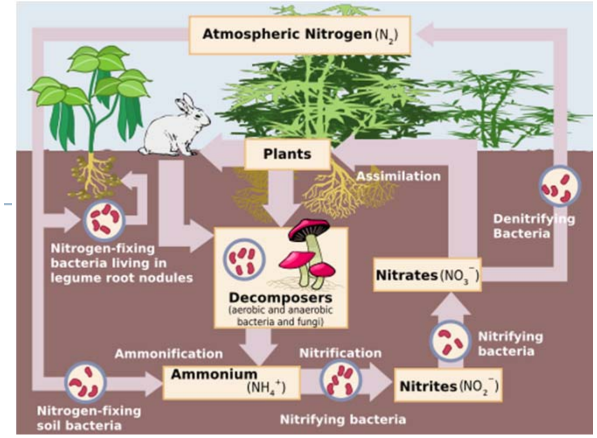
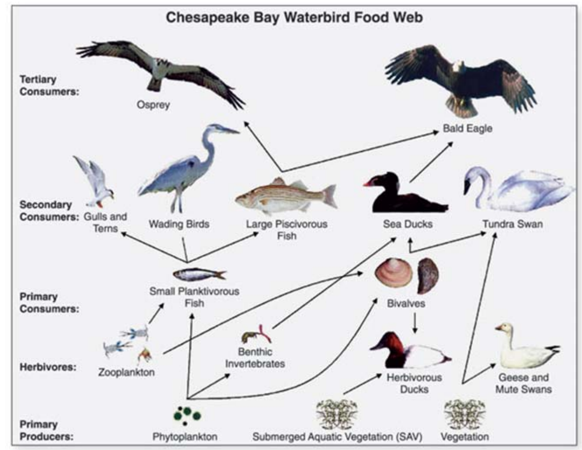
Machine Learning for Ecological Science and Environmental Policy

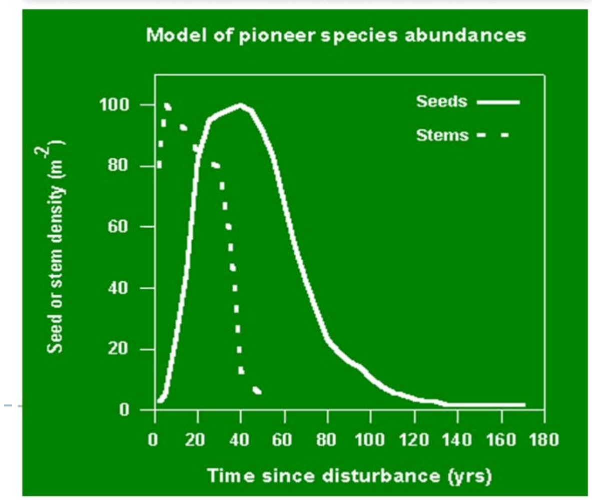
Tom Dietterich, Rebecca Hutchinson, Dan Sheldon
Oregon State University

Introduction

- Ecological Science
  - Processes governing the function and structure of ecosystems
    - Flows of energy and nutrients
      - Sunlight, water, carbon, nitrogen, phosphorus
    - Species distribution and interaction
      - Reproduction, Dispersal, Migration, Invasion
      - Competition, Food Webs, Mutualism
    - Non-equilibrium systems: Continual disturbance and system resilience
      - Many species depend on disturbances



