine Strategy Framework Directive, Australia's Environment Protection and Biodiversity Conservation Act, US National Environmental Policy Act), while others have turned to indicator suites to fulfill or guide a variety of interconnected regulatory mandates (Levin et al., 2013; Foran et al., 2016; Link et al., 2019). For example, in the US, indicators are a key aspect of regional integrated ecosystem assessments (IEAs) (Levin et al., 2013; Harvey et al., 2017)

* Corresponding author. Present address: U.S. Fish and Wildlife Service, Western New England Complex, 11 Lincoln Street, Essex Junction, Vermont 05452, USA

https://doi.org/10.1016/j.ecolind.2021.107522
Received 29 September 2020; Received in revised form 9 February 2021; Accepted 9 February 2021

Given their increasingly prominent role in ecosystem science and management of marine systems, a large body of research has focused on how to select the best indicators (Rice and Rochet, 2005; Queirós et al., 2016; Tam et al., 2017; Otto et al., 2018; Bundy et al., 2019). Other studies provide advice on how to use indicators in the context of marine ecosystem management (Large et al., 2013; Large et al., 2015; Levin and Möllmann, 2015; Burthe et al., 2016; Bal et al., 2018; Gaichas et al., 2018). This body of work emphasizes indicator selection, aggregation, interpretation, and use for management, yet the sensitivity of commonly used indicators to spatial scale is a relatively unexplored but very important topic.

To generate *any* quantitative indicator for testing and eventual use, one must first define a spatial region (or regions) over which to collect or summarize data. Regions for which indicator time series are calculated may be the entire ecosystem or different ecosystem subunits (e.g., ecological production units in the US or assessment regions of regional seas in the European Union). In any case, scientists and managers must decide: (1) how large of a spatial extent is appropriate to define the ecosystem; (2) how and where the exact boundaries should be drawn; (3) whether ecosystem subunits should be defined; and (4) how and where subunit boundaries should occur. These decisions should ideally integrate an understanding of spatial structure in key ecosystem processes with jurisdictional boundaries and policies (Levin et al., 2013). This latter consideration is a major challenge (Crowder et al., 2006). What works politically might make little ecological sense (or vice versa). Regardless of the reasoning or methods used to draw boundaries, any boundary system is likely to have some impact on how indicators perform (Fig. 1). Moreover, since indicators are central to many next steps in EBM or EAFM (e.g., ecosystem assessment, risk assessment, management strategy evaluation; Bunnefeld et al., 2011; Gaichas et al., 2018), the sensitivity of indicators to spatial scale could be manifested across the entire management process.

Ecological theory and a broader recognition of spatial structuring in ecosystems all but guarantee some indicators are sensitive to spatial scale. The 'problem of scale' (*sensu* Levin, 1992) is that processes can show different trends or relationships depending on the focal region of a study (i.e., the spatial extent) and the spatial units used to sample or analyze data (i.e., the spatial grain). For example, Rose and Leggett (1990) demonstrate that the correlation between a predator and its prey is scale dependent; at large sampling grain predators and prey show strong positive correlation (co-occur in space, as predators follow prey), but at fine sampling grain predator avoidance is detected and correlations become negative. The main lesson is that when processes are spatially structured, choices related to extent and grain of analysis can markedly influence results (Levin, 1992). And since advances in spatial statistics, geographic information systems, and genomics continue to reveal that spatial structure is pervasive in marine ecosystems (e.g., marine heatwaves, Oliver et al., 2018; zooplankton, Morse et al., 2017; fish, Ciannelli et al., 2008; Bradbury et al., 2013; penguins, Lynch et al., 2012; humans, Colburn et al., 2016), exploring the consequences of spatial structure to indicator behavior is a pressing need.

We conceptually illustrate some potential issues related to defining boundaries as they pertain to indicators (Fig. 1). For a process operating systematically across a large spatial domain it would matter little how the boundaries of an ecosystem were defined (Fig. 1A). All treatments of any data subset would reveal the same trend. Such an indicator might be considered robust to ecosystem delineation or the spatial scale of analysis. However, if there is spatial heterogeneity in a trend (i.e., a north to south gradient, an east to west gradient, or patchy populations and metapopulation dynamics, Fig. 1B–D) then how one defines the spatial boundaries of the ecosystem will alter the perceived trend. If the four processes are considered at the same time, we find that any particular boundary system chosen might work well for some processes, but not others.

Here, we recreate the conceptual analysis portrayed in Fig. 1 with data from the Northeast United States Continental Shelf (NES) large marine ecosystem (Sherman, 1991) to explore indicator scale sensitivity.

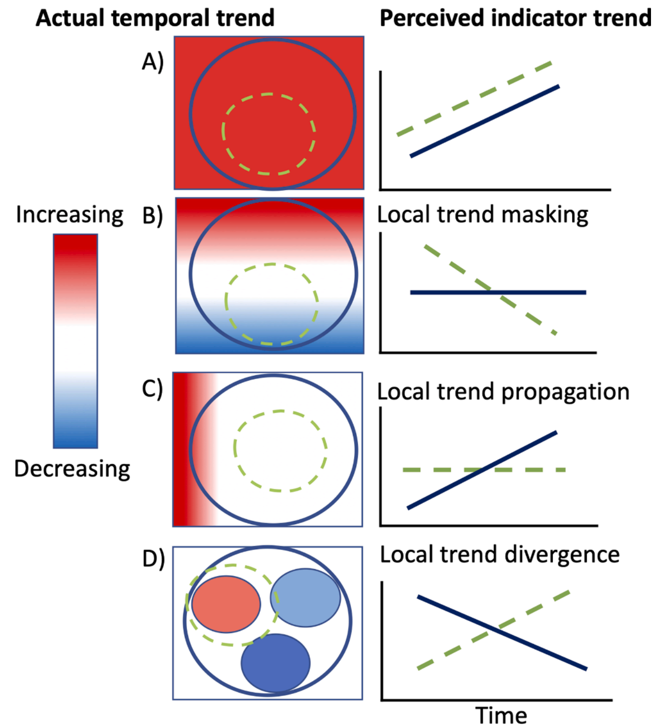

Fig. 1. Hypothetical scenarios of how spatial extent and boundaries influence indicator trends. A 'true' spatial pattern of temporal trend exists in each domain (left boxes, color gradient), but the perception of trend is scale dependent. Solid blue circles (left panels) represent a large region over which one might summarize data and generate an indicator (shown as solid blue trendline, right panels). Green hatched circle (left panels) represents a smaller region over which one might summarize data and generate an indicator (green hatched trendline, right panels). Scenarios include (A) consistent trends perceived at all spatial scales; (B) when no trend is perceived at a large extent, but strong local trend(s) exist within the spatial domain; (C) when a trend is perceived at large extent, but driven by a highly localized trend; (D) when a trend is present at large extent, but a different trend is perceived at a local scale within this region. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Specifically, our objectives were to: (i) determine how changing the spatial extent and region covered by an ecosystem assessment (EA) alters one's perception of trends; (ii) examine which indicators are most sensitive to spatial scale; and (iii) explore the relationship between grain, trend strength, and values of indicators. The NES is an ideal region for this exploration because there is abundant long-term data, and a great deal of empirical work on specific ecosystem processes (i.e., fisheries, zooplankton, oceanography) and EAs to provide context for our exploratory results (Link et al., 2002; EcoAP, 2009, 2011; NEFSC, 2017a, 2017b, 2018a, 2018b). Our goal therefore is not to elucidate novel ecological or oceanographic trends in the NES, re-define or evaluate how boundaries are currently defined in the NES, or develop new and thoroughly vetted indicators for application. Rather, our study is intended to highlight issues to be aware of when developing EAs that use indicators. The potential issues related to scale highlighted in this study will apply to any scenario where boundaries are drawn and indicators are calculated, and so should be globally applicable in both marine and terrestrial systems.

2. Methods

2.1. Study area background and existing spatial delineations

The NES is a large marine ecosystem (Sherman, 1991) that covers a spatial area of 250,000 km^2 extending from Cape Hatteras (USA, North

Carolina) in the South to Novia Scotia (Canada) in the North (34–45 degrees N, Fig. 2A). Fisheries are currently managed from a single-species perspective by two regional fisheries management councils, yet there is growing interest and commitment to implement EBFM (Levin et al., 2013). To this end, an IEA has been adopted as the cornerstone of the federal approach to integrating ecosystem information to inform management decisions (Levin et al., 2013; Harvey et al., 2017). Indicators are a foundational component of the IEA and are regularly calculated for four distinct spatial regions within the NES (Lucey and Fogarty, 2013). These include the Gulf of Maine (69,000 km^2), the Scotian Shelf (31,000 km^2), Georges Bank (58,000 km^2), and the Mid-Atlantic Bight (126,000 km^2). These are considered, at present, the appropriate spatial scale to conduct an IEA (DePiper et al., 2017). Indicators used are multidisciplinary time-series contributed by over 30 scientists each year, which are assembled, analyzed with a consistent trend fitting technique, and presented in an EA report. This provided a good model with well-developed indicators, based on extensive survey data, to explore our research questions.

2.2. General approach

We created a suite of 31 assessment region boundaries that can be thought of in a similar way as Ecological Production Units, or in the context of the MSFD, these are akin to subregions within a regional sea (Fig. 2A). We then developed the same 22 indicator time series for each of these units (31 units × 22 indicators = 682 total, Table 1). Each set of indicators is treated as an independent assessment (i.e., a mini EA). Conceptually, we follow the procedure shown in Fig. 1 but with real data, then explore the resulting time series in the context of our research questions.

2.3. Defining ecosystem subunit boundaries

We used the region covered by the NOAA bottom trawl survey offshore strata, that are regularly sampled, to define the total geographic extent of the NES for this study (Fig. 2A). This survey covers the entire NES with 350–400 trawl stations sampled annually (Azarovitz, 1981). Next the NES was split in half at the latitude which gave two polygons of exactly the same spatial area, resulting in a northern unit and a southern unit of ~ 120,000 km^2 each. Subsequent ecosystem splits were made within the northern and southern units by splitting each into 2, 3, 4, and 5 units using a more complicated procedure (Fig. 2A, supplementary materials). In short, subsequent splits were done in an automated manner that maximized evenness in sampling density of the survey data used for indicator calculations. Spatial units are an average of 245,000 (n = 1), 123,000 (n = 2), 59,000 (n = 4), 39,000 (n = 6), 29,000 (n = 8), and 23,000 (n = 10) km^2, though these numbers are rounded in figures and subsequent text for simplicity. These boundaries are naïve to the inherent biophysical structure present in the ecosystem, and existing governance structure. This was done to focus the results on broader issues related to scale sensitivity, rather than nuances of how this particular ecosystem is, or should be defined in practice.

