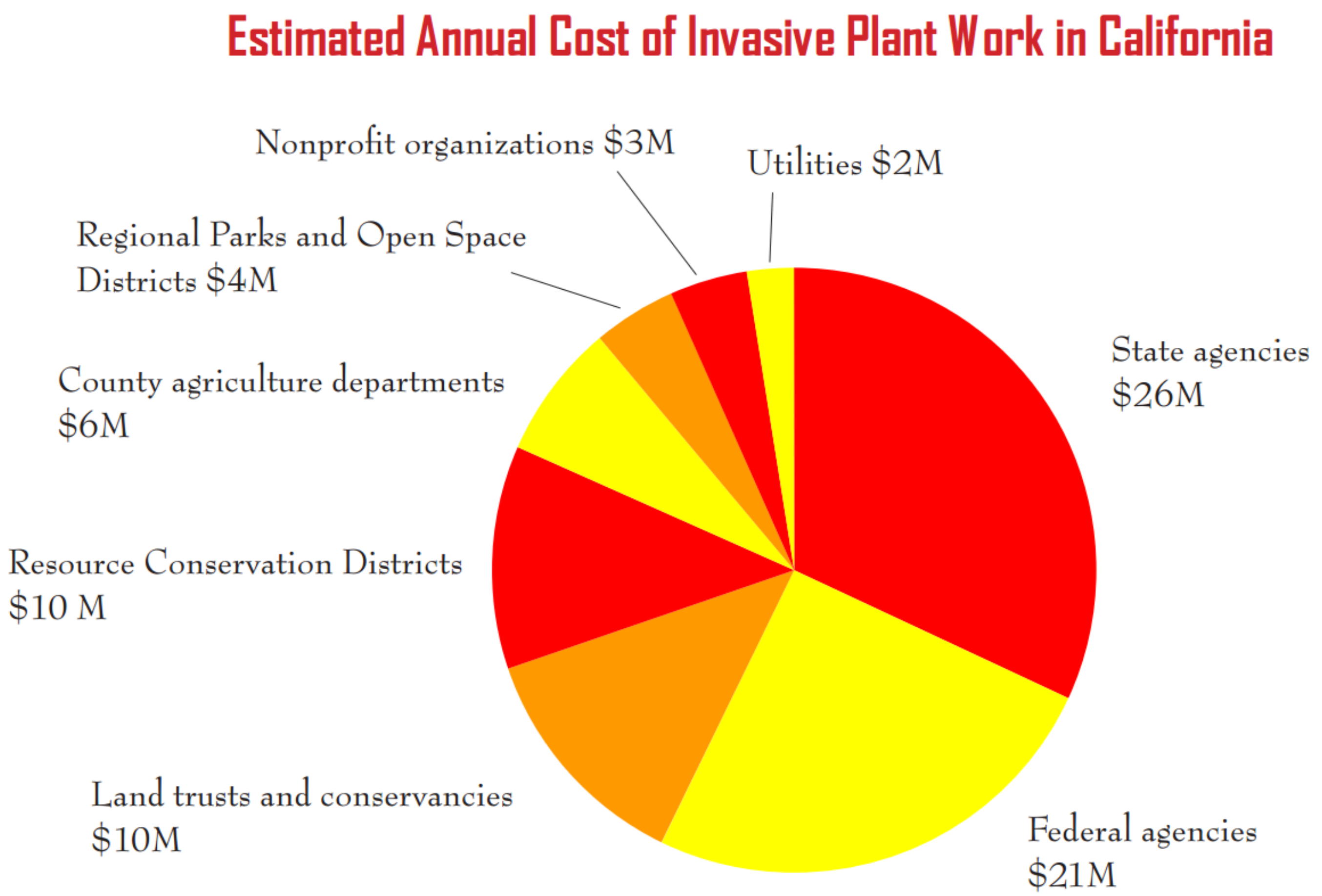
Computational botany for invasive species decision support, risk analysis, and policy

- John M. Drake & John Paul Schmidt (University of Georgia)
- David Finnoff (University of Wyoming)
- Mike Springborn (University of California Davis)

Odum School of Ecology
The University of Georgia

University of Wyoming
New Thinking

Invasive plants



Invasive plants cost the state of California >$82m annually

Image: California Invasive Plant Council

Civilian Conservation Corps workers planting Kudzu

25,000 non-indigenous plant species introduced in USA

- Only ~20% (*n*=5000) naturalized
- Only ~4.5% (*n=1,110*) are recognized as weeds
- Only ~1.7% (*n=435*) become pests

Leafy spurge *(Euphorbia esula)*

Pimental, D. et al. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics 52:273-288.

