RESEARCH ARTICLE

Scaling of classification systems—effects of class precision on detection accuracy from medium resolution multispectral data

Daniel Gann · **Jennifer Richards**

Abstract

Context Land-cover class definitions are scale-dependent. Up-scaling categorical data must account for that dependence, but most decision rules aggregating categorical data do not produce scale-specific class definitions. However, non-hierarchical, empirically derived classification systems common in phytosociology define scale-specific classes using species co-occurrence patterns.

Objectives Evaluate tradeoffs in class precision and representativeness when up-scaling categorical data across natural landscapes using the multi-dimensional grid-point (MDGP)-scaling algorithm, which generates scale-specific class definitions; and compare spectral detection accuracy of MDGP-scaled classes to 'majority-rule' aggregated classes.

Methods Vegetation maps created from 2-m resolution WorldView-2 data for two Everglades wetland areas were scaled to the 30-m Landsat grid with the MDGP-scaling algorithm. A full-factorial analysis evaluated the effects of scaled class-label precision and class representativeness on compositional information loss and detection accuracy of scaled classes from multispectral Landsat data.

Results MDGP-scaling retained between 3.8 and 27.9% more compositional information than the majority rule as class-label precision increased. Increasing class-label precision and information retention also increased spectral class detection accuracy from Landsat data between 1 and 8.6%. Rare class removal and increase in class-label similarity were controlled by the class representativeness threshold, leading to higher detection accuracy than the majority rule as class representativeness increased.

Conclusions When up-scaling categorical data across natural landscapes, negotiating trade-offs in thematic precision, landscape-scale class representativeness and increased information retention in the scaled map results in greater class-detection accuracy from lower-resolution, multispectral, remotely sensed data. MDGP-scaling provides a framework to weigh tradeoffs and to make informed decisions on parameter selection.

D. Gann (✉) · J. Richards
Department of Biological Sciences, Florida International University, Miami, FL, USA
e-mail: gannd@fiu.edu

D. Gann
Institute of Environment, Florida International University, Miami, FL, USA

Introduction

Classification systems represent generalized human perceptions of the world that group objects with similar properties. The properties and degree of generalization in a classification system are often determined by discipline and are scale specific. To model and understand physical and biological processes that operate at different spatial scales, ecologists frequently use categorical land-cover information that standardizes and categorizes land units (pixels or grid cells) across large spatial regions (e.g., landscape to continental or global scale). For classification systems to be useful when modeling a specific process in a spatially explicit fashion, the classes in a classification system need to be (1) recognizable and identifiable on the ground at the scale of analysis and (2) detectable from remotely sensed data across the entire (spatially exhaustive) landscape of interest. However, since data are often acquired and interpreted at different spatial scales, information from different sources often needs to be scaled up to become compatible among sources. Hence, up-scaling, the process of information aggregation also referred to as coarse-graining, has received much attention in the geographic branches of many scientific disciplines (Wu and Hobbs 2002; Lischke et al. 2007; Teng et al. 2020). Newman et al. (2019) identified the problem of up-scaling in a statistically unbiased manner as one of three intrinsic limitations to progress in landscape ecology. In the context of understanding land-cover dynamics and their effects on organisms, three aspects of information aggregation that are not sufficiently addressed are (1) scale-dependency of classification schemes, (2) effects of up-scaling on compositional information loss, and (3) how that reduction in information affects the accuracy of land-cover detection from remotely sensed data. These aspects are important to consider when modeling interactions of landscape patterns and ecological processes at the landscape scale.

Interdisciplinary approaches to problem-solving often lead to new methods in science. This study integrates concepts from three disciplines—landscape ecology, phytosociology, and remote sensing—to establish a new method to scale categorical information across spatial scales. The theory of scaling in landscape ecology is combined with concepts from phytosociology that construct scale-specific, representative vegetation classifications. Further coupling the process of defining a scaled classification system with the principles of accurate delineation of classes from remotely sensed data provides a feedback process for refining landscape-specific, cross-scale classification systems. The main objective of this study is to demonstrate the efficacy of the multi-dimensional grid-point (MDGP)-scaling algorithm (Gann 2019) in generating scale-specific and non-hierarchical classification schemes that can be effectively detected from multi-spectral remotely sensed data.

Up-scaling categorical data in landscape ecology.

Despite the ever-increasing spatial resolution of satellite data and aerial photography, in order to detect and quantify changes in land cover and land use across long temporal scales, up-scaling of more recent very high-resolution categorical maps to the courser resolution of older maps is required. Up-scaling categorical information aggregates information from multiple original map objects (e.g., pixels, grid cells) at the initial resolution, and information is generalized and lost in the process. Up-scaling categorical information is more problematic than up-scaling continuous data because of limited mathematical or statistical methods. The most frequently used decision rules to aggregate compositional information are the majority, random, and nearest-neighbor rules. Studies that evaluated the effect of these rules on class abundance and landscape metrics (Turner et al. 1989; O'Neill et al. 1996) showed that common and clumped classes were overrepresented, while rare and dispersed classes disappeared or were underrepresented (Turner et al. 1989; He et al. 2002; Wu et al. 2002; Raj et al. 2013; Coulston et al. 2014). Even more complex spatial aggregation methods that attempt to preserve rare classes (Coulston et al. 2014) only consider the original classes when class labels are assigned to the larger spatial units. When simply replacing mixes of multi-class landscape units at a lower resolution with a single non-mixed class label, the classification scheme of the original scale by default is assumed to be valid at the aggregated scale. However, even if this assumption is justifiable when scale change is small, for large scale factors and especially in heterogeneous landscapes, this assumption leads to a huge, often unquantified, loss of compositional information.

When aggregating a landscape, we can quantify the compositional information that is retained in the scaled landscape units with the Czekanowski Index (Czekanowski 1909), which we define here as the local compositional information retention (IR_c) as

$$IR_c = \Sigma_{i=1}^{N} \min \left(PC_i, PC_iS \right) \qquad (1)$$

where PC_i is the proportion of class i cells within the scaled grid cell and PC_iS is the proportion of class i retained in the scaled class label. Quantifying information loss using Eq. (1) when scaling a small landscape with the majority and nearest-neighbor rules is illustrated in Fig. 1. Aggregating at a small-scale factor of three leads to large compositional information loss (Fig. 1). Information loss increases as the scale factor increases to nine, where the application of these rules leads to a single class at the lower scale, although which class is retained differs between the rules (Fig. 1). Gann (2019) has demonstrated the impact on information loss that applying these decision rules has when scaling simulated landscapes with varying known characteristics. This over-simplification of land-cover information at the aggregated scale ultimately propagates to ecological models, where frequently co-occurring class mixtures that might be ecologically meaningful cannot be considered because they are lost in the scaling process. Ecological models that rely on scaled data then suffer from this oversimplification, when relationships between processes operating at the aggregated scale of landscape patterns are obscured or erroneously rendered significant.

The few aggregation methods that acknowledge mixed classes at coarser spatial scales often use hierarchical class systems that ignore non-hierarchical compositions of natural systems (Wu and David 2002a; Ju et al. 2005). Ju et al. (2005) developed a multi-scale, multi-granular framework that allows for scaling in the spatial domain using quad-tree data structures to increase flexibility for aggregation in the spatial and categorical domains. The categorical domain, however, was limited to hierarchical class labels that aggregated the finer scale class labels to coarser, predetermined class labels at the next hierarchical level (Ju et al. 2005). Hierarchical classification systems aggregate linearly, and groups from a lower level belong to only a single group at the higher level. However, class or species associations result

from processes that operate at different spatial and temporal scales and do not necessarily lead to hierarchical class systems. For instance, individual pixels labeled trees at a high spatial resolution (i.e., class: tree) can, at lower resolutions, become members of forests, woodlands, swamps, or savannas; the tree density, environmental conditions, and floristic characteristics of the co-occurrence with other species determine the coarser resolution class membership of individual trees. Thus, hierarchical classification systems can over-simplify complex patterns of spatial heterogeneity and obscure community assembly rules that determine species and class co-occurrences.

Class detection at varying spatial scales in remote sensing

Similarly, classification systems applied in land-cover mapping from remotely sensed data are often structured hierarchically, presenting over-generalized classes that were detected from remotely sensed data at medium to coarse resolutions. Classification accuracy, the label correctly representing the ground condition, depends on the spatial resolution of the sensor in relation to the local heterogeneity of the landscape and the thematic resolution (detail) of the classification system. Like the scaling of categorical maps, evaluation of class accuracy and quantification of class abundance across sensors with different spatial resolutions is generally restricted to coarse classification systems that do not vary with spatial scales, even when scales vary by magnitudes (Raptis et al. 2003; Knight et al. 2013; Xu et al. 2019, 2021). In remote sensing, the class that is dominant is often accepted as the correct class for mixed pixels (Ozdogan and Woodcock 2006), even when in cases of high local diversity, that dominance can be much smaller than 50%. Using sub-pixel fractions of the correct classes has been proposed to adjust accuracy estimates and to better estimate actual land-cover abundances (Latifovic and Olthof 2004; Pontius and Connors 2009), but they are still not the norm because of statistical challenges (Stehman and Foody 2019). The use of a single classification system across multiple scales and the practice of accepting the dominant class as correct at the coarser scale not only leads to gross over- or under-representation of land cover classes (Fig. 1) but also makes precise quantification of change in
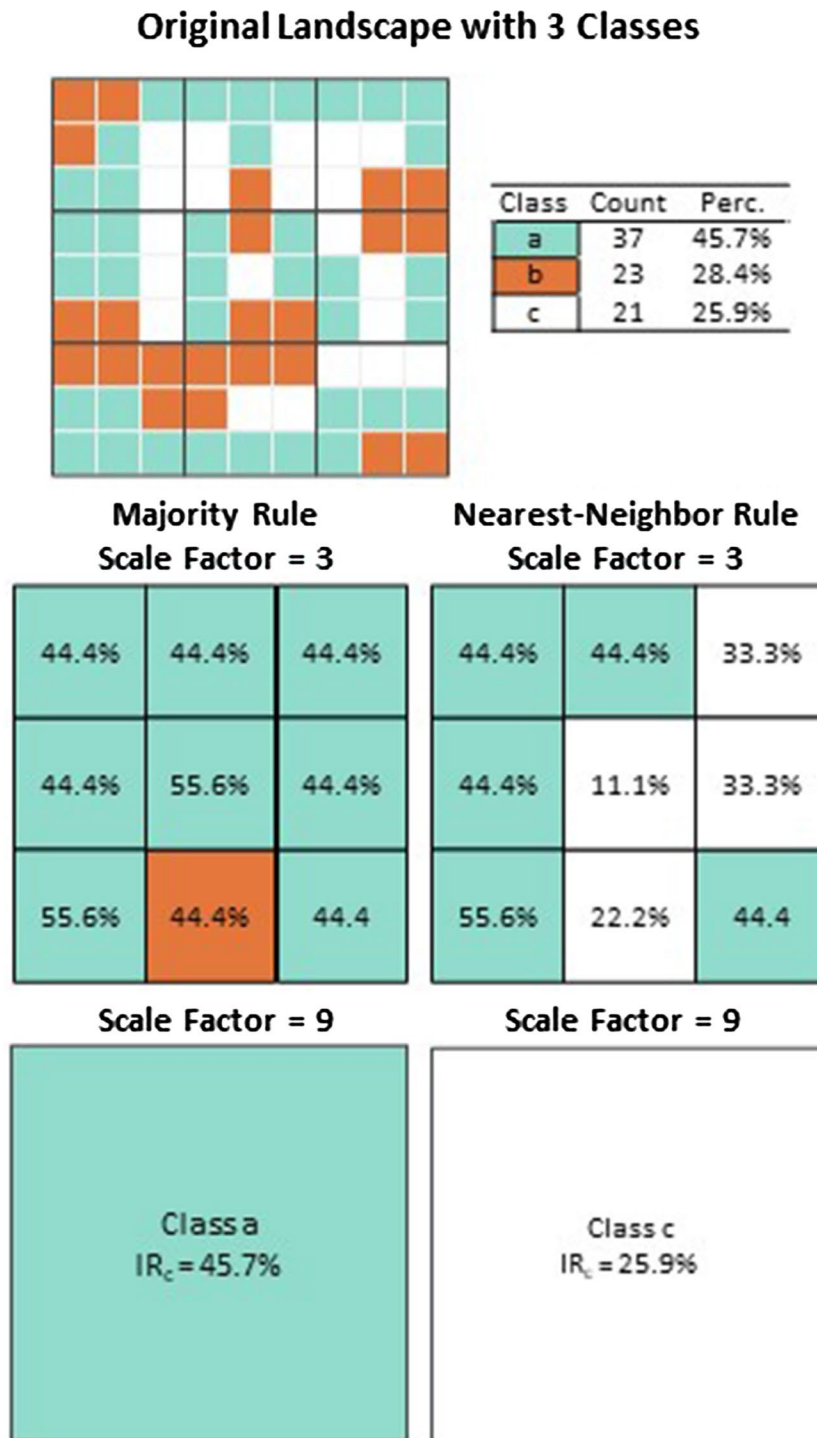
Original Landscape with 3 Classes

Fig. 1 Effects of majority and nearest-neighbor decision rules on compositional information retention (IR_c) in a spatially heterogeneous and dispersed landscape with three classes. Top: The original landscape with 81 grid cells (left) and class abundances as counts and percentages (right). Bottom: original landscape upscaled with the majority rule (left) and nearest-neighbor rule (right) with scale factors of 3 (top) and 9 (bottom). Percentage within each scaled grid cell is the compositional information retention (IR_c) for the scaled cell

very heterogeneous landscapes with mixed pixels impossible.