2.4. Indicators

We used a total of 22 indicators to represent each spatial unit. Five described the physical environment, eight represented lower trophic level processes, and nine represented mid trophic level processes. We required indicators with long-time series known to be useful for describing the ecosystem, so we used indicators commonly appearing in federal EAs for the NES. All but two of our final 22 indicators (Spiny Dogfish and Silver Hake) have been used in previous federal EA reports and have been vetted and selected from much broader lists of potential indicators (Link et al., 2002; Methratta and Link, 2006; EcoAP, 2009, 2011; NEFSC, 2018a, 2018b, 2019a, 2019b). Spiny dogfish and Silver Hake are abundant and important mid trophic level taxa that also provided interesting contrasts of spatial variation in trends (Nye et al., 2011). The methods used for all zooplankton indicators and those based on the NEFSC trawl survey (Table 1) are all identical to those from NEFSC (2019a) so are not described in detail in the main text

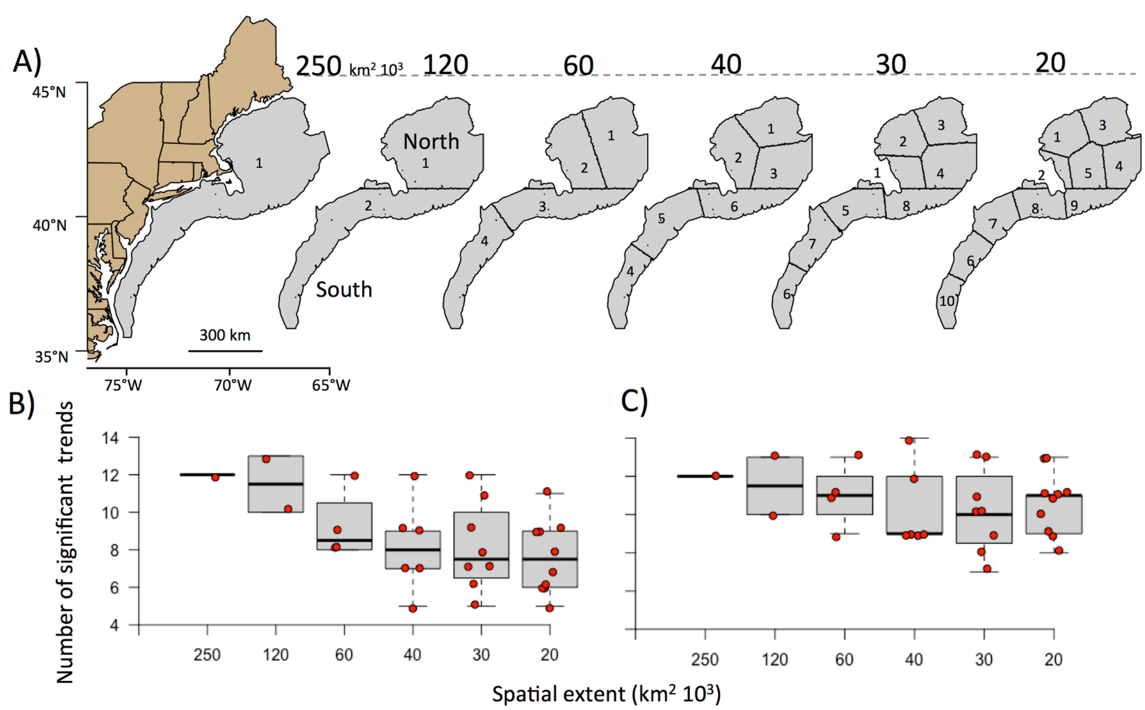

Fig. 2. The Northeast US Atlantic Continental Shelf Large Marine Ecosystem, split into 31 units (A) with the average spatial extents of ecosystem units shown (above). Panel B shows the number of significant indicator trends (out of 22 tested) within each spatial unit assessed with a Mann-Kendall test and a Bonferroni correction. Each point corresponds to a unit in panel A. The same results using a consistent alpha of 0.05 are shown in panel C.

Table 1
Indicators used in this study, all indicators run from the start year (shown) to 2018 unless otherwise noted. Abbreviations are as follows: BT, bottom temperature; lat, latitude; lon, longitude; mon, month; yr, year; WOD, world ocean database; SST, sea surface temperature; Strat, stratification defined as the difference in water density between the surface and at 50 m; yday, day of year in numeric form; OISST, optimally interpolated sea surface temperature; Chyl, chlorophyll-a; const., a small constant number; OCCI, Ocean Colour Climate Change Initiative; EcoMon, Ecosystem Monitoring survey by the North East Fisheries Science Center (NEFSC). Indicators in bold are calculated with the exact methods described in NEFSC (2019a) and data using the NEFSC trawl survey are from the spring sampling period.

Category	Indicator	Description	Data	Start
Physical	Bottom temperature[a]	log(BT + 1) ~ s(lat, lon) + s(mon) + yr	WOD	1975
	Surface temperature[a]	SST ~ s(lat, lon) + s(mon) + yr	–	1975
	Stratification[a]	log(Strat) ~ s(lat, lon) + s(mon) + yr	–	1975
	Summer SST	Mean (yday 182–273)	OISST	1981
	Marine heatwave days	Extensive analysis	–	1981
Lower TL	Chlorophyll-a[a]	log(Chyl + const.) ~ s(lat, lon) + s(mon) + yr	WOD	1980
	Chlorophyll-a (sat)	Annual mean Chyl	OCCI	1997
	Zooplankton volume	Annual volumetric anomaly	EcoMon	1977[b]
	Centropages typicus	Annual abundance anomaly	–	–
	***Pseudocalanaus* spp.**	–	–	–
	Temora longicornis	–	–	–
	Calanus finmarchicus	–	–	–
	Small/large copepod	–	–	–
Mid TL	**Trawl biomass**	Biomass tow^{-1}	NEFSC	1970
	Benthivore	Biomass tow^{-1} 42 benthivore sp.	–	–
	Benthos	Biomass tow^{-1} 9 benthos sp.	–	–
	Piscivore	Biomass tow^{-1} 63 piscivore sp.	–	–
	Planktivore	Biomass tow^{-1} 20 planktivore sp.	–	–
	Sea Scallop	Biomass tow^{-1} Sea Scallop	–	–
	American Lobster	Biomass tow^{-1} American Lobster	–	–
	Silver Hake	Biomass tow^{-1} Silver Hake	–	–
	Spiny Dogfish	Biomass tow^{-1} Spiny Dogfish	–	–

[a] The annual coefficient value for *Year* from the model was used as the indicator.
[b] end year for indicators using EcoMon data is 2017.

(supplementary materials). All indicators end at 2018 or 2017 and have variable start dates (Table 1). The minimum length for a time series was 37 years and the maximum was 48.

2.4.1. Indicators using world Ocean data

Indicators for in situ bottom temperature, surface temperature, stratification, and chlorophyll-a (in situ) were developed with data from the World Ocean Database, which includes contributions from various agencies (including all NMFS surveys) and researchers. There were 58,554 CTD casts available for temperature and stratification indicators, and 17,379 casts that contained chlorophyll-a measurements. Because these represent a broad collection of datasets (i.e., CTD casts are not random in space or time) we used generalized additive models (GAMs) to develop an annual index that accounted for spatial unevenness and the month the CTD casts were taken (Table 1, supplementary materials). The GAMs were fit using the "mgcv" package in the R statistical programming environment (Wood, 2001; R Development Core Team, 2019). We used the default thin plate regression spline option (bs = "tp") and restricted maximum likelihood to fit the models (method = "REML"). Models were first run including all data for the NES to assess fit and identify appropriate transformations before running for data subset to each spatial unit. The annual coefficient value for *Year* from the model was used as the indicator (Table 1). Similar methods are regularly used to standardize fisheries sampling data and assess chlorophyll trends (Maunder and Punt, 2004; Boyce et al., 2010).

2.4.2. Indicators using satellite data

We used the NOAA Optimum Interpolation ¼ degree Daily Sea Surface Temperature Analysis (OISST) to calculate an indicator of summer sea surface temperature (SST) and a count of days classified as Marine Heatwaves (MHWs) per year (Reynolds et al., 2007; Hobday et al., 2016). For the summer SST indicator, we averaged SST measures occurring within a spatial unit by year including only data from July to September (day-of-year 182 to 273). For the MHW indicator, we first built a single time series of average daily temperatures within a spatial unit. This was then used to detect marine heatwave events according to methods described in Hobday et al. (2016) and implemented in the R package heatwaveR (Schlegel and Smit, 2018). We defined a MHW as an anomalously warm water event (>90th percentile based on a climatology from 1981 to 2012) that lasted for more than 5 days. The indicator represents a count of all days per year, that were classified as MHW days. An indicator of chlorophyll-a was developed using satellite data from the Ocean Colour - Climate Change Initiative dataset (4.0), a recent data product that merges data from several satellite missions (Sathyendranath et al., 2019). We averaged daily chlorophyll-a estimates, occurring within each spatial unit annually to produce the indicator.

2.5. Data analysis

The resulting 682 time series (31 spatial units × 22 indicators) were assessed for a monotonic trend with a Mann-Kendall non-parametric test to classify each as increasing, decreasing, or stationary (Mann, 1945; Kendall, 1957). The Mann-Kendall test was chosen because it is commonly used in EAs to assess indicator trends (NEFSC, 2017a, 2017b; Gaichas et al., 2018). A Bonferroni correction was included to account for multiple testing within each scale by dividing alpha (0.05) by the number of tests done. For example, for indicators done at the 20,000 km^2 scale (i.e., 1/10th of the NES) we used an alpha of 0.05/10, for the 30,000 km^2 we used 0.05/8. Theil-sen slopes and intercepts were also calculated for each time series (Sen, 1968; Theil, 1992). All time series were independently scaled and centered (subtracting mean and dividing by standard deviation) prior to analysis to focus this analysis on trends, rather than absolute values.

To examine how the overall perception of trends change with the extent and region of an EA (*objective i*), we compared one simulated EA (i.e., all 22 trend results for a unit) to another EA that was nested within it. To assess how a broad-scale EA represents the same trends evaluated at a more local scale, trends were classified as significantly increasing, decreasing, or not-trending based on the Mann-Kendall test, and then assessed for consistency. For example, if trawl biomass was increasing in the 250,000 km^2 unit but not trending in a nested spatial unit, this is a classified as a disagreement. We counted the number of disagreements for each comparison and report this as a % of the total comparisons. We also tallied disagreements according to their cause (Fig. 1B–D). Masking is when a trend is *absent* at a broad extent but *present* within a nested unit (i.e., the trend occurring at a local scale is 'masked' by data aggregation); propagation is when a trend is *present* at a broad extent but *absent* within a nested unit (i.e., a trend in only a portion of subunits *'propagated'* to the broad scale analysis, thus creating inconsistencies between the broad scale trend and those nested units where trends are not occurring); divergence is when a trend is *present* at a broad extent *and* a nested unit but the direction of trends differed.

