Identifying *Limonium ramosissimum* populations using a species distribution model and ground searches in tidal marshes of South San Francisco Bay

Abstract:
Early detection of invasive species is required for cost effective management of plant invasions, but finding nascent populations often requires extensive field surveys. Species distribution models provide a tool to focus field surveys to habitat likely to be invaded, and GPS tracking and mapping can be used to quantify and document search results. A logistic regression based-species distribution model was developed for an invasive wetland plant, *Limonium ramosissimum,* in South San Francisco Bay salt marshes, and the results ground truthed using field searches. While the model's overall goodness of fit was not significant, many *L. ramosissimum* patches were found during ground searches and these were overwhelmingly located in cells the model predicts as moderate to high probability of potential *L. ramosissimum* habitat. Also, the model assigned high probability habitat with moderate accurately overall (Kappa = .68). These results suggest species distribution models are useful tools for identifying new invasions, even when the populations not in equilibrium are being modeled. This model could be refined using additional presence/absence data generated since model development and alternate predictor variables could be explored to improve model predictive power.

Introduction:
Wetlands are prone to plant species invasions because they aggregate seeds, nutrients, and are disturbed frequently (Zedler and Kercher 2004). However, San Francisco Bay's salt marshes are subject to additional factors that promote the spread of invasive plants, including habitat fragmentation, urbanization, horticulture (With, 2004; Burt et al, 2007) and an increase in available space via habitat creation and restoration (Davis et al, 2000). For these reasons, invasive plant species management in San Francisco Bay is a necessity- particularly in restored marshes where initial low competition and resource availability create opportunities for non-native plant establishment.

Once established, invasive wetland plant species may inhibit establishment of native plants, alter habitat structure, lower biodiversity, and change nutrient cycling (Zedler and Kercher 2004, Byers et al, 2002). Unchecked, invasive plants may lead to the development of alternative trophic relationships that may further damage native species if the invasive species is removed (Antonio and Meyerson, 2002) or legacy effects resulting in persistent, less desirable ecological states that resist restoration (Suding et al, 2004). For all these reasons, early identification and removal of harmful invasive species may be critical for successful restoration of native biota in marsh restoration projects.

In 2007, several densely growing populations of an invasive plant, Algerian sea lavender (*Limonium ramosissimum ssp proveniciale*- Figure 1, page 2), were discovered in salt marshes in San Francisco Bay. A perennial, salt-tolerant forb of Mediterranean origin, *L. ramosissimum ssp provenciale* has spread to marshes and tidal lagoons in southern California, from San Diego to Santa Barbara (Archabld and Boyer, unpublished). At Carpenteria Marsh in Santa Barbara, *L. ramosissimum* displays invasive characteristics including broad salinity tolerance, prolific seed production and the ability to compete with native plants (Page et al, 2007).

Figure 1: *Limonium ramosissimum ssp provenciale* at Coyote Pt. Marina, San Mateo in July (A,B,C,D) and in August (E), 2008; at Sanchez Marsh, Burlingame (F).

Figure 2: Known locations of *Limonium ramosissimum* in San Francisco Bay prior to this study.

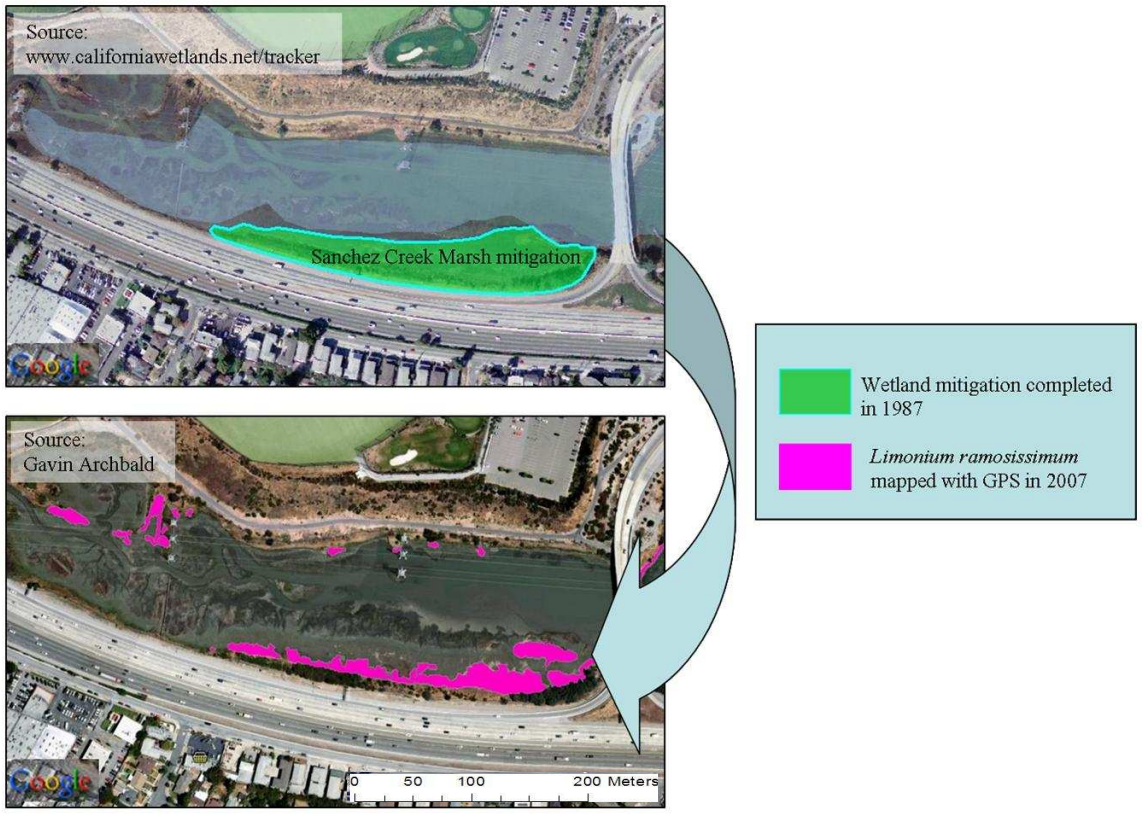

Figure 3: *L .ramisissimum* population at Sanchez Marsh in Burlingame is located at the site of a 1987 wetland mitigation project.

Prior invasion history is a key indicator of future invasion potential (Kolar and Lodge, 2001) and *L. ramosissimum*'s invasion history in California and in San Francisco Bay indicates the plant poses a threat to natural and restored tidal marsh habitat. Prior to this study, *L. ramosissimum* had been found in 12 marshes (Figure 2, page 3) in South San Francisco Bay covering a total of .8 hectares and most of these marshes have a history of either restoration or disturbance. The largest, densest known population is at a restored marsh, Sanchez Creek Marsh (Burlingame, CA) where *L. ramosissimum* densely covers approx. 45% of the marsh habitat created in a 1987 mitigation project (Figure 3, page 3). However, the total extent of the *L. ramosissimum* invasion in San Francisco Bay is unknown.

Planning for the conversion of approximately 15,000 acres of former salt ponds to tidal marsh and other habitats in San Francisco Bay, the South Bay Salt Pond Restoration Project issued a call for scientific research proposals in 2008 to address restoration goals and challenges- among them, the reduction of non-native plant species. This project was funded with the goal of both identifying *L. ramosissimum* populations in S. SF Bay, and developing detection methods for future monitoring. Mapping and monitoring *L. ramosissimum* populations is part of a larger invasive species management plan being carried out by the U.S. Fish and Wildlife Service at the Don Edwards Wildlife Refuge, the major property owner of SBSP Project lands. This paper provides an example and case study of how predictive habitat distribution models combined with ground searches provide a means to map invasive plants.

Predictive habitat distribution models are frequently used to predict a species' actual or potential distribution based on occurrence data and underlying environmental variables (Franklin, 1995, Guisan and Zimmerman, 2000), and many of these models have focused on invasive species (Václavík and Meentemeyer, 2007). Among the most widely used models for prediction is logistic regression, a form of the Generalized Linear Model used when the dependent variable is binary, such as presence/absence of an invasive species, and independent variables are relevant environmental predictors of species presence. The output of this model is the log odds of the binary event occurring which, using GIS, is converted to a spatially explicit probability of occurrence between 0 and 1 for every grid cell in the model. GIS is used to in this approach to generate and sample environmental variables based on species occurrence data and to apply the resulting model to generate a predictive map.

Predictive distribution models are based on the niche concept and assume that species occurrence data represents a population that has realized its fundamental niche and is in psuedoequilibrium (Guisan, 2000). Predictive habitat distribution models violate this assumption, particularly when an invasive species is at an early stage in the invasion process- as is the case for *L. ramosissimum*. In spite of this violation, many studies have used logistic regression to predict the potential range of invasive plant species, including *Lepidium latifolium* in South San Francisco Bay (Vanderhoof et al, 2009). However, because of the violation of this basic assumption, these models are best seen as an approximation of future distribution, not actual or current distribution (Holcombe et al, 2007, Vaclavick and Meentemeyer, 2009).

The goal of this model, therefore, is to develop a probabilistic map of habitat *L. ramosissimum* could invade in South San Francisco Bay marshes, not a predictive model of where founding populations of *L. ramosissimum* have established. By identifying habitats likely to be invaded, the model can be used to limit the scope of field searches and aid in early detection. The

goals of conducting field searches using this model were both to identify *L. ramosissimum* populations and to evaluate the model's accuracy.

Methods:

Study Area:

The model predicts probability of *L. ramosissimum* habitat at marshes and shoreline in South San Francisco Bay in and around the South Bay Salt Pond Restoration Project. This model's extent is limited to regions with Lidar data collected from a 2004 survey (Foxgrover, 2005) which is a key component of predictive layer in the model.

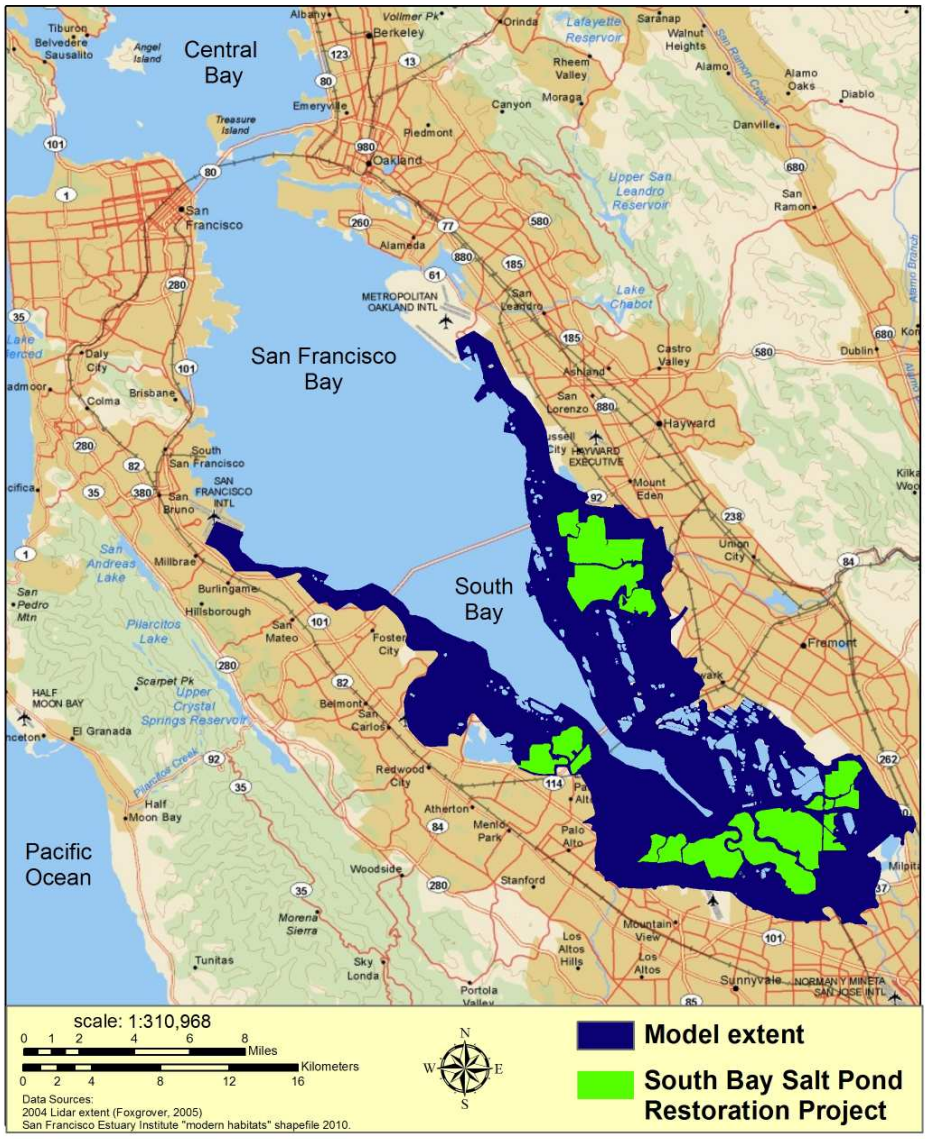

Figure 4: Model extent.

Model Approach:

The key steps involved in creating this model were:

1. Define and generate relevant environmental predictor variables using ArcGIS.

2. Using ArcGIS, sample predictor variables in and out of known *L. ramosissimum* populations.
3. In SPSS, fit the logistic regression model to the sampled data.
4. Apply the model in ArcGIS to predict potential *L. ramosissimum* habitat in the South Bay.
5. Conduct surveys to ground truth model predictions.
6. Based on survey results, generate accuracy assessment.

Step 1- Predictor variables and data acquisition:
All the predictor variables used in the model were chosen to replace a number of known or theorized ecological drivers of *L. ramosissimum's* abundance and distribution in a simple way (Guisan et al, 1999) and are discussed below. All variables were created as rasters (GRID files) in ArcMap 9.3. The predictor variables used in the model and their sources are shown in Table 1.

Table 1: Predictor variables used in logistic regression model and their sources.

Predictor variable	**Source(s)**
Elevation relative to average high tides	2004 South Bay Lidar dataset and interpolated NOAA tidal station data
Kernel density of 2009 *Spartina* hybrid point locations	Draft 2009 *Spartina* point locations from The Invasive *Spartina* Project.
Distance to high marsh habitat	Modern Baylands shapefile acquired from San Francisco Estuary Institute
Distance to water's edge	Heads -up digitization of the bay and channel edges
Topographic slope and aspect	Derived from LIDAR based digital elevation model

Elevation relative to average high tides:
Species distribution models rely on including environmental predictor variables that structure species distributions. Salt marsh species exhibit species distributions which are a function of, among other variables, tidal inundation (Chapman, 1934). Because marshes experience progressively higher tides in the South Bay relative to a fixed geodetic datum, and because marsh species also shift vertically up in response to these higher tides, a new variable was created to normalize elevation data to a measure of average high tides. The purpose of this "elevation relative to average high tides" layer in this model, therefore, is to provide a predictive environmental variable with less variability than elevation alone relative to *L. ramosissimum*'s vertical distribution.

