Ecological Indicators 104 (2019) 764–775

Contents lists available at ScienceDirect

Ecological Indicators

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

Original Articles

Development and evaluation of NatureServe's multi-metric ecological integrity assessment method for wetland ecosystems

Don Faber-Langendoen[a,*], Joanna Lemly[b], William Nichols[c], Joe Rocchio[d], Kathleen Walz[e], Regan Smyth[a]

[a] NatureServe, Conservation Science Division, 2511 Richmond Highway, Suite 930, Arlington, VA 22202, United States
[b] Colorado Natural Heritage Program, Colorado State University, 252 General Services Building, 1475 Campus Delivery, Fort Collins, CO 80523-1475, United States
[c] NH Natural Heritage Bureau, Division of Forests & Lands – DNCR, 172 Pembroke Road, Concord, NH 03301, United States
[d] Washington Dept. of Natural Resources, Natural Heritage Program, Olympia, WA 98504, United States
[e] Natural Heritage Program, NJDEP MC 501-04, Office of Natural Lands Management, 501E State Street, 4th Floor, Trenton, NJ 08625, United States

ARTICLE INFO

Keywords:
Ecological integrity
Rapid assessment
Discriminatory power
Redundancy
Wetland
Terrestrial
Metrics

ABSTRACT

Many ecological monitoring and assessment programs include rapid assessment methods that employ indicators or metrics to track the degree of divergence of ecosystem condition from reference conditions. Although these rapid assessment methods use a combination of metrics to rate overall ecological condition, they rarely include tests of either the merits of the component metrics being assessed or the method of aggregating the metrics into an overall rating. We used a conceptual model of ecological integrity for wetlands and field data to select and test 15 rapid assessment indicators (using specific metrics) across a spectrum of major ecological factors or MEFs (landscape, buffer, vegetation, hydrology, soil). We applied these metrics to 220 wetland sites across six states (Colorado, Indiana, Michigan, New Hampshire, New Jersey, and Washington), using two assessment area (AA) approaches: 106 sites used 0.5 ha point-based AAs; 114 sites used variable-sized polygon-based AAs. We statistically tested metric ratings and factor scores for their discriminatory power (DP) in relation to a stressor index using the Kruskal-Wallis test, and for redundancy using Spearman rank correlation and scattergrams. Of the 15 metrics, 12 had good or strong DP and were not redundant. Across all metrics, only two pairs (vegetation pair and buffer pair) were strongly correlated. The soil metric had the lowest DP, but it was among the least redundant of any metric. The DP of buffer metrics was lower for point-based approaches than for polygon-based approaches because the buffer for point-based AAs often included additional wetland area. Aggregating individual metrics into MEF scores (e.g., vegetation, hydrology, soil), primary factors scores (Landscape Context, on-site Condition) and overall ecological integrity ratings, either maintained or improved the interpretability of the ratings. Our analyses support the use of 12 rapid field-based metrics, spanning Landscape Context and on-site Condition, to assess the ecological integrity of wetlands. Although tested here for wetlands, the models and metrics are also being applied to upland terrestrial ecosystems. Our findings confirm the merits of our rapid assessment method in providing an intermediate level of assessment that is efficient and ecologically meaningful, within states and across watersheds and regions.

1. Introduction

Ecosystem monitoring and assessment programs are critical for resource management, conservation, and restoration activities, given the dramatic losses, degradation, and alteration that have occurred to many ecosystems. These programs increasingly emphasize not only the changing area of ecosystems (e.g., Dahl, 2011), but also their changing ecological health, condition, or, as we refer to here, ecological integrity (Woodley et al., 1993; US EPA, 2002, 2016; Stoddard et al., 2006; Wurtzebach and Schultz, 2016). Concepts of biotic and ecological integrity range from the specifically ecological to larger philosophical, religious, and ethical concerns (Leopold, 1966; Rolston, 1993) as well as aspects of ecological resilience and ecosystem services (Goodin et al., 2018). Here we address the core ecosystem aspects of ecological integrity as they relate to conservation and restoration activities, mitigation decisions at site and watershed or landscape scales, and land use

* Corresponding author.
E-mail addresses: don_faber-langendoen@natureserve.org (D. Faber-Langendoen), joanna.lemly@colostate.edu (J. Lemly), William.Nichols@dncr.nh.gov (W. Nichols), joe.rocchio@dnr.wa.gov (J. Rocchio), Kathleen.Walz@dep.nj.gov (K. Walz), Regan_Smyth@natureserve.org (R. Smyth).

https://doi.org/10.1016/j.ecolind.2019.04.025
Received 11 September 2018; Received in revised form 7 April 2019; Accepted 10 April 2019
Available online 18 May 2019

planning (Fennessy et al., 2007; Faber-Langendoen et al., 2008). However, ecosystems are complex, and not all aspects of a system's condition or integrity can be readily measured, if only for cost reasons. Thus, to meet program objectives, tested and validated indicators are needed to provide reliable, cost-effective, and ecologically meaningful information on ecological integrity. If well-chosen, these indicators can also provide early signs of ecosystem degradation and identify risks of ecosystem collapse.

The process of selecting indicators is challenging because ecological condition can be assessed at a variety of scales, from site to landscapes and continents. Foundational to most monitoring programs is the ability to assess condition at the site or local ecosystem scale, and to then integrate those assessments into other scales (these scales may also require their own independent indicators). Indicators are often chosen from the broad spectrum of ecosystem attributes, given the interaction of changes in the physical, chemical, and biological components of ecosystems. Multiple indicators are often needed to assess whether ecological condition is good, fair, or poor (National Research Council, 1994).

In the last 10–20 years, assessments of wetland condition have become particularly common, and a wide variety of methods have been developed, ranging from more qualitative rapid assessment methods (RAMs) to intensive, quantitative methods (Fennessy et al., 2007; US EPA, 2016). RAMs play an important role in monitoring programs because they are less costly and require less expertise compared to quantitative methods. Therefore, they can be repeated more often, providing more information for ongoing adaptive-based management. There are now a wide diversity of RAMs available, but many have had limited testing or the methodology and testing have been limited to a particular region, state, or set of wetland types (Mack, 2001; Brooks et al., 2006; Sutula et al., 2006; Wardrop et al., 2007, 2013). Validation of the suite of rapid indicators is often done by rolling the indicators into an index and then comparing the index against other independent, quantified variables, such as stressor gradients or rigorous quantitative data on vegetation or soils. If the overall index correlates well enough with these independent variables, then the method is judged to be validated (Mack, 2006; Sutula et al., 2006; Stein et al., 2009). This is a good first step; however, rarely are the component indicators within the rapid method evaluated. This next step is important because of the expense of collecting and managing site-based field data, even for RAMs. In addition, some RAM users find value in focusing on the individual metrics to provide insight into management priorities and restoration goals.

NatureServe and its Network partners from Natural Heritage Programs, in collaboration with a variety of agency partners, have developed a rapid assessment of ecosystem condition, structured around the concept of ecological integrity (Faber-Langendoen et al., 2016a). Our Ecological Integrity Assessment (EIA) method is a multi-metric approach, similar to the Index of Biotic Integrity for aquatic systems (Karr and Chu, 1999) and to a variety of state-based wetland RAMs (Fennessy et al., 2007). Here, we propose to test the performance of each of our specific indicators (hereafter "metrics"), using data collected across multiple states and a wide variety of wetland types. We assessed performance based on the ability of the metrics to discriminate a range of stressor levels (from low to very high), with minimal redundancy across metrics. We also assessed whether or not aggregating the scores of individual metrics into major ecological factors and an overall ecological integrity rating provided additional insights into the ecological condition of wetlands. Finally, we compared the performance of the metrics when data were collected using a point (0.5 ha)-based versus polygon-based sampling approach.

2. Background: the ecological integrity assessment method

2.1. Conceptual ecological models and metrics

We have described the Ecological Integrity Assessment (EIA) method in detail in other publications (Unnasch et al., 2009; Faber-Langendoen et al., 2012, 2016a; Mitchell et al., 2014). Here we provide a brief summary to set the stage for our analyses.

2.1.1. Conceptual model

We define our assessment of ecological integrity as "an assessment of the structure, composition, function, and connectivity of an ecosystem as compared to reference ecosystems operating within the bounds of natural or historical disturbance regimes" (Faber-Langendoen et al., 2016a). To have integrity, an ecosystem should be relatively unimpaired across a range of characteristics and spatial and temporal scales (Andreasen et al., 2001).

Conceptual models are a critical step in understanding how ecosystems function. They help us identify the major ecological factors (MEFs) that characterize the ecological drivers and dynamics of the ecosystem, and which we must address when making management decisions to maintain ecological integrity (Mitchell et al., 2014). Our model, summarized in Faber-Langendoen et al. (2016a), is based on characteristics of ecosystems known from historical conditions spanning the variety of wetland and upland terrestrial ecosystem types. We draw our understanding of these characteristics from known or historical reference sites, where conditions are not impacted or are minimally impacted by negative anthropogenic stressors (Davies and Jackson, 2006; Stoddard et al., 2006). The model includes three primary organizing factors: Landscape Context, (on-site) Condition, and Size. These primary factors are subdivided into the six MEFs: landscape, buffer, vegetation, hydrology, soil, and size, which capture the structure, composition, processes, and connectivity of most terrestrial systems. Other major factors, such as animals (e.g., birds, fish) or particular ecological processes (e.g., fire), can be assessed where resources, time and field sampling design permit. The model is fairly intuitive, but a key component is that the model includes both the "inner workings" (Condition) and the "outer workings" (Landscape Context) of an ecosystem (Parkes et al., 2003; Oliver et al., 2007).

