Supplementary Material

Methodology

For each set of imagery, we selected RGB cells of data for *S. alterniflora*, *S. patens*, and water that were representative of each class, including color extremes. Typically, machine learning to classify RGB images is performed with a candidate image set that a model uses to train and refine classifications for other images in a testing set, randomly partitioned from the candidate set prior to training (Cheng & Chen 2003, Stolar et al. 2017, Jiang et al. 2019). The images used for the candidate set are then selected for a given model based on user-defined criteria. The selection of training cells based on what is known as an ecological class or, in our case, a single species is very similar to the process of selecting images for a candidate set. In this way, we treat each data selection as an individual, albeit simplistic (single color), characterizing image. Evaluation of a trained model is frequently obtained as a measure of accuracy from Equation 1 (below), given a random partition ratio of 80% training to 20% testing data of the total selected dataset (Gholamy et al. 2018). Peak model accuracy is desirable as overfitting of model classes can be problematic and lead to a reduction in the accuracy measure (Ying 2019). We thus define accuracy equilibrium as the stabilization of the accuracy measure of a learned model for an increasing number of training data. The appropriate number of training data can then be represented as the intersection of stabilization and peak performance before reduction in accuracy due to overfitting.

$$\text{Accuracy} = \sum(C_{correct}) / N_{total} \qquad \text{Eq. 1}$$

where $C_{correct}$ is an evaluation of each model classification in the testing set versus corresponding manual assignments. N_{total} is the total number of testing entries.

We fulfilled our selection criteria by systematically increasing the number of training data, akin to a sensitivity analysis, for our models while iteratively building testing sets to maintain an 80-20 train-test relationship. Further, to avoid "cherry-picking" representative testing data, our only active selection was the total dataset, including both training and testing data, which was done with a stepwise increase of 50 entries per class. Our method thus allowed all classes to be equally represented across all years and for both marshes, versus other methods which have biased machine learning results to more accurately fit the most representative states of the most common and highest density classes on the landscape. Achieving a fully random design is not possible when ensuring equal representation between classes given that systematic boundaries cannot be placed on the landscape for classes while aiming for efficient

classification. Further, bounding classes would defeat the purpose of using machine learning to classify landscapes. Thus, we opted for a half-random design by using the above selection procedure and then randomly partitioning those selections into the 80-20 ratio at every step. This allowed us to build both datasets and ensure effective comparison between models with increasing training data until accuracy equilibrium was reached.

A CART classifier (Krzywinski and Altman, 2017) was then used in Google Earth Engine (GEE; Gorelick et al. 2017) to predict classes for *S. alterniflora*, *S. patens*, and water throughout our target regions. Sufficient steps were completed past peak performance to ensure that the overfitting clause was satisfied and not simply a trough in performance prior to a higher peak. Further, we visually evaluated our models at each step to correlate model accuracy with predictions for known land classes. This was done out of caution as high model accuracy with few training points was often indicative of poor class coverage and more errant model predictions, requiring further data before accuracy equilibrium could be reached. Our iterative selection of data was performed for all years of imagery to ensure model accuracy reached equilibrium and variance between years was captured. There should not be a standard for the total number of training data used as each model is unique and requires calibration for specific changes in environment, be they spatial or temporal. Furthermore, data must be selected from the region to be modeled as any fluctuation in community composition from physical, chemical, or hydrologic inputs that could result in dramatically different appearances and thus performance of the model.

Tables

Table S1 Proximal stream gauge data for our target marshes (modified from Lester et al. 2020). These parameters are reflective of relative inputs to the marshes and as such are used in this study as a proxy. The rivers used for each marsh are 1st order streams and are sustained via the confluence of other smaller rivers. Thus, they also represent the most dominant inputs of streamflow and nutrients to their respective marshes. Note the much larger streamflow and nitrogen inputs to the southern marsh. Total N included all measured nitrogen-based inputs. Total P included all measured phosphorus-based inputs.