Overview

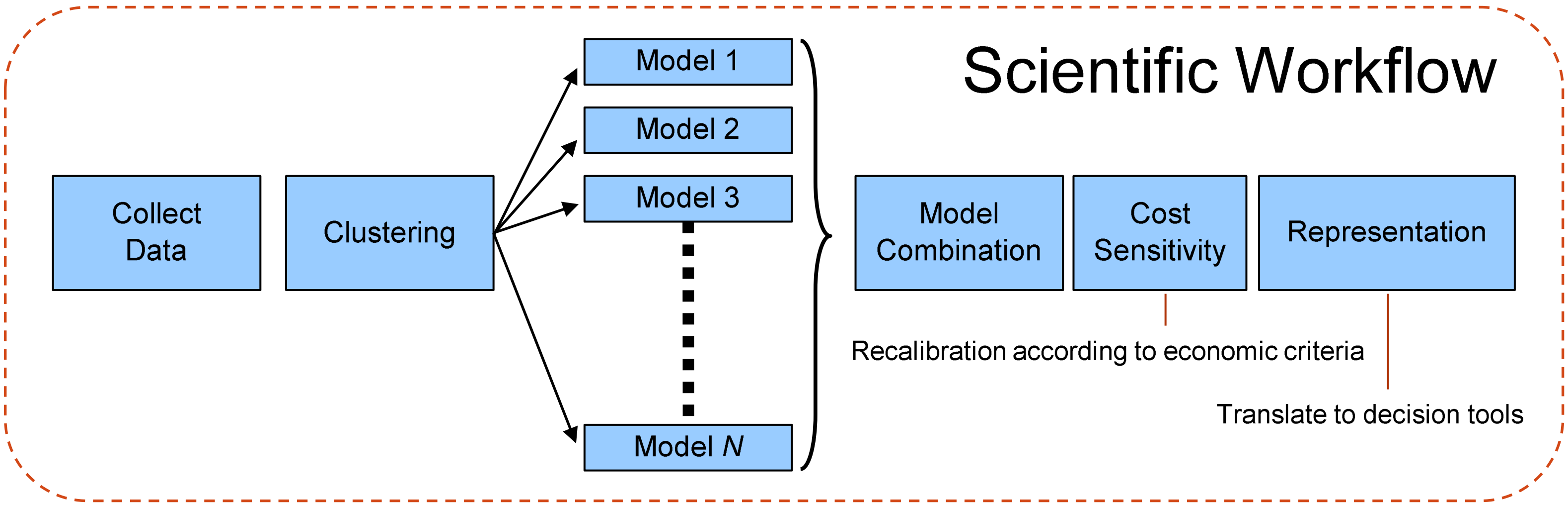
Problem

- 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

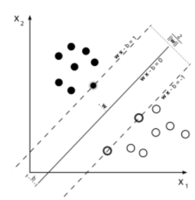
Objective

- 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

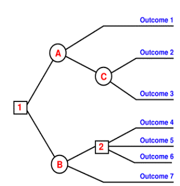
Scientific workflow



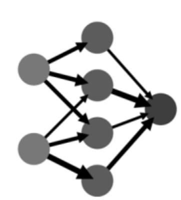
Relational database
•classes of invasive species
•species traits



Support vector machines



Regression trees



Neural nets

Relational database
•classes of invasive species
•revenue, damages, control costs

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

<u>145 Variables</u>

~10% relate to native distribution

~15% relate to introduced distribution

~70% relate to biological traits

<u>Large datasets</u>

Seed mass (>2000)

Chromosome number (>1000)

Maximum height (>500)

