Challenges in Predicting the Distributions of Exotics using Habitat Distribution Models

Chad C. Jones

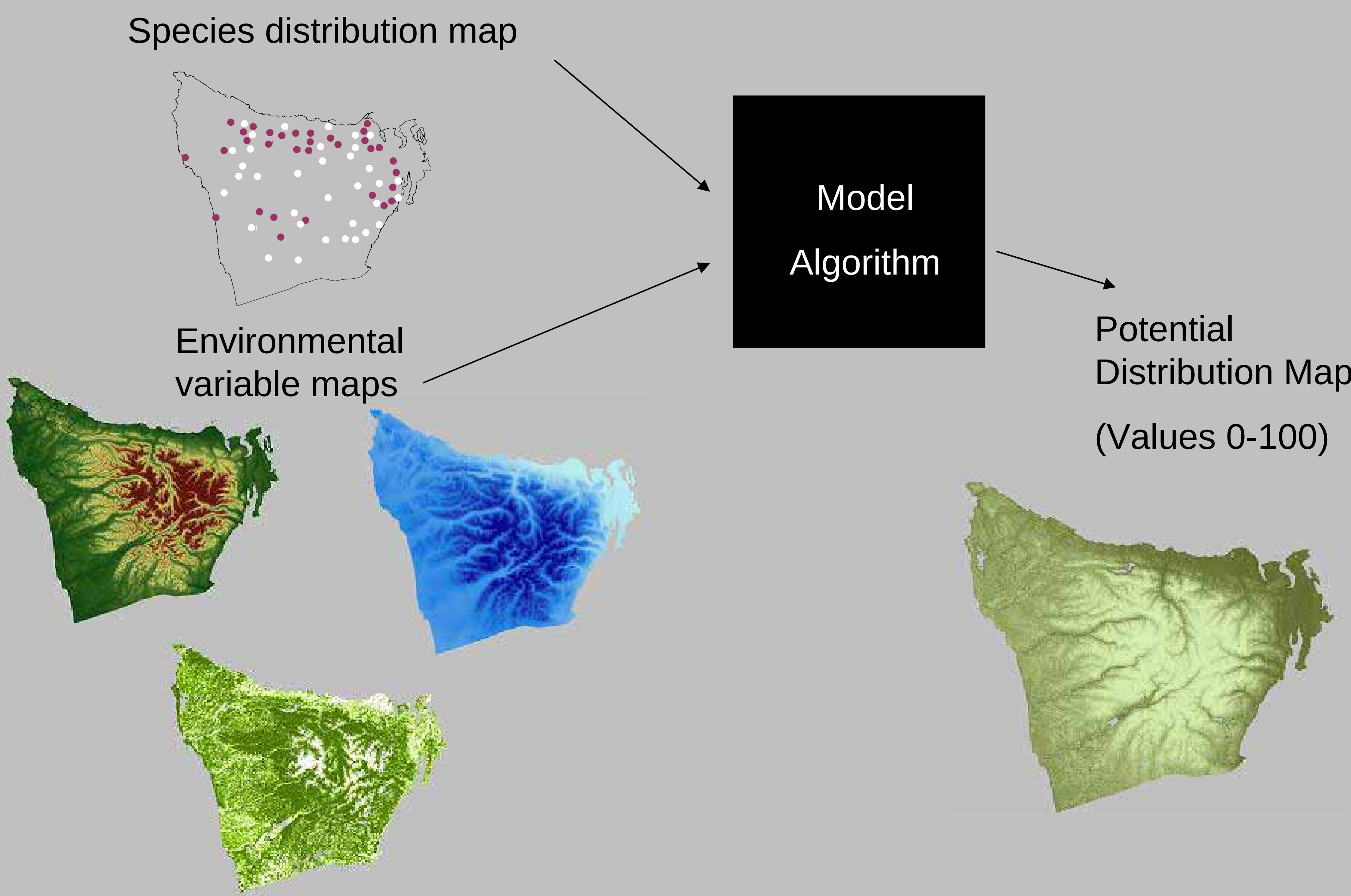
Steven A. Acker

Charles B. Halpern

Potential of Models

- Monitoring for invasive species is costly
- If we can predict where exotics have the potential to invade, we can target monitoring efforts in this area
- Habitat distribution models have been used to predict distributions of rare species and are starting to be applied to invasive species

