Duke NICHOLAS INSTITUTE FOR ENVIRONMENTAL POLICY SOLUTIONS

Working Paper
May 2019

Nicholas Institute for Environmental Policy Solutions nicholasinstitute.duke.edu

Trends in Measuring Biodiversity and Ecosystem Function in Mitigation Quantification Methodologies

Erika Zambello,[1] Lydia Olander,[2] Emma Glidden-Lyon,[3] Emily Meza,[4] and Jessica Wilkinson.[5]

CONTENTS

Author Affiliations
[1] Choctawhatchee Basin Alliance; [2] Nicholas Institute for Environmental Policy Solutions, Duke University; [3] Senior Analyst at Abt Associates; [4] Architect and Urban Planner at Arcadis; [5] The Nature Conservancy

Citation
Erika Zambello, Lydia Olander, Emma Glidden-Lyon, Emily Meza, and Jessica Wilkinson. 2018. "Trends in Measuring Biodiversity and Ecosystem Function in Mitigation Quantification Methodologies." NI-WP 19-01. Durham, NC: Duke University.

Review
We would like to thank Bruce McKenney, Joe Kiesecker, and others from The Nature Conservancy's Development by Design Team for working with us on this project. We would also like to thank our reviewers Betsy Albright, Carter Ingram, and Dean Urban, as well as Alexander Challen Hymen from the Mattie Kelly Institute of the Environment for coding advice.

Publication Number: NI WP 19-01

SUMMARY

Over the last decade, efforts to use compensatory mitigation to manage and ameliorate the impacts of development on biodiversity and ecosystems around the world have accelerated. Mitigation mechanisms provide a structured way to advance economic development and infrastructure while also achieving environmental goals. In order to operationalize mitigation programs, practitioners need a methodology for calculating or quantifying impacts and offsets (debits and credits). The methods currently employed in the U.S. and abroad are extremely varied. Surprisingly, the literature on best practices or standards for developing science-based approaches to the quantification of impacts and offsets is sparse and there is also no single broadly accepted best practice guidance.

This paper analyzes a sample of quantification methods currently in use to identify the most commonly used metrics and methods for quantifying impacts and offsets and evaluate the degree to which these methods conform to existing guidance. We reviewed 43 mitigation quantification protocols to assess both "what" these methodologies were measuring (i.e., the metrics used to quantify impacts and offsets) and "how" they were measuring it (i.e., direct or in direct measures of ecosystem functions).

Interestingly, we found no trends in "what" or "how" biodiversity and ecosystems have been measured over time and thus see no indication that these methodologies have been converging over the study period (1965 to 2014) on a common set of metrics and methods for quantifying impacts and offsets. This is an indication that information on best practices is poorly shared across different types of mitigation programs (e.g., species program to wetland program) and even within the same program types (e.g., wetland program to wetland program). When current measurement practices (both "what" and "how") are compared to the limited best practices outlined in professional and academic literature, we also find that few of the protocols adhere to these existing recommendations. This is another indication that existing technical guidance on quantification methodology development is poorly shared or somehow falls short. We conclude that there would be great benefit to investing in the development of further science-based literature on and best practices for the development of robust, reasonably precise mitigation quantification methodologies. We also offer several recommendations for further academic research on the topic that we believe would help to improve the ability of mitigation programs to achieve conservation objectives. Ultimately, we hope that this analysis can inform the development and dissemination of robust best practice guidance and the adoption of quantification tools that reflect high standards for conservation.

INTRODUCTION

As the friction between development and natural resource protection grows, mitigation has become a well-accepted way to protect environmental values, such as specific habitat types, species, or, more broadly, biodiversity, while meeting development needs. Mitigation is a structured decision-making process that seeks to achieve a defined conservation goal. The term refers to the step-wise process of project review that seeks to first avoid impacts to the most important resources, minimize impacts, and then offset remaining residual impacts, also known as compensatory mitigation (Quetier and Lavorel 2011). Some mitigation policies seek to apply the mitigation hierarchy to achieve no net loss to the resources under consideration, while others strive to achieve a net benefit. Put simply, offsets are an exchange: enhancing or protecting ecological values at an "offset site" in order to counterbalance the loss of these values at the "impact site" (Gordon et al. 2011; Moreno-Mateos et al. 2015). Mitigation quantification tools are applied to support decisions about how much and what kind of compensation is required to meet a program's goal, such as no net loss.

A growing number of countries are adopting mitigation policies and incorporating mitigation principles into development impact assessment and global development financing decision making. By 2014, at least 56 countries had policies for mitigation offsets, with 97 programs operating across the globe and another 15 under development (OECD 2014). In the United States, there are several policies in place that provide guidance on the application of mitigation under a wide range of authorities (Wilkinson 2016).

The importance of the first two steps of the mitigation hierarchy—avoidance and minimization of impacts—are critical and essential for meeting programs' conservation goals (McKenney and Wilkinson 2015). This paper, however, focuses on the last step in the mitigation hierarchy: compensatory mitigation.

Compensatory mitigation programs support a growing restoration economy (BenDor et al. 2015) and more conservation funding is generated through these programs for habitat protection and restoration (an estimated $3.8 billion a year) than through other conventional conservation funding sources; e.g., $435 million in funding through the Land and Water Conservation Fund in fiscal year 2019 (LWCF n.d.). The desire to ensure that compensatory mitigation dollars are maximized for conservation outcomes is, as a result, strong.

Compensatory mitigation has not, however, always proven to have a good track record for meeting conservation outcomes (Quetier and Lavorel 2011; Gardner et al. 2013; Moreno-Mateos et al. 2015). This track record can be attributable, in part, to the relative infancy of restoration science. Rey Benayas et al., for example, reviewed 89 assessments of restoration projects from all continents (except Antarctica) and found that although restoration improved degraded systems, it did not result in biodiversity or ecosystem services that were equal to the natural, nondegraded ecosystems (2009). The National Academy of Sciences determined in 2001 that the performance of wetland mitigation projects could be improved dramatically through the increased use of ecological performance standards in setting permit requirements and better site selection in a landscape context (Turner, Redmond, and Zedler 2001).

Without a doubt, there is significant risk and uncertainty in the compensatory mitigation endeavor and a certain degree of hubris in the assumption that we can replace natural systems that have developed over millennia. One of the natural reactions to the poor track record of compensatory mitigation is to develop habitat quantification tools that strive for greater degrees of precision. However, many conservation professionals recognize the inherent difficulty in "measuring" ecological impacts and uplifts. Moreno-Mateos et al. emphasize that commonly used metrics for ecosystem health, such as area restored, habitat area specified for certain species, and ecosystem function "reflect only a small part of local ecosystem composition, structure and function" (2015). As a result of past shortcomings and the desire to capture the full range of ecosystem properties, mitigation programs often fall into the "precision trap," seeking to reduce uncertainty and the risk of restoration failure by developing overly complicated quantification methods that are challenging, time consuming, and expensive to apply in practice (Tallis 2015; Moreno-Mateos et al. 2015; Willamette 2011). It is not clear, however, that greater precision provides more predictable and effective outcomes for conservation. The alternative may be a conservative approach that combines straightforward and science-based quantification tools that can demonstrate a roughly proportional relationship between the impacts and the offset (both in nature and extent) with an additional roughly proportional set aside or adjustment to cover risk of failure and uncertainty.

Surprisingly, only limited guidance exists to support the development of mitigation quantification methods (Willamette 2011; BBOP 2012; Kiesecker 2010; King and Price 2004; Tallis 2015; Naeem et al. 2015) and no one set of recommendations seems to have been widely applied as the standard. Given the growth of mitigation programs, there is merit in advancing the science and practice of mitigation quantification tool development. To that end, this paper analyzes a sample of methods currently in use, the degree to which these methods conform to existing best practice guidance, and whether these approaches are converging on a common set of metrics. Our hope is that this analysis can guide the development of additional guidance that supports the development of science-based, reasonable quantification tools that appropriately account for risk and uncertainty.

APPROACHES TO MITIGATION QUANTIFICATION

Mitigation quantification tools, at their simplest, utilize a set of metrics to evaluate habitat function or condition and quantify the loss of these values on the impact side and the gain in functions on the offset side. These metrics are designed to take a straight acreage measure—such as acres or linear feet—and assign to that unit a functional measure (e.g., 0.1 to 1.0) so the unit is, for example, a functional acre. These metrics are essentially proxies for more direct measures of habitat quality and function (Willamette 2011).

The choice of metrics influences how practitioners account for gains and losses (Goncalves et al. 2015). Early mitigation programs generally utilized area as the only metric and applied simple ratios to the area measure. This approach, however, may not have adequately accounted for risk of failure (Maron et al. 2012) or the functional value of habitat loss and offsets (Salzman and Ruhl 2006). Many of the early ratio-based approaches relied upon best professional judgement without any explicit discussion of how ecological factors were considered, if at all, in their development (Bruggeman and Jones 2008). As understanding of the importance and complexity of ecosystem function grew, area alone was considered inadequate to quantify impacts and offsets and mitigation programs sought to develop metrics that more directly quantified habitat or resource function and condition (Salzman and Ruhl 2006).