Biological traits

Growth form

Life history

Wetland habitat association

Maximum height

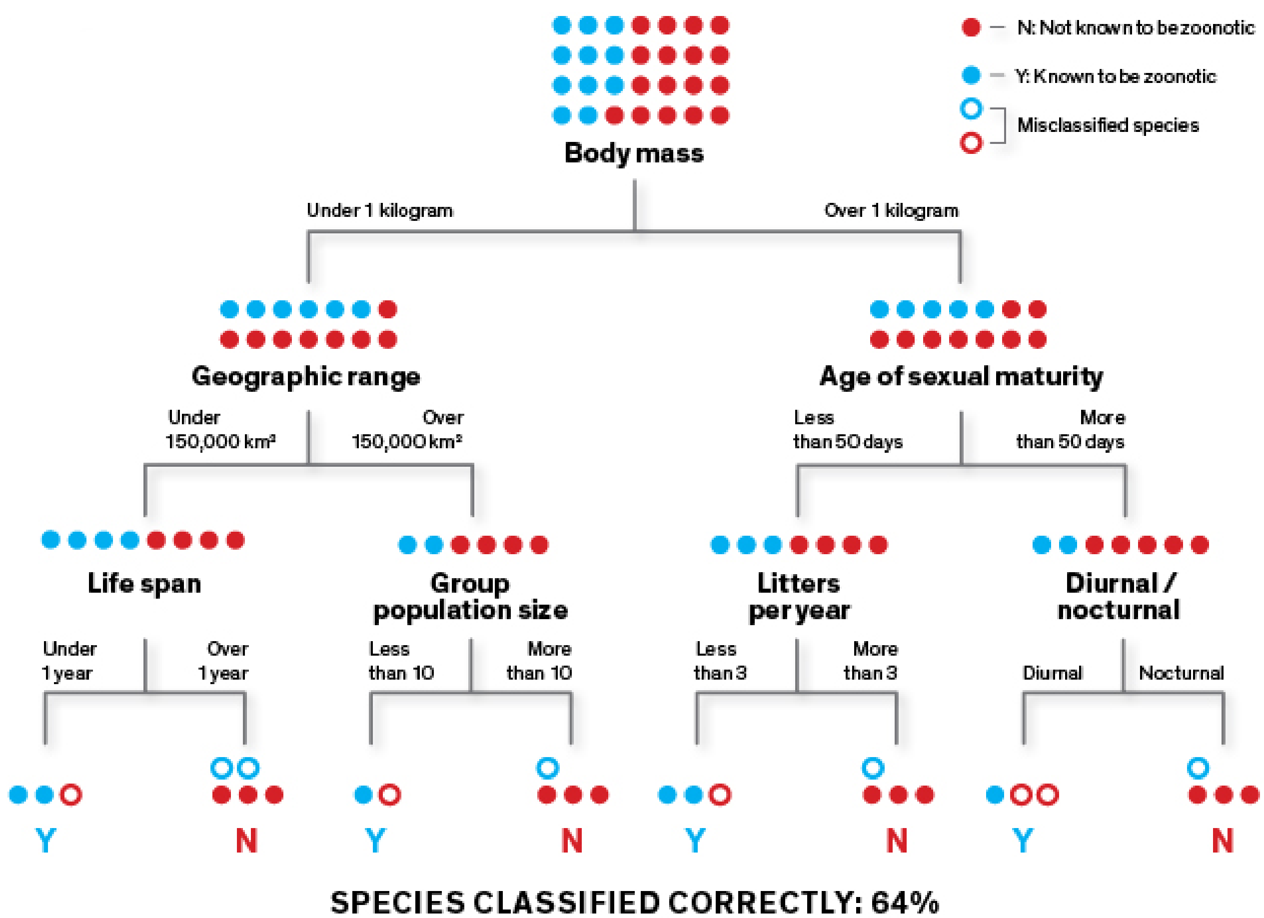
Maximum chromosome number/species

Seed mass

Leaf traits

Native latitudinal and longitudinal range

Machine learning with boosted regression trees



Pattern recognition algorithms based on ideas about human cognition

Image: Han, B. 2015. The algorithm that's hunting Ebola. IEEE Spectrum.