Habitat Distribution Models



Modeling Techniques

- Logistic Regression (LR)
  - Most commonly used technique
  - Requires presence and absence data
  - Requires many assumptions
- Ecological Niche Factor Analysis (ENFA)
- Genetic Algorithm for Rule-set Prediction (GARP)

Model Assumptions

HOWEVER

These models assume that the species are in equilibrium with the environment

AND

Presence-Absence models assume that the sampled frequency reflects the actual frequency

Data Challenges

- Invasive species data are often not systematically or randomly sampled
- Datasets are often clumped along roads and in areas currently invaded
- Absence data is often uncertain

Modeling Challenges

1) Clumped data points can have a strong effect on model results
2) Spreading species can lead to biased results
3) Species frequency is often unknown and may affect model results
4) We cannot test how well models predict future distributions

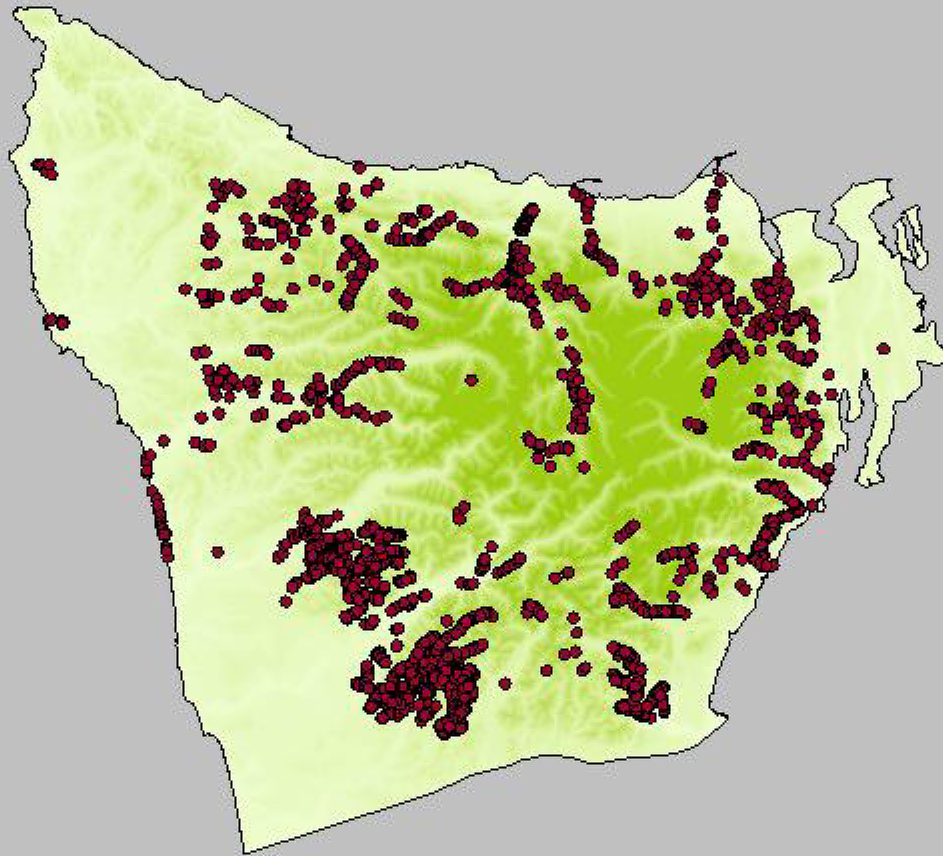
Project Goals

- Develop habitat distribution models for three invasive species on the Olympic Peninsula to inform development of monitoring
- Compare the results and sensitivities of three modeling techniques

Olympic Peninsula, WA

~12,500 km^2

13 datasets combined to provide 4142 +/- data points



Study Species

Geranium robertianum

Cirsium arvense

Photo: Roger del Moral

Rubus laciniatus

Modeling Methods

- For each species and model technique:
  - Data divided into 5 equal subsets
  - Developed the model with 4/5 of the data
  - Used the remaining 1/5 to assess accuracy
  - Repeated this for all 5 possible subsets*
- Used an equal number of absences as presences for each model run

*At present only one subset has been used for GARP

Habitat Variables

- Climate
  - # of frost days
  - Annual precipitation
  - Precipitation frequency
  - Water vapor
  - Radiation
  - Distance to nearest water
- Topography
  - Slope
  - Potential radiation
  - Heat load
  - Topographic Moisture Index
- Vegetation
  - Conifer cover
  - Vegetation cover

