esa ECOSPHERE

METHODS, TOOLS, AND TECHNOLOGIES: SYNTHESIS & INTEGRATION

Floristic Quality Assessment: a critique, a defense, and a primer

GREG SPYREAS†

Illinois Natural History Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820 USA

Citation: Spyreas, G. 2019. Floristic Quality Assessment: a critique, a defense, and a primer. Ecosphere 10(8):e02825. 10.1002/ecs2.2825

Abstract. Floristic Quality Assessment (FQA) measures have become extraordinarily influential ecological metrics in North America over the past 20 yr. Government agencies, conservation organizations, land managers, and researchers alike utilize this plant-based measure to evaluate habitat conservation value, ecological integrity, and naturalness. Its relative uniqueness, utility, and ease of use, among vegetation measures, portend the continued popularity of FQA going forward. FQA's use and influence far exceeds its study—where the literature addressing questions and criticisms regarding its methodology and ecological meaning has not kept pace with reliance upon it. Furthermore, the lack of literature review has led to disorder and confusion among its users. This review addresses these issues in three parts. First, it concisely explains the metrics and their methods, and most importantly, it synthesizes the often-misinterpreted conceptual basis behind FQA. The bulk of the review then tackles common questions from researchers and non-technical users alike regarding the measures. It does this with two lists. The first list reviews FQA's most common criticisms and summarizes evidence for and against them. The second list confronts the most common mistakes surrounding FQA, regarding both its application and misunderstanding in the literature. In each instance, straightforward guidelines and answers to uncertainties are emphasized.

Key words: anthropogenic disturbance; biological integrity; coefficient of conservatism; conservation value; ecological indicator; Floristic Quality Index (FQI); habitat assessment; human impact; Mean *C*; natural area assessment; natural area quality.

Received 12 February 2019; revised 19 June 2019; accepted 20 June 2019. Corresponding Editor: Debra P. C. Peters.

† **E-mail:** spyreas@illinois.edu

INTRODUCTION

A fundamental job of conservation practitioners and land managers is to quickly assess the relative biological value of land parcels and their species. They must evaluate the naturalness, ecological integrity, and conservation value of so much land, so frequently, that assessments can become rote—seemingly occurring unconsciously. However, the outcome of this initial site assessment, often the first step of conservation work, is critical as it dictates which lands are prioritized (e.g., for purchase, protection, restoration), and it identifies target communities for restoration. Repeatable, quantitative, measures—ecological indicators—are valuable to guide and standardize such work. While they are common today, the scientific development of ecological indicators has only occurred recently (Niemi and McDonald 2004). Therefore, the best choice for site- or community-level measures is rarely clear, and weaknesses in available measures impede their use.

To generalize, two kinds of vegetation-based habitat assessments are used for this type of work, judgment-based and metric-based. Professional judgment is fast and it takes advantage of accumulated field experience and knowledge, but it may not seem objective, repeatable, or precise. Alternatively, quantitative metrics can be repeatable and objective, but they too have flaws. For example, simple, familiar numbers like species richness are insufficient for quantifying more complex vegetation properties of interest. The properties responsible for a natural area's uniqueness and value are often immediately apparent to on the ground observers, but they are not captured by diversity measures (Smith and Theberge 1986, Mace 2005). More sophisticated ecological measures, such as those measuring endemism, rarity, regional uniqueness, taxonomic distinctness, specialization, intactness, completeness, uniquity, or more complex types of diversity (Izco 1998, Ricotta 2004, Scholes and Biggs 2005, Hawkins 2006, Devictor et al. 2008, Chapman et al. 2009, Filippi-Codaccioni et al. 2010, Ejrnæs et al. 2018), may give more nuanced and insightful results. But, these often require large datasets and/or complicated calculations, which few practitioners have the time for or expertise to use. Beyond unwieldy calculations and data, some of these can also seem overly derived, artificial, or synthetic when applied to characterizing a natural area.

It is in this context that Floristic Quality Assessment (FQA)-based habitat measures have successfully emerged as a bridge between judgment-based and quantitative assessment. Floristic Quality Assessment is widely used and is highly influential, especially in North America. It is increasingly relied upon to guide ecological policy, regulation, conservation, management, and research. However, its use, study, and philosophical basis have never been simultaneously reviewed. This has led to a disordered FQA literature, which has led to theoretical misunderstanding, methodological errors, and frequent misapplication. Justified skepticism and criticism of FQA has followed. An overview of FQA is long overdue. This review will characterize the relevant literature, research, and application of FQA for its users, with particular a focus on misunderstandings, misuses, and criticisms of it.

The Metrics

Floristic Quality Assessment uses the plant species found in an area to measure its floristic quality. It is calculated using one or both of FQA's primary metrics—the mean coefficient of conservatism (Mean *C*) or the Floristic Quality Index (FQI).

$$\bar{C} = \sum C/(S)$$

$$\mathrm{FQI} = C \times \left(\sqrt{S}\right)$$

where Mean *C* is the average conservatism value (*C* value) of the plant species in an area, and *S* is the number of plant species. Floristic Quality Index is the product of Mean *C* and the square root of the area's plant richness. These are the basic, standard FQA measures (Swink and Wilhelm 1994). By using them, only two pieces of information are needed to calculate an area's FQA value: a plant species list and the published *C* values for those species. Regional botanical experts pre-assign each plant species a *C* value. The value ranges from 0 to 10 for each plant in a region (often a state, which is used henceforth). Species values are only calibrated for, and valid in, the state they are assigned (Fig. 1). Native species may be assigned any value from 0 to 10, whereas non-native species receive zeros. Note that non-native species may or may not be included in either equation, with similar performance either way (Spyreas et al. 2012). Of further note is that many variant FQA measures have been created, but these show either inconsistent, or decreased performance, compared to the standard measures (reviewed in Spyreas 2014, Kutcher and Forrester 2018).

When sampling an area's floristic quality, either the flora of the entire habitat patch or plots within it (site sub-samples) can be used. If the FQI is used, the size of the sample area will influence values because species richness generally increases with the size of the sample area (Matthews et al. 2005). Mean *C*, however, is more immune to biases related to sample area, at least at plot, community, or site scales (Fig. 2; Spyreas 2016).

The Concept

There are two related concepts that underlie FQA and are critical to understanding it: a species' conservatism and a site's floristic quality. Species conservatism—estimated via *C* values—is assigned based on a species' fidelity to habitats that are more, or less, degraded by human disturbances (Fig. 3). High values are assigned to plants

Fig. 1. Species conservatism values can vary among states. Red pine (*Pinus resinosa*; left) is highly conservative in Illinois (C = 9), but not in neighboring Michigan (C = 3). Ebony spleenwort (*Asplenium platyneuron*; right) is highly conservative in Florida (C = 10), but not in Illinois (C = 4). Differences in regional ecologies are only one reason that a species' value may differ. The individual scoring process for each state's flora affects species values. It is crucial to remember that because C values were created as relative comparisons to other species and communities in that state, they are only intended for use there.

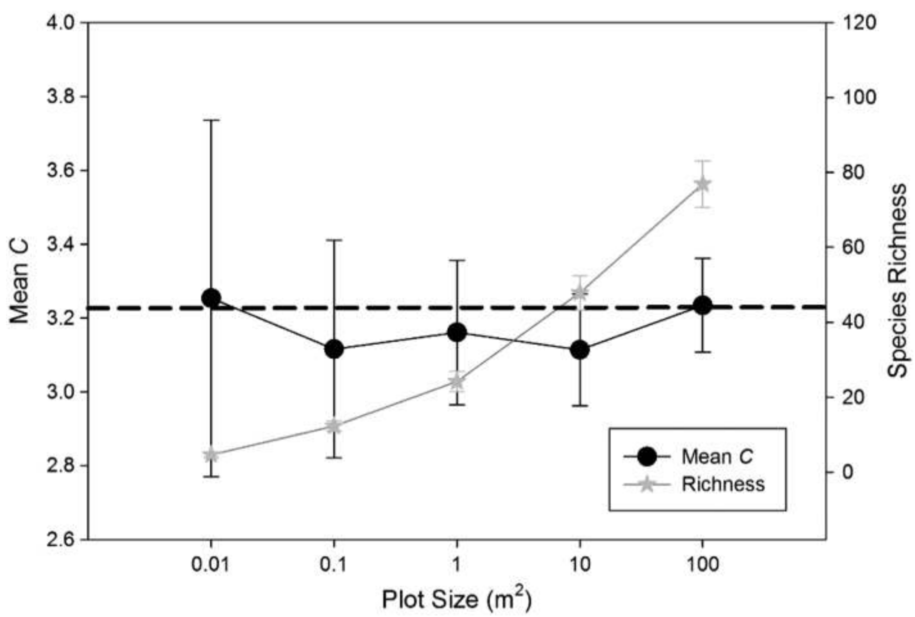

Fig. 2. The relationship between area and Mean C. Unlike species richness (gray line), Mean C across different sized, nested, grassland plots shows it to be largely unrelated to the size of the area sampled. Although it does become less precise in very small plots with few species (larger confidence intervals), dashed horizontal line is the 100-m^2 plot value for reference. ±95% CI; N = 20. Figure based on Spyreas (2016).

that are exclusive to undegraded, relictual, native habitats in the region (Fig. 4). Users can assume that such species are more sensitive to anthropogenic disturbances (Swink and Wilhelm 1994). The conservatism concept assumes that anthropogenic disturbances are generally of a type and intensity unlike that of natural disturbance regimes in their effects on native plant communities. While many other species traits or ecological characteristics—rarity, successional stage, *r* vs. *K* syndromes, niche breadth—are loosely correlated with species conservatism, none of these completely characterizes species values, and it is important to note that they do not define them (see *You're doing it wrong* below).

