# Distributional Patterns of Perennial Pepperweed, *Lepidium latifolium*, in the San Francisco Bay

Melanie Vanderhoof and Chris Rogers,
Environmental Science Associates

Photo: http://www.prbo.org/cms/docs/wetlands/lepidium04.pdf

# Research Questions

- Where does *L. latifolium* occur in the San Francisco Bay area?

- Can its distribution pattern be explained and predicted using environmental variables?

# *Lepidium latifolium*

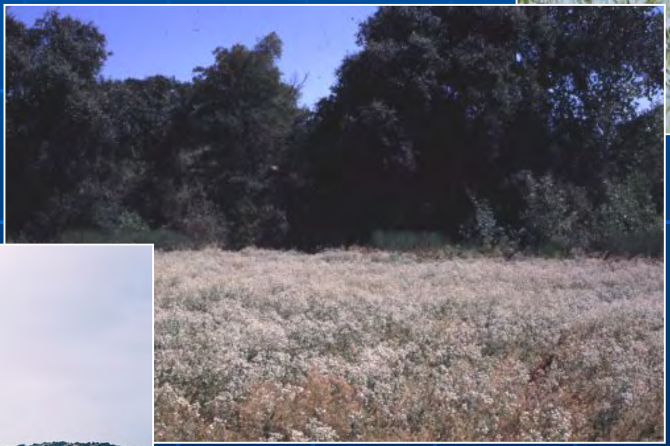
- A perennial weed
- Member of mustard family.
- Native to Eurasia.
- First recorded in CA in 1936, possibly from contaminated agricultural seed.
- Forms dense colonies or patches that grow 2-3 ft in height.


http://www.weedmapper.org/lela2picture.htm

# Invaded Habitats

Agricultural areas

Vernal Pools



Riparian Zones

Marshes

Photos: ESA

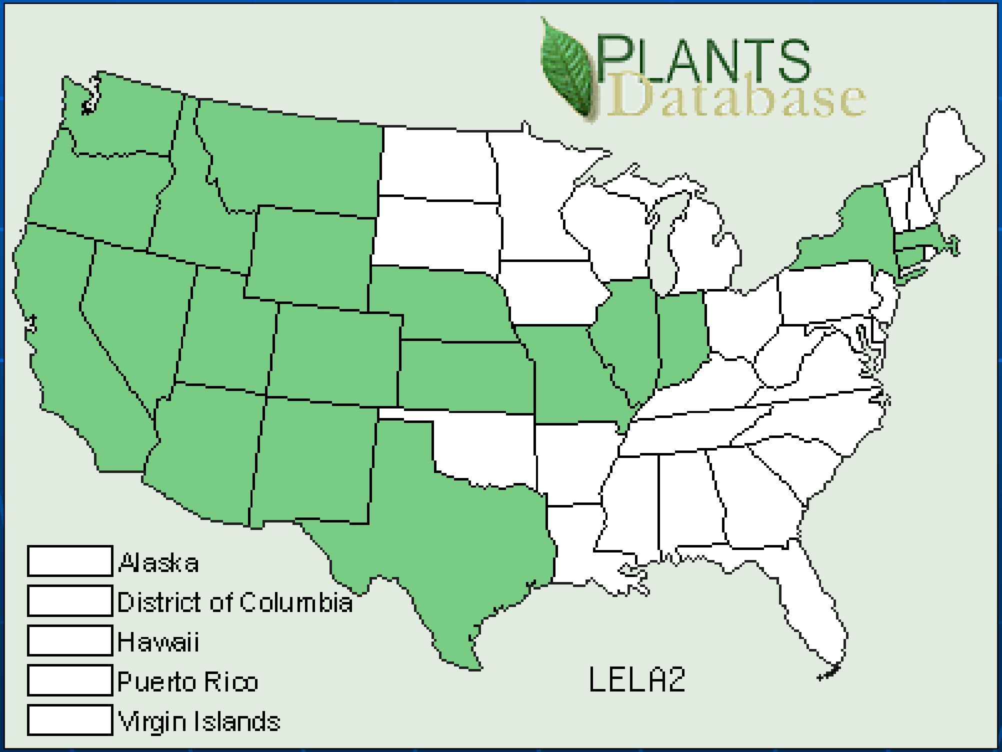
# Objectives

1. Map the distribution of *Lepidium latifolium* along the shoreline of the San Francisco Bay.

2. Develop a predictive model that identifies high risk areas in the larger San Francisco Bay Area based on environmental variables.

# Mapping Methods

- Limited mapping of *L. latifolium* has occurred in the bay area (Grossinger *et al.* 1998; May 1995)