Questions

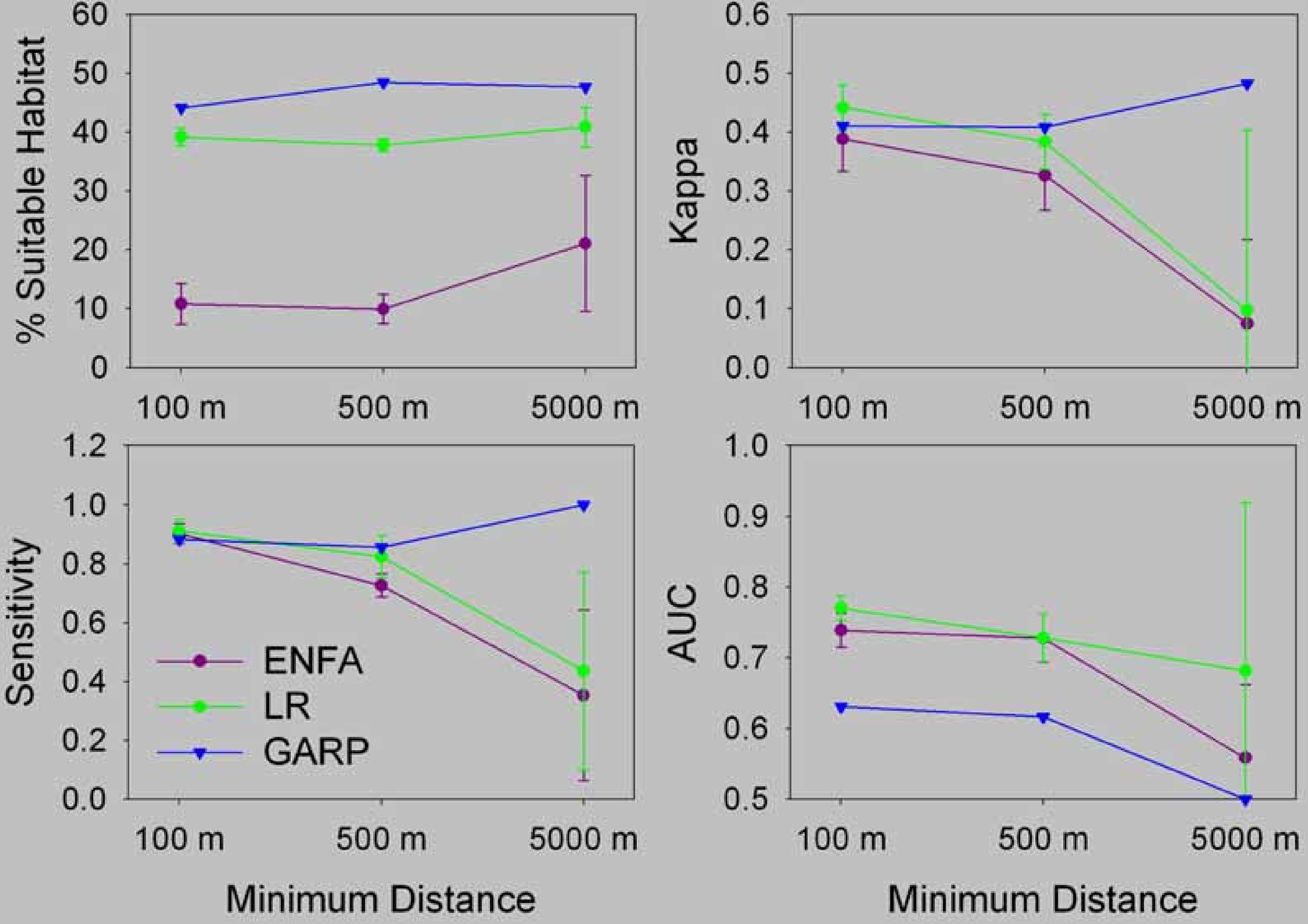
- How does a clumped sample affect model results?
- How are model results affected by the fact that species are still spreading?
- How does frequency affect logistic regression results?
- How do results from different models compare?