Elevation data:
Elevation data was derived from the South San Francisco Bay 2004 Topographic Lidar data set (Foxgrover and Jaffee, 2005). The data was used to create a 1-meter resolution digital elevation model (DEM) for analysis. Using ArcMap 9.3, Lidar ground return data was converted from xyz lattice files to multipoint features. Average point spacing was set to 1.2 meters. Multipoint features were converted to raster files, projected horizontally in NAD 1983 UTM Zone 10 and vertically in NAVD88.

Tidal data:
To obtain rasters of average high tides, values of predicted mean tide, mean spring range and mean range from 39 primary and secondary tidal stations in the South Bay were exported from http://tidesandcurrents.noaa.gov/tides09/tab2wc1a.html. These values were converted to shapefiles, and transformed from feet relative to the Mean Lower Low Water (MLLW) tidal datum to meters NAVD88 using NOAA's Vertical Datum Transformation java-script based tool (VDatum).

ArcGIS was then used to interpolate between values of mean tide, mean spring range and mean range from tidal stations. The regularized spline interpolator was selected (cell size =10m, weight = .5, number of points = 5). Spline creates a surface that minimizes curvature but that passes through the source points and is often the best method for representing smoothly varying variables (Childs, 2004).

The resulting rasters were then combined to an unconventional proxy for average high tides using raster calculator in ArcGIS. The average high tide raster was created using the follow raster calculation:

Average high tide raster = mean tide raster + ½ (average (spring tidal range raster + mean tidal range raster))

This proxy for average high tides was improvised for this model. It was chosen to capture both average tide heights and include the effect of the higher spring tides. One weakness to this data layer is that systematic errors affect Vdatum's accuracy in the S. Bay. Also, predicted average tides are likely to have higher error than measured tides. An alternative approach which relies on mean higher high water (MHHW) tidal datum measurements is described in a companion paper, "Merging tidal datums and Lidar for Species Distribution Modeling in South San Francisco Bay".

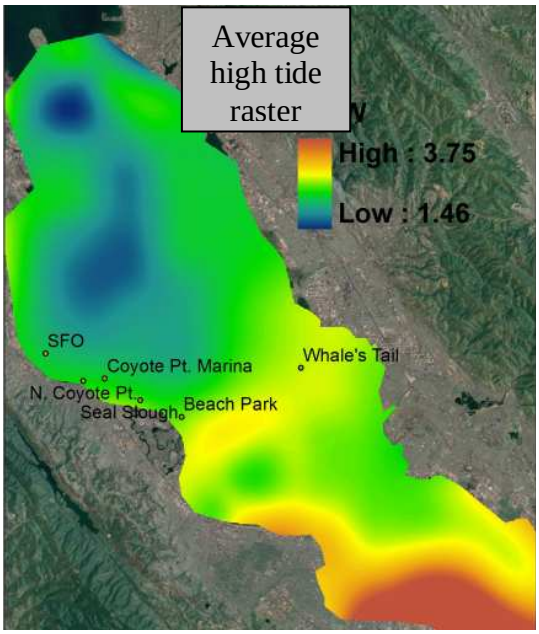

Figure 5: Interpolated average high tide raster generated for this model. *L. ramosissimum* populations prior to model generation within the model extent are shown.

Combining Elevation relative to average high tides as a predictor variable:
A new predictor variable was created by subtracting the Lidar-based DEM from the average high tide raster using raster calculator in ArcGIS. The resulting variable is the elevation value in meters relative to average high tides (Figure 6, page 8).

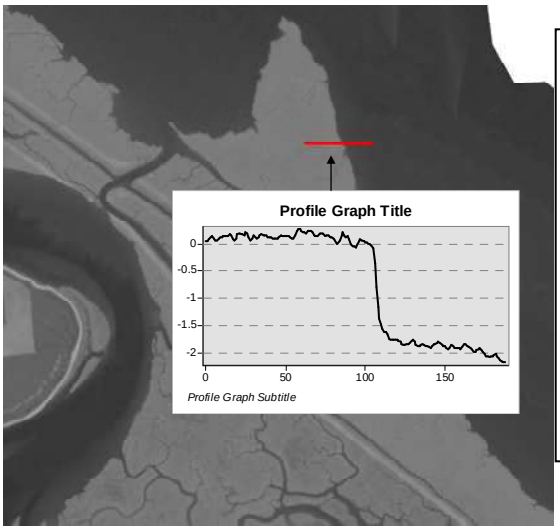

Figure 6: Output showing the variable "elevation relative to average high tide". In a the linear profile of this layer, (red) marsh elevation (light grey) hovers just above 0 m- which corresponds to where average high tide equals marsh elevation. Elevation drops to about -2 m on adjacent mudflats (dark grey).

Kernal Density of 2009 *Spartina* hybrid locations

A species' ability to deliver high propagule pressure to new locations is a key factor predicting invasion potential (Kolar and Lodge, 2001). Experiments have shown *L. ramosissimum* seeds can float for over two weeks in bay water then germinate (Archbald and Boyer, unpublished), indicating seeds have the potential for long distance dispersal events. However, where seeds are likely to travel is from existing source populations is unknown. To address this, based on the assumption that water current and wind patterns drive dispersal in San Francisco Bay, and that these forces act similarly on floating *L. ramosissimum* seeds and floating invasive *Spartina* seeds, a raster quantifying the densities of 2009 *Spartina* hybrid point locations using the quadratic kernel function (Silverman, 1986) in ArcMap was used as a predictor variable (Figure 7).

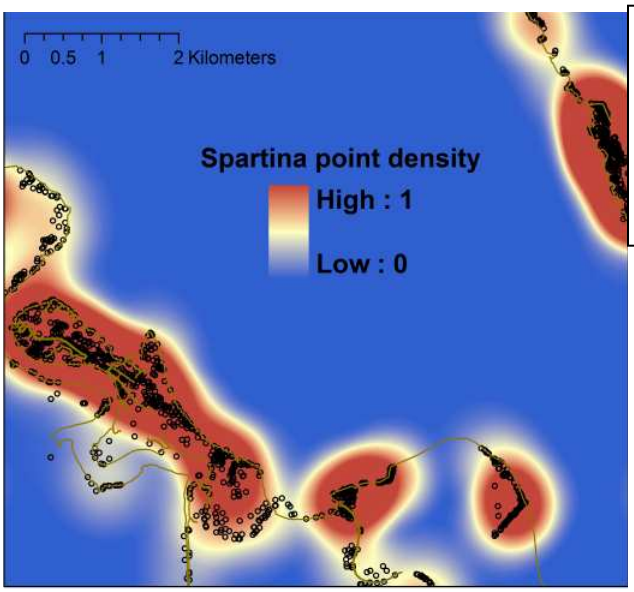

Figure 7: To identify regions where seeds may concentrate, the kernel density function was used with 2009 *Spartina* hybrid locations using an 800m search radius and results were normalized on a 0-1 scale. The results quantify point densities.

Distance to high marsh habitat

L. ramosissimum has exclusively invaded high marsh habitat where it has been observed. Therefore, if additional populations exist in the South Bay marshes, and they are spreading seeds to nearby habitat, distance from large high marsh areas may be a factor in predicting spread. To capture this variable, a Euclidean distance raster was calculated based on polygons delineating high marsh habitat from the "modern habitats" shapefile distributed by San Francisco Estuary Institute in ArcGIS and the resulting raster used as a predictor variable.

Water's edge:

Distance to water's edge was included as a variable to account for dispersal at the marsh scale. Vegetation traps seeds and the probability of seed capture may be a function of distance from the water's edge. Also, disturbance at the water's edge is often high due to wave action, geese grazing and seal haul out sites, and disturbance is strongly correlated with establishment of many invasive species.

To create this variable, shoreline and tidal channels 5 meters in width and greater were "heads up digitized" using ArcMap from a combination of ESRI ArcGIS Online World User Imagery, IKONOS 2009 imagery, and the 2004 South Bay Lidar Dataset in ArcMap. The bay's edge was defined as the transition zone between water or mudflat and marsh vegetation or riprap. The water's edge variable was then created using the Euclidean distance function to the resulting shapefile in ArcMap.

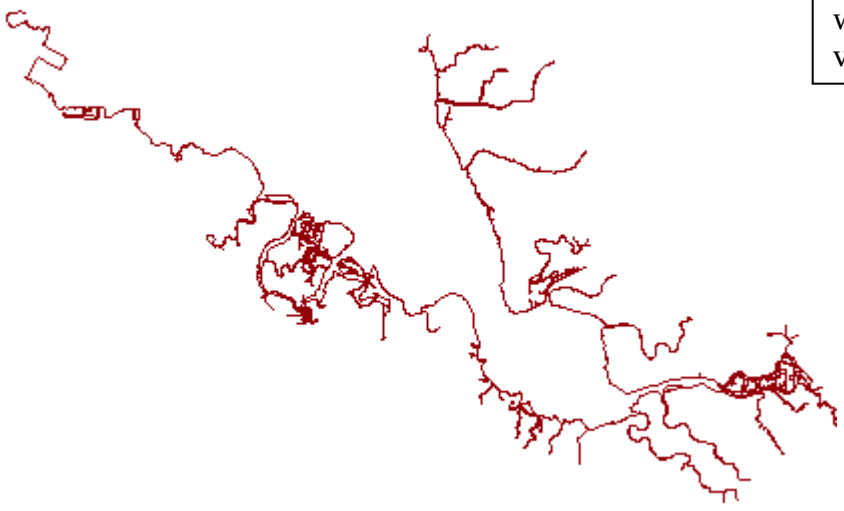

Figure 8: Shapefile used to create the water's edge variable

Topographic variables

Slope and aspect are indirect variables relating to resource and stress gradients. Increased slopes may be associated with higher 0_2 availability for plants as slopes may drain more rapidly after high tide events then flat surfaces. Aspect effects orientation relative to wind, wave action and light. Topographic variables were derived from 1m Lidar-based DEM in ArcMap using the Spatial Analyst extension.

Step 2- Sampling predictor variables for logistic regression:

Samples of predictor variables from searched locations where *L. ramosissimum* is and is not present are required for logistic regression analysis. Marshes and shoreline were searched in 2007-2008 from the San Francisco Airport to Beach Park, Foster City (Figure 9, page 11) and *L. ramosissimum* populations mapped using a handheld Trimble Geo-XH GPS with sub-meter accuracy. Patches larger than $1m^2$ were mapped as polygons and smaller patches were mapped as points. In ArcMap the random sample tool (Hawth's Tools) was used to create three sample points per mapped polygon. These points were merged with mapped *L. ramosissimum* point locations for a total of 328 sample points from 5 geographically distinct populations.

To sample *L. ramosissimum* absence locations in the same region, a "mask" was first created to limit the vertical range to draw samples from. This was done to focus absence sample points in regions relevant to potential ASL establishment (ie not underwater or far above the intertidal). To do this *L. ramosissimum*'s vertical range was compared to tidal datums and a maximum possible range was specified as 1.25m above average high tides and the minimum elevation was set to .2 m below average high tides (Figure 10, page 11) using raster calculator. Using this mask and excluding polygon locations with *L. ramosissimum*, 328 absence samples were generated using the random sample tool.

While power analysis was not conducted, power (the probability of not retaining a false null hypothesis) is expected to increase with balanced sample designs and when prevalence rates are small (Hsieh et al, 1998) and "rules of thumb" for sample sizes range from recommending the dependent variable have at least 10 samples per model parameter (Peduzzi et al., 1996) to 30 samples per model parameter (Pedhazur, 1997). Since this model includes 6 parameters, a conservative estimate is a minimum of 180 samples are required. This sampling design includes 328 samples in vs out of ASL populations, though since *L. ramosissimum* patch polygons were sampled with three points each, some samples from small polygons are likely pseudoreplicates.

Using these 328 sample points in and out of ASL populations, values for all six predictor variables were extracted from the raster layers using the sample tool in ArcMap, then transferred to SPSS for logistic regression analysis. All sample points were used for model fitting since an independent validation step via ground searches was planned.

Figure 9: Region searched for ASL in 07-08 (red) with populations displayed (pink).

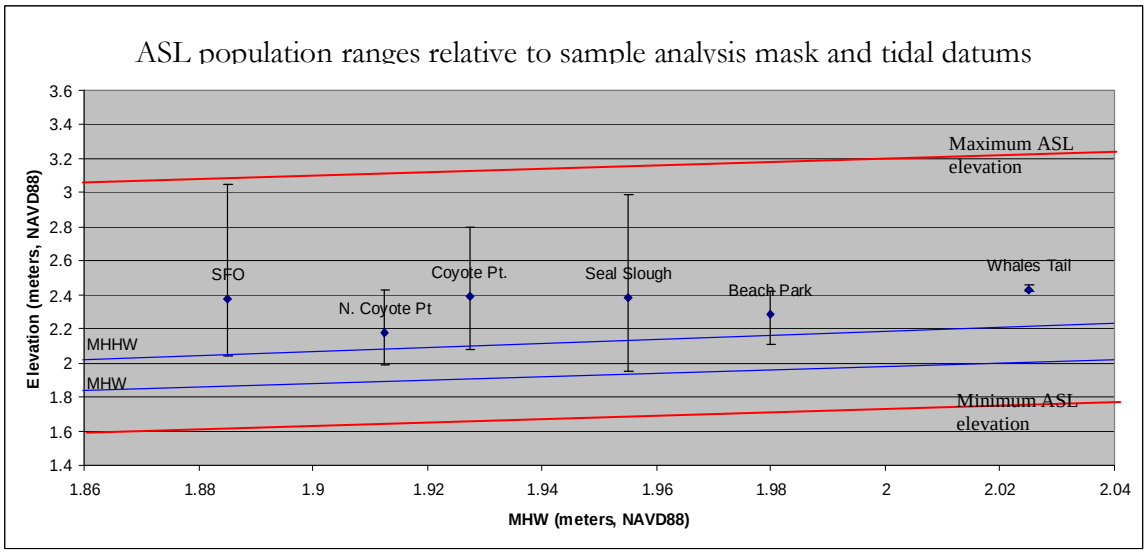

Figure 10: Vertical range of the sample analysis mask (red) relative to ASL's vertical range.

Step 3- Logistic regression analysis:

Logistic regression was carried out in SPSS. Independent variables were entered using backwards step-wise analysis using a significance value <= .05 to include individual variables in the model. Hosmer and Lemeshow tests were used to test for overall fit of binary logistic regression model and the Nagelkerke R^2 test was applied to measure model effect size. No test for spatial autocorrelation (Moran's I), covariates, or log likelihood ratio was conducted.