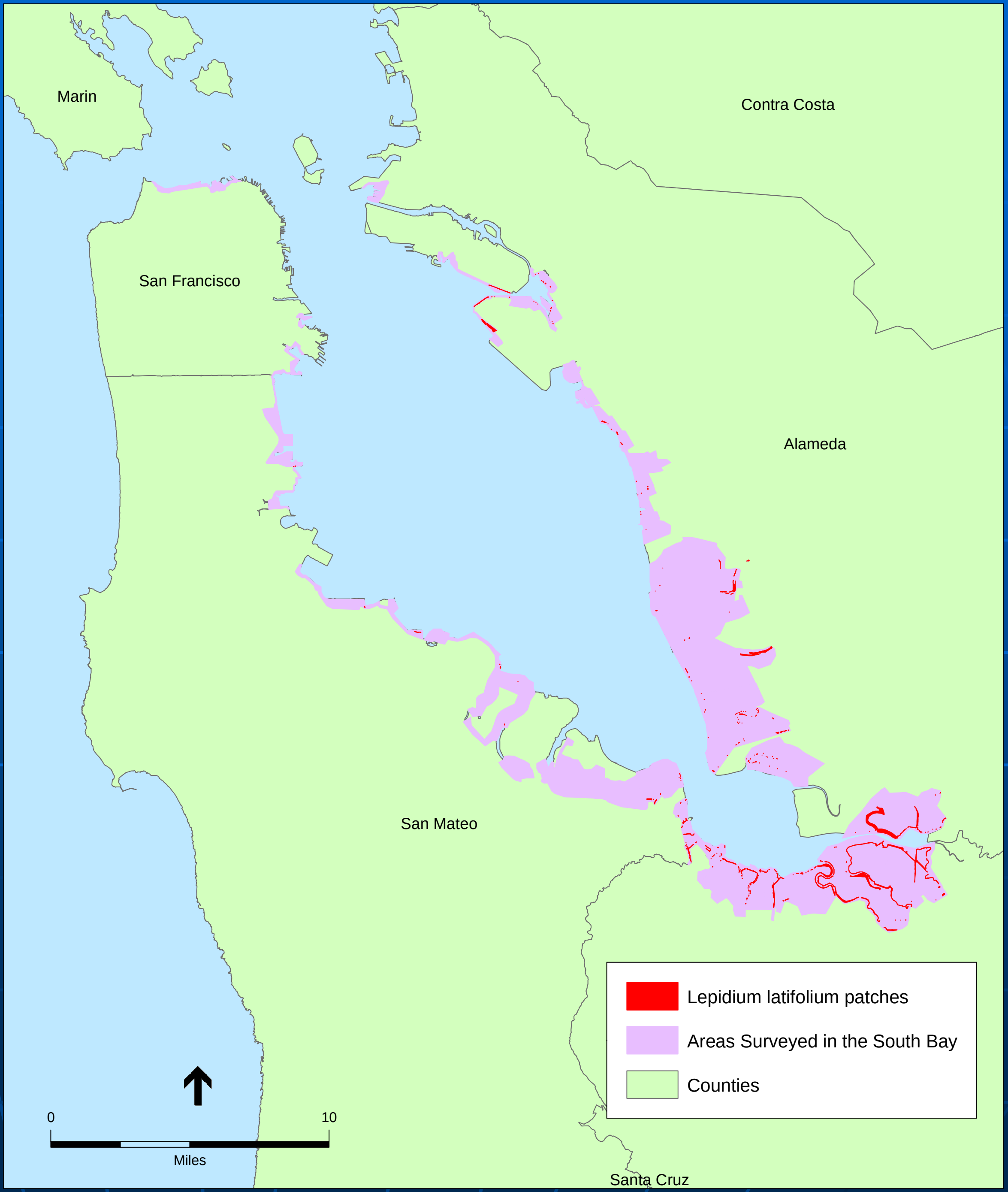

http://plants.usda.gov/maps/large/LE/LELA2.png

# Mapping the Shoreline

- GPS mapping method
  - the California Weed Mapping Handbook (CDFA)
- All patches larger than 1x1m were recorded.

Photo: ESA

# Patches of *L. latifolium*



# Predictive Modeling

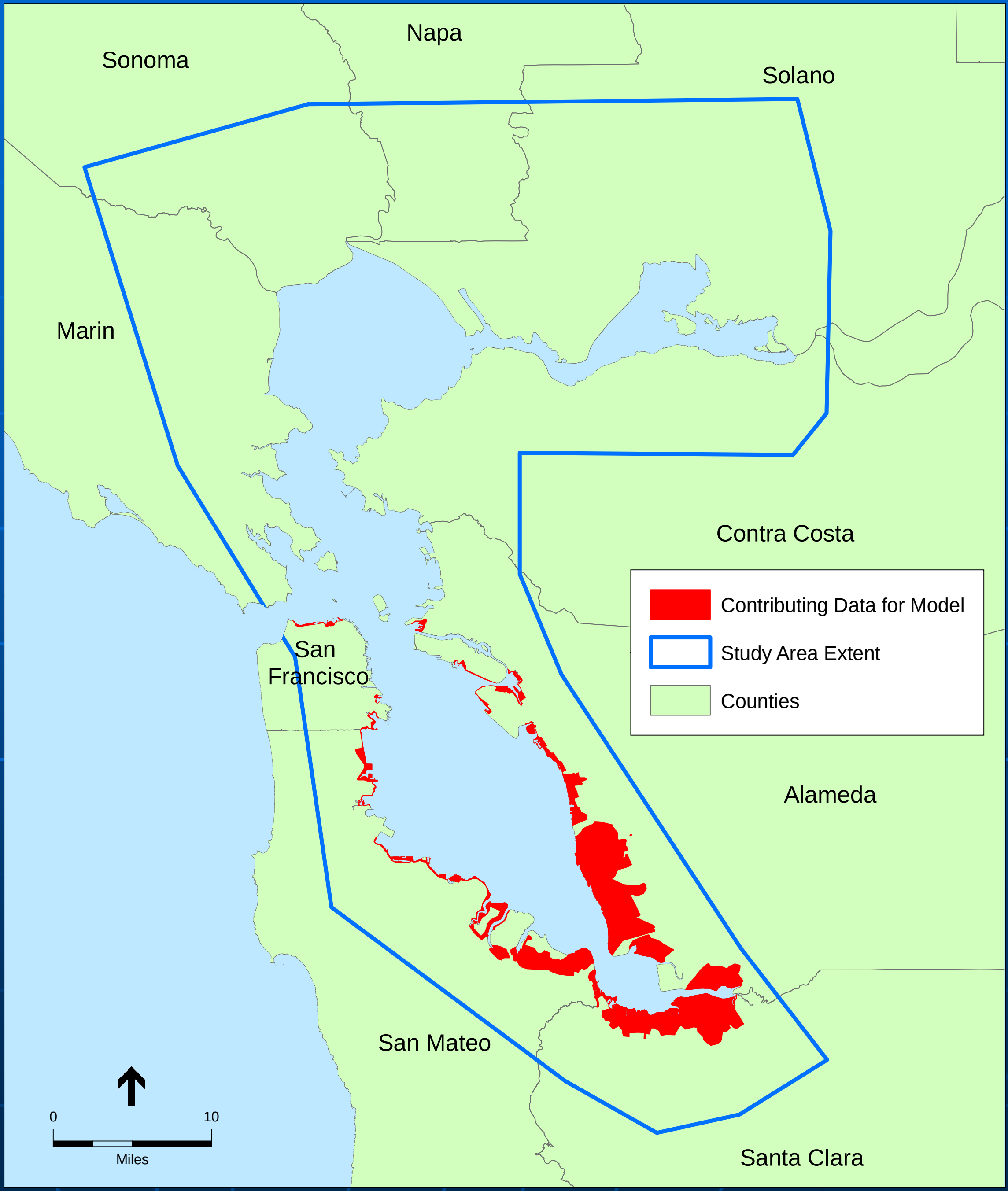
- Predict probabilities of occurrence or spatial distribution of species.
- Gain insight into species/environment relationships.
- Assess risk on local and/or regional scale.

# Predictive Modeling

1. Dependent variable: known presence/absence of species
2. Independent variables: environmental factors
3. Relate variables to distribution using a statistical model (Binomial logistic regression)
4. Transform model into a GIS probability map.

# Study Area

- Randomly selected 500 presence points and 500 absence points within areas surveyed.

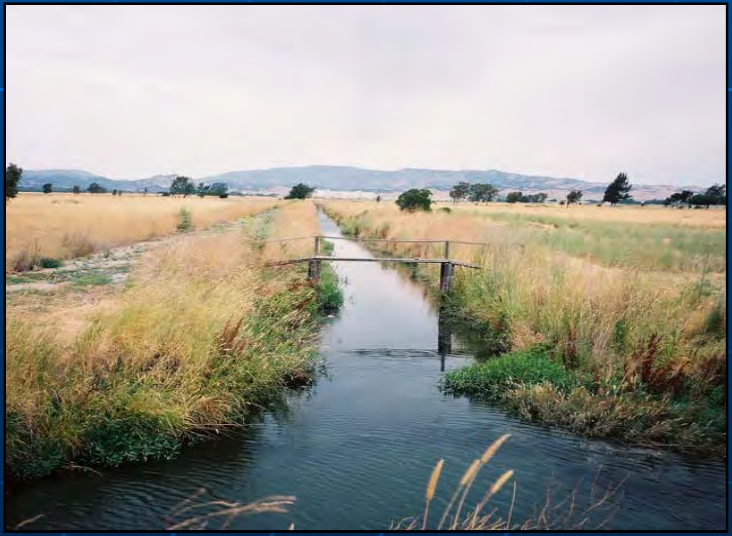


# Transferability

- Making predictions outside of the area in which the model was developed.
- Difficulties...
- Model valid for area surveyed.
- Model can provide guidance and risk assessment for outside areas.