In addition to visualizations, we calculated several metrics to assess which indicator trends were most sensitive to scale (*objective ii*). The first was the variance of the 31 Theil-Sen slopes for a given indicator. Low variance implies the monotonic trendline is consistent, regardless of the spatial unit considered. High variance means the slope of the trend changes depending on the spatial region considered. We also calculated the average pairwise Pearson correlation coefficient between all 31 units, for a given indicator (*Ave* r_p). High *Ave* r_p means the indicator time series data are consistent across space (i.e., spatially coherent), and this metric makes no assumption about the presence or absence of a trend (Östman et al., 2017). Lastly, for a subset of indicators (foraging groups) we graphically depict the relationship between indicator absolute values, indicator time series trend, and spatial extent of the EA (*Objective iii*).

3. Results

3.1. The degree of inconsistencies between ecosystem assessments at different scales

The number of significant indicator trends varied within each scale considered indicating that some parts of the ecosystem were changing more so than others (Fig. 2B). Treating the ecosystem as a single unit (extent = 250,000 km^2) resulted in 12 out of 22 indicators with significant trends (55%); splitting it in two showed that the north had 13 (59%) significant trends whereas the south had only 10 (45%). Further disaggregation at progressively smaller spatial extents revealed anywhere from 5 to 12 significant trends. This means that one way of defining a region for an assessment would show only 23% of indicators are trending, yet a different way of defining it would show 59% of indicators are trending. Comparisons across scales demonstrate a pattern where more significant trends are detected at larger spatial extents (Fig. 2B). Without the Bonferroni correction, this trend was weaker, but still visually present (Fig. 2C).

Comparisons of indicator trends among spatial units showed inconsistencies of between 4% and 36% (Fig. 3; Table 2). These comparisons relate the trend of an indicator (increasing, decreasing, stationary) calculated at a broad extent to the *same indicator* calculated at a smaller and nested scale and examined whether the results match. For example, if we treat the NES as a single region and calculate indicators and then compare these to the indicators for the northern unit, 5 out of the 22 trend results were inconsistent (23%). The whole NES version at 250,000 km^2 was also 23% inconsistent with the southern unit. Comparing the 250,000 km^2 unit to the ten nested 20,000 km^2 units gave a mean inconsistency of 30% (range 23% to 36%), implying a single EA would misrepresent local trends by as much as 36%. Overall there appeared to be a trend where mean and maximum inconsistencies increased when the 250,000 km^2 was compared to units of progressively smaller spatial extent (Fig. 3A, B), but this trend flattened once a scale of 40,000 km^2 was reached.

The southern 120,000 km^2 scale EA was more consistent with nested local trends than the northern 120,000 km^2 EA, but there was great variation within these comparisons (Fig. 3). Mean inconsistencies in the north were 23% for most scales, while they were closer to 15% for most comparisons in the south. Yet, the observed variation within a given spatial scale means that the 120,000 km^2 units reflected trends occurring in some nested spatial units quite well (Fig. 3C) while they poorly matched those occurring in other nested units (Fig. 3B). Like the comparisons with the 250,000 km^2 units, mean and maximum inconsistencies appeared to increase when comparisons were made with units of smaller spatial extent (Fig. 3A, B).

3.2. Common causes for cross-scale inconsistencies

Local trend propagation (Fig. 1C) was the dominant cause for cross-scale indicator inconsistencies, accounting for between 40% and 100% of them relative to the total for each comparison (Table 2). For example, trawl biomass increased when considered at the whole ecosystem scale but only increased in one of the two nested units at the 120,000 km^2 scale (the south) (Fig. 4C). The strong trend in the southern unit thus

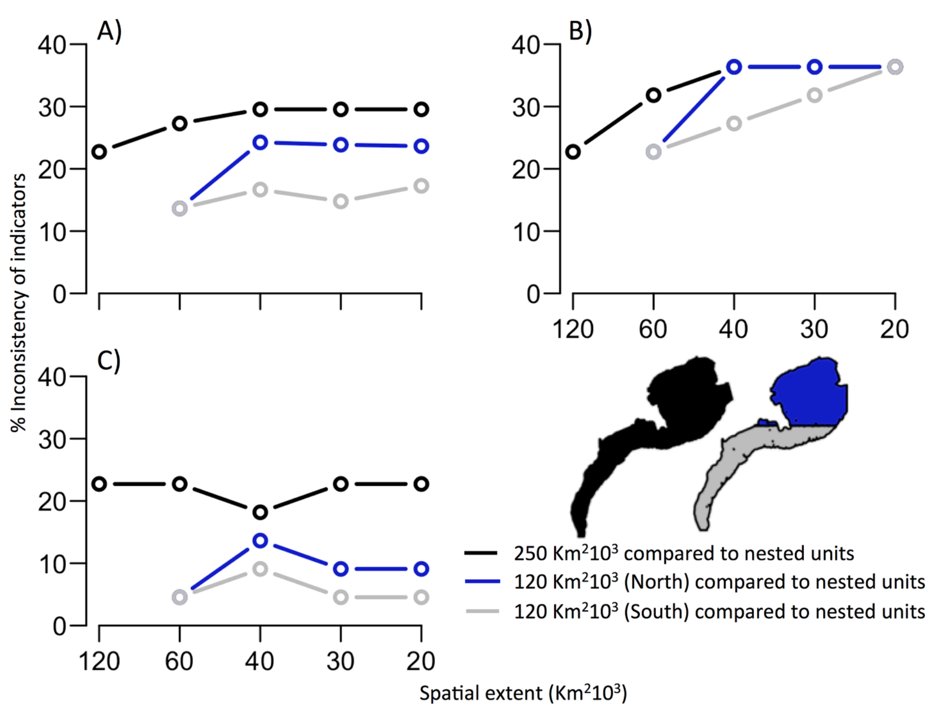

Fig. 3. Percent inconsistency of indicator trends when comparing those calculated for one spatial unit (250,000 km^2, 120,000 km^2 north, 120,000 km^2 south; colored lines) to nested spatial units in terms of the mean inconsistency (A), maximum inconsistency (B), and minimum inconsistency (C). Percent inconsistency is defined as the number of time series trends that differed, divided by the total number compared (n = 22).

Table 2
Indicator consistency and the reason for inconsistencies. The comparison column shows which units are being compared. For example, 250:120 designates that indicator trends (n = 22) for the whole ecosystem version (250,000 km^2) were compared to the same indicator trends calculated for the 120,000 km^2 units (i.e., a northern and southern unit, Fig. 2). Percent inconsistency is defined as the number of time series trends that differed, divided by the total number compared. Inconsistency (%) is followed by the minimum and maximum observed in a comparison. Total inconsistent (n) is followed by the total number of comparisons made. The proportion of total inconsistencies due to three causes as depicted in Fig. 1 is also shown.

Comparison	Inconsistency (%)	Total inconsistent (n)	Propagation (%)	Masking (%)	Divergence (%)
250:120	23 (23–23)	10/44	40	40	20
250:60	27 (23–32)	24/88	63	29	8
250:40	30 (18–36)	39/132	72	23	5
260:30	30 (23–36)	52/176	71	21	8
250:20	30 (23–36)	65/220	75	18	6
120(N):60	14 (5–23)	6/44	100	0	0
120(S):60	14 (5–23)	6/44	67	33	0
120(N):40	24 (14–36)	16/66	94	6	0
120(S):40	17 (9–27)	11/66	73	27	0
120(N):30	24 (9–36)	21/88	95	5	0
120(S):30	15 (5–32)	13/88	77	23	0
120 (N):20	24 (9–36)	26/110	100	0	0
120(S):20	17 (5–36)	19/110	79	21	0

'propagated' to the next spatial extent, creating an inconsistent result between the 250,000 km^2 unit (increasing) and the northern unit (no trend). Further disaggregation of the southern unit revealed this increasing trend was not widespread (increasing in only 2/10 units, Fig. 4C). In this case, 8 units exhibit no trend and are thus inconsistent with the increasing trend portrayed by the 250,000 km^2 unit. Other examples of local trend propagation in Fig. 4 include *Temora longicornis* abundance (increased in northern unit but not in some nested subunits) and piscivore biomass (increased in southern unit but not in some nested subunits).

Local trend masking (Fig. 1B) accounted for between 0% and 40% of inconsistencies (Table 2) and was the second most common cause of cross-scale inconsistencies. The indicator of *Temora longicornis* abundance (a copepod) presents two examples. First, there was no trend detected at whole ecosystem scale, but there was an increasing trend occurring in the spatially nested northern unit (Fig. 4B). This trend failed to propagate to the next spatial extent, and thus was masked by aggregation of data across a large region where no trend was occurring. Additionally, an increasing trend was detected at the southernmost unit in the 60,000 km^2 scale, and since this trend also did not propagate to units of larger extents, this created inconsistencies. Other notable examples of masking include Silver Hake (Fig. 4F), *Pseudocalanus spp.* abundance (Fig. S4), and benthivore biomass (Fig. S5).

Local trend divergence (Fig. 1D) accounted for between 6% and 20% of inconsistencies when comparing the 250,000 km^2 scale indicators to ones for nested units but did not occur at all when comparing northern and southern units to ones nested within them (Table 2). One example is piscivore biomass, which would be perceived as increasing from the whole ecosystem version (250,000 km^2) but after disaggregation, we find the northern unit is actually decreasing (Fig. 4D). Similarly, the whole ecosystem indicator gives a perception of an increasing trend in American Lobster biomass, while the southern unit portrays a strongly decreasing trend (Fig. S6).

3.3. Scale sensitivity of indicators and spatial coherence (Ave r^p)

We used two quantitative metrics to describe scale sensitivity of specific indicators and found that *Ave* r_p provided a more intuitive and useful description than Theil-Sen slope variability (Table 3). The two metrics were not correlated (p = 0.46, r = -0.16) which means that indicators with high spatial coherence of the raw time series (represented by *Ave* r_p) do not necessarily have more consistent monotonic trends. This is not intuitive, but we note that the Theil-Sen slope variability included 31 slope estimates (for each indicator), that were in many cases representing non-significant trends. Many slope estimates, for time series without a trend, lead to erratic slope estimates and thus high Theil-Sen slope variation. This occurred even when the time series, as determined by *Ave* r_p and visual inspection of the data, were quite coherent across space. We therefore favored *Ave* r_p to describe which indicators were most sensitive to scale but report both (Table 3).

In general, scale sensitivity appeared to follow a pattern where physical indicators of temperature were not sensitive, lower trophic level indicators were intermediately sensitive, and mid trophic level indicators were most sensitive (Table 3). No matter how the ecosystem was subdivided, temperature indicators were increasing (i.e., the large block of red on the right, Fig. 5), among spatial unit variability of Theil-Sen slopes was low, and *Ave* r_p was high (Table 3). All Indicators of chlorophyll-a had high spatial coherence (0.71–0.76), indicators for zooplankton had intermediate coherence (0.51–0.66), and those for trawl survey indicators (Mid-TL) had the lowest spatial coherence (0.18–0.54, Table 3). An outlier to the apparent pattern of physical indicators being more coherent was that water column stratification was not spatially coherent, having an *Ave* r_p of 0.27.