Stream Gauge (River)	Marsh	Year	Streamflow (ft^3 s^{-2})	Total N (mg L^{-1})	Total P (mg L^{-1})
01409387 (Mullica River)	Northern	2005	44.9	0.441	0.0117
		2006	49.6	0.466	0.0126
		2007	56.2	0.457	0.0128
		2008	28.1	0.421	0.0126
		2009	41.9	0.470	0.0144
		2010	62.3	0.458	0.0144
		2011	54.1	0.481	0.0169
		2012	39.9	0.470	0.0163
		2013	46.1	0.487	0.0180
		2014	45.5	0.483	0.0184
		2015	47.0	0.478	0.0190
		2016	31.5	0.473	0.0199
01411110 (Great Egg Harbor River)	Southern	2005	210.9	0.828	0.0191
		2006	220.7	0.839	0.0204
		2007	243.1	0.838	0.0194
		2008	157.6	0.865	0.0186
		2009	210.7	0.868	0.0214
		2010	304.9	0.840	0.0207
		2011	283.6	0.872	0.0229
		2012	213.7	0.882	0.0211
		2013	243.2	0.887	0.0234
		2014	259.8	0.881	0.0227
		2015	256.6	0.886	0.0228
		2016	171.3	0.928	0.0216

Table S2 Model comparison results for environmental predictors of the total area of *S. patens* patches in each marsh. Models were compared using leave-one-out cross-validation (Vehtari et al. 2017). ELPD stands for the expected log pointwise predictive density. The farther away a model is from 0 ELPD, the less relative predictive power it has. Thus, the bolded central model is the best model for both raw and interpolated (INTRP) data. All models converged for markov chains before achieving maximum iterations.

	Model Parameterization					
	Streamflow	**Total N**	**Total P**	**Streamflow × Total N**	**Marsh Intercept**	**ΔELPD**
	x	x	x	x	x	-4.4
R	x	x		x	x	-2.7
A	**x**	**x**			**x**	**0**
W	x				x	-1.6
		x			x	-5.2
I	x	x	x	x	x	-1.6
N	x	x		x	x	-0.3
T	**x**	**x**			**x**	**0**
R	x				x	-13.2
P		x			x	-8.4