Because of concerns that existing ratio approaches were not adequately capturing habitat function or condition, some mitigation programs express a preference for the use of methodologies that take into account the functionality of habitat lost and offset. The regulations governing the federal wetland and stream program, the 2008 Compensatory Mitigation Rule, for example, state that when appropriate functional or condition assessment methods or other suitable metrics are available, they should be used to measure impacts and offsets. The Rule, however, recognizes the challenges inherent in developing such approaches and therefore establishes that when such a methodology is not available, agencies should use a minimum one-to-one acreage or linear-foot compensation ratio. The one-to-one ratio should, however, be adjusted to account for: "the method of compensatory mitigation (e.g., preservation), the likelihood of success, differences between the functions lost at the impact site and the functions expected to be produced by the compensatory mitigation project, temporal losses of aquatic resource functions, the difficulty of restoring or establishing the desired aquatic resource type and functions, and/or the distance between the affected aquatic resource and the compensation site" (Compensatory Mitigation for Losses of Aquatic Resources 2008).

Quantification methods can also utilize adjustment factors to incentivize specific project attributes or account for risk and uncertainty. So, for example, if a mitigation program has a preference for offset projects to be located in close proximity to impacts, provide like-for-like offsets, or minimize temporal loss of functions, projects that don't have these attributes may receive fewer credits. Adjustment factors can also be used to address things like temporal loss of functions (e.g., habitat type that take a long time to become established) or restoration methods that do not have a long track-record of demonstrated success (McKenney and Wilkinson 2015).

RECOMMENDATIONS FOR BEST PRACTICES

This study sought to explore what metrics existing quantification methods use and how these metrics are measured. We looked first to the peer reviewed and gray literature (see References) to determine what, if any, best practices are offered by academics, government agencies, and nongovernmental organizations.

"WHAT"

The best practice papers include a number of recommendations for "what" mitigation quantification methodologies should measure. Many recommend that quantification tools measure the connections between the offset project and other features in the landscape, including hydrological connections, critical habitat areas, and other planned conservation priorities (Willamette 2011; BBOP 2012; USACE 2008; Kiesecker, 2010; King and Price 2004; Tallis 2015). The purpose of this is to place the offset project in the larger landscape-scale context, understanding not only its relative habitat value but also its connection to other ecological features, conservation efforts, and possible threats.

When projects (and protocols) are focused on an ecosystem or habitat, the Willamette Partnership (2011) recommends taking into consideration whether the offset project will affect species or habitat that is high risk or high priority (Willamette 2011). In these cases, a measurement of priority habitat and endangered or threatened species is recommended.

Many of these papers also recommend measurement of both ecosystem function and services (Willamette Partnership 2011; USACE 2008; King and Price 2004; Naeem et al. 2015; Tallis 2015), while others recommend the measurement of habitat function, condition, and/or quality (Gardner et al. 2013; USACE 2008). Tallis (2015) states that measurements of ecosystem services should include provisioning, regulating, supporting, and cultural services. This paper also suggests that measurements of ecosystem function allow for direct comparison between programs based on a functional level and allow for more accurate measurement of impacts and offsets (Tallis 2015, Naeem et al. 2015).

Gardner et al. (2013) suggest that the appropriate metrics for biodiversity offset programs are components of biodiversity that are of particular importance to people, ecosystem functions, and surrogate components that represented unmeasured biodiversity. King and Price (2004) and Naeem et al. (2015) recommend that quantification methods account for the lag time between project impacts and when offsets reach maturity.

Finally, the literature also recommends including metrics to account for uncertainty and risk (Willamette 2011; USACE, 2008; King and Price 2004; Naeem et al. 2015). This includes measurements for how threats like climate change and invasive species may affect impacts or offsets (Naeem et al. 2015) and how much risk or uncertainty is present in the type of project being implemented by the mitigation program (King and Price 2004).

"HOW"

Many of the best practice papers reviewed included specific recommendations for "how" quantification metrics should be measured. The papers state that both direct and proxy measures of habitat and ecosystem functionality may be appropriate (Willamette 2011; USACE 2008; Naeem et al. 2015). The literature recommends using qualitative measurements for habitat and species and limited quantitative measures to assess condition, extent, or other attributes relevant to habitat and species functionality (Willamette 2011; USACE 2008; King and Price 2004; Naeem et al. 2015).

How ecosystem services should be measured, however, was not frequently discussed. This indicates a possible disconnect between the recognized importance of ecosystem services and the difficulty in measuring them. Naeem et al. (2015), among others, notes that a gap exists between measurements of the production of ecosystem services and the actual delivery of these services. While authors suggest that ecosystem services should be measured, they are not currently included in most protocols. New approaches for defining ecosystem services measures, called benefit relevant indicators, are emerging and may prove useful for filling this gap (Olander et al. 2018).

This study asks two sets of questions. First, are there trends in "what" is measured and "how" it is measured in existing quantification methodologies? Have the measurements required evolved over time as new methods are developed? Do the measures vary by type of program? For example, are there specific measures used for wetlands and others for biodiversity? And second, how well do quantification methodologies meet best practice recommendations from the literature (Table 1)?

Table 1: Summary of best practice recommendations drawn from literature

Best practice recommendations for "what" protocols should measure	Best practice recommendations for "how" protocols should measure
Connectivity of important habitat	Direct quantitative measures of ecosystem functions or condition
Perceived ecosystem values for species assemblages and land cover type	Proxy quantitative measures of ecosystem function or condition
Ecosystem function	Qualitative assessment of species and habitat functions or condition
Ecosystem services	
Risk of mitigation failure	
Time required for offsets to equal development	

METHODS

Data Collection

For the purpose of this paper, a "quantification methodology" is a system of metrics and accounting rules used to quantify impacts and offsets for the purposes of implementing a mitigation program. "Projects" are individual compensatory mitigation projects carried out on the ground that seek to provide functional uplift of habitat through activities that restore, enhance, protect, or create habitat, and are measured by a specific quantification methodology.

For this analysis we collected information on "what" and "how" quantification methodologies measured environmental impacts and offsets. We reached out to a few experts to identify an initial set of methodologies for inclusion in our analysis. The bibliographies of the guidance documents for these protocols as well as professional and academic papers in our literature review directed us to additional mitigation offset protocols. We used this snowballing technique to assemble a list of sample protocols for our analysis. The informal snowballing approach did not yield a comprehensive data set, but rather a sample set.

Methodologies were rejected from the analysis when publicly available documents did not include enough information, or if they did not provide details on how mitigation was measured. We sought to include only methodologies that were actively in use and, as such, did not include in this analysis those that had been proposed but had not yet been adopted and implemented and those that were developed for a specific project rather than applied routinely for projects operating under a similar rule set. The sources reviewed included the primary documentation for the mitigation quantification methodology, methodology websites, and outside case studies. This resulted in a dataset of 43 quantification methodologies, 40 of which had enough publicly available information to support our "what" analysis, and 31 of which had enough publicly available information to support our "how" analysis (a full list of methodologies and references used in this analysis can be found in Appendix A).

We then reviewed the metrics used in each protocol and developed different "what" and "how" categories into which the measures could be logically grouped (Tables 2 & 3). These classifications came from our observations of the data and are consequently somewhat subjective. We created a list that is broader than simply dividing "functional" versus "nonfunctional" metrics, but not so broad that the groups became too small to analyze.

From the list of what was measured across all protocols, we created a list of 13 "what" categories (Table 2). A full list of the metrics that were included in each "what" category can be found in Appendix B.

Table 2: "What" measurement categories and descriptions

"What" category	Description
Riparian/Wetland structure	Riparian ecosystem structures (bank width, buffering, stream channels, etc.).
Acre/Area	Acre and area measurements.
Effects on biogeochemical cycles	Measures of biogeochemical cycles (sediment, nutrients, and thermoregulation).
Effects on hydrological cycle	Measures of the hydrological cycle, (water quality, volume retention, etc.).
Species composition	Biodiversity indicators, specific plant and animal communities.
Connectivity	Distance to other protected landscape, potential for wildlife migration, and other measures of overall distance and connectedness across the landscape.
Habitat type	A soil, wetland, or ecosystem type (coastal marsh, etc.)
Habitat for specific species	Habitat or habitat features that support specific plants or animals.
Vegetation cover/Structure	Measures such as grass cover, shrub cover, and canopy cover, as well as vegetation structures, such as number of trees of a certain age or fallen log. (When vegetation structures are included as habitat, such as tree hollows, they are included in the above category.)
Condition	Condition or health of an ecosystem.
Aesthetics and recreation	Measures of aesthetic value, recreational value.
Time for offsets to equal development	Measures of project lags, time to offset completion, etc.
Other	Measures of risk and vulnerability, management activities, size-by-species distribution, and others which were not common across protocols.

We analyzed the quantification methodologies based on six "how" characteristics (Table 3). A full list of measurements included in each "how" category can be found in Appendix B.