3.4. Patterns between spatial extent, trend magnitude, and absolute indicator values

We found a greater range of trend magnitudes and absolute values of indicators when smaller and smaller subunits were used (Fig. 6). In simple terms this means that localized 'hotspots' of trend or abundance were portrayed in full strength when indicators were calculated for a small area, but when indicators were generated for bigger regions these 'hotspots' were subdued by inclusion of areas with no trend (or contrasting trend) and lower abundance. The same could be said for 'cold spots' (e.g., areas of notably decreasing trend or low abundance).

A notable example is the pattern depicted by Benthivore biomass, which is increasing significantly in only one of ten 20,000 km^2 units (in the North East, in what is called the Scotian Shelf) and lead to widespread trend propagation (Fig. 4E). When trends are compared to spatial extent (Fig. 6) there is *always a single unit* that has a trend far higher than others. Any unit that by chance happened to include this local area of increasing abundance ended up being driven by this local trend. The mean of all trends (dashed line) is quite close to the trend that would be depicted by a single indicator for the whole region. Similarly, the mean values of indicators (kg tow^{-1}) show greater variation with decreasing spatial extent. For example, some units had a high density of benthivores (max 95 kg tow^{-1}) whereas other units have very few benthivores (min = 7 kg tow^{-1}).

4. Discussion

We provide the first evaluation to our knowledge of how boundaries

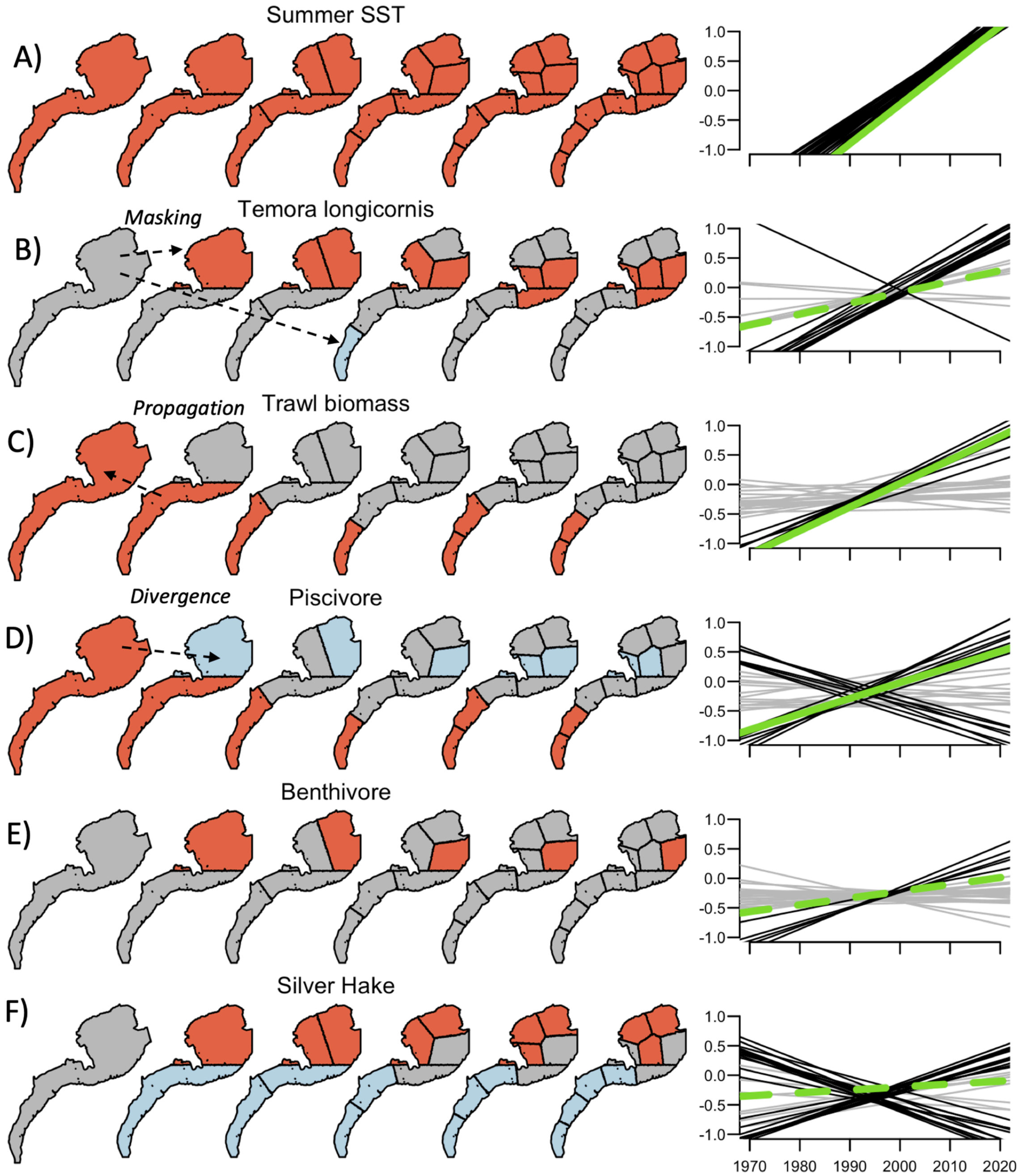

Fig. 4. Perceived trends in five indicators (A–E) based on six different ways of dividing the NES into units for an ecosystem assessment. The leftmost panel is the trend that would be perceived if the ecosystem was treated as a single unit (245,000 km^2), other panels in a row show progressive splitting, recalculation of the indicator, and re-analysis of the time series. Units are coded based on results of Mann-Kendall test for a monotonic trend on the indicator time series (red = increasing, blue = decreasing, grey = no trend). The farthest right panel shows the linear fits to the time series (31 lines appear in each panel, one corresponding to each spatial unit). Significant trends are shown in black, non-significant trends are in grey, and the trend for the whole ecosystem version (245,000 km^2) is green with a solid line if significant and a dashed line if not. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and ecosystem subunit delineation can influence an entire EA and found the perception of the ecosystem, in terms of 22 indicators, changed by as much as 38% as a function of scale. Although this specific estimate of change is a function of our own chosen boundaries and study design (*see below*, *section 4.1*), this study serves as an important demonstration of just how much scale and boundaries can influence a multidisciplinary EA. This emphasizes the importance of a preliminary step in developing *any* ecological or ecosystem indicator, whether part of an EA or not: Defining spatial scale and boundaries. Since the simple question "is there a trend?' is so strongly influenced by this step, which should be made early in the indicator development process, we expect many subsequent uses of indicators are also sensitive to scale. Thus, the scale that indicators are developed should be chosen carefully, and the consequences of different boundary systems explored to determine how they might influence indicator behavior and uses in management.

4.1. Important caveats to interpretation

Although we sought to imitate realistic steps in developing an EA, the fact that we did this quickly for 31 separate units at six spatial extents

Table 3
Indicators and metrics of their behavior including *Ave* r_p (a measure of spatial coherence), Slope variability (variability of Theil-Sen slopes fit to time series for the 31 units), increasing (number of significant increasing monotonic trends), and decreasing (number of decreasing monotonic trends). Rows are sorted according to *Ave* r_p within each category.

Category	Indicator	Ave r_p	Slope variability	Increasing	Decreasing
Physical	Summer SST	0.81	2.57	31	0
	MHW days	0.8	3.47	30	0
	Surface temperature	0.67	5.09	21	0
	Bottom temperature	0.66	10.02	23	0
	Stratification	0.27	25.51	0	0
Lower TL (chyl-a)	Chlorophyll	0.76	59.08	0	2
	Chlorophyll (sat)	0.71	42.61	0	0
Lower TL (zooplankton)	Sm/lg copepod ratio	0.66	19.02	0	0
	Zooplankton volume	0.52	11.27	0	0
	Pseudocalanus spp.	0.71	10.38	0	24
	Centropages typicus	0.61	13.74	0	0
	Temora longicornis	0.55	92.59	12	1
	Calanus finmarchicus	0.51	12.76	0	0
Mid TL (groups)	Benthivore	0.54	13.52	5	0
	Trawl biomass	0.3	24.00	8	0
	Benthos	0.28	3.01	14	0
	Planktivore	0.24	9.14	7	0
	Piscivore	0.18	47.55	8	7
Mid TL (species)	Silver Hake	0.3	58.52	11	12
	American Lobster	0.26	69.20	16	7
	Sea Scallop	0.26	3.66	12	0
	Spiny Dogfish	0.2	23.65	11	0

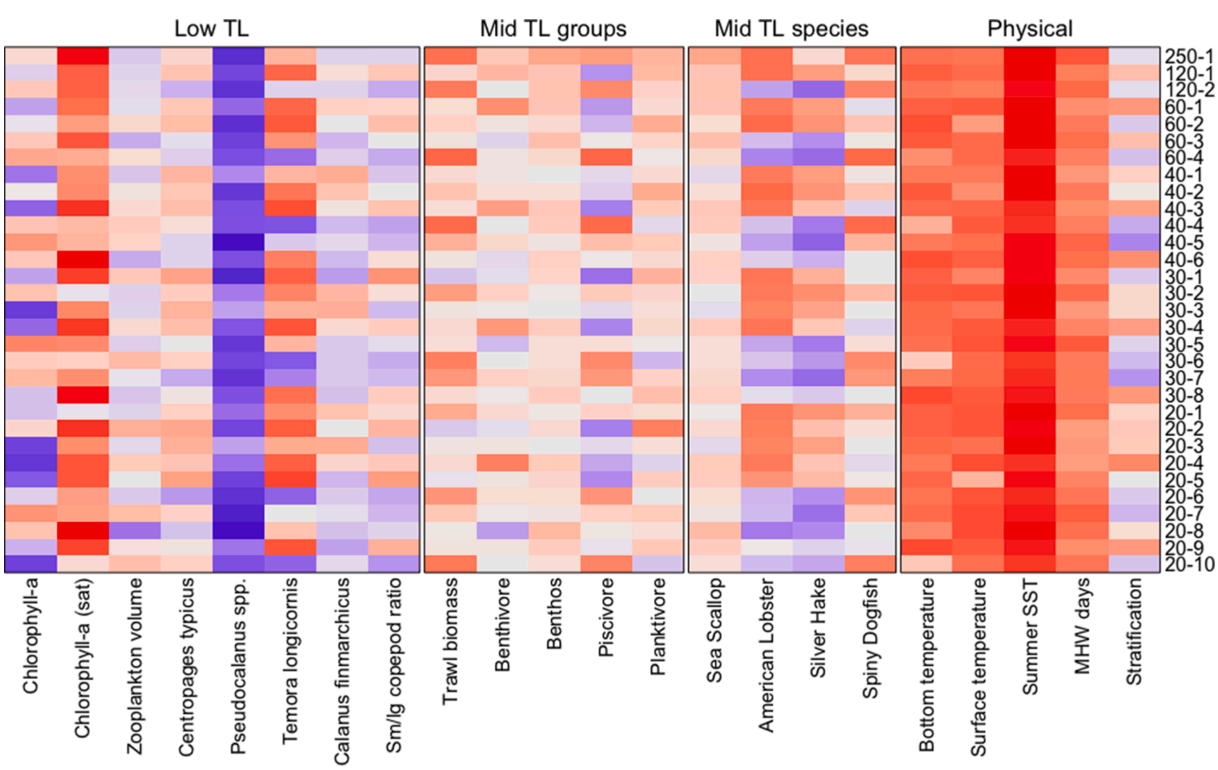

Fig. 5. Variation in Theil-Sen slopes of indicator time series (columns) for 31 different spatial units (rows, see Fig. 2). Slope values are depicted as shaded from dark red (strong increase) to dark blue (strong decrease). The TL used in headers refers to trophic level. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

