

● Cost-Sensitive Machine Learning Algorithms

Invasive Species Decision Support, Risk Analysis, and Policy

- John M. Drake
- University of Georgia

- David Finnoff
- University of Wyoming



Odum School
of Ecology
The University of Georgia

UNIVERSITY
OF WYOMING
New Thinking

Overview

Problem statement

- Costs of invasive species are \$US billions per year
- Risk analysis and policy mechanisms in place or under development (Quarantine 37")
- Ecological risk assessment procedures inadequate

Objective

- Develop cost-sensitive decision support tools to aid risk analysis for species proposed for introduction and rapidly evaluate new non-indigenous species concerning potential for further spread

Construction of the database

Species classes

	Species	Weeds	State listed	Federally listed
total	5954	1110	435	46
present in >1 state	3076	974	373	31
ornamental	1292	394	143	10

*Includes Alaska, Hawaii, Puerto Rico, and the Virgin Islands

145 Variables

~10% relate to native distribution

~15% relate to introduced distribution

~70% relate to biological traits

Large datasets

- seed mass (>2000)
- chromosome number (>1000)
- and maximum height (>500)

Biological traits

Growth form

Life history

Wetland habitat association

maximum height

maximum chromosome number/species

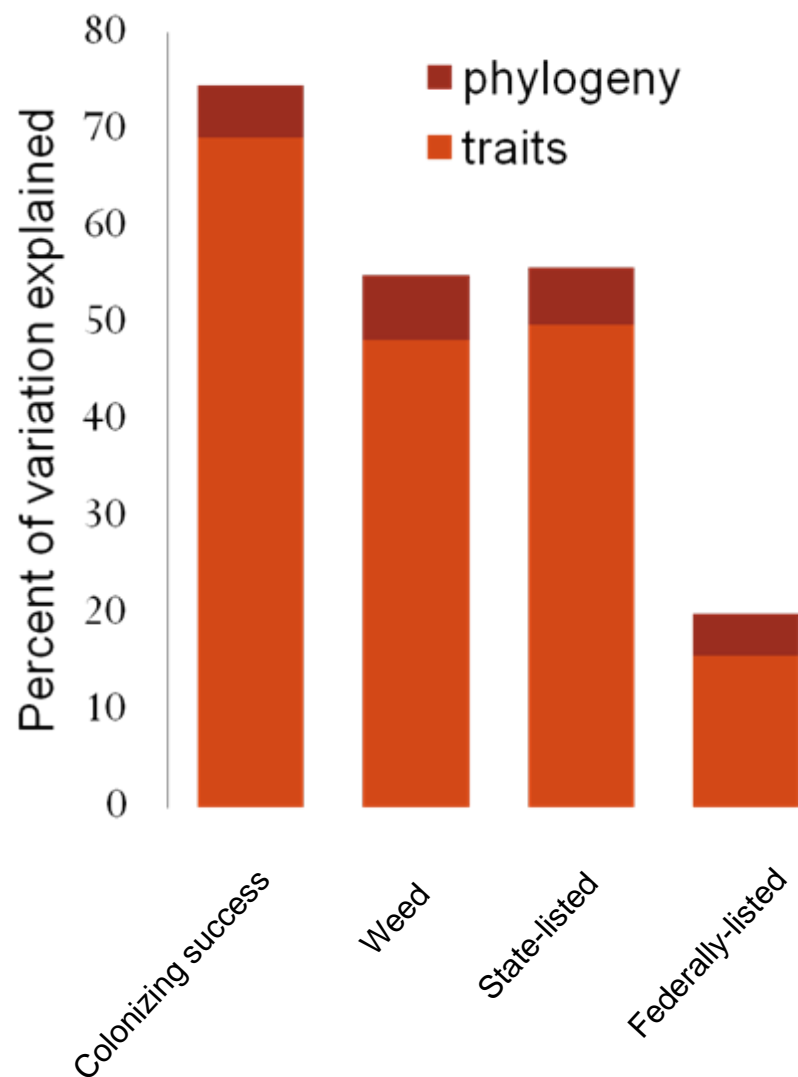
seed mass

Leaf traits

Native latitudinal and longitudinal range

Predictive covariates of invasive genera

*also Family-level and Order-level



Genus	Common	Invasive class
<i>Emex</i>		Federal
<i>Imperata</i>	Cogon	Federal
<i>Pennisetum</i>	Fountain grass, Kikuyu grass, Feathertop grass	Federal
<i>Prosopis</i>	Mesquite	Fed, State, Weed
<i>Carduus</i>	Thistle	State, Weed
<i>Centaurea</i>	Star-thistle, knapweed	State
<i>Dipsacus</i>	Teasel	State
<i>Miscanthus</i>	Silvergrass	State
<i>Onopordum</i>	Thistle	State
<i>Solanum</i>	Nightshade	State
<i>Salsola</i>	Tumbleweed	State, Weed

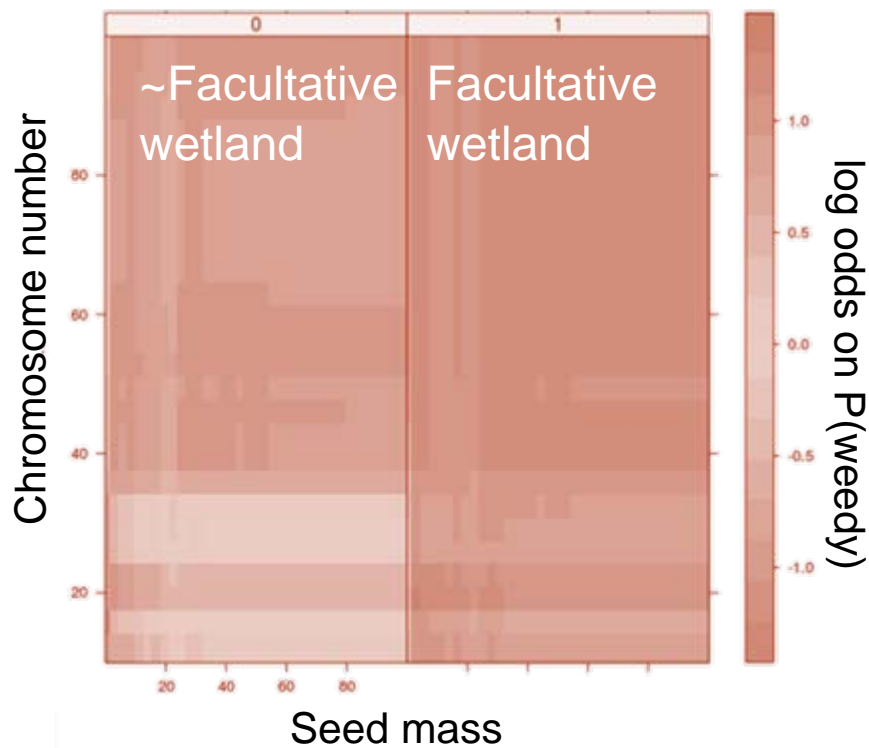
Species classification on a balanced dataset: Maximal performance

- Facultative wetland association
- Oligate wetland association
- Maximum height
- Seed mass
- Maximum chromosome number
- Leaf traits: evergreen-ness, C:N ratio, leaf specific area

Key biological traits

Performance Summary

	Weeds		State-listed	
	prediction on 20% holdout (95% CI)	sample size (test set)	prediction on 20% holdout (95% CI)	sample size (test set)
all species	0.74 (0.69,0.78)	447	0.69 (0.62,0.76)	174
species present in > 1 state	0.78 (0.73,0.82)	385	0.78 (0.70,0.84)	149
ornamentals	0.85 (0.79,0.9)	160	0.75 (0.62,0.86)	56