Although our overall model includes Size as one of the primary factors, we exclude it from our analyses here because it is not a consistent measure of ecological integrity. Ecosystem types typically have examples that vary in size for very natural reasons; e.g., an acidic bog may occur in a small isolated kettle or in an expansive depression. However, because there are often conservation values associated with larger examples (higher species richness, greater resilience to stressors, better habitat for area-dependent species), Size can be added to Landscape Context and Condition to produce a final rating, called an Element Occurrence Rank (EO RANK) (NatureServe, 2002; Regan et al., 2004; Faber-Langendoen et al., 2016c), which helps prioritize ecosystem occurrences of highest conservation or resource management value (Unnasch et al., 2009). But these conservation priorities go beyond considerations of ecological integrity *per se*.

2.1.2. A 3-level approach to metric selection

A multi-level approach is often used to assess ecological integrity (US EPA, 2002; Wardrop et al., 2013). Level 1 (remote assessment) relies primarily on remote sensing indicators of landscape integrity, such as used in NatureServe's Landscape Condition Model (Hak and Comer, 2017). Level 2 (rapid field assessment) uses relatively simple semi-quantitative or qualitative field-based indicators, often supplemented by a stressor checklist (see Section 2.1.4.). Level 3 (intensive field assessment) requires detailed field measurements, typically involving plot or transect based sampling of the vegetation, intensive soil sampling, and/or quantitative measures of hydrology. The metrics used at each of the three levels can also be calibrated with each other to

Table 1
Ecological integrity metrics for wetland rapid assessment (Level 2). Some metrics have variants based on either wetland type in the U.S. National Vegetation Classification (USNVC) (e.g., bog & fen, marsh, floodplain & swamp forests, mangrove) or on hydrogeomorphic (HGM) class (e.g., tidal, nontidal, riverine, depression-lacustrine-slope). Two optional metrics are only applicable to forested wetlands. See Faber-Langendoen et al. (2016b) for details on metric protocols (Size metrics are not shown).

Primary Rank Factors	Major Ecological Factors	Metrics	Metric Variant	Metric Variant Type
LANDSCAPE CONTEXT	LANDSCAPE	LAN1. Contiguous Natural Land Cover		
		LAN2. Land Use Index		
	BUFFER	BUF1. Perimeter with Natural Buffer		
		BUF2. Width of Natural Buffer		
		BUF3. Condition of Natural Buffer		
CONDITION	VEGETATION	VEG1. Native Plant Species Cover		
		VEG2. Invasive Nonnative Plant Species Cover		
		VEG3. Native Plant Species Composition	Y	USNVC
		VEG4. Vegetation Structure	Y	USNVC
		VEG5. Woody Regeneration [opt.]		
		VEG6. Coarse Woody Debris [opt.]	Y	USNVC
	HYDROLOGY	HYD1. Water Source	Y	HGM
		HYD2. Hydroperiod	Y	HGM
		HYD3. Hydrologic Connectivity	Y	HGM
	SOIL	SOI1. Soil Condition	Y	USNVC

provide a multi-level assessment (Stein et al., 2009).

2.1.3. Fine-tuning the model for specific ecosystem types

Given the wide variety of ecosystems, there is a balance in identifying indicators that are both widely applicable and sensitive to how different ecosystems function. Ecological classifications can be helpful tools in the indicator selection process because they help managers better understand natural variability within and among types, and thereby help us recognize differences between occurrences with good integrity and poor integrity (e.g., the evaluation of the hydrology of tidal salt marshes will be distinct from the hydrology of bogs or floodplain forests). We first develop widely usable metrics, then add variants as needed to ensure that they are sensitive to the particular processes of individual ecosystem types. Hydrologic metric variants correspond to the different hydrogeomorphic (HGM) class types (Brinson, 1993) and vegetation metric variants correspond to the U.S. National Vegetation Classification (USNVC) formation types (ESA Panel, 2015, usnvc.org). Wetlands are classified using both classification systems so that metric variants are properly applied (e.g., a forested swamp that is classified to the HGM depressional class would be evaluated using hydrological metric variants specific to that type; similarly, when classified to the USNVC “Floodplain & Swamp Forest” formation, it would be evaluated using vegetation metrics specific to that type). See Faber-Langendoen et al. (2016b) for a description of each metric and its variants.

2.1.4. Ecological integrity and stressors

Stressor checklists assist with further understanding the factors that affect individual metrics and the overall condition of an ecosystem (Sutula et al., 2006). Metrics are scaled based on a “stressor-dose response” to changes in stressor levels. The term “stressor” (or direct threats) is defined as “the proximate (human) activities or processes that have caused, are causing, or may cause the destruction, degradation, and/or impairment of biodiversity and natural processes” (Master et al., 2012). We restrict our focus to those stressors that are associated with observed impacts whenever the associated effects of the stressors are evident (i.e., we exclude potential future threats). For example, a direct stressor may be recent tree removal or mowing. Less recent mowing or tree removal would be included only if the effect of those stressors is still currently evident (e.g., old tree stumps, continued erosion from past logging).

Our EIA methodology includes a stressor checklist that systematically scores the Scope and Severity of each stressor present at a site (Faber-Langendoen et al., 2016b, adapted from Collins et al., 2006). Scope is defined as the proportion of an ecosystem or its buffer (0–100 m zone surrounding the focal ecosystem) that is currently affected by the stressor, including stressors that may have occurred in the past, but the effect is still currently evident (e.g., past logging that has removed all large trees from a stand, resulting in a current small tree size class structure, or increased runoff). Severity is the level of damage to the ecosystem or buffer from the stressor, based on recent existing evidence (i.e., approximately a 10 year time frame). Severity is assessed by known or inferred degree of degradation or decline in integrity to specific major ecological factors, such as the buffer, vegetation, soils, and hydrology. Together these define the level of Impact (Low, Medium, High, or Very High) that a given stressor has on the ecosystem (e.g., Small Scope and Slight Severity = Low Impact, whereas Pervasive Scope and Extreme Severity = Very High Impact).

2.2. Initial development of the model and rapid assessment metrics

Over the past 10 years, we used our conceptual ecological model to develop rapid assessment (Level 2) metrics that are informative of the ecological integrity of wetlands (Faber-Langendoen et al., 2008, 2012, 2016a). Briefly the development process included: a) identify the major ecological factors and key ecological attributes of the ecosystem; b) select indicators and metrics; c) determine assessment points or thresholds; d) calculate metric and overall integrity ratings; and e) create summary reports, including scorecards. This past work led to a working EIA model and draft set of metrics, which were applied and tested in 2008–2010 on 12 sites in Colorado (Lemly and Rocchio, 2009) and 277 wetland sites in Michigan and Indiana (Faber-Langendoen et al., 2012). As a result of those studies we proposed 15 metrics for landscape, buffer, vegetation, hydrology, and soil that were most responsive, practical, cost-effective, and ecologically meaningful in measuring the integrity of wetlands (Table 1).

Individual state programs began applying, testing, and refining the NatureServe 2012 EIA, typically through U.S. Environmental Protection Agency (EPA) funding of state-based applications of wetland condition assessments (Walz and Domber, 2011; Nichols and Faber-Langendoen, 2012; Nichols, 2013; Lemly and Gilligan, 2015; Rocchio et al., 2016). In all states, training sessions were held with field crews to ensure that ecological integrity metric protocols were consistently applied. For example, in the Michigan-Indiana study (Faber-Langendoen et al., 2012), a two-day training session was held; then, after the first 10 sites were visited, data were reviewed by team leaders, and further clarification given to achieve consistency in scoring. Finally, in mid-season, several sites were resampled and if the metric scores differed by more

Table 2
Examples of three metrics and ratings from three Major Ecological Factors (MEFs). See Table 1 for a full list of metrics for each MEF. Metrics typically have four assessment point ratings, but occasionally three or five are used (see VEG2), depending on the sensitivity and variability of the metric.

	LANDSCAPE MEF	VEGETATION MEF	HYDROLOGY MEF
Metric Rating	***LAN1. Contiguous Natural Land Cover: ALL WETLANDS***	***VEG2. Invasive Nonnative Plant Species Cover*: ALL WETLANDS**	***HYD1. Water Source*** ***Variant: DEPRESSION, LACUSTRINE, SLOPE***
EXCELLENT (A) (4 points)	**Intact:** Embedded in 90–100% natural habitat around AA.	Invasive nonnative plant species apparently absent.	Water source is natural: site hydrology is dominated by precipitation, groundwater, and/or natural runoff from an adjacent freshwater body. There is no indication of direct artificial water sources. Land use in the local drainage area of the site is primarily open space or low density, passive uses. Lacks point source discharges into or adjacent to the site.
GOOD (B) (3 points)	**Variegated:** Embedded in 60–90% natural habitat.	Invasive nonnative plant species in any stratum present but sporadic (1–3% cover).	Water source is mostly natural, but site directly receives occasional or small amounts of inflow from anthropogenic sources. Indications of anthropogenic input include developed land or agricultural land < 20% in the immediate drainage area of the site, small storm drains or other local discharges emptying into the site, or some road runoff. No large point sources discharge into or adjacent to the site.
FAIR (C) (C = 2 points, C- = 1.5 points)	**Fragmented**: Embedded in 20–60% natural habitat.	C. Invasive nonnative plant species in any stratum somewhat common (4–10% cover). C. Invasive nonnative plant species in any stratum common (11–30% cover).	Water source is moderately impacted by anthropogenic sources, but is still a mix of natural and non-natural sources. Indications of moderate contribution from anthropogenic sources include developed land or irrigated agriculture that comprises 20–60% of the immediate drainage many small storm drains or a few large ones, or moderate road runoff.
POOR (D) (1 point)	**Relictual**: Embedded in < 20% natural habitat.	Invasive nonnative plant species in any stratum abundant (> 30% cover).	Water source is substantially impacted by anthropogenic sources (e.g., urban runoff, direct irrigation, pumped water, artificially impounded water, or other artificial hydrology). Indications of substantial artificial hydrology include > 60% developed or agricultural land adjacent to the site, and the presence of major point sources that discharge into or adjacent to the site, or large amounts of road runoff.

than one categorical rating, and if overall EIA rating differed by 10%, a meeting was held to determine the basis for the differences. Other states conducted similar field training, and two states have published field testing results (Lemly and Rocchio, 2009; Nichols, 2013).