Effects of Clumping

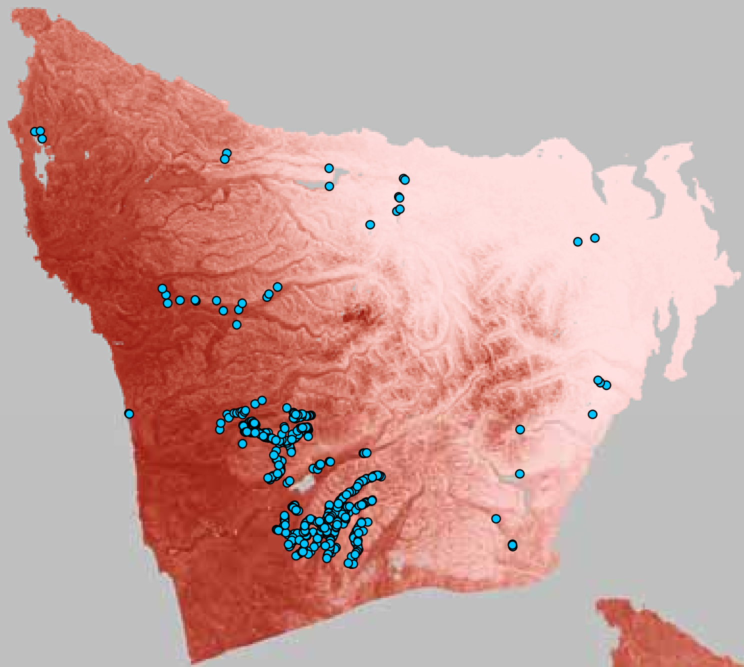
- Set minimum distances between plots and randomly removed plots closer than this minimum distance

Min Distance	*Geranium*	*Rubus*	*Cirsium*
100 m	219	559	549
500 m	80	245	271
5000 m	35	38	67

