WETLANDS, Vol. 29, No. 2, June 2009, pp. 648–665

VALIDATION OF A WETLAND RAPID ASSESSMENT METHOD: USE OF EPA'S LEVEL 1-2-3 FRAMEWORK FOR METHOD TESTING AND REFINEMENT

Eric D. Stein[1], A. Elizabeth Fetscher[1], Ross P. Clark[2], Adam Wiskind[3], J. Letitia Grenier[4],
Martha Sutula[1], Joshua N. Collins[4], and Cristina Grosso[4]

[1]*Southern California Coastal Water Research Project*
3535 Harbor Blvd., Suite 110, Costa Mesa, California, USA 92626
E-mail: erics@sccwrp.org

[2]*California Coastal Commission*
725 Front Street, Suite 300, Santa Cruz, California, USA 95060

[3]*Moss Landing Marine Laboratories*
8272 Moss Landing Road, Moss Landing, California, USA 95039

[4]*San Francisco Estuary Institute*
7770 Pardee Lane, Oakland, California, USA 94621

Abstract: Wetland rapid assessment has become popular in a variety of applications. Because rapid assessments rely on observable field indicators as surrogates for direct measures of condition, they must be validated against independent data. Here we present a case study of the validation of the riverine and estuarine modules of the California Rapid Assessment Method (CRAM). We evaluated responsiveness of the method to "good" vs. "poor" wetland condition, ability to represent a range of conditions, internal redundancy, alternative combination rules for constituent metrics, and reproducibility of results. Because no independent, concurrently collected measure of condition directly reflecting the same elements comprising CRAM was available for validation, we demonstrate the use of existing monitoring and assessment data on avian diversity, benthic macroinvertebrate indices, and plant community composition. Results indicate that CRAM is an effective tool for assessing general riverine and estuarine wetland condition based on its correspondence with multiple independent assessments of condition. Reproducibility analysis revealed several problematic metrics where ambiguous language or metric construction led to high inter-team error rates. Addressing these issues improved overall average error to within 5%. This study demonstrates that, when validated, rapid assessment methods provide a meaningful and reliable tool for assessing wetland condition.

Key Words: calibration, CRAM, method evaluation, user consistency, weight of evidence, wetland condition, wetland monitoring

INTRODUCTION

In recent years, rapid wetland assessment methods have been gaining popularity for use in a range of wetland regulatory, ambient assessment, and management applications (Stapanian et al. 2004, Cohen et al. 2005, Fennessy et al. 2007). Increases in the need for program accountability have resulted in expansion of ambient monitoring programs, more rigorous performance monitoring for mitigation and restoration projects, and an increased focus on landscape scale and cumulative impact assessment (USEPA 2002). In recognition that an intensive assessment is not always practical or desirable, the U.S. Environmental Protection Agency (USEPA) has proposed a three-tiered approach to monitoring and assessment, termed Level 1-2-3. Under this approach, Level 1 consists of habitat inventories and landscape-scale assessment, Level 2 consists of rapid assessment, and Level 3 consists of intensive assessment (Kentula 2007). Because it is less time-consuming and relatively inexpensive, Level 2, or rapid assessment, is emerging as a key element of many monitoring programs.

The intent of all rapid assessment methods (RAMs) is to evaluate the complex ecological condition of a selected ecosystem using a finite set of observable field indicators, and to express the relative condition of a particular site in a manner that informs ecosystem management. RAMs are structured tools combining scientific understanding of process and function with best professional judgment in a consistent, systematic, and repeatable manner (Sutula et al. 2006). They are based on the assumption that the ecological condition of wetlands will vary along a disturbance gradient, and that the resultant state can be evaluated based on a core set of visible field metrics. These metrics are typically qualitative measures of a specific biological or physical attribute that reflects some element of ecological condition and can be related to key ecosystem functions.

Because of their integrative nature and reliance on translating ecological theory into field indicators that reflect wetland condition, it is important that RAMs be calibrated and validated against independent measures of wetland condition in order to establish their scientific defensibility (Sutula et al. 2006). The goal of this process is not to maximize correlation between RAM metrics and any single measure of condition. Rather, the goal is to optimize RAM results against multiple independent measures of condition. In their review of RAMs, Fennessy et al. (2004) recommend that the calibration/validation process utilize results from more intensive wetland monitoring activities (i.e., Level 3 assessments). In this way, the assumptions behind the rapid assessment can be tested.

Given the cost and difficulty of collecting or compiling suitable intensive data that represent a gradient of wetland condition, very few RAMs are calibrated or validated, although excellent examples do exist. The Ohio Rapid Assessment Method (ORAM) has been validated against measures of ecological condition based on macroinvertebrate, bird, amphibian, and vascular plant diversity data (Mack 2001, Andreas et al. 2004, Micacchion 2004, Stapanian et al. 2004). The ORAM validation relied on measures of floral and faunal community structure as surrogates for direct measure of function. Similarly, Wardrop et al. (2007) used a floristic quality index (which measures richness of native plant communities) to validate a RAM for the Juniata watershed in Pennsylvania, USA. Numerous hydrogeomorphic (HGM) assessment methods have used a similar approach (Hruby et al. 1999, Hauer et al. 2002, Lee et al. 2003, Hill et al. 2006), although the HGM assessment typically includes a greater emphasis on physical and/or hydrological wetland features than do rapid assessments.

Given the complexity and diversity of wetland function and the inherent simplifications associated with RAMs, there is often no direct, mechanistic relationship between the RAM model and the validation data, hence there is no single "gold standard" measure of wetland condition that can be used for validation. However, decisions regarding modification of CRAM components can be made based on a "weight-of-evidence" approach. Weight-of-evidence is the process of combining information from multiple lines of evidence to reach a conclusion about an environmental system or stressor (Linthurst et al. 2000, Burton et al. 2002). Using multiple lines of evidence to make inferences about environmental condition is well established in ecological risk assessment, environmental toxicology, and contaminant research (Burton et al. 2002, Smith et al. 2002). The weight-of-evidence approach is less commonly used for wetland assessment, but examples exist for assessment of stream and riparian communities. Bryce et al. (2002) demonstrated the value of using multiple assemblages to assess stream and riparian habitats by comparing indices of biotic integrity based on fish, birds, and macroinvertebrates. Griffith et al. (2005) used metrics for fish, benthic macroinvertebrates, and algae to create a mixed assemblage index of biotic integrity. These investigators found that, although different indices may agree on the general level of disturbance or condition, indicators differed in their sensitivity to stressors and responded differently to conditions in stream substrate, water column chemistry, or channel and riparian habitat. Consequently, using multiple indicators together provided the most complete and robust understanding of overall stream condition.

The preferred approach for RAM validation is to collect independent measures of condition concurrently with conducting a RAM-based assessment. Although desirable, the collection of new data is often cost or time prohibitive. For this reason, relying on existing data sources and applying the weight-of-evidence approach is an attractive alternative. This approach allows for validation of the RAM via exploration of relationships between RAM output and independent measures of condition. These relationships can then be assessed vis-à-vis the expectations of a conceptual model that is developed *a priori*.

In this paper, we present the results of validation of the California Rapid Assessment Method (CRAM; Collins et al. 2006) in riverine and estuarine wetlands. We define "validation" as the process of documenting relationships between CRAM results and independent measures of condition in order to establish CRAM's defensibility as a meaningful and repeatable measure of wetland condition. The overall validation process includes several steps designed to meet the following objectives: 1) assure that the method is producing meaningful results based on a comparison between CRAM scores and independent measures of condition (*evaluation*), 2) make adjustments to the method, where needed, to improve the ability of the CRAM to discern differences in wetland condition (*calibration*), and 3) minimize observer bias by assessing repeatability between independent assessment teams and modifying metrics that lead to inconsistencies (*standardization*). The validation process involved evaluating CRAM in terms of its performance with regard to several factors: 1) *responsiveness*, a measure of the ability of the method to discern good vs. poor condition, 2) *range* and *representativeness*, the ability of the method to appropriately capture the distribution of condition states that exists in nature, 3) *redundancy*, the degree to which multiple metrics measure the same elements of condition, 4) *integration*, the effect of different means of combining CRAM's component metrics of condition to generate an overall score, and 5) *reproducibility*, the proportion of total variance attributable to user error. The approach presented for validating CRAM with the aid of existing data sources is applicable to any RAM that follows the general framework recommended by Fennessy et al. (2007) and Wardrop et al. (2007).