precluded the careful nuance that would be applied to a full EA in practice. We highlight several points to be aware of when interpreting our results. First, the specific boundaries that we used (i.e., Fig. 2A) undoubtedly influenced our results. With a different boundary system (e.g., size of units, location of units, number of units) the exact values describing among unit disagreement would be different (e.g., Table 2). We are keenly aware that the main message of this article (i.e., choices regarding spatial scale and ecosystem boundaries influence indicator behavior) also applies to our own study. Second, we acknowledge the inadequacy of the null hypothesis-testing framework to examine statistical significance of an assumed monotonic trend (Wasserstein and Lazar, 2016; Hardison et al., 2019). For our study, we required a simple way to classify trends for comparison across units, and thus we used the now-routine ‘$p < 0.05$″ approach (with a correction for multiple testing), which remains common in the treatment and analysis of indicators in practice. However, strict reliance on this approach to interpret

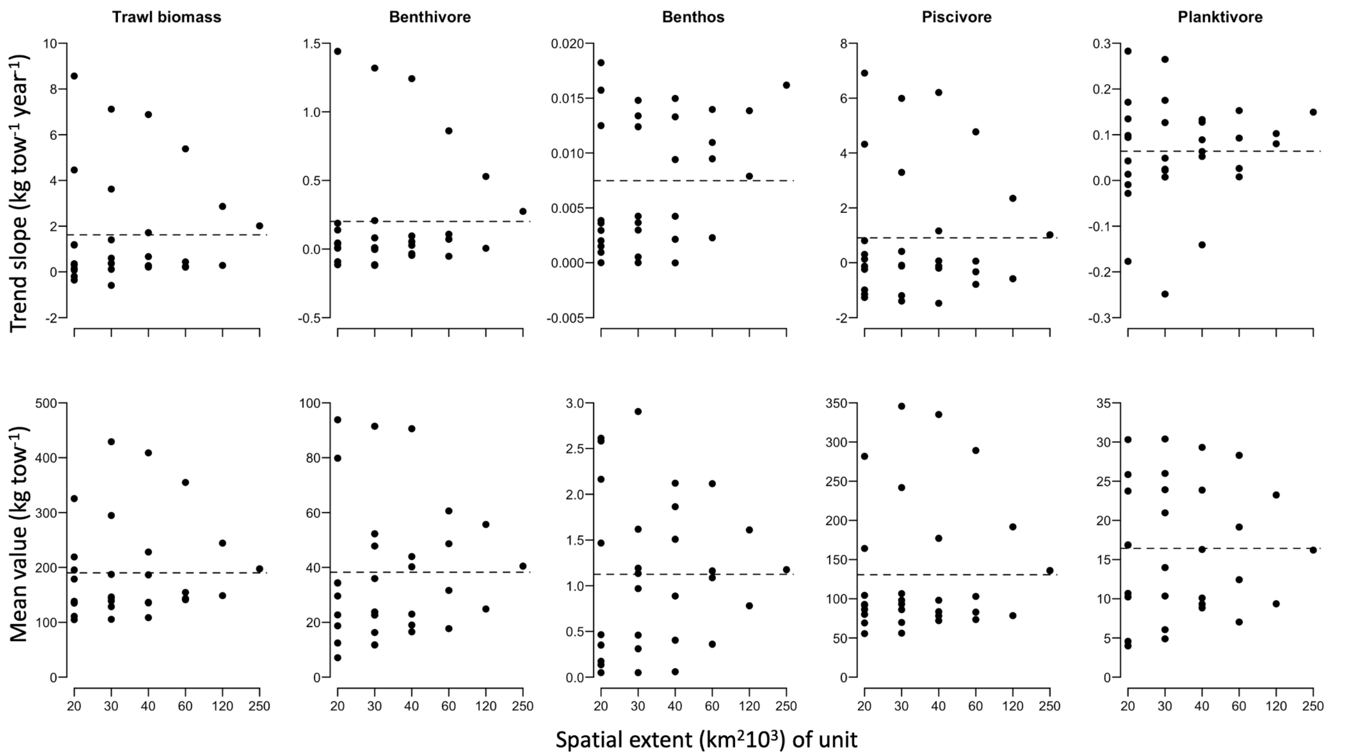

Fig. 6. Trends (Theil-Sen slopes) and mean indicator values (1970–2018) of aggregate species groups examined at multiple spatial scales. Points correspond to each spatial unit in Fig. 2A.

indicators and make consequential management decisions is not recommended in a real-world EA (Hardison et al., 2019). We do not expect, however, that these two methodological choices would have a substantive impact on our main findings.

Finally, the degree to which our results are generalizable to other ecosystems is not clear; the NES is highly complex, heavily impacted by anthropogenic causes, and is experiencing systematic warming at a rate greater than 99% of the worlds oceans (Pershing et al., 2015). That there is great spatial heterogeneity in the NES perhaps amplified the importance of scale, relative to a similar study done in a more homogenous LME. However, the main message we invoke (i.e., scale matters) is probably robust to all of these caveats, since the 'problem of scale' is a well-accepted issue across the biological sciences (Levin, 1992). The degree to which choices related to scale influence an indicator and an EA are no doubt context specific, and may be less important in certain contexts (e.g. a system with little or no warming) and perhaps even more than our study demonstrates in others (e.g., a system with heterogeneous warming). We encourage those developing indicators to assess the importance of scale in their applications.

4.2. Degree and common causes for inconsistencies

When preparing an ecosystem assessment or assessing environmental status, there are endless options for defining spatial boundaries and subunits (Queirós et al., 2016; Otto et al., 2018), and our results demonstrate that the specifics of these boundaries can markedly change the overall perception of the ecosystem. We found that the detected signals of ecosystem change depend on the way the ecosystem is defined. At the broadest spatial extent, the NES was treated as a single spatial unit to develop indicators, which is common in some research articles (Large et al., 2013) and EAs in the region (Fogarty et al., 2012). When we compared this whole ecosystem version of the 22 indicators to those for northern and southern subunits, results were 23% inconsistent (in both comparisons). Assuming the north and south units (each 120,000 km^2) represent regional dynamics well, we think of this as a 23% erosion of a potential signal due to averaging spatial variation. When the whole ecosystem version of indicators was compared to those for progressively smaller subunits, inconsistencies increased and reached a maximum of 36%. The main message is that a single EA will provide a big picture view of an ecosystem but can mis-represent local dynamics. The details of these inconsistencies (i.e. Fig. 1B–D) could have different implications for management.

The most common case for cross-scale inconsistencies was 'local trend propagation': Change is occurring, but not across the entire assessment region (sensu Fig. 1C). One management implication of local trend propagation is that a prescribed treatment for a problem is applied to an unnecessarily broad region. For example, if a broad scale indicator (e.g., the whole NES) portrays excessive fisheries bycatch, widespread fisheries closures could cause unnecessary economic losses if the bycatch was actually driven by a localized area or fishery gear type. This issue (and others) underly criticisms of static spatial management that promising tools, like dynamic ocean management, seek to resolve (Maxwell et al., 2015; Dunn et al., 2016; Hazen et al., 2018). In another scenario, localized trends in biodiversity, a notoriously scale-dependent and patchy ecosystem property (Peterson et al., 1998), could drive an undesirable trend. For example, such an instance could produce low scores for environmental status for the EU's MSFD biodiversity descriptor if spatial variation was not adequately recognized (Uusitalo et al., 2016).

The next most common cause for inconsistencies was local trend masking (sensu Fig. 1B). The critical point in these examples is that because of the chosen boundaries used, we fail to detect a trend. A well-understood example is given by Silver Hake, which is managed as two discrete stocks and is experiencing a poleward range shift (Nye et al., 2009, 2011). Because of these regionally divergent trends, we would not perceive a trend when aggregating to the whole NES when in fact important changes are happening (Link et al., 2011). Other instances of local trend masking are likely common in practice across the breadth of indicator applications. Fishing pressure may appear stable even as local fishing pressure intensifies (Preciado et al., 2019), broad scale indicators of pollution or eutrophication could appear stable as point sources vary (Gergel, 2005), or cultural service indicators could be insensitive to temporal change if societal interest shifted spatially within the study extent (Richards and Friess, 2015). In the NES few of our specific examples of masking are surprising because of an extensive body of research. Yet, this issue could be more problematic in poorly understood ecosystems where such a background knowledge is lacking.

The least common, but perhaps most consequential issue was when a broad scale indicator displayed a significant trend, but a nested spatial region showed a significant trend in the opposite direction (sensu Fig. 1D). This is exemplified by the American Lobster example. Again, we know that this species is experiencing a dramatic increase in the north and a decrease in the south of the NES (Steneck and Wahle, 2013). Our results demonstrate how this northern trend overwhelms the declines in the south, leading to a perception of increasing trends when aggregated to the whole NES. In terms of decision-making, this issue could be especially problematic. Imagine a very simple proposed management action based on a trigger (if indicator X is increasing, do A: if decreasing, do B); we might be doing A for a large region where locally the right approach is actually B. Although this was not a common finding in our case study, it could be a common pattern in other systems or with other indicators.