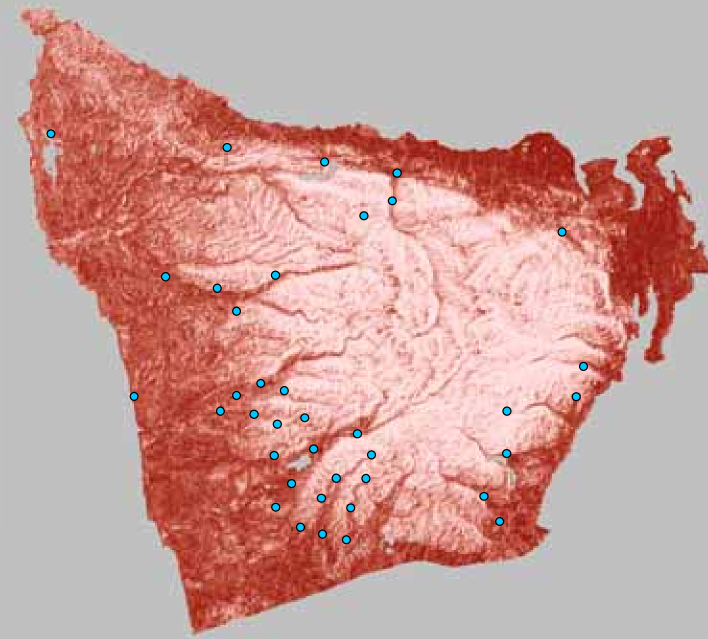
Rubus laciniatus



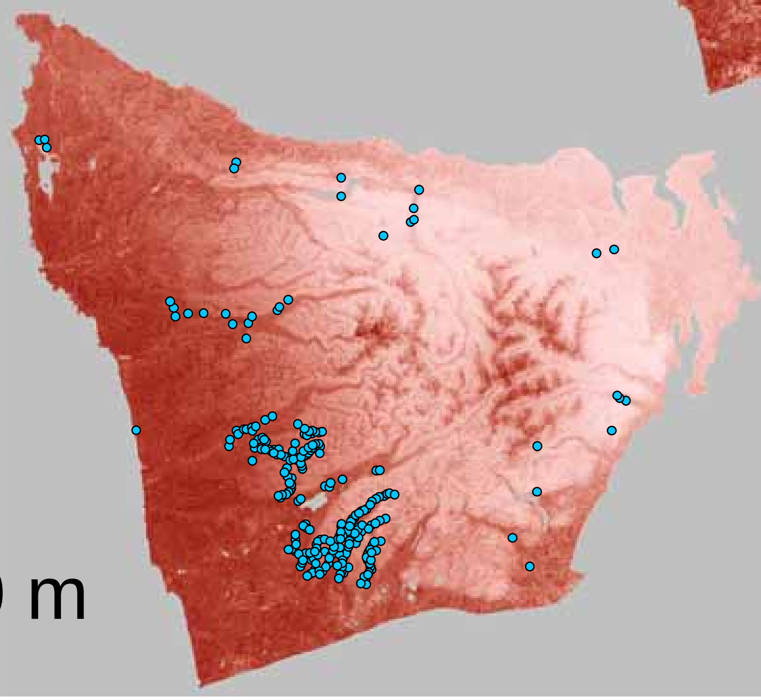
Rubus laciniatus ENFA models



100 m



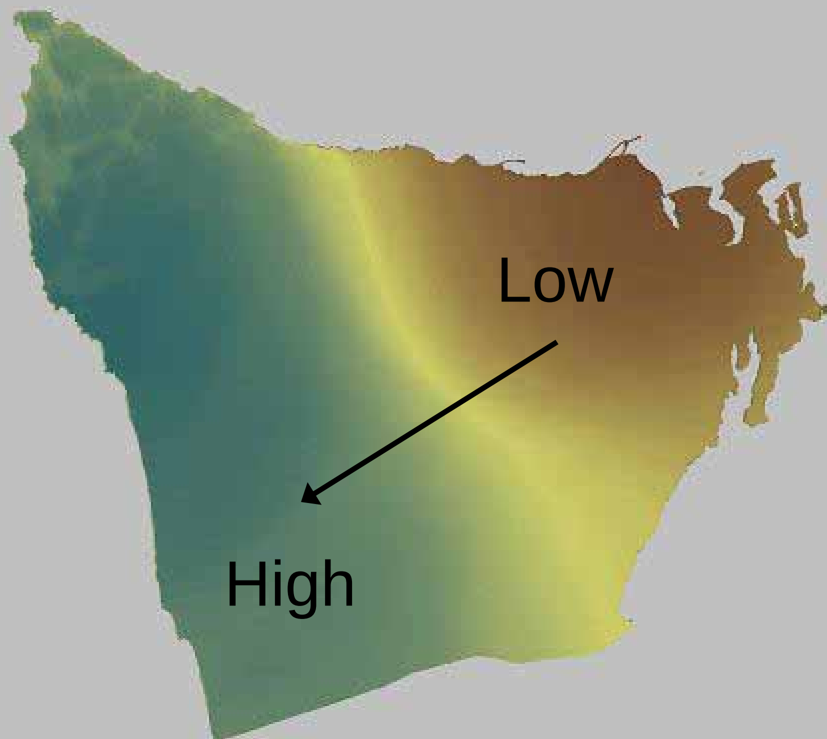
5000 m



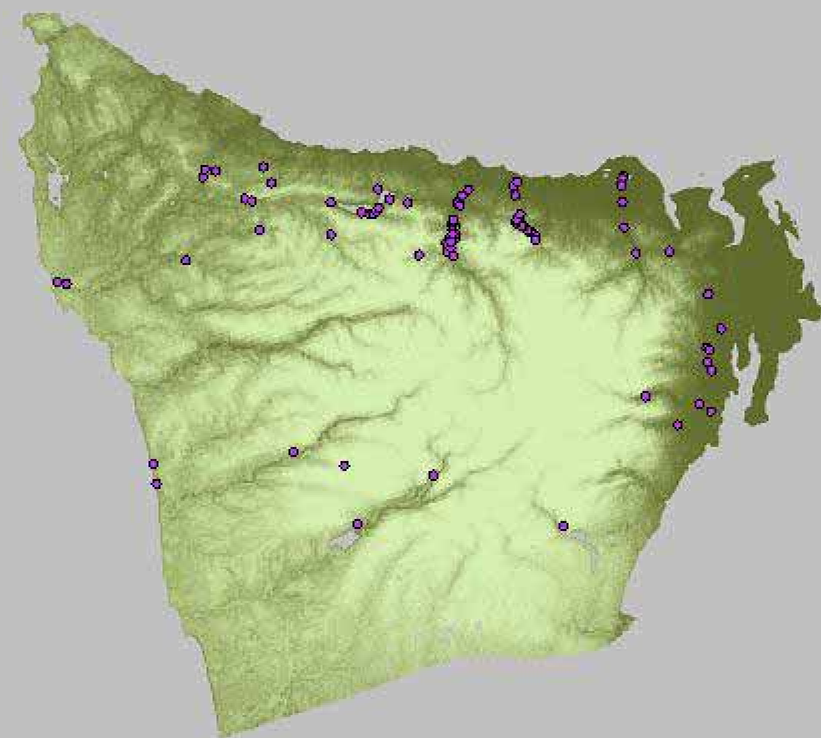
500 m

Effects of Spreading Species