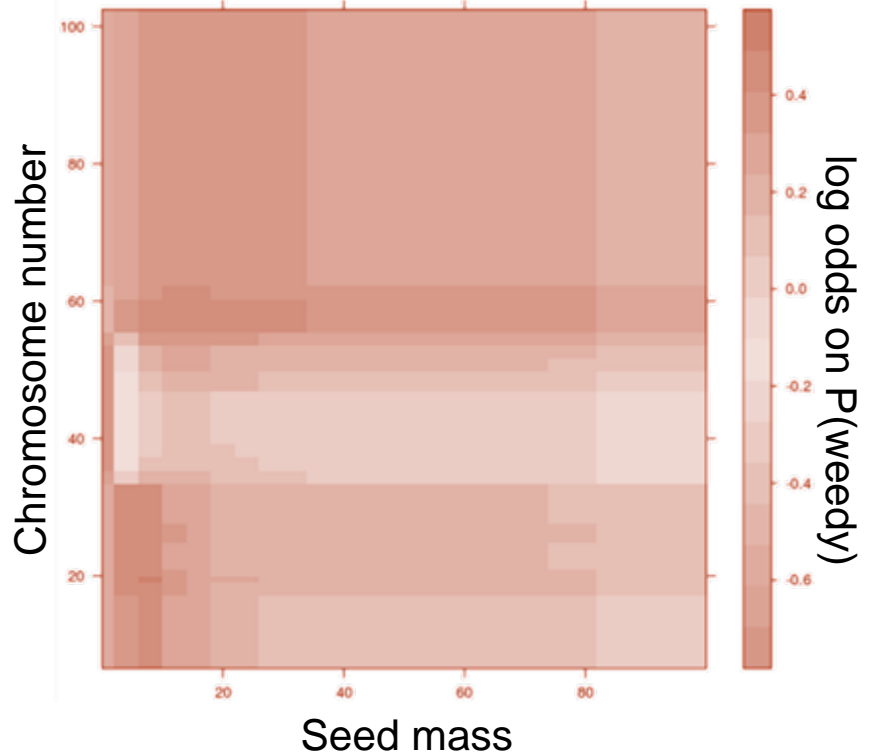
3-variable model

Variables

- seed mass
- chromosome number
- facultative wetland affinity

Performance on balanced withheld test dataset (n=385)

- Absolute accuracy: 72%
- 95% confidence interval: (66%,80%)



2-variable model

Variables

- seed mass
- chromosome number

Performance on balanced withheld test dataset (n=151):

- Absolute accuracy: 74%
- 95% confidence interval: (62%,86%)

Construction of Decision Boundaries

3-variable model

Variables

- seed mass
- chromosome number
- facultative wetland affinity

Performance on balanced withheld test dataset (n=385)

- Absolute accuracy: 72%
- 95% confidence interval: (66%,80%)