Species classification on a balanced dataset: Maximal performance

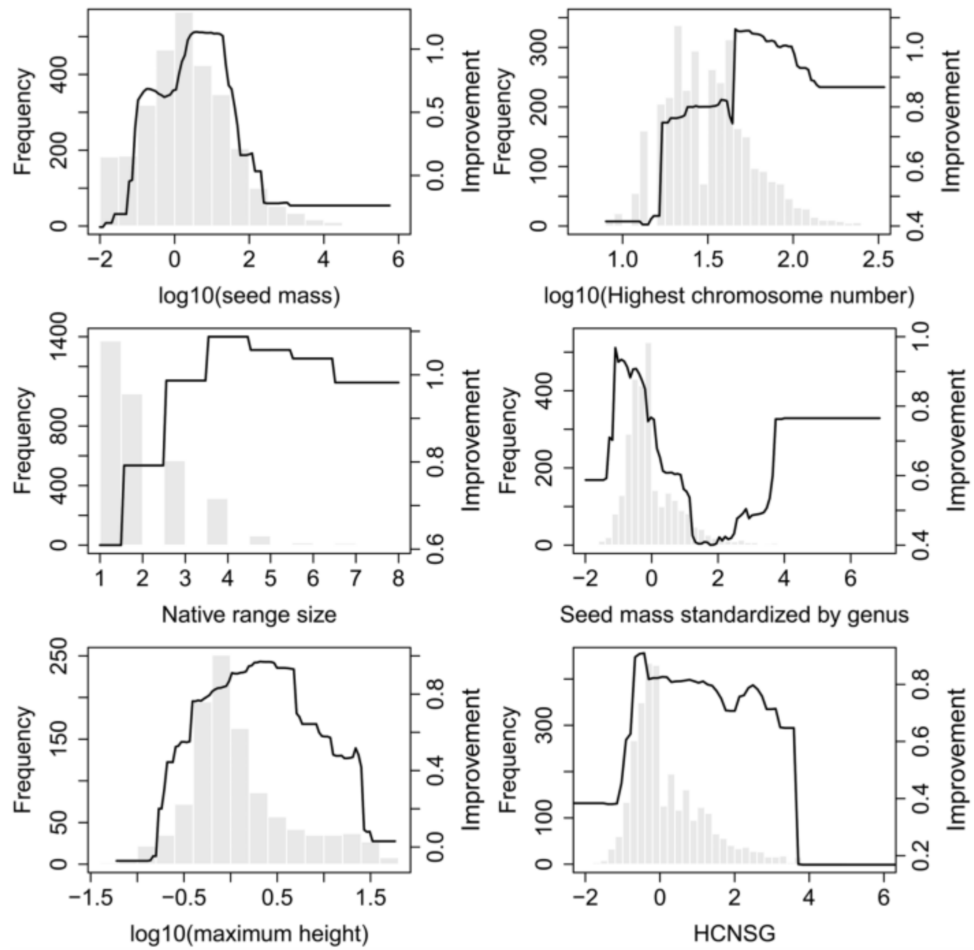
- Facultative wetland association
- Obligate wetland association
- Maximum height
- Seed mass
- Maximum chromosome number
- Leaf traits: evergreen-ness, C:N ratio, leaf specific area

Key biological traits

Performance Summary

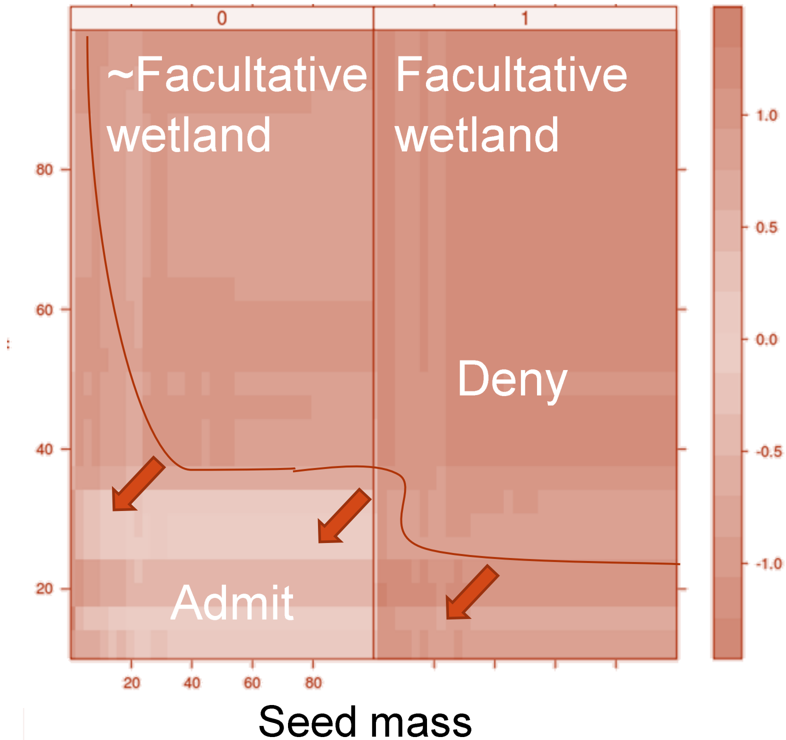
	Weeds		State-listed	
	prediction on 20% holdout (95% CI)	**sample size (test set)**	**prediction on 20% holdout (95% CI)**	**sample size (test set)**
all species	0.74 (0.69,0.78)	447	0.69 (0.62,0.76)	174
species present in > 1 state	0.78 (0.73,0.82)	385	0.78 (0.70,0.84)	149
ornamentals	0.85 (0.79,0.9)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