Generating classification systems in phytosociology

Solutions to the disconnect between thematic resolution of classification schemes and spatial resolution are provided by phytosociology, which has a long history of generating representative classification systems at different spatial scales from quantitative measures of species co-occurrence in relevés (Braun-Blanquet 1964; Van Der Maarel 1979). In phytosociology, association patterns of species sampled at a 1-m^2 scale differ from those at 900-m^2 (e.g., a Landsat pixel), resulting in classification schemes that recognize this scale-dependence of co-occurrence as analytical scales vary. Further, vegetation classification systems that are driven by quantitative analysis need to be robust to sampling error and to consistently assign class labels to random samples within the spatial domain they represent (De Cáceres et al. 2009; Wildi 2010; Tichý et al. 2011; De Cáceres and Wiser 2012). Samples drawn from a categorical raster map within a window of a specific size (e.g., 3×3 aggregation kernel) resemble relevé data of species abundance for quadrats or plots, where each sampled grid cell represents a plot of a relevé set with a relative abundance of each class. Hence, the principles and methods applied in phytosociology can be extended to classification systems when scaling categorical data and detecting mixed land-cover classes from remotely sensed data.

Unifying scaling framework

The multi-dimensional grid-point (MDGP)-scaling algorithm (Gann 2019) is, to our knowledge, the only published algorithm that generates scale-specific and non-hierarchical classification schemes that recognize non-hierarchical co-occurrence patterns or class/species associations across scales. The algorithm is founded on principles of compositional data analysis and phytosociology. Using local class abundances at the higher resolution, this algorithm generates a new classification system that reflects common class co-occurrence frequencies from the higher resolution data and assigns new class labels to the lower resolution grid cells. The two user-determined parameters that control information loss and class definitions at

the lower resolution are class-label precision (parts) and a representativeness threshold. Class-label precision determines how much detail of the original, location-specific compositional information (i.e., IR$_c$) is retained in a scaled grid cell. Because we are dealing with categorical raster data, the number of cells within a scaled grid cell is finite, and the relative abundance data are compositional count data or frequencies that fill the space of a polytope where class count (richness) determines the number of polytope axes (Gann 2019). Class-label precision, expressed as parts, proportions, or percentage cover, determines the number and location of grid points in the multi-dimensional polytopes that represent the potential class labels (Fig. 2, black dots in ternary plots).

The effects of the class-label precision parameter on information retention are demonstrated in Fig. 2, where the MDGP-scaling algorithm was applied to the original landscape in Fig. 1. In what follows, we use parts and the corresponding precision of percentage cover interchangeably (e.g., 1-part = 100%, 2-part = 50%, 3-part = 33%, 4-part = 25% and 5-part = 20%). Information retention increases rapidly from 45.7% for the 100% precision (majority rule) to 78.4%, 87.6% and 95.7% for the 50%, 33%, and 25% precision of class labels, respectively (Fig. 2). However, with increasing label precision, the number of potential scaled classes increases exponentially. In the three-class example with a scale factor of nine, increasing label precision from 1-part (100%) to 4-part (25%), the class number increased from three to 15 classes (Fig. 2). Increasing the number of original classes to 5, a 25% label precision produces 70 potential classes. To control the number of classes as class count and label precision increase, rare mixed classes are eliminated by the MDGP-scaling algorithm by applying a representativeness threshold, which sets the lowest acceptable proportion across the landscape for a scaled class to be retained in the scaled classification scheme (Gann 2019).

Further, to address sampling error related to the generation of a classification system from one randomly selected and arbitrary origin for the lower resolution grid (i.e., remote sensor grid), and to be able to consistently assign class labels to random samples within the spatial domain the labels represent, the MDGP-scaling algorithm evaluates class-label fidelity (CLF). Class-label fidelity is represented by the mean probability of a class to occur across
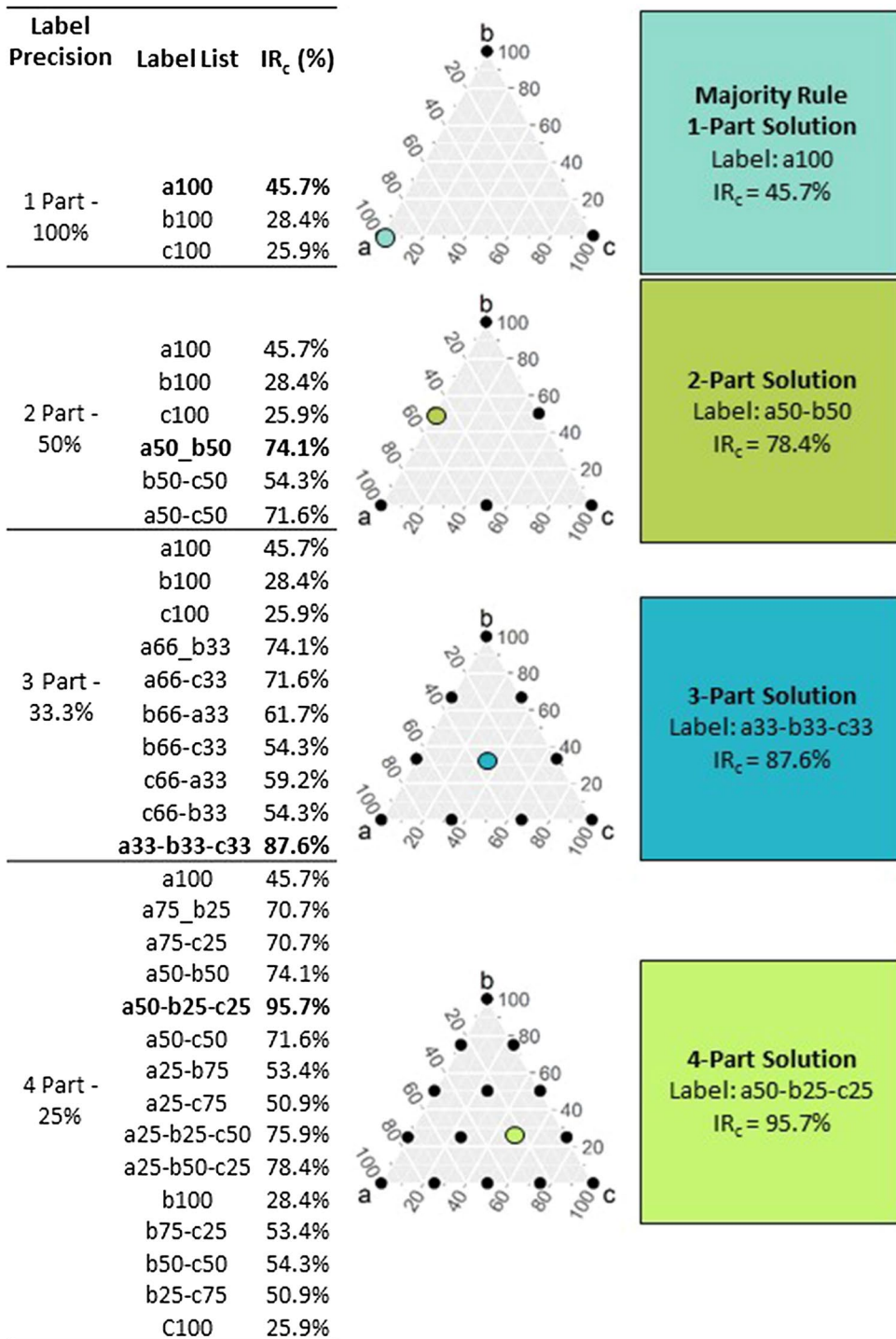
Label Precision	Label List	IR_c (%)
1 Part - 100%	**a100**	**45.7%**
	b100	28.4%
	c100	25.9%
2 Part - 50%	a100	45.7%
	b100	28.4%
	c100	25.9%
	a50_b50	**74.1%**
	b50-c50	54.3%
	a50-c50	71.6%
3 Part - 33.3%	a100	45.7%
	b100	28.4%
	c100	25.9%
	a66_b33	74.1%
	a66-c33	71.6%
	b66-a33	61.7%
	b66-c33	54.3%
	c66-a33	59.2%
	c66-b33	54.3%
	a33-b33-c33	**87.6%**
4 Part - 25%	a100	45.7%
	a75_b25	70.7%
	a75-c25	70.7%
	a50-b50	74.1%
	a50-b25-c25	**95.7%**
	a50-c50	71.6%
	a25-b75	53.4%
	a25-c75	50.9%
	a25-b25-c50	75.9%
	a25-b50-c25	78.4%
	b100	28.4%
	b75-c25	53.4%
	b50-c50	54.3%
	b25-c75	50.9%
	C100	25.9%

Fig. 2 MDGP solutions (left), ternary plots (center) and scaled grid cells (right) for majority rule (1-part, 100% precision) and MDGP 50% to 25% class label precisions. Left: Possible scaled class labels (Label List) for MDGP 100% through 25% label precisions with maximum compositional information retention (IR_c) in bold; the numbers in each label are the MDGP-percentages for the associated class. Center: Ternary plots for MDGP; black dots are the possible MDGP labels for each label precision (Label List), and the colored point is the MDGP that maximizes IR_c for the scaled grid cell. Right: Scaled solution for each precision; IR_c increases greatly with mixed classes

random-origin scaling results (CLF$_m$) or the proportion of classes that had a recurrence probability of one (CLF$_p$). To generate representative classification schemes, class-label precisions and representativeness thresholds that do not produce high fidelity (i.e., that have high sampling error for a specific landscape) can be easily identified and avoided (Gann 2019).

The MDGP-scaling algorithm was developed and tested in simulated landscapes (Gann 2019). Here, we evaluate its performance when up-scaling vegetation cover in real landscapes and detecting the scaled land-cover classes from remotely sensed, multispectral data. In this study, because of the repercussions that map precision (Meentemeyer and Box 1987; Quattrochi 1991; Buyantuyev and Wu 2007; Wickham and Riitters 2019; Halstead et al. 2022) and map accuracy (Langford et al. 2006; Kleindl et al. 2015) have on modeling landscape-scale patterns and processes, our main objectives were to quantify the effects of MDGP-scaling parameters on (1) information retention and class representativeness when upscaling categorical data across natural, heterogeneous landscapes and (2) the accuracy of detecting scaled classes from remotely sensed data from lower resolution multispectral satellite data.

Methods

Effects of the MDGP-scaling parameters on landscape-level information retention and class detectability using remotely sensed data were evaluated for two natural landscape types within the greater Everglades ecosystem (FL, USA) (Fig. 3). Plant communities for these two landscapes had been mapped from bi-season WorldView-2 (WV-2; Maxar Technologies, Westminster, CO) multispectral data (eight spectral bands ranging from 0.4 to 1.04 μm) at a spatial resolution of 2 m (Richards et al. 2015; Gann 2018). For this study, both mapped landscapes were scaled to 30 m using the MDGP-scaling algorithm (Fig. 4), and class detection accuracy of the scaled community classes was evaluated for Landsat Thematic Mapper multispectral data (six spectral bands ranging from 0.45 to 2.35 μm) that had been acquired for the same time (Fig. 4).

The integrated testing framework for scaling and spectral-detection analysis was coded in R (R Core Team 2016), making extensive use of packages

"raster" (Hijmans and van Etten 2010), "rgdal" (Bivand et al. 2013), "compositions" (van den Boogaart and Tolosana-Delgado 2008), "caret" (Kuhn et al. 2016), "ggplot2" (Wickham 2016) for all graphs, and the "MDGP-scaling" algorithm which was scripted in R (Gann, 2019) and is available as an R package at https://github.com/gannd/landscapeScaling. Maps were made in ArcGIS Pro 2.9.

Study areas

The two natural landscapes we studied were (1) a healthy, ridge-and-slough patterned landscape within southern Water Conservation Area 3A (WCA3A) and (2) a degraded, sawgrass-dominated, wet prairie in northeast Shark River Slough (NESRS) (Fig. 3), both within the larger Everglades wetland system in southern Florida, USA. The ridge-and-slough landscape dominated the undeveloped Everglades, while the degraded slough is disturbed habitat currently undergoing restoration (McVoy et al. 2011).

Water Conservation Area 3A—ridge and slough

The ridge-and-slough landscape of WCA3A is characterized by alternating sawgrass (*Cladium jamaicense*) ridges and deeper sloughs that are dominated by submerged aquatic, floating broadleaved, and emergent graminoid freshwater species; higher elevations can have woody shrubs and various tree species. The most common slough species is *Nymphaea odorata*, which forms dense, floating-leaved carpets and is often accompanied by species of *Utricularia* and floating mats of periphyton. The 2-m-scale community classification scheme for the ridge-and-slough landscape was composed of eight classes, including two each in aquatic-submerged, broadleaved-floating, graminoid- and broadleaved-emergent vegetation, a mixed shrub-marsh class, and a shrub-tree class (Fig. 5, Table 1).

Northeast Shark River Slough—human-influenced wet prairie

The second landscape, a 4.19 km^2, sawgrass-dominated wet prairie in NESRS (Fig. 3), is a degraded, former ridge-and-slough landscape that experienced decades of altered hydrology with decreased water depth and hydroperiod, causing a reduction in
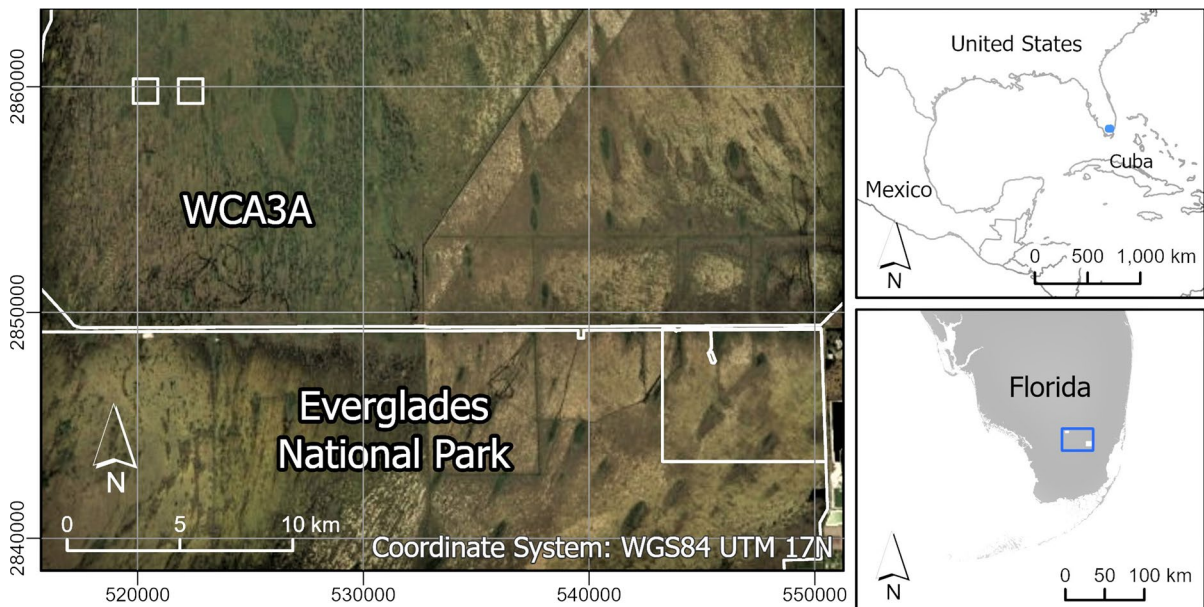
Fig. 3 Study areas (white outline) in the Everglades of South Florida, USA. Two 1.21 km² areas within Water Conservation Area 3A (WCA3A) and one 4.19 km² polygon in North-east Shark River Slough within the boundaries of Everglades National Park. Background imagery provided by Earthstar Geographics (Esri 2021, ArcGIS PRO 2.9.3)

topographic relief (McVoy et al. 2011). As a consequence, the slough communities transitioned into remnant shallow depressions dominated by sedges and other graminoids that form distinct patches within a matrix of sawgrass-dominated communities. The classification scheme for NESRS consisted of 14 community classes: two included broadleaved species; six, graminoid-dominated vegetation; four, shrub or tree components; and two were non-vegetation classes (Fig. 6, Table 2).