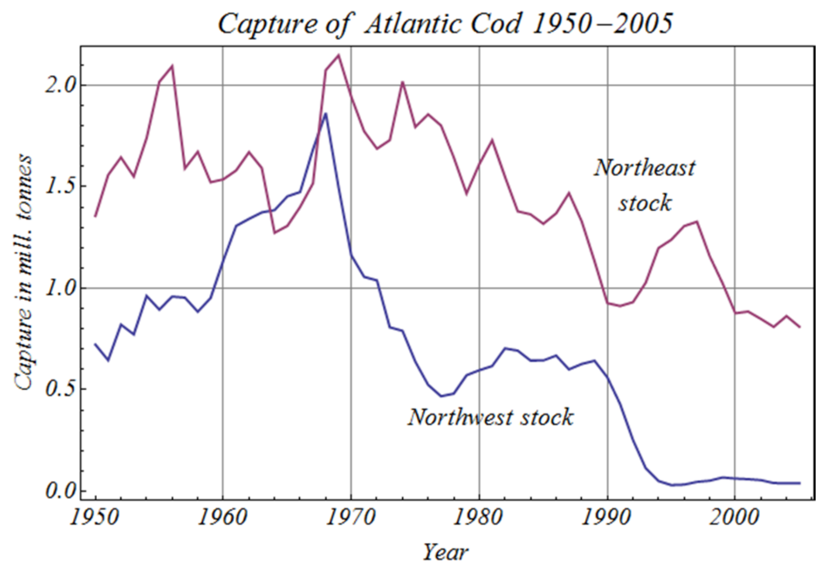




Wikipedia; http://www.plantbio.uga.edu/~chris/wind.html

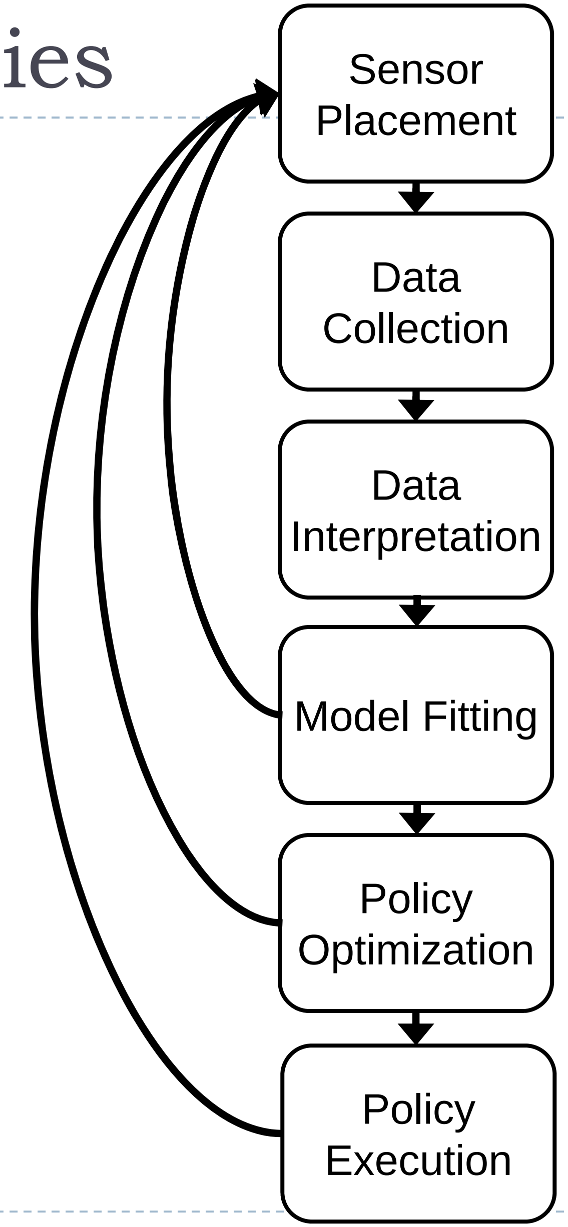
Introduction

- Environmental Policy
  - Natural resource management
    - Fisheries
    - Forestry
    - Water resources (rivers, aquifers, estuaries)
  - Conservation biology
    - Reserves and conservation easements
    - Endangered species
    - Endangered ecosystems
    - Invasive species management



Data → Models → Policies

- Data Acquisition
  - Sensor Placement
  - Data Interpretation
- Model Fitting
  - Species Distribution Models
  - Dynamical Models
- Policy Optimization
  - MDPs
  - POMDPs
  - Network cascades