4.3. What indicator trends are most sensitive to scale?

Our results suggest a pattern where temperature indicators trends were the least sensitive to scale, and then indicators of the marine community varied predictably along trophic levels (i.e., phytoplankton, zooplankton, fish and invertebrates). These results are largely consistent with theory and field-specific research. For example, climatic variation and global warming operate across relatively broad spatial scales (Belkin, 2009). Although there is variability in rates of warming in the NES, as well as marine heatwave frequency and duration (Pershing et al., 2015; Oliver et al., 2018), nearly any way of dividing the NES will still reveal an increasing trend. In our study, temperature and heatwave indicators displayed the highest spatial coherence in the time series (Ave r_p 0.66–0.81). Patterns in chlorophyll were similarly insensitive to scale. Although we detected no long-term trends in chlorophyll, both the satellite derived and in-situ based indicator time series were highly coherent across spatial units (Ave $r_p = 0.76, 0.71$). Most zooplankton indicators were quite coherent across spatial subunits, but we did detect some divergent trends and lower coherence in some indicators (Ave r_p as low as 0.51 for *Calanus finmarchicus*). This is consistent with recent studies demonstrating regional variation in zooplankton dynamics in the NES within the four Ecological Production Units defined by NMFS (Morse et al., 2017). Lastly, mid-trophic level indicators (fish and invertebrates) displayed the lowest spatial coherence and highest variation in trends, consistent with Östman et al. (2017) that found strong spatial coherence in physical pressures, but far less for indicators of the coastal fish community in the Baltic Sea. Whether this apparent pattern in indicator scale sensitivity is supported by other indicators, in other ecosystems, is a topic worth further investigation. In any case, this highlights again a fundamental challenge when assessing ecosystems; a spatial scale that is effective for capturing spatial variation in one indicator might be inappropriate for another.

4.4. Consistency with scaling theory

The pattern of heteroscedasticity observed when comparing indicator trends and mean values to spatial extent (e.g., triangular pattern in Fig. 6) is noteworthy and consistent with expectations from scaling theory. The six spatial breaks used in our study are akin to changing grain with extent held constant and Weins' (1989) predictions are applicable. When data are compiled for the entire NES to produce a single annual mean, all spatial variation is lost to averaging (e.g., maximal within subunit variance, Fig. 7) and this produced intermediate

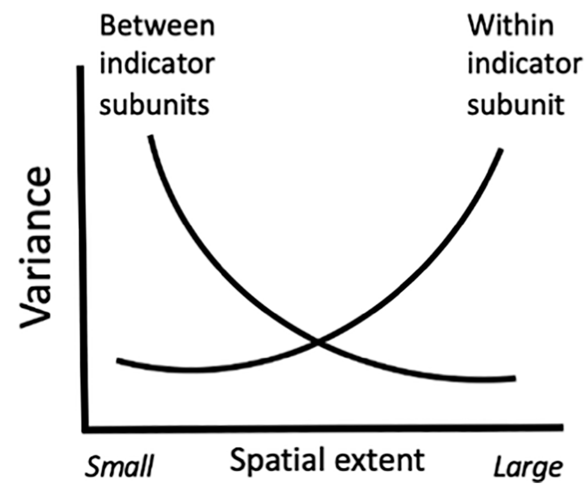

Fig. 7. Weins (1989) depiction of how variance changes with scale applied to indicators. Consider an indicator calculated for a single large spatial unit (large extent) versus the same indicator calculated for each of multiple smaller subunits (small extent). At a large extent, subunit values (or trends of indicator time series) are more similar to one another (low variance), but within subunit spatial variance is high and lost to averaging across a large region. As subunit extent is decreased, variance between subunits increases and within subunit spatial variance decreases.

trends and absolute values of indicators. At the smallest grain, 'hotspots' of local abundance and variability in temporal trends emerge (e.g., maximal between subunit variance, Fig. 7). These patterns arose by chance in our analysis, since subunit delineation was naïve to any spatial structure in the data. Had we based subunits on the spatial structure of indicator datasets, this variation may have been even more evident. For example, had we developed boundaries that matched distinct ecoregions (Lucey and Fogarty, 2013) or followed bathymetric features (e.g., on the continental shelf vs. shelf break), we would be maximizing within subunit similarity and exemplifying between subunit variation (Fig. 7). Our results suggest that when there is high spatial variation in an underlying process, broad scale indicators will produce subdued trends and values relative to indicators developed at smaller spatial scales. Whether capturing this variation is necessary will depend on the intended purpose and objectives for developing indicators in the first place.

A variety of spatially explicit methods would be useful to apply in an ecosystem indicator context to identify spatial variation. It should be noted that our work is not directly a spatial analysis; we repeatedly developed a simple EA using different boundary systems and conducted trend analysis in each case independently. This process matched how EAs are often conducted in practice (i.e., develop boundaries, calculate indicators, analyze indicators) to explore realistic issues that could arise; yet further application of spatial tools could be quite useful. For instance, semivariograms, a tool to explore the distance and strength of autocorrelation (Sokal and Oden, 1978) could be applied to either (1) raw survey data values or (2) mean values of indicators developed at different spatial extents to inform appropriate indicator scales. Or instead of defining distinct regions *a priori* (i.e., ecosystem boundaries), indicators could be assessed using spatially explicit analysis of raw survey datasets. The vector autoregressive spatio-temporal modeling tools (VAST, Thorson, 2019) for example, is being explored for this purpose in the Mid-Atlantic Bight (Scott Large, personal communication).

While we have focused on issues of spatial scale and invoked concepts from spatial scaling theory, many of these concepts could also be considered from a temporal perspective. For example, how does the length of an indicator time series influence the direction, strength, or statistical properties of a trend? Or, how would time series trends differ if indicators were calculated at the sub-annual scale (i.e., seasonal, monthly, or daily)? Holding space constant and assessing indicators at different temporal scales would be a promising extension of this work.

4.5. Implications for marine ecosystem assessments and management

We have stepped back to a preliminary step (i.e., one conducted before any subsequent analysis or decisions can be made *using* indicators) and found that different choices related to spatial scale and ecosystem boundaries can dramatically alter the perception of trends. Had we carried this analysis through subsequent steps in a management framework, we would undoubtedly have found that they too were sensitive to scale. Aside from broader awareness of this potential issue when using and interpreting indicators, we provide several recommendations to better incorporate spatial variation into indicator-based assessments.

First, spatial analysis should be done to inform the boundary systems used for indicator development and to compliment the simple, but integrative perspective provided by whole ecosystem assessments. By using an arbitrary boundary system, we showed that boundaries are quite important, and thus it is wise to delineate ecosystems and subunits based on the spatial structure of the ecosystem. For example, in the NES a spatial clustering analysis of physiographic, oceanographic, and primary production datasets was conducted by NMFS to define the aforementioned ecological production units (Lucey and Fogarty, 2013). Although NMFS recognizes limitations of these static boundaries (e.g., dependence on variables used, jurisdictional issues, and openness of the system, Gamble et al., 2016) they are ecologically meaningful and have proven useful for structured EAs. Also, while these integrated assessments may not portray spatial variation in all indicators, they are informed by and complimented with rigorous spatial analysis of survey datasets. Supporting mechanistic analyses help identify spatial variation that could lead to otherwise undetected instances of propagation, masking, and divergence (sensu Fig. 1, Nye et al., 2011; Link et al., 2011; Kleisner et al., 2017). Moreover, datasets are publicly available and so can be evaluated (as we have done) by other scientists in the context of more topic-specific research.

Our results also suggest scale considerations should be included in indicator selection routines, which often consider responsiveness to a known pressure as a desirable feature (Otto et al., 2018; Shin et al., 2018). We expect relationships between indicators and pressures are also scale dependent. For example, the strength of predator–prey relationships or the negative impacts of bottom trawling on food webs, can weaken or strengthen depending on the spatial scale considered (Frank and Leggett, 1981; Levin, 1992; Preciado et al., 2019). Thus, if a pressure operates and influences a process of interest at fine spatial scales, then indicators would be most responsive to this pressure if assessed at finer resolution (Preciado et al., 2019). An indicator may be judged as poor for myriad reasons, but one reason is simply that it was not developed at a characteristic spatial scale. While not considered explicitly here, we recommend that the influence of spatial scale in 'indicator vetting' routines be examined in more detail. The approach outlined in Bundy et al. (2019) is a promising advance in this regard, which suggests using a suite of indicators calculated at multiple spatial scales to a identify a consistently useful indicator suite. In doing this, the authors also recognized important spatial variation in the values and trends portrayed by the indicators reflective of environmental variation and differing anthropogenic pressures across the system.

For ecosystem indicators to be useful, they should be developed at scales matching those needed to make management decisions as well as those matching ecological processes. In most cases, goals and objectives should be defined for an EBM or EBFM framework and these should guide the scales at which indicators are developed. This could involve the same organization developing indicators at variety of scales, dependent on a weighted consideration of (1) the scales of specific ecosystem objectives and management needs and (2) the scales that the underlying processes operate at. For example, to provide ecosystem context for fisheries managers in a single-stock framework, NMFS annually re-calculates a suite of indicators (taken from the EAs) for regions matching the stock boundaries. Thus, indicators already developed in the EA process are spatially customized to match stock structure of managed species in the NES, that do not always not match Ecological Production Units used for EAs.

Yet, since the governance structure in many ecosystems is multiscale (i.e., federal, state, local user groups), there is also value in developing nested assessments of a large marine ecosystems to facilitate decision making by different groups (Steneck and Wilson, 2010). A state or local government may have a hard time making meaningful decisions when using indicators generated by federal decision makers that cover much broader regions (and are developed to address broader scale objectives). Likewise, federal agencies will find excessive subunits at small spatial extents to be unwieldly or uninformative relative to their own objectives. Thus, the idea that indicators are sensitive to scale (i.e., tell a different story, depending on the boundaries used) is not necessarily a problem, but an incentive to learn about ecosystems more deeply. Such a framework of 'nested' EA reports is also demonstrated in the NES by two regional reports by NMFS (Mid-Atlantic and New England), and several local assessments of connected systems (e.g., Long Island Sound Study, LISCCMP, 2015; and the Chesapeake Bay Program Hershner et al., 2007). Collectively, these can support integrated multiscale governance of ecosystems, but will require coordination across different levels of government (Steneck and Wilson, 2010) to avoid already complicated issues related to sector-based management (Crowder et al., 2006).

4.6. Conclusions

Indicators are sensitive to spatial scale, but some indicators are more sensitive than others. The implications of this fact will vary depending on the intended use of the indicator and the indicator suite. Deeper exploration of indicator trends, at multiple scales, allows for the recognition of masking, propagation, and asynchrony that may be occurring within a spatial domain of interest. That these cross-scale inconsistencies occur is not a problem in itself, but they are important to recognize when making decisions regarding the management of ecosystems.

CRediT authorship contribution statement

Kurt C. Heim: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Lesley H. Thorne:** Conceptualization, Resources, Writing - review & editing, Project administration, Funding acquisition. **Joseph D. Warren:** Conceptualization, Resources, Writing - review & editing, Project administration, Funding acquisition. **Jason S. Link:** Conceptualization, Writing - review & editing. **Janet A. Nye:** Conceptualization, Methodology, Data curation, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the NEFSC Ecosystem Dynamics and Assessment branch for making available computer code and data to reproduce indicators appearing in their reports. We acknowledge the efforts of all those at the NEFSC who have contributed to the planning and execution of the bottom trawl, plankton, scallop and many other surveys. We thank the entire NEFSC staff for their dedicated work in auditing and maintaining these databases, and thoughtful feedback on the ideas and results presented here. Specifically, we thank Scott Large, Lisa Methratta, and

Vincent Saba for feedback that improved this manuscript. This work was funded by the New York State Environmental Protection Fund Ocean and Great Lakes Program, NYS DEC.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2021.107522.