The second fundamental concept is an area's floristic quality (measured by Mean *C* or FQI). Sitewide floristic quality is traditionally used for two related purposes. The first is to ask, "How degraded by human disturbance is a site overall?" By definition, the presence/absence of disturbance-sensitive species provides an accurate measure of the disturbance and stress an area has incurred. The second more common use for site floristic quality stems from its relatively unique ability to quantify an area's conservation value. Because pristine, native plant communities are so rare in most modern landscapes, remnant habitats that harbor conservative plant assemblages are special and valuable for conservation purposes (often along with their associated fauna, Panzer and Schwartz 1998). This is the basic definition of floristic quality.

Use of FQA

Floristic Quality Assessment was originally developed in the late 1970s in the Chicago region (USA) to identify protection-worthy lands with a simple, repeatable, quantitative method (Swink

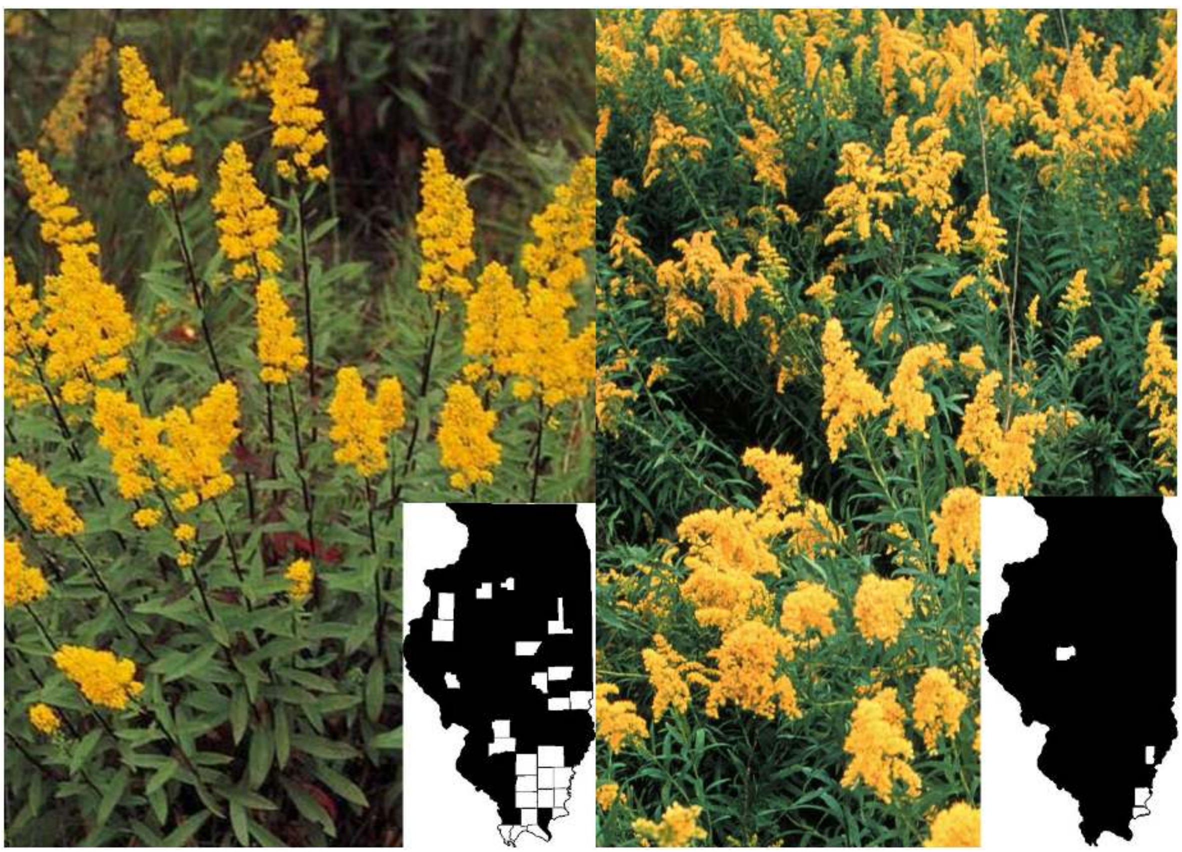

Fig. 3. Similar species, dissimilar conservatism values. Closely related species often vary in their *C* values. For example, *Solidago speciosa* (left) and *S. canadensis* (right) are widespread across Illinois (dark counties map inset), and they are closely related. However, *S. speciosa* is restricted to relatively pristine native habitats (C = 7), whereas *S. canadensis* can be found in almost any upland situation (C = 0). Phylogeny (relatedness), rarity, showiness, and other species traits are often incorrectly associated with species conservatism.

Fig. 4. What does floristic quality look like? Low floristic quality habitats may be beautiful, diverse, and valuable for ecosystem services and wildlife. However, the conservation value of their floras is not special. Two sets of habitats are pictured. The first wetland (A) had been used for row crop agriculture for 100 yr prior to being left fallow for two years before being photographed. The plants in view are not conservative and could be found in nearly any wet area in the region, disturbed or not (Mean C = 2.2). The second wetland (B) is a high-quality remnant wetland. Plants in view are indicative of an undegraded native habitat (Mean C = 6). The first upland (C) is an abandoned field. Its plants are not conservative and could be found in any upland field in the region, ruderal or not (Mean C = 1.5). The second upland (D) is a centuries-old assemblage of conservative tallgrass prairie plants that have never been plowed (Mean C = 6.5).

and Wilhelm 1979). It has since expanded geographically, and most of North America now has C values assigned to its flora (Fig. 5), as do areas of at least three other continents (Zinnen et al., *unpublished manuscript*). Its literature has grown exponentially, especially over the past 20 yr (Zinnen et al., *unpublished manuscript*). The breadth of application of FQA has also expanded; it is now used for basic ecological research, as well as many different types of habitat assessment. This includes habitat monitoring over time, assessing restoration success, and guiding ecological management practices and techniques. It is also a legally codified regulatory criterion used by municipal, county, state, and federal agencies (Matthews and Endress 2008, Chu and Molano-Flores 2013).

There are five primary reasons why FQA has become so popular (reviewed in Spyreas 2014): (1) its simplicity and ease of use. Only a plant species list and C values are needed; (2) its ability to quantify and distill years of botanical knowledge and field experience into an objectively applied number; (3) its flexibility and variety of uses; (4) its effectiveness; it accurately measures what it purports to; and (5) its uniqueness; it measures ecological information that is not captured by other ecological metrics.