Frequency of Precipitation

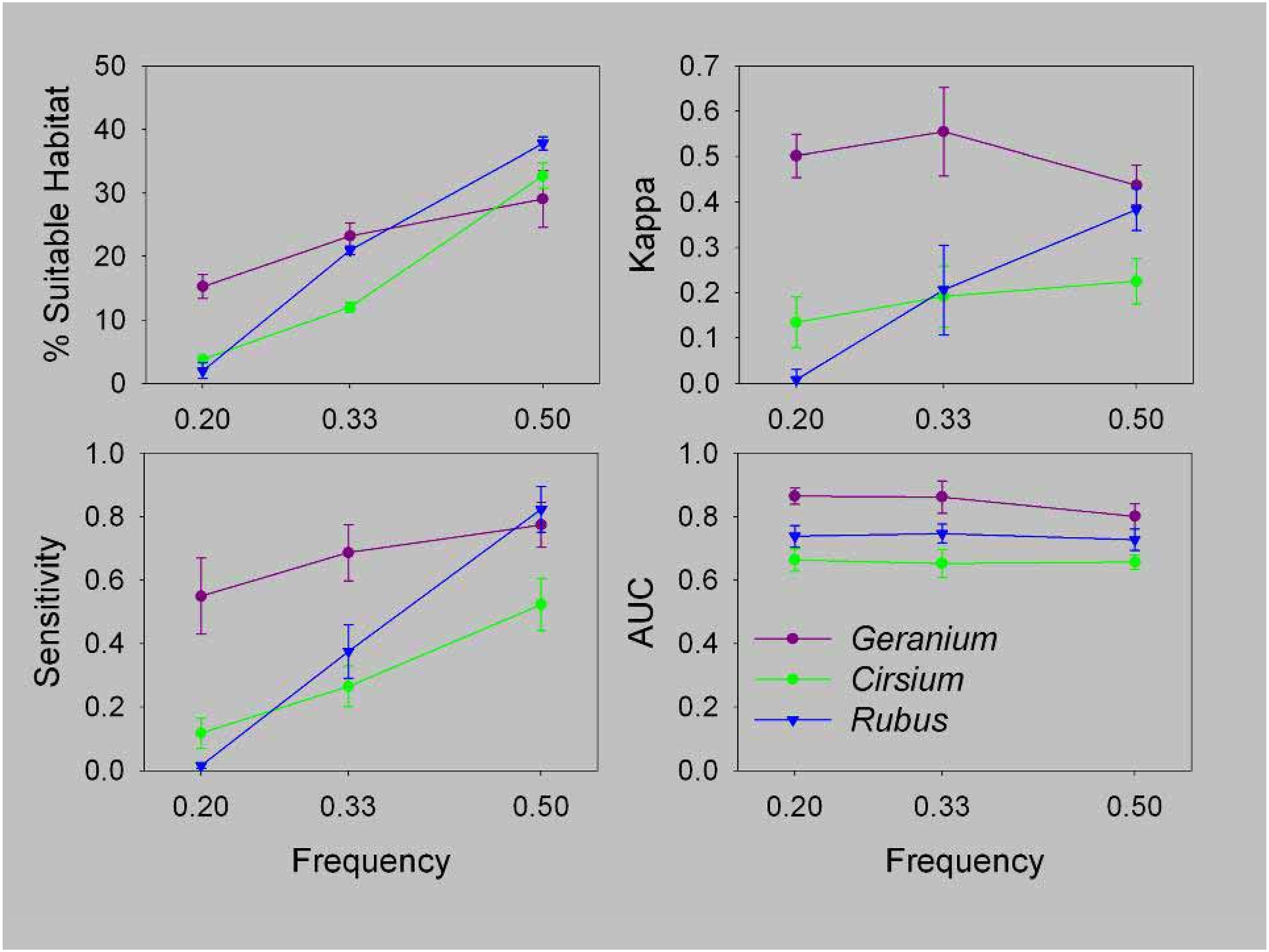


Geranium robertianum model results



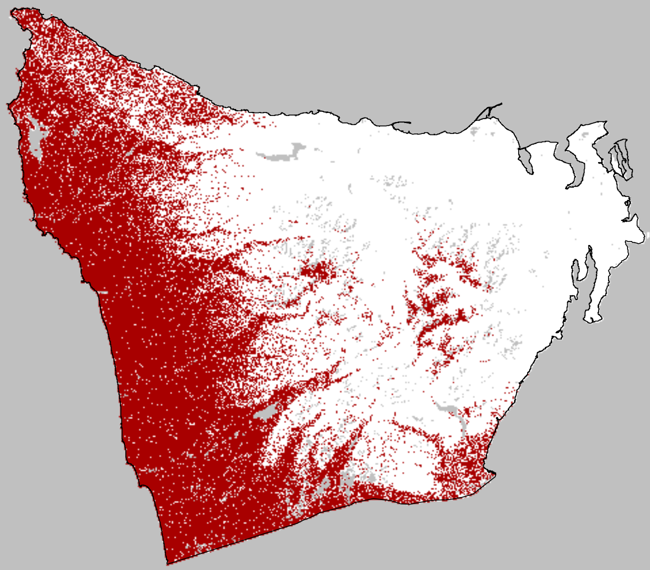
Effects of Frequency

- Sampled frequency ranged from 10-20%
- However, this was highly dependent on where the plots were