Unique Aspects

- Heterogeneity
  - Physical quantities (nutrients, temperature, wind)
  - Organisms and species (viruses, bacteria, fungi, plants, animals)
  - Spatial Scale (inside a single organism, watershed, continent, planet)
- Hidden dynamics
  - Virtually all interactions are not directly observed
  - Observations are noisy and incomplete
  - Most movement (dispersal, migration) is not directly observed
  - Non-stationary dynamics: climate change, land-use change, evolution
- Optimization wrt learned dynamic models
  - Large spatio-temporal MDPs
  - Essential POMDPs
  - Need for robust solutions
    - poorly-modeled dynamics
    - politics

Goals for the Tutorial

- Review the primary data sources, model types, and machine learning problems that arise in ecological science and environmental policy
- Provide examples of current ML work in each of these areas
- Point out open problems and opportunities for additional ML contributions
- Provide pointers to data sets and relevant literature

Outline

- Data Acquisition
  - Sensors: Physical sensors, human observers, repurposing data from other sources
  - Data interpretation: Extracting signals from data
- Ecological Models
  - Species Distribution Models
  - Dynamical Models: Dispersal, Migration, Invasion, Climate Change
- Policy Optimization
  - Conservation: Reserve design, Network design
  - Invasive species: Eradication, restoration, monitoring
  - Fisheries: Managing harvest levels

Part 1: Data Acquisition

- Data Sources
  - Instruments placed in the environment:
    - Weather stations: temperature, wind direction, wind speed, solar radiation, relative humidity, snow depth, precipitation
    - Stream gauging stations: water flow rate, temperature, height
    - Isotope and dye studies: Carbon, Nitrogen, Phosphorous
    - RFID tags: Fish
    - Radio collars: mammals and birds
    - Acoustic monitoring
      - Birds, insects, bats, whales

SensorScope

wildlifeacoustics.com

biomark.com

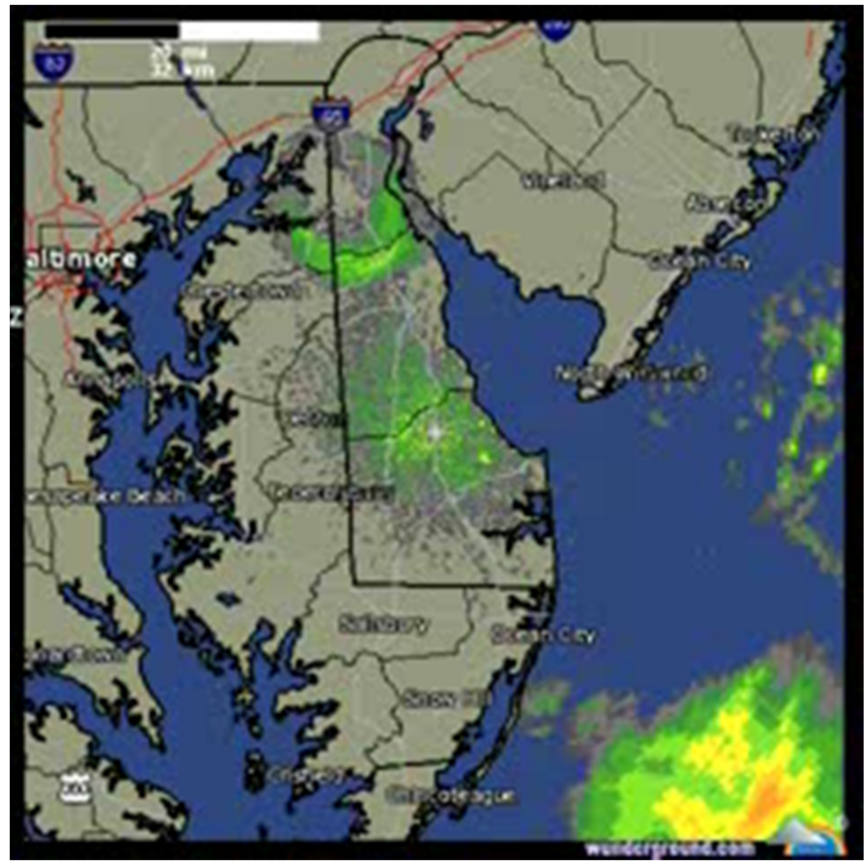
atstrack.com

Data Acquisition: Human Observation

- Trapping and identification
  - Insect traps: emergence, malaise, UV light
  - Electro-fishing
  - Kick nets
- Volunteers
  - Bird sightings
  - Whale observations