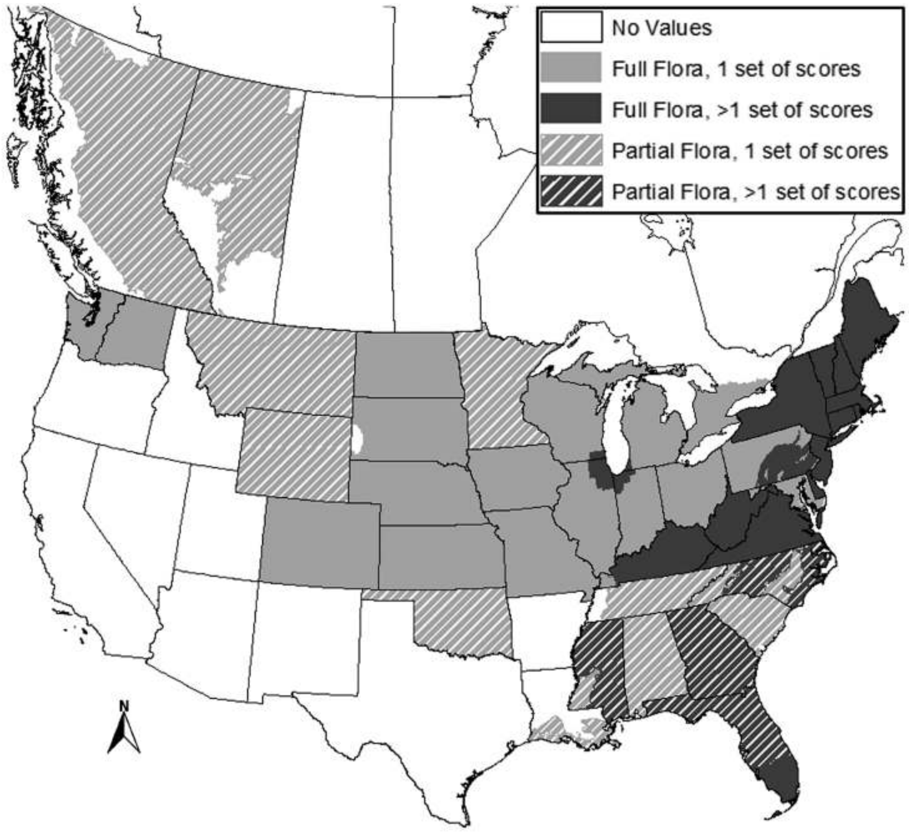

Fig. 5. North American regions with *C* values assigned to their flora. For several states/regions, there are multiple versions of *C* values available (dark gray). While this may make it confusing to know which version to use, different versions result in similar relative results. Stripes indicate regions where only parts of the flora have been assigned values (e.g., wetland plants). There are ongoing efforts to assign values to the floras of the remaining sections of the western USA. Most *C* value lists can be found at the Universal Floristic Quality Assessment website, (https://universalfqa.org/view_databases) or by contacting the author directly.

Everyone's a Critic

With FQA's popularity has come skepticism. It has been criticized as imprecise, inconsistent, biased, subjective, romantic, tautological, untested, and unsubstantiated by ecological theory. This section addresses the more common and trenchant criticisms, moving from the conceptually focused to the more methodological. A synopsis of the 10 criticisms (italics) is followed by a response.

Circularity

A species' conservatism value is based on its propensity to occur in high-quality habitats. However, these nice habitats are identified by their floras in the first place. This tautology to C values highlights how FQA does not measure an independent ecological property; it essentially measures itself.

Some circularity or non-independence is inherent in species *C* values; it is also not necessarily problematic. Part of assessing the natural quality or biological degradation of a habitat is observing the community of plants in it. As long as FQA users (1) understand that to some extent, the observed species co-occurrences in high-quality habitats underlie how species *C* values were assigned, and (2) agree that the native remnant habitats identified as high quality (where conservative species co-occur) are correctly identified as high quality; then, circularity is not problematic. Indeed, one desirable aspect of FQA is that it provides a very simple way to measure the interesting propensity for conservative plants to occur with each other, which is an ecological phenomenon that is otherwise not easily measured (Matthews et al. 2015).

To the second point, few would argue with the ability of experienced botanists to identify exceptional, undegraded native habitats. But, habitat properties besides the plant community influence a botanist's identification of high-quality natural areas. Observed disturbance legacies from pollution, farming, grazing, logging, alteration of hydrology, and other site factors factor into how expert botanists have judged habitats they have encountered over the years, and therefore, they inform species *C* values. Not surprisingly, studies have found that FQA measures correlate very highly with measured site disturbance variables and indices (Fig. 6). Indeed, these measured anthropogenic disturbance levels could be seen as the independent and objective property that FQA captures.

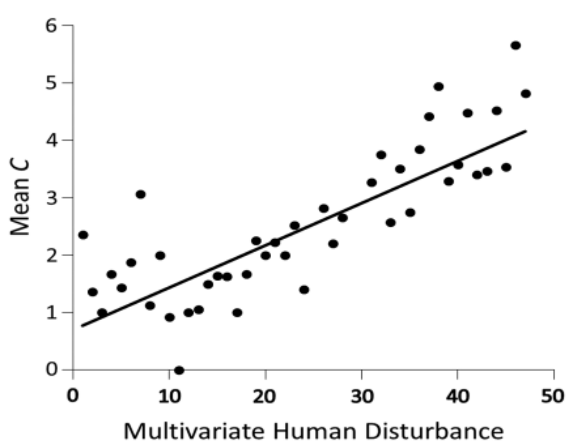

Fig. 6. How well do Floristic Quality Assessment (FQA) measures perform? Site Mean *C* vs. anthropogenic disturbance across 47 emergent wetlands in Illinois (r^2 = 0.61). The *x*-axis was created by combining dozens of human disturbance measures into a disturbance index. When compared with other measures, FQA metrics are almost always found to be the best predictor of site disturbance, typically capturing between 60% and 90% of the variation. Note that predictability declines at the low (left) and high (right) end of the spectrum, either suggesting weak Mean *C* performance there (the highest and lowest quality sites are difficult to predict) or failure to quantify disturbance levels there. Figure adapted from Spyreas (2014).

It works in practice, but not in theory

Floristic Quality Assessment is invalid because it is not based in, or substantiated by, ecological theory.

Many authors have recently tried to move beyond the standard, basic, definition for floristic quality or species conservatism by inserting ecological theories or properties into their definitions (e.g., succession, rarity, niche breadth, specialization, Grime's CSR). Such theory-based floristic quality definitions have proliferated, but they suffer from two problems. First, they lack empirical support or evidence, whereas the standard, non-theoretical conceptual grounding in anthropogenic disturbance accurately predicts FQA values (Mack 2007; Fig. 6).

The second problem with new FQA explanations is that they are incomplete. Their theories do not fit, or explain, every species' *C* value. (For example, some high *C* value species in the flora will have a wide niche breadth or they will be early-successional. But, such inconsistent species are never explained.) To bolster the weakness of any one theory, some authors have incorporated multiple theories into new FQA definitions, where FQA is attributable to some indeterminate combination of them. Some have even used different theories for the different integer levels. For example, species with values from 0 to 3 relate to rarity, whereas values from 8 to 10 are explained by niche breadth, but not rarity. These frameworks have never explained why one theory would relate to some *C* values and species behaviors, but not others. Alternatively, within the standard definition framework, every *C* value, for every species, can be explained by remnant habitat dependence and disturbance tolerance.

Unfortunately, once published, new definitions have proliferated. This has muddled the literature and confused FQA users. There is little doubt that existing ecological theories, concepts, and properties will eventually prove to useful for expanding our understanding of why species are more, or less, conservative (Bauer et al. 2018), but their ability to define FQA is currently speculative and incomplete. One hundred and fifty years of the study of ecology shows that no theory will explain the behavior of thousands of plant species across untold habitat types (Lawton 1999). Whether FQA's originators were heeding this lesson, or whether it was because FQA was foremost intended to be practical, they explicitly avoided associating it with existing ecological theory. Floristic quality "... is clearly a human concept and not a true ecosystem property" (Bourdaghs et al. 2006), albeit one that characterizes the ecology and behavior of species occurrences in natural vs. anthropogenic habitats very well. Thus, the simple standard conceptual framework remains the soundest approach to using and understanding it.

Subjectivity

Plant species C values are subjectively assigned, yet FQA is presented as an objective measure. It is completely opinion-based and it is therefore not scientific.

C values are subjectively assigned. However, they are based on botanists' experience with where plants occur and what types of habitats they occur in, which is accumulated over decades of study. Subjectivity in any professional judgment is a problem when it leads to inaccuracy or bias. Biases exist in even the best-trained professionals. However, bias has yet to be shown for FQA (Matthews et al. 2015). Inaccuracies (or biases) that do occur in *C* values can be mitigated at two levels. First, during the original *C* value assignment process other botanists on the panel moderate any one person's biases. Second, the collective species values at a site are aggregated by FQA metrics, which moderates the bias contained in any individual species value (Matthews 2003).

The bias of rarity

Rare species in particular are biased toward higher C value assignments.

The earliest versions of FQA actually did incorporate rarity directly into species *C* values, where designated threatened or endangered species automatically received the highest values (Swink and Wilhelm 1979, Wilhelm and Ladd 1988, Ladd 1993). However, those values were subsequently replaced and modern evaluations consciously eschew considerations of rarity. It is true, however, that *C* values and species rarity are highly correlated (Fig. 7). In modern highly disturbed landscapes, this general trend is to be expected, but there are also many examples of non-conservative species that are quite rare, and vice versa. Indeed, conservative species can either be common, naturally rare (e.g., edge of their range), or rare because their habitat has been destroyed.