In 2013, we initiated a multi-state collaboration to further refine our EIA methodology. We formed a team of NatureServe and Network staff from Natural Heritage Programs in Colorado, New Hampshire, New Jersey, and Washington, spanning multiple EPA regions, and together we refined the metric criteria and assessment points (or thresholds) from 2012 (Table 1). Examples of the metrics are provided in Table 2, and a full description of each metric is provided in Faber-Langendoen et al. (2016b). Of the 15 metrics, 13 were treated as core (always used) and 2 were optional (used if applicable to the wetland type, such as forested wetlands). In addition, as shown in Table 1, eight metrics have variants based on broadly distinct ecological patterns and processes, as recognized by vegetation formation types of the USNVC (e.g., marsh, floodplain forest, and bog) or hydrologic distinctions of HGM classes (e.g., riverine - nonriverine and tidal - nontidal). Full details of each metric are provided in Faber-Langendoen et al. (2016b).

3. Methods

3.1. Site data

We compiled 220 sites across six states in which the EIA had been applied: Colorado (60 sites, 2013–2014 field sampling), Indiana (26, 2009–2010), Michigan (61, 2009–2010), New Hampshire (10, 2009–2012), New Jersey (46, 2012–2014), and Washington (17, 2012–2104). These sites typically had brief field notes describing the condition of the site, and an assigned condition rating from A (Excellent) to D (Poor), along with a full EIA. We selected sites that included a wide range of major wetland types, preferentially occurred on public lands, were logistically accessible, contained both Level 2 data and quantitative vegetation plot data, and which spanned the range of A to D condition ratings. The latter criterion was included to increase the likelihood that our data contained a wider range of ecological integrity and stressor levels, but we also independently conducted a standard assessment of stressors (see Section 3.4).

3.2. Sampling design method

Despite the overall similarity in applying the EIA method in the six states, there were two different sampling designs used across these states – a polygon approach and a point-based approach.

Polygon approach (116 sites): Four of the states (IN, MI, NH, and WA) were sampled using a polygon-based approach, where the goal was to assess the ecological integrity and conservation value of a particular wetland type at a site. The assessment area (AA) was delineated based on the full extent of the wetland type at that location with general uniformity in condition.

Point-based approach (104 sites): Two of the states (CO and NJ) were sampled using a point-based approach, where the goal was to assess wetland condition across a watershed. Sites were chosen using a spatially balanced random sample design based on the Generalized Random Tessellation Stratified (GRTS) method (Stevens and Olsen, 2004) to statistically represent the condition of wetland area across the watershed. With this approach, the AA surrounding the point was restricted to a small area (in our case, 0.5 ha), similar to the US EPA (2016) protocols for the National Wetland Condition Assessment (NWCA). The purpose was to represent condition at the point rather than condition within the entire wetland type. The full extent of the wetland type being sampled was not delineated, so size was not recorded.

The two methods differed most strongly in the way in which the buffer and surrounding landscape were assessed; the 100 m buffer surrounding AAs sampled with the point-based approach often included the same wetland type as the point (especially if the wetland type was large), whereas AAs sampled with the polygon approach extended to the edge of the wetland type and the buffer was a different wetland, an adjacent upland, or some combination of both.

3.3. Field data collection

For all sites, the six state programs conducted similar field data collection methods. First, a pre-field assessment was conducted on the polygon or point using the most current aerial photography. Landscape and buffer metrics were given preliminary ratings based on information from the imagery, and notes were made of on-site conditions (ditches, ATV tracks, logging, etc.).

Second, an on-site field survey of the AA was conducted, including checking for any notable variation in condition. A basic, narrative site description was completed, summarizing features of each MEF. The site was classified to a wetland type, initially based on the primary ecological classification used by the state program, and then to the USNVC (usnvc.org) and HGM (Brinson, 1993) classifications. All 15 Level 2 EIA metrics were scored. Metric variants were applied, as needed, based on the wetland type (Faber-Langendoen et al., 2016b). Metric ratings were later checked for consistency in the office.

Third, for all 220 sites, a vegetation plot (though not required for Level 2 assessments) was also sampled in a representative part of the polygon. Plots were either a minimum of 10x10 m in shrub-herb wetlands or 20 × 20 in tree wetlands, or a nested 0.1 ha plot design that was applicable to all wetlands (Peet et al., 1998). A list of species was compiled, and their percent cover estimated. These data were used both for providing quantitative support for the Level 2 vegetation metrics (e.g., VEG2 Invasive Nonnative Plant Species Cover) and to calculate Level 3 metrics based on floristic quality (DeBerry et al., 2015). Depending on the project, additional Level 3 data on soils were also collected, including soil texture, pH, soil color and depth, and depth of water in the soil pit.

3.4. Stressor data and indices

Each of the states applied a stressor checklist in the field. Checklists were reviewed and standardized for all 220 sites to the format of Faber-Langendoen et al. (2016b) to allow consistent rating of stressors to the four MEFs - buffer, vegetation, hydrology, and soil - and to the overall site. Scope and Severity (Table 3) were rated for each observed stressor.

3.5. Data management and scorecards

Metrics data and stressor checklists were entered into NatureServe's Ecological Observations Database (EcoObs). EcoObs is currently managed by NatureServe staff and is used by a number of state natural heritage programs. The database is structured to match field data protocols, including fields for general site description, environmental description, vegetation plot data, Level 1–3 metrics, stressor checklists, scorecards, and descriptive soils data.

3.5.1. Ecological integrity scorecard

For this project, metrics were used to generate an EIA scorecard for each site (Fig. 1). The protocols for the ecological integrity metrics specify the assessment points (or thresholds) that indicate increasing departure from reference conditions. Metrics typically have a 4-point rating system from A (4 points) to D (1 point) (Table 2). After metrics are rated in the field and office, their ratings are incorporated into a scorecard that uses the metric rating points to generate ratings for the MEFs, for Landscape Context and Condition, and for overall Ecological Integrity (Fig. 1). At the outset, each metric was considered to contribute equally to our understanding of ecological integrity and received the same weight. The metric scores were typically averaged to generate a rating or for each MEF. The one exception was the buffer MEF, in which a two-step geometric mean was used to integrate the three metrics in order to increase the sensitivity of the lowest metric rating to the score (following Collins et al., 2006). In turn, the MEFs were weighted, as follows: for the Condition rank factor, vegetation was given more weight (0.55) than hydrology (0.35), as it was the most visible and accurately measured of the MEFs. The soil MEF was given a low weight (0.1), because initial testing in 2012 showed it to be sensitive to only the most heavily stressed sites (see Faber-Langendoen et al., 2012). Within the Landscape Context rank factor, the buffer MEF, which captured the inner 100 m effects of stressors, was weighted twice (0.66) that of the 500 m landscape MEF (0.33). Landscape Context and Condition scores were then weighted (0.3 and 0.7 respectively) to generate an EIA rating or grade.

Table 3
Stressor categories for Scope and Severity (from Master et al., 2012). The combination of Scope and Severity determine the Impact rating (e.g., Small Scope and Small Severity = Low Impact, whereas Pervasive Scope and Extreme Severity = Very High Impact).

SCOPE of Threat (% of AA or Buffer affected by direct threat)	
1 = Small	Affects a small (1–10%) proportion of the AA or Buffer
2 = Restricted	Affects some (11–30%)
3 = Large	Affects much (31–70%)
4 = Pervasive	Affects most or all (71–100%)
SEVERITY of Threat within the defined Scope (degree of degradation to AA or Buffer)	
1 = Slight	Likely to only slightly degrade/reduce
2 = Moderate	Likely to moderately degrade/reduce
3 = Serious	Likely to seriously degrade/reduce
4 = Extreme	Likely to extremely degrade/destroy or eliminate

3.5.2. Human stressor index (HSI)

From the stressor checklists, we entered the Scope and Severity ratings assigned to each stressor identified in the AA or its buffer into EcoObs (Table 3). The EcoObs database automatically assigned an Impact rating based on the combination of Scope and Severity ratings (e.g., Small Scope and Small Severity = Low Impact, whereas Pervasive Scope and Extreme Severity = Very High Impact; see Faber-Langendoen et al., 2016b). In turn, each rating was scored (Low – 1 point, Medium – 4 points, High – 7 points, Very High – 10 points). Points were summed across stressors within each MEF. Points were then weighted by the MEF (buffer 0.3, vegetation 0.3, hydrology 0.3, soil 0.1) to produce an overall Human Stressor Index (HSI_{total}) score and rating (0– < 1 = Absent, 1– < 4 = Low, 4– < 7 = Medium, 7– < 10 = High, and ≥10 = Very High) (Faber-Langendoen et al., 2016b). We used the HSI as a means of testing how well the individual metrics, the MEFs, and overall EIA rating responded to stressor levels. We used two forms of the HSI to establish stressor categories: HSI_{total} combined the stressor ratings from all four MEFs into an overall score and rating. This form of the HSI potentially contained autocorrelation with some of the EIA metrics that include stressor information to assign the condition rating (see e.g., HYD1 Water Source in Table 2). For that reason, we also calculated $HSI_{abiotic}$, which combined buffer (0.4), hydrology (0.4), and soil (0.2) MEF scores, with which we tested the effectiveness of on-site vegetation (biotic) metrics in responding to these abiotic stressors.