References

Azarovitz, T.R., 1981. A brief historical review of the Woods Hole laboratory trawl survey time series. Canad. Spec. Publ. Fish. Aquat. Sci. 58, 62–67.

Bal, P., Tulloch, A.I.T., Addison, P.F.E., McDonald-Madden, E., Rhodes, J.R., 2018. Selecting indicator species for biodiversity management. Front. Ecol. Environ. 16 (10), 589–598.

Belkin, I.M., 2009. Rapid warming of large marine ecosystems. Progr. Oceanogr. 81 (1–4), 207–213. Elsevier Ltd.

Boyce, D.G., Lewis, M.R., Worm, B., 2010. Global phytoplankton decline over the past century. Nature 466(7306), 591–596. Nature Publishing Group.

Bradbury, I.R., Hubert, S., Higgins, B., Bowman, S., Borza, T., Paterson, I.G., Snelgrove, P.V.R., Morris, C.J., Gregory, R.S., Hardie, D., Hutchings, J.A., Ruzzante, D.E., Taggart, C.T., Bentzen, P., 2013. Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. Evol. Appl. 6 (3), 450–461.

Bundy, A., Gomez, C., Cook, A.M., 2019. Scrupulous proxies: defining and applying a rigorous framework for the selection and evaluation of a suite of ecological indicators. Ecol. Ind. 104, 737–754.

Bunnefeld, N., Hoshino, E., Milner-Gulland, E.J., 2011. Management strategy evaluation: a powerful tool for conservation? Trends Ecol. Evol. 26 (9), 441–447.

Burthe, S.J., Henrys, P.A., Mackay, E.B., Spears, B.M., Campbell, R., Carvalho, L., Dudley, B., Gunn, I.D.M., Johns, D.G., Maberly, S.C., May, L., Newell, M.A., Wanless, S., Winfield, I.J., Thackeray, S.J., Daunt, F., 2016. Do early warning indicators consistently predict nonlinear change in long-term ecological data? J. Appl. Ecol. 53 (3), 666–676.

Ciannelli, L., Fauchald, P., Chan, K.S., Agostini, V.N., Dingsør, G.E., 2008. Spatial fisheries ecology: recent progress and future prospects. J. Mar. Syst. 71 (3–4), 223–236.

Colburn, L.L., Jepson, M., Weng, C., Seara, T., Weiss, J., Hare, J.A., 2016. Indicators of climate change and social vulnerability in fishing dependent communities along the Eastern and Gulf Coasts of the United States. Marine Policy 74, 323–333. Elsevier.

Crowder, L.B., Osherenko, G., Young, O.R., Airamé, S., Norse, E.A., Baron, N., Day, J.C., Douvere, F., Ehler, C.N., Halpern, B.S., Langdon, S.J., McLeod, K.L., Ogden, J.C., Peach, R.E., Rosenberg, A.A., Wilson, J.A., 2006. Resolving mismatches in U.S. ocean governance. Science 313 (5787), 617–618.

DePiper, G.S., Gaichas, S.K., Lucey, S.M., Da Silva, P.P., Anderson, M.R., Breeze, H., Bundy, A., Clay, P.M., Fay, G., Gamble, R.J., Gregory, R.S., Fratantoni, P.S., Johnson, C.L., Koen-Alonso, M., Kleisner, K.M., Olson, J., Perretti, C.T., Pepin, P., Phelan, F., Saba, V.S., Smith, L.A., Tam, J.C., Templeman, N.D., Wildermuth, R.P., 2017. Operationalizing integrated ecosystem assessments within a multidisciplinary team: lessons learned from a worked example. ICES J. Mar. Sci. 74 (8), 2076–2086.

Dunn, D.C., Maxwell, S.M., Boustany, A.M., Halpin, P.N., 2016. Dynamic ocean management increases the efficiency and efficacy of fisheries management. PNAS 113 (3), 668–673.

EcoAP (Ecosystem Assessment Program), 2009. Ecosystem Status Report For the Northeast U.S. Continental Shelf Large Marine Ecosystem. Northeast Fisheries Science Center Reference Document, 09-11: 34.

EcoAP (Ecosystem Assessment Program), 2011. Ecosystem status report for the northeast shelf large marine ecosystem-2011. Northeast Fish. Sci. Center Reference Document 12–07, 38.

Fogarty, M.J., Friedland, K.D., Col, L., Gambel, R., et al., 2012. Status of the northeast US continental shelf large marine ecosystem: an indicator-based approach. Am. Fish. Soc. Symp. 79, 1–28.

Foran, C.M., Link, J.S., Patrick, W.S., Sharpe, L., Wood, M.D., Linkov, I., 2016. Relating mandates in the united states for managing the ocean to ecosystem goods and services demonstrates broad but varied coverage. Frontiers in Marine Science 3 (FEB).

Frank, K.T., Leggett, W.C., 1981. Wind regulation of emergence times and early larval survival in Capelin (Mallotus villosus). Can. J. Fish. Aquat. Sci. 38(m):215–223.

Gamble R., Fogarty M., Keith, C., 2016. Ecological Production Units for the Northeast U. S. Continental Shelf. NEFSC Ecosystem and Climate Science Review, Woods Hole, MA. Accessed at https://www.integratedecosystemassessment.noaa.gov/sites/default/files/pdf/ne-ecological-production-units-paper.pdf.

Gaichas, S.K., DePiper, G.S., Seagraves, R.J., Muffley, B.W., Sabo, M.G., Colburn, L.L., Loftus, A.J., 2018. Implementing ecosystem approaches to fishery management: risk assessment in the US Mid-Atlantic. Front. Mar. Sci. 5 (November), 1–25.

Gergel, S.E., 2005. Spatial and non-spatial factors: when do they affect landscape indicators of watershed loading? Landscape Ecol. 20 (2), 177–189.

Hardison, S., Perretti, C.T., DePiper, G.S., Beet, A., 2019. A simulation study of trend detection methods for integrated ecosystem assessment. ICES J. Marine Sci. 76 (7), 2060–2069.

Harvey, C.J., Kelble, C.R., Schwing, F.B., 2017. Implementing "the IEA": using integrated ecosystem assessment frameworks, programs, and applications in support of operationalizing ecosystem-based management. ICES J. Mar. Sci. 74 (1), 398–405.

Hazen, E.L., Scales, K.L., Maxwell, S.M., Briscoe, D.K., Welch, H., Bograd, S.J., Bailey, H., Benson, S.R., Eguchi, T., Dewar, H., Kohin, S., Costa, D.P., Crowder, L.B., Lewison, R. L., 2018. A dynamic ocean management tool to reduce bycatch and support sustainable fisheries. Sci. Adv. 4 (5), 1–8.

Hershner, C., Havens, K., Bilkovic, D.M., Wardrop, D., 2007. Assessment of Chesapeake Bay Program selection and use of indicators. EcoHealth 4 (2), 187–193.

Hobday, A.J., Alexander, L.V., Perkins, S.E., Smale, D.A., Straub, S.C., Oliver, E.C.J., Benthuysen, J.A., Burrows, M.T., Donat, M.G., Feng, M., Holbrook, N.J., Moore, P.J., Scannell, H.A., Sen Gupta, A., Wernberg, T., 2016. A hierarchical approach to defining marine heatwaves. Progr. Oceanogr. 141, 227–238. Elsevier Ltd.

Kendall, M.G., 1957. Rank correlation methods. Biometrika.

Kleisner, K.M., Fogarty, M.J., McGee, S., Hare, J.A., Moret, S., Perretti, C.T., Saba, V.S., 2017. Marine species distribution shifts on the U.S. Northeast Continental Shelf under continued ocean warming. Progr. Oceanogr. 153, 24–36. Elsevier Ltd.

Large, S.I., Fay, G., Friedland, K.D., Link, J.S., 2013. Defining trends and thresholds in responses of ecological indicators to fishing and environmental pressures. ICES J. Mar. Sci. 70 (4), 755–767.

Large, S.I., Fay, G., Friedland, K.D., Link, J.S., 2015. Quantifying patterns of change in marine ecosystem response to multiple pressures. PLoS ONE 10 (3), 1–15.

Levin, P.S., Kelble, C.R., Shuford, R.L., Ainsworth, C., DeReynier, Y., Dunsmore, R., Fogarty, M.J., Holsman, K., Howell, E.A., Monaco, M.E., Oakes, S.A., Werner, F., 2013. Assessments: a US perspective. ICES J. Mar. Sci. 71, 1198–1204.

Levin, P.S., Möllmann, C., 2015. Marine Ecosystem regime shifts: challenges and opportunities for ecosystem-based management. Philos. Trans. Royal Soc. B: Biol. Sci. 370 (1659), 1–8.

Levin, S.A., 1992. The problem of pattern and scale in ecology. Ecology 73 (6), 1943–1967.

Link, J.S., Brodziak, J.K., Edwards, S.F., Overholtz, W.J., Mountain, D., Jossi, J.W., Smith, T.D., Fogarty, M.J., 2002. Marine ecosystem assessment in a fisheries management context. Can. J. Fish. Aquat. Sci. 59 (9), 1429–1440.

Link, J.S., Brodziak, J.K.T., editors, and Brodziak, J.K.T., Dow, D.D., Edwards, S.F., Fabrizio, M.C., Fogarty, M.J., Hart, D. Jossi, J. W., Kane, J., Lang, K.L., Legault, C.M., Link, J.S., MacLean, S.A., Mountain, D.G., Olson, J., Overholtz, W.J., Palka, D.L., Smith, T.D., contributors, 2002. Status of the Northeast U. S. Continental Shelf Ecosystem: a report of the Northeast Fisheries Science Center's Ecosystem Status Working Group. Northeast Fish. Sci. Cent. Ref. Doc. 02-11; 245 p. Link, J.S., Bundy, A., Overholtz, W.J., Shackell, N., Manderson, J., Duplisea, D., Hare, J., Koen-Alonso, M., Friedland, K.D., 2011a. Ecosystem-based fisheries management in the Northwest Atlantic. Fish Fish. 12(2), 152–170.

Link, J.S., Nye, J.A., Hare, J.A., 2011. Guidelines for incorporating fish distribution shifts into a fisheries management context. Fish Fish. 12, 461–469.

Link, J.S., Dickey-Collas, M., Rudd, M., McLaughlin, R., Macdonald, N.M., Thiel, T., Ferretti, J., Johannesen, E., Rae, M., 2019. Clarifying mandates for marine ecosystem-based management. ICES J. Mar. Sci. 76, 41–44.

LISCCMP, 2015. Long Island Sound Comprehensive Conservation and Management Plan 2015. Accessed at. https://longislandsoundstudy.net/wp-content/uploads/2015/09/CCMP_LowRes_Hyperlink_singles.pdf.