Beyond individual species, some have suggested an FQA bias toward particular habitat types, especially rarer habitats. For example, pristine fens, a naturally rare wetland type, may achieve higher FQA scores than pristine marshes, a rather common wetland community (Bried et al. 2013). Assumedly this happens because the *C* values of species found in these unique communities were overvalued by scorer bias. However, it is also quite possible that these habitat types simply have different relative disturbance levels (e.g., different disturbance legacies that are difficult to detect). Future tests will need to determine whether certain vegetation types, or even certain regions within states, have higher (biased) floristic quality scores because of higher *C* values assigned to their species, or whether these habitats are simply less degraded.

Imprecision

Species C values and site FQA scores are imprecise.

Imprecision is inherent in any measure. However, for such a broad, highly encompassing ecological measure, FQA has shown to be accurate and precise enough for its users. Tests of several

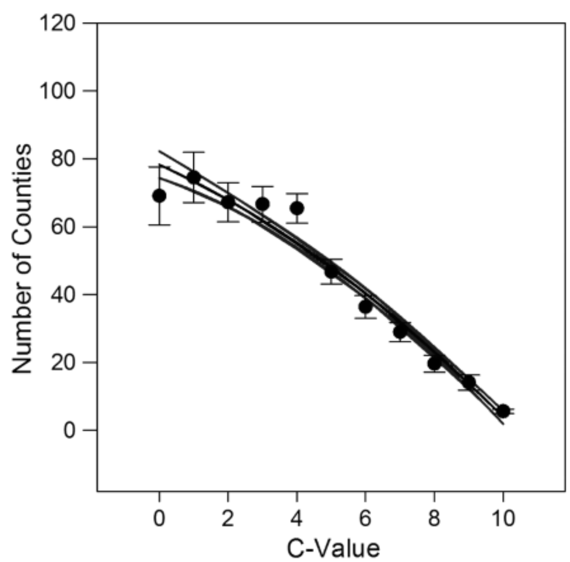

Fig. 7. Rarity and *C* values. The relationship between *C* values and the number of Illinois counties in which a species occurs (natives only). How widely a species is distributed is highly correlated with conservatism. $P < 0.0001$, $r_{adj}^2 = 0.50$, ±95% CI, $N = 2090$. Figure adapted from Spyreas (2014).

hundred individual species values have shown high precision in them (Matthews et al. 2015, Bried et al. 2018). Furthermore, any imprecision in species-level values is dampened by FQA metrics that average all the plants at a site, because the "... overlap of ecological tolerances of multiple species is smaller than the amplitude of a single one" (LaPaix et al. 2009). There is some evidence that FQA metrics are less able to discern floristic quality differences at the highest and lowest ranges—that is, small difference among the highest conservation value, least biologically degraded sites (Bowles and Jones 2006; Fig. 6). Therefore, while FQA has frequently been shown to be an accurate measure, its predictive ability and precision at the extremes requires further study.

Species abundance

There is no accommodation in FQA for how abundant or frequent species are in habitats. This has the effect of over-weighing the importance of rare species at sites. It also ignores that relative abundances within communities can by themselves reflect a sites conservation value and disturbance level (Nielsen et al. 2007).

There are several reasons why FQA's original authors created a measure that could be used irrespective of relative plant abundances or frequencies. The primary one is that the abundance of any given species was considered "... irrelevant or only ambiguously related to" a habitat's floristic quality (Swink and Wilhelm 1994). Second, potentially over-emphasizing diminutive or less-common species could be as problematic as over-emphasizing large or frequent species. Third, presuming to know the ideal evenness or relative species abundances in natural areas is untenable. Finally, the botanists who initially assign *C* values assumedly do not incorporate community abundance into their valuations.

The relative abundance of species at a site is clearly important for both ecological and conservation matters. However, the original authors of FQA suspected that abundance was largely redundant information within FQA metrics. It was their observation that when "a few plants are inordinately abundant in a community, the mean floristic quality is also depressed" (Wilhelm and Ladd 1988). Nonetheless, subsequent authors have created several new FQA calculations that incorporate abundance. However, because few of these calculations have even marginally improved upon the performance of simpler standard metrics when tested, the original author's intuition about redundancy has proven correct (Francis et al. 2000, Cohen et al. 2004, Bourdaghs et al. 2006, Hopple and Craft 2013). While there is little harm in incorporating abundance, users should consider qualifications. First, abundance is rather susceptible to the natural variation that occurs in species by season, year, or with natural disturbance. This would make abundance-weighted scores less comparable across time or across different sites sampled in different years (Ervin et al. 2006). Second, collecting and calculating abundance values takes more time and effort, all without consistent evidence for improved performance.

Overreliance

This single measure is both too influential (e.g., dictating regulatory and land conservation decisions), and incomplete (e.g., does not consider faunas). Floristic Quality Assessment is inadequate for assessing habitats on its own (Taft et al. 1997).

This criticism is justifiable. Distilling the complexity and value of natural systems into a single ecological measure is not ideal, especially considering the breadth of uses and consequences to those uses that FQA has come to assume. There are many reasons that FQA should not be used as a stand-alone measure to evaluate areas. For example, it cannot account for wildlife, ecosystem function, rare plants, or population sizes in an area. Furthermore, users now assume a level of exactness and certainty to scores that is not justified. For example, laws frequently mandate exact score thresholds (e.g., FQI = 20; Streever 1999, Herman et al. 2001, Matthews and Endress 2008), which implies a precision that ignores, among other things, that there is normal variability (intra- or inter-annual) in natural areas. Nevertheless, the charge of overreliance is a critique of how FQA is interpreted and used, rather than what it measures or how it performs.

Restorations and manipulability (or, how to cheat using FQA)

FQA numbers are vulnerable to manipulation. As the legal requirements and financial incentives for achieving high FQA scores have increased, so have *the incentives for abuse. For example, consider a planted restoration that is required to reach or maintain a target FQA score to trigger payment for restoration work. Planting a few high C value species/individuals that are cheap/easy to cultivate and establish at the last minute, artificially inflates the score of an otherwise sub-par restoration. Such quick-fix species may be used regardless of whether they are ecologically appropriate for the region, site, or whether they are likely to survive over the long term. Floristic Quality Assessment's overemphasis of high C value species leaves it more prone to manipulation than other ecological metrics.*

A fundamental downside to any numerical indicator is their tendency to incentivize gaming the system (Muller 2018). Shortcuts to compliance, or even cheating, are a concern for FQA and habitat restoration (DeBerry and Perry 2015), and anecdotal accounts of it are common. But, it can be difficult to prove such malfeasance. Steps to prevent this could include (1) critical examination of site lists for inappropriate species, and (2) requiring long-term monitoring to assure sustained site scores (Matthews and Spyreas 2010).

The possibility of manipulating scores warrants attention, but again, it is mainly a function of how FQA is used rather than a criticism of whether it works as intended. Regarding FQA's use, it is worth noting here that it has never been clear how planted restorations fit within FQA's original conception (e.g., species are scored with regard to occurrence in remnant habitats not modern restorations). Therefore, it can be unclear how scores from novel, created, restored, or de novo habitat should be interpreted. Nonetheless, because they are now widely used to evaluate these habitats, the utility and meaning of FQA metrics need study and consideration regarding them.

When (and where) to draw the line

There are no guidelines for what constitutes the area to sample for FQA. The site is offered as the operant term, but sites are not ecological entities.

Determining the appropriate sample unit in any ecological inquiry is often not an obvious or trivial matter. And, FQA has its own specific set of sampling considerations. Chief among these is the question of where to draw the sample area boundary. (This is not to be confused with the choice of creating the species list from site subsampling, i.e., plots, or sampling the full habitat.) When the area of interest is small and has well-defined boundaries such as property lines, rivers, and roads, the sample area is obvious; it is usually the whole habitat patch (Taft et al. 1997). The species from two or more discrete habitat fragments like this should not be combined in a single site score calculation. On the other hand, determining the boundary within a large, contiguous habitat is more complicated. Should users confine sampling to a single plant community type (e.g., relevé), or physiographic unit (e.g., hydrology, soil type, slope, aspect)? Should users include the species from ecotones (habitat edges or boundaries) in the species list? Few have weighed in on these questions. Floristic Quality Assessment's originators did not warn against crossing vegetation types when compiling a species list, and they suggested including ecotones in the sample area (Swink and Wilhelm 1994). This liberal perspective probably arose because FQA was originally used to assess the value of entire land parcels. Since it has expanded in scope and use, stricter or clearer conventions are needed going forward.