3.6. Statistical tests

3.6.1. Discriminatory power of metrics and major ecological factors

A variety of statistical methods are available to help assess the statistical rigor of metrics, applicable to both rapid and intensive metrics. Three commonly assessed criteria include comprehensive range of metric variation, discriminatory power (particularly in response to stressors), and redundancy (Blocksom et al., 2002; Klemm et al., 2003; Jacobs, 2010; Wang et al., 2015).

Discriminatory power (DP) is the ability of a metric, MEF, or rank factor, to distinguish levels of stress, from Absent to High. We compared how well each metric and factor (Table 1) was able to distinguish these stressor levels based on HSI ratings. We examined their response using box-and-whisker plots and one-way analysis of variance (ANOVA) (appropriate for Level 2 categorical data). We used the Kruskal-Wallis rank sum test, which is the non-parametric analogue of a one-way ANOVA and makes no assumptions about normality. We then applied the Pairwise Wilcoxon Rank Sum Tests to calculate pairwise comparisons between group levels and used the Hochberg method for p-values adjustments for multiple comparisons. Variables with non-significant F-values from the ANOVA were considered non-responsive and candidates for removal. We used the R software for statistical analyses (R

State/Prov: CO **ObsArea Code:** LA-PF04
ObsArea Name: Arkansas River Floodplain **ObsDate:** 2014/07/07
Project: CO Lower Ark **County:** CO - Bent
Observers: C. Wiechmann, S. Rubin

Macrogroup: M028 Great Plains Flooded & Swamp Forest
Ecological System: CES303.678 Western Great Plains Floodplain
HGM: Riverine

Human Stressors Index (HSI) Scores: *Total:* 2.7 (High) *Abiotic:* 2 (Medium) *Onsite:* 2 (Medium)
by Major Ecological Factor (MEF): *Buffer:* 1 (Low) *Veg:* 4 (Very High) *Soil:* 0 (Absent) *Hydro:* 4 (Very High)

Protocol: NatureServe Wetland 2016

	Wt	Field Rating	Field Pts	Calc Pts	Calc Rating
ECOLOGICAL INTEGRITY				2.24	C+
Rank Factor: LANDSCAPE CONTEXT	0.3			2.89	B-
MEF: LANDSCAPE	0.33			3.00	B+
LAN1. Contiguous Natural Land Cover	1	A	4		
LAN2. Land Use Index	1	C	2		
MEF: BUFFER	0.66			2.83	B-
BUF1. Perimeter with Natural Buffer	n/a	A	4		
BUF2. Width of Natural Buffer	n/a	A	4		
BUF3. Condition of Natural Buffer	n/a	C	2		
Rank Factor: CONDITION	0.7			1.97	C-
MEF: VEGETATION	0.55			1.79	C-
VEG1. Native Plant Species Cover	1	C-	1.75		
VEG2. Invasive Nonnative Plant Species Cover	1	D	1		
VEG3. Native Plant Species Composition	1	D	1		
VEG4. Overall Vegetation Structure	1	C	2		
VEG5. Woody Regeneration	1	D	1		
VEG6. Coarse Woody Debris	1	A	4		
MEF: HYDROLOGY	0.35			1.67	C-
HYD1. Water Source	1	C	2		
HYD2. Hydroperiod	1	D	1		
HYD3. Hydrologic Connectivity	1	C	2		
MEF: SOIL	0.1			4.00	A+
SOI1. Soil Condition	1	A	4		

Fig. 1. Ecological integrity scorecard. Weights, field ratings and points, calculated points (scores), and ratings are shown for each metric, Major Ecological Factor (MEF), rank factor, and overall Ecological Integrity for a wetland site. Additional site information is provided in the heading. The scorecard is generated from NatureServe's Ecological Observations Database (EcoObs).

Core Team, 2015). When conducting our analyses, we first looked at DP across all 220 sites, then by method (points – 106 sites; polygons – 114 sites). Metrics with poor discriminatory power were candidates for removal, pending redundancy correlations (below, Section 3.6.2.)

3.6.2. Redundancy correlations among metrics

We screened all metrics for redundancy. If two or more metrics have a strong linear relationship to one another, then one of them is contributing no new information to the assessment. We used the method of Fulton et al. (2005) to screen for redundancy. First, we assessed the degree of correlation between all pairs of metrics. Categorical levels of correlation were: $r \geq 0.9$ very strongly redundant, $r = 0.7–0.89$ (strongly redundant), $r = 0.5–0.69$ (moderately redundant), $r < 0.5$ (minimal or low redundancy). Second, we examined the scatterplots of all pairs with moderate, strong, or very strong correlations ($r \geq 0.5$) to determine the nature of the relationship. If the relationship between metrics was curved, cone-shaped, or had a wide spread in some ranges, we concluded that the metrics were minimally or not redundant. Absolute value of the correlation, and its linearity or triangularity, are more meaningful than statistically significant correlations, especially when relatively large data sets are involved. This is because, as Fulton et al. (2005) note, when there are a large number of data points involved, even bi-plots with fairly flat relationships could have statistically significant correlation coefficients. Thus, only tight relationships where both metrics fell close to a straight line were considered redundant. If there were redundant metrics after these assessments, the metric chosen was the one with better DP. Conversely, a metric with low DP (and thus candidate for removal), but low redundancy, could still be retained.

3.7. Landscape context and on-site condition

Finally, we tested the degree to which Landscape Context and its MEFs (landscape and buffer) discriminated (using the Kruskal-Wallis test) or correlated with on-site Condition (the integrated scores of vegetation, hydrology, and soil).

4. Results

4.1. Wetland types and reference gradient

Our 220 sites spanned the various wetland formations and macrogroups found in the U.S. (Table 4). Our sites also spanned the stressor gradient, based on HSI_{total}: Absent (26 sites), Low (93), Medium High/

Table 4
The 220 sites classified to USNVC Formation and Macrogroup. Type descriptions are available on usnvc.org.

FORMATION Macrogroup	No. of Sites
FLOODED & SWAMP FOREST (81)	
Central & Appalachian Floodplain Forest	13
Central & Appalachian Swamp Forest	3
Great Plains Floodplain Forest	29
Northern Flooded & Swamp Forest	28
Rocky Mountain & Great Basin Montane Riparian Forest	3
Southern Coastal Plain Basin Swamp & Flatwoods	1
Vancouverian Flooded & Swamp Forest	4
MARSH, WET MEADOW, & SHRUBLAND (104)	
Arid West Interior Freshwater Emergent Marsh	12
Atlantic & Gulf Coastal Plain Wet Prairie & Marsh	18
Eastern Cool Temperate Seep	2
Eastern North American Marsh, Wet Meadow, & Shrubland	40
Eastern North American Ruderal Wet Meadow & Marsh	7
Great Plains Wet Meadow, Marsh & Playa	13
Vancouverian Lowland Marsh, Wet Meadow, & Shrubland	7
Western North American Montane-Subalpine Wet Shrubland & Wet Meadow	5
SALT MARSH (3)	
North American Atlantic & Gulf Coast Salt Marsh	3
BOG & FEN (32)	
North American Boreal & Sub-Boreal Acidic Bog & Fen	17
North American Boreal & Sub-Boreal Alkaline Fen	11
North Pacific Bog & Fen	4
Total	**220**

Very High (27). The latter category was combined because we had few (3) sites with “Very High” stress. Our scores for EIA also showed a full range of scores: A+ (23 sites), A− (49), B+ (55), B− (49), C+ (25), C− (16), D (3). These data indicate that our metrics span a comprehensive range of ecological integrity and stressor variation.

4.2. Discriminatory power

4.2.1. Discriminatory power among metrics

The degree of discriminatory power (DP) for each metric is summarized in Table 5. Box plots for each metric were reviewed to validate DP values. For the overall data set, 10 of the 13 core metrics had High to Very High DP, indicating that they successfully discriminated between either 3 (High) or all 4 (Very High) levels of stressors. Two of the buffer metrics (B1, B2) and the soil metric (S1) had moderate or low DP. The two optional vegetation structure metrics (V5, V6) also had moderate to low DP. Although these metrics performed poorly, they still had statistically significant DP. For the soil metric, its DP was still statistically significant ($p < 0.01$) because it helped separate highly stressed sites from sites with moderate to absent stressors.

The effect of the two sampling designs (point vs polygon) was most notable for landscape and buffer metrics. The polygon-based data showed an improvement in the DP for the two buffer metrics (B1 and B2) (Table 5). Conversely, for the point-based approach, B1 performed poorly, and L1 had lower DP. Conversely, the buffer condition metric (B3) was the most discriminatory of any metric. The soil condition metric (S1) had low DP with either sampling method. DP of on-site vegetation, hydrology, and soil metrics were comparable between the two approaches, with somewhat better DP for the Condition metrics when using the point-based sampling design.

4.2.2. Discriminatory power among major ecological factors

Using the scorecard approach (Fig. 1), we generated rating scores for each MEF and tested their DP against the HSI_{total} stressor ratings (Table 5, Fig. 2). For all sites combined, the vegetation, hydrology, buffer, and landscape MEFs all showed High or Very High DP in separating stressor levels. The DP of the vegetation MEF was also greater than its component vegetation metrics; that is, the best individual vegetation metric had a DP of F = 83 (V3), whereas the vegetation MEF had F = 89. Similarly, DP of the hydrology MEF (F = 118) was well above the individual hydrology MEFs (F = 76–88); it also had the best DP of any MEF (Table 5). Not surprisingly, the soil MEF, which was based on the one metric that performed poorly, had low DP, but it helped distinguish highly stressed sites from others. Both the buffer MEF and landscape MEF had very high DP, with buffer somewhat higher than landscape DP (F = 92, 75 respectively). Rolling them together into an overall Landscape Context rank factor score further improved the DP (F = 97).