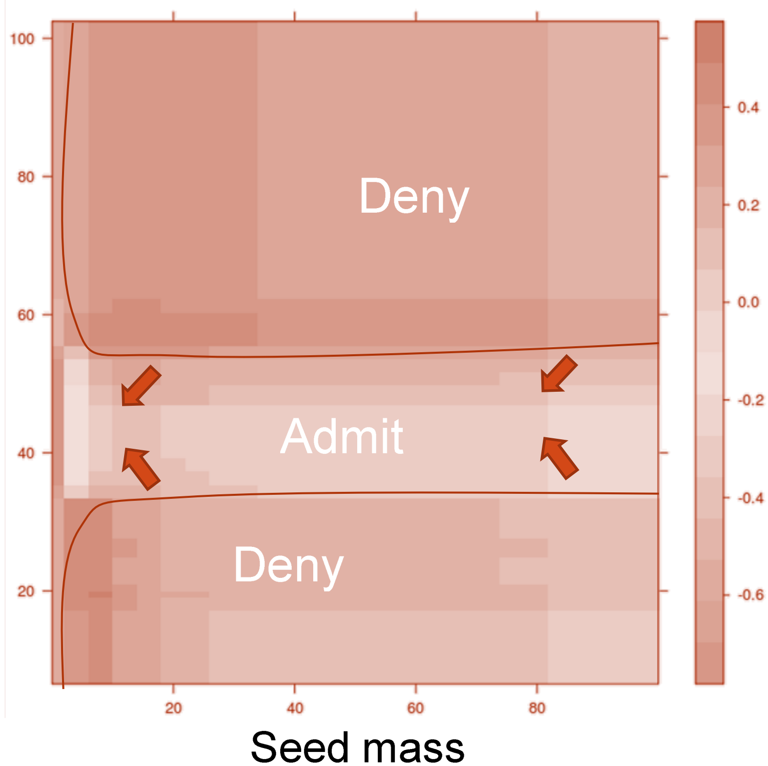
Permutation methods facilitate visualization of partial dependence



Computational methods diagnose how learned models make predictions

Schmidt, J.P., M. Springborn & J.M. Drake. 2012. Bioeconomic forecasting of invasive species by ecological syndrome. Ecosphere 3:46.





Construction of Decision Boundaries

3-variable model

Variables
- seed mass
- chromosome number
- facultative wetland affinity

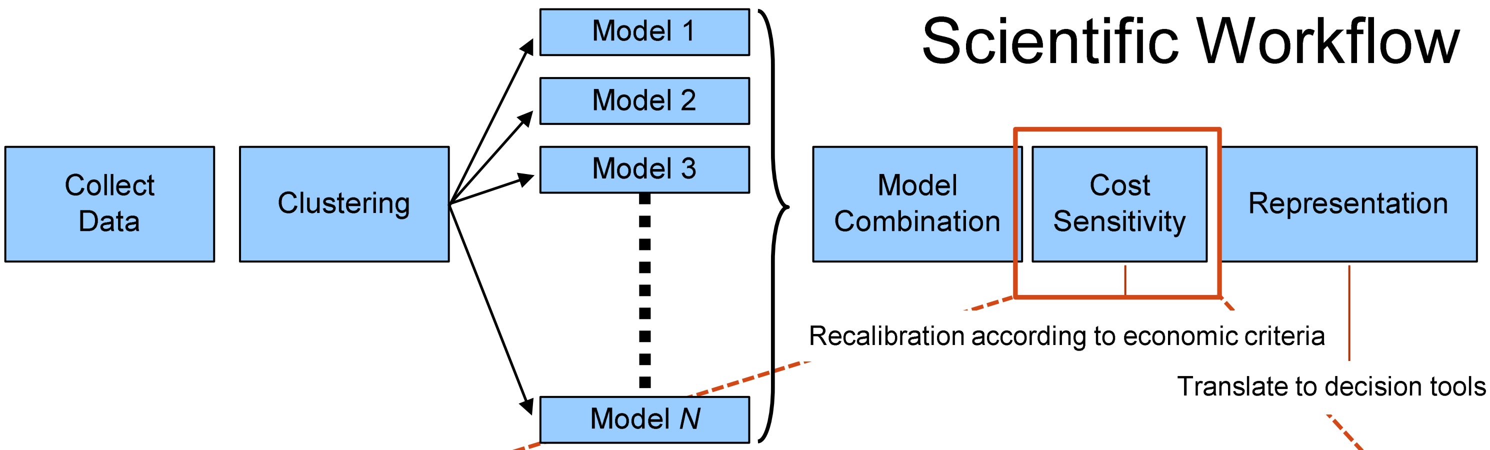
Performance on balanced withheld test dataset (n=385)
- Absolute accuracy: 72%
- 95% confidence interval: (66%,80%)