Figures

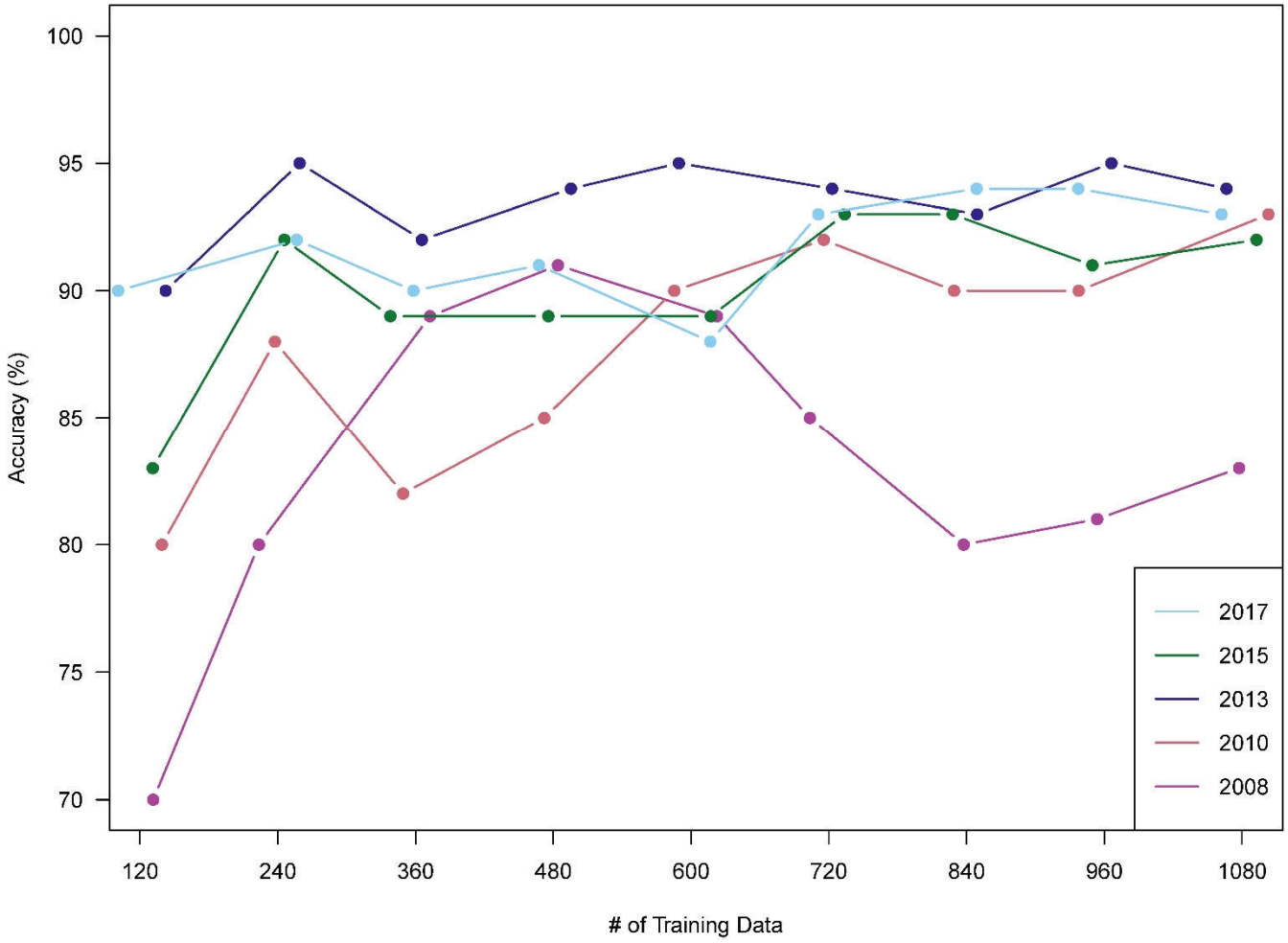

Fig. S1 Model accuracy for increasing numbers of randomly partitioned training data used in machine learning supervision. Analyses were continued at least 5 steps past the selected # of training data to ensure that we satisfied the overfitting clause. X = 1080 was chosen as a graphical cutoff as it was the entry before overfitting for the year with the greatest numbers of training data to reach equilibrium. Created in R, version 4.1.3 (R Core Team 2022).