# Environmental Predictor Variables

- Habitat type
- Tidal regime
- Elevation
- Distance to open water
- Distance to road
- Distance to levees
- Distance to agricultural land



Photos: ESA

# Spatial Extents

**Spatial Extent 1**


All variables considered

**Spatial Extent 2**


Elevation and Tidal Regime excluded

# Results and Discussion

Photo: ESA, Pittsburg, CA

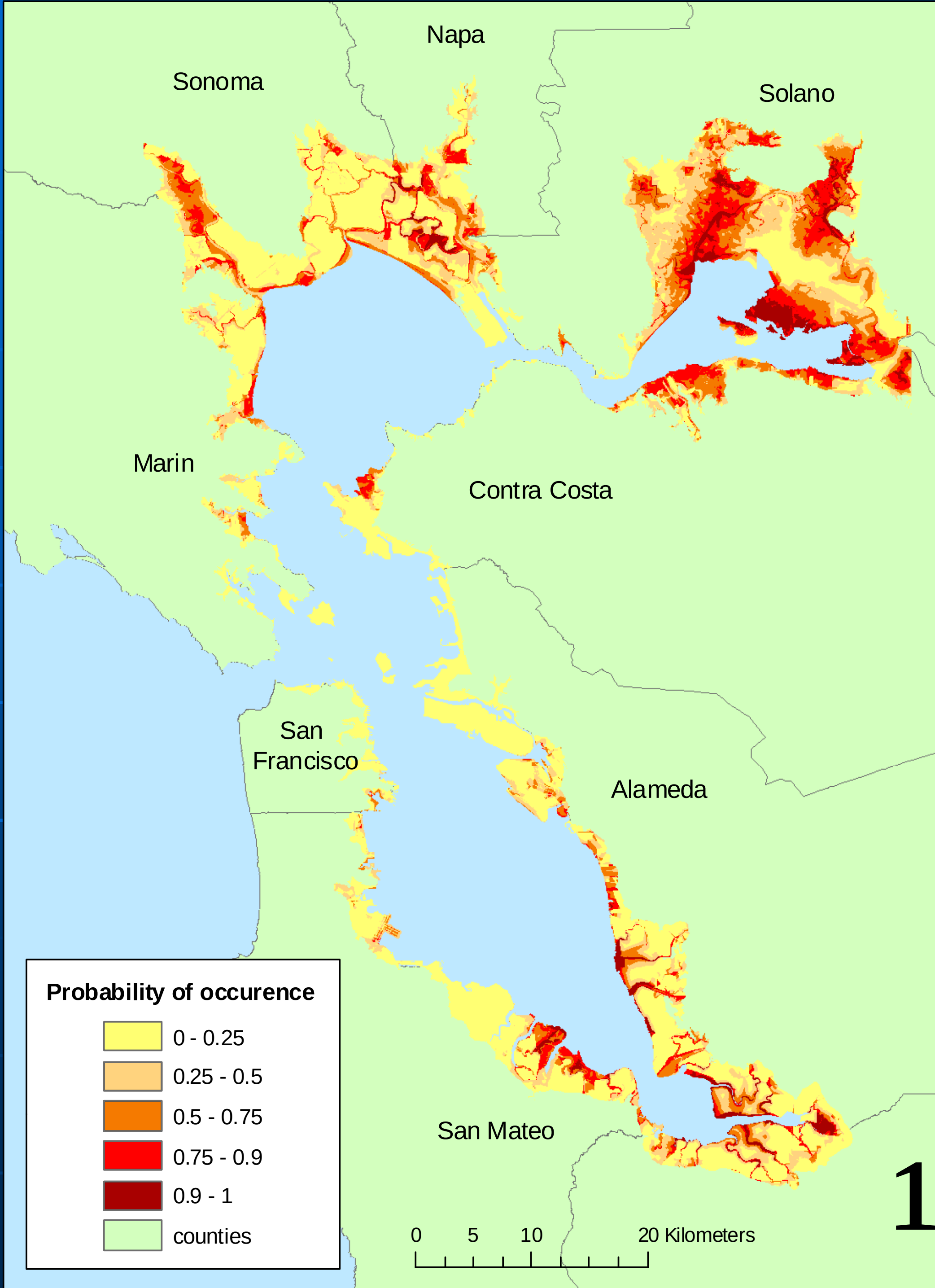
# Individual Variables: Nagelkerke $R^2$

| Variable | Nagelkerke $R^2$ |
|---|---|
| **Distance to *L. latifolium* patch** | **0.92** |
| **Tidal, tidal** | **0.44** |
| **Cover, estuarine wetland** | **0.42** |
| **Cover, water** | **0.36** |
| **Tidal, diked** | **0.36** |
| **Bay habitat, salt ponds** | **0.32** |
| **Bay habitat, marsh** | **0.31** |
| **Distance from water** | **0.25** |
| Distance from paved roads | 0.094 |
| Bay habitat, developed | 0.04 |
| Distance from levee | 0.036 |
| Cover, high-intensity development | 0.028 |
| Cover, palustrine wetland | 0.021 |
| Tidal, muted | 0.013 |
| Bay habitat, water | 0.012 |
| Tidal, non-tidal | 0.008 |
| Bay habitat, mud flat | 0.007 |
| Distance from agriculture | 0.006 |
| Cover, other estuarine habitat | 0.002 |
| Bay habitat, other | 0.001 |
| DEM | 0.001 |
| Cover, low-intensity development | 0.001 |
| Cover, other | <0.001 |
| Cover, bare land | <0.001 |
| Cover, grassland | <0.001 |