Data Acquisition: Repurposing Data Gathered for Other Purposes

- Repurposing information gathered for other purposes
  - Fish catch data
  - Doppler weather radar



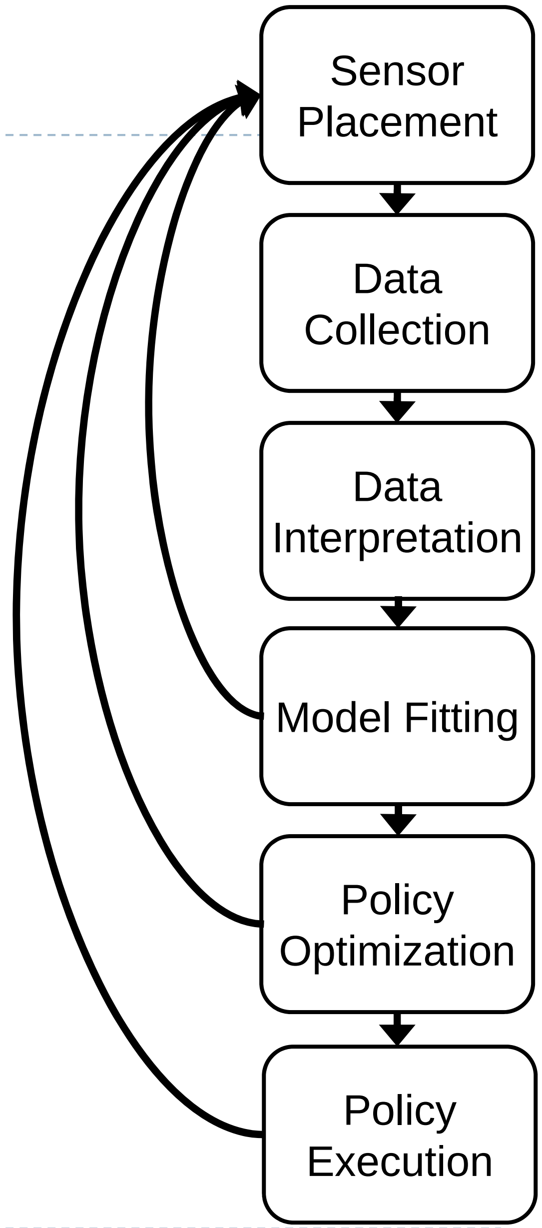
Data Acquisition: Remote Sensing

- Satellite-borne Sensors
  - Landsat 7
    - 15m resolution; whole planet coverage every 16 days
  - MODIS
    - 250m-1km resolution; whole planet coverage every 1-2 days

Sensor Placement

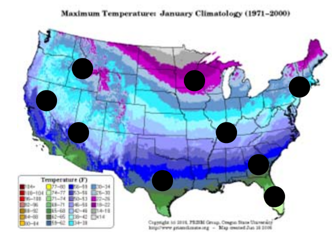
- Where should we place sensors to gain the best information for...
  - improving our models
  - improving our policies
  - guiding policy execution

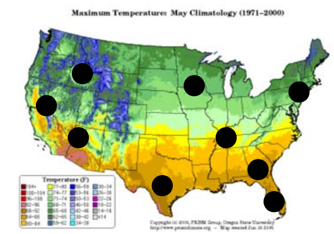
- Related questions in ML
  - Active Learning
  - Exploration in RL
  - Optimal POMDP policies