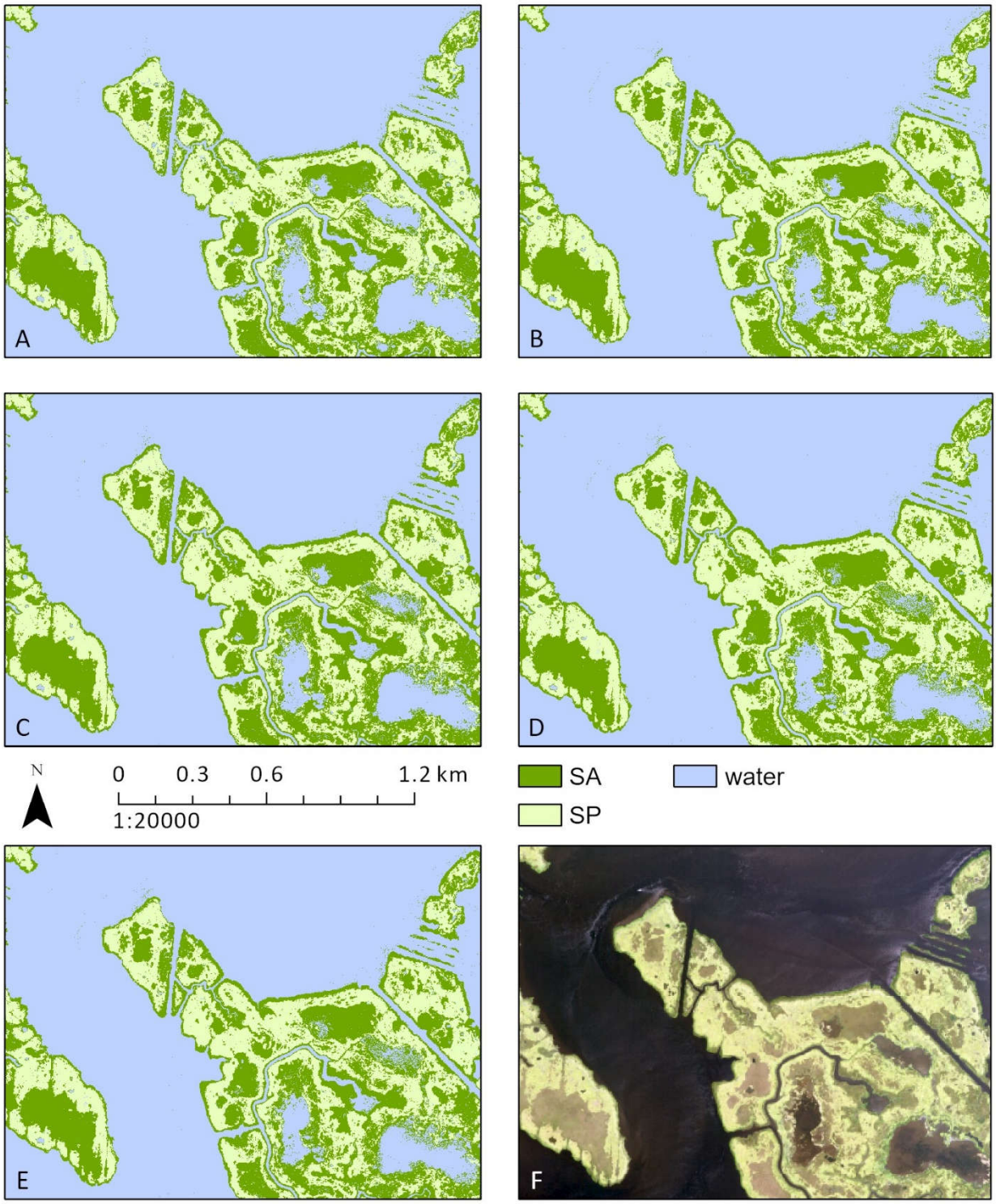

Fig. S2 2019 RGB image classification of a roughly 350 ha sample region (*Spartina* spp. marsh in Galloway and Bass River, NJ; 39°32′54″ N, 74°25′44″ W) for cases with increasing breadths of training data. Pane A was trained with 240 total entries from the surrounding area (Accuracy = 93.3%), Pane B with 480 (Accuracy = 97.5%), Pane C with 720 (Accuracy = 98.3%), and Pane D with 960 (Accuracy = 98.8%). Pane E is the 960 (most accurate) training entry case for comparison with Pane F, a true color image. SA is *S. alterniflora* and SP is *S. patens*. Projection: WGS 84 / UTM Zone 18N. Created in ArcGIS Pro, version 2.9.1 (Esri Inc.).