# Spatial Extent 1

Significant Variables

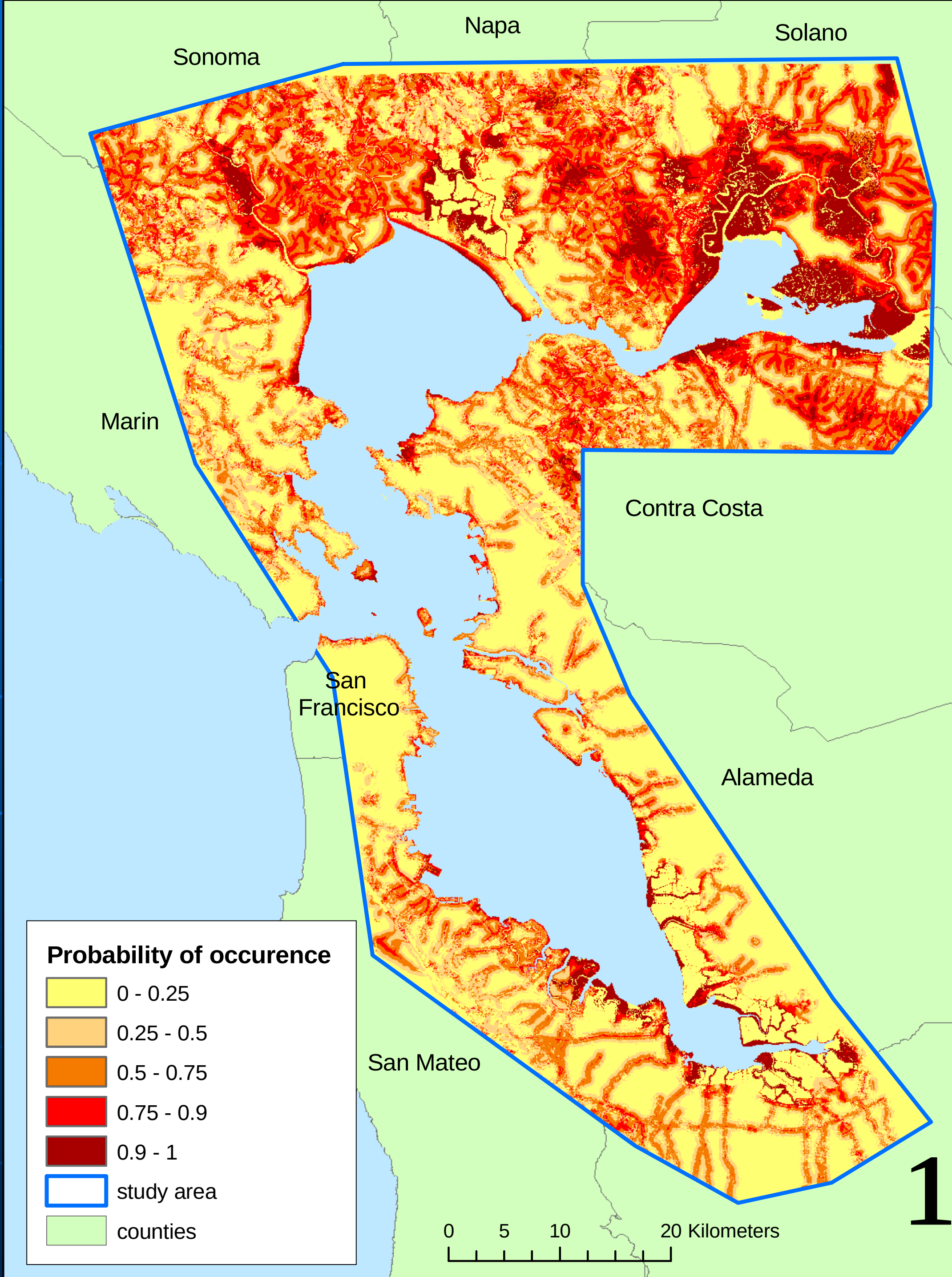
- Found in wetlands
- Outside of diked tidal areas
- Closer to water and levees
- Further from roads



# Spatial Extent 2

Significant Variables

- Found in wetlands
- Found in grassland, low-intensity devel. and bare ground
- Not in water
- Closer to water
- Further from road and agriculture



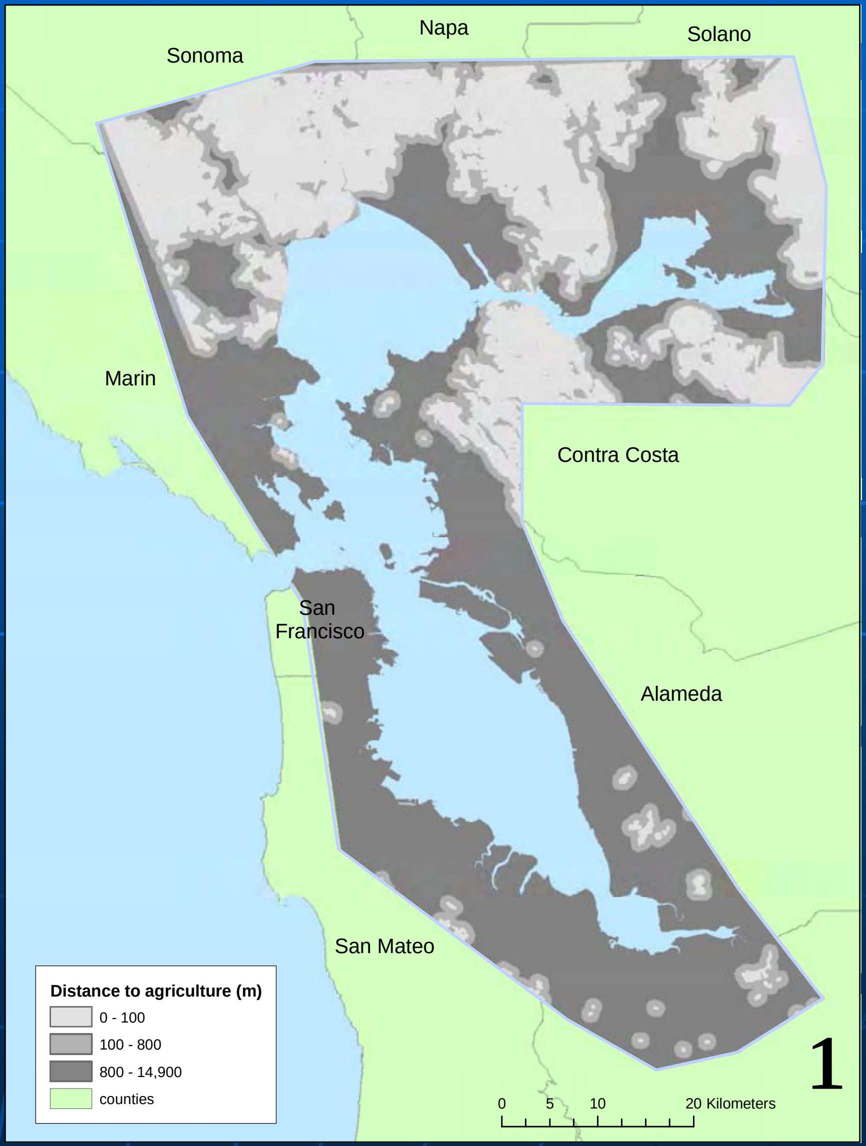
# Significant in Both Models

- Wetlands
- Shorter distance to water
- Longer distance to roads

# Unexpected Relationships

- Distance to Agriculture
- Distance to Roads
- Pattern may differ in San Pablo Bay and Suisun Marsh.

Variable, Distance to Agriculture →



# Internal Accuracy Tests (SPSS)

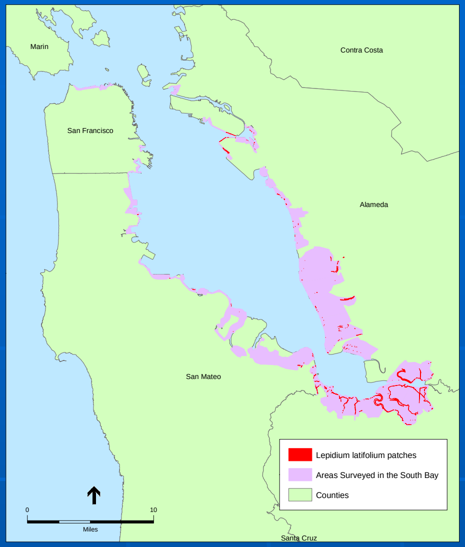
- Nagelkerke $R^2$
  - **Model 1:** 0.542
  - **Model 2:** 0.623

- Classification Table
  - **Model 1:**
    - 80.4% absent correct
    - 82.8% present correct.
  - **Model 2:**
    - 79.3% absent correct
    - 87.4% present correct.