METHODS

Overview of CRAM

The overall goal of CRAM is to provide a rapid, scientifically defensible, and repeatable assessment method that can be used routinely for wetland monitoring and assessment. CRAM consists of assessing wetlands with respect to four overarching attributes: Buffer/Landscape Context, Hydrology, Physical Structure, and Biotic Structure. Within each of these attributes are a number of metrics that address more specific aspects of wetland condition (Table 1). Each of the metrics is assigned a numeric score based on either narrative or schematic descriptions of condition, or thresholds across continuous values. Metric descriptions are based on characteristics of wetlands observed across a range of reference conditions (Smith et al. 1995), such that the highest score for each metric represents the theoretical optimum condition obtainable for the wetland feature being evaluated for a given wetland type in California. Although wetlands perform a suite of functions, CRAM is designed to assess condition based on the capacity of a wetland to support characteristic native flora and fauna. In other words, hydrology and physical structure are assessed based on their contribution to supporting plant and animal habitat rather than on the ability of the wetland to provide services such as flood attenuation or water quality improvement. The underlying assumption of CRAM is that the "living-resource support" function is a common management endpoint, is easily discernable, and integrates the contributions of hydrogeomorphic, physicochemical, and biotic interactions within a wetland. The relationship between habitat and physical and biological processes has been demonstrated for a variety of taxa including fish, amphibians, and invertebrates (Talmage et al. 2002, Baber et al. 2004). The selection of CRAM metrics and attributes reflects the underlying assumption that such relationships exist. For this reason, CRAM was validated using Level 3 data that reflect capacity to provide the living-resource support function.

Table 1. CRAM attributes and metrics from the pre-calibration version of CRAM. AA=Assessment Area.

Attributes		Metrics
Buffer and Landscape Context		Landscape Connectivity
		Percent of AA with Buffer
		Average Width of Buffer
		Buffer Condition
Hydrology		Water Source
		Hydroperiod
		Hydrologic Connectivity
Structure	Physical	Physical Patch Richness
		Topographic Complexity
	Biotic	Organic Matter Accumulation
		Biotic Patch Richness
		Vertical Biotic Structure
		Interspersion and Zonation
		Percent Non-native Plant Species
		Native Plant Species Richness

CRAM is applicable to wetlands (including riverine wetlands and their associated in-stream and riparian habitats) throughout California. The general approach and metric categories are consistent across wetland types that roughly correspond to the classes articulated by Cowardin et al. (1979), but the specific narratives used to score each metric are customized, as needed, for the characteristics of the specific wetland type being assessed. Metric scores are aggregated up to the level of attributes as well as into a single overall score via simple arithmetic relationships. Categories have been developed based on implied equivalence in the sense that the incremental increase in condition associated with moving from one category to the next higher category is the same across attributes. A detailed description of the method is provided in the CRAM manual (Collins et al. 2006).

Conceptual Approach to RAM Validation

The performance of CRAM was evaluated by comparing CRAM scores to field data on biotic community structure, which are believed to be indicative of the level of ecosystem function. Because these data integrate over time and through space in ways analogous to the CRAM attributes, the evaluation and adjustment of CRAM took place largely at the attribute level. Changes to metrics and to combination algorithms were made to improve relationships between CRAM attributes (as opposed to metrics) and independent measures of condition and to provide more consistency between independent assessment teams (i.e., to improve standardization).

Selection of Validation Data Sets

Existing data sets were screened for suitability for use in validation. In addition to providing an independent measure of ecological condition, the data sets needed to meet the following criteria:

1. The data set should have statewide coverage to allow for validation to the same data sources across the study area.
2. The data set should represent a range of conditions across a gradient of disturbance.
3. The site locations of the surveys should be accessible to the CRAM assessment crews.
4. The data should be reflective of defined element(s) of wetland function (i.e., living-resources support) that can be related to specific attributes of the rapid assessment method.
5. The data should be readily available and include metadata describing the original purpose and objectives for the data set, sampling methods and location, procedures for data collection, and analysis. Quality control information should accompany the data set or be available through consultation with the data authors.
6. The authors of the data set should be available for consultation about such issues as missing data, filling data gaps, the meaning of zero counts, interpretation of outlier data points, and limitations on interpretation of the data set, including the degree to which the data can be extrapolated to sites for which data do not exist.
7. The data set should be scientifically credible and clear of any controversy about its validity, integrity, and ownership; and it should not be currently withheld from distribution because of legal or proprietary concerns.
8. Consistent data collection and analysis methods and quality assurance procedures should apply to the entire data set.
9. The data should be recently collected so that they reflect existing field conditions. For the purposes of CRAM validation, "recently collected" means that data were not older than 3 years old. It is assumed that this period is an acceptable interval within which to expect only negligible changes in condition at the site, assuming no major impacts (anthropogenic or natural) have occurred (e.g., major flood, fire, change in land use practices). Sites where major impacts are known to have occurred during the intervening time period should be excluded.

Three sources of Level-3 data were identified for use in validation of CRAM: 1) Riparian bird capture data from the Monitoring Avian Productivity and Survivorship Program (MAPS); 2) Benthic macroinvertebrate data from the statewide bioassessment database; and 3) Plant community composition data from a recent USEPA assessment under the Environmental Monitoring and Assessment Program (EMAP) West Coast Pilot. Each of these data sets is described in more detail below.

The MAPS Program is a nationwide effort, managed by the Institute for Bird Populations, that collects annual data on bird populations during the breeding season using a constant-effort, mist net approach at fixed-site locations (IBP 2006). A detailed description of the MAPS program objectives and approach can be found at http://www.birdpop.org/maps.htm. MAPS data provide species-specific information about trends in demographics, productivity, survival rates, and rates of recruitment into the adult populations. MAPS data on diversity of bird species from captures in riparian sites during the 2003 breeding season were used for CRAM validation of riverine wetlands. The six "MAPS metrics" used for the validation are described in detail in Table 2.

Throughout California, efforts are underway to collect bioassessment data in wadeable streams for use in a variety of programs. Data collected include information about benthic macroinvertebrate (BMI) species diversity and abundance. These data can be used to calculate an Index of Biotic Integrity (IBI; Ode et al. 2005). The results of bioassessment provide information about water quality and in-stream benthic habitat condition resulting from perturbations such as contamination, hydromodification, and sedimentation from upstream sources (Resh and Jackson 1993). IBI scores from bioassessment data collected by the California Department of Fish and Game (CDFG) in 2003 using the California Stream Bioassessment Procedure (Harrington 1999) were used for CRAM validation. A detailed description of CDFG bioassessment objectives and approaches can be found at http://www.dfg.ca.gov/abl/Field/datacollection.asp.

Table 2. Expected correlations between CRAM attributes and Level-3 data metrics. The nature of expected relationships (i.e., positive or negative) is indicated by "+" and "−" signs.

Level-3 Metric	Definition	Overall CRAM Score	Buffer/ Landscape	Hydrology	Physical Structure	Biotic Structure
BMI IBI	Benthic macroinvertebrate Index of Biotic Integrity for riverine wetlands	+	+	+	+	+
MAPS 1	Species richness of all birds	+				
MAPS 2	Species richness of riparian-associated species	+				+
MAPS 3	Species richness of non-riparian-associated species		+			
MAPS 4	Reproductive index (ratio of young to adults) for all species	+				
MAPS 5	Reproductive index (ratio of young to adults) for riparian-associated species	+				+
MAPS 6	Reproductive index (ratio of young to adults) for non-riparian-associated species		+			
EMAP 1	Relative percent cover of non-native plants across the marsh plain	−	−	−		−
EMAP 2	Relative percent cover of invasive plants across the marsh plain	−	−			−
EMAP 3	The total number of native plant species found along transects across the marsh plain	+		+	+	+
EMAP 4	Relative percent cover of non-native plants along the backshore border of the assessment area		−			−
EMAP 5	The total number of plant species (native plus non-native) found along transects across the marsh plain	+		+	+	+

EPA's EMAP-Estuaries West Coast Pilot conducted a probability-based ambient assessment of intertidal wetlands in Washington, Oregon, and California in 2002 (Sutula et al. 2001). As an intensification of this survey, comprehensive plant community composition data were collected in southern California and the San Francisco Bay Area. Resources providing a description of the EMAP-Estuaries West Coast Pilot objectives and approaches can be found at http://www.epa.gov/region09/water/wemap. Assessment of plant community composition at these locations involved collecting point-intercept data along a series of transects oriented in a stratified manner designed to cover a variety of elevation gradients and geomorphic features throughout the coastal marsh plain. These data provide information about the species richness, diversity, and relative percent cover. Five "metrics" were calculated from the EMAP data for use in CRAM validation, and are described in detail in Table 2.

Validation Analysis

Of the six wetland classes covered by CRAM, the riverine and estuarine classes were selected as the priority for calibration and validation based on current assessment needs and availability of appropriate Level 3 data. Validation of the remaining CRAM wetland classes (depressional wetlands, vernal pools, seeps and springs, lake and lagoon fringe wetlands) will occur in the future. Three regional field teams used CRAM to assess the condition of 95 riverine sites distributed throughout California. Of these, 54 had benthic macroinvertebrate data and 41 had MAPS bird data. For estuaries, assessments were conducted at 38 sites statewide, all of which had EMAP vegetation data. CRAM Assessment Area sizes ranged from 0.13–74 ha for estuarine wetlands and 0.04–25 ha for riverine wetlands. These ranges were influenced by locations where Level 3 data were collected and the need for coincident assessment areas. At each site, CRAM Assessment Area(s) were identified that corresponded to the area where the Level-3 data had been collected. A CRAM assessment was conducted for each site and the results were used in combination with the existing intensive (Level 3) data to conduct the following validation analyses, which are adapted from analyses used by others to test indices of biotic integrity (Whittier et al. 2007).