Table 3: "How" measurement categories and description

"How" category	Description
Counts	Number of days, length of stream, area of habitat, number of species on a site, etc.
Presence/Absence	Binary yes/no category. For example: endangered species present or not, breeding habitat present or not, etc.
Proportion	Percent measures (eg., area of tree coverage per square km), densities (eg., stalks per square meter), ratios (eg., 85% coverage compared to reference site).

“How” category	Description
Descriptions	High/medium/low categorical bins based on qualitative, pure narratives, or approximate, best-judgements calls (eg., such as when filling out boxes on forms, etc.).
Expert judgment	Identification of key elements (ex. habitat, species, other) by trained professionals, specifically called out as such (eg., CapeNature biodiversity specialist); trained wetland scientists.
Proxy ecosystem function	Measure based on indicative wording in the paper (eg., measuring dissolved oxygen) to infer fish health, or number of scrub jay families per area as a proxy for population viability. Additionally, the use of complex models with complex inputs (eg., Shadelator, GIS Spatial Reference Models with complex inputs (species, density, stressor layers, etc.) to calculate something like “proximity”) above and beyond simple formulas.
Weighted Metrics	Formulas or equations that define extent of mitigation required, credits received, etc. Includes multipliers (such as numbers assigned to good, moderate, poor).

ANALYSIS

To assess whether and to what degree the metrics used in the quantification methodologies changed over time we used generalized linear models (GLM), using both poisson and binomial families. If the number of metric categories used per method increases or decreases over time—which we viewed as an indication of the complexity of the methodology—we used R to run a separate GLM comparing the year to the total number of metric categories used for both “what” and “how” metrics. To assess whether the use of each individual metric category was more likely to be used as new methodologies were developed over time, we ran a logistic GLM from the binomial family. With this test, we were able to statistically analyze whether use of specific “what” and “how” metric categories were more likely to be used as time progressed. Both tests would indicate patterns over time. If the protocols seem to better follow recommendations over time, this would provide some evidence that the programs may be building or learning from past experience over time and converging on a set of best practice metrics.

Descriptive analysis of projects across “what” and “how” categories was performed in Excel to assess how different quantification methodologies differ in category use. This enabled us to determine, for example, if terrestrial protocols more frequently use different “what” categories compared to aquatic protocols and whether there were differences in categories used for wetland based versus forest based versus biodiversity-based protocols?

Finally, to go beyond a descriptive assessment and explore whether protocols formed statistically distinct groups based on which “what” or “how” they measured things, we ran a nonmetric multidimensional scaling ordination followed by a clustering technique in R. In each category, protocols were given a “1” to indicate that the protocol includes a measurement in that category, and a “0” for when it does not. To find relationships between the various “what” and “how” categories, we used nonmetric multidimensional scaling (NMS), an ordination technique that aims to summarize the pattern of multivariate data onto as few axes as possible, while retaining as much variance, or information, as possible. Nonmetric multidimensional scaling explicitly projects the ecological distance, a measure of dissimilarity, of the samples into the new ordination space.

Once the NMS was run, we used a cluster analysis to analyze data for “what” and “how” category assemblages. Once groups were created, we analyzed both the methodologies that appeared in each group as well as the characteristics of each group to assess if there was any meaning in the cluster groups. For a detailed NMS and clustering methodology, see Appendix C.

RESULTS

This study sought to compare the metrics used by current mitigation quantification methodologies. We focused on two characteristics of the metrics. The first was "what" quantification methodologies measure. The second was "how" the quantification methodologies measured the metrics. We characterized "what" and "how" ecosystems and biodiversity were measured for 40 and 31 mitigation protocols, respectively, used in a diversity of programs ranging from wetland mitigation in the U.S. to koala trees in Australia (US ACOE 2010; Adamus et al. 2010; Dept. of Environmental Heritage and Protection 2014). A full list of the protocols analyzed can be found in Appendix A. The protocols varied in the resources for which mitigation was being required or recommended (i.e., biodiversity, wetland loss, species populations, etc.), when the methodology was developed, regulatory authority or program for which it was developed, and geographic location where it is applied.

We assessed trends in "how" and "what" protocols measured, both over time and among different types of ecosystem mitigation programs. We attempted to determine if there was any convergence among methodologies in what they were measuring. Finally, we assessed how well the methodologies analyzed conformed to the best practice recommendations described in the literature.

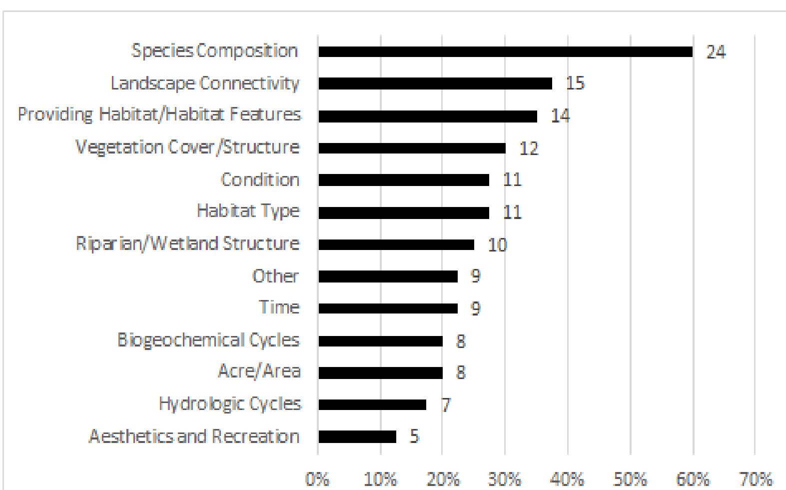
For 37 of the 40 protocols we were able to identify the year the methodology was adopted. The methodologies analyzed were developed over an almost 50-year period spanning from 1965–2014 with most protocols developed after 2005. For the "what" metric analysis, 17 protocols were developed for wetlands, for specific species, for forests or other vegetative categories, for biodiversity, and that measured multiple or other types of ecosystem mitigation programs. The protocols can also be categorized by whether they focused on terrestrial (23) or aquatic (17) systems, and whether they were applied in the United States (25) or other countries (15).

When analyzing "how" metrics, 13 protocols were developed for wetlands, 6 for specific species, 3 for forests or other vegetative categories, 5 for biodiversity, and 4 that measured multiple types or other types. Thirteen focused on aquatic systems, and 18 focused on terrestrial systems.

We found no significant pattern in the number of "what" measurement categories used in protocols over time. Additionally, none of the categories of metrics were more likely than another to be used as time progressed.

Across the 13 "what" categories identified, only one—species composition—was used in more than half of the methodologies. Species composition was utilized in 24 of the 40 protocols (Figure 1). Individual protocols had anywhere from one to nine different types of "what" measurements and most averaged between three and four categories.

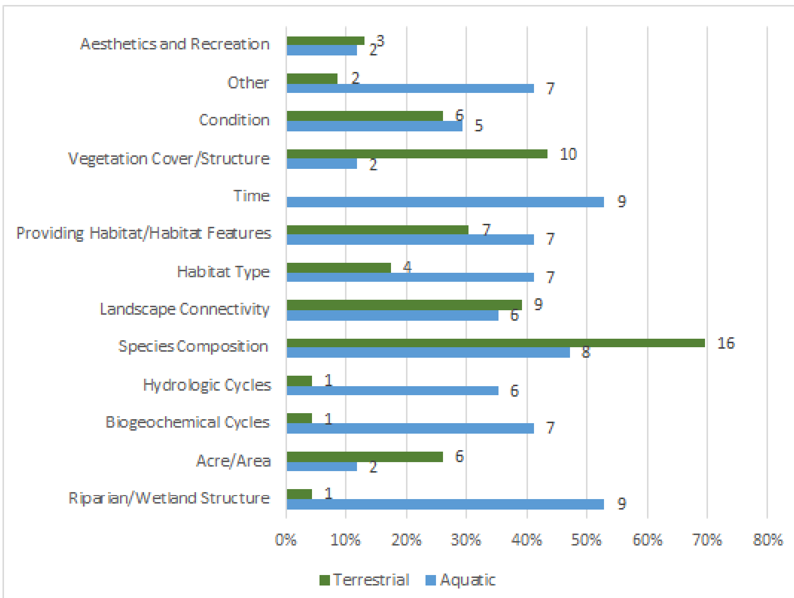
Figure 1: Percent of protocols that include each "what" measurement category. Numbers indicate how many protocols made up each percentage.



Unsurprisingly, a larger percentage of methodologies developed for aquatic resources included metrics for riparian/wetland structure and the hydrologic cycle, but they also had a larger percentage of protocols measuring biogeochemical cycles and the only ones measuring time lag for restored resources to mature. By contrast, methodologies developed for application in terrestrial systems had a higher percentage of metrics that measured acre/area, species composition, and vegetation cover/structure (Figure 2). Forty-six percent of terrestrial methodologies used measurements in up to two "what" categories, while 50% of aquatic protocols used measurements in five to six categories. As a whole, aquatic protocols measure a more diverse group of "what" metrics.