Photo: Melanie Vanderhoof

# External Accuracy

(Random 30% of original data)

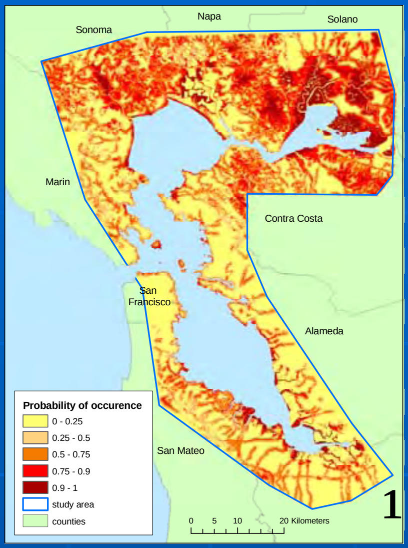



SOURCE: ESA

Distribution of *Lepidium latifolium*, CalFed ERP-02-P09. 202392

**Figure 3.3**

Distribution of *L. latifolium* in San Francisco Bay, California




1. The percent of cells coded correctly (either present or absent).
2. The percent of absent cells correctly coded.
3. The percent of present cells correctly coded.
4. The number of present cells correctly coded compared to the total number of cells predicted to be present (as a percent). *(# of false positives or over-predicting presence)*

# External Accuracy Findings

ACCURACY OF BINARY LOGISTIC REGRESSION MODELS AT DIFFERENT CUTPOINTS

| Spatial Extent | Cutpoint | 1(%) | 2(%) | 3(%) | 4(%) |
|---|---|---|---|---|---|
| 1 | 0.5 | 86.01 | 85.9 | 93.28 | 9.23 |
| 1 | 0.75 | 92 | 92.16 | 91.5 | 15.22 |
| 1 | 0.9 | 98 | 98.74 | 50.2 | 37.97 |
| 2 | .5 | 65.06 | 64.62 | 93.85 | 3.87 |
| 2 | 0.75 | 77.68 | 77.5 | 89.48 | 5.69 |
| 2 | 0.9 | 89.05 | 89.4 | 68.12 | 8.88 |

1. The percent of cells coded correctly (either present or absent).
2. The percent of absent cells correctly coded.
3. The percent of present cells correctly coded.
4. The number of present cells correctly coded compared to the total number of cells predicted to be present (as a percent).

# Models: extent 1 vs. extent 2

- Model 2 tested slightly higher on internal accuracy.
- Model 1 tested better with external accuracy.
- Tidal regime categories.

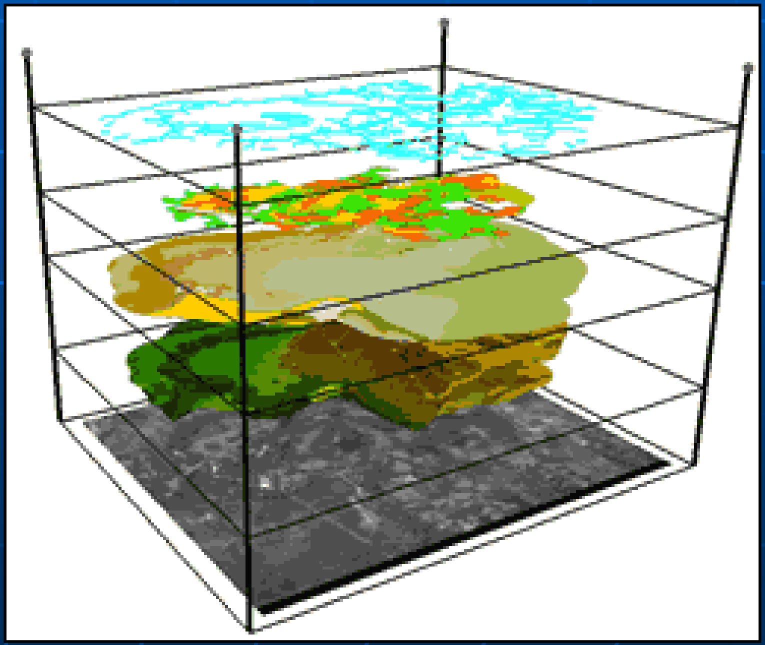
# Weaknesses of Models

- Resolution not ideal (limits accuracy and precision of relationship)
- Additional variables
- Site specific relationships

Photo: ESA

# Threats to Validity

- Absence data (hasn't spread yet)
- Conversions of patches to 30 m raster grid.
- Errors and inaccuracies within the data layers

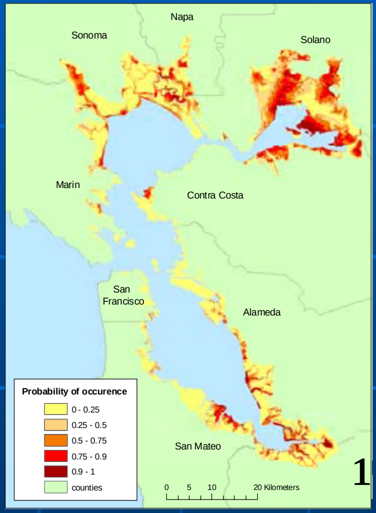


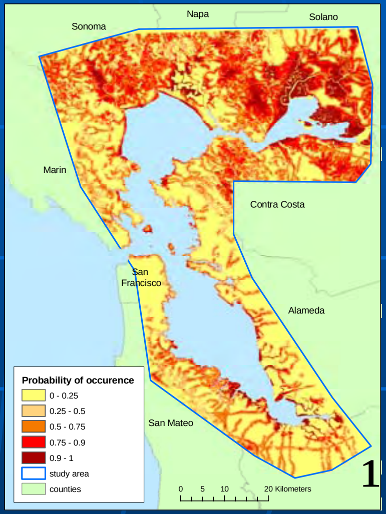
http://www.ruraltech.org/gis/images/gis_layers.gif

# Risk Assessment

Photo: ESA

- Is largely restricted to tidal marsh and riparian habitat.