High-resolution plant community maps

The high-resolution plant communities that served as the basis for the scaling evaluation had been mapped from bi-season WV-2 data at a 2-m spatial resolution using the random forest classifier (Breiman 1984). The vegetation map for WCA3A used wet-season data acquired on October 20, 2012, and dry-season data from May 5, 2011, and had a design-based estimated 95% confidence accuracy of 91.2% (Fig. 5) (Gann 2018). WV-2 satellite data for the NESRS map had been acquired on November 6 and 9, 2010 (wet season) and on May 6, 2013 (dry season), and the

design-based overall accuracy for the map was 89.2% (Fig. 6) (Richards et al. 2015).

Scaling parameter evaluation for information retention and class-label fidelity

The three parameters of MDGP-scaling that control information retention in the scaled classification scheme are (1) scale factor, (2) class-label precision, and (3) class representativeness, where

(1) Scale factor is the ratio of the lower resolution to that of the higher resolution.
(2) Class-label precision is the minimum proportion of a class that is retained in aggregated class labels.
(3) Representativeness of a scaled class is the minimum proportion of the larger landscape that a newly generated class must occupy to not get dropped from the scaled classification scheme.

The scale factor for this study was 15 for scaling from WV-2 2-m resolution to a 30-m Landsat grid-cell. Relative abundances of classes for the 225 WV-2 grid cells in each coarse-resolution grid cell
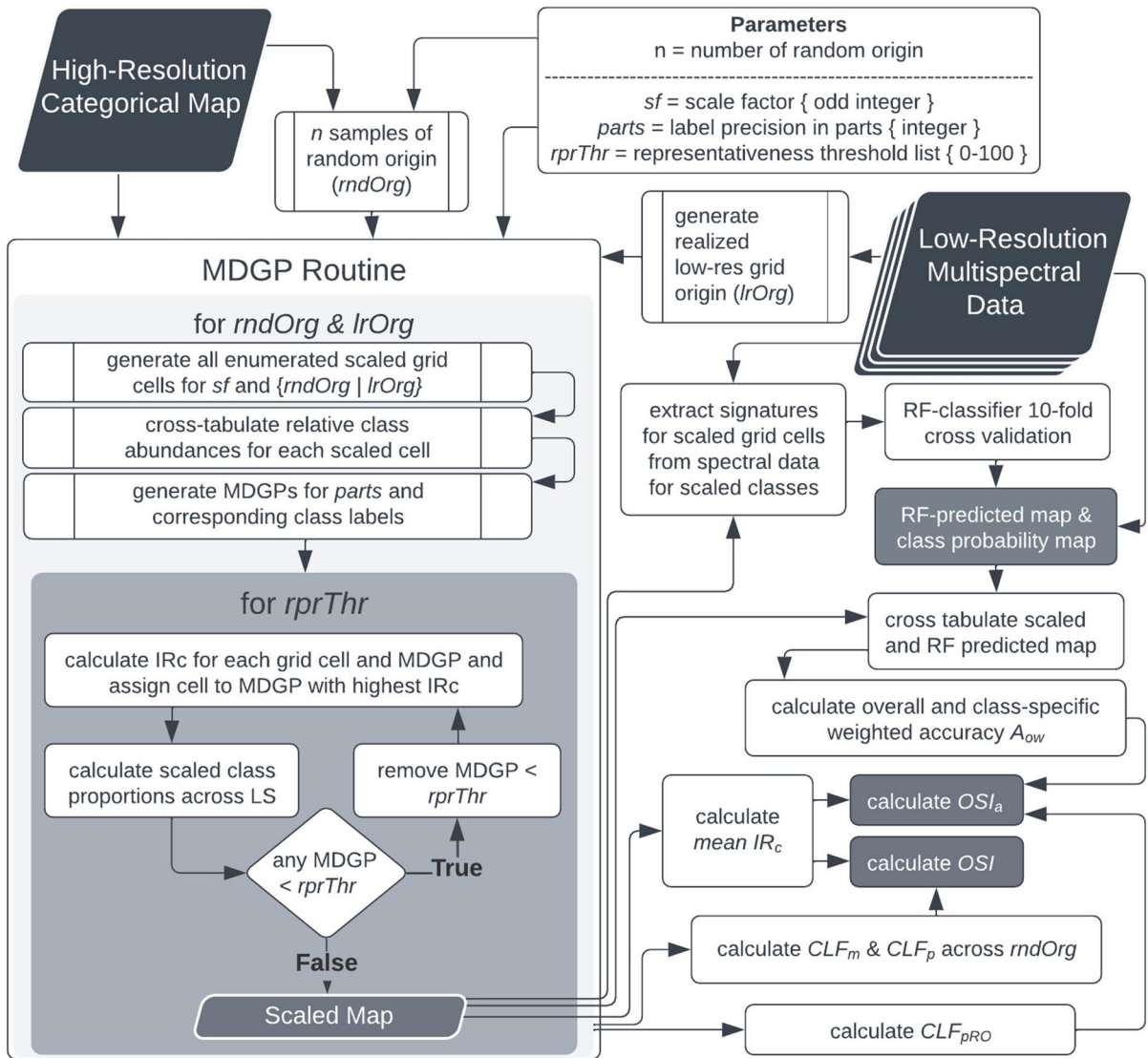
Fig. 4 Flowchart for evaluating effects of MDGP-scaling parameters for class precision (*parts*) and representativeness thresholds (*rprThr*) on *mean IR$_c$*, *CLF$_m$*, *CLF$_p$*, *CLF$_{pRO}$* and on scaled-class detection accuracy (A_{ow}) from low-resolution multispectral data. Algorithm inputs in black boxes; algorithm outputs in darker gray boxes. *lrOrg* low-resolution grid origin; *mean IR$_c$* the mean of compositional information retention across all cells of the scaled map of the realized grid of the low-resolution multispectral data, *CLF$_m$* mean probability of a class to occur across random origin scaling results; *CLF$_p$* class-label fidelity as a proportion of classes recurring across all random map origin (RO) samples; *CLF$_{pRO}$* proportion of classes generated for the realized Landsat grid that occur across all classes generated for 10 random origin solutions; A_{ow} weighted overall accuracy; *OSI* optimal scaling index; *OSI$_a$* optimal scaling solution considering weighted class-detection accuracy, *mean IR$_c$*, and *CLF$_{pRO}$*

were determined, and MDGP scaled class labels were generated. The algorithm was applied in a full factorial design of five options each for class-label precision and landscape representativeness. The options considered were class-label precisions of 1, 2, 3, 4, and 5 parts, which translate to 100% (equivalent to the majority rule), 50%, 33%, 25% and 20% label precisions. For landscape representativeness, thresholds of 1%, 5%, 10%, 15% and 20% were evaluated. Monotypic classes that were below the landscape
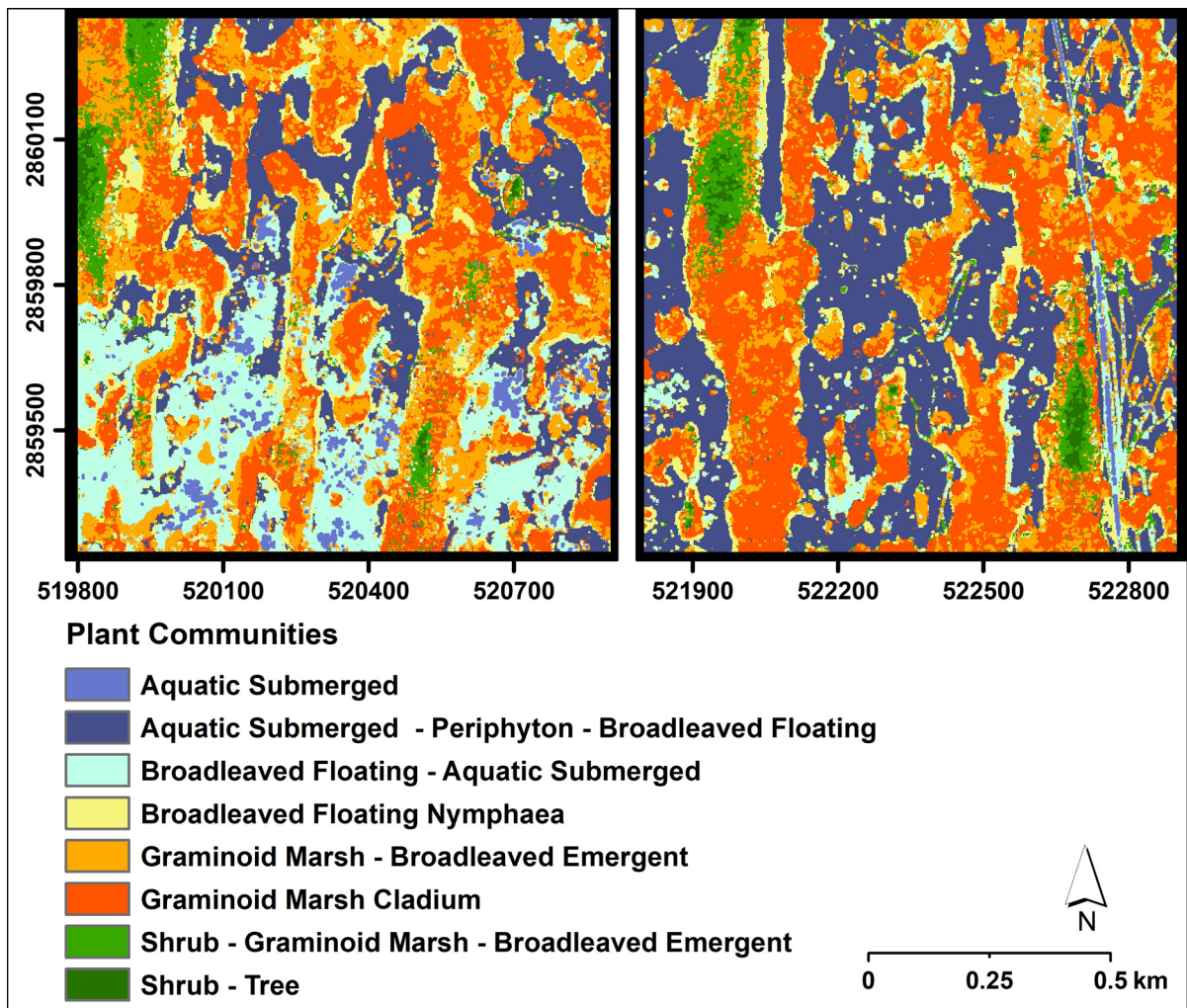
Fig. 5 High-resolution vegetation map for WCA3A study area (Fig. 3) derived from WV-2 2-m satellite data. Coordinates in Meters WGS84 UTM 17 N

Table 1 Classification scheme and mapped class proportions for high-resolution plant community map for WCA3A

Class Code	Class Name	Class Proportion (%)
aS	Aquatic Submerged	1.8
aSpblF	Aquatic Submerged—Periphyton—Broadleaved Floating	28.2
blF	Broadleaved Floating—Aquatic Submerged	10.8
blFNy	Broadleaved Floating *Nymphaea*	10.3
gMblE	Graminoid Marsh—Broadleaved Emergent	18.9
gMCl	Graminoid Marsh *Cladium*	24.7
s_gMblE	Shrub—Graminoid Marsh—Broadleaved Emergent	4.5
s	Shrub—Tree	0.8