Because the set of U.S. methodologies reviewed were dominated by aquatic quantification methodologies, many of the differences identified between U.S. and international methodologies were attributable to the differences one might expect between aquatic and terrestrial metrics. Effects of the aquatic versus terrestrial split were also seen in the categories used by wetlands/estuary/marine, species, forest/vegetation, biodiversity, and other protocols; biogeochemical and hydrological cycles metrics were measured exclusively by wetlands/estuary/marine protocols. Interestingly, time required for offsets to mature was a metric unique to wetlands/estuary/marine protocols.

Figure 2: Percentage of protocols using each "what" category divided into aquatic and terrestrial protocols. Numbers indicate how many protocols made up each percentage.



Qualitative measurements were found to differ significantly between aquatic and terrestrial systems. Nonmetric multidimensional scaling and cluster analysis were used to determine if additional groupings of the protocols could be identified (Appendix C). No distinct groups were revealed around what methodologies measured, not even a simple two-group distinction between aquatic and terrestrial protocols. The lack of grouping indicates there are no major patterns in metric groupings for the different methodologies.

We found no significant pattern in the different ways ("how") quantification methods measured different factors. Additionally, none of the "how" categories were more likely to be used as time progressed. Some protocols used only one "how" measurement category while others included multiple approaches. Of all measurement approaches used, weighted metrics and counts were the most common, appearing in 20 and 21 of the 31 protocols, respectively (Figure 3, next page).

Figure 3: Percent of protocols using each "how" measurement category. Numbers indicate how many protocols made up each percentage.

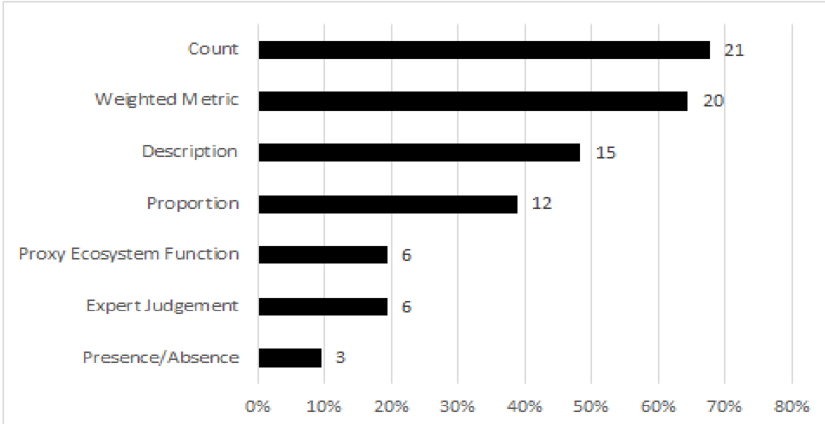
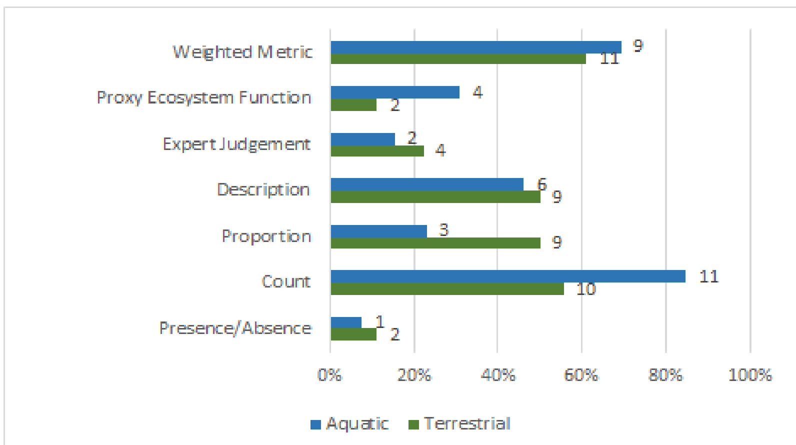


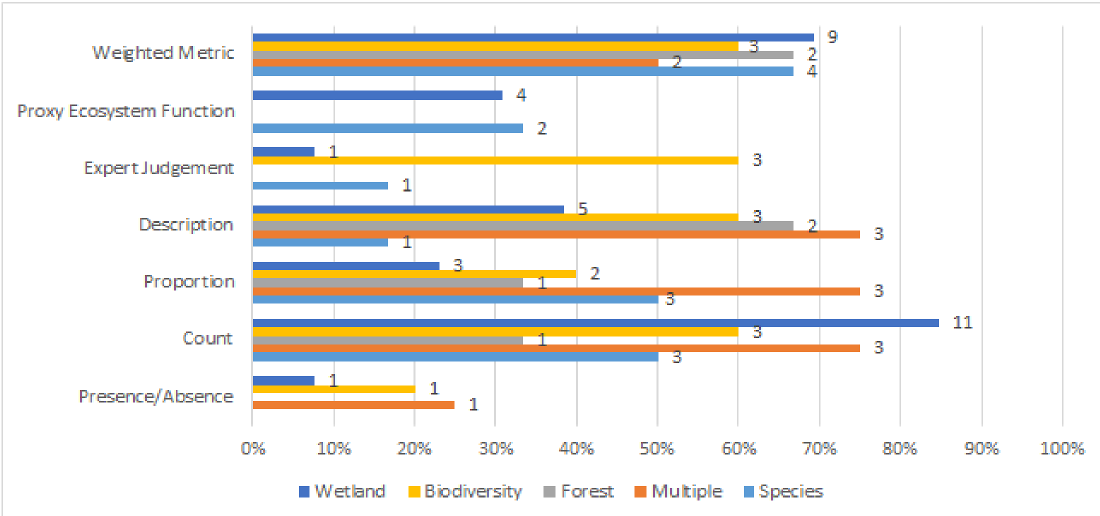
Figure 4: Comparisons of "how" measurement categories used between terrestrial and aquatic protocols. Numbers indicate how many protocols made up each percentage.



Once again, the "how" measurement categories used were, as one might expect, different for terrestrial and aquatic protocols (Figure 4). Aquatic methodologies utilized a higher percentage count, proxy ecosystem function, and weighted metrics, and terrestrial projects used a higher percentage of proportion metrics. The use of other categories was similar across aquatic and terrestrial programs. Once again, differences in how metrics were measured between U.S. and international methodologies were largely attributable to differences between terrestrial and aquatic methodologies.

We also grouped methodologies into one of four categories—biodiversity, forest/vegetation, wetlands/estuary/marine, species, and protocols that used multiple categories—and compared how metrics were measured across the different categories. Certain project types, such as species-specific initiatives, used nearly all "how" measurement types. Weighted metrics were used by all project types, as were descriptive, proportion, and count measures. (Figure 5, next page).

Figure 5: Comparisons of "how" measurement category use between protocol types. Numbers indicate how many protocols made up each percentage.



As with the "what" measurement categories, the NMS and Custer analysis found some groupings, but these did not follow any clear logic for what might distinguish the measurement approach used in protocols, such as geographic location where the methodology is applied, ecosystem type for which the tool was designed, or regulatory program for which the methodology was developed.

DISCUSSION

This review reinforces the findings of previous studies, which have concluded that existing quantification methodologies vary widely and there is little consistency across different approaches (Tallis 2015; Naeem 2015). While the quantification tools analyzed here were developed across five decades, neither "what" they measured nor "how" they measured different metrics showed any growing consistency. Our analysis suggests that as new methodologies are developed, they are not beginning to adhere to a set of best practices nor are those developing these tools converging toward a set of commonly accepted metrics.

No one set of metrics is likely appropriate for all types of mitigation programs, which may explain why this analysis shows significant variation in both "what" metrics are utilized and "how" these metrics are measured. However, this variation in measurement may not be deliberate as it does not seem to follow clear patterns, except for distinguishing aquatic and terrestrial systems. Our analysis of what is measured does not show any clustering by the type quantification methodology (e.g., wetlands/streams), date of adoption, regulatory authority for which the methodology was developed, or geography for which the tool was developed. And while the "how" categories did statistically cluster there was no clear link to type of methodology. If quantification methodologies designed to measure impacts and offsets for particular habitat types are not consistent with others designed for the same habitat type and cannot be differentiated from other types of methodologies by what they are measuring (e.g., wetlands protocols are not distinct from coastal marsh protocols), then the implication is that the methodologies are not being tailored to the specific goals of the program. We suspect that the variation across and among methodology types results from there being an absence of clear standards or best practice examples that are followed as new quantification methodologies are developed.

The majority of the reviewed methodologies also do not adhere to the best practices described in the literature (see References). For the "what" is measured categories, recommendations that deal specifically with perceived ecosystem value of certain species assemblages, "species composition," were followed the most often, by 60% of the protocols (Table 4). By contrast, only 38% used connectivity as a metric, and uncertainty of mitigation outcomes and risk of mitigation failure were rarely incorporated as separate metrics. Time until the mitigation offset equals the development was measured by just under a fourth of the protocols.