Step 4- Predict potential *L. ramosissimum* habitat in the South Bay:
Using raster calculator in GIS, coefficients which significantly improved model performance during stepwise model fitting in SPSS were used to calculate probability of *L. ramosissimum* habitat and a 1 m cell by cell basis using the following equations (Garson, 2006):

1. z = 3.437 + (-.004 * [Distance to high marsh]) + (- 1.588 * [average high tide relative to elevation]) + (- 2.478 * [*Spartina* kernel]) + (- .027 * [Distance to water's edge])

2. odds = exp(z)

3. probability = odds/1 + odds

Step 5- Ground truth model predictions:
Model results were uploaded onto a handheld Trimble Geo-XH GPS and displayed as a background to guide field searches of marshes and shoreline across the model extent. 18 days of field searches were carried out including 11 days of boat based searches and 7 days of levy based searches. Searches were carried out in order to cover as much high probability habitat as possible. En route to these locations, significant low and 0 probability habitat was searched, enabling an accuracy assessment of the model.

Boat based searches began approximately two hours before high tide and ended approximately two hours after high tide. All boat paths were recorded by GPS. Searches at high tide enabled high marsh habitat to be seen from the boat. Boat based searches were particularly useful because levies are not passable after rain, and because clapper rail breeding season prevented marsh access on foot for much of the time dedicated to ground truthing. Levy searches were conducted on foot or by slow moving car.

Boat and levy searches were carried out by traveling linear routes- either long shoreline via boat or along levy or shoreline via foot or car. Marsh approximately 20 meters from the search location was visually scanned for *L. ramosissimum*. It was judged that any large *L. ramosissimum* patch would be seen within this distance, but small seedlings would be virtually impossible to see in dense canopy.

Periodically, in both haphazard locations and where wrack, bare ground or other disturbance indicators were seen, more detailed searches were conducted. If the search was conducted by boat, the vessel was nosed against the high marsh and a 30 second search was carried out with bare eyes and binoculars from the bow of the boat. In each location, a GPS position and an estimate of the number of meters of marsh habitat visible from the search point was recorded. If the search was conducted from a car, the car was stopped and the same procedure used.

When found, *L. ramosissimum* populations were mapped using the handheld GPS at the patch scale and percent cover, flowering, and co-occurring species, were recorded.

After searches are complete, search locations and search paths were differentially corrected to improve accuracy and transferred to shapefiles.

In order to evaluate the amount of likely habitat that has been searched, in ArcMap, 20 meter radius buffers were added to search paths recorded with GPS, and survey point locations were buffered by the amount of visible marsh recorded at each search point. These buffered search locations were then used to identify how many cells, at what probability, were searched.

Figure 11: Searching shoreline by boat (left). Survey points and search paths were displayed on model results to quantify extent searched (right).

Step 6- Accuracy assessment:
Two model accuracy assessments were conducted.

First, an assessment was made to determine how accurately the model assigned probability of *L. ramosissimum* habitat to locations where searches found *L. ramosissimum*. To do this, model probability values were extracted from all *L. ramosissimum* point and patch locations not used for model generation at one meter resolution and compared against a random sample of model values outside of *L. ramosissimum* populations.

The second model assessment method tested the degree to which the model accurately assigned "high probability of *L. ramosissimum* habitat" versus "low probability of *L. ramosissimum* habitat". To test this, model probabilities of .7 – .97 were considered high probability, and 0-.3 considered low probability. 25 random points were generated within each of these two probability ranges. Random point locations were then visually assessed using high resolution aerial imagery (NAIP 2009 and Google maps) to determine the accuracy of their assignment. If a sample point was located in mid-marsh, high-marsh or transitional upland, it was considered high probability of *L. ramosissimum* habitat. If it was in low marsh, upland or water, it was considered low probability of *L. ramosissimum* habitat.

The second assessment allowed a Cohen's Kappa statistic to be generated, providing an accuracy rating for the model's probabilistic classification of *L. ramosissimum* habitat.

Results
Predictor variables and logistic regression model:
Results of backwards step-wise logistic regression found four predictor variables significantly improved the performance of the model: *Spartina* kernel density, elevation relative to average

high tides, distance to high marsh, and distance to water's edge (Table 2). The topographic variables slope and aspect did not improve the model's performance and were rejected.

The logistic regression equivalent of an R^2 value, Nagelkerke $R^2 = .596$, indicating about 60% of the variation in the dependent variable is explained by the model. The overall test of how well the model fits the data, the Hosmer and Lemeshow goodness-of-fit test, indicated the model does not fit the data at an acceptable level (statistic < .05).

Table 2: Significant variables included in the model.

Significant variables	**Coefficient**	**SE**	**Significance**
Elevation relative to average high tides	-1.588	.334	< .001
Kernel density of 2009 *Spartina* hybrid point locations	-2.478	.933	.008
Distance to high marsh habitat	-.004	.000	< .001
Distance to water's edge	-.027	.003	< .001
constant	3.473	.349	< .001

These coefficients were then used to generate a probabilistic map of *L. ramosissimum* potential habitat.

Model output of potential *L. ramosissimum* habitat:
Applying the logistic regression coefficients in ArcGIS produced a raster layer of spatially explicit probabilities of potential *L. ramosissimum* habitat across the model extent (Figure 12, page 15). The extent of potential habitat by probability is summarized in Table 3.

Table 3: Areal extent of potential *L. ramosissimum* habitat

Total model extent:	110.9 km^2
Area with > .75 probability	16.7 km^2
Area with > .9 probability	7.12 km^2

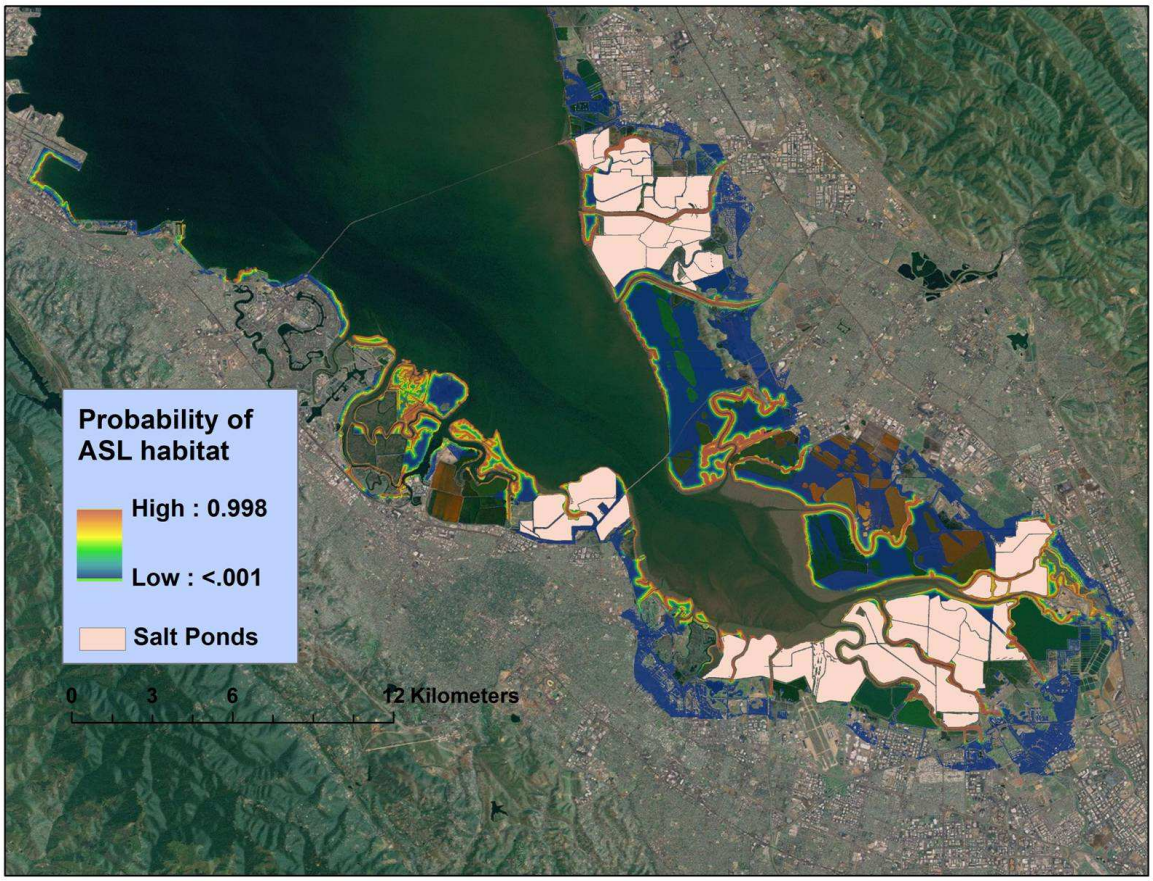

Figure 12: Full extent of logistic regression model of potential *L. ramosissimum* (Algerian sea lavender, or ASL) habitat.

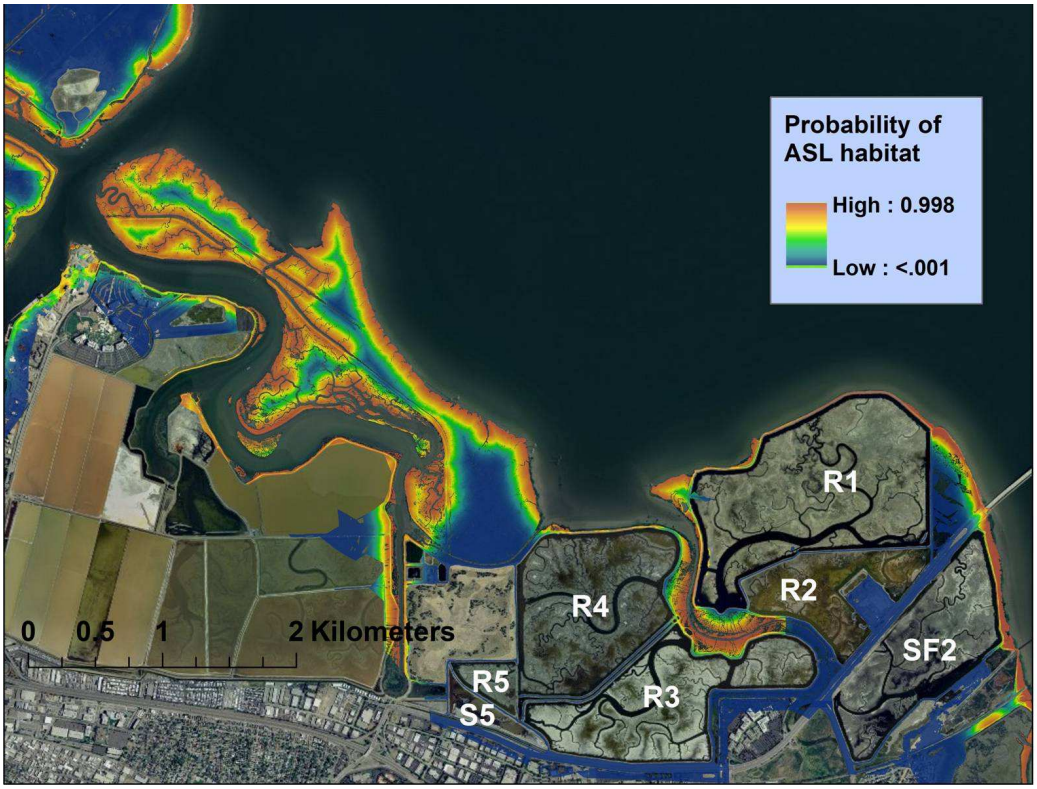

Figure 13: Model results at Ravenswood.

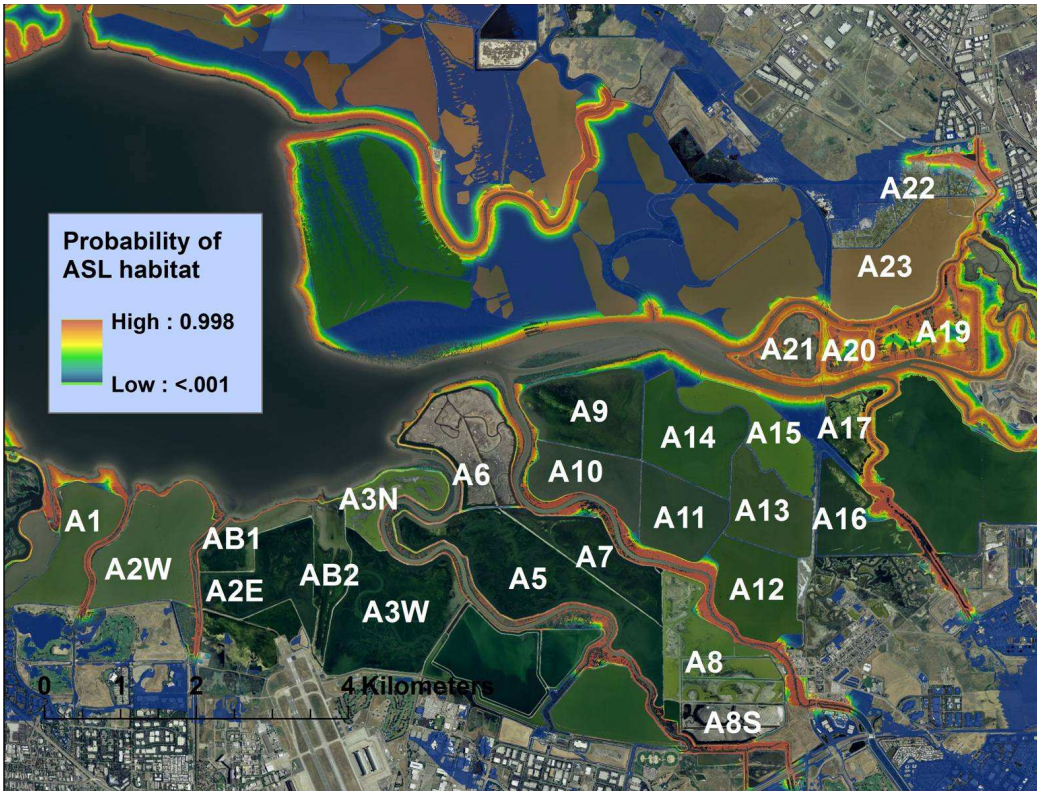

Figure 14: Model results at Alviso.