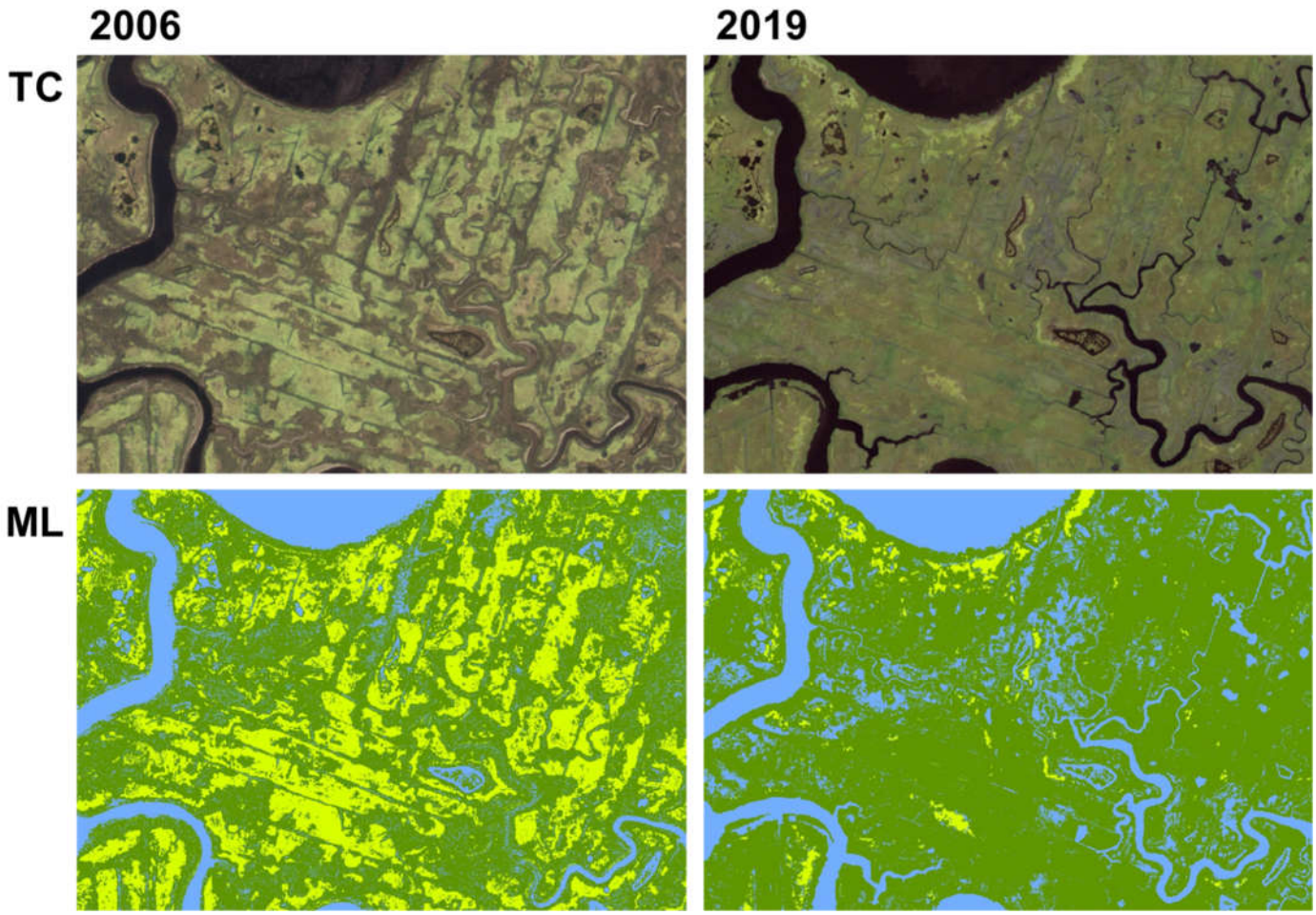

Fig. S3 Geospatial example comparison of NAIP imagery (TC for true color) and our supervised classifications (ML for machine learning) used to confirm machine learning accuracy measures. These depictions are roughly 1/100th the total area of our marshes. *S. alterniflora*: dark green and *S. patens*: light green. Note the consistently accurate predictions for *S. patens*, even capturing exact shapes of patches. Furthermore, note the increase in marsh ponding captured for 2019 as a proxy for fine-scale dynamics accurately represented in our final models. Relative centroid is at 39° 17′ 12.8″ N, 74° 40′ 26.3″ W. Created in Google Earth Engine (Gorelick et al. 2017).