Responsiveness is a measure of the ability of the method to discern good vs. poor condition, and was tested in two ways. First, correlation (using Spearman's ρ) and simple regression analyses were used to characterize the relationships between Level-3 data and CRAM overall, attribute, and metric scores. Consistent patterns of correlations between CRAM metrics or attributes and multiple Level-3 variables, in the expected directions, were interpreted as indicating responsiveness. Where the relationship between attribute scores and Level-3 data differed from expected based on the CRAM conceptual model, modifications were explored to improve the relationship. Modifications included changes to metric scaling or metric combination rules and were based on the ecological models underlying CRAM. The metrics within each attribute were also investigated for consistency of response to varying condition. Divergent metrics were modified (or in some cases eliminated or combined with other metrics) to improve overall method performance.

The second test of the ability of CRAM to reflect overall condition was based on investigation of the relationship between CRAM scores and the Landscape Development Index (LDI; Brown and Vivas 2005), which is a (Level 1) landscape measure of human disturbance. The LDI analysis was conducted using land use data from the 2001 National Land Cover Database (http://www.epa.gov/mrlc/nlcd-2001.html) and the emergy values as published by Brown and Vivas (2005). Correlations between CRAM attribute and overall scores and LDIs were used as an additional measure of CRAM responsiveness to condition along a gradient of stress. Relationships were tested against human disturbance at various spatial scales, including within a 200 m buffer, a 500 m upstream area (for riverine wetlands), the upstream drainage area, and the entire watershed.

Range and representativeness are measures of the ability of the method to appropriately capture the distribution of conditions that exist in nature. Distributions of scores for metrics and attributes were graphed and compared against the normal distribution as well as distributions of the Level-3 data types. There were *a priori* assumptions about distributions of the Level-3 data, based on the goals of the studies for which the data were collected. For instance, for the EMAP vegetation data, scores were not expected to be normally distributed, but rather representative of the range of conditions in the region because the samples were selected probabilistically from all possible locales within the study region. The distributions of scores from validation sites were also compared to the distribution of scores at "reference standard" sites (i.e., known good condition sites, based on studies independent of CRAM and the Level-3 data sets; Ambrose et al. 2006). Metrics and attributes that were severely skewed relative to expectations were modified to improve their distribution. Modifications typically entailed adjusting the metric categories by redefining the thresholds between scores within a metric. All modifications were informed by both the distribution of the data and the underlying conceptual models that govern CRAM.

Redundancy assesses the degree to which multiple metrics measure the same elements of condition. High redundancy between specific metrics constitutes implicit weighting of that aspect of the wetland and should be taken into consideration when interpreting CRAM results. Redundancy was measured in two ways. First a correlation matrix, using Spearman's ρ, was generated to investigate relationships between metrics. Second, a Principal Components Analysis (PCA) was conducted using the individual CRAM metric scores. Correlations were tested for BMI IBI scores both on the first principle component (PC1) of the PCA, which represents the metrics that most influence variability in CRAM scores, and on the CRAM overall score, to test for fidelity of the results across hierarchical levels of CRAM. Redundant metrics were not necessarily eliminated, but were acknowledged in order to improve the transparency of the method, to inform combination rule development, and to aid in interpretation of results.

Integration measures the effect of different metric combination rules on attribute scores. Between one and four potential combination rules were constructed for each attribute based on conceptual model(s) of how the metrics relate to each other to represent the component of condition being assessed by each attribute (Table 3). In all cases, a simple arithmetic mean of metric scores was included as the neutral model with which to compare any alternative combination models. Alternatives to the neutral model consisted of more mathematically complex combinations of metrics based on assumed mechanistic relationships. Combination rules were tested by correlating the resultant attribute scores against the appropriate Level-3 data. Alternatives to the neutral model were selected only if they were mechanistically justified and either provided stronger correlations between attribute scores and Level-3 data or helped meet other validation objectives (i.e., range, responsiveness). All combination rules were tested with Level-3 data to ensure that score calculation processes did not undermine other validation objectives, such as responsiveness.

Reproducibility is a measure of the proportion of total variance attributable to user error, and is a reflection of the precision of CRAM results. Numerous duplicate CRAM assessments were completed by teams of wetland scientists trained in

Table 3. Metric combination rules tested for each CRAM attribute. BPR=Biotic Patch Richness; IZ=Interspersion and Zonation; OMA=Organic Matter Accumulation; VBS=Vertical Biotic Structure; NN=Non-Native.

Attribute	CRAM Metric Combination Rule
Buffer and Landscape Context	(% w/Buffer + Avg Width + Buffer Condition + Connectivity) / 4
	[(% w/Buffer * Avg Width * Condition)$^{1/3}$ + Connectivity] / 2
Hydrology	(Water Source + Hydroperiod + Connectivity) / 3
	[Water Source * ((Hydroperiod + Connectivity)/ 2)] $^{1/2}$
Physical Structure	(Patch richness + Topographic complexity) / 2
Biotic Structure	(OMA + BPR + VBS + IZ + Native spp. + % NN) / 6
	[(OMA+ BPR + VBS + IZ + Native spp.)/5 * % NN)] $^{1/2}$
	[(OMA+ BPR + (% NN*VBS *IZ) $^{1/3}$ + (Native spp. * % NN) $^{1/2}$] / 4
	[(OMA+ BRP + VBS + IZ) + (Native spp. * % NN) $^{1/2}$] / 5

the use of CRAM, to determine the sampling error of the method in terms of multiple potential sources: 1) within-team variability (the same team conducted two CRAM assessments of the same Assessment Area within a month), 2) between-team variability (two teams completed a CRAM assessment within the same Assessment Area within a month), 3) among-region variability (CRAM teams from each of the regions evaluated the same Assessment Area within a month), and 4) temporal variability (the same team returned to conduct a second CRAM assessment 4–5 months later).

Sampling error from each identified source was estimated using a simple tally system that recorded the magnitude of discrepancy between paired assessments. For paired metric scores that differed by one metric category, the difference was enumerated as one (1) discrepancy. If scores for a metric differed by two categories, the discrepancy was enumerated as two (2). The metric discrepancies were summed for each attribute and then expressed as a percentage of total possible differences (i.e., the number of differences that would have occurred if every metric differed by the maximum possible categories) to provide an estimate of error. Sampling error rates were used as a guide to determine when adjustments were necessary to address ambiguity within metrics. In general, adjustments were made when error rates exceeded 10%, or where systematic errors occurred.

RESULTS

Responsiveness

CRAM overall scores were significantly correlated with several of the Level-3 variables in ways that were consistent with the CRAM conceptual model. For riverine wetlands, CRAM overall scores were strongly significantly correlated with BMI IBI scores (Spearman's $\rho = 0.642$, $P < 0.0001$; Figure 1a) and were negatively correlated with the relative percent cover of non-native plants in estuarine wetlands (EMAP 4: Spearman's $\rho = -0.359$, $P = 0.0379$; Figure 1b). In addition, positive relationships were observed between CRAM overall score and some MAPS-derived measures of wetland function in terms of avian support in riverine wetlands (e.g., MAPS 1: Spearman's $\rho = 0.303$, $P = 0.055$; Figure 1c).

Individual CRAM attribute and metric scores were also correlated with elements of the Level-3 data sets that represent analogous aspects of wetland condition (Table 2). There were significant correlations between CRAM scores and multiple measures of biological community structure at both the attribute level (Table 4) and the metric level (Table 5). Correlations with the MAPS data were stronger at the attribute level than at the overall score level, particularly between avian richness (MAPS 1 and MAPS 2) and CRAM biotic structure, which reflect similar aspects of condition (Figure 2). The one exception to the positive correlations was the significant negative correlation between the CRAM physical structure attribute and the MAPS metric measuring the reproductive ratio of non-riparian birds (MAPS 6). BMI scores correlated strongly and positively with all CRAM attributes for riverine wetlands. For estuaries, the strongest correlations with Level-3 data at the attribute level were for Buffer and Landscape Context and Biotic Structure, the latter of which correlated with a number of the EMAP vegetation metrics. The positive relationships with EMAP 1 and 2 were unexpected, as these are measures of the relative representation of non-native and invasive plant species, respectively, on the marsh plain. No significant relationships were observed between any Level-3 estuarine data and the Hydrology or Physical Structure attributes.