located, actual frequency is unknown and likely increasing
- Ran Logistic regression models at different frequencies (50%, 33%, and 20%).

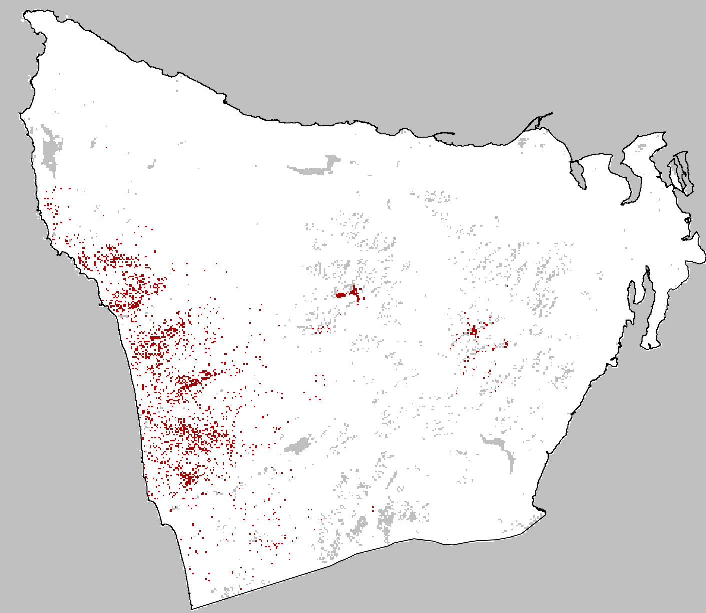


% Suitable Habitat
Kappa
Sensitivity
AUC
Geranium
Cirsium
Rubus
0.20
0.33
0.50
Frequency
Frequency