The vegetation MEF correlation with $HSI_{abiotic}$ is not as strong as with HSI_{total} ($F = 62 < F = 89$, and $r = -0.56 < r = -0.68$, respectively) but still High. This is suggestive of some autocorrelation between the ecological integrity metric ratings and the stressor ratings.

When separating the data by sampling design (point vs polygon approach), DP was comparable for landscape and buffer MEFs (Table 5), unlike for individual metrics (Section 4.2.1 above). As with individual metrics, the sampling design had little effect on DP scores for vegetation, hydrology, or soil MEFs.

4.2.3. Overall ecological integrity

Our overall rating of ecological integrity integrated on-site Condition score with the Landscape Context score into an EIA score and rating. Our correlation of HSI_{total} with EIA scores was very robust ($r = -0.81$) (Fig. 3), indicating that our combination of ecological metrics was responsive to the changing levels of stressors.

4.3. Redundancy correlations among metrics

We assessed all pairs of metrics with moderate to very strong correlations ($r > 0.5$) (Table 6). Encouragingly, no metric pairs had “very strong” correlations and only three metrics had “strong” correlations: B1 (Perimeter with Natural Buffer) with B2 (Width of Natural Buffer), V1 (Native Plant Species Cover) with V2 (Invasive Nonnative Plant Species Cover), and V2 with V3 (Native Plant Species Composition). We carefully examined each of the scattergrams for any additional evidence of redundancy (Fig. 4). All metrics with moderately redundant correlations had scattergrams that were triangular or scattered, and thus they have minimal redundancy (Fig. 4). The strong correlation of metric B1 (Perimeter with Natural Buffer) with B2 (Width of Natural Buffer) and their relatively tight scattergram is not unexpected, in that they were both designed to assess aspects of buffer dimension; whenever the perimeter is missing a natural buffer, the width at that point also becomes zero. The strong correlation of V1 (Native Plant Species Cover) with V2 (Invasive Nonnative Plant Species Cover), and their relatively tight scattergram (Fig. 4) indicated that they were largely redundant. This is also not unexpected in that V2 is to some degree a subset of the information used for V1. Although V2 and V3 (Native Plant Species Composition) were strongly correlated, their scattergram was not tight (Fig. 4); therefore, we did not consider them redundant.

Notably, the soil metric, despite its low DP, was only one of two metrics (the other being V6) that showed minimal redundancy with any other metric (Table 6).

4.4. Landscape context and on-site condition

Our comparison of Landscape Context metrics and MEFs with overall on-site Condition showed strong correlations between the two (Table 7). MEFs always performed as well or marginally better than individual metrics. For example, for all sites, the correlations for the overall Landscape Context rank factor (r = 0.63) was higher than for the buffer MEF (r = 0.61) and landscape MEF (r = 56), and these in turn were as high as or higher than the correlations for individual metrics. These support the use of a multi-metric approach to

Table 5
Discriminatory Power (DP) of metrics and Major Ecological Factors (MEF),.based on the Kruskal-Wallis (KW) test. KW values are categorized separately for all sites versus sites by sampling design because DP thresholds differ based on number of sites in the analysis, i.e.: All Sites: Low DP < 25, Moderate DP 25–49, High DP 50–74 (Bold), Very High DP ≥ 75 (Bold underlined). Sites by Sampling Design: Low DP < 15, Moderate DP 15–24, High DP 25–34 (Bold), Very High DP ≥ 35. (Bold, underlined)

Metric Code	Metric Name	Combined (n = 220)		Polygon (n = 114)		Points (n = 106)	
		KW	P	KW	P	KW	p
	LANDSCAPE	**74.9**	***	**43.2**	***	**31.4**	***
LAN1	Contiguous Natural Land Cover	**50.6**	***	**37.3**	***	15.8	**
LAN2	Land Use Index	**74.5**	***	**35.1**	***	**39.0**	**
	BUFFER	**92.2**	***	**46.0**	***	**49.3**	***
BUF1	Perimeter with Natural Buffer	37.0	***	**39.4**	***	6.4	ns
BUF2	Width of Natural Buffer	41.0	***	**32.9**	***	23.8	***
BUF3	Condition of Natural Buffer	**80.8**	***	**29.6**	***	**52.2**	***
	VEGETATION	**89.1**	***	**40.2**	***	**50.2**	***
VEG1	Native Plant Species Cover	**50.8**	***	**31.2**	***	**40.3**	***
VEG2	Invasive Nonnative Plant Species Cover	**75.0**	***	**30.4**	***	**46.0**	***
VEG3	Native Plant Species Composition	**83.3**	***	**37.7**	***	**46.3**	***
VEG4	Vegetation Structure	**60.7**	***	**25.0**	***	**38.2**	***
VEG5	Woody Regeneration [opt.]	(105) 31.0	***	(46) 10.0	*	(59) **28.0**	***
VEG6	Coarse Woody Debris [opt.]	(111) 17.7	***	(52) 10.1	*	(59) 11.0	*
	HYDROLOGY	**117.7**	***	**48.5**	***	**55.0**	***
HYD1	Water Source	**76.3**	***	**33.0**	***	**46.0**	***
HYD2	Hydroperiod	**83.1**	***	**45.1**	***	**47.0**	***
HYD3	Hydrologic Connectivity	**88.1**	***	**43.7**	***	**46.6**	***
	SOIL	13.1	**	3.8	ns	13.1	**
SOI1	Soil Condition	13.1	**	3.8	ns	13.1	**

understanding ecological integrity. Equally important, the Landscape Context ratings (A–D) had a strong ability to discriminate between different levels of on-site Condition (Fig. 5).

5. Discussion

5.1. Testing the strength of indicators and indices

Our data set from six states spanned large geographic regions of the U.S., included multiple wetland types, and covered a wide range of stressor levels. The data provided a good test of our approach to assess wetland condition within and across states in the country. In general, the metrics performed very well. All metrics had statistically significant discriminatory power among at least two levels of HSI stressor ratings, and most metrics were able to discriminate three or four. Almost all metrics were largely non-redundant with other metrics, apart from two pairs of metrics (B1 – Perimeter with Natural Buffer and B2 – Width of Natural Buffer; V1 – Native Plant Species Cover and V2 – Invasive Nonnative Plant Species Cover). The two buffer metrics were originally developed as part of a buffer index (Collins et al., 2006), here reflected in our buffer MEF, so we recommend that the emphasis for rating buffers be placed on the MEF rather than on the individual metrics. For the two vegetation metrics, we suggest that V1 is a candidate for

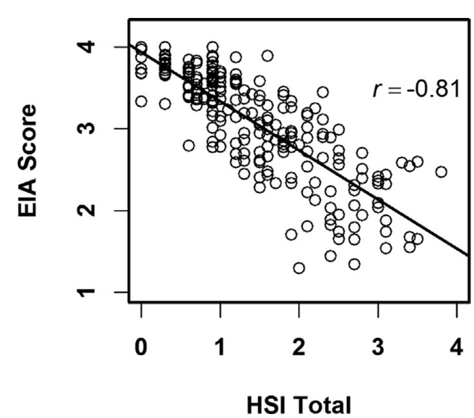

Fig. 3. Correlation of Ecological Integrity Assessment (EIA) scores with Human Stressor Index $(HSI)_{total}$ scores. Ecological integrity scores decline as stressor levels increase.

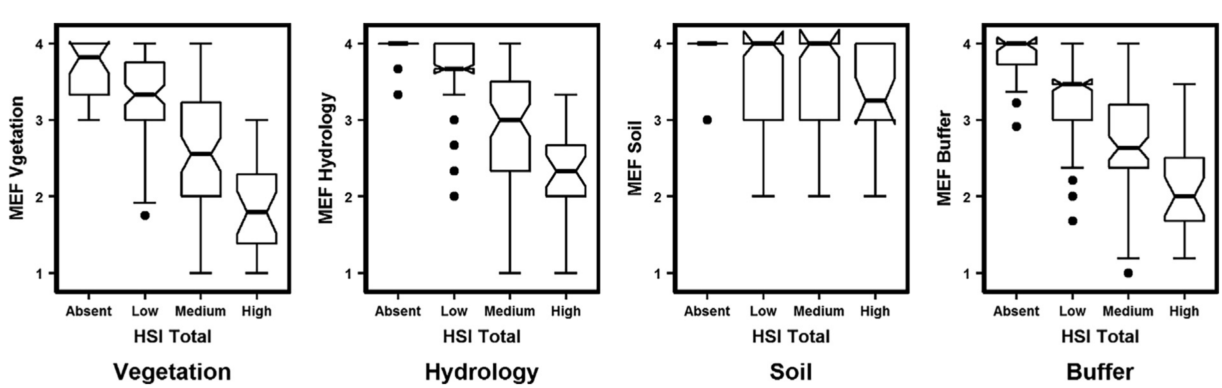

Fig. 2. Response of Major Ecological Factors to stressors. Stressor categories are based on the Human Stressor $Index_{total}$. Using the Hochberg method, all means are significantly different at $p < 0.01$ for vegetation, hydrology, and buffer, whereas for soil, only the Absent category is distinct from other means.

Table 6
Spearman Rank Correlations among metrics. Categories for redundancy were: r ≥ 0.9, very strongly redundant (*VSTR*); r = 0.7–0.89, strongly redundant (STR); r = 0.5–0.69, moderately redundant (Mod); r < 0.5, minimal or no redundancy (m). There were no metrics with "very strongly redundant" values. Metric codes in the columns are simplified for space reasons (i.e., BUF1 = B1, etc.). Full metric names are provided in Table 1.