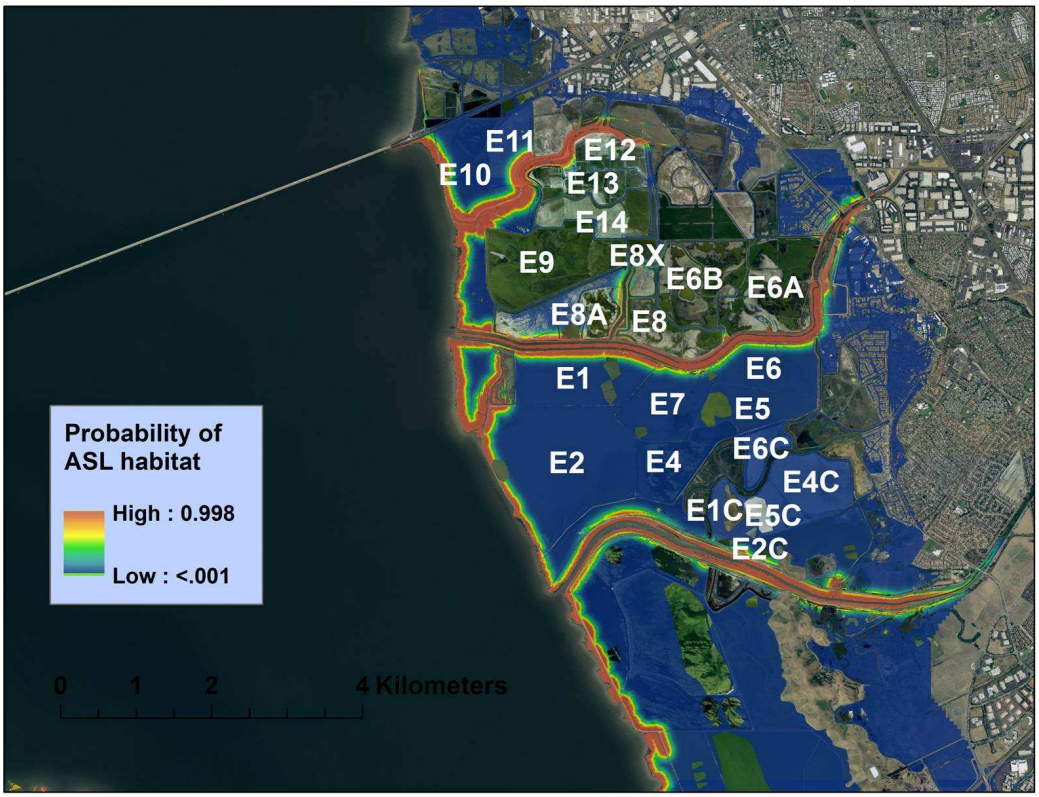

Figure 15: Model results at Eden's Landing.

Survey results:
18 days of boat and levy surveys resulted in 2.2 km^2 of marsh and shoreline searched within the model extent (Figure 16, page 18). Search extent is summarized by probability of potential *L. ramosissimum* habitat in Table 4.

Table 4: Model extent searched

Total area searched in model	2.2 km^2 (2% of total model extent)
Area searched with >.75 probability	1.4 km^2 (8% of >.75 extent)
Area searched with > .9 probability	.9 km^2 (13% of > .9 extent)

Within the search extent, *L. ramosissimum* populations were found in seven distinct locations (Figure 16, page 18). Three of these populations were small (one to a few dozen plants) and were removed after mapping. Three populations are considerably larger and were not removed. The *L. ramosissimum* population at Plummer Creek Marsh has not been field mapped and was indentified through personal communication with Invasive *Spartina* Project staff member, Whitney Thornton. Some *L. ramosissimum* patches at Ideal Mash and Whale's Tail were previously mapped in 2008 but searches led to mapping of several new patches at each marsh.

All *L. ramosissimum* population locations within the model extent are shown against background imagery and relative to model results and search locations in Figures 17- 30. All

known *L. ramosissimum* populations in San Francisco Bay are shown in Figure 31, page 33 and population sizes in Table 5, page 34.

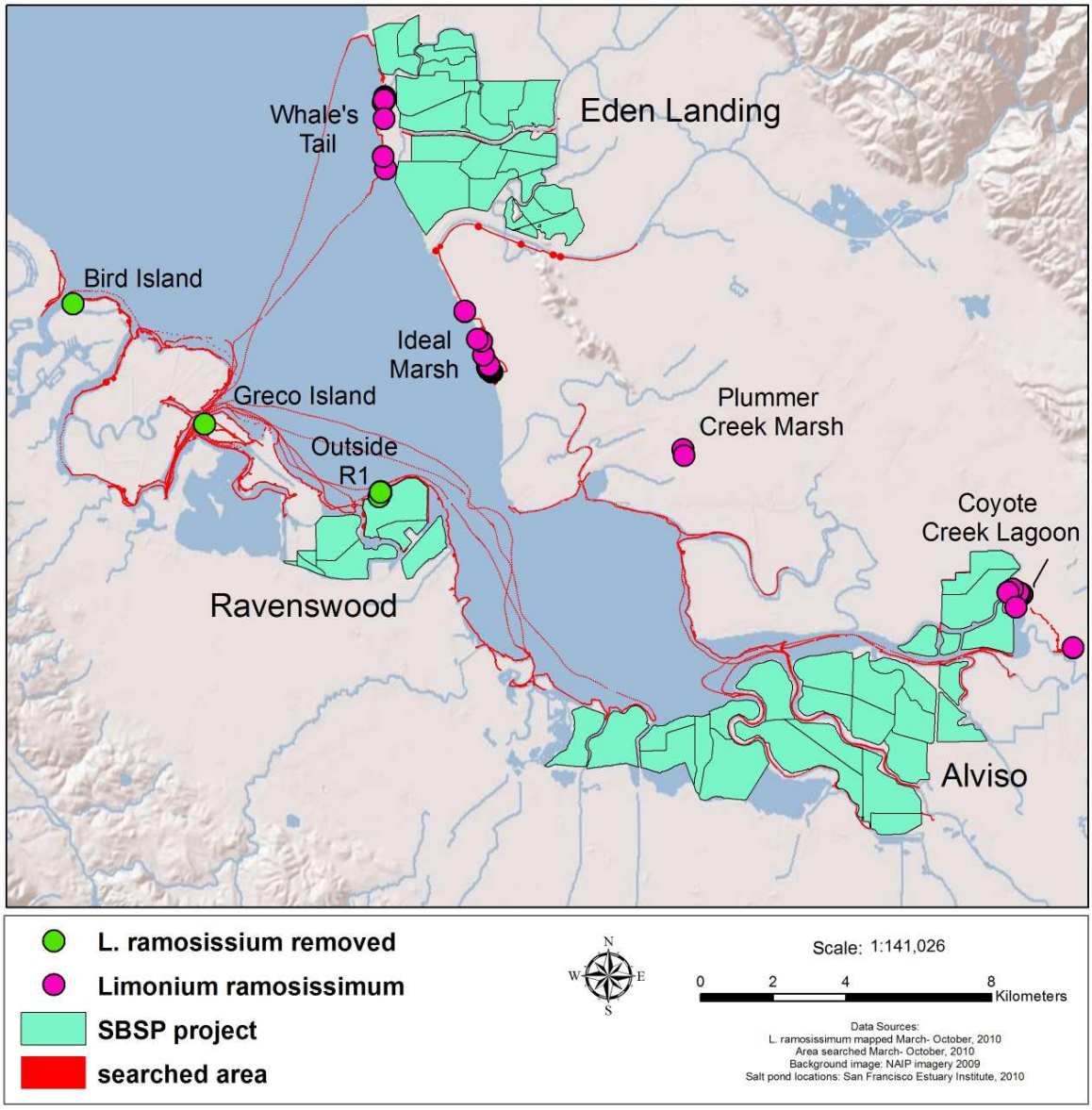

Figure 16: Map of *L. ramosissimum* presence and absence within the model extent.


San Mateo- Hayward Bridge
E11
E12
E10
E13
E14
E9
San Francisco Bay
Whale's
Tail
North
E8A
Old Alameda Creek
Whale's
Tail
South
Cargill
Mitigation
Marsh
E1
E2
Whale's Tail
Limonium ramosissimum
SBSP project
N
W
E
S
Scale: 1:18,828
0 250 500 1,000
Meters
Data Sources:
L. ramosissimum- mapped October, 2010
Background image: IKONOS imagery acq. July, 2010
SBSP pond locations: San Francisco Estuary Institute, 2010

Figure 17

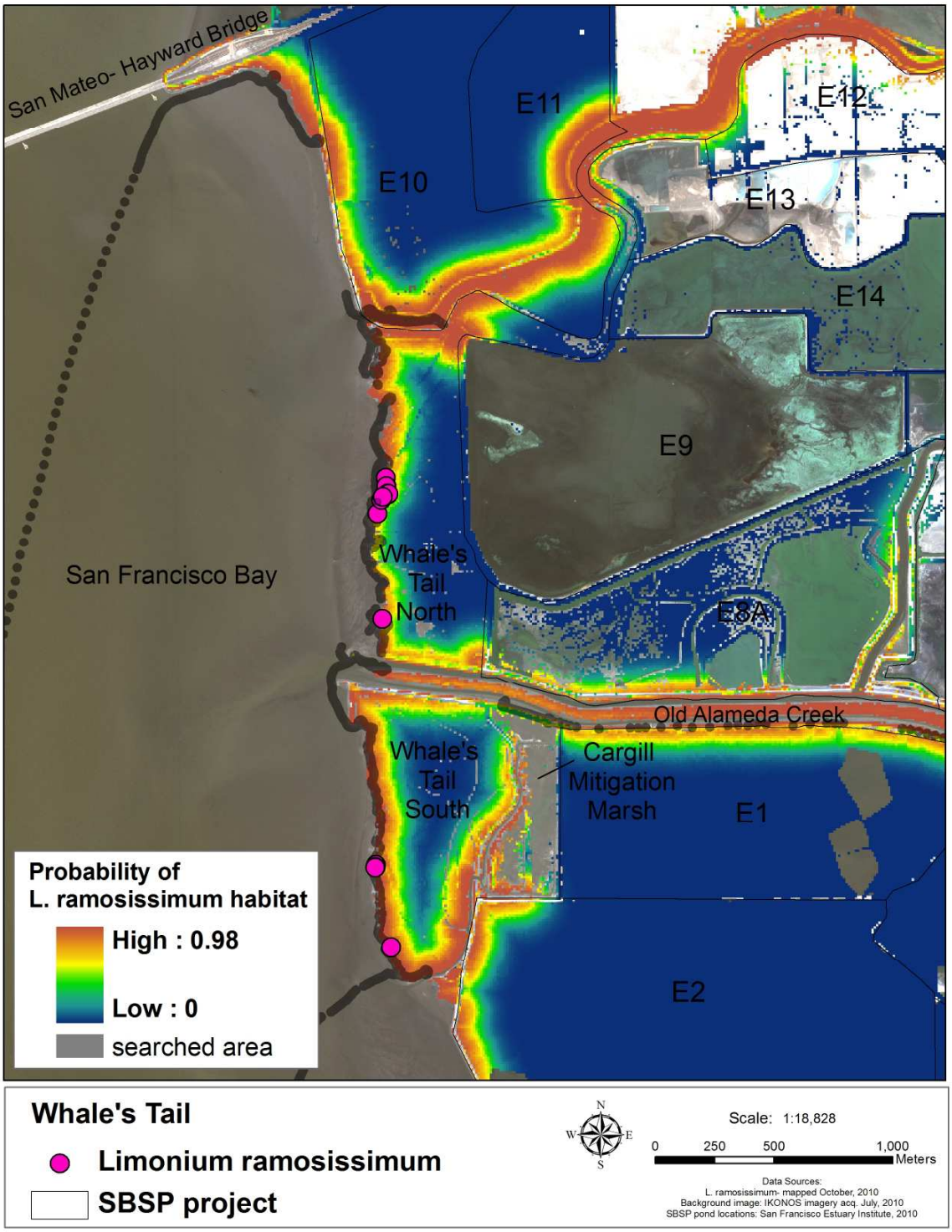
San Mateo- Hayward Bridge
E11
E12
E10
E13
E14
E9
San Francisco Bay
Whale's Tail North
E8A
Old Alameda Creek
Whale's Tail South
Cargill Mitigation Marsh
E1
E2
Probability of L. ramosissimum habitat
High : 0.98
Low : 0
searched area
Whale's Tail
Limonium ramosissimum
SBSP project
N
W
E
S
Scale: 1:18,828
0 250 500 1,000
Meters
Data Sources:
L. ramosissimum- mapped October, 2010
Background image: IKONOS imagery acq. July, 2010
SBSP pond locations: San Francisco Estuary Institute, 2010

Figure 18

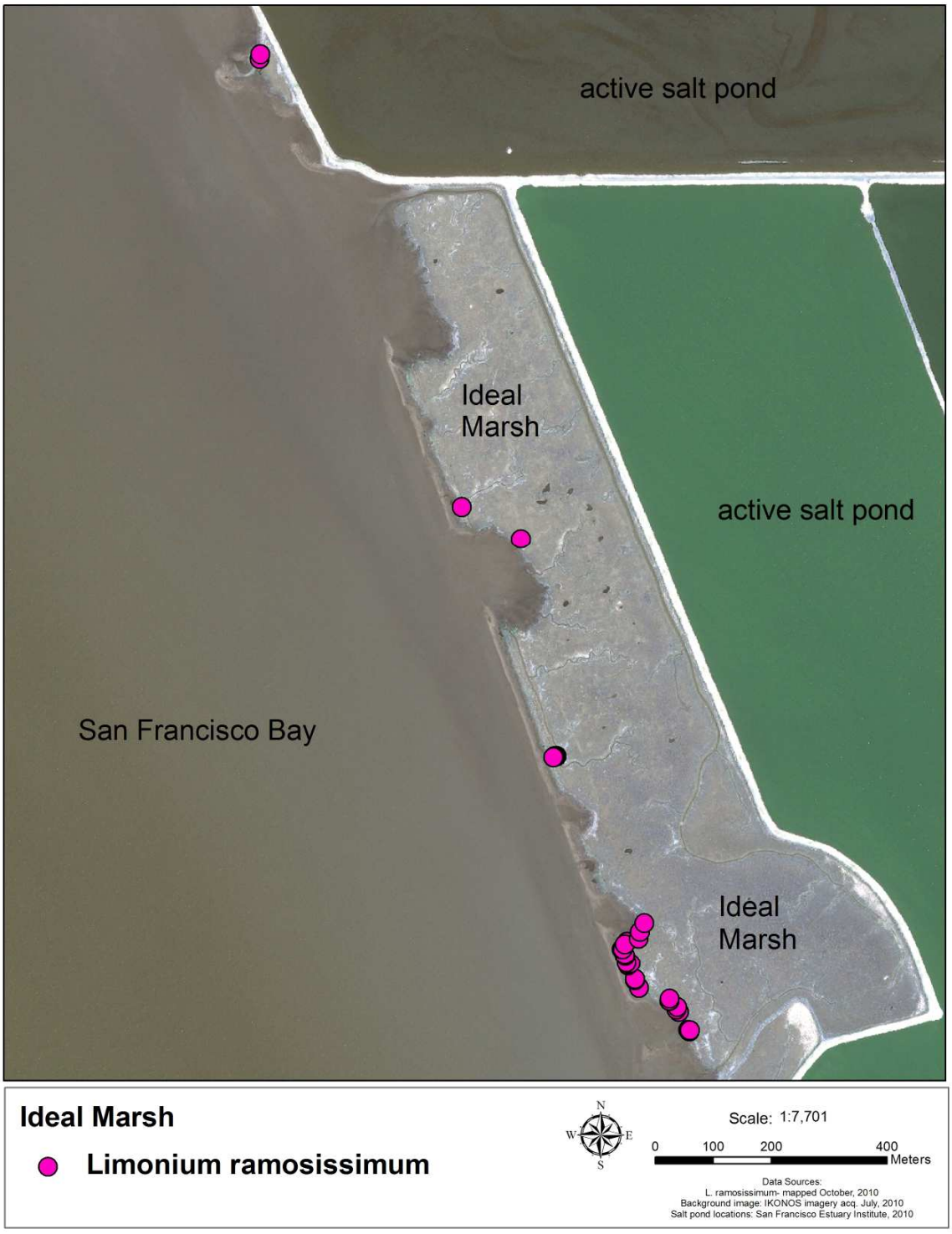
active salt pond
Ideal
Marsh
active salt pond
San Francisco Bay
Ideal
Marsh
Ideal Marsh
Limonium ramosissimum
N
W
E
S
Scale: 1:7,701
0
100
200
400
Meters
Data Sources:
L. ramosissimum- mapped October, 2010
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 19