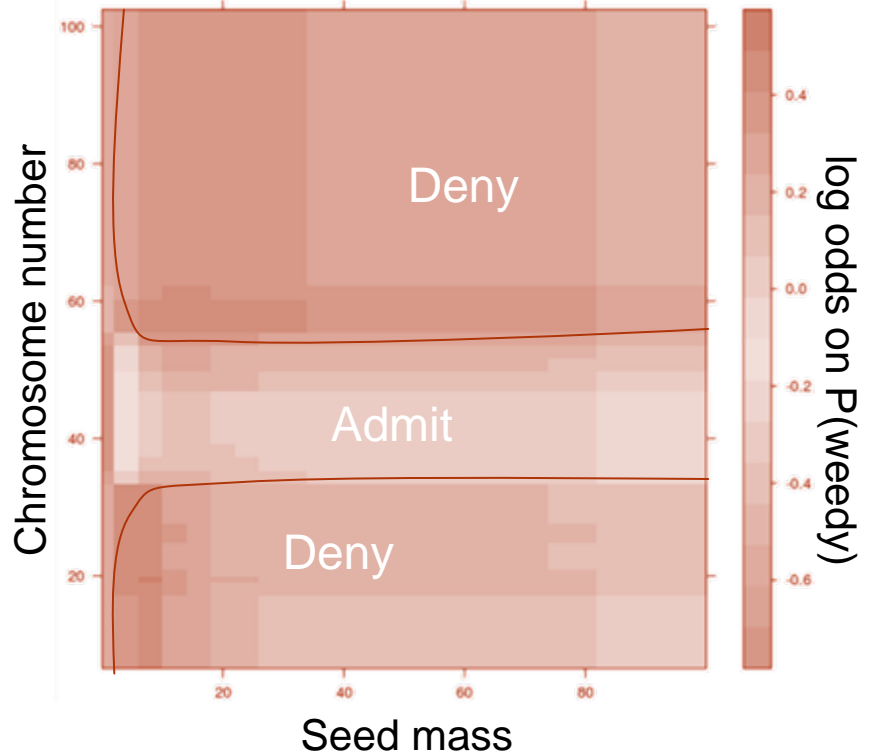
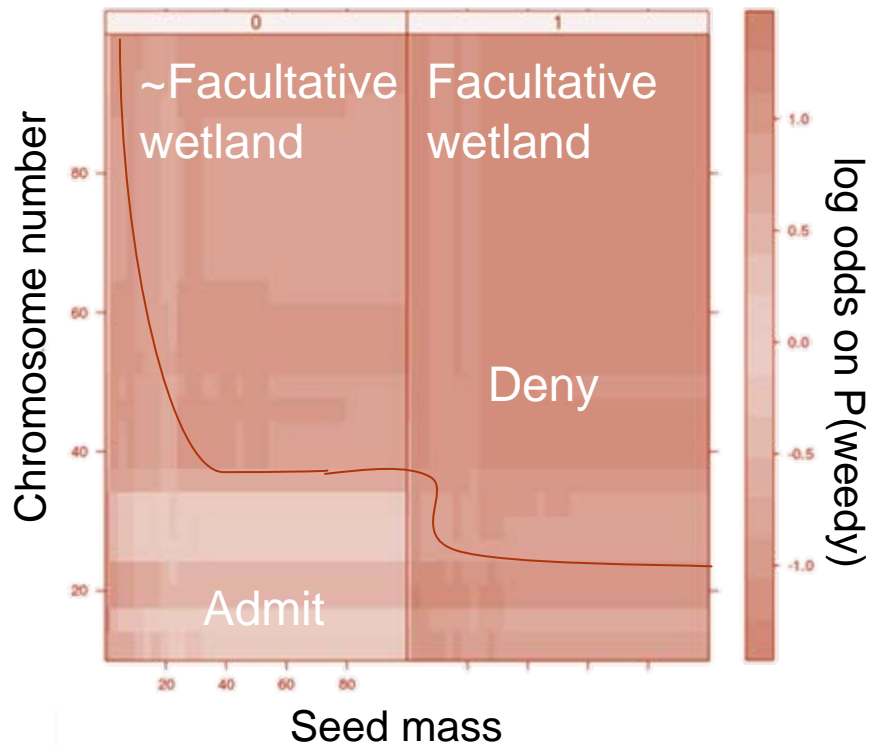
2-variable model