	B1	B2	B3	H1	H2	H3	L1	L2	S1	V1	V2	V3	V4	V6	
BUF1		**STR**	m	M	m	m	m	**Mod**	m	m	m	m	m	m	BUF1
BUF2			m	M	m	m	**Mod**	**Mod**	m	m	m	m	m	m	BUF2
BUF3				M	m	m	m	**Mod**	m	m	**Mod**	**Mod**	m	m	BUF3
HYD1					**Mod**	**Mod**	m	m	m	m	m	**Mod**	m	m	HYD1
HYD2						**Mod**	m	m	m	m	m	m	m	m	HYD2
HYD3							m	m	m	m	m	**Mod**	m	m	HYD3
LAN1								**Mod**	m	m	m	m	m	m	LAN1
LAN2									m	m	m	m	m	m	LAN2
SOI1										m	m	m	m	m	SOI1
VEG1											**STR**	**Mod**	m	m	VEG1
VEG2												**STR**	**Mod**	m	VEG2
VEG3													**Mod**	m	VEG3
VEG4														m	VEG4
VEG6															VEG6

removal from future EIA applications, though further exploration of their redundancy across the entire range of stressor conditions is needed.

By far the weakest metric and MEF was that for soil condition. Soils are difficult to evaluate in a rapid assessment, as only the most visible evidence of disturbance is rated, such as evidence of compaction, ruts, or filling. Furthermore, these signs are only evident in wetlands where the water draws down during the season. When soil disturbances are visibly evident, most other metrics will also be rated low. Conversely, because soil stressors are not as common or visible in wetlands, most sites rate high, which could tend to inflate the overall EIA. Nonetheless, despite the low DP, the soil metric was one of the least redundant metrics (Table 6). These finding support retaining the metric, but giving it a low weight (0.1) in the scorecard (Fig. 1).

Building on the strength of our metrics, we were also able to develop aggregate indices for MEFs, rank factors, and an overall ecological integrity rating (comparable to the Index of Ecological Integrity of Andreasen et al., 2001) that retained or improved on the strength of our individual metrics. Together they increased our ability to discriminate among stressor levels. These aggregate indices also allowed us to interpret the integrity of MEFs (vegetation, hydrology, and soils), and compare the integrity of on-site condition versus the adjacent buffer and landscape. Our scorecard (Fig. 1) brings this information together in a fairly transparent way, allowing users to understand the status of various components of ecological integrity. This information can be further summarized in dashboards or other types of scorecards (Tierney et al., 2009; Mitchell et al., 2014).

5.2. Vegetation and hydrology indicators

Vegetation indicators are an important aspect of any rapid assessment method. We know a great deal about the ecological behavior of individual species, species assemblages, and vegetation structure, as well as their spatial and historical variability. Indeed, the EPA National Wetland Condition Assessment relied strongly on vegetation, and more specifically vegetation composition, for its Vegetation Multi-Metric Index (US EPA, 2016). As they note, vegetation is a major component of the biodiversity and structure found in wetlands, provides important habitat and food sources for birds, fish, and other wildlife, and both responds to and influences other physical features (e.g., soil, hydrology) and chemical processes (e.g., nutrient cycling) in wetland systems. "Thus, vegetation can reflect and integrate different components of wetland ecosystem integrity and serve as an effective indicator of wetland condition" (US EPA, 2016). Our rapid assessment methods support the use of vegetation as a major component of the overall score, and it is weighted more heavily than other factors in our scorecard (Fig. 1). Still, given the fundamental role of hydrology in wetlands, we find that hydrologic indicators are worth incorporating in the assessment. Indeed, departures from optimal hydrologic conditions may well serve as an important signal that vegetation conditions could decline in the future.

5.3. Use of stressor index to test EIA

Stressors are a key explanatory reason for why ecosystems degrade. Thus, although not necessary in routine applications of rapid assessments, we encourage the collection of stressor data to assist with the interpretation of ecological integrity and as a means of guiding management activities to maintain or improve ecological condition. Still, observed stressors may not account for all the observed levels of ecological integrity. For example, coarse woody debris levels may reflect past logging events that leave few signs of stress today. Similarly, lack of woody regeneration may relate to an altered fire regime that is not always readily assessed. Thus, on-site stressor evaluations may not capture all of the anthropogenic stressors that have affected the ecological integrity of an ecosystem.

In addition, we recognize that our EIA metrics rely, to varying degrees, on stressor information, creating a certain degree of autocorrelation in our testing. Where we were able to assess this (e.g., vegetation metrics against $HSI_{abiotic}$) these effects were not strong. Nonetheless, it will be valuable to search out additional independent measures of stressors and ecological integrity to further refine our model.

5.4. Sampling design

No substantial differences in metric performance were detected with on-site Condition ratings between the point or polygon-based approaches, suggesting that they are largely complementary for on-site evaluations. Further, it suggests that a combination of Level 2 polygon-based sampling can complement a Level 3 intensive point sampling. But for Landscape Context ratings, the point-based approach is less likely to detect the effects of stressors on the wetland from adjacent wetlands or uplands with different conditions, because the buffer is more likely to assess the same wetland type and condition that the point is placed in. Buffers are usually understood to be buffers of a wetland unit, not of a point within a wetland; thus, point-based approaches may need to consider expanding their buffer assessment to include nearby non-wetland areas. This may require some calibration when comparing the two approaches. It may suggest that a larger radius be used for point-based sampling (though a 1 km radius showed no differences compared to the standard 0.5 km radius with the New Jersey data (K. Walz,

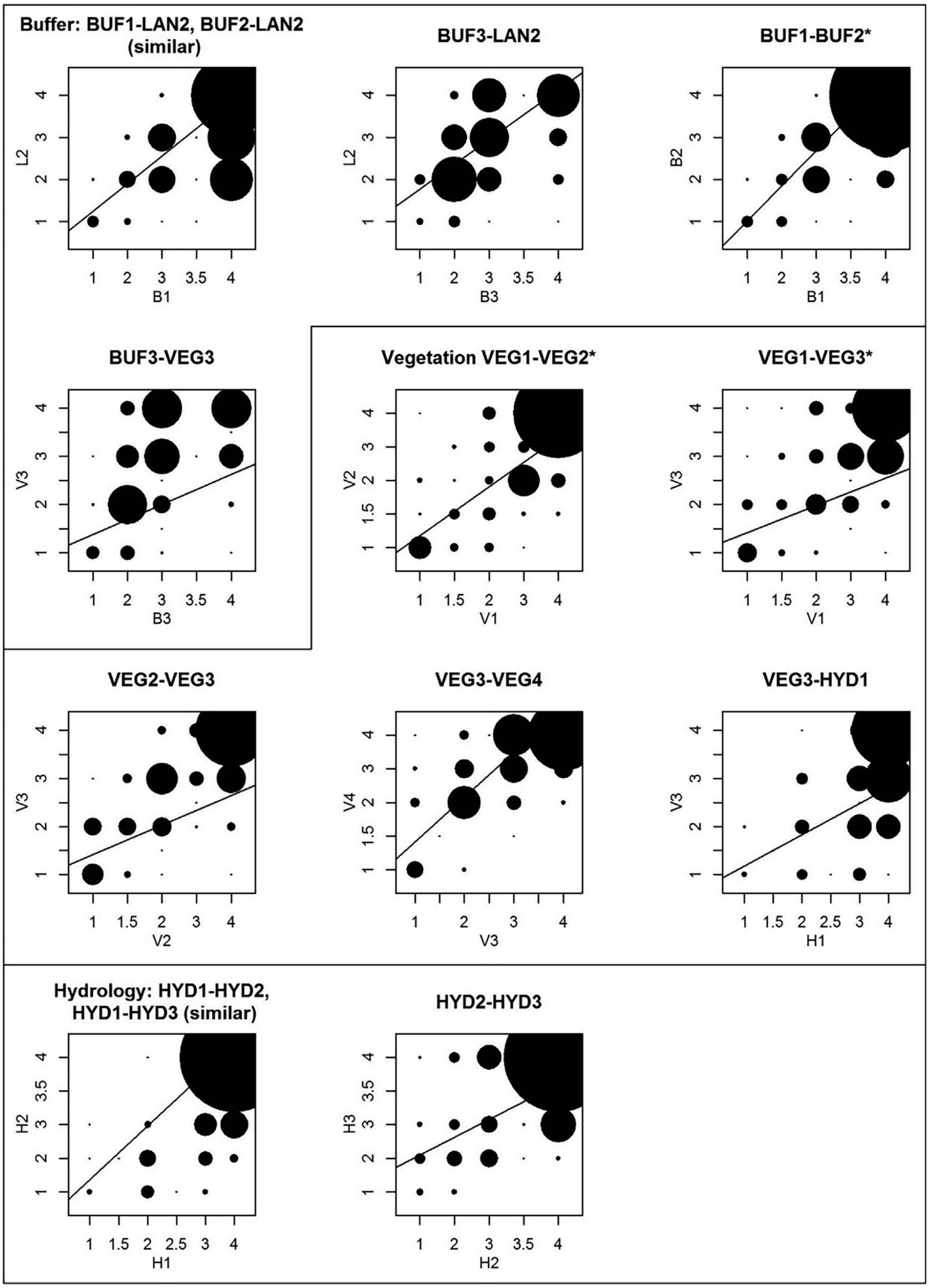

Fig. 4. Scattergrams of metrics with moderate to strong correlations. Size of circles reflects the density of points. Metrics with asterisks had strong correlations (r = 0.7–0.89); other metrics had moderate correlations (0.5–0.69) (see Table 6).

unpublished data).

5.5. Linking landscape context and on-site condition

One of the key findings of our work is that we can meaningfully aggregate or disaggregate metric information, whether that be for specific metrics within an MEF, in comparing MEFs, or exploring the interrelations of Landscape Context and on-site Condition. Our analyses showed that combining metrics in these ways typically increases the DP of our analyses. For example, integrating the three buffer metrics into a Buffer MEF (the Buffer Index of Collins et al., 2006) provides a stronger DP than do the individual metrics.