The strength of correlations at the metric level varied from metric to metric, but in general, the pattern observed for most metrics was consistent

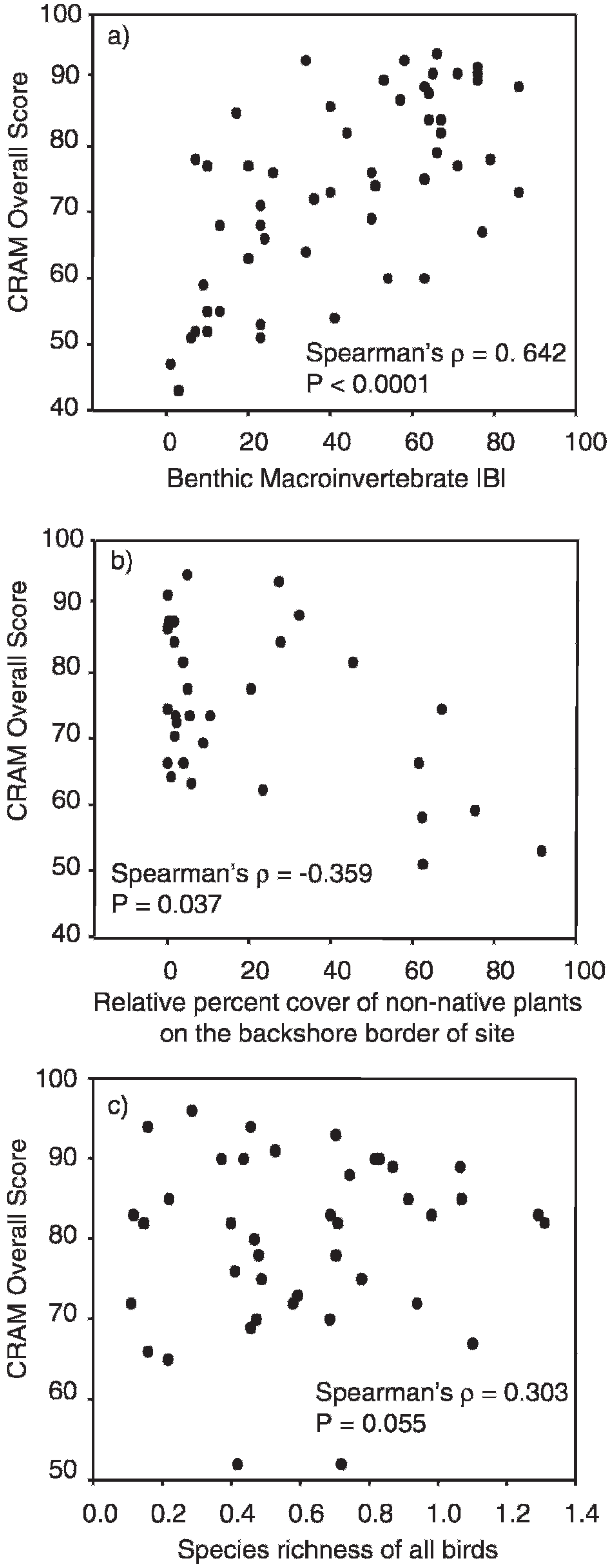

Figure 1. Relationships between CRAM overall score and A) benthic macroinvertebrate Index of Biotic Integrity scores for riverine wetlands, B) the relative percent cover of non-native plant species along the backshore of estuarine wetlands, EMAP4, MAPS1, and C) bird species richness in riverine wetlands.

with those observed at the attribute level (Table 5). At least one Level-3 metric correlated significantly, and in the expected direction, for every riverine metric with the exception of Hydroperiod, although its relationship to BMI IBI scores was nearly significant (Spearman's $\rho = 0.197$, $P = 0.069$). For many of the riverine metrics, there were significant correlations not only with multiple metrics, but with metrics from two distinct data sets (i.e., both MAPS and the BMI IBI). For estuaries, the expected significant correlations between metrics and Level-3 data were observed for over half the metrics. There were only two relationships that ran contrary to expectations: the significant positive relationships between the Water Source CRAM metric and the EMAP 1 and 2 metrics (which reflect the relative percent cover of non-native and invasive plant species in estuarine Assessment Areas).

There were consistent, significant negative correlations between LDI scores and CRAM overall and attribute scores. These relationships were apparent at varying spatial scales ranging from buffers of varying widths around the wetland up to the entire contributing catchment (Table 6). As the index of human disturbance increased, the CRAM scores decreased. For riverine wetlands, relationships were strongest for the Buffer and Landscape Context and Biotic Structure attributes, regardless of the spatial scale investigated. For estuarine wetlands, relationships were strongest for the Buffer and Landscape Context and Hydrology attributes, while the relationships with the Physical Structure and Biotic Structure attributes were not significant at the $\alpha = 0.05$ level.

Range and Representativeness

For riverine wetlands, CRAM Biotic Structure attribute scores were normally distributed (Shapiro-Wilk's $W = 0.971$; $P = 0.189$). The Physical Structure attribute scores had a high Shapiro-Wilk's W, but were not statistically normal ($W = 0.905$; $P < 0.0001$; Figure 3a). For estuarine wetlands, both the Buffer and Landscape Context and Biotic Structure attributes were nearly normally distributed ($W = 0.939$; $P = 0.049$ and $W = 0.934$; $P = 0.035$, respectively; Figure 3b). These distributions are consistent with the distributions based on various Level-3 indicators, suggesting that the distribution in CRAM scores is representative of the actual range of condition at the validation sites. In contrast, the Hydrology attribute scores were positively skewed for both wetland classes ($W = 0.863$; $P < 0.0001$ for riverine and $W = 0.849$; $P = 0.0001$ for estuarine; Figures 3a and b). Buffer and Landscape Context

Table 4. Relationships between CRAM attributes and Level-3 metrics. Correlations are presented in terms of Spearman's ρ. All relationships that are significant at the $\alpha = 0.05$ level are shown. Level 3 metrics are as defined in Table 2.

CRAM Attribute	Wetland Class	Level-3 Metric	ρ	P
Buffer and Landscape Context	estuarine	EMAP 4	−0.352	0.042
	riverine	BMI IBI	0.433	<0.0001
		MAPS 3	0.474	0.002
Hydrology	riverine	BMI IBI	0.464	<0.0001
Physical Structure	riverine	BMI IBI	0.262	0.012
		MAPS 6	−0.343	0.035
Biotic Structure	estuarine	EMAP 1	−0.338	0.038
		EMAP 2	−0.368	0.023
		EMAP 3	0.502	0.001
	riverine	BMI IBI	0.324	0.001
		MAPS 1	0.342	0.029
		MAPS 2	0.328	0.037

was also highly skewed for riverine wetlands ($W = 0.822$; $P < 0.0001$; Figure 3a). Further investigation revealed that the component metrics of the Buffer and Landscape Context attributes were also positively skewed. Similarly, the Hydroperiod metric was positively skewed, resulting in the shift of the distribution of the Hydrology attribute scores (to which the Hydroperiod metric contributes).

Redundancy

Correlation among metrics within an attribute was generally high, particularly for the Buffer and Landscape Context and Physical and Biotic Structure attributes. Although not unexpected, such correlations can result in implicit weighting of certain wetland features by ''double counting'' them via several metrics. Results of the Principal Components Analysis (PCA) indicate that the level of redundancy inherent in CRAM does not obscure the overarching patterns of wetland condition. Both overall CRAM scores and BMI IBI scores were positively correlated with the first principle component (PC1) of the PCA (Spearman's $\rho = 0.970$, $P < 0.0001$ and $\rho = 0.647$, $P < 0.0001$, respectively), and with each other ($\rho = 0.642$; $P < 0.0001$). Figure 4 shows an example of results that are indicative of all the PCA correlations. The results indicate fidelity across hierarchical levels of CRAM and suggest that the manner in which CRAM metrics are combined into attributes, and attributes combined into overall scores, captures the overall variance in wetland condition.

Integration

For all CRAM attributes, there were no significant differences between the neutral metric combination model (i.e., an arithmetic mean) and the more complex mechanistic models in terms of their relationship with Level-3 data. We calculated the difference between attribute scores generated by the neutral model for combining the CRAM metrics and all alternative model(s). We then regressed the percent difference between models on the neutral model scores. The regression tested the null hypothesis that the slope of the relationship between the percent difference between models and the neutral model scores is 0. Our failure to reject the null hypothesis indicates the models behave similarly across a range of CRAM scores. Because there were no differences among combination rules, the neutral model was selected to ease the use and interpretation of CRAM by a broad range of practitioners.

Reproducibility

The average error in overall CRAM results following repeated independent assessment ranged from 7–23% prior to modifications in the method protocols and support materials. Error rates were lowest within a single assessment team (7–11%) and higher when a site was assessed by two different teams (9–23%). Investigation of error by attribute revealed the most likely causes of discrepancy between individual assessments. The highest error rates among different regions were in the Physical and Biotic Structure attributes. Further investigation of specific metrics revealed that for riverine wetlands, the majority of error was due to three metrics: Hydroperiod, Vertical Biotic Structure, and Percent Non-Native Plant Species (which were also problematic for estuarine wetlands). Comparison of assessments conducted months apart resulted in a 25% error rate for the Biotic Structure attribute compared to an error of 7% when assessments were

Table 5. Relationships between CRAM metrics and Level-3 metrics. Correlations are presented in terms of Spearman's ρ. All relationships that are significant at the $\alpha = 0.05$ level are shown. Level-3 metrics are as defined in Table 2.