Variables

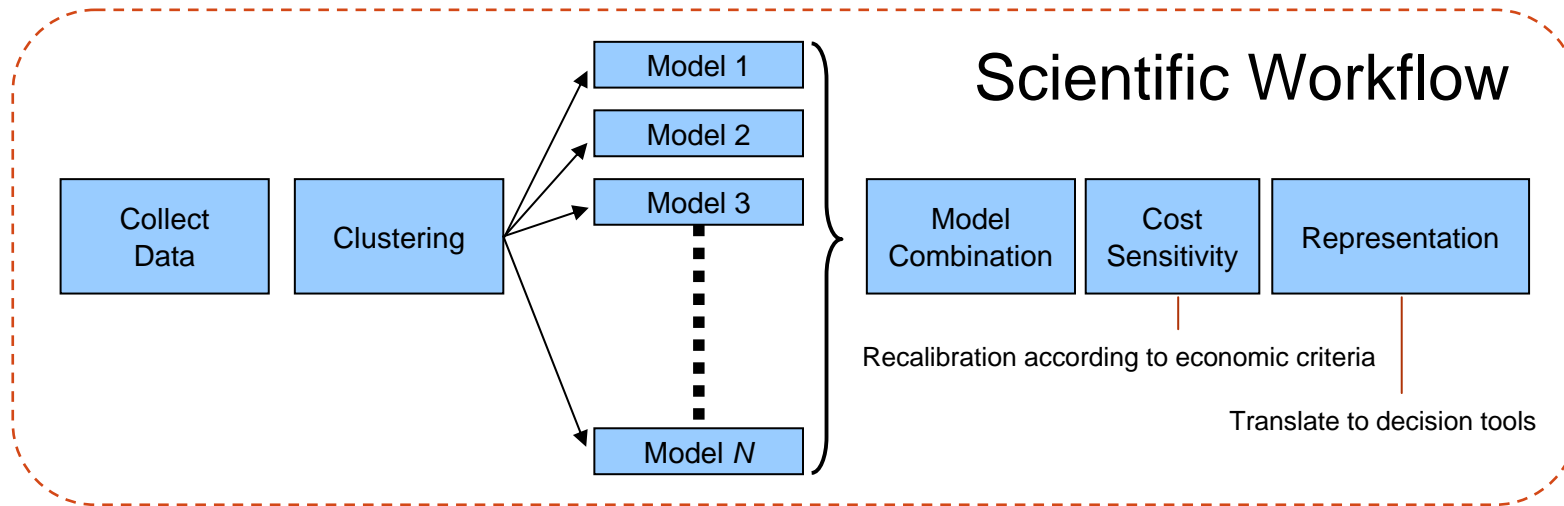
- seed mass
- chromosome number

Performance on balanced withheld test dataset (n=151):

- Absolute accuracy: 74%
- 95% confidence interval: (62%,86%)



Economic data



Cost/Benefit data on 311 pairs of species (one invasive, one non-invasive con-generic)

Cost of control: 16 spp

Marginal Cost: 1 sp.

Total Cost: 9 spp.

Benefit of control: 8 spp.

Economic benefit: 2 spp.

Construction of Decision Boundaries

3-variable model

Variables

- seed mass
- chromosome number
- facultative wetland affinity

Performance on balanced withheld test dataset (n=385)

- Absolute accuracy: 72%
- 95% confidence interval: (66%,80%)