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



Need to know the costs ("expected net benefits") of invasive species

Expected net benefits per species assessed

$$ENB = \underbrace{\pi}_{\substack{\text{Prob. that a}\\ \text{species is}\\ \text{an invasive}}} \cdot \underbrace{TPR}_{\substack{\text{Prob. that an}\\ \text{invasive is so}\\ \text{classified}}} \cdot \underbrace{\left(V_L - V_T\right)}_{\substack{\text{Benefit}\\ \text{of excluding}\\ \text{an invasive}}} - \underbrace{(1-\pi)}_{\substack{\text{Prob. that}\\ \text{a species}\\ \text{is benign}}} \cdot \underbrace{FPR}_{\substack{\text{Prob. that a}\\ \text{benign species}\\ \text{is mistakenly}\\ \text{classified}}} \cdot \underbrace{V_T}_{\substack{\text{Cost of}\\ \text{excluding a}\\ \text{benign species}}}$$

$$\propto \pi \cdot TPR \cdot \left[\frac{V_L - V_T}{V_T}\right] - (1-\pi) \cdot FPR.$$

$V_t > 0$: Per species expected benefit of traded import
$V_l > 0$: Per species expected losses conditional on the species truly belonging to the invasive class
π: Base frequency of invasive species
TPR: True positive rate of proposed policy
FPR: False positive rate of proposed policy

Estimates

- π=4.40% (weeds)
- π=1.74% (noxious)

Estimated as fraction of introduced species (~25,000)

Following Feenstra (2004), calculated from import demand estimated from US Dept. Commerce Global Ag. Trade System (GATS) data on 8 sub-groups within the plants-for-planting collection of goods

- V_t=\$281,200 (low)
- V_t=\$410,100 (high)

Benchmarked at value where decision-maker indifferent to randomly selected species

- V_l= ranges from \$6m to \$24m (conservative)
- V_l= ranges from \$630m to \$1.6b (pub'd)

Based on controversial published tallies

Expected net benefits under different modeling assumptions

Lower bound

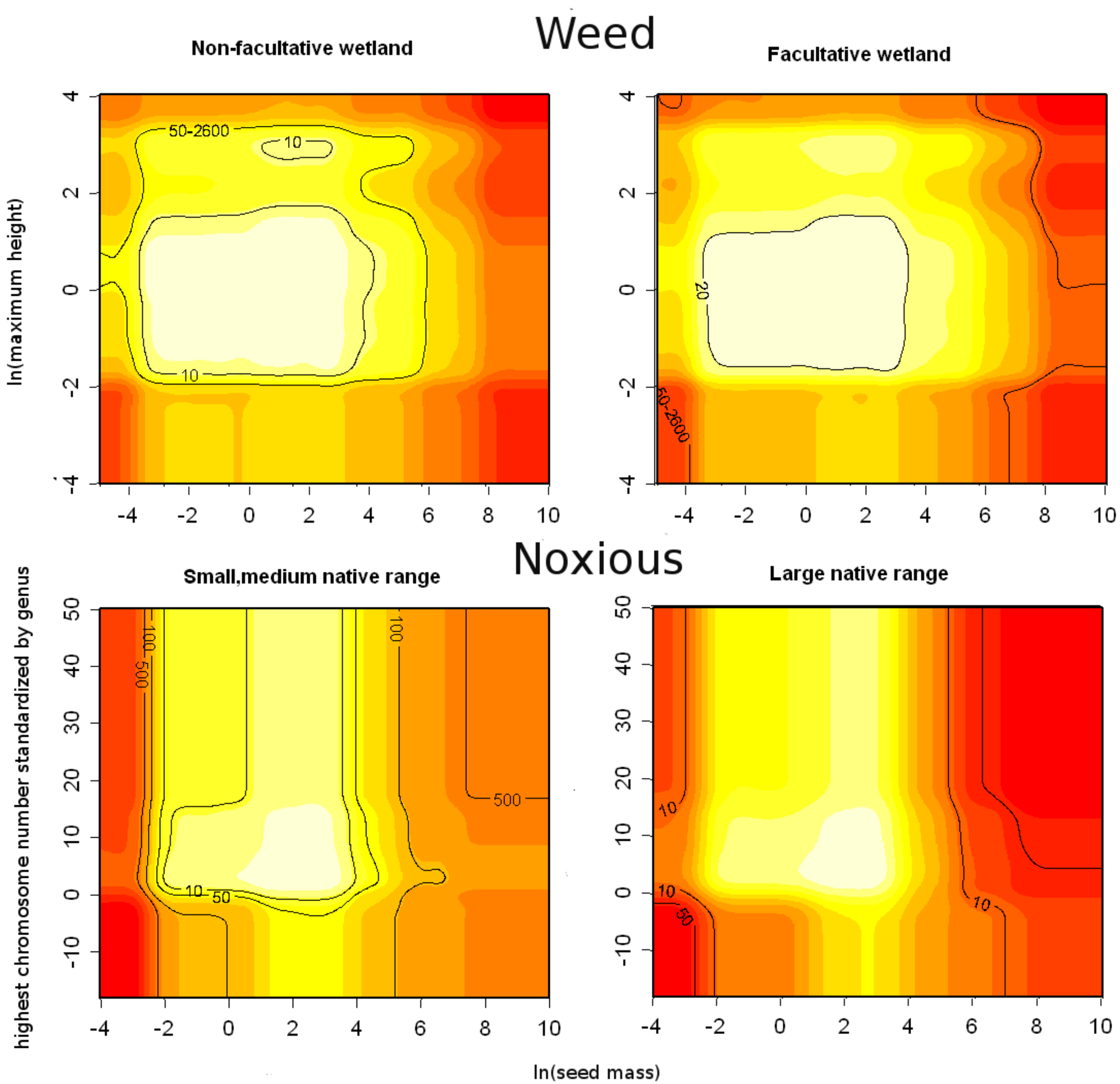
		V_T	V_L (minimal)	π	φ	TPR	FPR	ENB
Noxious	High	$410,100	$23,568,966	0.0174	56	0.94	0.66	$110,000
Noxious	Low	$281,200	$16,160,920	0.0174	56	0.94	0.66	$80,000
Weed	High	$410,100	$9,320,455	0.044	22	0.59	0.23	$140,000
Weed	Low	$281,200	$6,390,909	0.044	22	0.59	0.23	$100,000