Attribute	CRAM Metric	Wetland Class	Level-3 Metric	ρ	Prob>|ρ|
Buffer and Landscape Context	Connectivity	riverine	BMI IBI	0.355	0.001
			MAPS 3	0.346	0.027
	% of Assessment Area with Buffer	riverine	BMI IBI	0.278	0.013
	Average Buffer Width	estuarine	EMAP 4	−0.496	0.003
		riverine	BMI IBI	0.451	<0.0001
	Buffer Condition	estuarine	EMAP 4	−0.561	0.001
		riverine	BMI IBI	0.440	<0.0001
			MAPS 3	0.409	0.008
			MAPS 6	−0.389	0.016
Hydrology	Water Source	estuarine	EMAP 1	0.328	0.044
			EMAP 2	0.464	0.003
		riverine	BMI IBI	0.567	<0.0001
			MAPS 3	0.455	0.003
	Hydrologic Connectivity	estuarine	EMAP 1	−0.416	0.009
			EMAP 4	−0.464	0.006
		riverine	BMI IBI	0.447	0.001
Physical Structure	Physical Patch Richness	riverine	MAPS 6	−0.368	0.024
	Topographic Complexity	riverine	BMI IBI	0.323	0.003
Biotic Structure	Organic Matter Accumulation	estuarine	EMAP 1	−0.366	0.024
			EMAP 2	−0.504	0.001
		riverine	BMI IBI	0.407	<0.0001
			MAPS 4	0.309	0.050
			MAPS 5	0.320	0.042
	Biotic Patch Richness	riverine	MAPS 1	0.419	0.006
			MAPS 2	0.481	0.001
	Vertical Biotic Structure	estuarine	EMAP 3	0.336	0.039
	Interspersion/Zonation	estuarine	EMAP 3	0.513	0.001
			EMAP 5	0.428	0.007
		riverine	BMI IBI	0.253	0.016
	% Non-Native Plant Species	estuarine	EMAP 1	−0.462	0.004
			EMAP 2	−0.383	0.018
		riverine	BMI IBI	0.401	<0.0001
			MAPS 3	0.340	0.030
	Native Plant Species Richness	estuarine	EMAP 1	−0.383	0.018
			EMAP 2	−0.416	0.009
			EMAP 3	0.461	0.004
		riverine	BMI IBI	0.217	0.039
			MAPS 1	0.315	0.045
			MAPS 2	0.425	0.006

conducted only weeks apart. This suggests that seasonal differences in plant communities may contribute to variability in this attribute.

To address initial reproducibility problems, CRAM was modified to reduce ambiguous language in the metric descriptions and additional guidance was provided for metrics subject to high error rates. For several metrics (Native Plant Species Richness, Percent Non-Native Plant Species, and Vertical Biotic Structure) the basic evaluation method was simplified or changed. Table 7 provides a summary of the types of changes made to CRAM metrics.

Reproducibility of the revised CRAM was re-evaluated in 2007 in preparation for its use in a statewide ambient survey. The average attribute error between independent assessment teams ranged from 6–12% and average error in overall CRAM score was 5% for estuaries and 7% for riverine wetlands. More importantly, the error rate in previously problematic metrics was substantially reduced. For several of the metrics related to plant community composition, a substantial simplification of the metrics dramatically improved reproducibility, yet still provided adequate ability to discern biotic condition. For example, the original error

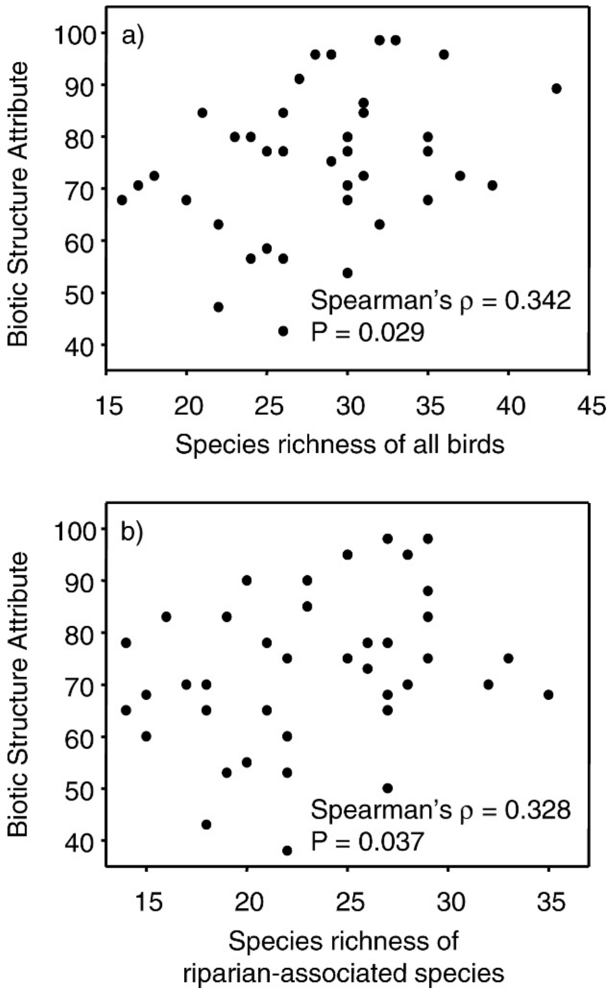

Figure 2. Relationship between CRAM Biotic Structure attribute score and measures of A) overall bird species richness, MAPS1, and B) riparian-associated bird species richness, MAPS2.

rates for the Vertical Biotic Structure and Percent Non-native Plant Species metrics were 26% and 28%, respectively. Following the modifications made during the validation process, these error rates were reduced to 11% and 8%, respectively. As a result of the modifications described above, the overall error rates met the pre-determined objective of < 10% error between assessment teams.

DISCUSSION

The analyses presented in this paper demonstrate how existing data can be used to evaluate, refine, and standardize RAMs using a weight-of-evidence approach. CRAM attributes generally corresponded well to multiple independent measures of biologic condition (BMI, IBI, MAPS) and to indicators of landscape disturbance (LDI). These results validate the underlying conceptual models of CRAM and provide scientific defensibility for the method that will be important for future regulatory and management applications.

Evaluation

The results of this analysis show that CRAM is an effective tool for assessing general wetland condition based on field indicators of a wetland's ability to support characteristic flora and fauna. Specifically, CRAM meets key features suggested by Brooks et al. (1998) for an acceptable index of ecological integrity, such as the ability to discern biological communities with high integrity, inclusion of metrics with biological, chemical, and physical bases, inclusion of indicators that are related to specific stressors that can be managed, and protocols that can be rapidly applied. Conclusions about the validity of CRAM are based on its correspondence with previously validated independent measures of condition that reflect biotic integrity in terms of bird and macroinvertebrate indices for riverine wetlands and plant indices for estuarine wetlands. Furthermore, CRAM results were strongly (negatively) correlated with independent measures of landscape

Table 6. Correlations between CRAM attribute and overall scores and the Landscape Development Index (LDI). Correlation coefficients are shown for riverine and estuarine wetlands at various spatial scales. All coefficients shown are significant at the p value noted. NS = not significant.

	Riverine						Estuarine	
	200 m Buffer	P	500 m Upstream	P	Watershed	P	200 m Buffer	P
Overall CRAM	−0.594	0.01	−0.586	0.01	−0.358	0.01	−0.332	0.05
Buffer and Landscape Context	−0.648	0.01	−0.53	0.01	−0.404	0.01	−0.407	0.01
Hydrology	−0.249	0.01	−0.266	0.01	−0.227	0.05	−0.427	0.01
Physical Structure	−0.304	0.01	−0.287	0.01	−0.251	0.01	NS	
Biotic Structure	−0.465	0.01	−0.483	0.01	−0.262	0.01	NS	