We are particularly encouraged by how well the Landscape Context ratings discriminated between differences in on-site Condition (Fig. 5). Other studies have found similar results (Brooks et al., 2006; Mack, 2006; Hychka et al., 2007; Mita et al., 2007). These findings validate the use of Landscape Context metrics as a rapid assessment method for identifying the level of ecological integrity found on sites (where field

Table 7
Spearman Rank Correlation (r) of Landscape Context metrics with on-site Condition. Values are reported for individual buffer and landscape metric scores, the combined MEF scores, and overall Landscape Context rank factor scores. Correlations are shown for all sites combined, and for sites by sampling design: polygon-based assessment areas (AAs) and point-based AAs. All correlations are significant at $p < 0.001$.

	CONDITION		
	All Sites (r) 220 sites	Polygon AA (r) 114 sites	Point AA (r) 106 sites
LANDSCAPE CONTEXT	0.63	0.64	0.75
Buffer MEF	0.61	0.64	0.75
BUF1. Perimeter with Natural Buffer	0.33	0.58	0.37
BUF2. Width of Natural Buffer	0.39	0.55	0.52
BUF3. Condition of Natural Buffer	0.61	0.51	0.72
Landscape MEF	0.56	0.56	0.63
LAN1. Contiguous Natural Land Cover	0.48	0.52	0.47
LAN2. Land Use Index	0.54	0.51	0.66

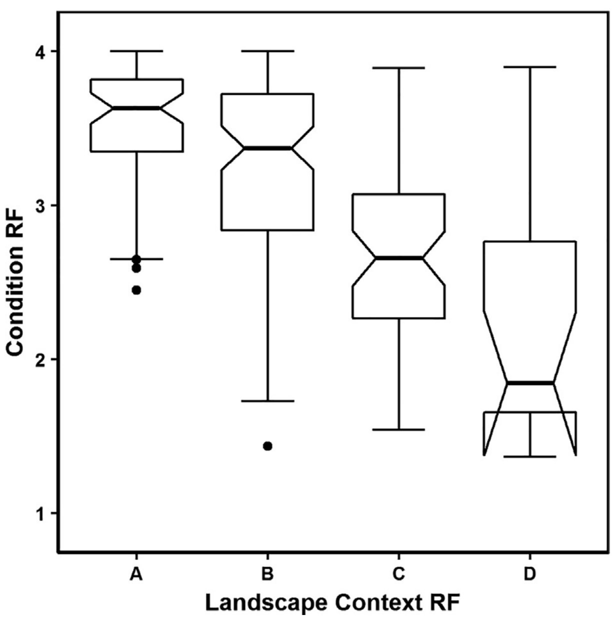

Fig. 5. Response of on-site Condition rank factor scores to Landscape Context rank factor (RF) ratings. All means are significantly different at $p < 0.01$, using the Hochberg method.

crews can check the photo interpreted land use/land cover maps). These site-specific metrics can also be used as part of Level 1 assessments alongside our Landscape Condition Model (LCM), which uses remote-sensing based interpretations of land use/land cover to assess condition (Comer and Faber-Langendoen, 2013; Hak and Comer, 2017). The two methods provide complimentary approaches, where LCM provides a powerful way to characterize the overall condition of landscapes and watersheds, and the Landscape Context metrics characterize the land use/land cover patterns at specific sites. The Landscape Context metrics can also be used as part of 3-level watershed or state-based assessment approach, whereby the Landscape Context metrics are applied to the AAs as part of a Level 1 assessment, generating an initial EIA rating, from which a subset of AAs can be selected for increasingly accurate Level 2 and Level 3 assessments (Stein et al., 2009; Walz and Faber-Langendoen, 2018).

5.6. Ecological integrity assessments for all terrestrial ecosystems

Our findings affirm the merits of our rapid assessment method for wetlands in providing an intermediate level of assessment that is efficient and ecologically meaningful, across a wide range of geographies. In addition, the EIA methodology is applicable to both uplands and wetlands. Working with partners, we have recently developed various applications for uplands, including Level 1 in western upland ecosystems (Comer et al., 2013), Level 2 for all upland ecosystems in Washington (Rocchio et al., 2017), and for southern pine ecosystems (Nordman et al., 2016); as well as contributing to Level 3 monitoring programs for northeastern upland forests (Tierney et al., 2009). Completion of EIAs for both wetlands and uplands will provide the basis for assessing changes over time for all terrestrial ecosystems and in ways that support environmental decision-making.

Acknowledgements

We gratefully acknowledge funding support for our work from the U.S. Environmental Protection Agency (FAIN: 83543901, May 01, 2013–March 29, 2016). We appreciate the support from Chris Faulkner, EPA's project manager, and discussions with Gregg Serenbetz. We also thank EPA for their vision for rigorous standards in wetland assessments, nationally and within states. The data we collected and used for this study would not have been possible without the enthusiasm and dedication of our field crews, and we thank them. We are also grateful for the many discussions with coworkers and colleagues over many years that have helped improve our methodology. We thank Kristin Snow and Mary Harkness for their expertise in designing EcoObs and managing the data for this project. We thank Patrick McIntyre and Michael Lee for their assistance with final production of the figures.

Declarations of interest

None.

References

Andreasen, J.K., O'Neill, R.V., Noss, R., Slosser, N.C., 2001. Considerations for the development of a terrestrial index of ecological integrity. Ecol. Indic. 1, 21–35.
Blocksom, K.A., Kurtenbach, J.P., Klemm, D.J., Fulk, F.A., Cormier, S.M., 2002. Development and evaluation of the lake macroinvertebrate integrity index (LMII) for New Jersey lakes and reservoirs. Environ. Monit. Assess. 77, 311–333.
Brinson, M.M., 1993. A Hydrogeomorphic Classification for Wetlands. U.S. Army Corps of Engineers Technical Report WRP-DE-4, Washington DC.
Brooks, R.P., Wardrop, D.H., Cole, C.A., 2006. Inventorying and monitoring wetland condition and restoration potential on a watershed basis with examples from Spring Creek watershed, Pennsylvania, USA. Environ. Manage. 38, 673–687.
Collins, J.N., Stein, E.D., Sutula, M., Clark, R., Fetscher, A.E., Grenier, L., Grosso, C., Wiskind, A., 2006. California Rapid Assessment Method (CRAM) for Wetlands and Riparian Areas. Version 4.5. San Francisco Estuary Institute, Oakland, CA.
Comer, P., Faber-Langendoen, D., 2013. Assessing ecological integrity of wetlands from national to local scales: Exploring the predictive power, and limitations, of spatial models. Natl. Wetlands Newslett. 35 (3), 20–22.
Comer, P.J., Crist, P.J., Reid, M.S., Hak, J., Hamilton, H., Braun, D., Kittel, G., Varley, I., Unnasch, B., Auer, S., Creutzburg, M., Theobald, D., Kutner, L., 2013. A Rapid Ecoregional Assessment of the Central Basin and Range Ecoregion. Report, appendices, and databases provided to the Bureau of Land Management. NatureServe, Washington DC.
Dahl, T.E., 2011. In: Status and Trends of Wetlands in the Conterminous United States 2004–2009. U.S. Department of the Interior; Fish and Wildlife Service, Washington DC, pp. 108.
Davies, S.P., Jackson, S.J., 2006. The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. Ecol. Appl. 16, 1251–1266.
DeBerry, D.A., Chamberlain, S.J., Matthews, J.W., 2015. Trends in Floristic Quality Assessment for wetland evaluation. Wetland Sci. Pract. 2015, 12–22.
ESA Panel (Ecological Society of American Vegetation Classification Panel), 2015. How a national vegetation classification can help ecological research and management. Front. Ecol. Environ. 13, 185–186.
Faber-Langendoen, D., Kudray, G., Nordman, C., Sneddon, L., Vance, L., Byers, E., Rocchio, J., Gawler, S., Kittel, G., Menard, S., Comer, P., Muldavin, E., Schafale, M., Foti, T., Josse, C., Christy, J., 2008. Ecological Performance Standards for Wetland Mitigation based on Ecological Integrity Assessments. + Appendices. NatureServe, Arlington, VA.