When using the FQI metric specifically, users must consider factors that affect richness if they break up habitats into sample units. This includes the size of the area, environmental gradients, ecotones, or any other contributors to α- or β-diversity within or across the units. Discerning sample boundaries is further complicated by suggestions that different types of vegetation (e.g., forest vs. prairie) may inherently differ in their FQA scoring potential (see *The bias of rarity*). One rule that can universally guide where to delineate sample units is to consider whether they encompass a single land-use history or disturbance level.

A final consideration is what should limit the size of sites or sample areas? At the largest scale, FQA scores generated from an entire state's flora are not useful because state scores are not comparable with one another. But, the floristic quality of county-level floras has successfully been compared over time (Leitner et al. 2008). When comparing habitats, the practicality of surveying increasingly large areas will limit site size. At smaller scales, the minimum size of effective sample plots can be quite small, although usefulness quickly declines at plot sizes of 0.1 m^2 or less, because having so few species in plots greatly inflates the variance (Fig. 2). Swink and

Wilhelm (1994) advise that Mean *C* is a meaningful measure from areas 0.25 m^2 up to 100 acres or more, which is a reasonable rule of thumb.

Species misidentification and detection

FQA relies on correctly identified species. The demise of field botany in science curricula and the loss of botanical expertise (Ahrends et al. 2011) suggest that incomplete and mistake-ridden species lists, and therefore, inaccurate FQA scores predominate. The unfulfillable demand for expert botanists makes widespread and accurate use of FQA prohibitive.

Despite the dearth of botanical expertise, the use of FQA expands unfettered. It is possible that ill-trained botanists are conducting surveys, and that most species lists and scores generated are highly inaccurate. However, studies suggest that FQA is forgiving of both misidentified and undetected species (Figs. 8, 9). Of course, complete, accurate, species lists ensure accurate FQA scores, but there appears to be room for error in the metrics (Spyreas 2016 and citations within).

You're Doing It Wrong: Mistakes in Understanding and Using FQA

The increasingly wide reach of FQA has come with growing pains. This section addresses 10 common and avoidable problems surrounding FQA's use, methods, and understanding.

Wrong definition

There is no more common error among FQA users than misunderstanding what site floristic quality and species conservatism mean. This is understandable given the number of definitions introduced into the recent literature. They have created uncertainty and confusion as to what it is exactly that FQA measures, hindering its utility (Medley and Scozzafava 2009). A review of 100 randomly selected FQA publications published between 1988 and 2013 found that 88% incorrectly defined FQA, and 84% incorrectly interpreted what FQA measures (Spyreas, *unpublished data*). In nearly every case, this stemmed from adding a new theory or concept to the core floristic quality or conservatism definition (see *It works in practice* above). While such concepts may help us understand the ecology of conservatism for some species, and while FQA has been used as a proxy measure for some of them (e.g., succession; Bauer et al. 2015), they should not be confused with its fundamental meaning. At this point, tolerance of anthropogenic disturbance and exclusivity to remnant habitats are the only validated criteria for defining FQA.

Conservatives only need apply

FQA users can become overly fixated on high *C* value species. For example, in restoration work there is concern that a specific subset of species are being over-promoted because they have high *C* values. Restoration plantings are often used to replace destroyed original habitats. These must often meet floristic quality score targets. However, achieving target scores by repeatedly planting the same small group of highly conservative species that are cheap and easy to cultivate and establish (commonly referred to as workhorse species) could be lowering overall regional diversity. This process, termed biological homogenization, occurs when unique local habitat remnants are lost, and repeatedly replaced with the same few species (in this case, planted high *C* value species; Price et al. 2019). Discouraging overused or reoccurring species in planted habitats may be a future consideration.

It is important to remember that FQA does not posit that less conservative species are not ecologically important. Every native plant has value and a place in its respective ecosystem. Instead, FQA conveys two things about high conservative species: (1) All else being equal, they have greater conservation value, and (2) they reflect a site's history of minimal disturbance and degradation.

Man vs. nature: natural vs. human disturbance

Floristic Quality Assessment is intended to separate sites by anthropogenic disturbance levels, including the effects of hydrological alteration, plowing and tilling, heavy livestock grazing, heavy logging, pollution, introduced exotic species, and fire suppression. Natural disturbance regimes include fire, tree windfall, insect outbreaks, and flooding. Some of these could function as either disturbance type depending on the circumstance (e.g., grazing, fire). However, the overwhelming majority of evidence thus far has shown that it is the footprint of modern human disturbances that FQA registers. Further research will be needed to determine if, where, or how often natural disturbances affect FQA measures.

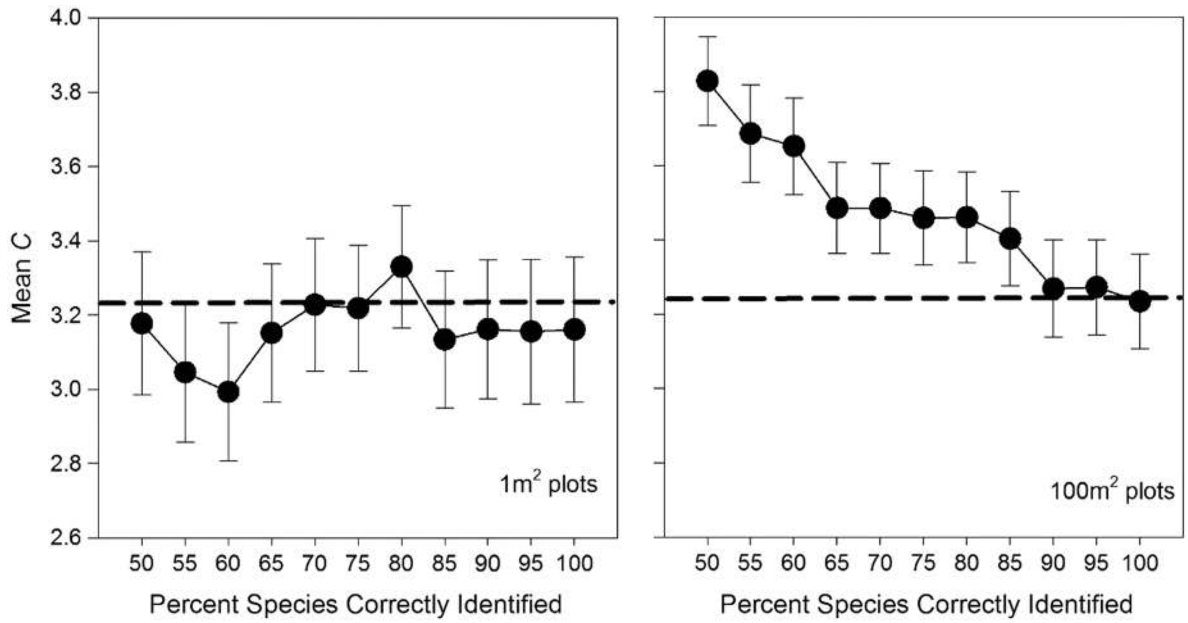

Fig. 8. Species misidentification and Mean *C*. Mean *C* of grassland plots with between 50% and 100% of species correctly identified. Misidentifications were created by randomly replacing species with randomly chosen species from the county flora. The figure shows that Mean *C* is relatively unsusceptible to moderate levels of species misidentification in plots (i.e., <10%). Mean *C* values for 100-m^2 plots are converging upon the Mean *C* of the overall county flora at high misidentification levels. Dashed horizontal line is 100% correct identification value in the 100-m^2 plot for reference. ±95% CI; N = 20. Figure adapted from Spyreas (2016).

Assigning C *values*

The initial process of assigning *C* values is especially important because errors are carried forward every time the values are used. This process has not been equally consistent, transparent, or rigorous across regions. However, most of the procedural mistakes that have been made are easily avoided. For example, some regions have inexplicably deviated from the 0–10 scoring scale (e.g., 1–5, 1–10). Others use unsubstantiated, often disjointed multi-theory scoring criteria (discussed above). Species should be judged on their conservatism; exclusively relying on this single gradient not only simplifies the scoring process, but it avoids inclusion of unsubstantiated scoring criteria.