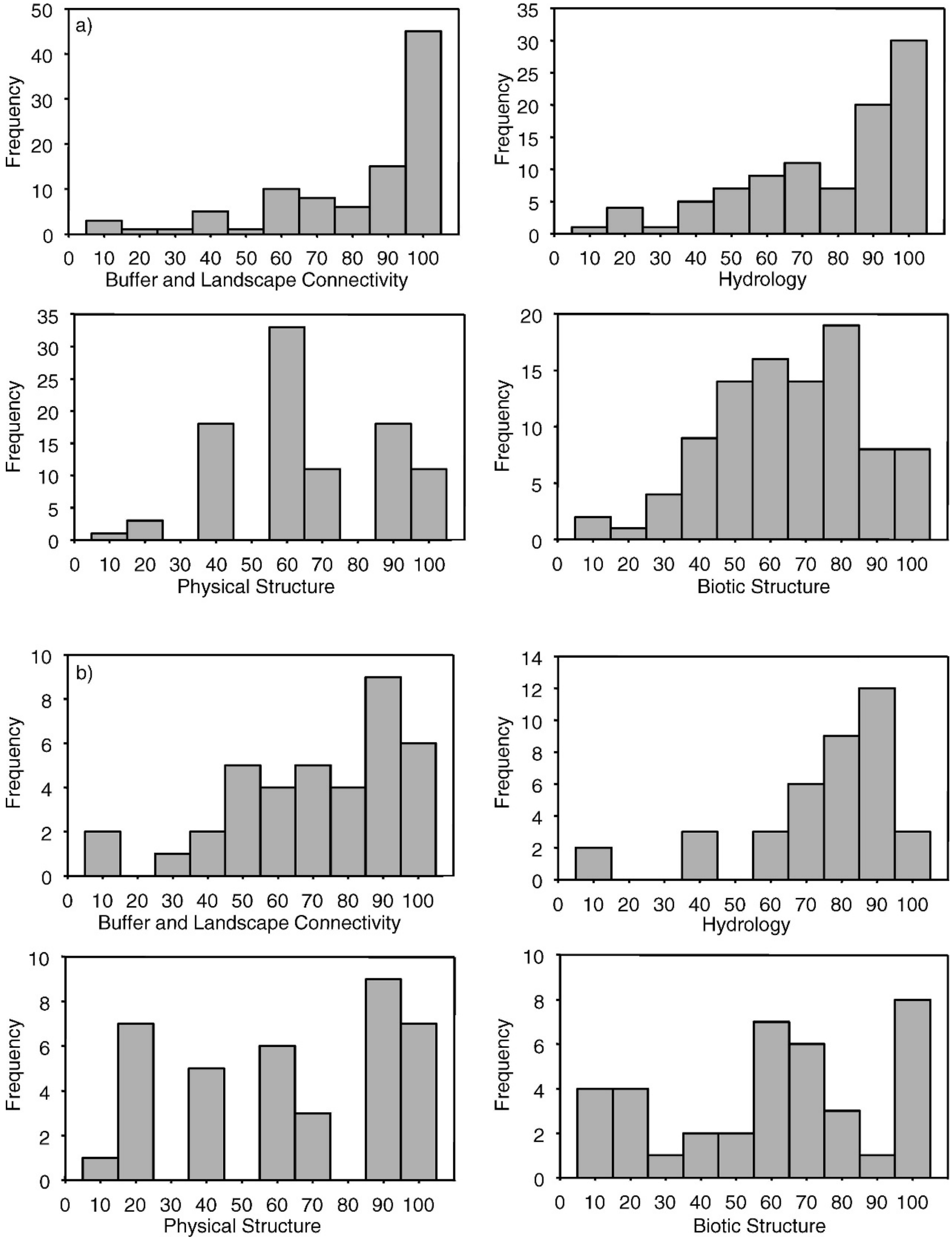

Figure 3. Distributions of CRAM attribute scores for A) riverine wetlands and B) estuarine wetlands.

disturbance based on the LDI, which has previously been shown to indicate stress and to correlate with lower wetland condition as measured in both Level-2 (RAM) and Level-3 (intensive) assessments (Brown and Vivas 2005, Mack 2006, Reiss and Brown 2007). Similarly, in their study of wetland compensatory mitigation, Ambrose et al. (2006) found that CRAM scores reliably reflected overall

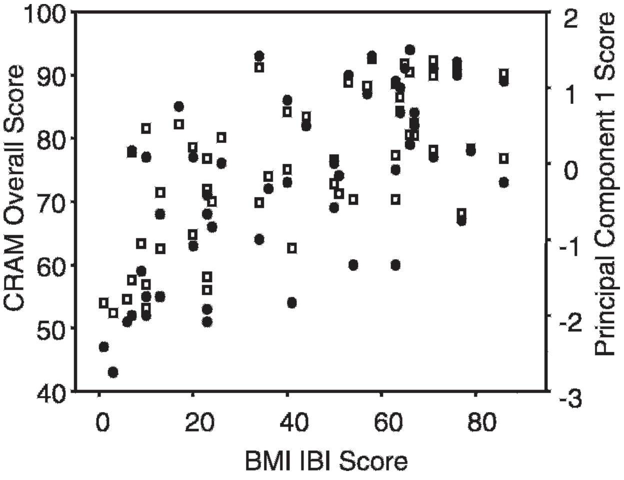

Figure 4. Relationship between CRAM overall score and principal component scores derived from a model with all CRAM metrics loaded. Solid black circles correspond to CRAM Overall Score as calculated per the CRAM protocol, and open squares correspond to Principal Component 1 Scores.

condition. Ambrose et al. (2006) also conducted CRAM assessments at 47 "reference sites" that represented the best attainable conditions within each wetland class and found that "reference standard" sites exhibited CRAM scores that were clustered toward the upper end of possible scores.

In aggregate, Level-3 data corroborated overall CRAM performance; however, individual Level-3 measures and CRAM scores were not always significantly correlated. Deviations between CRAM and MAPS and IBI data may result from the fact that CRAM is scaled to a theoretical optimum condition while the MAPS/IBI indices are scaled to the least disturbed condition sampled for these studies (Stoddard et al. 2006). However, it should be noted that site selection for the MAPS dataset tended to be skewed toward higher-quality habitat areas. This is because MAPS monitoring sites are selected based on the specific needs of individual monitoring proponents, and not probabilistically or systematically. This may mitigate the aforementioned phenomenon to some degree, but still represents a potential deficiency of the data set from the standpoint of our purposes. Bird and macroinvertebrate indicators used by MAPS and the BMI IBI can integrate external stressors (water quality, predation intensity, lack of food from adjacent habitats or upstream areas) in ways that CRAM does not. CRAM integrates impacts from upstream and other adjacent areas through Landscape and Hydrologic Connectivity and Water Source metrics, which reflect a small portion of the overall condition score. However, poor water quality can have a significant impact on benthic macroinvertebrates and potentially bird populations, overwhelming other site-specific condition attributes. Therefore, MAPS and IBI results may respond to specific stressors that make other components of the wetland condition irrelevant. This is illustrated by the fact that the strongest correlations for the MAPs data were with the CRAM Biotic Structure attribute, while weaker correlations were seen for CRAM attributes that are less closely related to aspects of the riparian community measured by MAPS. MAPS and IBI metrics may also respond to stressors not associated with the wetland being assessed. For example, MAPS data may be confounded by the presence of young migrant birds from other areas that were captured during the mist net surveys or by population effects at overwintering habitat. Nevertheless, the fact that very few sites with high MAPS or IBI scores were found in association with low CRAM scores is a key indication of CRAM's ability to discern condition in a robust manner despite different underlying theoretical assessment models.

The relationships between Level-3 data, LDI scores, and CRAM scores were less significant for estuarine wetlands than for riverine wetlands. This difference could be due to smaller sample sizes, exaggerated regional differences in estuaries, and a smaller range in condition compared to riverine wetlands. A potential deficiency of the EMAP data set, from the standpoint of our purposes, is the fact that the data do not capture very high quality sites. This is because much of the California coastline has been degraded, and there are very few estuaries in the state that have not been impacted by anthropogenic activities, particularly in southern California and the San Francisco Bay Area. The influence of regional differences is illustrated by the unexpected positive relationship between the CRAM water source metric and the EMAP metrics that reflect the relative percent cover of non-native and of invasive (which is a subset of non-native) plant species. This relationship likely resulted from the effect of different forcing functions in different regions of California. In northern California, EMAP assessments yielded higher percent cover values for invasive plant species, mainly due to the prevalence of invasive cordgrass in this area. In southern California, invasive species are less of a problem within the tidal marsh plain. However, the CRAM Water Source metric scores are lower in southern California, which tends to have more intense coastal development impacting estuaries. These regional differences in key forcing functions likely led to a spurious relationship between CRAM and some of the EMAP metrics. Another possible

Table 7. Summary of changes to CRAM based on the calibration analyses.

	Type of Change					
Attributes/Metrics	Rescored Metric Based on Calibration Results	Refined Scaling of Metrics	Revised Wording/Clarified Narratives	Created Separate Narratives for a Wetland Type or Sub-type	Added more Guidance to Definitions and Field Indicators Sections	Added New Wetland Types: Vernal Pools and Playa
Buffer and Landscape Context						
Landscape Connectivity		X	X	X	X	X
Percent of Assessment Area with Buffer			X		X	X
Average Buffer Width		X	X		X	X
Buffer Condition		X	X		X	X
Hydrology						
Water Source			X		X	X
Hydroperiod or Channel Stability		X	X	X	X	X
Hydrologic Connectivity		X	X	X	X	X
Physical Structure						
Structural Patch Richness	X	X	X	X	X	X
Topographic Complexity			X	X	X	X
Biotic Structure						
Organic Matter Accumulation				Deleted Metric		
Vertical Biotic Structure	X		X	X	X	
Horizontal Interspersion and Zonation		X		X	X	X
Plant Community:	X	X	X	X	X	X
Number of Plant Layers Present	X	X	X	X	X	X
Percent Non-Native Plant Species	X	X	X	X	X	X
Native Plant Species Richness	X	X	X	X	X	X

explanation for the relatively poor relationships between CRAM and EMAP data is that estuarine plant communities may respond to stressors in a non-linear manner compared to CRAM metrics which were constructed based on a linear response model. Future analysis could employ a different set of statistical approaches to investigate potential nonlinear responses (Bedford and Preston 1988).