Table 4: Comparisons between "what" the literature recommends measuring, versus what current protocols are actually measuring

What *should* be measured?	What *is* measured.
Connectivity of important habitat	38% of protocols.
Perceived ecosystem values for species assemblages and land cover type	60% measure species composition, 35% habitat and habitat features, 28% habitat type.
Ecosystem function	Yes, spread out across what categories.
Ecosystem Services	Aesthetics and recreation were used in just over 16% of the protocols.
Risk of mitigation failure	Rarely.
Time required for offsets to equal development	23% of protocols.

When analyzing which of the "how" to measure guidelines were followed, less than half the protocols followed recommendations (Table 5). Noting that one protocol can follow multiple best practices, proxy measurements of ecosystem function was included in 19% (6 total) of protocols and almost 50% of the protocols (15 total) used descriptive assessments of ecosystem importance of species and habitat types. Aesthetics and recreation were the only ecosystem services specifically measured by five protocols. In general ecosystems services benefiting human populations were noted as co-benefits that improved a "what" metric, but was not its sole purpose for inclusion in the mitigation protocol. So, despite recommendations that ecosystem services considering impacts to people be included, we did not see that reflected in the protocols we reviewed.

Table 5: Comparisons between "how" the literature recommends measuring, versus how protocols are actually measuring

How should measurements be done?	How measurements are done.
Proxy measurement of ecosystem function	19% of protocols.
Descriptive assessments of ecosystem importance of species and habitat types	48% of protocols.

Our analysis of the functional measurements used by the reviewed methodologies addressed mostly regulating services, such as nutrient and sediment levels associated with water (USACOE 2010), which may be relatively easy for programs to measure when compared to other services like cultural and aesthetic services (US Fish and Wildlife Service 1980; Queensland Government n.d.; USACOE 2010).

CONCLUSION

In our review of mitigation quantification methodologies, we found no discernable pattern or change over time in "what" is measured and "how" it is measured, nor did we find a clear pattern when we analyzed methodologies designed to measure the same type of resources (e.g., wetland or species). Although several authors have offered recommendations for what and how quantification methodologies should measure to best ensure that mitigation programs meet their conservation goals (e.g., no net loss), such recommendations are inconsistently followed and there is no indication that the degree to which they are followed is increasing over time. For example, we saw no consistent increase in the use of functional or connectivity measures over time. In fact, the only pattern that emerges when we looked across all of the methodologies is that, quite logically, aquatic programs used different measurement categories than terrestrial protocols.

As a practical matter, the lack of consistency across methodologies (i.e., what constitutes a credit) makes it difficult to compare whether offsets from different regions or programs are meeting shared conservation objectives and how robust different mitigation markets are (i.e., how many credits are being generated in different regions).

The lack of consistency and progression toward a common set of metrics and approaches suggests limited sharing of best practices. It also suggests that practitioners developing different quantification tools are either unaware of the literature on best practices or find it lacking. In addition, the lack of consistency across mitigation quantification methods, adherence to best practices, or dearth of well-accepted best practices may also undermine confidence in the scientific rigor of these methodologies and the ability of mitigation programs more generally to support conservation objectives. Our findings also suggest that there may be a significant amount of “reinventing the wheel” in development of mitigation quantification methodologies. Given the lack of clearly accepted standards and no evidence of shared learning across programs over time, we suggest that there would be significant benefit to investing in an effort for actors in the mitigation community (government, field practitioners, private sector) to develop collaboratively a common set of best practices for the process of developing and substance of robust, reasonably precise mitigation quantification methodologies.

Answering the question about how much precision is needed to achieve conservation objectives will be important to inform the development of best practice guidance. Although the desire to quantify impacts and offsets with a high degree of precision is laudable, the evidence that doing so better supports the achievement of conservation objectives is unproven. We suggest that there would be great value in studying whether more complex methodologies and those that strive to measure functions and services directly are more or less likely to achieve intended conservation outcomes than those that are easier to apply or use proxy measures, indirect measures, or condition to quantify impacts and offsets in combination with conservative factors to account for risk and uncertainty.

REFERENCES

Alagona, P., and S. Pincetl. 2008. "The Coachella Valley Multiple Species Habitat Conservation Plan: A Decade of Delays." *Environmental Management* 41: 1–11.

Benayas, J.M.R., A.C., Newton, and J.M. Bullock. 2009. "Enhancement of Biodiversity and Ecosystem Services by Ecological Restoration: A Meta-Analysis." *Science* 325: 1121.

BenDor, T. and J. Riggsbee. 2011. "A Survey of Entrepreneurial Risk in U.S. Wetland and Stream Compensatory Mitigation Markets." *Environmental Science and Policy* 14(3): 301–314.

Bruggeman, D. and M. Jones. 2008. "Should Habitat Trading Be Based on Mitigation Ratios Derived from Landscape Indices? A Model-Based Analysis of Compensatory Restoration Options for the Red-Cockaded Woodpecker." *Environmental Management* 42(4): 591–602.

Business and Biodiversity Offsets Programme. 2012. Standard on Biodiversity Offsets. Washington, D.C.: BBOP.

Cochran, B., N. Robinson Maness, and E. Alcott. 2011. "Measuring Up: Synchronizing Biodiversity Measurement Systems for Markets and Other Incentive Programs." Hillsboro, Oregon: Willamette Partnership.

Compensatory Mitigation for Losses of Aquatic Resources. 2008. Federal Register 73(70) (to be codified at 40 CFR pts. 230).

Gardner, T., et al. 2013. "Biodiversity Offsets and the Challenge of Achieving No Net Loss." *Conservation Biology* 27(6): 1254–1264.

Goncalves, B. et al. 2015. "Biodiversity offsets: from current challenges to harmonized metrics." *Current Opinion in Environmental Sustainability* 14: 61–67.

Gordon, A. et al. 2011. "Assessing the impacts of biodiversity offset policies." *Environmental Modelling & Software* 26: 1481–1488.

Kiesecker, J. M., H. Copeland, A. Pocewicz, and B. McKenney. 2009. "Development by Design: Blending Landscape-Level Planning with the Mitigation Hierarchy." *Frontiers in Ecology and the Environment* 8(5): 261–266.

King, D. M., and E. W. Price. 2004. Developing Defensible Wetland Mitigation Ratios. Solomons Island, MD.

Land and Water Conservation Fund. n.d. "Land and Water Conservation Fund." U.S. Department of the Interior. Retrieved from: https://www.doi.gov/lwcf.

Madsen, B., N. Carrol, D. Kandy, and G. Bennett. 2011. Update: State of Biodiversity Markets. Washington, DC: Forest Trends.

Maechler, M. et al. 2014. Cluster: Cluster Analysis Basics and Extension. R package version 1.15.2.

Maron, M. et al. 2012. "Faustian bargains? Restoration Realities in the Context of Biodiversity Offset Policies." *Biological Conservation* 155: 141–148.

McKenney, B., and J. Wilkinson. 2015. "Achieving Conservation and Development: 10 Principles for Applying the Mitigation Hierarchy." The Nature Conservancy.

Moreno-Mateos, D. et al. 2015. "The True Loss Caused by Biodiversity Offsets." Biological Conservation 192.

Naeem, S., et al. 2015. "Get the Science Right When Paying for Nature's Services." *Science* 347(6227): 1206–1207. doi: 10.1126/science.aaa1403

OECD 2014. "Biodiversity Offsets: Effective Design and Implementation." The Organisation for Economic Cooperation and Development. Retrieved from: http://www.oecd.org/env/resources/Biodiversity%20Offsets_Highlights_for%20COP12%20FINAL.pdf

Olander, L. et al. 2018. "Benefit Relevant Indicators: Ecosystem Services Measures That Link Ecological and Social Outcomes." *Ecological Indicators* 85: 1262–1272. https://doi.org/10.1016/j.ecolind.2017.12.001

Oksanen, J. e. a. 2015. Vegan: Community Ecology Package. R package version 2.2-1. Quetier, F. and S. Lavorel. 2011." Assessing Ecological Equivalence in Biodiversity Offset Schemes: Key Issues and Solutions." *Biological Conservation* 155: 2991–2999.

R Core Team. 2014. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Salzman, S. and J.B. Ruhl. 2006. "'No Net Loss' – Instrument Choice in Wetlands Protection," in *Moving to Markets in Environmental Regulation: Lessons from Twenty Years of Experience* (Jody Freeman and Charles D. Kolstad eds.)

Tallis, H., C.M. Kennedy, M. Ruckelshaus, J. Goldstein, and J.M. Kiesecker. 2015. "Mitigation for One & All: An Integrated Framework for Mitigation of Development Impacts on Biodiversity and Ecosystem Services." *Environmental Impact Assessment Review* 55(0): 21–34. doi: http://dx.doi.org/10.1016/j.eiar.2015.06.005

Turner, R.E., A.M. Redmond, and J.B. Zedler. 2001. Count IT by Acre of Function – Mitigation Adds Up to Net Loss of Wetlands. *National Wetlands Newsletter* 23(6): 5–6, 14–16.