- Predicted areas:
  - N. part of Suisun Bay
  - Grizzly Bay
  - Petaluma River
  - Napa-Sonoma marshes
  - Marshes in Don Edwards Wildlife refuge

# Risk and Public Lands

**Distribution of medium, high and very high risk land:**

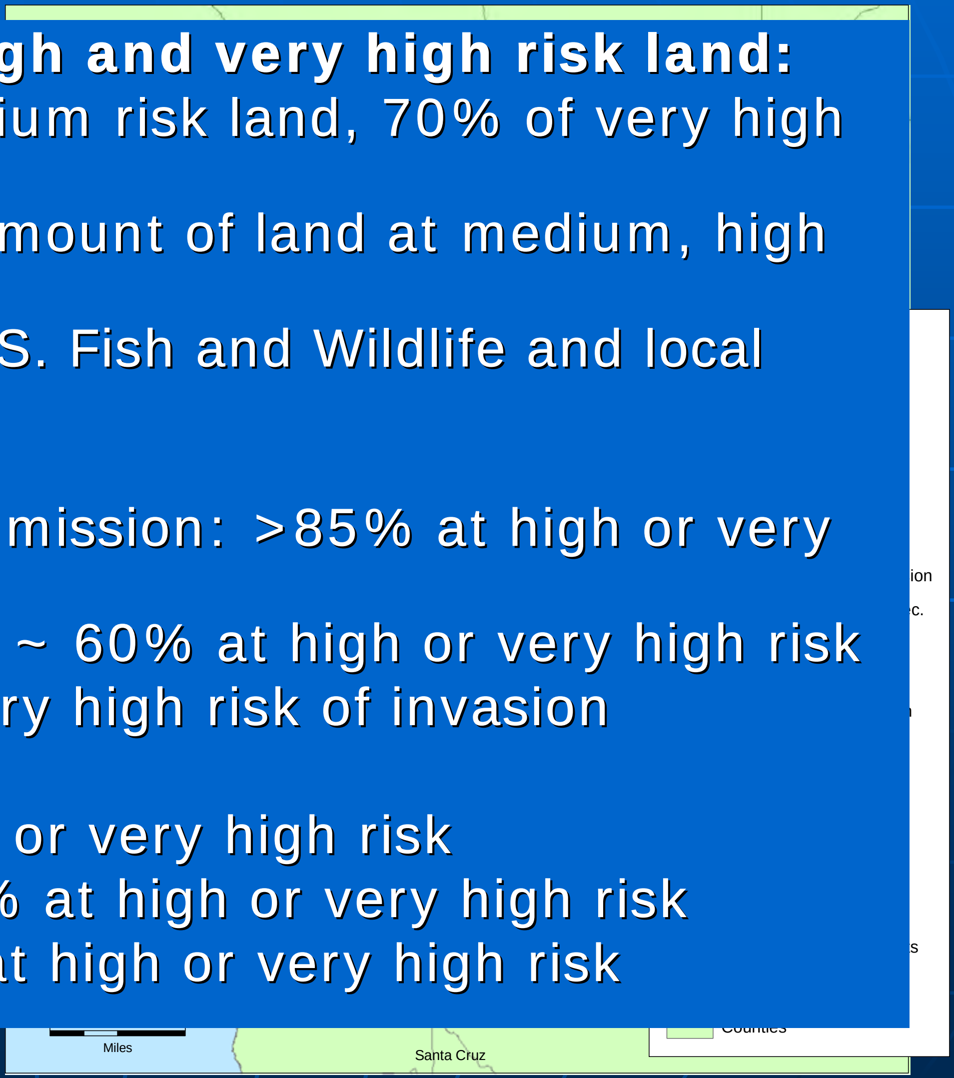
- Private land – 85% of medium risk land, 70% of very high risk land.
- CDFG manages the most amount of land at medium, high and very high risk.
- Department of Defense, U.S. Fish and Wildlife and local water districts.

High Risk

- California State Lands Commission: >85% at high or very high risk
- Conservancy / Land Trusts ~ 60% at high or very high risk
- CDFG: >40% at high or very high risk of invasion

Low Risk

- NASA 3% at medium, high or very high risk
- National Park Service: <3% at high or very high risk
- Open Space District ~4% at high or very high risk



# Management Recommendations

- Identify high risk lands.
- Control infestations early.
- Monitoring priorities:
  - Conservation/Restoration areas
  - Marsh habitat
  - Border between marshes and grasslands or low-intensity development.
  - Close to water

http://www.cherrug.se/galleri/vaxter/images/Lepidium%20latifolium%20Bitterkrassing%20Tygelsjo%20angar%2020050714%20001.jpg

# Conclusions

- Mapping efforts established baseline distribution data.

- Within the S.F. Bay Delta:
  - *L. latifolium* prefers to grow within the tidal zone
  - In brackish and salt marshes
  - Close to water

Photo: ESA

# Conclusions

- 1st modeling attempt for *L. latifolium* in California.
- Significant relationships discerned in model can contribute to knowledge of *L. latifolium*.
- Modeling attempt shows promise for other invasive species.