Faber-Langendoen, D., Hedge, C., Kost, M., Thomas, S., Smart, L., Smyth, R., Drake, J., Menard, S., 2012. Assessment of Wetland Ecosystem Condition Across Landscape Regions: A Multi-metric Approach. Part A. Ecological Integrity Assessment Overview and Field Study in Michigan and Indiana. EPA/600/R-12/021a. U.S. Environmental Protection Agency Office of Research and Development, Washington DC.
Faber-Langendoen, D., Nichols, W., Rocchio, J., Walz, K., Lemly, J., Smyth, R., Snow, K., 2016a. Rating the condition of reference wetlands across states: NatureServe's Ecological Integrity Assessment method. Natl. Wetlands Newslett. 38 (3), 12–16.
Faber-Langendoen, D., Nichols, W., Walz, K., Rocchio, J., Lemly, J., Gilligan, L., 2016b. NatureServe Ecological Integrity Assessment: Protocols for Rapid Field Assessment of Wetlands. v2.0. NatureServe, Arlington, VA.
Faber-Langendoen, D., Nichols, W., Rocchio, J., Walz, K., Lemly, J., 2016c. In: An Introduction to NatureServe's Ecological Integrity Assessment Method. NatureServe, Arlington, VA, pp. 33.
Fennessy, M.S., Jacobs, A.D., Kentula, M.E., 2007. An evaluation of rapid methods for assessing the ecological condition of wetlands. Wetlands 27, 543–560.
Fulton, E.A., Smith, A.D.M., Punt, A.E., 2005. Which ecological indicators can robustly detect effects of fishing? ICES J. Mar. Sci. 62, 540–551.
Goodin, K.L., Faber-Langendoen, D., Brenner, J., Allen, S.T., Day, R.H., Congdon, V.M., Shepard, C., Cummings, K.E., Stagg, C.L., Gabler, C.A., Osland, M., Dunton, K.H., Ruzicka, R.R., Semon-Lunz, K., Reed, D., Love, M., 2018. In: Ecological Resilience Indicators for Five Northern Gulf of Mexico Ecosystems. NatureServe, Arlington, VA, pp. 381.
Hak, J.C., Comer, P.J., 2017. Modeling landscape condition for biodiversity assessment: Application in temperate North America. Ecol. Indic. 82, 206–216.
Hychka, K.C., Wardrop, D.H., Brooks, R.P., 2007. Enhancing a landscape assessment with intensive data: A case study in the upper Juniata watershed. Wetlands 27, 446–461.
Jacobs, A., McLaughlin, E., Havens, K., 2010. Mid-Atlantic Tidal Wetland Rapid Assessment Method Version 3.0. June 2010. Delaware Department of Natural Resources and Environmental Control, Dover, DE.
Karr, J.R., Chu, E.W., 1999. In: Restoring Life in Running Waters: Better Biological Monitoring. Island Press, Washington DC, pp. 206.
Klemm, D.J., Blocksom, K.A., Falk, F.A., Herlihy, A.T., Hughes, R.M., Kaufmann, P.R., Peck, D.V., Stoddard, J.L., Thoeny, W.T., Griffith, M.B., Davis, W.S., 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing mid-Atlantic highlands streams. Environ. Manage. 31, 656–669.
Lemly, J., Rocchio, J., 2009. Field Testing of the Subalpine-Montane Riparian Shrublands Ecological Integrity Assessment (EIA) in the Blue River Watershed, Colorado. Unpublished report prepared for U.S. Environmental Protection Agency, Region 8, Denver, CO. Colorado Natural Heritage Program, Colorado State University, Fort Collins, CO.
Lemly, J., Gilligan, L., 2015. Ecological Integrity Assessment (EIA) for Colorado Wetlands Field Manual, Version 2.0. Colorado Natural Heritage Program, Colorado State University, Fort Collins, CO.
Leopold, A., 1966. A Sand County Almanac with Essays on Conservation from Round River. Ballantine Books, NY.
Mack, J.J., 2001. Ohio Rapid Assessment Method for Wetlands, Manual for Using Version 5.0. Ohio EPA Technical Bulletin Wetland 2001-1-1. Ohio Environmental Protection Agency, Division of Surface Water, 401 Wetland Ecology Unit, Columbus, OH.
Mack, J.J., 2006. Landscape as a predictor of wetland condition: An evaluation of the Landscape Development Index (LDI) with a large reference wetland dataset from Ohio. Environ. Monit. Assess. 120, 221–241.
Master, L.L., Faber-Langendoen, D., Bittman, R., Hammerson, G.A., Heidel, B., Ramsay, L., Snow, K., Teucher, A., Tomaino, A., 2012. NatureServe Conservation Status Assessments: Factors for Evaluating Species and Ecosystem Risk. NatureServe, Arlington, VA.
Mita, D., DeKeyser, E., Kirby, D., Easson, G., 2007. Developing a wetland condition prediction model using landscape structure variability. Wetlands 27, 1124–1133.
Mitchell, B.R., Tierney, G.L., Schweiger, E.W., Miller, K.M., Faber-Langendoen, D., Grace, J.B., 2014. Getting the message across: Using ecological integrity to communicate with resource managers. Chapter 10, 327 pp In: Guntenspergen, G.R. (Ed.), Application of Threshold Concepts in Natural Resource Decision Making. Springer, pp. 199–230.
National Research Council, 1994. In: Rangeland Health: New Methods to Classify, Inventory, and Monitor Rangelands. National Academy Press, Washington DC, pp. 180.
NatureServe, 2002. Element Occurrence Data Standard. Available from: http://www.natureserve.org/prodServices/eodata.jsp.
Nichols, W.F., 2013. Comparison of Alternative Wetland Assessment Methods at Numerous Sites in New Hampshire. NH Natural Heritage Bureau, Concord, NH.
Nichols, W.F., Faber-Langendoen, D., 2012. Level 2 Ecological Integrity Assessment Manual: Wetland Systems. New Hampshire Natural Heritage Bureau & NatureServe, Concord, NH + Appendices.
Nordman, C., White, R., Wilson, R., Ware, C., Rideout, C., Pyne, M., Hunter, C., 2016. Rapid Assessment Metrics to Enhance Wildlife Habitat and Biodiversity within Southern Open Pine Ecosystems, Version 1.0. U.S. Fish and Wildlife Service and NatureServe, for the Gulf Coastal Plains and Ozarks Landscape Conservation Cooperative.
Oliver, I., Jones, H., Schmoldt, D.L., 2007. Expert panel assessment of attributes for natural variability benchmarks for biodiversity. Austral Ecol. 32, 453–475.
Parkes, D., Newell, G., Cheal, D., 2003. Assessing the quality of native vegetation: The 'habitat hectares' approach. Ecol. Manage. Restor. 4 (Supplement), S29–S38.
Peet, R.K., Wentworth, T.R., White, P.S., 1998. The North Carolina vegetation survey protocol: A flexible, multipurpose method for recording vegetation composition and structure. Castanea 63, 262–274.
R Core Team, 2015. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Regan, T.J., Master, L.L., Hammerson, G.A., 2004. Capturing expert knowledge for threatened species assessments: A case study using NatureServe conservation status ranks. Acta Oecol. 26, 95–107.
Rocchio, F.J., Crawford, R.C., Ramm-Granberg, T.C., 2016. Field Manual for Applying Rapid Ecological Integrity Assessments in Wetlands and Riparian Areas. Washington Natural Heritage Program, Department of Natural Resources, Olympia, WA NHR-2016-01.
Rocchio, F.J., Ramm-Granberg, T., Crawford, R.C., 2017. Field Manual for Applying Rapid Ecological Integrity Assessments in Upland Plant Communities of Washington State. Washington Natural Heritage Program, Washington Department of Natural Resources, Olympia, WA 98504-7014.
Rolston III, H., 1993. Ecosystem integrity and health values. In: Rolston, H. (Ed.), Conserving Natural Value. Columbia University Press, NY, pp. 68–100.
Stein, E.D., Fetscher, A.E., Clark, R.P., Wiskind, A., Grenier, J.L., Sutula, M., Collins, J.N., Grosso, C., 2009. Validation of a wetland rapid assessment method: Use of EPA's Level 1-2-3 framework for method testing and refinement wetlands. Wetlands 29, 648–665.
Stevens Jr., D.L., Olsen, A.R., 2004. Spatially-balanced sampling of natural resources in the presence of frame imperfections. J. Am. Stat. Assoc. 99, 262–278.
Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K., Norris, R.H., 2006. Setting expectations for the ecological condition of streams: The concept of reference condition. Ecol. Appl. 16, 1267–1276.
Sutula, M.A., Stein, E.D., Collins, J.N., Fetscher, A.E., Clark, R., 2006. A practical guide for development of a wetland assessment method: The California experience. J. Am. Water Resour. Assoc. 42, 157–175.
Tierney, G.L., Faber-Langendoen, D., Mitchell, B.R., Shriver, G., Gibbs, J., 2009. Monitoring and evaluating the ecological integrity of forest ecosystems. Front. Ecol. Environ. 7, 308–316.
Unnasch, R.S., Braun, D.P., Comer, P.J., Eckert, G.E., 2009. The Ecological Integrity Assessment Framework: A Framework for Assessing the Ecological Integrity of Biological and Ecological Resources of the National Park System. Report to the National Park Service.
U.S. Environmental Protection Agency (US EPA), 2002. Methods for Evaluating Wetland Condition: Using Algae to Assess Environmental Conditions in Wetlands. EPA-822-R-02-021. US EPA, Washington DC.
U.S. Environmental Protection Agency (US EPA), 2016. National Wetland Condition Assessment: 2011 Technical Report. U.S. Environmental Protection Agency, Washington DC EPA-843-R-15-006.
Walz, K.S., Domber, S., 2011. Developing a Wetland Condition Monitoring Network for New Jersey: Application of New Assessment Methods. New Jersey Department of Environmental Protection, Trenton, NJ.
Walz, K.S., Faber-Langendoen, D., 2018. In fulfillment of Wetlands Protection Development Grant Section 104(B)(3); CFDA No. 66.461, CD97225809 In: New Jersey's Wetland Condition Assessment Intensification Study: A Multi-Tiered Assessment of Wetlands and Watersheds. Summary Report. New Jersey Department of Environmental Protection, Division of Parks and Forestry, Office of Natural Lands Management, Natural Heritage Program, Trenton, NJ, pp. 100.
Wang, X., Zheng, B., Liu, L., Wang, L., 2015. Development and evaluation of the lake multi-biotic integrity index for Dongting Lake, China. J. Limnol. 74, 594–605.
Wardrop, D.H., Kentula, M.E., Stevens, D.L., Jensen, S.F., Brooks, R.P., 2007. Assessment of wetland condition: An example from the Upper Juniata watershed in PA, USA. Wetlands 27, 416–431.
Wardrop, D.H., Kentula, M.E., Brooks, R.P., Fennessy, M.S., Chamberlain, S.J., Havens, K.J., Hershner, C., 2013. Monitoring and assessment of wetlands: Concepts, case studies, and lessons learned. In: Brooks, R.P., Wardrop, D.H. (Eds.), Mid-Atlantic Freshwater Wetlands: Advances in Wetlands Science, Management, Policy, and Practice. Springer, NY, pp. 318–419.
Woodley, S., Kay, J., Francis, G., 1993. In: Ecological Integrity and the Management of Ecosystems. St. Lucie Press, pp. 220.
Wurtzebach, Z., Schultz, C., 2016. Measuring ecological integrity: History, practical applications, and research opportunities. Bioscience 66, 446–457.