Fewer significant relationships between CRAM and LDI for estuarine wetlands (compared to riverine) also reflect the variable response of wetlands to landscape stressors. All riverine CRAM attributes were significantly correlated with LDI. However, LDI relationships were not significant for estuarine Physical and Biotic Structure attributes. This is likely because tidal forcing, rather than watershed stressors, largely controls the condition of estuarine wetlands. In addition, many California estuaries tend to be more intensively managed to promote wildlife functions (e.g., via active invasive plant control or treatment of watershed inflow), which further decouples landscape stressor effects from wetland condition. These discrepancies between CRAM validation results for estuarine and riverine wetlands further demonstrate the importance of using multiple validation measures in a weight-of-evidence approach, and understanding the factors that control each measure of condition or stress.

Calibration

Results were used to modify CRAM to improve its performance and validity. The main changes included providing better support documentation, guidance, and instructions; revising narratives for metric scoring; rescaling metrics; rescoring or re-binning metrics; eliminating or combining metrics; and creating new submetrics (Table 7). The most substantive changes included 1) rescaling the Buffer and Hydroperiod metrics to rectify the skewed distribution observed in the validation analysis and to better represent the distribution of scores across the range of condition; 2) restructuring the riverine Hydroperiod metric to focus on floodplain geomorphology; 3) combining the Physical and Biotic Structure metrics; and 4) refining the Buffer and Plant Community Composition metrics by creating submetrics. The submetric scores represent specific elements of the metrics and are aggregated to metric scores, which are then aggregated to attribute scores. For example, the ‘‘Percent of Assessment Area with Buffer’’, ‘‘Average Buffer Width’’, and ‘‘Buffer Condition’’ submetrics are combined into a single, ‘‘multidimensional’’ buffer metric, which is then combined with the Landscape Connectivity metric to generate an attribute score. This reduces the double counting of the buffer submetrics, as their combined weight is equal to that of the other metric in the Buffer and Landscape Context attribute. A similar approach was taken for the Plant Community Composition metric, for which submetrics were created to evaluate species richness, percent invasion, and structural complexity (based on the number of distinct plant layers present). Correlations between attributes and Level-3 data were re-analyzed following these changes to ensure that the modifications improved CRAM overall performance. This process will continue iteratively to provide for ongoing refinement of CRAM.

Standardization

As with any assessment, CRAM results should be viewed in light of the expected precision of the method. The repeatability analysis conducted allows for bounding of the confidence in CRAM output. From a management perspective, quantification of precision helps decision makers determine when differences in CRAM scores likely represent a true difference in condition as opposed to being within the expected error of the method. Following the modifications made as a result of this study, CRAM attribute scores should generally be considered precise within ± 10%, while overall CRAM scores should be considered precise within ± 6%. Higher precision at the overall score level results from the internal redundancies and ‘‘smoothing’’ of variability associated with combining attributes into an overall score. However, as with any multimetric assessment, a specific overall score can result from various combinations of attribute scores, and likewise for attribute scores resulting from various metric combinations. Therefore, CRAM results are best considered at both the overall score and attribute level to provide a more complete understanding of wetland condition.

Implications for Other Calibration/ Validation Efforts

This study demonstrates how data from existing monitoring and assessment programs can be used to calibrate RAMs. Ideally, validation would be done against an independent measure of condition that reflects the same elements as the RAM attribute of interest (e.g., Hydrology, Physical Structure). This ‘‘gold standard’’ measure would be independent of confounding factors associated with other elements of condition and would be collected concurrently with the RAM assessments. Obtaining this gold standard is difficult due to the challenge of identifying a unique measure of a single element of condition and the cost associated with creating this new data set. However, multiple indices that reflect condition along a gradient of disturbance can be used to provide a weight-of-evidence approach (Miller et al. 2004, DeZwart et al. 2006). Use of multiple validation measures is important because a precise match between RAM model output and validation data is not expected due to: 1) the inherent variability in natural systems; 2) different indices integrating different aspects of condition; 3) each index responding to different stressors and forcing functions; and 4) the fact that data are often collected over different spatial and temporal scales. The relationships between individual indices of condition will often be biased in one direction or another because of variable responses to natural environmental gradients and sensitivity to stressors (Hawkins 2006). It is virtually impossible to find response variables affected by a single forcing function or stressor (Karr and Chu 1999). If multiple relationships are concordant and consistent, it is reasonable to assume that the RAM results are accurately reflecting changes in condition relative to stressors on the wetland (Reiss and Brown 2007). The goal of validation should not be to maximize correlation with any one measure of biologic condition, but to optimize the method to achieve reasonable correlations with multiple measures of condition. This approach does not eliminate uncertainty in our conclusions; rather it provides a sound, transparent process for reducing uncertainty by integrating the best scientific information available at the time (Burton et al. 2002).

The analysis of CRAM relative to Level-1 or Level-3 data sources does not fit the traditional definition of calibration or validation. The purpose of calibration is to optimize the correspondence between RAM results and quantitative data for wetlands across a gradient of condition within a reference network (Brinson and Rheinhardt 1996) or to generate numeric scaling of metrics or variables (Hruby et al. 1999). In contrast, validation uses independent data sources to evaluate the accuracy of a RAM at assessing condition. True validation of assessment models of natural systems is impossible because natural systems are never closed and because model results are always non-unique (Oreskes et al. 1994). Furthermore, available Level-3 data sets are themselves indices of wetland condition based on floral and faunal community composition. Assessment models can only be evaluated in relative terms, and based on heuristic evidence from multiple independent measures of condition. Consequently, the overall RAM validation process includes elements that resemble both traditional calibration and validation (Oreskes et al. 1994, Janssen and Heuberger 1995). The ability to explain relationships observed in the data with well established ecological principles and understanding of wetland condition can serve to further validate RAM results. As with most biological models, CRAM performance should be continually refined as understanding of wetland condition improves and additional Level-3 data sets become available.

Final Thoughts

It is important to understand the limitations of RAMs. Despite rigorous validation that demonstrates the validity of a method, RAMs are only one tool for wetland monitoring and assessment. They are valuable in that they provide an inexpensive method that can be routinely and rapidly applied in a consistent manner across a range of wetland types. These features make RAMs valuable and reliable tools for general condition assessments, screening-level evaluations, and assessment of program performance. RAMs are not intended to replace intensive Level-3 data or to provide detailed information on specific wetland functions, support or health of particular species or communities, or detailed success of mitigation or restoration projects. When used in combination with Level-1 and Level-3 tools, validated RAMs fill a valuable niche in integrated assessment programs.

ACKNOWLEDGMENTS

CRAM development was funded by Section 104 Wetlands Protection Grants (Nos. CD-97918701-0 and CD-979-186-01) from the U.S. Environmental Protection Agency. CRAM could not have been developed without the assistance of Richard Sumner, Mary Kentula, Paul Jones, John Mack, M. Siobhan Fennessy, Cheryl McGovern, and the members of the California Rapid Assessment Method Core and Regional Teams, including Richard Ambrose, Ruben Guieb, Bobby Jo Close, John Dixon, Carl Wilcox, John Callaway, Andree Breaux, Dan Martel, Raymond Jay, Don Stevens, and Aaron Allen. Thanks are also extended to the multiple agencies and organizations that provided site access and validation data, particularly the Institute for Bird Populations and the California Department of Fish and Game. The calibration/validation approach and analysis benefited greatly from the technical assistance of Don Stevens, Michael Graham, and Max Coleman. John Callaway and Mary Kentula provided comments and suggestions on early drafts of the manuscript, which greatly improved the overall quality of the paper. We also thank two anonymous reviewers whose comments improved the overall clarity of this paper. This paper is dedicated to the memory of Jeff Robinson, our fellow CRAM team member who passed away unexpectedly while working in the estuarine wetlands he cared so deeply about.

LITERATURE CITED

Ambrose, R. F., J. C. Callaway, and S. F. Lee. 2006. An evaluation of compensatory mitigation projects permitted under Clean Water Act Section 401 by the California State Water Quality Control Board, 1991–2002. Report to the State Water Resources Control Board. University of California, Los Angeles, CA, USA.

Andreas, B. K., J. J. Mack, and J. S. McCormac. 2004. Floristic Quality Assessment Index (FQAI) for vascular plants and mosses for the State of Ohio. Ohio Environmental Protection Agency, Division of Surface Water, Wetland Ecology Group, Columbus, OH, USA.

Baber, M. J., E. Fleishman, K. J. Babbitt, and T. L. Tarr. 2004. The relationship between wetland hydroperiod and nestedness patterns in assemblages of larval amphibians and predatory macroinvertebrates. Oikos 107:16–27.

Bedford, B. L. and E. M. Preston. 1988. Developing a scientific basis for assessing cumulative effects of wetland loss and degradation on landscape functions: status, perspectives, and prospects. Environmental Management 12:751–72.

Brinson, M. M. and R. Rheinhardt. 1996. The role of reference wetlands in functional assessment and mitigation. Ecological Applications 6:69–76.

Brooks, R. P., T. J. O'Connell, D. H. Wardrop, and L. E. Jackson. 1998. Towards a regional index of biological integrity: the example of forested riparian ecosystems. Environmental Monitoring and Assessment 51:131–43.