United States Army Corps of Engineers. 2008. Compensatory Mitigation for Losses of Aquatic Resources; Final Rule. Washington, DC: Federal Registry Retrieved from http://www.usace.army.mil/Portals/2/docs/civilworks/regulatory/final_mitig_rule.pdf.

Wilkinson, J. 2016. "The Third Wave: Mitigation as a Mechanism for Supporting Public Land Management Decision-Making." *Public Land and Resources Committee Newsletter*. 13(1). American Bar Association or the Section of Environment, Energy, and Resources.

Wilkinson, J., J. Raepple, R. Rossner, A. Singler, T. Moberg, K. Kennedy, and S. Harold. 2017. "Environmental Markets and Stream Barrier Removal: An Exploration of Opportunities to Restore Freshwater Connectivity Through Existing Mitigation Programs." The Nature Conservancy.

Willamette Partnership. 2011. Measuring Up: Synchronizing Biodiversity Measurement Systems for Markets and Other Incentive Programs. U.S. Department of Agriculture, Office of Environmental Markets. Retrieved from: http://willamettepartnership.org/wp-content/uploads/2015/04/Measuring-Up-w-appendices-final.pdf

APPENDIX A: LIST OF PROTOCOLS ANALYZED

Table A1: List of protocols used in analysis

Year	Terrestrial /Aquatic	Bibliography
1965	Terrestrial	Species Banking. 2010. Forest code offsets (Brazil). Ecosystem Marketplace Network. http://www.speciesbanking.com/program/forest_code_offsets_cdigo_florestal
1980	Terrestrial	U.S. Fish and Wildlife Service. 1980. Habitat Evaluation Procedures. http://www.fws.gov/policy/esm102.pdf
1991	Aquatic	NOAA. 1991. Southern California eelgrass mitigation policy. http://www.westcoast.fisheries.noaa.gov/publications/habitat/california_eelgrass_mitigation/eelpolrev11_final.pdf
2002	Aquatic	Strange, E. et al. 2002. Determining Ecological Equivalence in Service-To-Service Scaling of Salt Marsh Restoration. Environmental Management 29(2): 290–300.
2004	Terrestrial	Delfs, C. 2004. A Citizen's Guide to the Forest Conservation Act in Maryland. Chesapeake Bay Foundation. https://www.cbf.org/document-library/cbf-guides-fact-sheets/Citizens-Guide-to-Forest-Conservationcd1a.pdf
2005	Terrestrial	National Fish and Wildlife Foundation. n.d.. Acres for America. http://www.nfwf.org/acresforamerica/Pages/home.aspx#.VETrlPldWuk; http://www.nfwf.org/acresforamerica/Documents/acres_2016_rfp.pdf
2006	Terrestrial	Texas A&M Institute of Renewable Resources. n.d. What Is the Golden-Cheeked Warbler Recovery Credit System? http://rcs.tamu.edu/what-is-it/
2007	Terrestrial	Brownlie, S. et al. 2007. Provincial Guideline on Biodiversity Offsets: Revised Draft—March 2007. Provincial Government of the Western Cape: Department of Environmental Affairs and Development Planning. https://www.westerncape.gov.za/text/2007/3/pgwcoffsetsguidelinedraft_5march_07.pdf
2007	Terrestrial	Department of Environment & Climate Change NSW. 2007. BioBanking: Biodiversity Banking and Offsets Scheme. https://www.environment.nsw.gov.au/topics/animals-and-plants/biodiversity/biobanking
2007	Terrestrial	Native Vegetation Council & Government of South Australia. 2007. Guidelines for a Native Vegetation Significant Environmental Benefit Policy for the Clearance of Scattered Trees. http://webcache.googleusercontent.com/search?q=cache:9SEIEB39K14J:https://www.environment.sa.gov.au/files/f8b2bce6-af07-4361-8c5a-9efa00f8d78a/con-nv-guideline-sebscatteredtree.pdf+&cd=1&hl=en&ct=clnk&gl=us
2007	Aquatic	USAC. 2007. Unified Stream Methodology. http://www.deq.virginia.gov/Portals/0/DEQ/Water/WetlandsStreams/USMFinal_01-18-07.pdf
2008	Terrestrial	Malua BioBank. n.d. The Malua Biobank. http://www.maluabiobank.com/explore.php?id=The_Malua_BioBank
2009	Terrestrial	Gibbons, P. et al. 2009. An Operational Method to Assess Impacts of Land Clearing on Terrestrial Biodiversity. Ecological Indicators. 9(1): 26–40.
2009	Terrestrial	U.S. Fish and Wildlife Service. 2009. Guidelines for the Establishment, Management, and Operation of Gopher Tortoise Conservation Banks. http://www.fws.gov/MississippiES/pdf/USFWSGopherTortoiseBankGuidance_27Jan2009.pdf
2009	Aquatic	Jacobs, A.D., D.F. Whigham, D. Fillis, E. Rehm, and A. Howard. 2009. Delaware Comprehensive Assessment Procedure Version 5.2. Delaware Department of Natural Resources and Environmental Control, Dover, DE 72pp. http://www.dnrec.delaware.gov/Admin/DelawareWetlands/Documents/DECAP%20Protocol%20v5.2%20July%2009.pdf
2009	Aquatic	Willamette Partnership. 2009. Ecosystem Credit Accounting: Pilot General Crediting Protocol, Willamette Basin Version 1.0. www.willamettepartnership.org
2009	Terrestrial	Willamette Partnership. 2009. Ecosystem Credit Accounting: Pilot General Crediting Protocol, Willamette Basin Version 1.0. www.willamettepartnership.org

Year	Terrestrial /Aquatic	Bibliography
2009	Aquatic	Primozich, D. 2009. Ecosystem Credit Accounting. Willamette Partnership. http://willamettepartnership.org/wp-content/uploads/2014/09/procedure-for-shadelator-v6.pdf
2009	Terrestrial	City of Bainbridge Island. 2009. BBOP Pilot Project Case Study. Bainbridge Island. Washington State, USA. http://www.forest-trends.org/documents/files/doc_3120.pdf
2009	Terrestrial	Hankla, D., and P. Souza. 2009. Florida Scrub-Jay Mitigation Guidance. U.S. Fish and Wildlife Service, North Florida Ecological Services Office. http://www.fws.gov/northflorida/Scrub-Jays/fsj_mit_guide.htm
2010	Aquatic	Adamus, P., J. Morlan, and K. Verble. 2010. Manual for the Oregon Rapid Wetland Assessment Protocol (ORWAP). Version 2.0.2. Oregon Dept. of State Lands, Salem, OR. http://www.oregon.gov/dsl/WETLAND/docs/orwap_guide.pdf; https://www.oregon.gov/dsl/WW/Documents/ORWAP_3_1_Manual_Nov_2016.pdf
2010	Aquatic	U.S. Army Corps of Engineers, New England District, Regulatory Division. 2010. New England District Compensatory Mitigation Guidance. https://www.justice.gov/sites/default/files/enrd/pages/attachments/2016/09/28/us_v._fkt_resort_management_llc_et_al_notice_of_lodging_with_proposed_cd_part_2.pdf
2010	Terrestrial	Kingborough. 2010. Biodiversity offset policy. http://www.kingborough.tas.gov.au/webdata/resources/files/Biodiversity_Offset_Policy.pdf
2011	Terrestrial	Government of Western Australia. 2014. WA Environmental Offsets Guidelines. http://www.epa.wa.gov.au/sites/default/files/Policies_and_Guidance/WA%20Environmental%20Offsets%20Guideline%20August%202014.pdf
2012	Terrestrial	Department for Environment Food and Rural Affairs. 2012. Biodiversity Offsetting Pilots: Guidance for Offset Providers. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/69530/pb13742-bio-guide-offset-providers.pdf
2012	Terrestrial	U.S. Fish and Wildlife Service. 2012. Panther Habitat Assessment Methodology. http://www.fws.gov/verobeach/MammalsPDFs/20120924_Panther%20Habitat%20Assessment%20Method_Appendix.pdf
2012	Aquatic	U.S. Army Corps of Engineers. 2012. "Corps of Engineers, New Orleans District Modified Charleston Method." http://www.mvn.usace.army.mil/Portals/56/docs/regulatory/2012MVN_MCMGuidebook_2_09_2012.pdf
2013	Aquatic	US Army Corps of Engineers: South Pacific Division. 2013. Regulatory Program Standard Operating Procedure for Determination of Mitigation Ratios. https://www.spa.usace.army.mil/Portals/16/docs/civilworks/regulatory/Mitigation/SOP%20for%20Determination%20of%20Mitigation%20Ratios.pdf
2013	Terrestrial	Victorian Government Department of Environment and Primary Industries. 2013. Native Vegetation Gain Scoring Manual. http://www.depi.vic.gov.au/__data/assets/pdf_file/0005/198968/Gain_manual_NVR.pdf
2013	Aquatic	NOAA. 2013. NOAA Discussion Draft: Coral Propagation and Active Species Enhancement Program. https://www.yumpu.com/en/document/view/39928285/noaa-mitigation-proposal-port-everglades-feasibility-study
2013	Aquatic	Government of Canada. 2013. Fisheries Productivity Investment Policy: A Proponent's Guide to Offsetting. Fisheries and Oceans Canada. http://www.dfo-mpo.gc.ca/pnw-ppe/offsetting-guide-compensation/index-eng.html
2013	Aquatic	U.S. Army Corps of Engineers. 2013. State of Missouri Stream Mitigation Method. http://www.mvr.usace.army.mil/Portals/48/docs/regulatory/Permits/MissouriStreamMitigationMethodAmendedApril2013.pdf
2014	Terrestrial	Government of New Zealand. 2014. Guidance on Good Practice Biodiversity Offsetting in New Zealand. http://www.doc.govt.nz/Documents/our-work/biodiversity-offsets/the-guidance.pdf
2014	Terrestrial	Department of Environment and Heritage Protection. 2014. Queensland Environmental Offsets Policy (Version 1.0). Queensland Government. https://www.ehp.qld.gov.au/assets/documents/pollution/management/offsets/offsets-policy.pdf