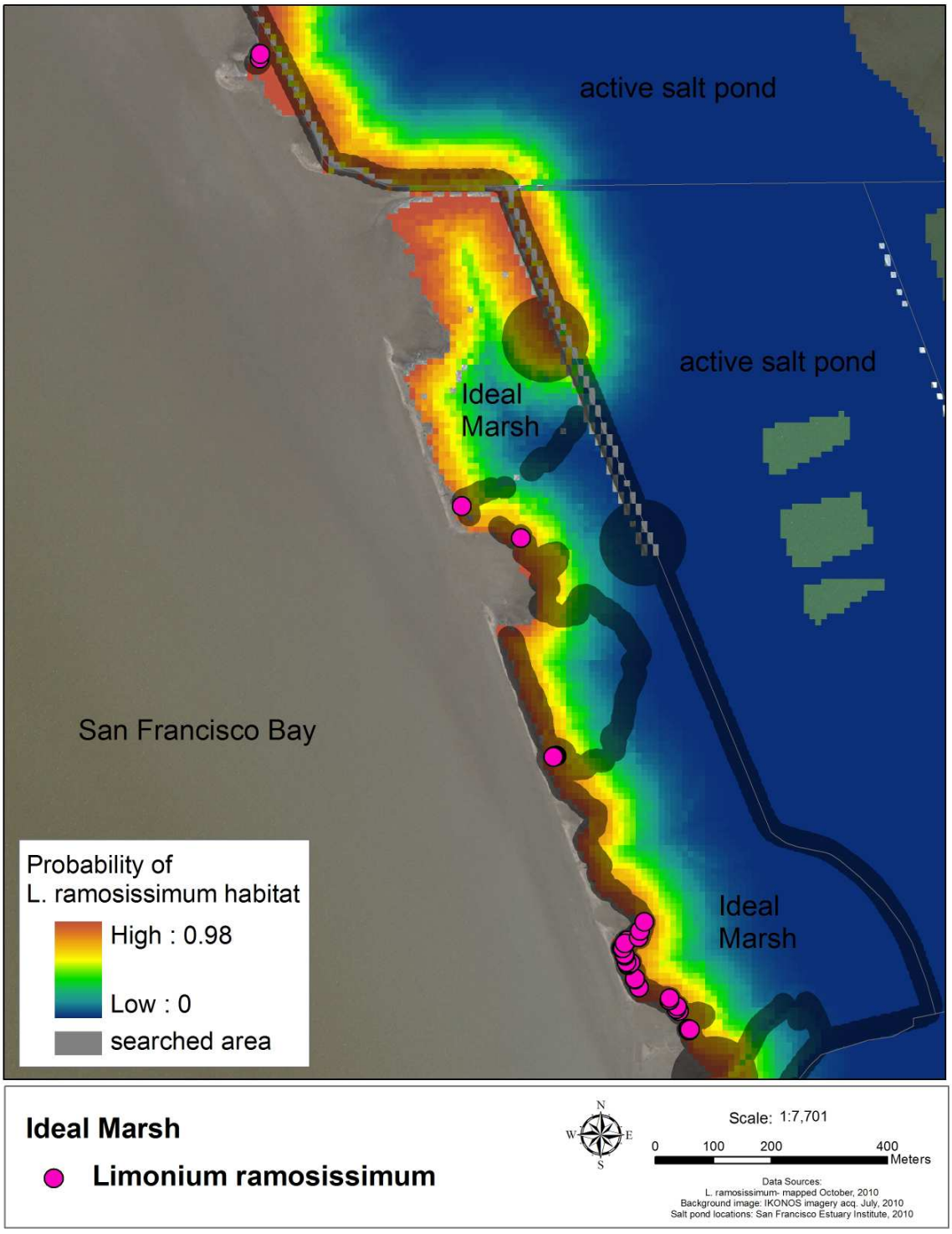
active salt pond
active salt pond
Ideal
Marsh
San Francisco Bay
Ideal
Marsh
Probability of
L. ramosissimum habitat
High : 0.98
Low : 0
searched area
Ideal Marsh
Limonium ramosissimum
N
W
E
S
Scale: 1:7,701
0
100
200
400
Meters
Data Sources:
L. ramosissimum- mapped October, 2010
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 20

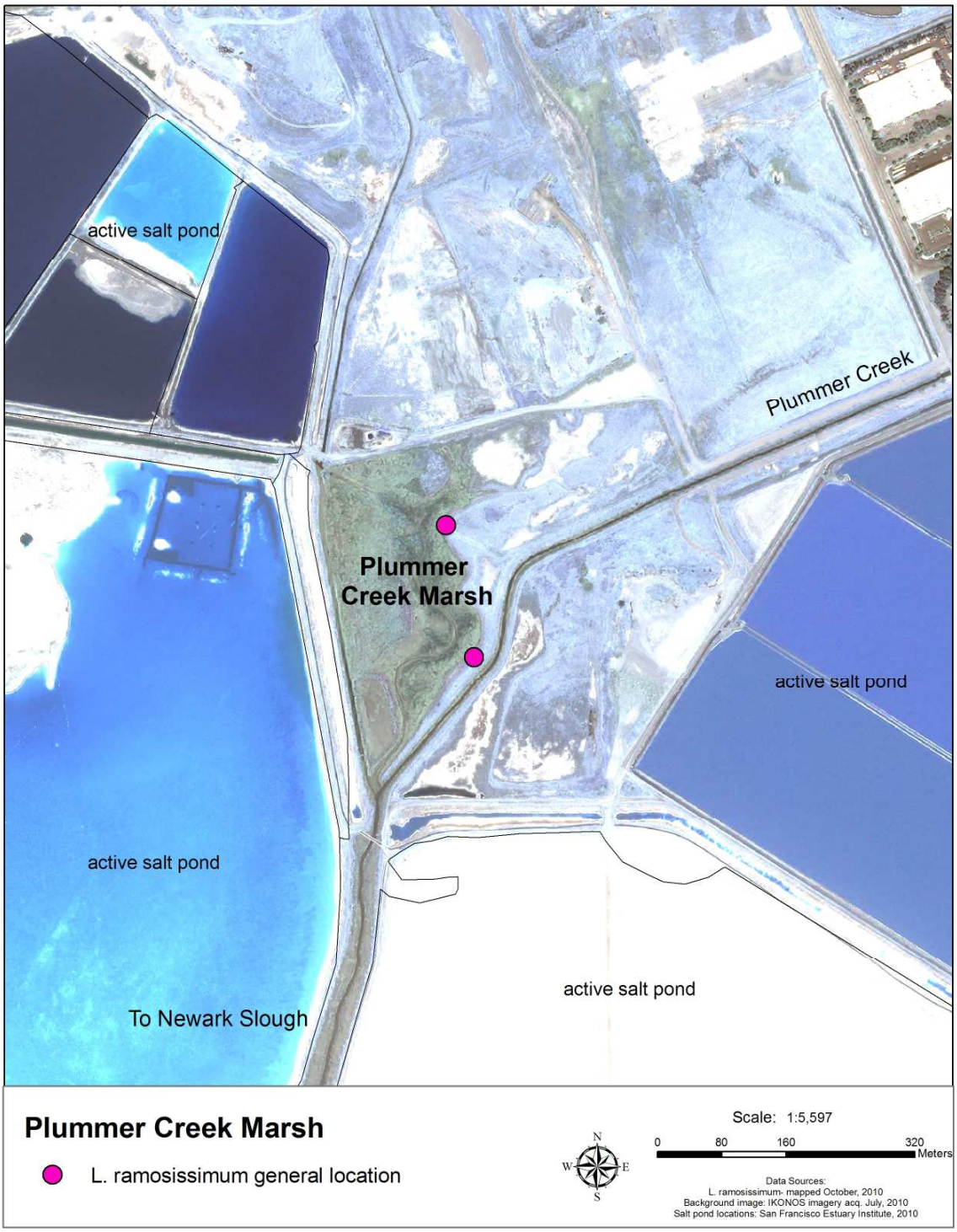
active salt pond
Plummer Creek
Plummer
Creek Marsh
active salt pond
active salt pond
active salt pond
To Newark Slough
Plummer Creek Marsh
L. ramosissimum general location
N
W
E
S
Scale: 1:5,597
0
80
160
320
Meters
Data Sources:
L. ramosissimum- mapped October, 2010
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 21

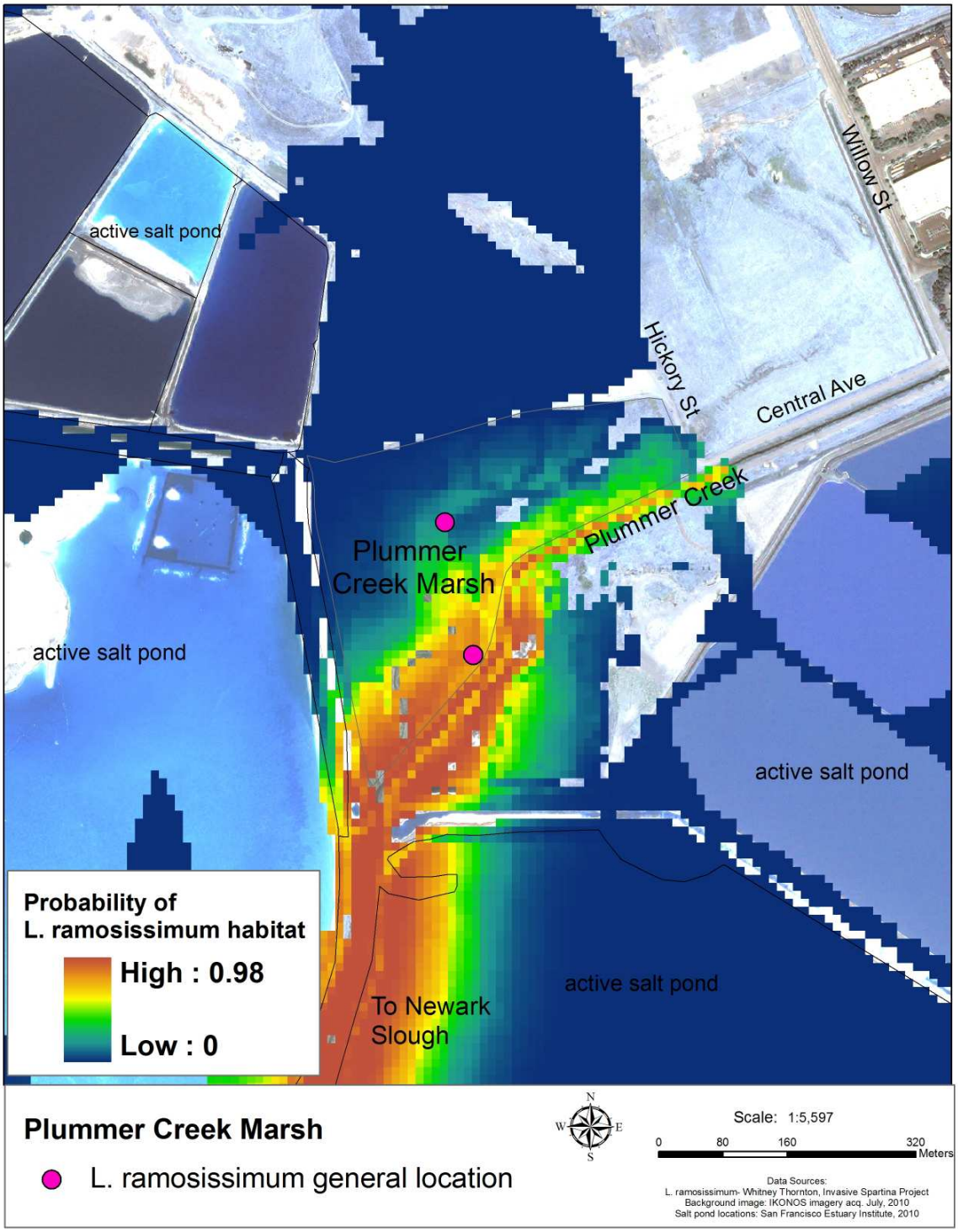
active salt pond
Willow St
Hickory St
Central Ave
Plummer Creek
Plummer
Creek Marsh
active salt pond
active salt pond
Probability of
L. ramosissimum habitat
High : 0.98
Low : 0
To Newark
Slough
active salt pond
Plummer Creek Marsh
L. ramosissimum general location
N
W
E
S
Scale: 1:5,597
0
80
160
320
Meters
Data Sources:
L. ramosissimum- Whitney Thornton, Invasive Spartina Project
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 22


A23
Gateway Blvd
Coyote Creek
Lagoon
A19
Coyote Creek
Coyote Creek
Newby Landfill
Coyote Creek Lagoon
Limonium ramosissimum
N
W
E
S
Scale: 1:7,995
0
100
200
400
Meters
Data Sources:
L. ramosissimum- mapped October, 2010
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 23