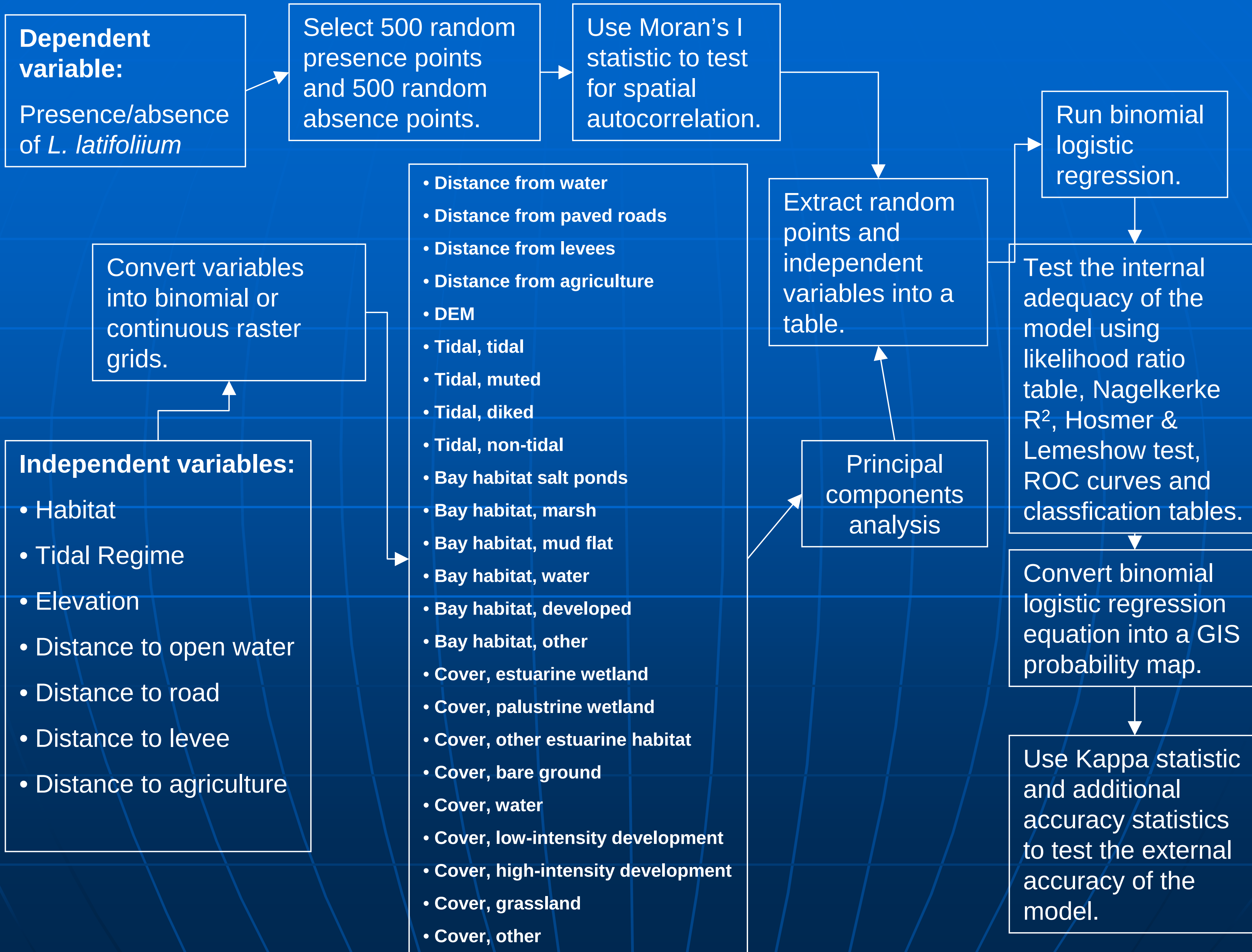

http://images.statemaster.com/images/motw/us_2001/california_ref_2001.jpg

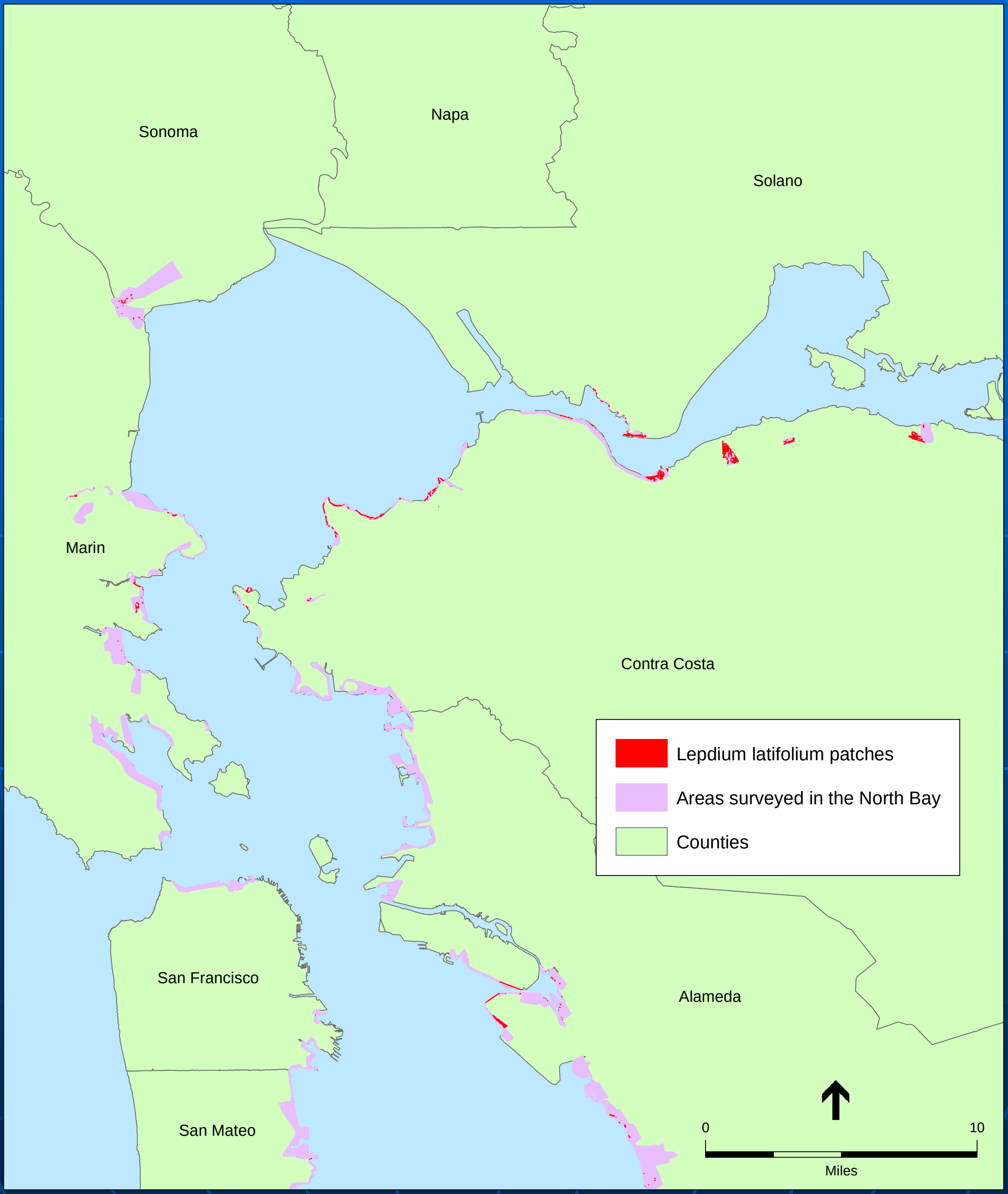
# Acknowledgements

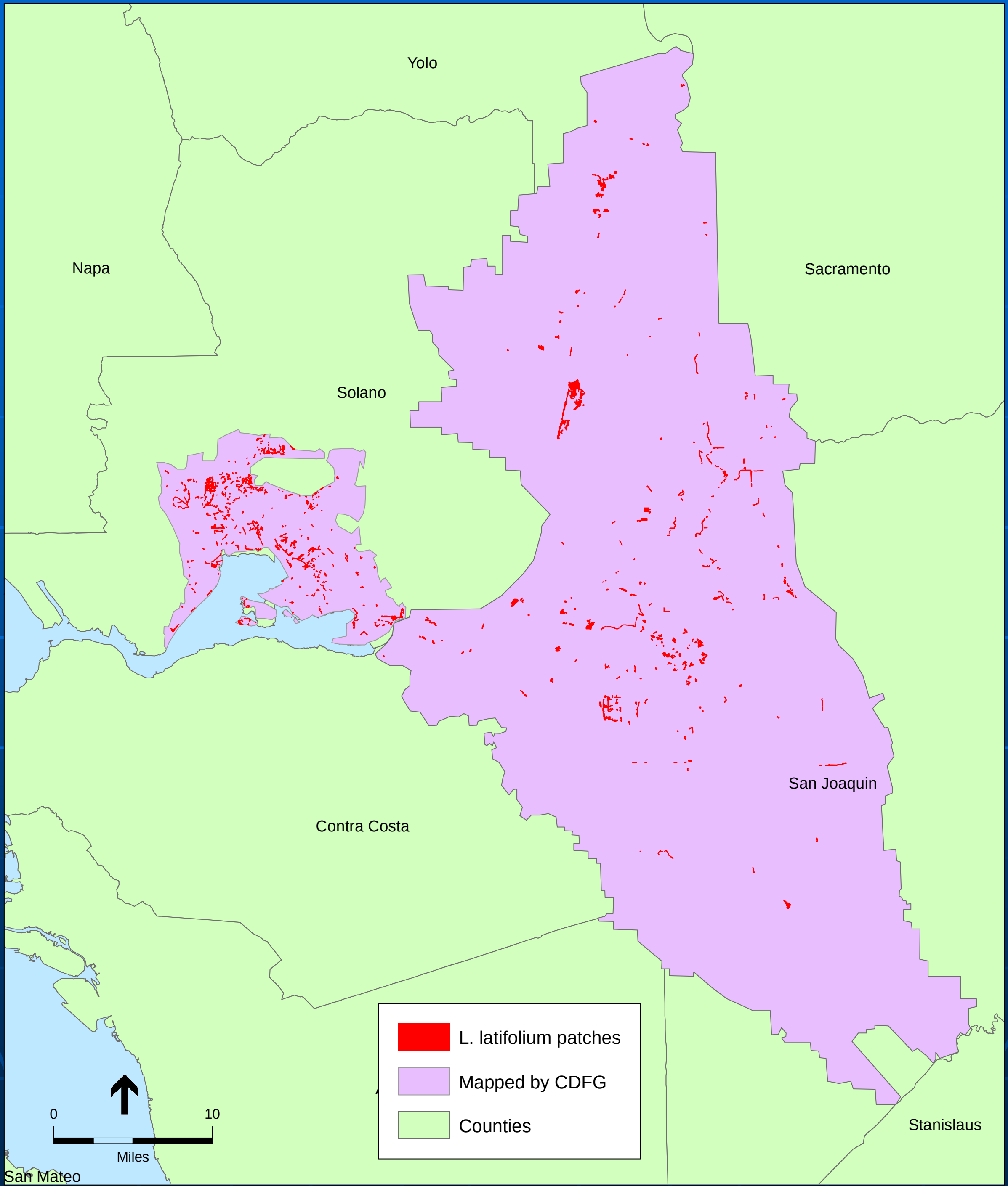

- CALFED
- Graduate school advisors, Barbara Holzman and Ellen Hines (SFSU)
- Martha Lowe, Mark Fogiel, field work
- Jerry Davis, Barry Nickel and Casey Cleve, GIS assistance
- Clyde Morris (USFWS) and John Krause (CDFG), access


- GIS data will be available through CDFG's BIOS.


Questions…


Dependent variable:
Presence/absence of L. latifoliium
Select 500 random presence points and 500 random absence points.
Use Moran's I statistic to test for spatial autocorrelation.
Run binomial logistic regression.
Convert variables into binomial or continuous raster grids.
• Distance from water
• Distance from paved roads
• Distance from levees
• Distance from agriculture
• DEM
• Tidal, tidal
• Tidal, muted
• Tidal, diked
• Tidal, non-tidal
• Bay habitat salt ponds
• Bay habitat, marsh
• Bay habitat, mud flat
• Bay habitat, water
• Bay habitat, developed
• Bay habitat, other
• Cover, estuarine wetland
• Cover, palustrine wetland
• Cover, other estuarine habitat
• Cover, bare ground
• Cover, water
• Cover, low-intensity development
• Cover, high-intensity development
• Cover, grassland
• Cover, other
Extract random points and independent variables into a table.
Test the internal adequacy of the model using likelihood ratio table, Nagelkerke R2, Hosmer & Lemeshow test, ROC curves and classfication tables.
Independent variables:
• Habitat
• Tidal Regime
• Elevation
• Distance to open water
• Distance to road