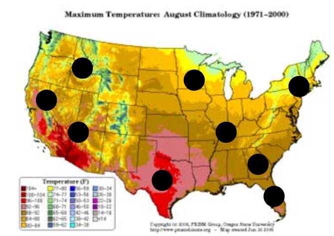


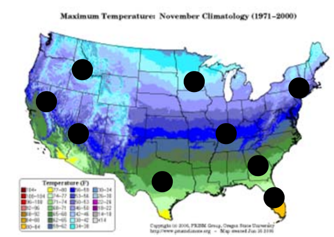
Basic Case: Spatial Sensor Placement for Real-Valued Function f [Krause, et al., 2008]

- Examples: Temperature, Rainfall, Nutrient Density, Pollutant Density
  - Goal: At each time t, we will observe the sensor readings (at the chosen locations) and estimate the complete spatial map of the target function f
- Given:
  - Initial model or initial set of observations
  - Budget: # of sensors k
- Find:
  - Locations at which to put the sensors in order to best estimate the function at future times



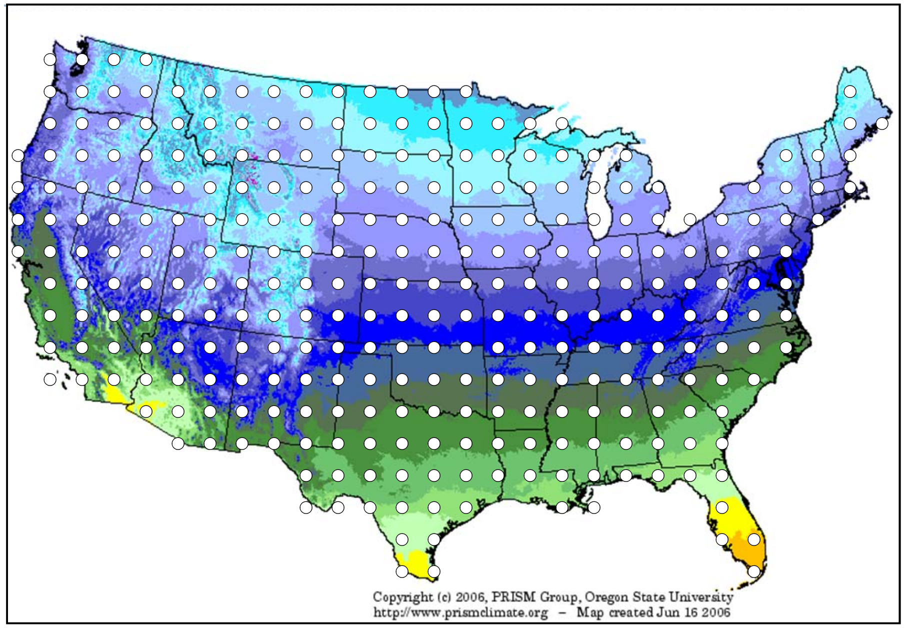






Approach

- Discretize space:
  - Let S be a set of points $(s_1, \ldots, s_N)$
    - where sensors can be placed
    - where we will make predictions
- Assume joint Gaussian
  - $(f(s_1), \ldots, f(s_N)) \sim \mathrm{Norm}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
    - where $\boldsymbol{\mu}$ has dimension N and
    - $\boldsymbol{\Sigma}$ has dimension $N \times N$
- Use the initial observations to estimate $\boldsymbol{\Sigma}$
- Choose an objective function $J(A)$ for evaluating the quality of a set of sensor locations A
- Formulate an optimization problem to choose a set $A \subset S$ of size k that optimizes $J(A)$.



What Criterion to Optimize?