Year	Terrestrial /Aquatic	Bibliography
2014	Terrestrial	The Habitat Exchange. 2014. Lesser Prairie-Chicken Habitat Quantification Tool: A Multi-Scaled Approach for Assessing Habitat Conditions for Lesser Prairie Chicken. http://www.fws.gov/southwest/es/Documents/R2ES/LPC_Guidelines_for_LPC_Mitigation_Lands_Dec2014.pdf
2014	Terrestrial	Department of Environment and Heritage Protection. 2014. Community Infrastructure Offsets for Koala Habitat. Queensland Government. http://www.ehp.qld.gov.au/wildlife/koalas/legislation/pdf/fs-community-infrastructure-offsets-koala-habitat.pdf
2014	Aquatic	City of Washington, D.C. 2014. Stormwater Retention Credit Trading Program. http://green.dc.gov/src
n.d.	Aquatic	Queensland Government. n.d. Marine Fish Habitat Offset Policy. https://www.daf.qld.gov.au/__data/assets/pdf_file/0003/68601/Marine-Fish-Habitat-Offset-Policy-12.pdf
n.d.	Aquatic	NC DENR. n.d. 401 & Buffer Permitting Unit. Division of Water Resources. http://portal.ncdenr.org/web/wq/nutrientbuffermitigation; http://deq.nc.gov/about/divisions/mitigation-services/dms-customers
n.d.	Aquatic	Florida Department of Environmental Protection. n.d. Uniform Mitigation Assessment Method. http://www.dep.state.fl.us/water/wetlands/mitigation/umam/index.htm; http://sfrc.ufl.edu/ecohydrology/UMAM_Training_Manual_ppt.pdf

APPENDIX B: CATEGORIES USED IN ANALYSIS

Table B1: "What" categories and example measurements within each category

"What" Category	Protocols
Riparian/Wetland structures	Presence of streams and their buffers; instream structures; riparian buffer; location of fish passage barriers; percent of riparian buffer to stream bank; increasing shoreline complexity; riverbank stabilization; re-vegetation of riparian areas; wetland physical attributes; creation or expansion of natural stream channels.
Area/Acres	Area of suitable habitat; area; acres; hectares.
Hydrological cycle	Hydrologic connection; volume; groundwater flow or discharge; water-level fluctuations; water storage and delay.
Biogeochemical cycles	Export of organic carbon; removal of elements and compounds; bioremediation; dissipation of energy; retention of particulates; sediment erosion; cycling of nutrients; carbon sequestration; thermoregulation.
Species Composition	Habitat supports species rarely found in other habitats; habitat diversity; relative rarity of components; invasive and noninvasive species; rare species; "special" habitats; native species; nationally threatened species.
Connectivity	Connectivity; habitat that enhances biological corridors; proximity to known populations; hydrologic connection; distance from project; distance from ocean connection; proximity to natural grass or shrub cover; native vegetation providing key landscape linkages or buffering; located in priority area; spatial orientation of patches.
Habitat type	Landcover type; nationally threatened ecosystems; soil type; nontidal wetlands; ecosystem type; floodplain; contiguous forest; coastal bays and buffers.
Providing habitat/Habitat features	Increasing structure through the placement of course material or large debris; habitat refugia; spawnable substrate; number of trees with hollows; removal of barriers to migration; food provisioning; habitat provisioning; total length of fallen logs; potential threatened species habitat; native trees with known potential for nesting hollows; native trees with known potential for nesting hollows; bird nesting.
Time for offsets to equal development	Temporal lag; years of loss; project lifespan; time to maturity; years to complete rehabilitation; species recovery time.
Vegetation cover/Structure	Herbaceous cover; shrub cover; percent cover of above-ground vegetation; percent of vegetative cover depending on habitat; percent overstory; native upper and midstory cover; native vs. exotic ground cover; dominant strata structure; canopy architecture of vegetation; biomass and stem density; grass height; basal area, height.
Condition	Habitat conditions; current condition; hydrologic condition (affected to unaffected); health; condition of vegetation; level of depletion of vegetation type; condition of affected habitat.
Aesthetics and recreation	Recreation value; aesthetic value; area or features of interest.
Other	Management activities; reasonable or foreseeable effects; mitigation risk; size-by-species distribution; disturbance regulation; type of impact (clearing, draining, etc.); benthic algae; macrophyte detritus; biomass; biological control; protected area; plant flowering; local importance.

Table B2 "How" categories and example measurements within each category

"How" Category	Protocols
Counts	200 acres of contiguous scrub habitat; patches of scrub habitat at least 25 acres in size; critical habitat; total length of falling logs; years of loss; area of proposed rehabilitation/offset; years to complete rehabilitation; hectares; area of features of interest (e.g., bat colonies); acreage of suitable habitat; habitat area; herbaceous ground cover; total canopy cover; elevation; aspect; wetted width; bankfull width; predominant length; channel incision; riparian buffer; temporal lag; length of disturbance from development; area of beds; density of beds; number of rare, threatened, and endangered species; instream structures, habitat structures, bankfull bench creation, lay back bank, bioremediation, streambank plantings, grass height, preventing nutrient loading, area of riparian buffer; eucalyptus DBH or canopy spread; native trees with nesting hollows; richness of plant species; vegetation; hydroperiods; volume retained; time to maturity; project lifespan.
Presence/Absence	Slope; vegetation listed as rare, threatened, or endangered; DBH; nationally endangered or threatened species; fish and wildlife utilization.
Proxy measurement of ecosystem function	Spatial orientation of patches so that demographic isolation doesn't occur; successful dispersal of scrub jays between habitat patches; size/frequency distributions; disturbance regulation; habitat refugia; food production; recreation; water storage and delay; phosphorus retention; nitrate removal; thermoregulation; carbon sequestration; organic matter export; aquatic invertebrate habitat; anadromous fish habitat; nonanadromous fish habitat; amphibian and reptile habitat; waterbird feeding habitat; waterbird nesting habitat; songbird, raptor, and mammal habitat; pollinator habitat; native plant diversity; elevation; aspect; wetted width; bankfull width, predominant length; channel incision; proximity to natural grass and shrub landcover.
Proportion	Percent of vegetative cover depending on habitat; groundcover; percent overstory; species count; soil type; basal area; canopy; shrub cover; herbaceous cover; rarity of vegetation type; level of depletion of vegetation type; condition of native vegetation; richness of biodiversity; native upper and midstory cover; native ground cover; exotic plant cover; overstory species occurring as regeneration; habitat conditions; riparian buffer; density of beds; canopy cover; fragmenting features; invasive and nonnative species; shading.
Expert judgment	Condition of affected habitat; presence of threatened species; presence of special habitats; biodiversity process value of affected habitat; compatibility with adjacent land uses and other management plans; reasonable or foreseeable effects; landcover type; richness of biodiversity; current condition; hydrologic connection; location; mitigation risk; floral and faunal components.
Descriptions	Floodplain; politically identified priority areas; trees associated with historic site or structure; condition of affected habitat; presence of threatened or endangered species; presence of special habitats; biodiversity process value of affected habitat; importance of biodiversity underpinning valued ecosystem services; the degree to which a habitat supports species rarely found in other habitats; nationally threatened ecosystems; hydrological conditions; soil characteristics; habitat type; connectivity score; connects remnants; located in a priority area; habitat provisioning; food provisioning; biodiversity of trees; vulnerability; replicability; aesthetic value; relative rarity of components; health of wetland; wetland physical attributes; sediment retention; dominant strata structure; existing vegetation type and cover as a proxy of site for providing habitat; hydrologic condition; type of impact; native vegetation providing key landscape linkages or buffering; within or directly adjacent to a known forty-spotted pardolate colony; hydrological, sediment; special habitat features; corridors.