Fig. 6 High-resolution vegetation map for NESRS study area (Fig. 3) derived from WV-2 2-m satellite data; class labels as in Table 2. Coordinates in Meters WGS84 UTM 17 N
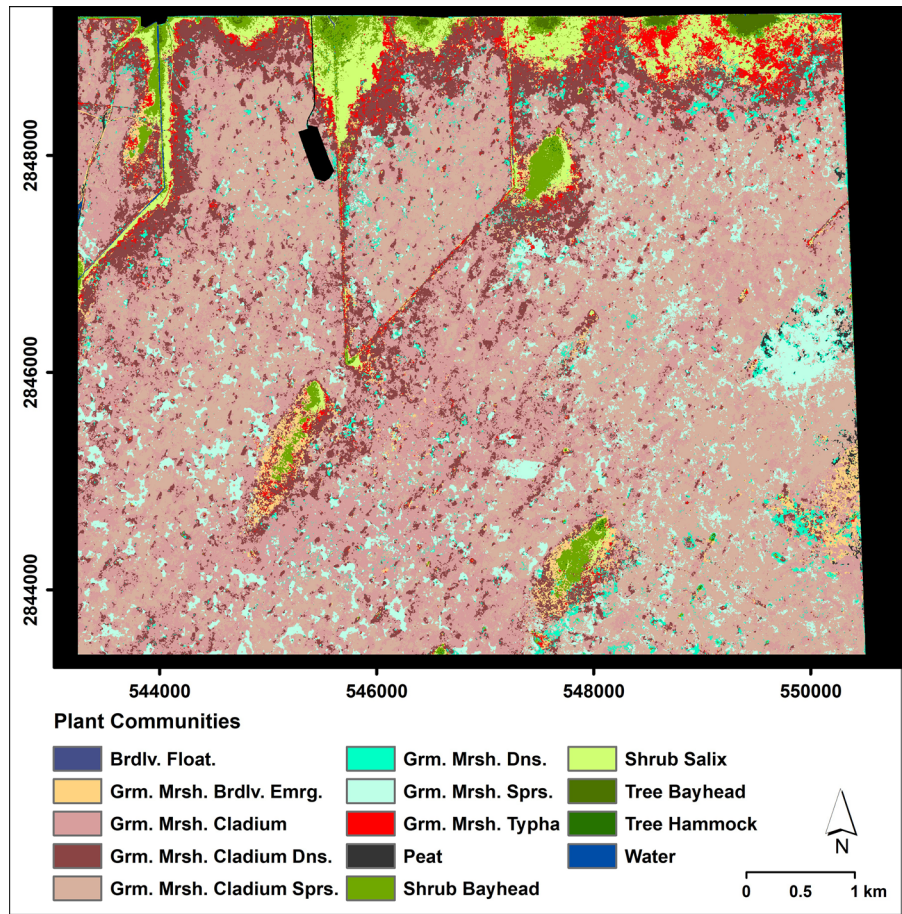

Table 2 Classification scheme and mapped class proportions for high-resolution plant community map for NESRS

Class Code	Class Name	Class Proportion (%)
blF	Broadleaved Floating (Brdlv. Float.)	<0.1
gMblE	Graminoid Marsh—Broadleaved Emergent (Grm. Mrsh. Brdlv. Emgr.)	2.9
gMS	Graminoid Marsh Sparse (Grm. Mrsh. Sprs.)	8.0
gMD	Graminoid Marsh Dense (Grm. Mrsh. Dns.)	2.3
gMCl	Graminoid Marsh *Cladium* (Grm. Mrsh. *Cladium*)	29.3
gMClD	Graminoid Marsh *Cladium* Dense (Grm. Mrsh. *Cladium* Dns.)	13.7
gMClS	Graminoid Marsh *Cladium* Sparse (Grm. Mrsh. *Cladium* Sprs.)	35.5
gMTy	Graminoid Marsh *Typha* (Grm. Mrsh. *Typha*)	2.7
sB	Shrub Bayhead	2.0
sSa	Shrub *Salix*	3.0
tB	Tree Bayhead	0.3
tH	Tree Hammock	0.1
wtr	Water	0.1
pt	Peat	0.3

representativeness threshold were retained in the scaled classification scheme, since they have high information retention and are expected to generate pure spectral signatures with high detection probability and accuracy. The algorithm thus generated class labels that reflected the constrained relative abundances of classes from the fine-scale map, and each scaled class had a defined minimum relative abundance across the landscape. This process resulted in 25 scaled maps (Fig. 4) and their associated scale-specific classification schemes for each study area (Fig. 4).

To account for sampling error related to the arbitrary grid origin of Landsat, effects of class-precision and landscape representativeness thresholds on class-label fidelity, the recurrence of class labels across scaling solutions, were evaluated for 10 random origins of each of the 25 landscapes (Fig. 4). This random origin sampling also provides a mean and confidence interval for the landscape-scale information retention.

Significance of differences in label precision by representativeness thresholds was tested with a pairwise-paired Wilcoxon signed-rank test, where data were paired by random origin iteration. Optimal scaling parameter solutions for each landscape were identified with an optimal scaling index (OSI) that weighted per-class information retention IRc_{min} above a user-defined, minimum-expected threshold multiplied by the two class label fidelity metrics CLF_m and CLF_p (Fig. 4, Eq. 2). Information retention above the expected minimum was normalized to per-class IR_c gain (Eq. 2) above the minimum to only credit models that reached the minimum expected information retention. The optimal-solution model was determined by the maximum OSI across all compared models.

$$OSI = CLF_m * CLF_p * \left(\frac{IRc - IRc_{\min}}{class\,count} \right) \qquad (2)$$

Scaling parameter evaluation for spectral detection accuracy

Accuracy of detecting scaled classes from Landsat Thematic Mapper (TM) multispectral reflectance data (Fig. 4) was evaluated for cloud-free Landsat 5 TM images acquired close to the acquisition dates of the WV-2 images used for the high-resolution maps. A November 11, 2011, scene was used for the WCA3A map and a December 25, 2010, scene for the NESRS map. Landsat data were atmospherically corrected using the FLAASH module in ENVI. For each study area, high-resolution maps were scaled for the realized 30-m grid specific to the Landsat scene path 015 row 042 (World Reference System 2). The MDGP-scaling algorithm generated relative class abundances from the high-resolution map for each Landsat grid cell, using the same class precision and representativeness threshold combinations as for the random-origin evaluation, and assigned up-scaled class labels. Scaled class labels for each of the 25 models were joined with Landsat spectral reflectance data of the corresponding pixel in the processed Landsat reflectance data (Fig. 4). Overall and class-specific spectral detection accuracy for each scaled map were estimated for each study area using a model-based, tenfold cross-validation when applying the random forest classification algorithm (Breiman 1984) as implemented in the "caret" package (Kuhn et al. 2016). The number of trees was set to 200. To determine the optimal number of randomly selected features at each node, the "mtry" parameter was evaluated for a range of two to six features, the number of features (spectral bands) in the TM dataset. Overall accuracy was used to evaluate spectral detectability and separability between classes for each of the 25 models.

As class-label precision and class count increase, misclassifications are more likely, purely by chance. Hence, a tradeoff exists between class-label precision and accuracy. To account for less severe misclassifications and to more accurately present the actual class proportions on the ground (Latifovic and Olthof 2004; Pontius and Connors 2009), partial credit for accurate proportions of class labels was given by weighting label errors with a weighted Kappa statistic (Cohen 1968). The portions of the partially matching class labels were used to calculate the weights of the weight matrix. The weight matrix was then applied to the confusion matrix, generating partial-credit class accuracies and weighted overall accuracy (A_{ow}) and their 95% confidence intervals (Rossiter 2014). Class-label fidelity for the realized Landsat grid was calculated as the proportion of classes identified for the Landsat grid over the classes generated from the 10 random origin solutions (CLF_{pRO}). As CLF_{pRO} increases, it is more likely that the scaled

classes for Landsat are representative classes for random locations across the landscape at that scale. For all 25 scaled landscapes per study area, A_{ow}, IR_c, and CLF_{pRO} were used to evaluate and select optimal class-label precision and representativeness threshold (Fig. 4).

The trade-offs that must be negotiated in the case of spectral detection from a realized grid are class-detection accuracy, information retention, and representativeness of the scaled classes of a realized grid. An index was developed to select the optimal scaling solution (OSI_a), defined as

$$OSI_a = CLF_{pRO} * \left(\frac{IRc - IRc_{min}}{class\,count} \right) * A_{ow}. \quad (3)$$

The optimal model solution was defined as the maximum of OSI_a across all evaluated models. Final maps were generated for solutions that maximized OSI_a.

Results

When applying the MDGP-scaling algorithm across the two natural landscapes, interaction effects of class-label precision and landscape representativeness on information retention and class-label fidelity were similar but not uniform. Therefore, results for scaling and classification accuracy are presented by study area.

WCA3A: information retention and class-label fidelity

Scaling produced 250 scaled landscapes and associated scale-specific class schemes. Evaluating the effects of class-label precision and representativeness thresholds for scaled maps for WCA3A showed that scaled class count and mean IR_c ($p < 0.05$) increased with increasing class-label precision for a minimum class representativeness of 1% (Fig. 7, Table 3). Increase in IR_c, however, diminished with increasing class-label precision. As class representativeness threshold increased to 5 and 10%, the increase of IR_c with increasing class-label precision diminished until no significant increases for class-label precisions greater than 4-parts (25%) were observed. As representativeness thresholds increased to 15 and 20%,

significant IR_c increase was observed only for label precisions below 3-parts (33%) (Fig. 7, Table 3).

Class-label fidelity generally decreased with increased class-label precision and representativeness. However, higher CLF occurred when precision exceeded 50% and minimum representativeness increased above 10% (Table 3). Setting the minimum expected IR_c threshold to 60% (Table 3), a class-label precision of 25% with a representativeness threshold of 10% (OSI = 0.73) or a 33% class-label precision with a landscape representativeness of 15% (OSI = 0.66) scored high on the OSI. The 33%-precision solution on average yielded 7.0 scaled classes, with an average IR_{cRO} of 73.5% across the landscape and a mean probability of class-label recurrence of 0.78 (Fig. 7 Table 3, Mean Prob.), with 44% of classes recurring with a probability of 1 (Table 3, Prop.1). The 25% class-label precision solution produced on average 7.9 scaled classes, which on average retained 77.7% of information and had a mean probability of class-label recurrence of 0.72 (Table 3, Mean Prob.), with 45% of classes re-occurring with a probability of 1 (Table 3, Prop. 1). In both cases, the higher CLF increased the optimal scale index.

WCA3A: spectral-detection accuracy

Scaling the WCA3A landscape to the specific Landsat grid, IR_{cRL} averaged 1.1% greater than the mean IR_c across the 10 random-origin grids (IR_{cRO}). Overall accuracy ranged from 66.6% for majority rule (1-part) with a 1% class representativeness to 78.2% for a 20% class-label precision and minimum landscape representativeness of 15% for each of the five classes the classification scheme produced (Fig. 8, Table 4).

All scaling solutions of the Landsat grid with a majority rule had a mean IR_{cRL} of less than 63%, which was significantly lower than the MDGP-scaled solutions for the 2- to 5-part label precisions and produced significantly lower overall accuracies than the corresponding multi-part solutions ($p < 0.05$). The three 2-part class precision models above 70.4% overall accuracy were those with representativeness thresholds of 10% and greater (Fig. 8, Table 4). Comparing the three solutions showed that the classification schemes were identical and that the differences in accuracy were minor (Table 4).

Adding the spectral detection accuracy to the optimal scaling index provided the same scaling solutions

Fig. 7 Mean probability of class label recurrence vs. information retention for WCA3A. Point shape specifies scaled landscape representativeness threshold (Rep.), while point color specifies class-label precision (Parts). Point labels represent the mean number of classes generated for each model across the 10 random origins. Horizontal bars are 95% confidence intervals

as those identified by the OSI. The 3-part class-label precision with a 15% representativeness threshold produced a classification scheme with 8 classes, an IR_{cRL} of 74.2%, and a class ratio of realized Landsat grid to random origin class solutions of 0.89 (Table 4). The overall classification accuracy was 73.9%, with an OSI_a of 1.17 (Fig. 8, Table 4). The second highest OSI_a was 1.12 for the 4-part class-label precision and 10% representativeness threshold model (Table 4). This solution also produced eight scaled classes, retained a slightly higher IR_{cRL} of 76.4%, and had a higher classification accuracy of 75.3% with a class-label fidelity of 0.73 (Fig. 8, Table 4).

The maps for the two optimal solutions indicate that only the 3-part class-label precision solution maintained the shrub/tree label in the scaled classes (Table 5). This solution was selected as the best-scaled map for a minimum requirement of 60% information retention when compared to the original high-resolution input map. This solution had eight classes, of which five were monotypic input classes and the other three were mixed classes that occupied 60.1% of the landscape (Fig. 9, Table 5). Two of the high-resolution community classes, "Aquatic Submerged"

and "Shrub-Tree", which accounted for 2.58% cover of the high-resolution map (Table 1), were not maintained in the scaled community class labels. Except for "trees", all original class names were included in pure or mixed class names (Fig. 9, Table 5). The scaled map and its associated location-specific, information-retention map are presented in Fig. 9, and the spectrally classified map and location-specific classifier confidence in Fig. 10.

NESRS: information retention and class-label fidelity

Results for NESRS were similar to those for WCA3A. The original high-resolution map of NESRS had 14 plant community classes, six more than the WCA3A map. Applying MDGP-scaling for all 25 class-label precision and representativeness-threshold combinations confirmed the expected increase in class number and mean IR_{cRO} when class-label precision increased (Fig. 11, Table 6). With increasing class-label precision, the increase in IR_{cRO} diminished, and the differences for consecutive pairwise comparisons became insignificant ($p \geq 0.05$) when representativeness was greater than 1% (Fig. 11, Table 6). For representativeness of 5%, the 4- and 5-part label precisions had

Table 3 WCA3A scaling results for varying class-label precisions (Parts) and landscape representativeness thresholds (Repr.) for 10 random origins (RO)

| Parts | Repr. | Class Fidelity | | Total Classes | Mixed Classes | Mono. Classes | Mean Classes$_{RO}$ | SD Classes$_{RO}$ | Information Retention | | OSI |
		Mean Prob.	Prop. 1						Mean IR$_{cRO}$ (%)	SD IR$_{cRO}$ (%)	IR$_c$ > 60%
1	1	1.00	1.00	7	0	7	7.0	0.00	57.3	0.45	− 0.39
1	5	0.89	0.71	7	0	7	6.2	0.63	58.0	0.57	− 0.21
1	10	0.80	0.71	7	0	7	5.6	0.70	58.6	0.63	− 0.15
1	15	0.80	0.71	7	0	7	5.6	0.70	58.6	0.63	− 0.15
1	20	0.80	0.71	7	0	7	5.6	0.70	58.6	0.63	− 0.15
2	1	0.89	0.79	19	12	7	17.0	1.15	74.8	0.62	0.61
2	5	0.80	0.53	15	8	7	12.0	0.67	72.0	0.53	0.43
2	10	0.76	0.40	10	3	7	7.6	0.97	70.0	0.92	0.40
2	15	0.72	0.30	10	3	7	7.2	1.14	69.5	0.87	0.29
2	20	0.67	0.30	10	3	7	6.7	0.95	67.8	2.07	0.23
3	1	0.89	0.82	34	29	5	30.1	1.20	80.5	0.29	0.50
3	5	0.67	0.37	19	14	5	12.8	0.63	76.8	0.62	0.33
3	10	0.61	0.29	14	9	5	8.5	0.71	74.9	0.95	0.30
3	*15*	*0.78*	*0.44*	*9*	*4*	*5*	*7.0*	*0.67*	*73.5*	*1.43*	*0.66*
3	20	0.66	0.22	9	4	5	5.9	0.57	70.5	1.74	0.26
4	1	0.78	0.58	52	48	4	40.3	1.16	83.4	0.20	0.26
4	5	0.67	0.35	20	16	4	13.4	1.26	79.7	0.55	0.34
4	*10*	*0.72*	*0.45*	*11*	*7*	*4*	*7.9*	*0.74*	*77.7*	*0.73*	*0.73*
4	15	0.57	0.18	11	7	4	6.3	0.67	75.9	1.57	0.26
4	20	0.54	0.22	9	5	4	4.9	0.32	71.6	1.44	0.29
5	1	0.61	0.36	78	75	3	47.8	1.62	85.2	0.19	0.12
5	5	0.48	0.14	29	26	3	13.8	1.14	80.6	0.76	0.10
5	10	0.57	0.07	14	11	3	8.0	0.82	77.6	1.33	0.09
5	15	0.46	0.08	12	9	3	5.5	0.53	74.4	1.74	0.10
5	20	0.58	0.13	8	5	3	4.6	0.52	72.8	1.65	0.20