- Mutual information between sensed and un-sensed locations:

$$J(A) = H\left(X_{S\setminus \mathrm{A}}\right) - H(X_{S\setminus \mathrm{A}} | X_A)$$

  where $H(X)$ is the entropy of the joint distribution $P(X)$

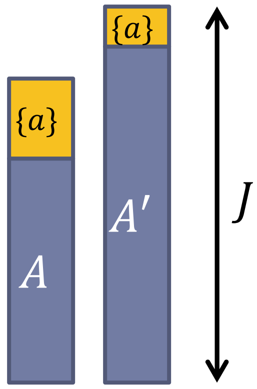
  - J measures the reduction in uncertainty about $S\setminus\mathrm{A}$ given the observations at A.
  - Easy to compute for multivariate Gaussians

- Rationale:
  - empirical: gives good results
  - computational/analytical: objective is sub-modular
    - Greedy Algorithm with provable bounds

- Submodularity:
  - J is submodular if forall $A \subseteq A'$ and all $a \in S\setminus\mathrm{A}'$,

$$J(A \cup \{a\}) - J(A) \geq J(A' \cup \{a\}) - J(A')$$

  "diminishing returns of adding a"



Krause, Singh & Guestrin, 2008

Greedy Algorithm

- Input:
  - Sites: S
  - Number of sensors: k
  - Estimated covariance matrix of joint Gaussian: $\mathbf{\Sigma}$
- Output: sensor locations $A \subset S, |A| = k$
- begin
  - $A \leftarrow \emptyset$
  - for $j = 1$ to k do
    - $a^* \leftarrow \underset{a \in S \setminus \mathrm{A}}{\operatorname{argmax}} \; J(A \cup \{a\})$
    - $A \leftarrow A \cup \{a^*\}$
- end

Analytical Bound

- Monotonicity assumption:

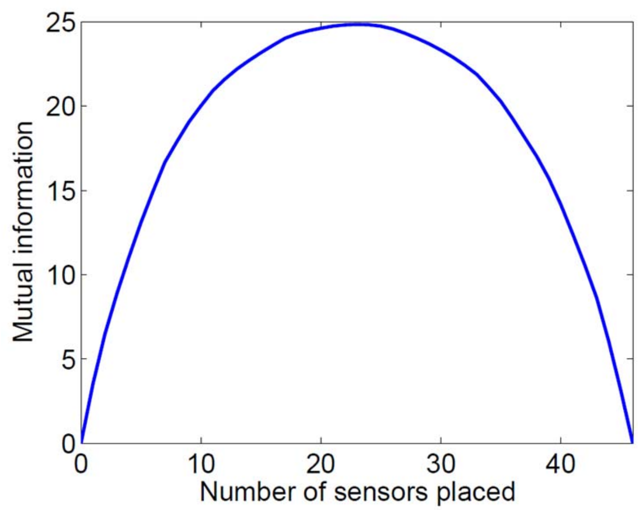
  $\forall\, a \in S \backslash A \quad J(A \cup \{a\}) > J(A) + \epsilon$

- Let $\hat{A}$ be the greedy solution and A^* be the optimal solution

  $$J(\hat{A}) \geq \left(1 - \frac{1}{e}\right) J(A^*) - k\epsilon$$

  - $\left(1 - \frac{1}{e}\right) \approx 0.632$

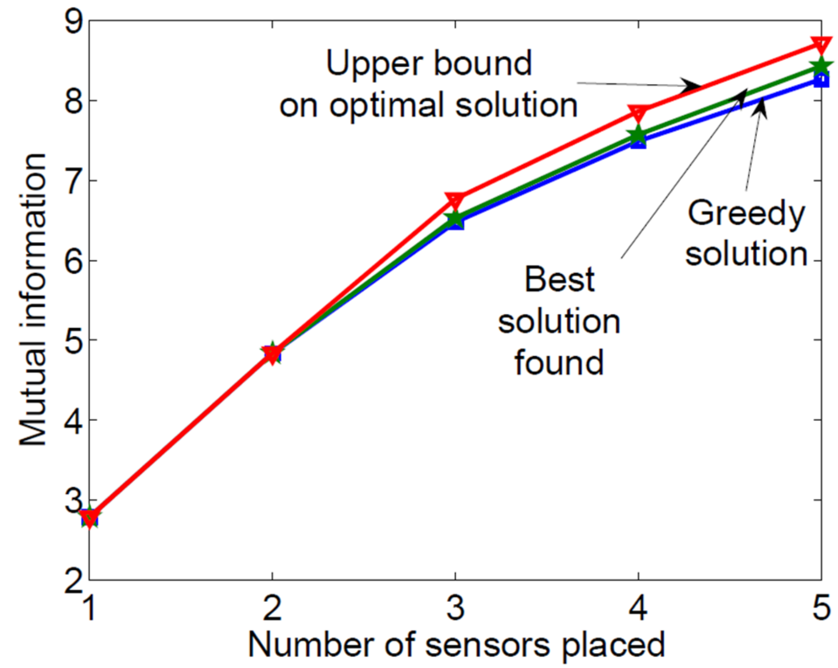
- Assumption will hold if S is discretized sufficiently finely