"How" Category	Protocols
Weighted metrics	Percent of forest land out of total development site determines whether mitigation needs to occur; area; condition of habitat; groundcover; native plant species richness;percent overstory; landcover type; number of gopher tortoises already in residence; connectivity score; native upper and midstory cover; native ground cover; exotic ground cover; number of trees with hollows, proportion of overstory species occurring as regeneration; total length of fallen logs; proximity to existing vegetation, species; height; hollows; density; abundance or scarcity; vulnerability; replaceability; aesthetic value; current condition; hydrologic connection; location; time lag; mitigation risk; floral and faunal components; species-by-size distribution; species recovery time; time to maturity; project lifespan; disturbance regulation; biological control; habitat refugia; food production; recreation; protected area; areas or features of interest (e.g., bat colonies); acreage of suitable habitat; water storage and delay; sediment retention; phosphorus retention; nitrate removal; thermoregulation; carbon sequestration; organic matter export; aquatic invertebrate habitat; anadromous fish habitat; nonanadromous fish habitat; amphibian and reptile habitat; waterbird feeding habitat; songbird, raptor, and mammal habitat; pollinator habitat; native plant diversity; public use and recognition; elevation; riparian buffer; ecosystem type; existing vegetation type and cover used as a proxy for condition of site for providing wildlife habitat; hydrologic condition; length of disturbance from development; type of impact; temporal lag; length of stream in project; percent of riparian buffer to stream bank; number of trees; native vegetation providing key landscape linkages or buffering; known threatened species habitat.

APPENDIX C: STATISTICAL ANALYSIS METHODS

To analyze whether any natural groups of protocols formed based on which "what" and "how" categories were used, we used R (R Core Team, 2014) statistical software to run both a nonmetric multidimensional scaling analysis as well as a cluster analysis. The steps for both are below.

Data Preparation for NMS and Clustering

We relativized datasets by row sums and columns. By relativizing by column, we are relativizing by species maximum, so that every species varies from 0 to 1. By doing this, all the species are the same in terms of how they contribute to the statistics, and there is a level playing field for all species. Additionally, we are also relativizing by row, or sample sum. This converts the values to proportions between 0 and 1, which makes the plots that are similar in species composition the same.

Methods: Nonmetric Multidimensional Scaling

Nonmetric multidimensional scaling has no assumptions. However, the distance measure used to project the multivariate swarm of points into ordinational space will have a great deal of impact on the results of the NMS, as certain distance measures can cause distortion. In this study, we have used Bray-Curtis extended distance as a distance measure, as this is the best distance measure for species data. Importantly, even when the points have been projected in ordination space they remain in least rank order.

In NMS, in order to achieve the results, there is not an eigen solution, but successive numeric approximations of the fit between ecological distances and ordination distances. Samples are moved in ordination space until the fit can improve no more, calculated as a level of "stress" or "badness of fit." This process is repeated through numerous iterations to reduce stress. To calculate stress, researchers sum the squared deviations, similar to a regression residual. To determine the number of NMS axes, I chose the lowest number of axes that would still yield a stress level near 0.15 or below.

We used the Vegan package (Oksanen et al. 2015) to build an extended Bray-Curtis distance matrix. Using Ecodist (Goslee and Urban 2007), we used the nmds call to create 10 ordinations for all dimensions from 1 to 6. Using this simple function we have created the NMS, but it still requires a great deal of refining and then postprocessing.

The multiple ordinations in the NMS work through different fit options for the NMS, and the more configurations attempted the more likely that the configuration with the minimum stress levels, the optimum, will be tried. To find the best configuration of the axes, it is important to calculate stress levels, or "badness of fit" of the relationship between the ecological and ordinational distances, for the ordinations of dimensions 1–6. After calculating the stress levels for the dimensions using a step-down procedure, I created a scree plot for both datasets to visualize the stress levels plotted against the number of dimensions, as well as the R2 plotted against the dimensions. The R2 represents the relationship between the two distance matrices, done as a regression, and the higher the R2, the higher number of sample points that are captured by the regression.

To have no stress (and thus an R2 of 100%) we would have to use all dimensions, which is not interesting as this would just reproduce the data. Instead, we visually chose the dimensions that reduce the most stress and maximize R2; a natural break point. For both the binomial and abundance datasets we used three NMS axes.

Once we decided on the number of dimensions to retain, we re-ran the nmds function with the selected number of dimensions and 20 iterations instead of 10, to make sure it was as accurate as possible. When testing the stress level for the best fit, we extracted the configuration with the lowest stress level.

At this point, we do not know which of the three dimensions captures the most variance, nor if there is alignment with the axes and sample points. To fix this problem, we used principle component analysis (PCA) to rotate the axes. We ran Principle Component 1 through the longest dimension of the points, which also contains the most variance, and Principle Component 2 to create an axis with the second highest variance, and so on. The NMS ordination is thus rotated.

We also calculated the R2 for each of the ordination axes. The first axis R2 can be calculated with Euclidean distances, but because the axes are calculated simultaneously the distances on axis two is dependent on axis one. To find the R2 for axis two, the R2 of axis one is subtracted from the total, and to find the R2 for axis three, the R2 of axis one and two are subtracted from the total. In this way we calculated the separate axes' R2 values.

Methods: Clustering

If clustering is run to conclusion, there will just be one large group. Thus, a researcher must choose where in the clustering process to stop analysis, either when there are only a handful of high-level groups, or a larger number of low-level groups. This part of the process is subjective and depends on the end-goal of the analysis.

Like NMS, clustering results depend on the choice of distance measure when defining the dissimilarity between groups, and also by the criteria used for the joining of groups when there are many samples within a single group. As we are using species data, we are again using extended Bray-Curtis distances as well as the average method of joining, described further in the Methods section below.

Using a Cluster package (Maechler et al. 2014) in R, we ran an agglomerative nesting function, which shows hierarchical clustering. As aforementioned, we used extended Bray-Curtis distance, as well as a dissimilarity method and the average linkage method, which calculates all the pairwise distances between two sample groups and joins them only if a specific percentage of the pairs meets a predetermined threshold value.

We created different levels of groupings which yielded cluster levels from 2 to 10 groups. To decide which level of groups to retain, we assessed the within-group variability versus the among-group variability using a Mantel's test and Mantel's correlations to reveal a curve that shows how many clusters we should retain for further analysis. In order to use Mantel's test we constructed a new distance matrix, and coded the elements in a binary 0, 1 system; 0 if two samples are in the same group, 1 if they are not in the same group. The new distance matrix is Euclidean and is calculated using absolute distances on cluster group membership. Using the Mantel's test, we calculated the strength of contrast between the among and within-group distances based on their Bray-Curtis dissimilarities. We repeated this process over several cluster levels to find the "best" level, and then ran the code to reveal our final groups across the three NMS axes.

NMS and Clustering: "What"

After calculating the NMS, we proceeded with just dimensions one through three, capturing over 75% of variability of the data points.

After rerunning the NMS with just three dimensions, the lowest stress level was calculated at 0.16; ideally we would like to have a stress level lower than 0.1, but four dimensions makes analysis much more complicated than three, and thus we retained three dimensions. The R2 is 0.83, and thus accounted for a large percentage of the variance in the sample points.

We calculated the agglomerative coefficient at 0.71, which measures the structure of the clustering of the dataset; the ideal is 1.0. We used the Mantel's test to decide how many groups to retain. Based on the returned Mantel values, the highest value is on six groups, and we proceeded with six groups, yielding a Mantel R of around 0.60.

When looking at NMS 1 vs. NMS 2 and NMS 2 vs. NMS 3 (Figure 5, Figure 6), Group 1 is clearly the largest, with Group 2 as the next largest, and Groups 3, 4, 5, and 6 smaller and spread out across NMS Axis 2. There is no clear separation between groups, though Group 1 and Group 2 appear to occupy different areas of the NMS. However, when looking at the protocols in Group 2, there is no obvious geographic, ecosystem, or any other type of link (Appendix D).

NMS and Clustering: "How"

The NMS and Cluster analysis calculates if there are groupings of the protocols based on which "how" measurement categories they use. Looking at the scree plots we used three dimensions in the NMS ordination, capturing over 90% of the variability of the data points.

After rerunning the NMS with just three dimensions, the lowest stress level was calculated at 0.10; The R2 is 0.94, and thus accounted for a large percentage of the variance in the sample points. We calculated the agglomerative coefficient at 0.86, which measures the structure of the clustering of the dataset. Once again, the ideal agglomerative coefficient is 1.0, and then used the Mantel's test to decide how many groups to retain.

Based on the returned Mantel values, the highest value is on five groups, and we proceeded with five groups, yielding a Mantel R above 0.60. The groups are more distinct here than the "what" groups in the previous section.

When looking at the clusters along NMS 1 vs. NMS 2, the largest are Group 1 and Group 3, and are fairly distinct (Figure 15). While 2, 4, and 5 are composed of only two or three projects, they are also relatively disjunct from the rest of the groups. While the groups mix together when looking at NMS 3 vs NMS 1 (Figure 16), the groups are once again disjunct along NMS 2 vs. NMS 3 (Figure 17).