Other weaknesses to the scoring process seem to have been more a product of expediency than misguided choices. The ability to create scores for a state's flora hinges on the availability or willingness of local botanical experts to score them. Some lists that encompass large multi-state regions have been created by averaging values from the surrounding states (United States Environmental Protection Agency 2016). The precision that reflects local ecological information is sacrificed when importing numbers from states for which they were not created. Similarly, having too few botanists assign state scores is not ideal (e.g., one or two people). The more botanists that are involved, the more knowledge of species and habitats that are available, and the greater ability there is to moderate outlier values (Matthews et al. 2015). Finally, more score-refining steps improve the final list. Ideally, this would include both in-person discussion of scores, as well as opportunities for anonymous, written comments by panelists (detailed in Spyreas 2014). Nonetheless, despite the procedural weaknesses outlined here, regional *C* value lists have thus far proven to be effective.

Species richness

Not controlling or accounting for differences in factors that affect species richness, such as sample area, sample intensity, and time of year,

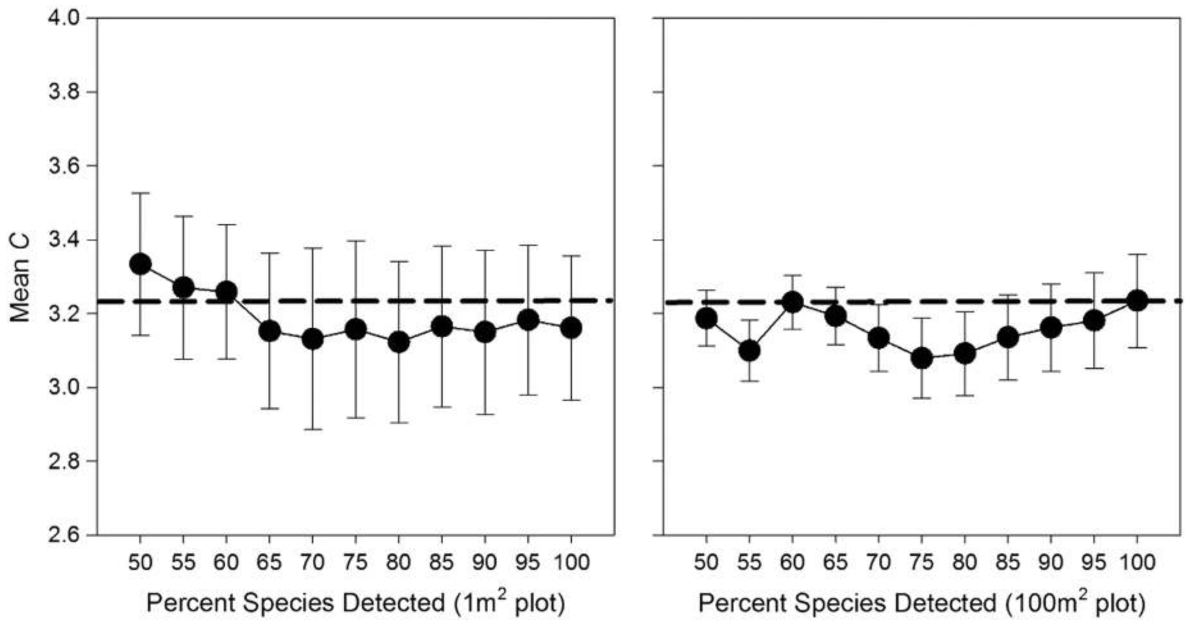

Fig. 9. Species detection and Mean *C*. Mean *C* score for grassland plots with between 50% and 100% of their species being detected. Undetected species were randomly chosen for removal from plot list. Figure illustrates that Mean *C* is remarkably unsusceptible to incomplete species lists. Dashed horizontal line is 100% detection value in the 100-m^2 plot for reference. ±95% CI; N = 20. Figure adapted from Spyreas (2016).

is a common mistake when comparing FQI scores. This highlights the importance of reporting complete, detailed, sample methods when using FQA so that scores are able to be put in context and accurately compared by the reader.

Comparisons among states

Comparing site scores from different states is a mistake because states have their own uniquely assigned *C* values that are only relative to other species in the state. They are assigned to reflect a species' behavior in a specific state, which will differ both for ecological and procedural reasons between states. For example, different botanist panels have used slightly different criteria, baselines, or protocols when assigning *C* values in different states.

Comparisons among vegetation types

Caution may be warranted when comparing site scores from different vegetation types (e.g., forest vs. grassland), as some have suggested that vegetation types have inherently different scoring potentials (see *The bias of rarity* above). While this has not been adequately studied, users should consider this possibility when comparing scores among different habitats within a state.

Yearly variation

The year and season of sampling are potential considerations when interpreting site FQA scores. In younger habitats (where succession is proceeding rapidly), site scores will vary with the habitats age. Therefore, the sample year matters when comparing site scores among habitats or across time where succession in ongoing (Fig. 10; Spyreas et al. 2012). In stable, mature habitats, scores seem not to vary across years, provided human disturbances are not being incurred (Spyreas 2016). However, some authors have warned of year effects in certain circumstances: fluctuations in habitat scores across years due to natural disturbance regimes (e.g., intense floods, droughts, wind storms). Natural year effects have only been discerned in a few, relatively unique situations thus far (Tulbure et al. 2007).

Intra-annual or seasonal effects within a year are also a potential complication, where comparing FQA scores from samples taken at different times of the year may not be appropriate. However, as long as it falls within the growing season (e.g., May–September), sample date has proven largely irrelevant to FQA scores thus far (Matthews et al. 2005). This is not surprising given the weak effect of species detectability (which varies across seasons) on scores (Fig. 9).

Watch your language

One substantial impediment to confident use of FQA is language—navigating the jumble of terms, labels, and even formulas for floristic quality metrics that are loosely spread across the literature. Lack of adherence to standard terms, abbreviations, or metric formulations has created considerable, unnecessary confusion. Different names for the same concepts and measures have been repeatedly re-created and re-named as new in different publications. Floristic Quality Assessment users should avoid adding to this confusion by following the naming conventions in the original, standard texts (Swink and Wilhelm 1994, Taft et al. 1997) and presented here. This especially includes formulas and abbreviations for FQA and its two standard measures the FQI and Mean *C* (including its component coefficients of conservatism [*C*], or *C* values).

A related concept to floristic quality is biological (or ecological) integrity. This concept underpins Indices of Biological Integrity (IBI). Occasionally, biological integrity is used interchangeably with FQA terminology. While not incorrect, FQA users should be mindful that although these are conceptual analogs, biological integrity, and IBIs have their own particular history of study, methodology, and legal frameworks, and to avoid confusion the two should be interchanged with care.

Report FQA-specific methods

Methodological details that are irrelevant to other vegetation measures can be of particular consequence for FQA because they may affect its values and render FQA results uninterpretable when not reported. Such methods include the following: How, when, and by whom sites were

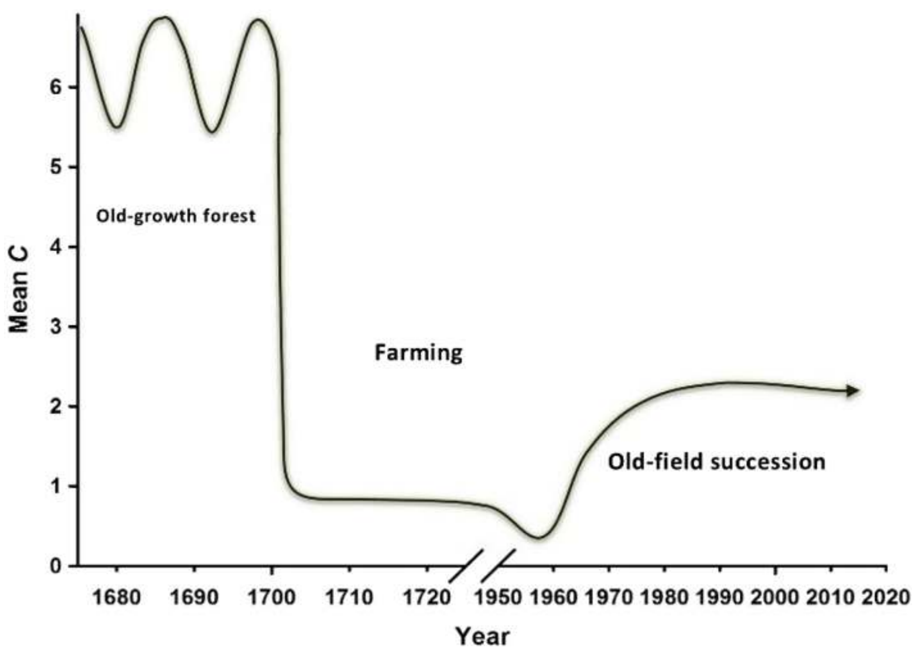

Fig. 10. Old field succession and floristic quality. A timeline of the floristic quality score within a New Jersey (USA) farm field over five centuries. The pre-European settlement Mean *C* value of the forest probably hovered between five and seven (based on contemporary reference forests). While the site was farmed, it would have had little native floristic quality beyond farm weeds. Fifty-five years of succession following agricultural abandonment have only yielded moderate gains in floristic quality. Figure based on Spyreas et al. (2012).

sampled; how sample intensity was controlled for (i.e., number of site visits, amount of time for surveys); what the size the site was (area), or the size/number of sample plots; how site or sample-unit boundaries were delineated; whether species across different community types were pooled in calculations; whether ecotones were sampled; how species lists from samples were assembled to calculate scores (e.g., cumulative site list, plot-based averages); whether exotic species were included in calculations; how unidentified species were treated; and where species *C* values were obtained from if more than one source for scores were available for the region.