Krause, Singh & Guestrin, 2008

Experimental Accuracy

- Theoretical bound is 63.2% of optimal
- Greedy algorithm is closer to 95% of optimal in this case



Intel Berkeley Temperature Sensors

Krause, Singh & Guestrin, 2008

Further Work

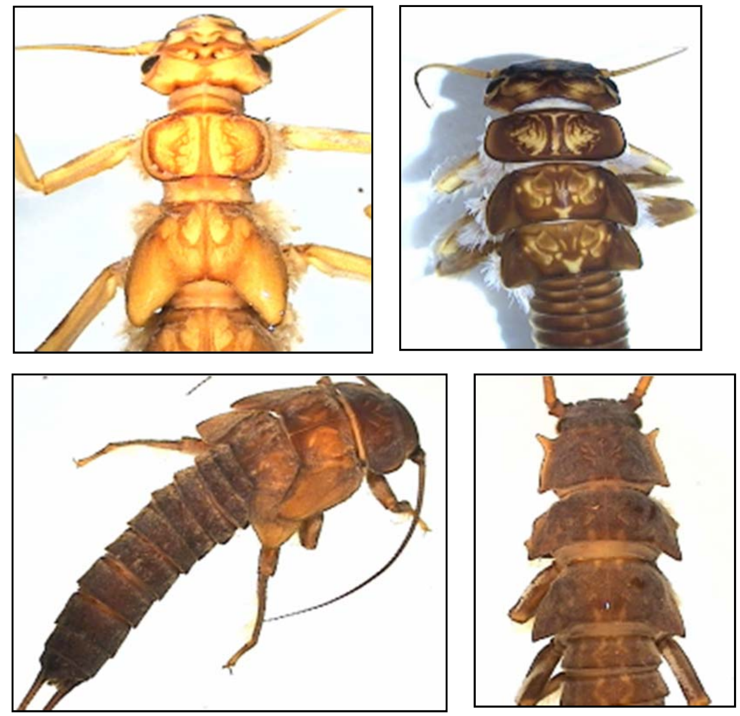
- Extensions:
  - Differential costs
  - Cost of wireless network communications
  - Robustness to failed sensors
  - Robustness to uncertainty in initial model of Σ

Data Interpretation

- Extracting high level interpretation from low-level sensor data
  - Example 1: Arthropod Population Counting
  - Example 2: Finding Swallow Roosts in Doppler Weather Radar

Arthropod Population Surveys

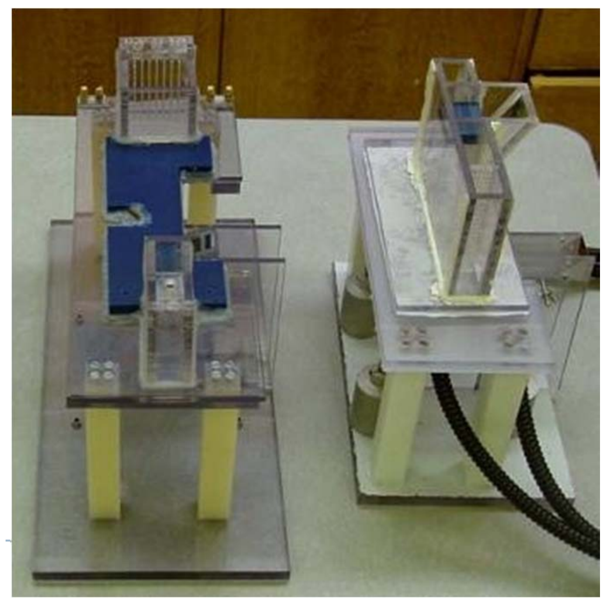
- Arthropods are a powerful data source
  - Found in virtually all environments
    - streams, lakes, oceans, soils, birds, mammals
  - Provide valuable information on ecosystem function
    - Standard tool for evaluating stream health in EPA biomonitoring and stream restoration efforts
- Problem: Identification is time-consuming and requires hard-to-find expertise
- Solution: Combine robotics, computer vision, and machine learning to automate classification and population counting