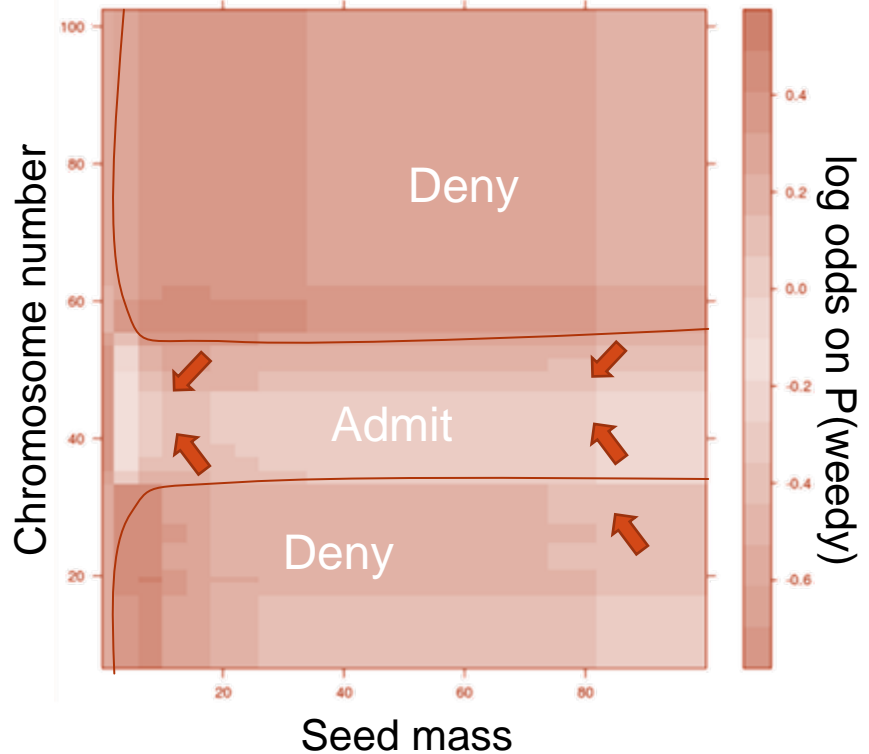
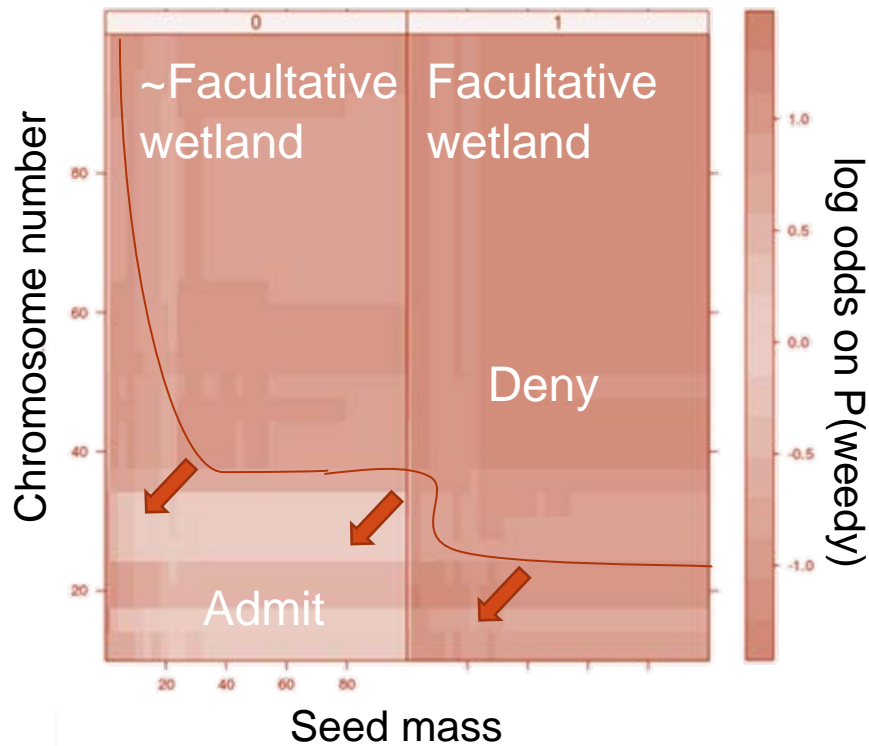
2-variable model

Variables

- seed mass
- chromosome number

Performance on balanced withheld test dataset (n=151):

- Absolute accuracy: 74%
- 95% confidence interval: (62%,86%)



Summary

- Findings to date
 - Invasive species are clustered within taxonomic groupings
 - Invasive propensity predictable to the level necessary to justify cost efficient trait-based admit/deny decision analysis
- Next Steps
 - Model calibration
 - Complete cost-benefit analysis and cost-sensitive learning
 - Visualization of more complex models