Known Unknowns and the Future

Floristic Quality Assessment will continue to gain influence and expand in use, despite the gaps in understanding (Medley and Scozzafava 2009). Much of what is unexamined about FQA relates to large-scale patterns in, and understanding of, site scores. Practitioners and researchers need to be better able to put site scores in their

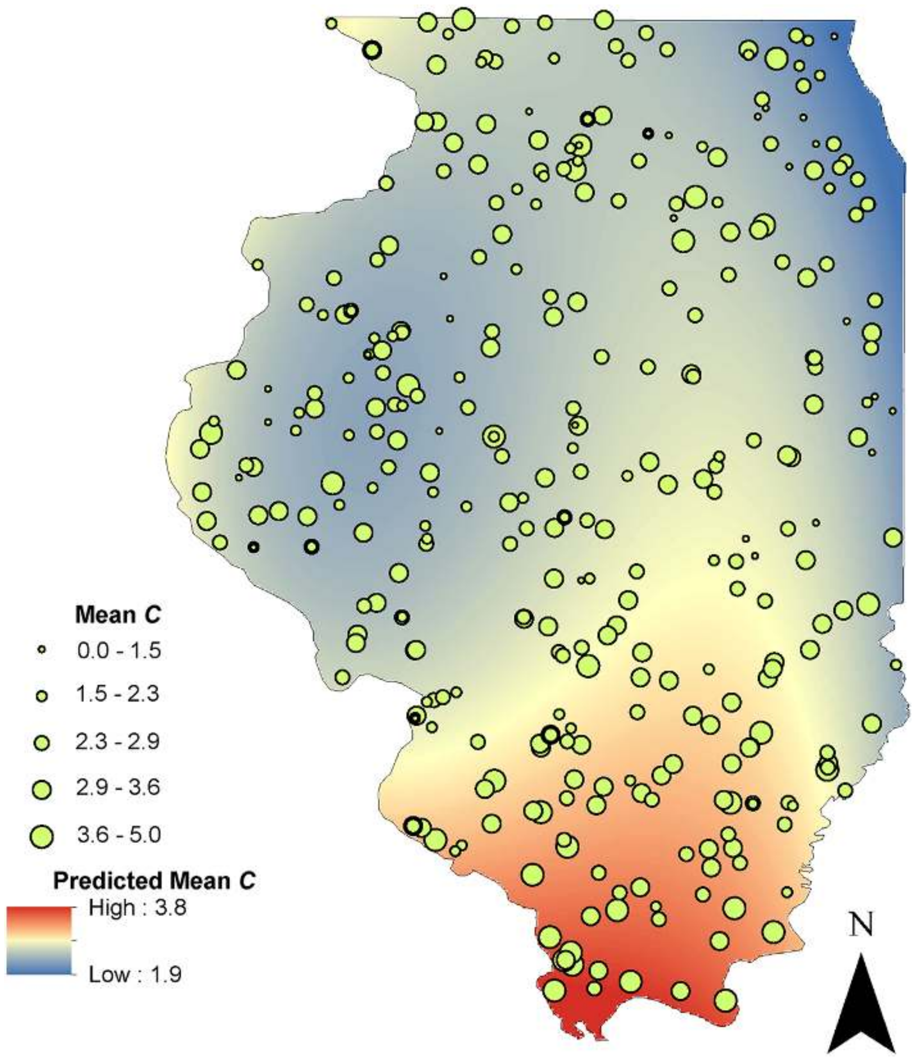

Fig. 11. A latitudinal gradient to floristic quality? Floristic quality of randomly chosen wetlands and forests across Illinois showing higher Mean *C* levels in the south. Yellow circle size indicates Mean *C* of sample sites, which were used to interpolate (i.e., predict) overall regional trends (color shading). Figure adapted from Spyreas (2014).

state into context. For example, what constitutes a high, medium, or low score for a given habitat type, region, or state? What is the highest score attainable? Are there ecological thresholds to site scores in communities? And, if so, are strict FQA score cutoffs appropriate or meaningful (e.g., for regulatory use)? These questions will require large datasets to set regional reference values.

Other questions regarding how site scores should be compared with one another abound. For example, there is some evidence that site floristic quality exhibits a latitudinal gradient similar to the well-established global gradient in species richness (Fig. 11). While it is unclear whether this is due to the inherent differences in the ecology of the vegetation, or if as some studies suggest, it is an artifact of regional human disturbance patterns (Nichols 1999), such a pattern affects how comparisons of site scores within a state should be interpreted. No doubt one of the most exigent FQA questions relates to habitat restorations. Because FQA was created with regard to remnant native habitats, its place alongside habitat plantings, recreations, or other novel habitats is unclear. Swink and Wilhelm (1994) suggested that scores of habitat plantings never reach levels found in the best habitat remnants, but there is some evidence to the contrary (Taft et al. 2006). Nonetheless, there is little guidance available for how comparisons of restorations vs. remnants should be made and interpreted.

As it continues to expand, considerations for how FQA will work when applied to new geographic regions may arise. For example, will FQA be applicable where few pristine remnant habitats remain to provide benchmarks for setting species values (e.g., long developed or highly urbanized areas in Europe)? Or, vice versa, how can *C* values be set in regions with little degraded habitat to serve as low-quality references (e.g., arctic, boreal, or montane zones)? The potential utility of FQA in hyper-diverse habitats such as tropical forests also seems improbable because species-level identification can be so difficult. Conversely, low diversity regions/habitats may not provide enough species at the site level to give scores sufficient resolution or power.

Finally, as remnant habitats degrade over time, or as species migrations due to global climate change scramble the foundational reference assemblages that form the basis for existing *C* values, do species values need to be reconsidered and reset as ecosystems alter and shift? Regional *C* value lists should, and in some cases, already have been refined to keep up with taxonomic changes, and perhaps to tweak initial valuations. But, the decision to explicitly account for increased degradation to the landscape over time with updated iterations of *C* value lists (Wilhelm and Rericha 2017) is not a simple or insignificant one.

Acknowledgments

Thanks to Eric Jansen, Michael Jeffords, Paul Marcum, Jeffrey W. Matthews, Michael Murphy, Dennis Nyberg, Edward Price, Cassi Saari, Jessica Stern, Gary Sullivan, John Taft, and Jack Zinnen for use of photographs or for comments on earlier versions of the manuscript. Thanks to Andy Olnas for the excellent North American map.

Literature Cited

Ahrends, A., et al. 2011. Conservation and the botanist effect. Biological Conservation 144:131–140.

Bauer, J. T., L. Koziol, and J. Bever. 2018. Ecology of Floristic Quality Assessment: testing for correlations between coefficients of conservatism, species traits and mycorrhizal responsiveness. AoB Plants 10:1–13.

Bauer, J. T., K. M. L. Mack, and J. D. Bever. 2015. Plant-soil feedbacks as drivers of succession: evidence from remnant and restored tallgrass prairies. Ecosphere 6:1–12.

Bourdaghs, M., C. A. Johnston, and R. R. Regal. 2006. Properties and performance of the Floristic Quality Index in Great Lakes coastal wetlands. Wetlands 26:718–735.

Bowles, J. M., and M. Jones. 2006. Testing the efficacy of species richness and Floristic Quality Assessment of quality, temporal change, and fire effects in tallgrass prairie natural areas. Natural Areas Journal 26:17–30.

Bried, J. T., B. E. Allena, E. T. Azeria, V. E. Crisfield, and M. J. Wilson. 2018. Experts and models can agree on species sensitivity values for conservation assessments. Biological Conservation 225:222–228.