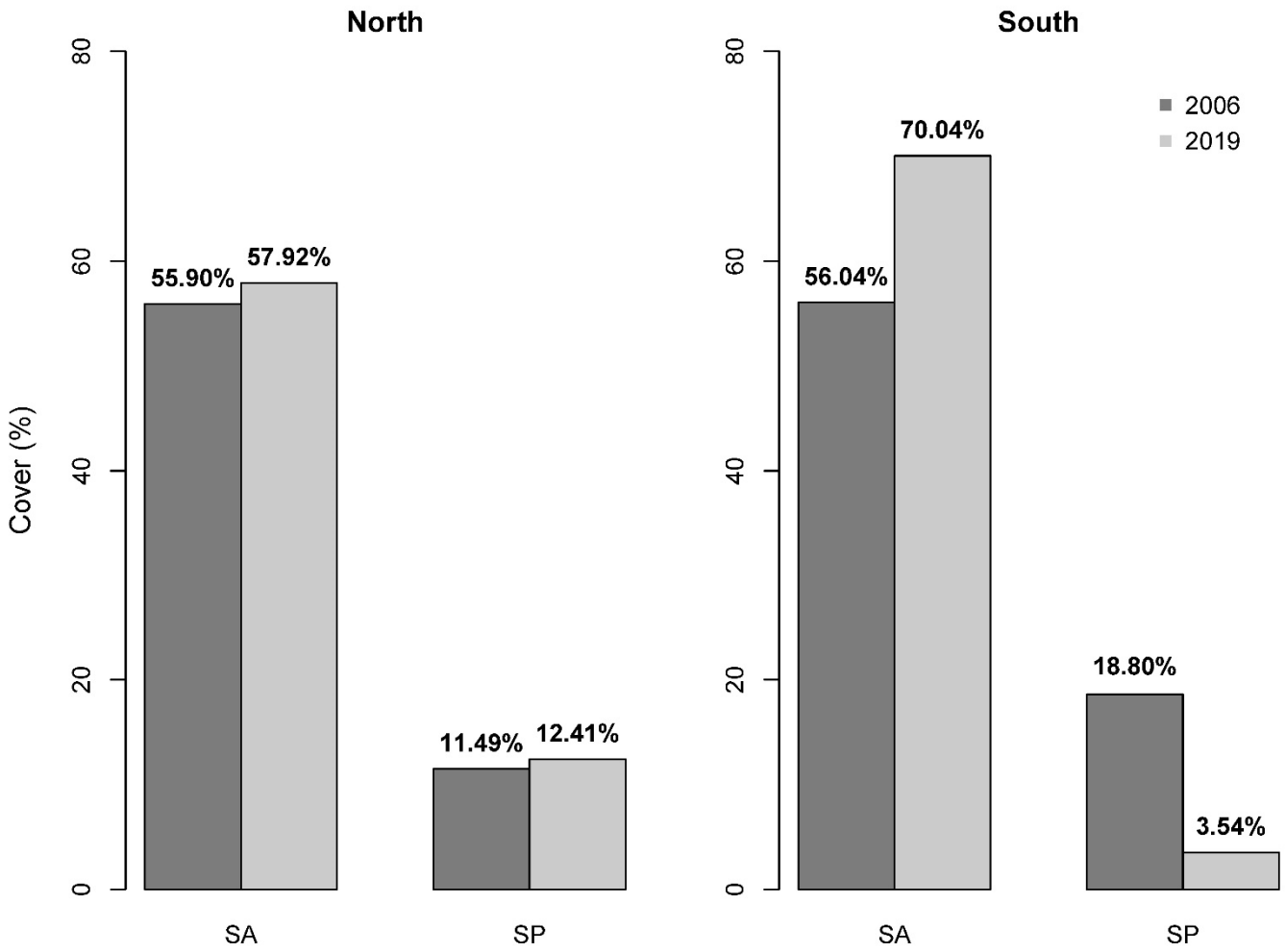

Fig. S4 Relative cover of *S. alterniflora* (SA) and *S. patens* (SP) for our target marshes from 2006 to 2019. Percentage values do not sum to 100 as water is not included, an aesthetic chosen to emphasize the target relationship. Created in R, version 4.1.3 (R Core Team 2022).