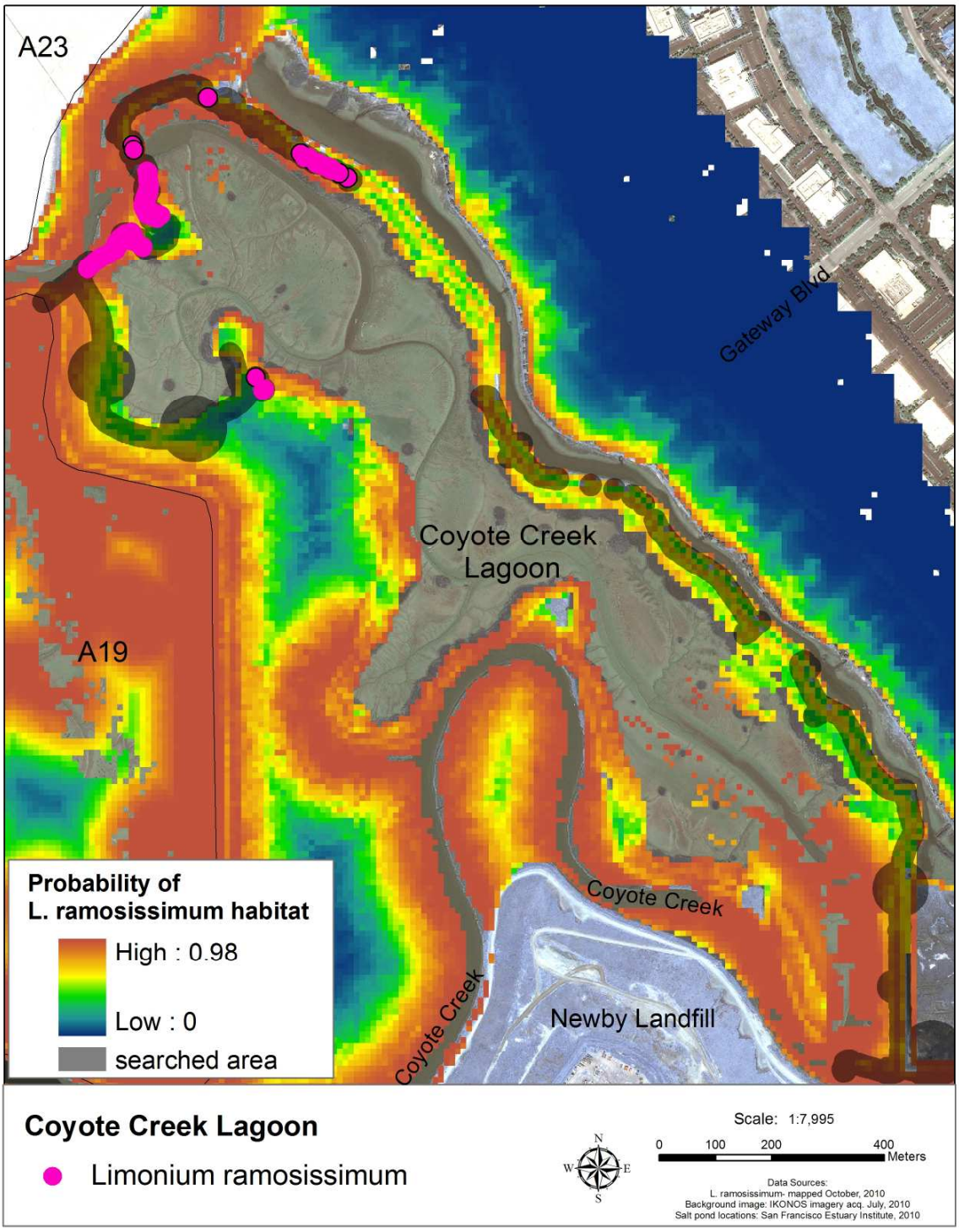
A23
A19
Gateway Blvd
Coyote Creek
Lagoon
Coyote Creek
Coyote Creek
Newby Landfill
Probability of
L. ramosissimum habitat
High : 0.98
Low : 0
searched area
Coyote Creek Lagoon
Limonium ramosissimum
Scale: 1:7,995
0 100 200 400
Meters
N
W
E
S
Data Sources:
L. ramosissimum- mapped October, 2010
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 24

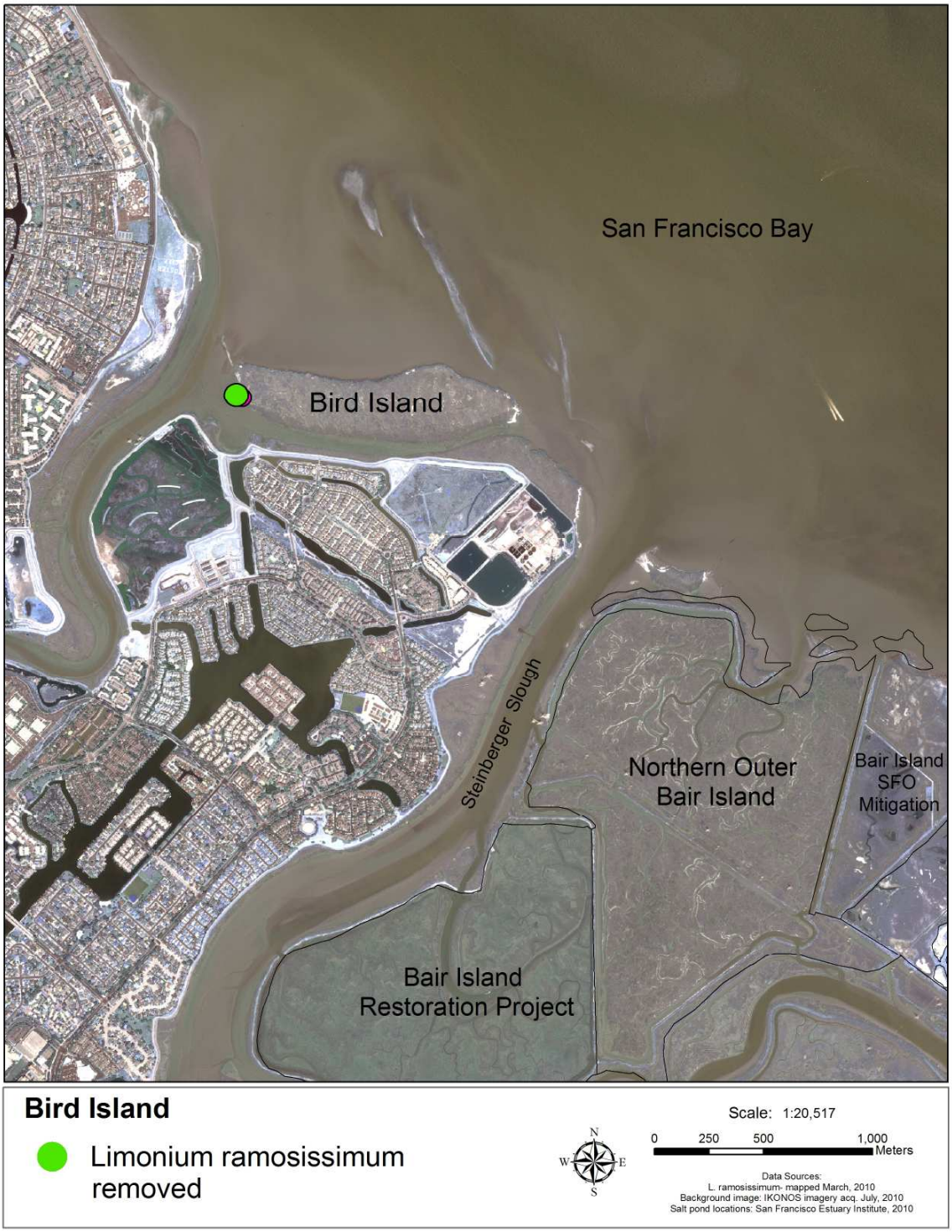
San Francisco Bay
Bird Island
Steinberger Slough
Northern Outer
Bair Island
Bair Island
SFO
Mitigation
Bair Island
Restoration Project
Bird Island
Limonium ramosissimum
removed
Scale: 1:20,517
0 250 500 1,000
Meters
N
W
E
S
Data Sources:
L. ramosissimum- mapped March, 2010
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 25

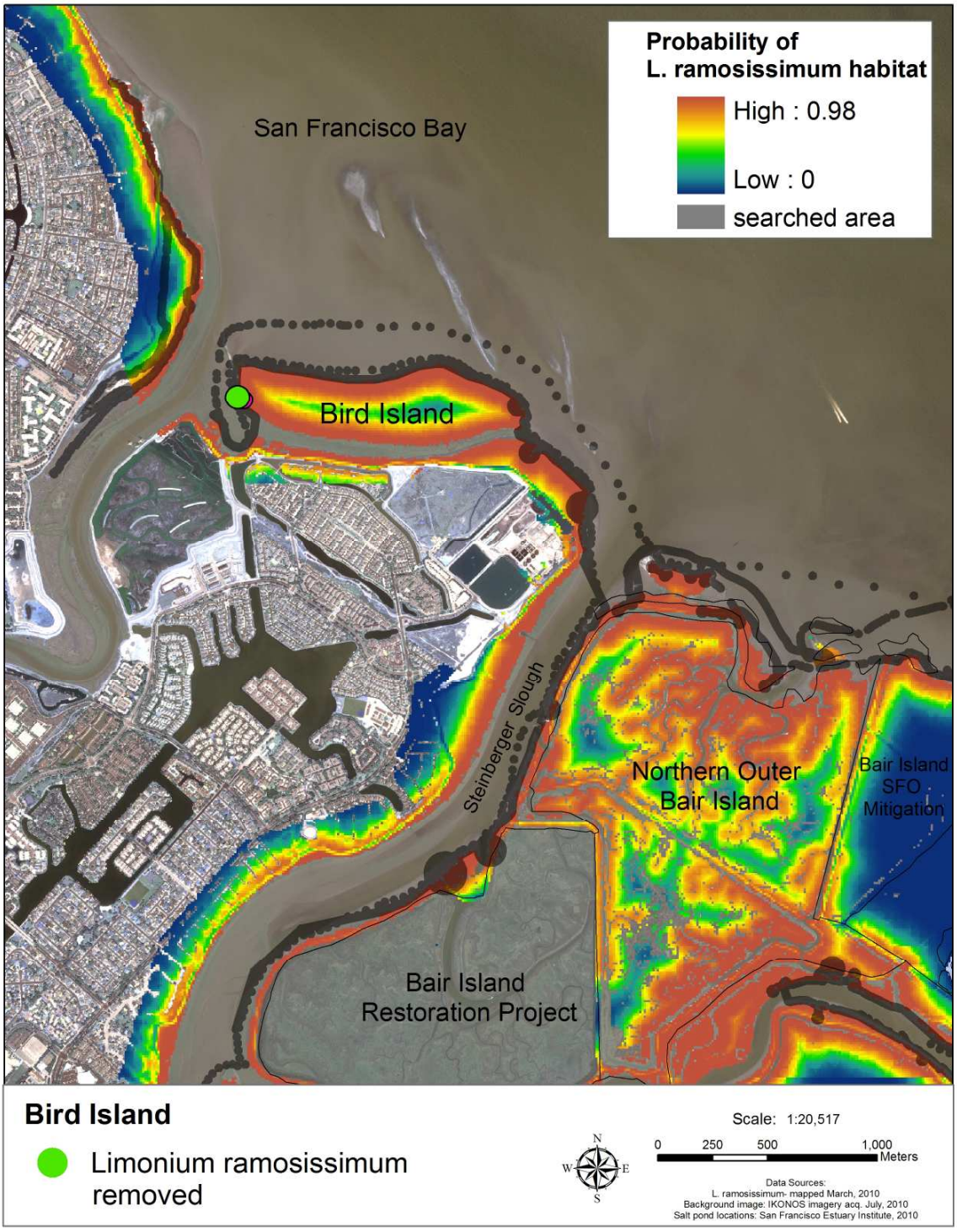
Probability of
L. ramosissimum habitat
High : 0.98
Low : 0
searched area
San Francisco Bay
Bird Island
Steinberger Slough
Northern Outer
Bair Island
Bair Island
SFO
Mitigation
Bair Island
Restoration Project
Bird Island
Limonium ramosissimum
removed
Scale: 1:20,517
0 250 500 1,000
Meters
N
W
E
S
Data Sources:
L. ramosissimum- mapped March, 2010
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 26


Northern
Outer
Bair Island
Bair Island
SFO
Mitigation
San Francisco Bay
Greco
Island
Pacific
Shores
Center
Westpoint Slough
salt pond
Greco
Island
salt pond
salt pond
Greco Island
Limonium ramosissimum
removed
N
W
E
S
Scale: 1:20,517
0 250 500 1,000
Meters
Data Sources:
L. ramosissimum- mapped March, 2010
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 27

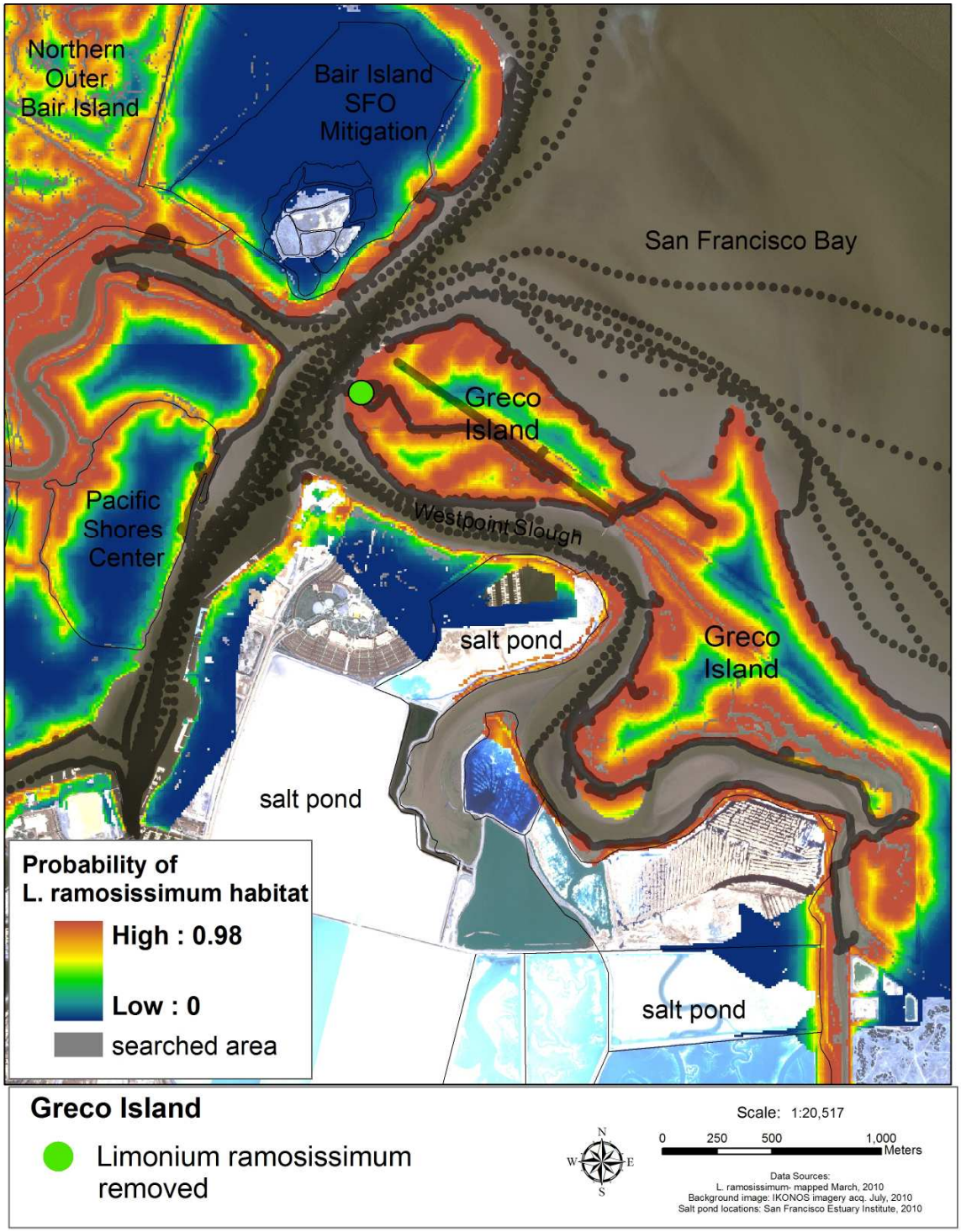
Northern Outer Bair Island
Bair Island SFO Mitigation
San Francisco Bay
Greco Island
Pacific Shores Center
Westpoint Slough
salt pond
Greco Island
salt pond
salt pond
Probability of L. ramosissimum habitat
High : 0.98
Low : 0
searched area
Greco Island
Limonium ramosissimum removed
Scale: 1:20,517
0 250 500 1,000 Meters
N
W
E
S
Data Sources:
L. ramosissimum- mapped March, 2010
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 28