Lucey, S.M., Fogarty, M.J., 2013. Operational fisheries in New England: Linking current fishing patterns to proposed ecological production units. Fish. Res. 141, 3–12.

Lynch, H.J., Naveen, R., Trathan, P.N., Fagan, W.F., 2012. Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. Ecology 93 (6), 1367–1377.

Mann, Henry, B., 1945. Nonparametric tests against trend. Econometrica 13(3), 245–259.

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70(2–3 SPEC. ISS.):141–159.

Maxwell, S.M., Hazen, E.L., Lewison, R.L., Dunn, D.C., Bailey, H., Bograd, S.J., Briscoe, D.K., Fossette, S., Hobday, A.J., Bennett, M., Benson, S., Caldwell, M.R., Costa, D.P., Dewar, H., Eguchi, T., Hazen, L., Kohin, S., Sippel, T., Crowder, L.B., 2015. Dynamic ocean management: defining and conceptualizing real-time management of the ocean. Marine Policy 58, 42–50.

Methratta, E.T., Link, J.S., 2006. Evaluation of quantitative indicators for marine fish communities. Ecol. Ind. 6 (3), 575–588.

Morse, R.E., Friedland, K.D., Tommasi, D., Stock, C., Nye, J., 2017. Distinct zooplankton regime shift patterns across ecoregions of the U.S. Northeast continental shelf Large Marine Ecosystem. J. Marine Syst. 165, 77–91. Elsevier B.V.

NEFSC, 2017a. State of the Ecosystem—Mid-Atlantic Bight. Northeast Fisheries Science Center, Woods Hole, MA.

NEFSC, 2017b. State of the Ecosystem—Gulf of Maine and Georges Bank. Northeast Fisheries Science Center, Woods Hole, MA.

NEFSC, 2018a. State of the ecosystem 2019: Mid-Atlantic. Retrieved from https://www.integratedecosystemassessment.noaa.gov/regions/northeast/reports. 2019.

NEFSC, 2018b. State of the ecosystem 2019: New England. Retrieved from https://www.integratedecosystemassessment.noaa.gov/regions/northeast/reports. 2019.

NEFSC, 2019a. State of the ecosystem 2019: Mid-Atlantic. Retrieved from https://www.integratedecosystemassessment.noaa.gov/regions/northeast/reports. 2019.

NEFSC, 2019b. State of the ecosystem 2019: New England. Retrieved from https://www.integratedecosystemassessment.noaa.gov/regions/northeast/reports. 2019.

Nye, J.A., Joyce, T.M., Kwon, Y.O., Link, J.S., 2011. Silver hake tracks changes in Northwest Atlantic circulation. Nat. Commun. 2 (1), 412–416. Nature Publishing Group.

Nye, J.A., Link, J.S., Hare, J.A., Overholtz, W.J., 2009. Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf. Mar. Ecol. Prog. Ser. 393, 111–129.

Oliver, E.C.J., Donat, M.G., Burrows, M.T., Moore, P.J., Smale, D.A., Alexander, L.V., Benthuysen, J.A., Feng, M., Sen Gupta, A., Hobday, A.J., Holbrook, N.J., Perkins-Kirkpatrick, S.E., Scannell, H.A., Straub, S.C., Wernberg, T., 2018. Longer and more frequent marine heatwaves over the past century. Nat. Commun. 9 (1), 1–12. Springer US.

Östman, Ö., Lingman, A., Bergström, L., Olsson, J., 2017. Temporal development and spatial scale of coastal fish indicators in reference ecosystems: hydroclimate and anthropogenic drivers. J. Appl. Ecol. 54 (2), 557–566.

Otto, S.A., Kadin, M., Casini, M., Torres, M.A., Blenckner, T., 2018. A quantitative framework for selecting and validating food web indicators. Ecol. Indic. 84 (March 2017), 619–631. Elsevier.

Pershing, A.J., Alexander, M.A., Hernandez, C.M., Kerr, L.A., Le Bris, A., Mills, K.E., Nye, J.A., Record, N.R., Scannell, H.A., Scott, J.D., Sherwood, G.D., Thomas, A.C., 2015. Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. Science 350 (6262), 809–812.

Peterson, G., Allen, C.R., Holling, C.S., 1998. Ecological resilience, biodiversity, and scale. Ecosystems 1 (1), 6–18.

Preciado, I., Arroyo, N.L., González-Irusta, J.M., López-López, L., Punzón, A., Muñoz, I., Serrano, A., 2019. Small-scale spatial variations of trawling impact on food web structure. Ecol. Indic. 98 (June 2018), 442–452. Elsevier.

Queirós, A.M., Strong, J.A., Mazik, K., Carstensen, J., Bruun, J., Somerfield, P.J., Bruhn, A., Ciavatta, S., Flo, E., Bizsel, N., Özaydinli, M., Chuševė, R., Muxika, I., Nygård, H., Papadopoulou, N., Pantazi, M., Krause-Jensen, D., 2016. An objective framework to test the quality of candidate indicators of good environmental status. Front. Mar. Sci. 3 (May).

R Development Core Team, 2019. R: A Language and Environment for Statistical Computing.

Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., Schlax, M.G., 2007. Daily high-resolution-blended analyses for sea surface temperature. J. Clim. 20 (22), 5473–5496.

Rice, J.C., Rochet, M.J., 2005. A framework for selecting a suite of indicators for fisheries management. ICES J. Mar. Sci. 62 (3), 516–527.

Richards, D.R., Friess, D.A., 2015. A rapid indicator of cultural ecosystem service usage at a fine spatial scale: content analysis of social media photographs. Ecol. Indic. 53, 187–195. Elsevier Ltd.

Sathyendranath, S., Brewin, R.J.W., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., Cipollini, P., Couto, A.B., Dingle, J., Doerffer, R., Donlon, C., Dowell, M., Farman, A., Grant, M., Groom, S., Horseman, A., Jackson, T., Krasemann, H., Lavender, S., Martinez-Vicente, V., Mazeran, C., Mélin, F., Moore, T.S., Müller, D., Regner, P., Roy, S., Steele, C.J., Steinmetz, F., Swinton, J., Taberner, M., Thompson, A., Valente, A., Zühlke, M., Brando, V.E., Feng, H., Feldman, G., Franz, B.A., Frouin, R., Gould, R.W., Hooker, S.B., Kahru, M., Kratzer, S., Mitchell, B.G., Muller-Karger, F.E., Sosik, H.M., Voss, K.J., Werdell, J., Platt, T., 2019. An ocean-colour time series for use in climate studies: the experience of the ocean-colour climate change initiative (OC-CCI). Sensors (Switzerland) 19 (19).

Schlegel, R.W., Smit, A.J., 2018. HeatwaveR: a central algorithm for the detection of heatwaves and cold-spells. J. Open Source Softw. 3 (27), 821.

Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc. 63, 1379–1389.

Sherman, K., 1991. The large marine ecosystem concept: research and management strategy for living marine resources. Ecol. Appl. 1 (4), 349–360.

Shin, Y.J., Bundy, A., Shannon, L.J., Simier, M., Coll, M., Fulton, E.A., Link, J.S., Jouffre, D., Ojaveer, H., MacKinson, S., Heymans, J.J., Raid, T., 2010. Can simple be useful and reliable? Using ecological indicators to represent and compare the states of marine ecosystems. ICES J. Mar. Sci. 67 (4), 717–731.

Shin, Y.J., Houle, J.E., Akoglu, E., Blanchard, J.L., Bundy, A., Coll, M., Demarcq, H., Fu, C., Fulton, E.A., Heymans, J.J., Salihoglu, B., Shannon, L., Sporcic, M., Velez, L., 2018. The specificity of marine ecological indicators to fishing in the face of environmental change: a multi-model evaluation. Ecol. Ind. 89 (February), 317–326.

Shin, Y.J., Shannon, L.J., 2010. Using indicators for evaluating, comparing, and communicating the ecological status of exploited marine ecosystems. 1. the indiSeas project. ICES J. Mar. Sci. 67 (4), 686–691.

Sokal, R.R., Oden, N.L., 1978. Spatial autocorrelation in biology 2: some biological implications and four examples of evolutionary and ecological interest. Biol. J. Linn. Soc. 10, 229–249.

Rose, G.A., Leggett, W.C., 1990. The importance of scale to predator-prey spatial correlations: an example of atlantic fishes. Ecology 71 (1), 33–43.

Steneck, R.S., Wahle, R.A., 2013. American lobster dynamics in a brave new ocean. Can. J. Fish. Aquat. Sci. 70 (11), 1612–1624.

Steneck, R.S., Wilson, J.A., 2010. A fisheries play in an ecosystem theater: challenges of managing ecological and social drivers of Marine fisheries at multiple spatial scales. Bull. Mar. Sci. 86 (2), 387–411.

Tam, J.C., Link, J.S., Rossberg, A.G., Rogers, S.I., Levin, P.S., Rochet, M.J., Bundy, A., Belgrano, A., Libralato, S., Tomczak, M., Van De Wolfshaar, K., Pranovi, F., Gorokhova, E., Large, S.I., Niquil, N., Greenstreet, S.P.R., Druon, J.N., Lesutiene, J., Johansen, M., Preciado, I., Patricio, J., Palialexis, A., Tett, P., Johansen, G.O., Houle, J., Rindorf, A., 2017. Towards ecosystem-based management: identifying operational food-web indicators for marine ecosystems. ICES J. Mar. Sci. 74 (7), 2040–2052.

Theil, H., 1992. A rank-invariant method of linear and polynomial regression analysis. In: Raj, B., Koerts, J. (Eds.), Henri Theil's Contributions to Economics and Econometrics. Springer, Dordrecht, pp. 345–381.

Thorson, J.T., 2019. Guidance for decision using the Vector Autoregressiv Spatio-Temoral (VAST) package in stock, ecosystem, habitat and climate assessments. Fish. Res. 210, 143–161.

Turnhout, E., Hisschemöller, M., Eijsackers, H., 2007. Ecological indicators: between the two fires of science and policy. Ecol. Ind. 7 (2), 215–228.

Uusitalo, L., Blanchet, H., Andersen, J.H., Beauchard, O., Berg, T., Bianchelli, S., Cantafaro, A., Carstensen, J., Carugati, L., Cochrane, S., Danovaro, R., Heiskanen, A. S., Karvinen, V., Moncheva, S., Murray, C., Neto, J.M., Nygård, H., Pantazi, M., Papadopoulou, N., Simboura, N., Srebaliene, G., Uyarra, M.C., Borja, A., 2016. Indicator-based assessment of marine biological diversity-lessons from 10 case studies across the european seas. Frontiers in Marine. Science 3 (SEP).

Wasserstein, R.L., Lazar, N.A., 2016. The ASA's statement on p-values: context, process, and purpose. Am. Statist. 70, 129–133.

Weins, J.A., 1989. Spatial scaling in ecology. Funct. Ecol. 3 (4), 385–397.

Wood, S.N., 2001. mgcv: GAMs and generalized ridge regression for R. R News.