The best two solutions are highlighted in italics

Prob probability of class recurrence, *Prop* 1 proportion of classes with recurrence probability of 1, *Mono* monotypic, *SD* standard deviation; *IR$_c$* compositional information retention, *IR$_{cRo}$* IR$_c$ across RO random origin, *OSI* Optimal Scaling Index (i.e., credit for information retention greater than 60%)

Fig. 8 Cross-validated weighted overall accuracy vs. information retention for WCA3A scaled maps. Point shape specifies scaled landscape representativeness threshold (Rep.), while point color specifies class-label precision (Parts). Point labels represent the number of classes generated for the realized Landsat grid for each model. Vertical bars are 95% confidence intervals
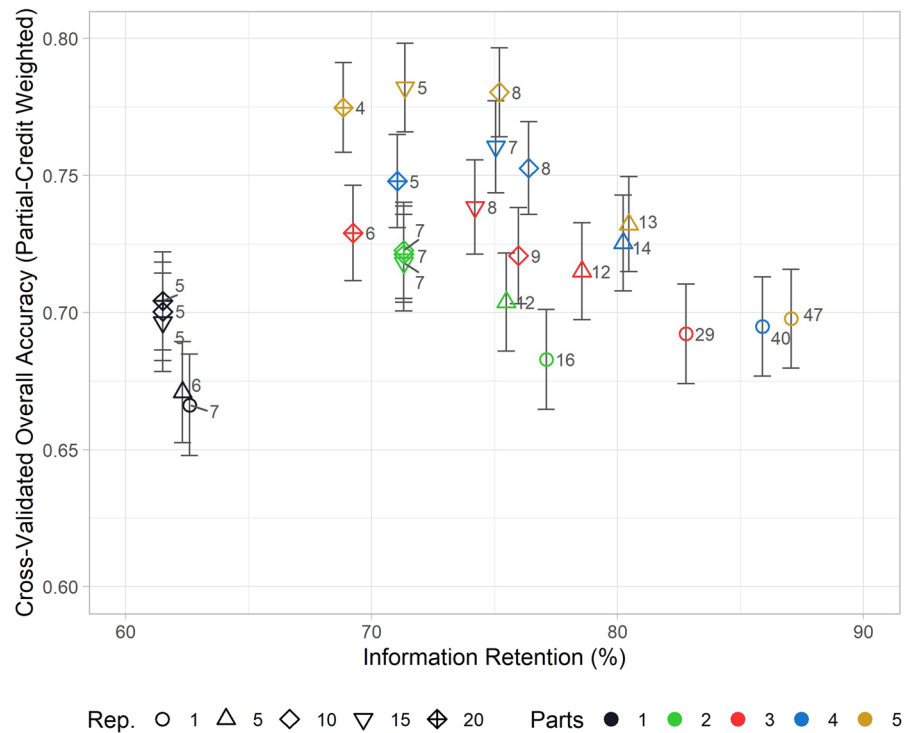
insignificant differences in IR_{cRO}. As the representativeness threshold increased above 5%, insignificant differences also occurred for 2- and 3-part label precisions ($p \geq 0.05$) (Fig. 11, Table 6).

Class-label fidelity in NESRS was significantly higher than for WCA3A (Tables 3, 6). With a minimum expected IR_c threshold of 60%, two solutions, the 3-part class-label precision with a representativeness threshold of 15% and a 2-part class-label precision with representativeness of 5% both had an OSI of 0.79, which was higher than the other 23 models (Table 6). The 3-part label-precision solution on average yielded 11.3 scaled classes, with an average IR_{cRO} of 72.7% and a mean probability of class-label recurrence of 0.94 (Fig. 11, Table 6, Mean Prob.), with 75% of classes recurring across all random-origin iterations (Table 6, Prop. 1). The 2-part class-label precision solution produced 13.3 scaled classes, with an average IR_{cRO} of 74.2% and a mean probability of class-label recurrence of 0.95 (Fig. 11, Table 6, Mean Prob.), with 79% of classes recurring across all random origin landscapes (Table 6, Prop. 1).

NESRS: spectral-detection accuracy

Information retention for the Landsat-grid scaled maps (IR_{cRL}) averaged 4.6% higher than the mean IR_c of the random-origins scaled maps (IR_{cRO}). Overall accuracy ranged from ~69% for majority-rule solutions to the highest accuracy of 73.2% (Fig. 12, Table 7). As in WCA3A, the highest cross-validated overall accuracy was achieved for a 3-part class-label precision and minimum landscape representativeness of 10%. This solution had 13 scaled classes (Table 8).

All scaling solutions with a 1-part label precision had a mean IR_{cRL} of 70.3%, which was significantly lower than the MDGP-scaled solutions for the 2- to 5-part label precisions (Fig. 12, Table 7). Accuracy was significantly higher for all multi-part solutions with class representativeness greater than 5% ($p < 0.05$) (Table 7). For class-label precisions of 50% and less, the 15% and 20% representativeness thresholds produced identical classification solutions.

Adding spectral-detection accuracy to the optimal scaling index indicated that the 33% class-label

Table 4 Scaling effects on spectral detection accuracies for WCA3A for varying class-label precisions (Parts) and landscape representativeness thresholds (Repr.)

Parts	Repr.	IR_{cRL} (%)	Total Classes	CLF_{pRL}	Overall Accuracy (%)	OA-CIL (%)	OA-CIU (%)	OSI_a
1	*1*	62.6	7	1.00	66.6	64.8	68.5	0.25
1	*5*	62.3	6	0.86	67.1	65.3	68.9	0.22
1	*10*	61.5	5	0.71	70.0	68.2	71.8	0.15
1	*15*	61.5	5	0.71	69.6	67.8	71.4	0.15
1	*20*	61.5	5	0.71	70.4	68.6	72.2	0.15
2	*1*	77.1	16	0.84	68.3	66.5	70.1	0.61
2	*5*	75.5	12	0.80	70.4	68.6	72.2	0.73
2	*10*	71.3	7	0.70	72.3	70.5	74.0	0.82
2	*15*	71.3	7	0.70	71.8	70.1	73.6	0.81
2	*20*	71.3	7	0.70	72.1	70.4	73.9	0.82
3	*1*	82.8	29	0.85	69.2	67.4	71.0	0.46
3	*5*	78.6	12	0.63	71.5	69.7	73.3	0.70
3	*10*	76.0	9	0.64	72.1	70.3	73.8	0.82
3	*15*	*74.2*	*8*	*0.89*	*73.9*	*72.1*	*75.6*	*1.17*
3	*20*	69.2	6	0.67	72.9	71.2	74.6	0.75
4	*1*	85.9	40	0.77	69.5	67.7	71.3	0.35
4	*5*	80.2	14	0.70	72.5	70.8	74.3	0.73
4	*10*	*76.4*	*8*	*0.73*	*75.3*	*73.6*	*77.0*	*1.12*
4	*15*	75.1	7	0.64	76.1	74.4	77.7	1.04
4	*20*	71.0	5	0.56	74.8	73.1	76.5	0.92
5	*1*	87.1	47	0.60	69.8	68.0	71.6	0.24
5	*5*	80.5	13	0.45	73.2	71.5	75.0	0.52
5	*10*	75.2	8	0.57	78.0	76.4	79.7	0.85
5	*15*	71.4	5	0.42	78.2	76.6	79.8	0.74
5	*20*	68.9	4	0.50	77.5	75.8	79.1	0.86

The two best solutions are highlighted in italics

IR_{cRL} mean information retention across the landscape for the realized landscape (Landsat grid), CLF_{pRL} the ratio of scaled classes for the realized landscape to the number of all classes identified across 10 random grid origins, *OA-CIL* and *OA-CIU* lower and upper confidence estimates of the tenfold cross-validated overall accuracy, OSI_a Optimal Scaling Index for class-detection accuracy

precision with a 10% class representativeness threshold produced the best scaling result ($OSI_a = 1.10$), generating 13 scaled classes (Table 8) that were detected from multispectral Landsat data with an accuracy of 70.7% (Table 7). The information retained for this solution was 80.3%, and the class-label count ratio was 1, indicating that all classes derived for the Landsat grid were represented in the random origin solutions (Fig. 13, Table 7).

Scaled community classes for the optimal solution of 33% class-label precision included three mixed classes (47.7% of the landscape) and 10 monotypic input classes (Table 8). The original community classes that were omitted in the scaled class labels were "Broadleaved Floating", "Tree Hammock", "Water", and "Peat" (Tables 2, 8). These four classes, however, accounted for only 0.4% cover in the original map (Table 2). The small class of "Tree Bayhead" was maintained as a monotypic class with the same cover percentage (0.28%) as the original map and mean information retention of 82.5%. The scaled map and its associated information retention by grid cell are presented in Fig. 13 and the spectrally classified map with location-specific classifier confidence in Fig. 14.

Table 5 WCA3A class schema and class proportions (Class Prop.) for two good scaling solutions of a 33% class-label precision with landscape representativeness of 15% (top) and a 25% class-label precision with representativeness of 15% (bottom)

Class name—33% Precision 15% Representativeness	Class Prop. (%)	Mean IR_c (%)	SD IR_c (%)
100 Aquatic Submerged—Periphyton - Broadleaved Floating*	15.7	88.9	9.2
67 Aquatic Submerged—Periphyton - Broadleaved Floating	19.9	68.0	14.7
33 Broadleaved Floating *Nymphaea*			
100 Broadleaved Floating—Aquatic Submerged*	9.7	65.6	19.2
33 Broadleaved Floating *Nymphaea*	21.8	65.1	11.8
33 Graminoid Marsh—Broadleaved Emergent			
33 Graminoid Marsh *Cladium*			
100 Graminoid Marsh—Broadleaved Emergent*	5.2	69.0	11.7
100 Graminoid Marsh *Cladium**	6.6	89.5	7.2
67 Graminoid Marsh *Cladium*	18.4	81.2	11.0
33 Graminoid Marsh—Broadleaved Emergent			
100 Shrub—Graminoid Marsh—Broadleaved Emergent*	2.6	63.2	17.6
Class name—25% Precision 10% Representativeness	**Class Prop. (%)**	**Mean IR_c (%)**	**SD IR_c (%)**
100 Aquatic Submerged—Periphyton—Broadleaved Floating*	13.0	89.3	16.6
25 Aquatic Submerged—Periphyton—Broadleaved Floating	12.8	62.6	17.0
25 Broadleaved Floating—Aquatic Submerged			
25 Broadleaved Floating *Nymphaea*			
25 Graminoid Marsh—Broadleaved Emergent			
25 Aquatic Submerged—Periphyton—Broadleaved Floating	15.9	68.6	15.0
25 Broadleaved Floating—Aquatic Submerged			
25 Graminoid Marsh—Broadleaved Emergent			
25 Graminoid Marsh *Cladium*			
75 Aquatic Submerged—Periphyton—Broadleaved Floating	14.0	78.1	12.4
25 Broadleaved Floating *Nymphaea*			
100 Broadleaved Floating—Aquatic Submerged*	7.0	74.0	14.5
50 Graminoid Marsh—Broadleaved Emergent	16.0	72.8	9.9
25 Broadleaved Floating *Nymphaea*			
25 Graminoid Marsh *Cladium*			
100 Graminoid Marsh *Cladium**	5.0	92.4	5.1
75 Graminoid Marsh *Cladium*	16.2	82.9	10.4
25 Graminoid Marsh—Broadleaved Emergent			

Left column gives the relative abundance of original classes in the scaled class

Prop proportion, *IR_c* compositional information retention, *SD* standard deviation

*Indicate original classes. Mixed classes make up 60.1% of the landscape for 33% precision and 74.9% for 25% precision solutions