Brown, M. T. and M. B. Vivas. 2005. Landscape development intensity index. Environmental Monitoring and Assessment 101:289–309.

Bryce, S. A., R. M. Hughes, and P. R. Kaufmann. 2002. Development of a bird integrity index: using bird assemblages as indicators of riparian condition. Environmental Management 30:294–310.

Burton, G. A., Jr., P. M. Chapman, and E. P. Smith. 2002. Weight-of-evidence approaches for assessing ecosystem impairment. Human and Ecological Risk Assessment 8:1657–73.

Cohen, M. J., C. R. Lane, K. C. Reiss, J. A. Surdick, E. Bardi, and M. T. Brown. 2005. Vegetation based classification trees for rapid assessment of isolated wetland condition. Ecological Indicators 5:189–206.

Collins, J. N., E. D. Stein, M. Sutula, R. Clark, A. E. Fetscher, L. Grenier, C. Grosso, and A. Wiskind. 2006. California rapid assessment method (CRAM) for wetlands and riparian areas. Version 4.5, San Francisco Estuary Institute. Oakland, CA. www.cramwetlands.org.

Cowardin, L., V. Carter, F. Golet, and E. Laroe. 1979. Classification of wetlands and deepwater habitats of the United States. Washington, D. C.: United States Fish and Wildlife Service, Biological Services Program. FWS/OBS-79/31.

DeZwart, D., S. D. Dyer, L. Posthuma, and C. P. Hawkins. 2006. Predictive models attribute effects on fish assemblages to toxicity and habitat alteration. Ecological Applications 16: 1295–1310.

Fennessy, M. S., A. D. Jacobs, and M. E. Kentula. 2004. Review of rapid methods for assessing wetland condition. National Health and Environmental Effects Research Laboratory, Office of Research and Development, Research Triangle Park, North Carolina, EPA/620/R-04/009.

Fennessy, M. S., A. D. Jacobs, and M. E. Kentula. 2007. An evaluation of rapid methods for assessing the ecological condition of wetlands. Wetlands 27:543–60.

Griffith, M. B., B. H. Hill, F. H. McCormick, P. R. Kaufmann, A. T. Herlihy, and A. R. Selle. 2005. Comparative application of indices of biotic integrity based on periphyton, macroinvertebrates, and fish to southern Rocky Mountain streams. Ecological Indicators 5:117–36.

Harrington, J. M. 1999. California stream bioassessment procedure. California Department of Fish and Game, Water Pollution Control Laboratory. Rancho Cordova, CA.

Hauer, F. R., B. J. Cook, M. C. Gilbert, E. J. Clairain, Jr., and R. D. Smith. 2002. A regional guidebook for applying the hydrogeomorphic approach to assessing wetland functions of intermontane prairie pothole wetlands in the Northern Rocky Mountains. U.S. Army Engineer Research and Development Center, Vicksburg, MS, Technical Report ERDC/EL TR-02-7.

Hawkins, C. P. 2006. Quantifying biological integrity by taxonomic completeness: its utility in regional and global assessment. Ecological Applications 16:1277–94.

Hill, A. J., V. S. Neary, and K. L. Morgan. 2006. Hydrologic modeling as a development tool for HGM functional assessment models. Wetlands 6:161–80.

Hruby, T., T. Granger, K. Brunner, S. Cooke, K. Dublanica, R. Gersib, L. Reinelt, K. Richter, D. Sheldon, E. Teachout, A. Wald, and F. Weinmann. 1999. Methods for assessing wetland functions volume I: riverine and depressional wetlands in the lowlands of western Washington. Washington State Department Ecology Publication #99-115.

Institute for Bird Populations (IBP). 2006. The monitoring avian productivity and survivorship (MAPS) program, Point Reyes Station, CA, http://www.birdpop.org/maps.htm.

Janssen, P. H. M. and P. S. C. Heuberger. 1995. Calibration of process-oriented models. Ecologic Modeling 83:55–66.

Karr, J. R. and E. W. Chu. 1999. Restoring life in running waters: better biological monitoring. Island Press, Washington, DC.

Kentula, M. E. 2007. Monitoring wetlands at the watershed scale. Wetlands 27:412–15.

Lee, L. C., P. L. Fiedler, S. R. Stewart, D. J. Partridge, J. A. Mason, E. M. Inlander, and M. C. Rains. 2003. Draft operational guidebook for assessment of the functions of riverine waters/wetlands in the Santa Margarita watershed, Riverside & San Diego counties, California. San Diego Regional Water Quality Control Board, Technical Publication. San Diego, CA.

Linthurst, R. A., L. A. Mulkey, M. W. Slimak, G. D. Veith, and B. M. Levinson. 2000. Ecological research in the office of research and development at the U.S. environmental protection agency: an overview of new directions. Environmental Toxicology and Chemistry 19:1222–29.

Mack, J. J. 2001. Vegetation Indices of Biotic Integrity (VIBI) for Wetlands: Ecoregional, Hydrogeomorphic, and Plant Community Comparison with Preliminary Aquatic Life Use Designations. Final Report to U.S. EPA Grant No. CD985875, Volume 1. Wetland Ecology Group, Division of Surface Water, Ohio Environmental Protection Agency, Columbus, OH.

Mack, J. J. 2006. Landscape as a predictor of wetland condition: an evaluation of the landscape development index (LDI) with a large reference wetland dataset from Ohio. Environmental Monitoring and Assessment 120:221–41.

Micacchion, M. 2004. Integrated wetland assessment program. part 7: amphibian index of biotic integrity (AmphIBI) for Ohio wetlands. Ohio Environmental Protection Agency, Wetland Ecology Group, Division of Surface Water, Columbus, OH, Ohio EPA Technical Report WET/2004-7.

Miller, S. W., Q. D. Skinner, and K. J. Reddy. 2004. Stream assessment using biotic indices: responses to physicochemical variables. Journal of the American Water Resources Association 40:1173–87.

Ode, P. R., A. C. Rehn, and J. T. May. 2005. A quantitative tool for assessing the integrity of southern coastal California streams. Environmental Management 35:493–504.

Oreskes, N., K. Shraderfrechette, and K. Belitz. 1994. Verification, validation, and confirmation of numerical-models in the earth-sciences. Science 263:641–46.

Reiss, K. C. and M. T. Brown. 2007. Evaluation of Florida palustrine wetlands: application of USEPA levels 1, 2, and 3 assessment methods. EcoHealth 4:206–18.

Resh, V. H. and J. K. Jackson. 1993. Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. p. 195–233. *In* D. M. Rosenberg and V. H. Resh (eds.) Freshwater Biomonitoring and Benthic Macroinvertebrates. Chapman and Hall, New York, NY, USA.

Smith, R. D., A. Ammann, C. Bartoldus, and M. M. Brinson. 1995. An approach for assessing wetland functions using hydrogeomorphic classification, reference wetlands, and functional indices. Waterways Experiment Station, U.S. Army Corps of Engineers, Vicksburg, MS, Technical Report WRP-DE-9.

Smith, E. P., I. Lipkovich, and K. Ye. 2002. Weight-of-evidence (WOE): quantitative estimation of probability of impairment for individual and multiple lines of evidence. Human and Ecological Risk Assessment 8:1585–96.

Stapanian, M. A., T. A. Waite, G. Krzys, J. J. Mack, and M. Micacchion. 2004. Rapid assessment indicator of wetland integrity as an unintended predictor of avian diversity. Hydrobiologia 520:119–26.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological Applications 16:1267–76.

Sutula, M., J. N. Collins, J. Callaway, T. Parker, M. Vasey, and E. Wittner. 2001. Quality assurance project plan (QAPP) for environmental monitoring and assessment program, west coast pilot 2002 intertidal assessment: California intensification. Southern California Coastal Water Research Project. Costa Mesa, CA.

Sutula, M. A., E. D. Stein, J. N. Collins, A. E. Fetscher, and R. Clark. 2006. A Practical Guide for the Development of a Wetland Assessment Method: The California Experience. Journal of the American Water Resources Association 42: 157–75.

Talmage, P. J., J. A. Perry, and R. M. Goldstein. 2002. Relation of instream habitat and physical conditions to fish communities of agricultural streams in the northern Midwest. North American Journal of Fisheries Management 22: 825–33.

U.S. Environmental Protection Agency (USEPA). 2002. Research strategy: environmental monitoring and assessment program. Office of Research and Development. Research Triangle Park. p. 78, http://www.epa.gov/emap/html/pubs/docs/resdocs/resstrat02.html.

Wardrop, D. H., M. E. Kentula, D. L. StevensJr, S. F. Jensen, and R. P. Brooks. 2007. Assessment of wetland condition: an example from the Upper Juniata watershed in Pennsylvania, USA. Wetlands 27:416–30.

Whittier, T. R., R. M. Hughes, J. L. Stoddard, G. A. Lomnicky, D. V. Peck, and A. T. Herlihy. 2007. A structured approach for developing indices of biotic integrity: three examples from streams and rivers in the western USA. Transactions of the American Fisheries Society 136:718–35.

Manuscript received 26 December 2007; accepted 5 December 2008.