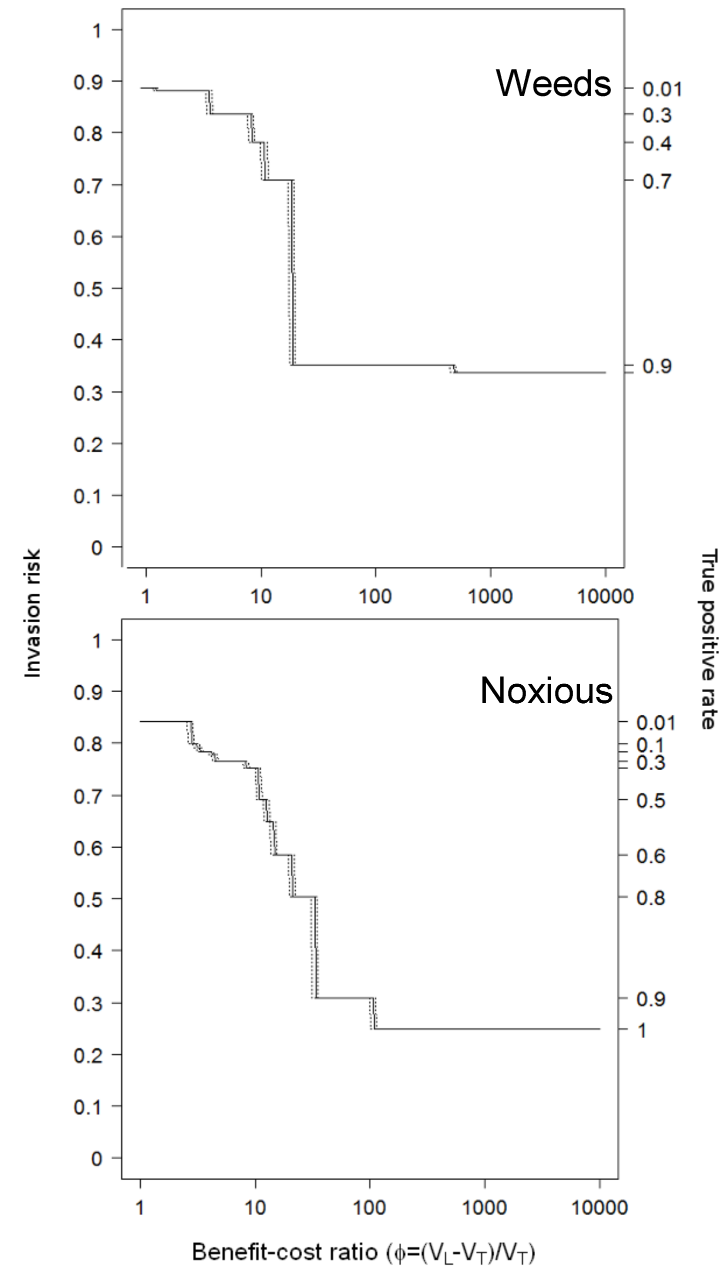
Upper bound

		V_T	V_L (pub'd)	π	φ	TPR	FPR	ENB
Noxious	High	$410,100	$1,593,563,218	0.0174	5666	1	0.98	$27,000,000
Noxious	Low	$281,200	$1,593,563,218	0.0174	3885	1	0.98	$27,000,000
Weed	High	$410,100	$630,181,818	0.044	2240	0.99	0.58	$27,000,000
Weed	Low	$281,200	$630,181,818	0.044	1536	0.99	0.58	$27,000,000

Schmidt, J.P., M. Springborn & J.M. Drake. 2012. Bioeconomic forecasting of invasive species by ecological syndrome. Ecosphere 3:46.