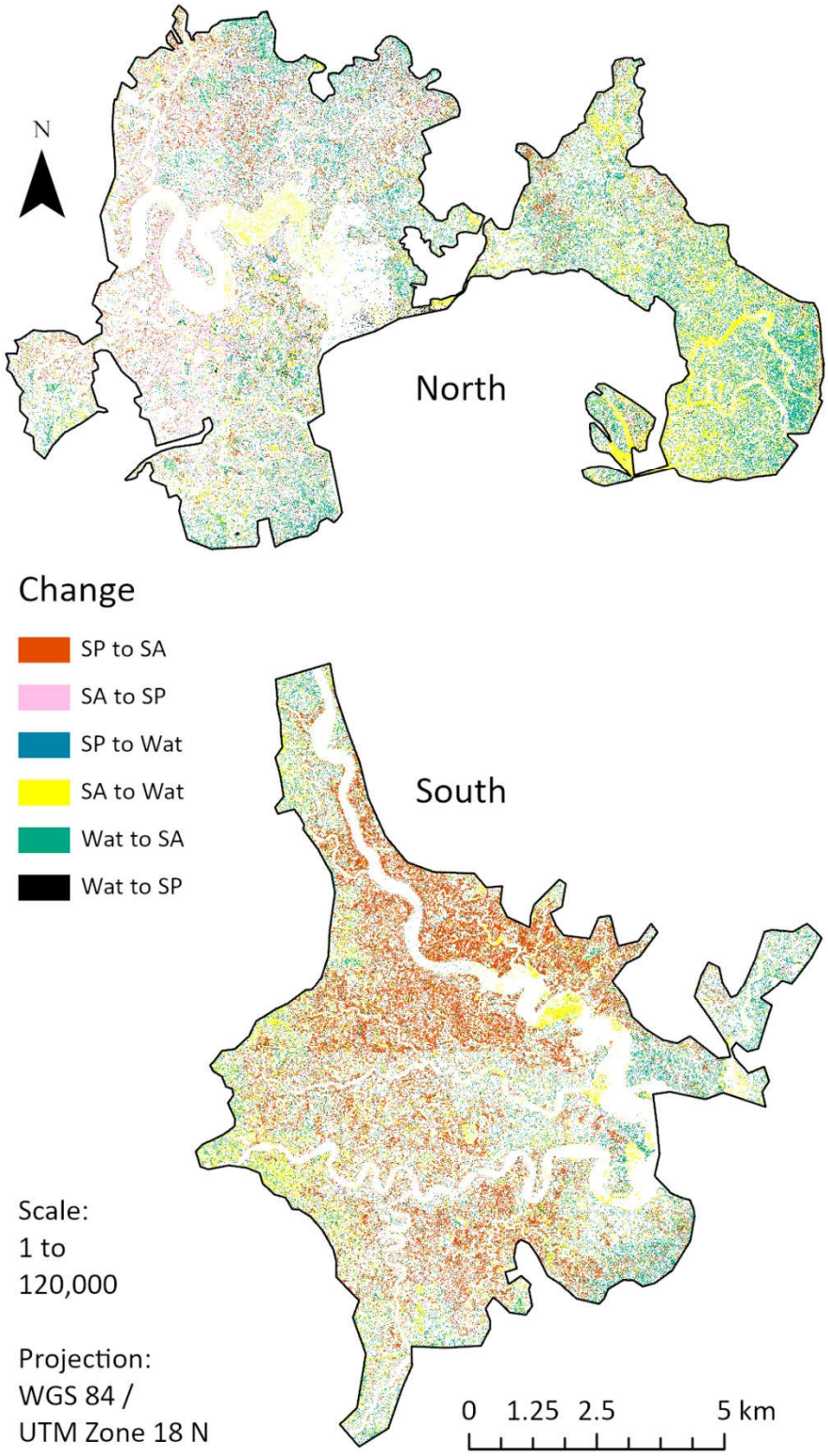

Fig. S5 Geospatial community change (2006-2019) for our target marshes. Note that most obvious changes occur from *S. patens* (SP) to *S. alterniflora* (SA) in the southern marsh. Northern marsh relative centroid is at 39° 33′ 20″ N, 74° 26′ 26″ W; southern marsh relative centroid is at 39° 18′ 22″ N, 74° 41′ 42″ W. Created in ArcGIS Pro, version 2.9.1 (Esri Inc. 2021).