Discussion

We have shown that in natural landscapes the MDGP-scaling algorithm generates thematic classes at a coarser resolution that retain high levels of compositional information for the mapped area that was represented in the higher resolution map. The scaled classes are quantitatively derived from the finer resolution data rather than being predetermined and reflect common vegetation associations present at the coarser resolution. Below, we discuss the scaling parameter effects on information retention and class-label fidelity and how the scaling results influence the scaled-class detection from coarser resolution Landsat spectral reflectance data. We conclude the discussion with applications that demonstrate the benefits of MDGP-scaling over decision rules that do not modify
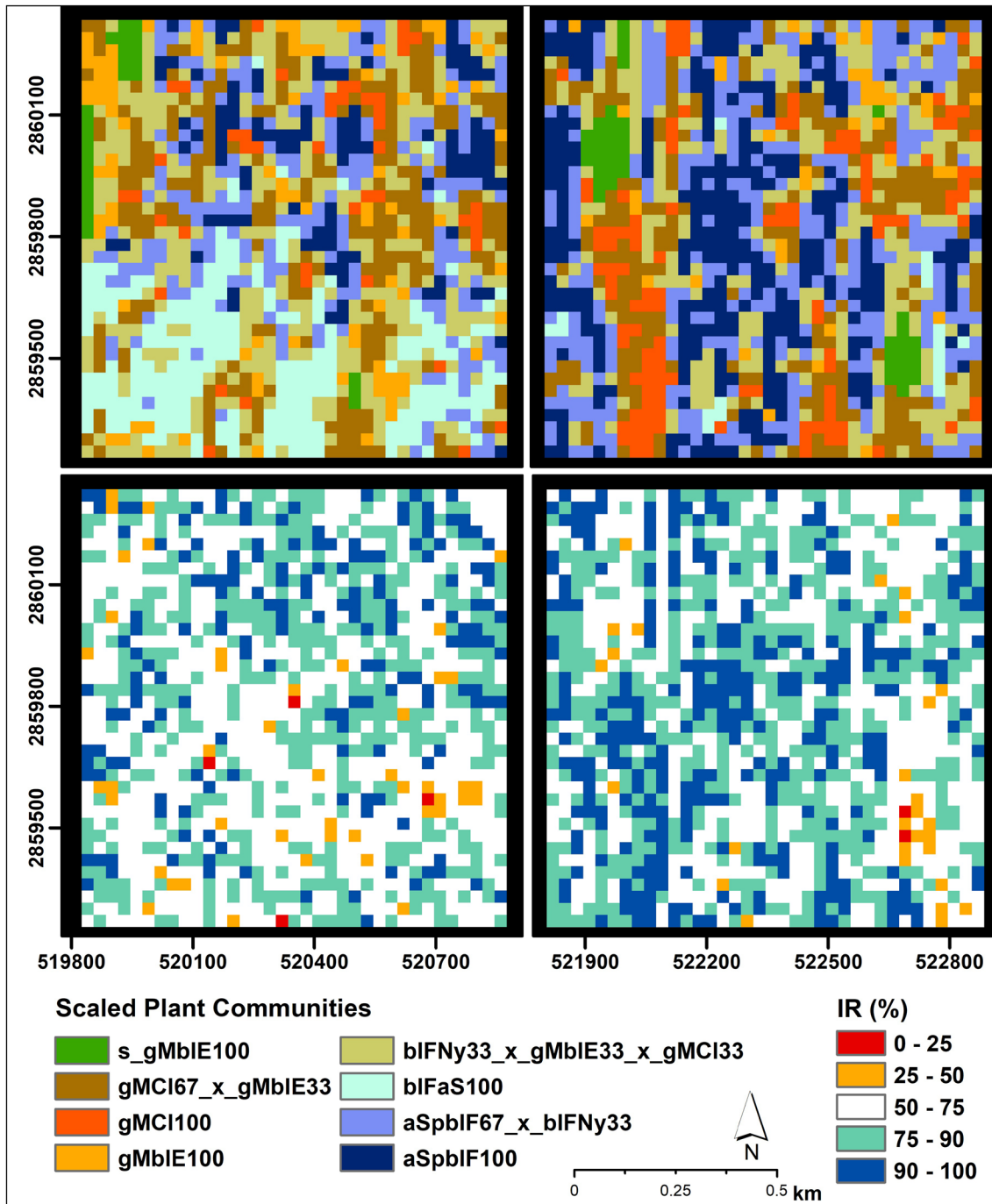
Fig. 9 Scaled plant communities for WCA3A (top) and location-specific information retention (IR) for the assigned community class label when compared to the high-resolution map (Fig. 5) (bottom). Class codes in Table 1. Numbers in labels give the class-label precision percent. Coordinates in Meters WGS84 UTM 17 N
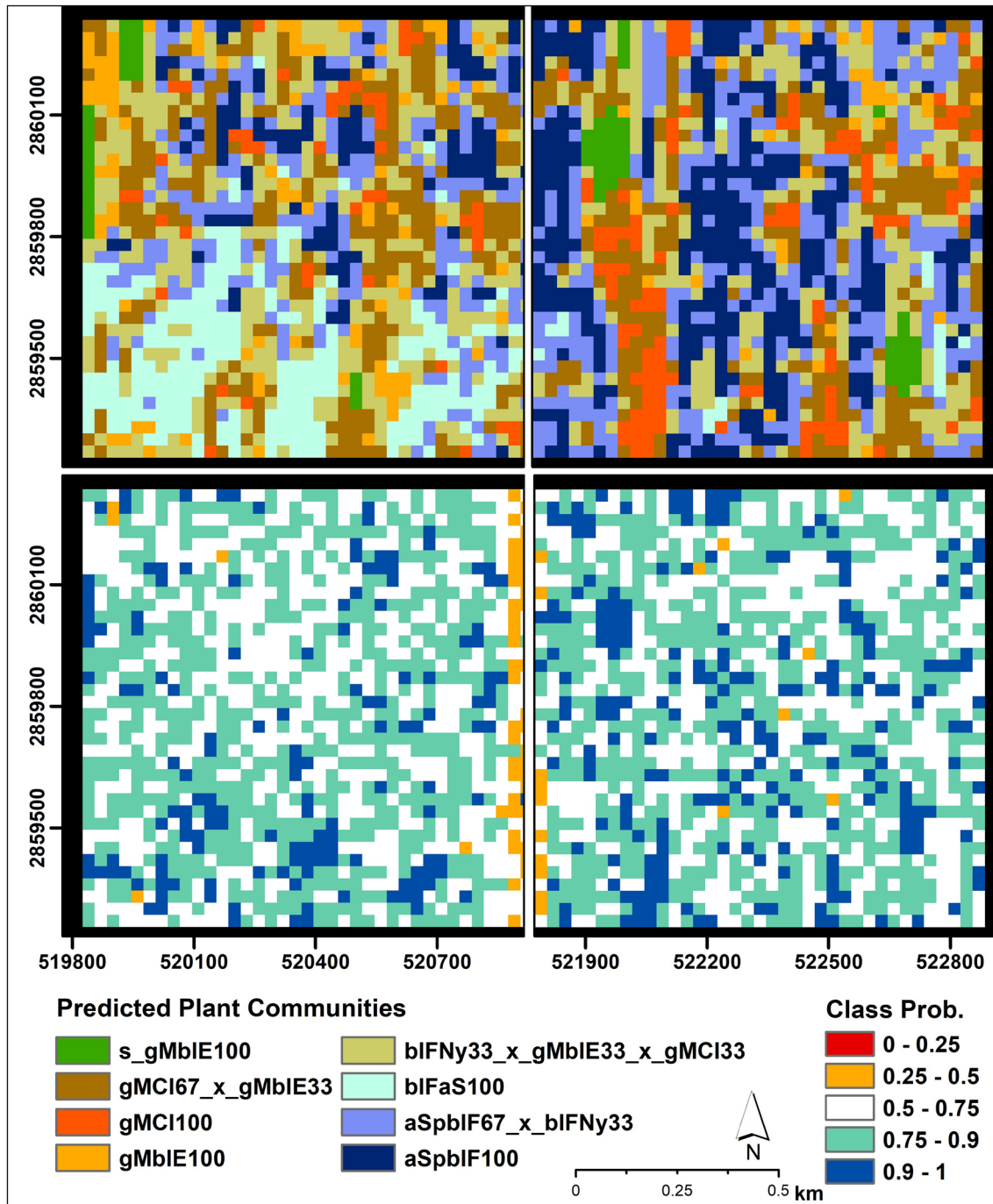
Fig. 10 Scaled plant-community classes predicted from Landsat spectral data (top) and location-specific classifier probability for class-label assignment (bottom) for WCA3A. Classes and abbreviations as in Fig. 9. Coordinates in Meters WGS84 UTM 17 N

Fig. 11 Mean probability of class label recurrence vs. information retention for NESRS. Point shape specifies scaled landscape representativeness threshold (Rep.), while point color specifies class-label precision (Parts). Point labels represent the mean number of classes generated for each model across the 10 random origins. Horizontal bars are 95% confidence intervals
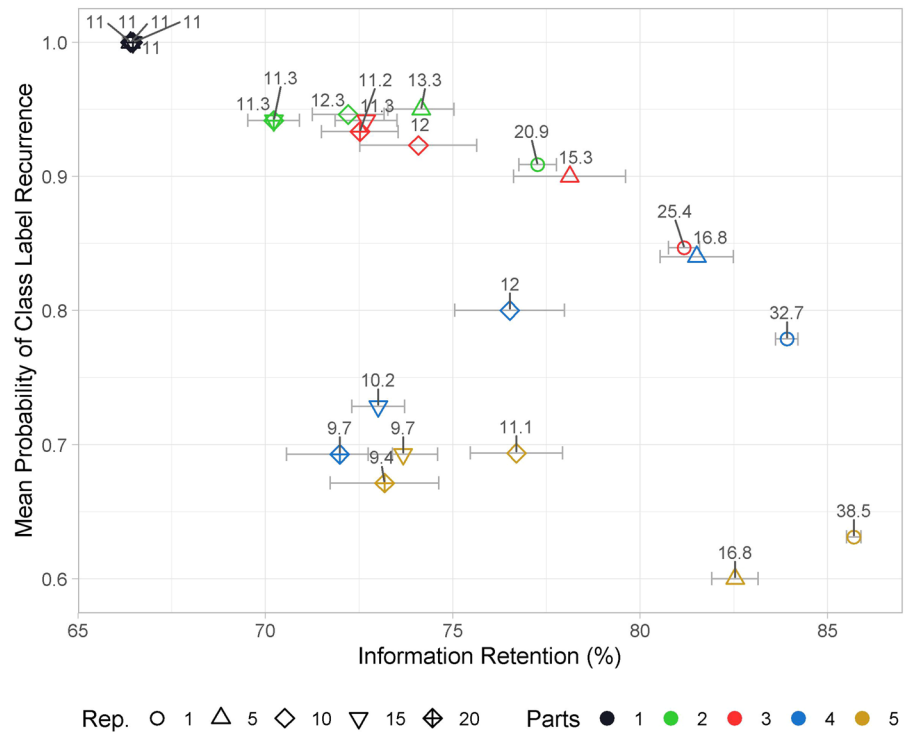
the classification scheme in the data aggregation process.

Categorical data integration across spatial scales

Analyses that integrate categorical data from different spatial resolutions require scaling the high-resolution data to a coarser resolution. Combining coarse-resolution categorical maps with maps generated from high spatial but low temporal resolution is more effective when information retention of the scaled product is optimized. The MDGP-scaling algorithm allows the user to optimize parameter selection, negotiating the trade-offs between information retention and class-label fidelity in natural landscapes. Information retention of MDGP-scaling was consistently and significantly higher for both natural landscapes when compared to the majority-rule solution, as was expected from results for artificial landscapes presented in Gann (2019). Class-label fidelity for both landscapes was high, which demonstrates that the algorithm can generate classification systems in natural landscapes that consistently assign new scaled labels as recognizable classes at the scaled spatial resolution.

Compared to majority-rule aggregation, the increases in information retention and class-label fidelity were always higher for WCA3A than for NESRS. On average, the pairwise mean IR_c difference was $3.4 \pm 1.7\%$ and mean CLF_{pRO} was 0.15 ± 0.11 greater for WCA3A than for NESRS. The reasons for the differences were either related to a lower number of classes in the original classification system of WCA3A or the higher spatial heterogeneity of the WCA3A landscape, leading to larger gains in information retention when using mixed class labels.

Scaled class detection from remotely sensed multispectral data

Classification systems derived from quantitative analysis of species or class co-occurrence patterns is integral to several scientific disciplines (e.g., phytosociology, community ecology). However, as sample area size increases (e.g., 1 m^2 to 900 m^2), sampling ground units becomes increasingly difficult. If high-resolution categorical maps with adequate class detail exist, application of the MDGP-scaling algorithm can produce high precision classification systems that provide representative mixed classes for medium to low

Table 6 NESRS scaling results for varying class-label precisions (Parts) and landscape representativeness thresholds (Repr.)

| Parts | Repr. | Class Fidelity | | Total Classes | Mixed Classes | Mono. Classes | Mean Classes$_{RO}$ | SD Classes$_{RO}$ | Information Retention | | OSI |
		Mean Prob.	Prop. 1						Mean IR$_{CRO}$ (%)	SD IR$_{CRO}$ (%)	IR$_c$ > 60%
1	1	1.00	1.00	11	0	11	11.0	0.00	66.4	0.21	0.58
1	5	1.00	1.00	11	0	11	11.0	0.00	66.4	0.21	0.58
1	10	1.00	1.00	11	0	11	11.0	0.00	66.4	0.21	0.58
1	15	1.00	1.00	11	0	11	11.0	0.00	66.5	0.20	0.59
1	20	1.00	1.00	11	0	11	11.0	0.00	66.5	0.20	0.59
2	1	0.91	0.78	23	12	11	20.9	0.57	77.3	0.50	0.59
2	*5*	*0.95*	*0.79*	*14*	*3*	*11*	*13.3*	*0.67*	*74.2*	*0.88*	*0.79*
2	10	0.95	0.77	13	2	11	12.3	0.67	72.2	0.96	0.72
2	15	0.94	0.83	12	1	11	11.3	0.67	70.2	0.69	0.71
2	20	0.94	0.83	12	1	11	11.3	0.67	70.2	0.69	0.71
3	1	0.85	0.67	30	20	10	25.4	1.17	81.2	0.41	0.47
3	5	0.90	0.65	17	7	10	15.3	0.82	78.1	1.50	0.69
3	10	0.92	0.62	13	3	10	12.0	0.94	74.1	1.56	0.67
3	*15*	*0.94*	*0.75*	*12*	*2*	*10*	*11.3*	*0.82*	*72.7*	*0.82*	*0.79*
3	20	0.93	0.67	12	2	10	11.2	0.79	72.5	1.02	0.70
4	1	0.78	0.62	42	32	10	32.7	0.95	83.9	0.30	0.35
4	5	0.84	0.65	20	10	10	16.8	1.03	81.5	0.98	0.70
4	10	0.80	0.53	15	5	10	12.0	0.82	76.5	1.47	0.59
4	15	0.73	0.50	14	4	10	10.2	1.32	73.0	0.71	0.46
4	20	0.69	0.43	14	4	10	9.7	1.49	72.0	1.41	0.37
5	1	0.63	0.48	61	52	9	38.5	1.08	85.7	0.19	0.20
5	5	0.60	0.36	28	19	9	16.8	1.03	82.5	0.62	0.29
5	10	0.69	0.38	16	7	9	11.1	1.20	76.7	1.23	0.39
5	15	0.69	0.36	14	5	9	9.7	0.67	73.7	0.93	0.35
5	20	0.67	0.36	14	5	9	9.4	0.70	73.2	1.45	0.34