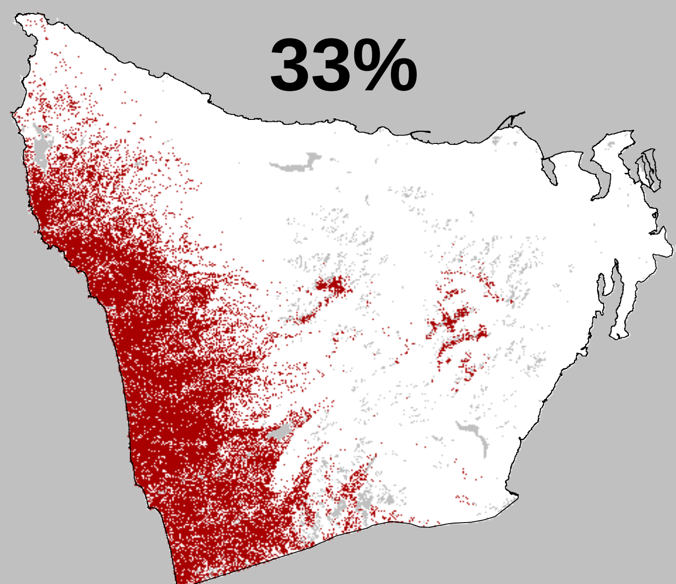
Rubus laciniatus – 500 m



50%

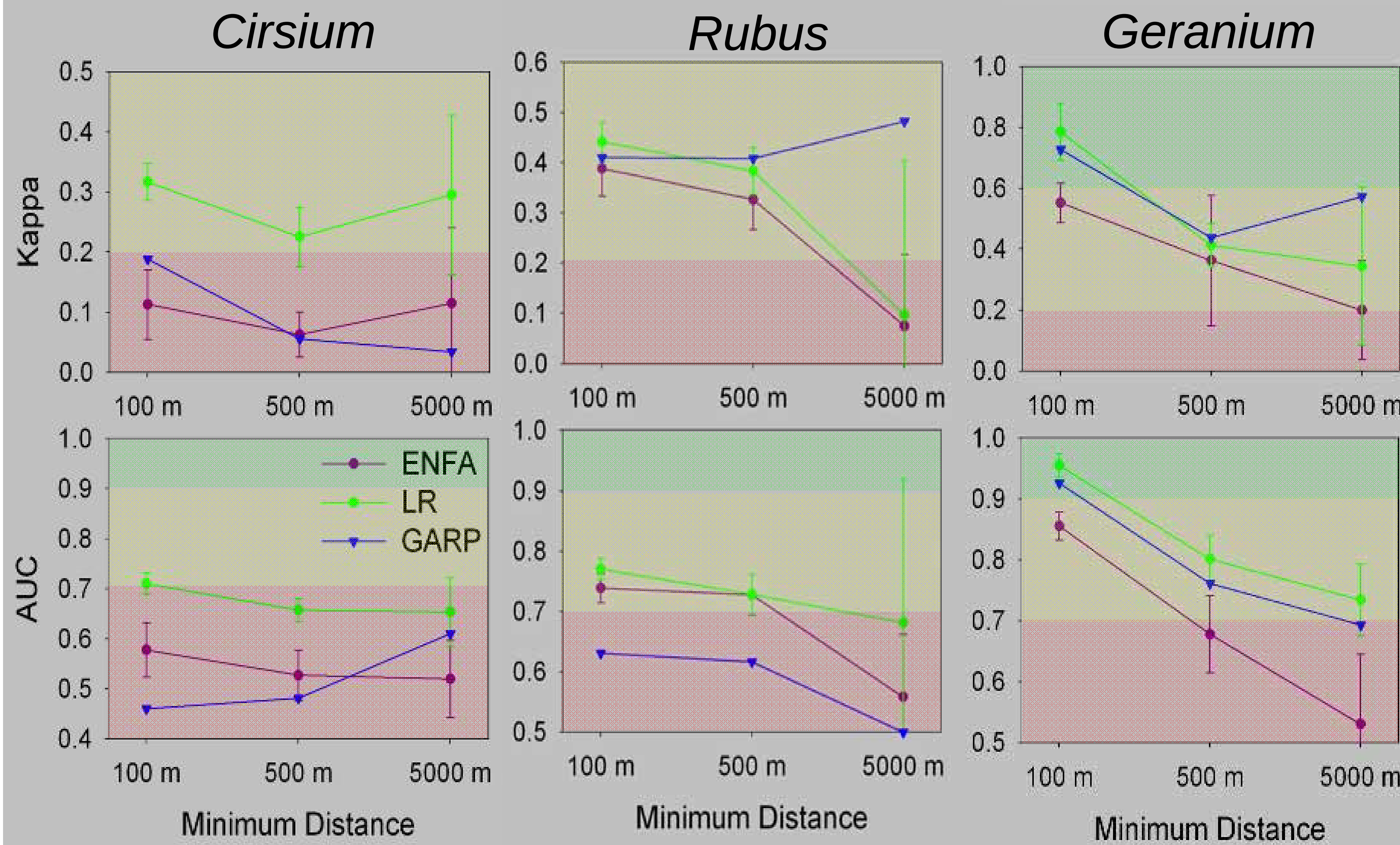


20%



33%

Comparison of Models



Conclusions

- Spurious correlations and clumping in the datasets potentially pose serious problems
- Logistic regression tends to perform better than ENFA or GARP, but is sensitive to frequency
- Measures of accuracy based on the current distribution (e.g. Kappa and AUC) may not accurately assess how well the models predict future distributions

Recommendations

- For Logistic Regression use a high frequency
- Sample design should try to capture the invasive across its entire distribution rather than focusing on the areas with greatest invasion
- Accuracy measures should be treated with care – look at the distribution maps and consider information about the biology of the species

Future Directions

- Finish modeling for two additional species
- Combine models to create a priority map for monitoring
- Add dispersal into the models
- Compare with models based on information from the native range

Acknowledgements

- Data Sources
  - Olympic National Park
  - ONP Exotic Plant Management Team
  - Olympic National Forest
  - Ann Lezberg
  - Robin Lesher, USDA Forest Service
  - Collette DeFerrari
  - Olympic Habitat Development Study
  - Ed Schreiner, USGS
  - Andrea Woodward, USGS
- Funding
  - National Park Service