San Francisco Bay
R1
Outside R1
Limonium ramosissimum removed
N
W
E
S
Scale: 1:2,346
0
30
60
120
Meters
Data Sources:
L. ramosissimum- mapped March, 2010
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 29

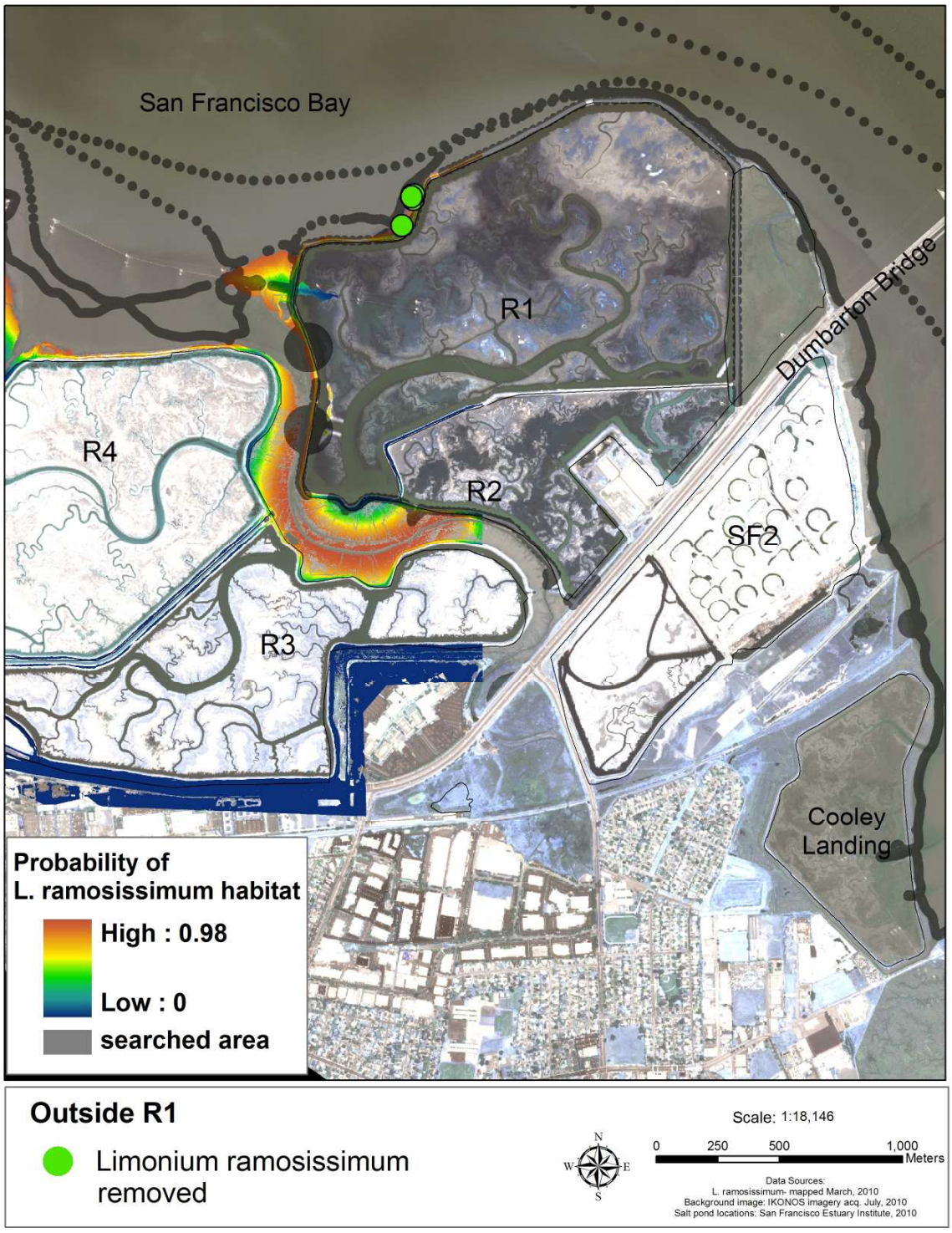
San Francisco Bay
R1
Dumbarton Bridge
R4
R2
SF2
R3
Cooley Landing
Probability of L. ramosissimum habitat
High : 0.98
Low : 0
searched area
Outside R1
Limonium ramosissimum removed
Scale: 1:18,146
0 250 500 1,000
Meters
N
W
E
S
Data Sources:
L. ramosissimum- mapped March, 2010
Background image: IKONOS imagery acq. July, 2010
Salt pond locations: San Francisco Estuary Institute, 2010

Figure 30

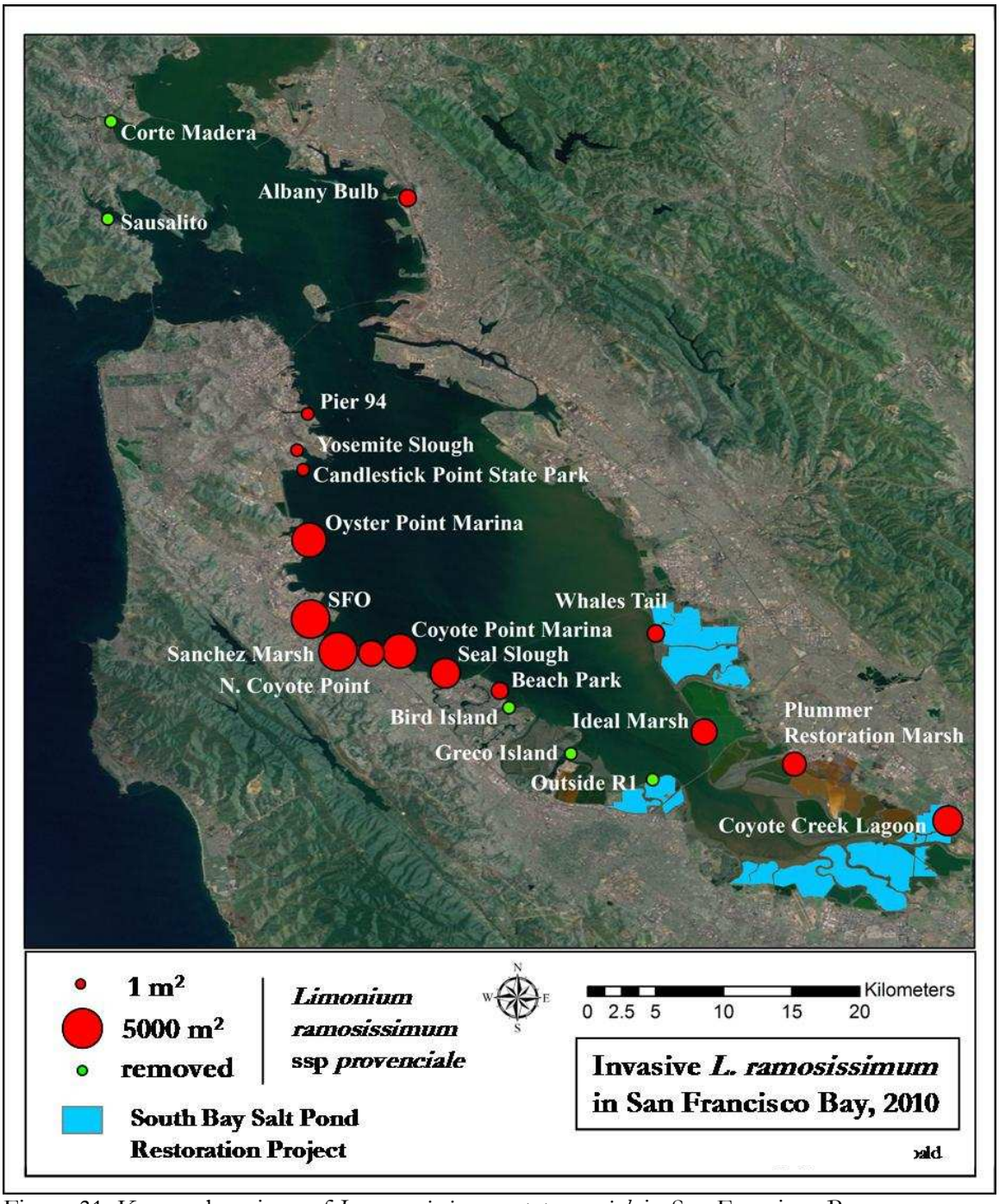

Figure 31: Known locations of *L. ramosissimum ssp provenciale* in San Francisco Bay.

Table 5: Relative sizes (determined by GPS mapping) of *L. ramosissimum ssp provenciale* populations in San Francisco Bay.

Location	**Area (m2)**
Sausalito	1
Corte Madera	1
Yosemite Slough	2
Candlestick Point State Park	1
Pier 94	1
Oyster Point Marina	1592
SFO	3859
Sanchez Marsh	4361
N. Coyote Point	449
Coyote Point Marina	2300
Albany Bulb	32
Seal Slough	519
Beach Park	13
Bird Island	1
Greco Island	1
Outside R1	2
Whales Tail	36
Ideal Marsh	239
Coyote Creek Lagoon	1117
Plummer Restoration Marsh	unknown

Model accuracy assessment

To determine how accurately the model assigned high probability of *L. ramosissimum* habitat to locations where *L. ramosissimum* was found during the survey, model probability values were extracted from all *L. ramosissimum* point and patch locations not used for model generation at one meter resolution. Results (Figure 32) show that the model assigned an average probability of .77 to locations the plant was found versus an average probability of .24 in random locations.

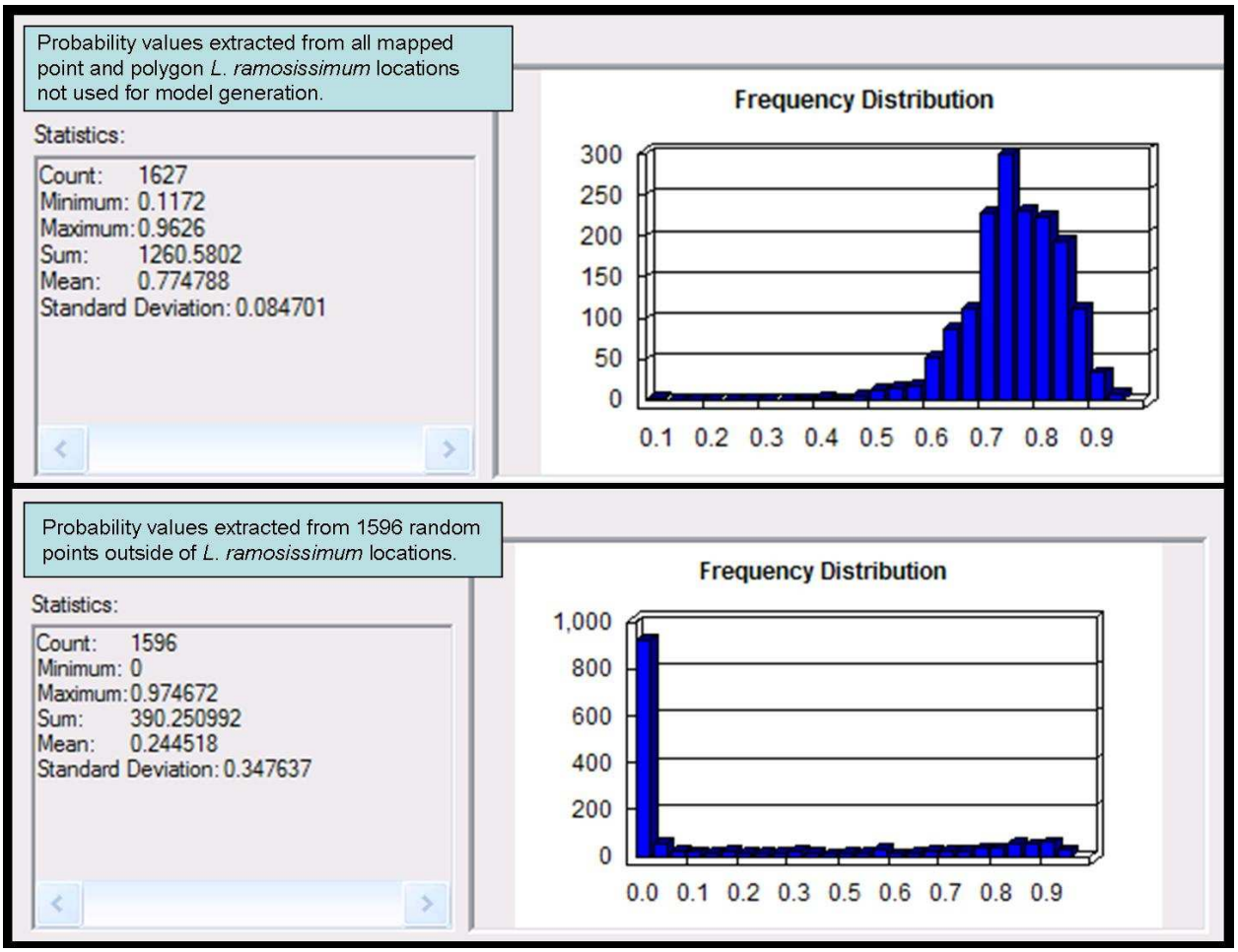

Figure 32: Cells in the model where *L. ramosissimum* was mapped had an average probability of .77 versus an average probability of .24 in random locations.