Main result





Schmidt et al. (2015)

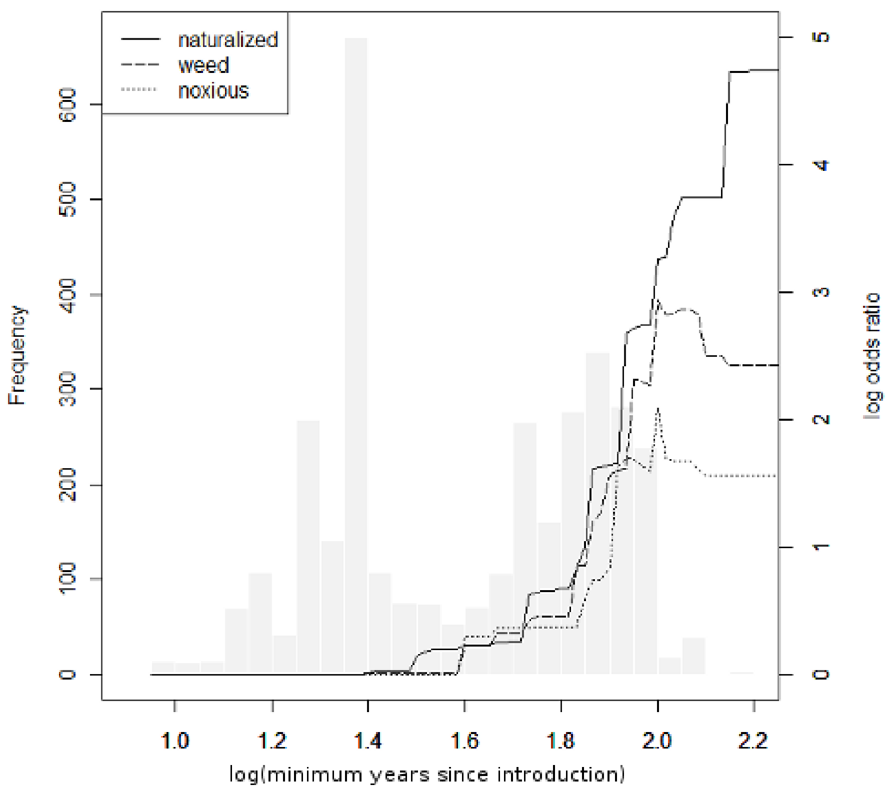
OPEN ACCESS Freely available online PLoS one

Time since Introduction, Seed Mass, and Genome Size Predict Successful Invaders among the Cultivated Vascular Plants of Hawaii

John Paul Schmidt*, John M. Drake

Odum School of Ecology, University of Georgia, Athens, Georgia, United States of America

Minimum time since introduction



	1840-1999	1840-1930
Total introduced	7,866	815
Naturalized	420 (5.3%)	252 (31%)
Weed	141 (1.8%)	90 (11%)
Noxious (HI)	39 (0.5%)	22 (3%)

Table 1. Model performance measured by area under the ROC curve (AUC) values for models of the invader classes as a function of key predictors.

model	model performance (AUC)			num. species
	naturalized	weed	noxious	
full model	0.92	0.91	0.88	4861
years since introduction	0.82	0.80	0.76	3460
seed mass + HCNSG	0.76	0.75	0.82	3180* (864)
HCNSG	0.69	0.68	0.80	2009
seed mass	**0.71**	**0.71**	**0.74**	**2077**

*species for which data contains values for either term, number of species with values for both in parentheses.

doi:10.1371/journal.pone.0017391.t001

Summary

- Findings
  - Some genera disproportionately represented in invasive species (rationale for regulation at higher taxonomic levels)
  - Species specific rapid screening protocol now available with conservatively estimated benefits of $80,000 to $140,000 per species assessed

    http://daphnia.ecology.uga.edu/drakelab/?p=306

  - Lag times of 10-100 years between introduction and realization of costs

- Next Steps
  - Development of an iPad app for rapid risk screening
  - The genomics of invasiveness