Column heading abbreviations as in Table 3

The two best solutions are highlighted in italics

Fig. 12 Cross-validated weighted overall accuracy vs. information retention for NESRS. Point shape specifies scaled landscape representativeness threshold (Rep.), while point color specifies class-label precision (Parts). Point labels represent the number of classes generated for the realized Landsat grid for each model. Vertical bars are 95% confidence intervals
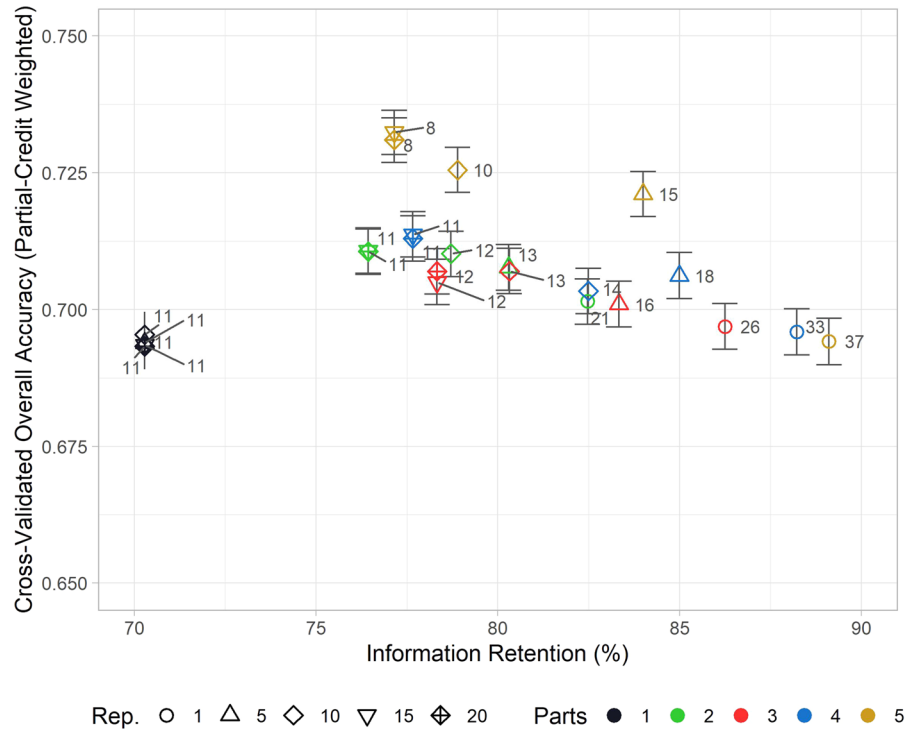
resolution pixel sizes. Our analysis also demonstrates that the detection of these more precise and representative land-cover classes from medium resolution spectral data was more accurate than for classification schemes that did not include the mixed classes.

Class detection accuracy from spectral data increased with higher class-label precision and with higher class representativeness thresholds. There are trade-offs, however, among the parameters. As the class-representativeness threshold increased, class count decreased, producing a reduced chance probability for class confusion and higher classification accuracy. However, as class count decreased, information retention at the grid-cell level was reduced, and grid cells that were further from the nominal class label increased the thematic heterogeneity of the mixed classes, and with it, spectral variability, reducing classification accuracy. In a similar fashion, class-label precision increased information retention and, therefore, more clearly associated defined thematic classes to spectral classes, so separability among classes increased. Because the number of thematic classes also increased with class-label precision, the chance probability for class confusion increased as well, reducing classification accuracy. Our study

indicates that no single best solution exists across study areas, but that the MDGP-scaling method integrates the quantitative evaluation of scaling parameter selection and its effects on representativeness of classification systems, information retention at the local (pixel) and landscape level, and spectral-detection of the scaled classes. The optimal scaling index that includes the class detection accuracy in its calculation is a useful index to determine parameter selection for MDGP-scaling. Applying this index when selecting optimal parameters allows user-specific and preference-optimized solutions.

Detection accuracy of the scaled classes from spectral data could further increase when including multi-season spectral data because hydrological and phenological cycles and the associated spectral reflectance patterns vary among plant communities. We used single-season Landsat data to detect the scaled classes, which is the most likely scenario for change detection applications in tropical regions because it is difficult to acquire cloud-free wet season data.

Table 7 Scaling effects on spectral detection accuracies for NESRS for varying class-label precisions (Parts) and landscape representativeness thresholds (Rpr.)

Parts	Repr.	IR_{cRL} (%)	Total Classes	CLF_{pRL}	Overall Accuracy (%)	OA-CIL (%)	OA-CIU (%)	OSI_a
1	*1*	70.3	11	1.00	69.4	68.9	69.8	0.65
1	*5*	70.3	11	1.00	69.4	69.0	69.8	0.65
1	*10*	70.3	11	1.00	69.5	69.1	70.0	0.65
1	*15*	70.3	11	1.00	69.4	68.9	69.8	0.65
1	*20*	70.3	11	1.00	69.3	68.9	69.8	0.65
2	*1*	82.5	21	0.91	70.2	69.7	70.6	0.69
2	*5*	80.3	13	0.93	70.8	70.4	71.2	1.03
2	*10*	78.7	12	0.92	71.0	70.6	71.4	1.02
2	*15*	76.4	11	0.92	71.1	70.7	71.5	0.97
2	*20*	76.4	11	0.92	71.1	70.6	71.5	0.97
3	*1*	86.2	26	0.87	69.7	69.3	70.1	0.61
3	*5*	83.3	16	0.94	70.1	69.7	70.5	0.96
3	*10*	*80.3*	*13*	*1.00*	*70.7*	*70.3*	*71.1*	*1.10*
3	*15*	78.3	12	1.00	70.5	70.1	70.9	1.08
3	*20*	78.3	12	1.00	70.5	70.1	70.9	1.08
4	*1*	88.2	33	0.79	69.6	69.2	70.0	0.47
4	*5*	85.0	18	0.90	70.6	70.2	71.0	0.88
4	*10*	82.5	14	0.93	70.3	69.9	70.8	1.05
4	*15*	77.7	11	0.79	71.4	71.0	71.8	0.90
4	*20*	77.7	11	0.79	71.3	70.9	71.7	0.90
5	*1*	89.1	37	0.61	69.4	69.0	69.8	0.33
5	*5*	84.0	15	0.54	72.1	71.7	72.5	0.62
5	*10*	78.9	10	0.63	72.6	72.1	73.0	0.86
5	*15*	77.1	8	0.57	73.2	72.8	73.6	0.90
5	*20*	77.1	8	0.57	73.1	72.7	73.5	0.90

Column heading abbreviations as in Table 4. Best solution is highlighted in italics

Land-cover change and biophysical parameter estimation using remote sensing

Ever-increasing spatial resolution of remote sensors has led to land-cover maps with very high spatial and thematic precisions. Since thematic map classification schemes are not uniform across the spatial scales of sensors, mixed-pixel classes with a coarse class label from an earlier time can be represented by pure pixels of their constituent class components in more recent maps that have a finer spatial resolution. Change detection over long periods, therefore, must reconcile the thematic class schemes that were used at each spatial scale. A change detection method that generates a representative classification scheme from high resolution thematic data and that can be detected

from the multispectral data of the coarser resolution (e.g., Landsat) can facilitate the detection of scaled classes across time.

Our application of the MDGP-scaling algorithm to upscaling vegetation cover in two Everglades wetland landscapes produced a classification scheme that effectively generated classes that can be detected from coarser resolution spectral data and can now be used to examine temporal change using recent vs. historical Landsat data without having to forfeit the high compositional information content of high-resolution maps derived from other sensors. Our study showed that including the generated mixed classes not only retained more information in class labels, representing the ground conditions in much higher detail, but also led to higher detection accuracies from Landsat

Table 8 NESRS class schema and class proportions (Class Prop.) for an optimal scaling solution of a 3-part (33%) class-label precision with landscape representativeness of 10%

Class name—33% Precision 10% Representativeness		Class Prop. (%)	Mean IR$_c$ (%)	SD IRc (%)
100	Graminoid Marsh—Broadleaved Emergent*	2.1	56.4	16.1
100	Graminoid Marsh *Cladium**	8.5	89.8	7.6
67	Graminoid Marsh *Cladium*	12.1	76.1	14.8
33	Graminoid Marsh *Cladium* Dense			
67	Graminoid Marsh *Cladium*	14.9	82.8	12.9
33	Graminoid Marsh *Cladium* Sparse			
100	Graminoid Marsh *Cladium* Dense *	9.6	74.5	16.6
100	Graminoid Marsh *Cladium* Sparse*	18.0	90.3	8.1
67	Graminoid Marsh *Cladium* Sparse	20.7	54.8	16.3
33	Graminoid Marsh *Cladium*			
100	Graminoid Marsh Dense*	1.1	70.8	17.6
100	Graminoid Marsh Sparse*	5.5	60.1	17.7
100	Graminoid Marsh *Typha**	2.0	74.1	19.9
100	Shrub Bayhead*	1.9	74.1	19.9
100	Shrub *Salix**	3.3	74.1	20.2
100	Tree Bayhead*	0.3	82.5	19.6

Left column gives the relative abundance of original classes in the scaled class. Column heading abbreviations as in Table 5

*Indicate original classes. Mixed classes make up 47.7% of the landscape

data. The MDGP-scaling algorithm is the first algorithm that facilitates the exploration of label precision of mixed classes on the detectability of those classes from a medium or low-resolution sensor. In our case, the high vegetation heterogeneity of the Everglades landscape makes mixed Landsat pixels the norm rather than the exception (e.g., 60.1% mixed classes in WCA3A, 47.7% in NESRS for the optimally scaled classification systems). Applying the new Landsat-scale classification system that was derived from the high-resolution co-occurrence patten of plant communities to historic Landsat scenes now allows for change detection at higher thematic precision.

Higher information retention in more detailed classification systems is also of great interest when estimating biophysical variables from remotely sensed data. The difficulties produced by spatial heterogeneity on the reliable estimation of biophysical variables using remotely sensed data have been identified and described for a suite of parameters and applications (Lu 2006). For example, Leaf Area Index (LAI), which estimates green leaf area per unit ground, and Fraction of Photosynthetically Active Radiation (FPAR) are two important biophysical variables in ecosystem gross primary productivity (GPP) models.

Estimation of these variables from Moderate Resolution Imaging Spectroradiometer (MODIS), which has a high temporal resolution (daily) but coarse spatial resolution (500 m), relies on land-cover knowledge of each pixel (Steltzer and Welker 2006; Zhao et al. 2016). Feagin et al. (2020) acknowledged the difficulty of modeling GPP for wetlands that display high heterogeneity of land cover relative to the coarse resolution of MODIS. Lotsch et al. (2003) demonstrated the sensitivity of LAI and FPAR to land-cover information and how the heterogeneity of vegetation types within a pixel affects LAI estimates in a non-linear fashion (Garrigues et al. 2006). Tian et al. (2002) showed that LAI errors at a coarse resolution are inversely related to the proportion of the dominant land cover in a pixel and that large errors were introduced when the woody component made up only a small proportion of otherwise non-woody pixels.

Lack of knowledge about mixed-pixel composition arises from coarse classification schemes or aggregation of detailed maps with algorithms that do not modify the classification scheme to accommodate mixed pixels (e.g., majority rule) at the scale of modeling the biophysical variable. The MDGP-scaling algorithm generalizes classes but retains much higher

Fig. 13 Scaled plant communities for NESRS (top) and location-specific information retention (IR) for the assigned community class label when compared to the high-resolution map (Fig. 6) (bottom). Class codes in Table 2. Numbers in labels give the class-label precision percent. Coordinates in Meters WGS84 UTM 17 N
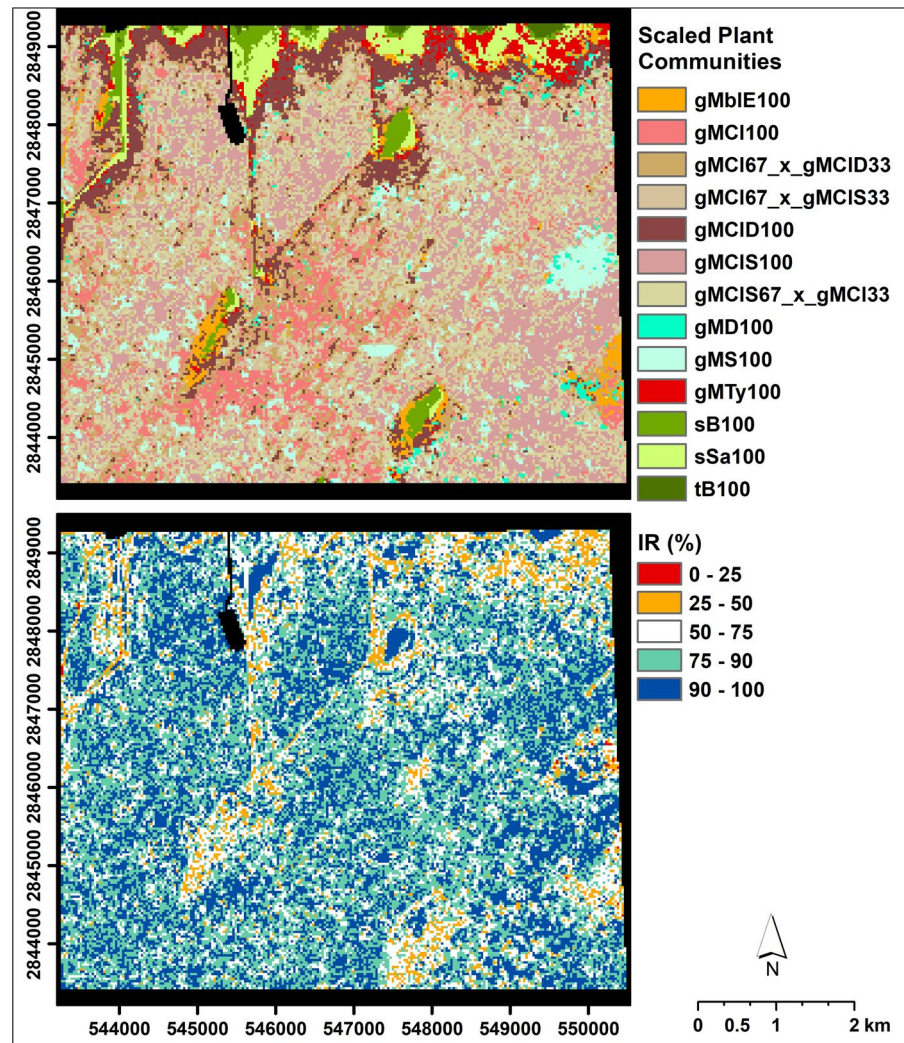

precision in land-cover mixes, which propagates to more accurate calculation or modeling of biophysical variables. Especially for very heterogeneous landscapes such as wetlands, knowing the approximate relative abundance of vegetation cover types within each response unit (pixel) of moderate-resolution remotely sensed data will allow us to reduce error and uncertainty of biophysical variable estimates.