The second model assessment method, a Cohen's Kappa statistic, quantified the accuracy which the model assigned "high probability of *L. ramosissimum* habitat" versus "low probability of *L. ramosissimum* habitat". The required confusion matrix and calculations are shown below:

Confusion matrix:

	0 -.3 probability	.7 - .97 probability	Total
Not *L. ramosissimum* habitat	19 cells	2 cells	25 cells
L. ramosissimum habitat	6 cells	23 cells	25 cells
Total	25 cells	25 cells	50 cells

$$\text{observed agreement } (P_o) = \frac{\text{sum of diagonal}}{\text{sum of matrix}} = \frac{19 + 23}{19 + 6 + 23 + 2} = .84$$

$$\text{chance agreement } (P_c) = P_1P_1 + P_2P_2 = \left(\frac{25}{50}\right)\left(\frac{25}{50}\right) + \left(\frac{25}{50}\right)\left(\frac{25}{50}\right) = .5$$

$$\text{Kappa statistic} = \frac{P_o - P_c}{1 - P_c} = \frac{.84 - .5}{1 - .5} = \mathbf{.68}$$

Discussion:

The model:

The relative size of the coefficients associated with each significant predictor variable indicates the degree of effect the variable has on the probability of *L. ramosissimum* habitat. Larger coefficients change the probability of *L. ramosissimum* habitat more rapidly assuming the predictor variables have equal ranges. In this model, all coefficients were negative, meaning as predictor variable values increase, the probability of *L. ramosissimum* habitat decreases. However, this leads to different interpretations for each independent variable.

In the case of the distance to *Spartina* variable, which had the largest negative coefficient, probability of *L. ramosissimum* habitat decreases rapidly as the variable value increases at closer distances to *Spartina* locations, indicating *L. ramosissimum* is negatively correlated with *Spartina* locations, the opposite of the hypothesized relationship. This result indicates that dispersal events are probably not driving distribution patterns similarly for both species and this variable could be left out of future model iterations.

The next largest coefficient, elevation relative to high tides, indicates that the probability of *L. ramosissimum* habitat increases with lower elevations relative to average tides. This means that while probability of *L. ramosissimum* habitat increases as cell values decrease from upland elevations to high marsh elevations (as hypothesized), probabilities continue to increase at negative, low marsh elevations, which runs counter to *L. ramosissimum*'s ecology. This mixed result may have the effect that the model is more likely to over predict habitat at low marsh elevations, than at upland elevations.

The two smallest coefficients were distance to water's edge and distance to high marsh. In both cases, variable values increase with greater distances from these features, lowering probability of *L. ramosissimum* habitat further from water's edge and high marsh, as predicted. While the sizes of these coefficients are small, visual interpretation of model results suggest that distance to water's edge is a primary variable driving *L. ramosissimum* habitat probability. This suggests that the size of coefficients alone does not determine the importance of an

independent variable, but also the range of the variable itself. Distance to *Spartina* values range from 0 to 1 while distance to water's edge ranges from 0 to approx 1000 (meters from water). The small coefficient associated with distance to water's edge may strongly drive *L. ramosissimum* habitat probabilities because of this larger range.

The finding that distance to water's edge appears to be strong driver of *L. ramosissimum* habitat likelihood is supported by field observations. At Ideal Marsh, for example, *L. ramosissimum* is found along the slight rise in marsh elevation at the bayward edge of the marsh (Figure 33). Marsh elevations decrease landward from the water's edge, increasing marsh inundation and limiting *L. ramosissimum* spread into the marsh plain. At Coyote Creek Lagoon, distance to water's edge is also an effective predictor by itself, though for different reasons. High marsh habitat in this case is located close to the water's edge, but levies preclude *L. ramosissimum* growing further shoreward away from the water's edge.

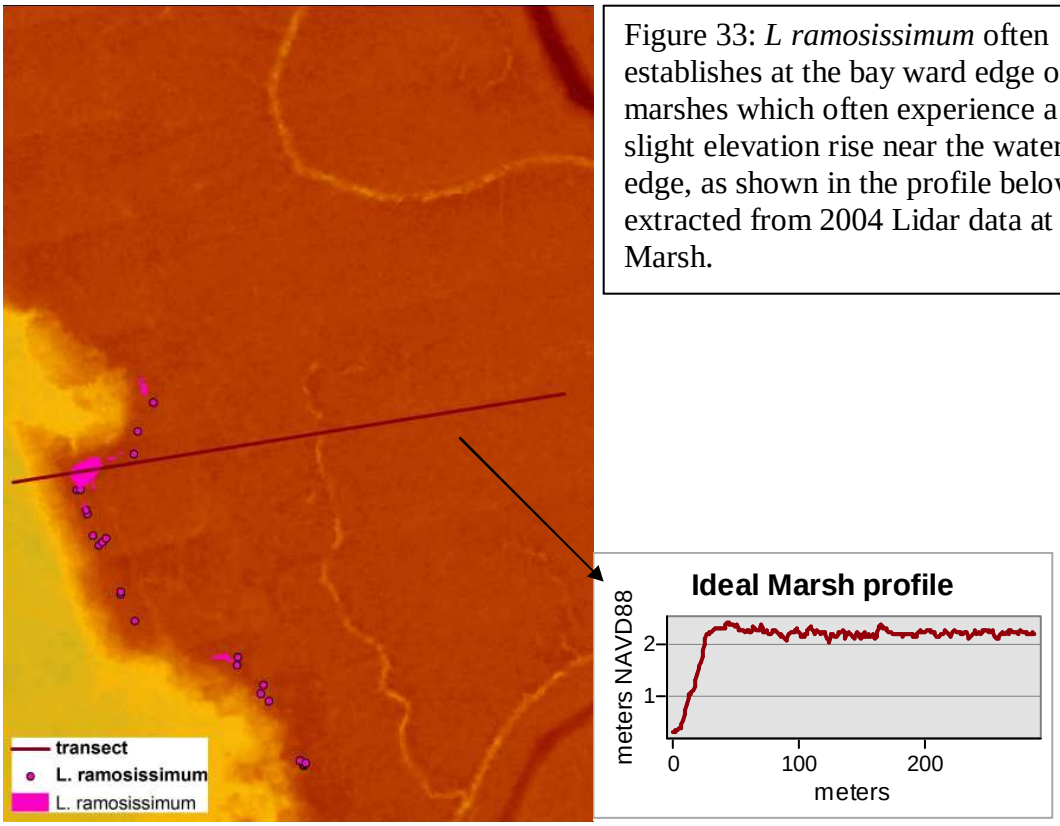

Figure 33: *L ramosissimum* often establishes at the bay ward edge of marshes which often experience a slight elevation rise near the water's edge, as shown in the profile below extracted from 2004 Lidar data at Ideal Marsh.

The overall test of how well the model fits the data, the Hosmer and Lemeshow goodness-of-fit test, indicates the model does not fit the data at an acceptable level (statistic < .05). This may be a function of the fact that the invasion has not reached equilibrium. As a result, many locations outside of current *L. ramosissimum* population locations sampled for analysis as absence points are actually suitable for invasion. In those locations, the variable values are similar to invaded areas leading to a poor overall goodness of fit. Rerunning this model with a larger sample size, and therefore power to not retain what may be a false null hypothesis of no model effect, could improve the model's overall measure of fit.

One limitation of this model is the accuracy of the elevation relative to the average high tide raster. Lidar elevation has known inaccuracies in brackish marshes where dense vegetation falsely raises marsh elevations. Also, because this model relied on a Vdatum conversion between MLLW and NAVD88 elevation, values for the average high tides layer were falsely increased in the southern reaches of the model. These inaccuracies likely result in model biases.

In addition, two key factors missing from this model that may improve prediction are a better disturbance variable and, conversely, native plant cover. Assuming *L. ramosissimum* seeds can arrive at any location with an equal chance, establishment would be effected by the degree of vegetative cover. In a future iteration of this model, results of remote sensing could be used to characterize disturbance and vegetation cover which may help focus the model's ability to predict likely invasion locations. Records of marsh restoration and other disturbances might also prove to be useful for improving prediction.

The survey:
18 days of surveys resulted in only 8% of the model habitat greater than .75 probabilities being searched. This highlights the difficulty associated with searching a large area where a broad habitat class is being modeled. Additionally, because of difficulty ground truthing many marsh areas due to weather and seasonal access restrictions, random sampling of the model's results was not carried out. Instead, surveys focused on surveying high probability areas which were accessible. As a result, estimates of how many more *L. ramosissimum* populations may exist in the study extent are difficult to project.

Accuracy assessment:
Populations found during mapping provide an independent means of assessing the accuracy with which the model assigned cells later found to contain *L. ramosissimum.* Extracting model values from all 1627 cells with *L. ramosissimum* found the average probability was .77 with a standard deviation of .08, versus 1596 random cell locations which had a probability of .24 and standard deviation of .34. This indicates the model accurately assigned high probability of *L. ramosissimum* habitat.

The Kappa statistic of .68 accuracy of high probability versus low probability model assignment supports this finding. The kappa statistic takes into consideration probability that a given cell will be accurately assigned the appropriate classification by chance.

Visually assessing the model, model accuracy may be compromised by the tendency to exclude the interior of marshes as possible *L. ramosissimum* habitat. While no *L. ramosissimum* populations to date have been found in the interior of marshes, this possibility should not be excluded, particularly if high, disturbed habitat exists within the marshes interior. Also, because the model depends on the distance to bay's edge and because channels less than 5 m in width were not digitized, marshes located along small channels may have higher probability of *L. ramosissimum* habitat then the model currently shows.

Additional *L. ramosissimum* populations in the study area
It is likely that additional *L. ramosissimum* populations exist within the study extent. Edges of marshes along the bay and major sloughs have largely been searched, and undisturbed marsh

plains where seedling recruitment is likely limited by dense marsh canopy probably are at low risk of new invasive populations.

However, interior high marshes, such as Plummer Creek Marsh, where seed dispersal may come from the local watersheds rather than distributed by Bay water, are likely locations where large populations may yet be undiscovered. Many of these "inland" high marshes are difficult to access due to property access restrictions and additional effort should be undertaken to identify and investigate these areas.

L. ramosissimum is a halophyte, and it is also a xerophyte. Because of this, invasions are likely to be most severe in areas where conditions are dry, though receiving some inundation, and where disturbances have been large.

Model's utility for early detection:
While species distribution models rely on the assumption that a species range has reached equilibrium, this work shows that logistic regression modeling is a useful tool for early detection of invasive species. The primary utility of this model is to focus and limit searches, which accuracy assessments show is justified in this study. Ideally, as searches progress and populations are discovered, new presence/absence data can be added to improve the model. Additionally, predictor variables can be refined and new ones created to improve model results. Creating a database of predictor variables facilitates developing models for other species of management concern.

Literature Cited:

Boorman, L.A. 1971. Studies in Salt Marsh Ecology with Special Reference to the Genus Limonium. *Journal of Ecology*, 59:1, 103-120.

Collingham, Y. C., R. A. Wadsworth, B. Huntley, and P. E. Hulme. 2000. Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology*. 37(Suppl. 1):13–27.

Collins, J.N., 2002. Invasion of San Francisco Bay Smooth Cordgrass, Spartina alterniflora: A Forecast of Geomorphic Effects On the Intertidal Zone. San Francico Estuary Institute,

Chapman, V. J. 1934. The ecology of Scolt Head Island. In Scolt Head Island. Ed. J. A. Steers. Cambridge: W. Heffer & Sons, Ltd. 234 pp.

Foxgrover, A. C., and B. E. Jaffe, 2005. South San Francisco Bay 2004 Topographic Lidar Survey: Data Overview and Preliminary Quality Assessment. U.S. Geological Survey Pacific Science Center, Santa Cruz, CA. Open-File Report 2005–1284

Franklin, J., 1995. Predictive vegetation mapping: geographical modeling of biospatial patterns in relation to environmental gradients. Prog. Phys. Geogr. 19, 474–499.

Garson, G. D. 2006. Logistic Regression. http://www2.chass.ncsu.edu/garson/PA765/logistic.htm. Accessed: February 25, 2010.

Guisan, A. and N. Zimmerman, 2000. Predictive habitat distribution models in Ecology. Ecological modeling, 135: 147-186.

Guisan, A., Weiss, S.B., Weiss, A.D., 1999. GLM versus CCA spatial modeling of plant species distribution. Plant Ecol.
143, 107–122.

Hickey, D., and E. Bruce. Examining Tidal Inundation and Salt Marsh Vegetation Distribution Patterns using Spatial Analysis (Botany Bay, Australia). *Journal of Coastal Research*, 26: 94-102.

Holcomb, T, Stohlgren, T.J., and C. Jarnevich. 2007. Invasive species management and research using GIS. Managing Vertebrate Invasive Species: *Proceedings of an International Symposium* (G. W. Witmer, W.C. Pitt, K.A. Fagerstone, Eds). USDA/APHIS/WS, National Wildlife Research Center, Fort Collins, CO.

Hsieh, F.Y., Bloch, D. A. and M. D. Larsen, 1998. A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, 17: 1623-1634.

Kana, T.W., Baca, B.J. and M.L. Williams. 1988. Charleston case study. In Greenhouse Effect, Sea Level Rise, and Coastal Wetlands, J.G. Titus (ed.). U.S. EPA, Washington, DC.

Kolar, S, C, and D. M. Lodge. 2001. Progress in invasion biology: predicting invaders. - *TRENDS in Ecology and Evolution.* Vol. 16:4

National Oceanic and Atmospheric Administration, National Ocean Service, and Center for Operational Oceanographic Products and Services, 2003. Computational Technique for Tidal Datums Handbook: NOAA Special Publication NOS CO-OPS 2. U.S. Department of Commerce. Silver Spring, Maryland.

Nielsen, C., P. Hartvig, and J. Kollmann. 2008. Predicting the distribution of the invasive alien Heracleum mantegazzianum at two different spatial scales. *Diversity and Distributions,* 14:307–317.

Peduzzi, P., J. Concato, E. Kemper, T. R. Holford, and A. Feinstein, 1996. A simulation of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 99: 1373-1379.

Pedhazur, E. J., 1997. *Multiple regression in behavioral research, 3rd ed.* Orlando, FL: Harcourt Brace.

Silverman, B.W., 1986 *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall

tidesandcurrents.noaa.gov/tides09/tab2wc1a.html. Accessed April, 2010.

Vaclavik, T., and R. K. Meentemeyer, 2009. Invasive species distribution modeling

(iSDM): Are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modeling*, 220, 3248-3258.

vdatum.noaa.gov/. Accessed February, 2010.