• Distance to levee
• Distance to agriculture
Principal components analysis
Convert binomial logistic regression equation into a GIS probability map.
Use Kappa statistic and additional accuracy statistics to test the external accuracy of the model.


Sonoma
Napa
Solano
Marin
Contra Costa
San Francisco
San Mateo
Alameda
Lepdium latifolium patches
Areas surveyed in the North Bay
Counties
0
10
Miles


Yolo
Napa
Sacramento
Solano
San Joaquin
Contra Costa
Stanislaus
San Mateo
L. latifolium patches
Mapped by CDFG
Counties
0
10
Miles

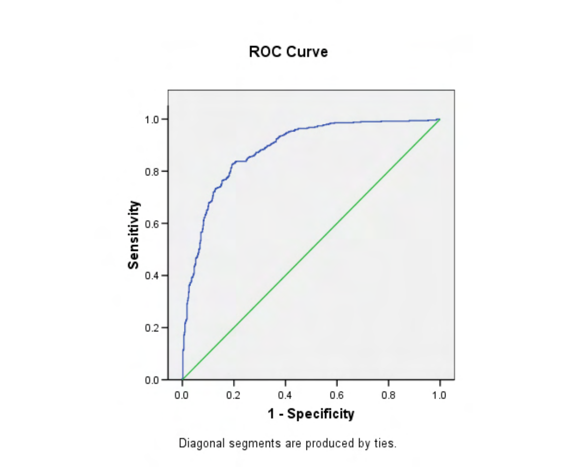
# Internal Accuracy Tests

| Spatial Extent | Likelihood ratio table | Nagelkerke $R^2$ | Hosmer & Lemeshow | % Absent Classified Correctly | % Present Classified Correctly |
|---|---|---|---|---|---|
| 1 | <0.001 | 0.542 | 0.059 | 80.4 | 82.8 |
| 2 | <0.001 | 0.623 | 0.631 | 79.3 | 87.4 |

# ROC Curves

**Table 12.** The area under the ROC curve and its statistical significance for both *L. latifolium* predictive models.  The cutpoint was assumed to be P =0.5.

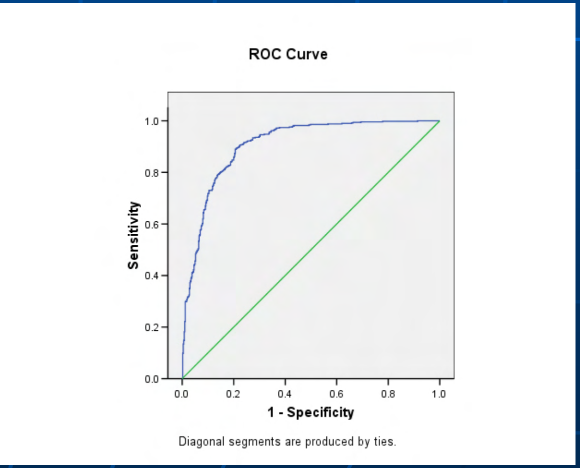
| Spatial Extent | ROC Area | ROC Asymp |
|---|---|---|
| 1 | 0.884 | <0.001 |
| 2 | 0.907 | <0.001 |




Spatial extent 1

**Table 13.** The sensitivity and specificity values on the ROC curvec at different cutpoints.  The goal is to maximize sensitivity and minimize specificity.

| Spatial Extent | Cutpoint | Sensitivity | Specificity |
|---|---|---|---|
| 1 | 0.5 | 0.826 | 0.196 |
| 1 | 0.75 | 0.679 | 0.104 |
| 1 | 0.9 | 0.217 | 0.012 |
| 2 | 0.5 | 0.874 | 0.206 |
| 2 | 0.75 | 0.714 | 0.1 |
| 2 | 0.9 | 0.31 | 0.02 |




Spatial extent 2