Conclusion

Understanding the effects of scale on process/pattern feedback is often the objective of landscape ecological studies, and much attention has been drawn to

defining and determining appropriate scales. The effect of the scaling process itself, however, is rarely considered, and the loss of information is usually unknown or unquantified because default methods in GIS software do not offer sophisticated choices for scaling categorical data. We demonstrated that the application of the MDGP-scaling algorithm when up-scaling natural landscapes enables the selection of scaling parameters that preserve or retain more information and that an increase in class-label precision also leads to an increase in detection accuracy of scaled classes from multispectral data. Because the algorithm generates scale-representative classification schemes with frequently occurring mixed classes, transition or expansion of ecotones is more

Fig. 14 Scaled plant-community classes predicted from Landsat spectral data (top) and location-specific classifier probability for class-label assignment (bottom) for NESRS. Classes as in Fig. 13. Coordinates in Meters WGS84 UTM 17 N
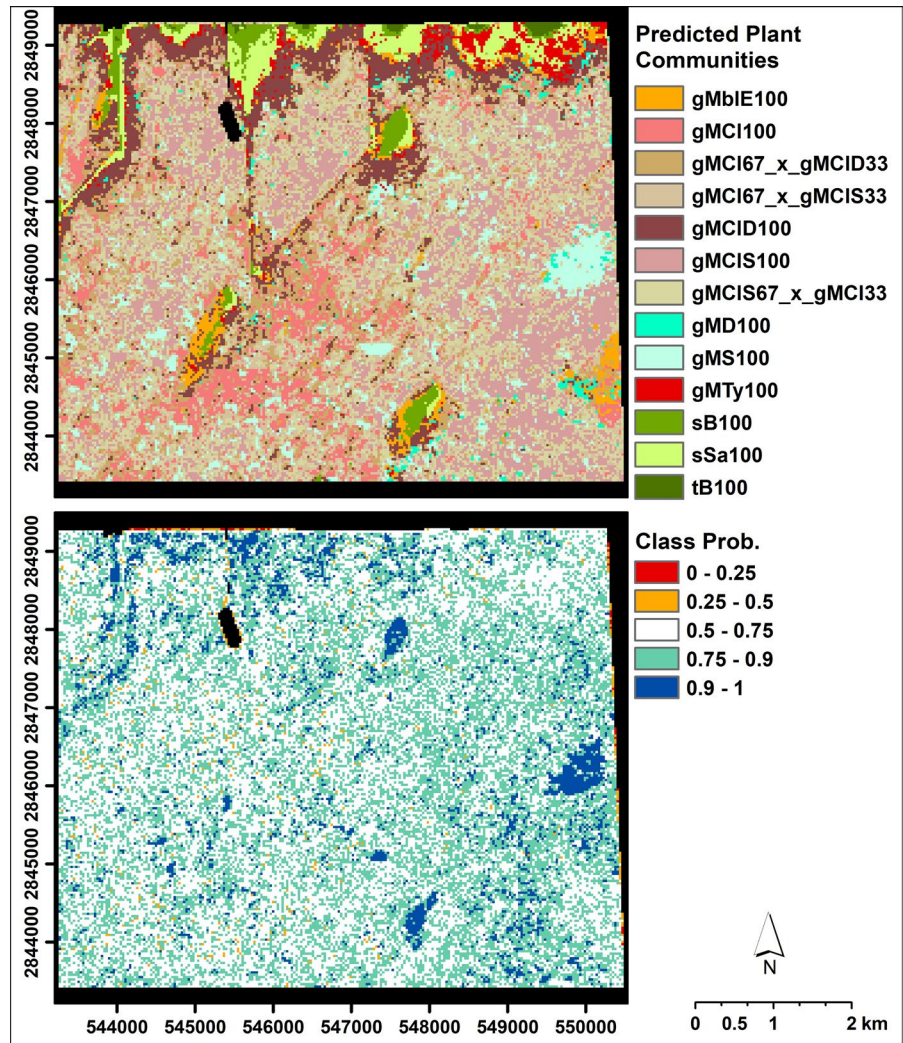
likely to be detected when comparing two categorical maps that have been generated at different spatial resolutions.

The suite of scaling solutions that can be generated by varying scaling parameters in MDGP-scaling showed that information retention, class-label fidelity, and detection accuracy need to be evaluated together to negotiate trade-offs for a specific application. The analysis reported here demonstrated that detection of scaled classes from lower resolution spectral data was possible and that the evaluation framework facilitates parameter selection that optimizes scaling results. Quantifying class-specific and location-specific information retention for the scaled products also enables estimation of spatially explicit confidence or error at the low-resolution grid cell level and thus of error propagation to model results.

References

Bivand R, Keitt T, Rowlingson B (2013) Rgdal: bindings for the geospatial data abstraction library R. package Vienna, Austria. R Found Stat Comput 0:8–10

Braun-Blanquet J (1964) Pflanzensoziologie, Grundzüge der Vegetationskunde, 3rd edn. Springer, Wien

Breiman L (1984) Classification and regression trees. Wadsworth International Group, Belmont

Buyantuyev A, Wu J (2007) Effects of thematic resolution on landscape pattern analysis. Landsc Ecol 22(1):7–13

Cohen J (1968) Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol Bull 70:213–220

Coulston JW, Zaccarelli N, Riitters KH, Koch FH, Zurlini G (2014) The spatial scan statistic: a new method for spatial aggregation of categorical raster maps [Quantitative Methods]. ArXiv E-Prints, pp 1–14. http://arxiv.org/abs/1408.0164

Czekanowski J (1909) Zur differential Diagnose der Neandertalgruppe. Korrespondenzblatt Der Deutschen

Gesellschaft Für Anthropologie, Ethnologie Und Urgeschichte. 40:44–47

De Cáceres M, Wiser SK (2012) Towards consistency in vegetation classification. J Veg Sci 23:387–393

De Cáceres M, Font X, Vicente P, Oliva F (2009) Numerical reproduction of traditional classifications and automatic vegetation identification. J Veg Sci 20:620–628

Gann D (2018) Quantitative spatial upscaling of categorical data in the context of landscape ecology: a new scaling algorithm biology. Florida International University PhD

Gann D (2019) Quantitative spatial upscaling of categorical information: the multi-dimensional grid-point scaling algorithm. Methods Ecol Evol 10:2090–2104

Garrigues S, Allard D, Baret F, Weiss M (2006) Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. Remote Sens Environ 105:286–298

Halstead BJ, Rose JP, Clark D, Kleeman PM, Fisher RN (2022) Multi-scale patterns in the occurrence of an ephemeral pool-breeding amphibian. Ecosphere. https://doi.org/10.1002/ecs2.3960

He HS, Ventura SJ, Mladenoff DJ (2002) Effects of spatial aggregation approaches on classified satellite imagery. Int J Geogr Inf Sci 16(1):93–109

Hijmans RJ, van Etten J (2010) raster: geographic analysis and modeling with raster data Version,. R Package Vers 1:7–29

Ju J, Gopal S, Kolaczyk ED (2005) On the choice of spatial and categorical scale in remote sensing land cover classification. Remote Sens Environ 96(1):62–77

Kleindl WJ, Powell SL, Hauer FR (2015) Effect of thematic map misclassification on landscape multi-metric assessment. Environ Monit Assess 187(6):321

Knight JF, Tolcser BP, Corcoran JM, Rampi LP (2013) The effects of data selection and thematic detail on the accuracy of high spatial resolution wetland classifications. Photogramm Eng Remote Sens 79(7):613–623

Kuhn M, Wing J, Weston S et al (2016) caret: classification and regression training version. R Package Vers 6:0–82

Langford WT, Gergel SE, Dietterich TG, Cohen W (2006) Map misclassification can cause large errors in landscape pattern indices: examples from habitat fragmentation. Ecosystems 9(3):474–488

Latifovic R, Olthof I (2004) Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data. Remote Sens Environ 90(2):153–165

Lischke H, Löffler TJ, Thornton PE, Zimmermann NE (2007) Model up-scaling in landscape research. In: Kienast F, Wildi O, Ghosh S (eds) A changing world: challenges for landscape research. Springer, Netherlands, pp 249–272

Lotsch A, Tian Y, Friedl MA, Myneni RB (2003) Land cover mapping in support of LAI and FPAR retrievals from EOS-MODIS and MISR: classification methods and sensitivities to errors. Int J Remote Sens 24:1997–2016

Lu D (2006) The potential and challenge of remote sensing-based biomass estimation. Int J Remote Sens 27:1297–1328

McVoy CW, Said WP, Obeysekera J et al (2011) Landscapes and hydrology of the predrainage everglades. University Press of Florida, Gainesville

Meentemeyer V, Box EO (1987) Scale effects in landscape studies. In: Turner MG (ed) landscape heterogeneity and disturbance. Springer, New York, pp 15–34

Newman EA, Kennedy MC, Falk DA, McKenzie D (2019) Scaling and complexity in landscape ecology. Front Ecol Evol 7:293

O'Neill RV, Hunsaker CT, Timmins SP et al (1996) Scale problems in reporting landscape pattern at the regional scale. Landsc Ecol 11:169–180

Ozdogan M, Woodcock CE (2006) Resolution dependent errors in remote sensing of cultivated areas. Remote Sens Environ 103(2):203–217

Pontius RG Jr, Connors J (2009) Range of categorical associations for comparison of maps with mixed pixels. Photogramm Eng Remote Sens 75(8):963–969

Quattrochi DA (1991) Remote sensing for analysis of landscapes: an introduction. In: Turner MG, Gardner RH (eds) quantitative methods in landscape ecology: the analysis and Interpretation of landscape heterogeneity, vol 82. Springer, New York, pp 51–76

R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Raj R, Hamm NAS, Kant Y (2013) Analysing the effect of different aggregation approaches on remotely sensed data. Int J Remote Sens 34(14):4900–4916

Raptis VS, Vaughan RA, Wright GG (2003) The effect of scaling on land cover classification from satellite data. Comput Geosci 29(6):705–714

Richards JH, Gann D, Sadle J (2015) Vegetation trends in indicator regions of Everglades National Park. Everglades National Park, Florida City

Rossiter DG (2014) Technical Note: Statistical methods for accuracy assessment of classified thematic maps. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.77.247&rep=rep1&type=pdf

Stehman SV, Foody GM (2019) Key issues in rigorous accuracy assessment of land cover products. Remote Sens Environ 231:111199

Steltzer H, Welker JM (2006) Modeling the effect of photosynthetic vegetation properties on the NDVI-LAI relationship. Ecology 87:2765–2772

Teng SN, Svenning J-C, Santana J et al (2020) Linking landscape ecology and macroecology by scaling biodiversity in space and time. Curr Landsc Ecol Rep 5:25–34

Tian Y, Woodcock CE, Wang Y et al (2002) Multiscale analysis and validation of the MODIS LAI product: I. Uncertain Assess Remote Sens Environ 83:414–430

Tichý L, Chytrý M, Šmarda P (2011) Evaluating the stability of the classification of community data. Ecography 34:807–813

Turner MG, O'Neill RV, Gardner RH, Milne BT (1989) Effects of changing spatial scale on the analysis of landscape pattern. Landsc Ecol 3(3):153–162

van den Boogaart KG, Tolosana-Delgado R (2008) compositions: a unified R package to analyze compositional data. Comput Geosci 34:320–338

Van Der Maarel E (1979) Transformation of cover-abundance values in phytosociology and its effects on community similarity. Plant Ecol 39:97–114

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer, New York

Wickham J, Riitters KH (2019) Influence of high-resolution data on the assessment of forest fragmentation. Landsc Ecol 34:2169–2182

Wildi O (2010) Data analysis in vegetation ecology. Wiley-Blackwell, West

Wu J, David JL (2002) A spatially explicit hierarchical approach to modeling complex ecological systems: theory and applications. Ecol Model 153(1–2):7–26

Wu J, Hobbs R (2002b) Key issues and research priorities in landscape ecology: an idiosyncratic synthesis. Landsc Ecol 17:355–365

Wu J, Shen W, Sun W, Tueller PT (2002) Empirical patterns of the effects of changing scale on landscape metrics. Landsc Ecol 17(8):761–782

Xu K, Tian Q, Yang Y, Yue J, Tang S (2019) How up-scaling of remote-sensing images affects land-cover classification by comparison with multiscale satellite images. Int J Remote Sens 40(7):2784–2810

Xu K, Zhang Z, Yu W et al (2021) How spatial resolution affects forest phenology and tree-species classification based on satellite and up-scaled time-series IMAGES. Remote Sens 13(14):2716

Zhao J, Wang Y, Zhang H et al (2016) Spatially and temporally continuous LAI datasets based on the mixed pixel decomposition method. Springerplus 5:516