Bried, J. T., S. T. Jog, and J. W. Matthews. 2013. Floristic quality assessment signals human disturbance over natural variability in a wetland system. Ecological Indicators 34:260–267.

Chapman, M. G., A. J. Underwood, and K. R. Clarke. 2009. New indices for ranking conservation sites using 'relative endemism'. Biological Conservation 142:3154–3162.

Chu, S., and B. Molano-Flores. 2013. Impacts of agricultural to urban land-use change on Floristic Quality Assessment indicators in Northeastern Illinois wetlands. Urban Ecosystems 16:235–246.

Cohen, M. J., S. Carstenn, and C. R. Lane. 2004. Floristic quality indices for biotic assessment of depressional marsh condition in Florida. Ecological Applications 14:784–794.

DeBerry, D. A., and J. E. Perry. 2015. Using the floristic quality concept to assess created and natural wetlands: ecological and management implications. Ecological Indicators 53:247–257.

Devictor, V., R. Julliard, and F. Jiguet. 2008. Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation. Oikos 117:507–514.

Ejrnæs, R., T. G. Frøslev, T. T. Høyea, R. Kjøller, A. Oddershede, A. K. Brunbjerg, A. J. Hansen, and H. H. Bruun. 2018. Uniquity: a general metric for biotic uniqueness of sites. Biological Conservation 225:98–105.

Ervin, G., B. Herman, J. Bried, and D. C. Holly. 2006. Evaluating non-native species and wetland indicator status as components of wetlands floristic assessment. Wetlands 26:1114–1129.

Filippi-Codaccioni, O., V. Devictor, Y. Bas, and R. Julliard. 2010. Toward more concern for specialisation and less for species diversity in conserving farmland biodiversity. Biological Conservation 143:1493–1500.

Francis, C. M., M. J. W. Austen, J. M. Bowles, and W. B. Draper. 2000. Assessing floristic quality in southern Ontario woodlands. Natural Areas Journal 20:66–77.

Hawkins, C. P. 2006. Quantifying biological integrity by taxonomic completeness: its utility in regional and global assessments. Ecological Applications 16:1277–1294.

Herman, K. D., L. A. Masters, M. R. Penskar, A. Reznicek, G. Wilhelm, W. W. Brodovich, and K. P. Gardiner. 2001. Floristic Quality Assessment with wetland categories and examples of computer applications for the state of Michigan. Revised Second edition. Michigan Department of Natural Resources, Wildlife Division, Natural Heritage Program, East Lansing, Michigan, USA.

Hopple, A., and C. Craft. 2013. Managed disturbance enhances biodiversity of restored wetlands in the agricultural Midwest. Ecological Engineering 61:505–510.

Izco, J. 1998. Types of rarity of plant communities. Journal of Vegetation Science 9:641–646.

Kutcher, T. E., and G. E. Forrester. 2018. Evaluating how variants of floristic quality assessment indicate wetland condition. Journal of Environmental Management 217:231–239.

Ladd, D. 1993. The Missouri floristic quality assessment system. The Nature Conservancy, St. Louis, Missouri, USA.

LaPaix, R., B. Freedman, and D. Patriquin. 2009. Ground vegetation as an indicator of ecological integrity. Environmental Reviews 17:249–265.

Lawton, J. H. 1999. Are there general laws in ecology? Oikos 84:177–192.

Leitner, L. A., J. H. Idzikowski, and G. S. Casper. 2008. Urbanization and ecological change in Milwaukee County. *In* D. M. Waller and T. P. Rooney, editors. The vanishing present: Wisconsin's changing lands, waters, and wildlife. University of Chicago Press, Chicago, Illinois, USA.

Mace, G. M. 2005. An index of intactness. Nature 434:32–33.

Mack, J. 2007. Developing a wetland IBI with statewide application after multiple testing iterations. Ecological Indicators 7:864–881.

Matthews, J. W. 2003. Assessment of the Floristic Quality Index for use in Illinois, USA, wetlands. Natural Areas Journal 23:53–60.

Matthews, J. W., and A. G. Endress. 2008. Performance criteria, compliance success and vegetation development in compensatory mitigation wetlands. Environmental Management 41:130–141.

Matthews, J. W., and G. Spyreas. 2010. Convergence and divergence in plant community trajectories as a framework for monitoring wetland restoration progress. Journal of Applied Ecology 47:1128–1136.

Matthews, J. W., G. Spyreas, and C. M. Long. 2015. A null model approach for testing the validity of plant species' Coefficients of Conservatism. Ecological Indicators 52:1–7.

Matthews, J. W., P. A. Tessene, S. M. Wiesbrook, and B. W. Zercher. 2005. Effect of area on isolation on species richness and indices of floristic quality in Illinois, USA wetlands. Wetlands 25:607–615.

Medley, L., and M. Scozzafava. 2009. Moving toward a national Floristic Quality Assessment: Considerations for the EPA National Wetland Condition Assessment. National Wetlands Newsletter 31(6–9):22.

Muller, J. Z. 2018. The tyranny of metrics. Princeton University Press, Princeton, New Jersey, USA.

Nichols, S. A. 1999. Floristic Quality Assessment of Wisconsin lake plant communities with example applications. Lake and Reservoir Management 15:133–141.

Nielsen, S. E., E. M. Bayne, J. Schieck, J. Herbers, and S. Boutin. 2007. A new method to estimate species and biodiversity intactness using empirically derived reference conditions. Biological Conservation 137:403–414.

Niemi, G. J., and M. E. McDonald. 2004. Application of ecological indicators. Annual Review of Ecology and Systematics 35:89–111.

Panzer, R., and M. W. Schwartz. 1998. Effectiveness of a vegetation-based approach to insect conservation. Conservation Biology 12:693–702.

Price, P. F., G. Spyreas, and J. W. Matthews. 2019. Wetland compensation and its impacts on β-diversity. Ecological Applications 29:1–11.

Ricotta, C. 2004. A parametric diversity measure combining the relative abundances and taxonomic distinctiveness of species. Diversity and Distributions 10:143–146.

Scholes, R. J., and R. Biggs. 2005. A biodiversity intactness index. Nature 434:45–49.

Smith, P. G. R., and J. B. Theberge. 1986. A review of criteria for evaluating natural areas. Environmental Management 10:715–734.

Spyreas, G. 2014. An examination of temporal trends, regional variation, and habitat-type differences in site-level floristic quality, and their implications for the use of Floristic Quality Assessment. University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois, USA.

Spyreas, G. 2016. Scale and sampling effects on floristic quality. PLOS ONE 11:e0160693.

Spyreas, G., S. J. Meiners, J. W. Matthews, and B. Molano-Flores. 2012. Successional trends in floristic quality. Journal of Applied Ecology 49:339–348.

Streever, W. J. 1999. Examples of performance standards for wetland creation and restoration in Section 404 permits and an approach to developing performance standards. U.S. Army Engineer Research and Development Center, Vicksburg, Mississippi, USA.

Swink, F., and G. Wilhelm. 1979. A method for environmental assessment of open land. Pages 850–861 *in* Plants of the Chicago region. The Morton Arboretum, Lisle, Illinois, USA.

Swink, F., and G. Wilhelm. 1994. Floristic Quality Assessment. Pages 11–18 *in* Plants of the Chicago region. Indiana Academy of Science, Indianapolis, Indiana, USA.

Taft, J. B., C. Hauser, and K. Robertson. 2006. Estimating floristic integrity in tallgrass prairie. Biological Conservation 131:42–51.

Taft, J. B., G. Wilhelm, D. Ladd, and L. A. Masters. 1997. Floristic Quality Assessment for vegetation in Illinois, a method for assessing vegetation integrity. Erigenia 15:1–24+ Appendix.

Tulbure, M. G., C. A. Johnston, and D. L. Auger. 2007. Rapid invasion of a Great Lakes coastal wetland by non-native *Phragmites australis* and *Typha*. Journal of Great Lakes Research 33:269–279.

United States Environmental Protection Agency. 2016. National Wetland Condition Assessment: 2011 Technical Report. EPA-843-R-15-006. US Environmental Protection Agency, Washington, D.C., USA.

Wilhelm, G. S., and D. Ladd. 1988. Pages 361–375 *in* Natural area assessment in the Chicago Region. Transactions of the 53rd North American Wildlife & Natural Resources Conference, March 8–23, Louisville, Kentucky. Wildlife Management Institute, Washington, D.C., USA.

Wilhelm, G., and L. Rericha. 2017. Flora of the Chicago region. A floristic and ecological synthesis. Indiana Academy of Science, Indianapolis, Indiana, USA.