OSU BugID Project

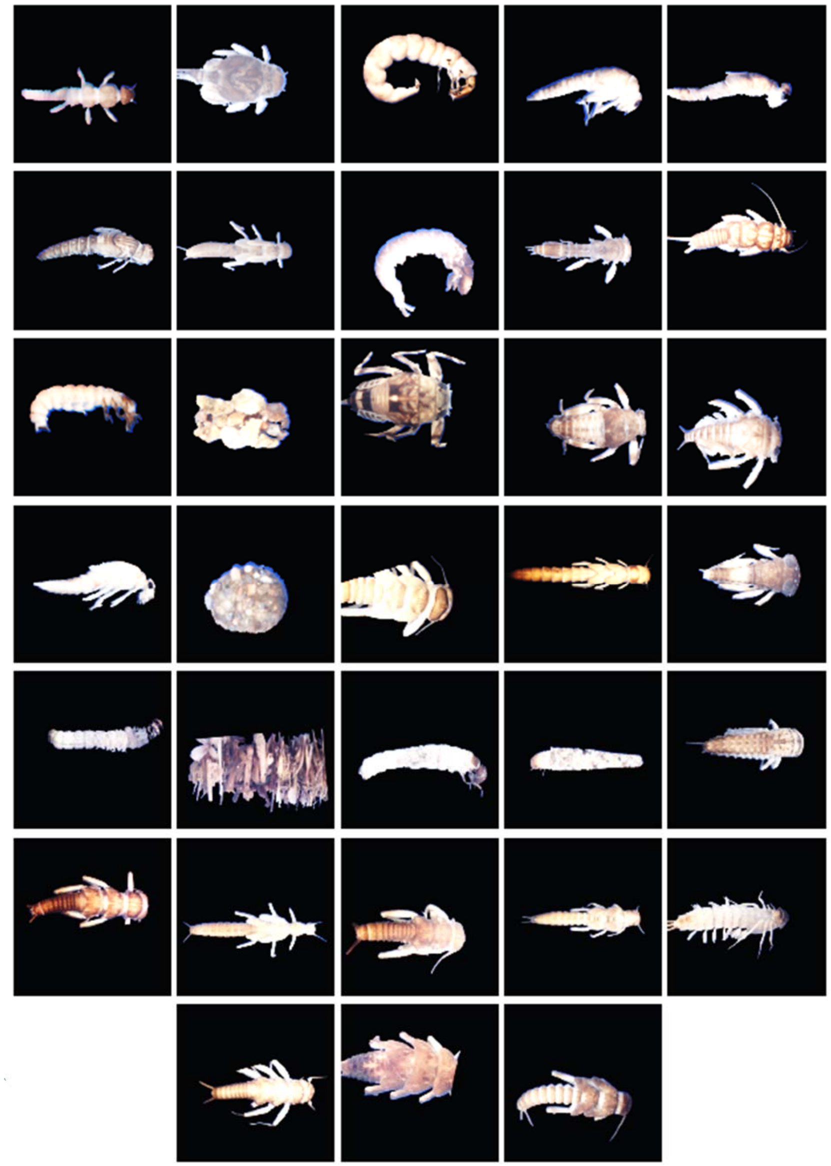
- Human technician gathers field sample
- Semi-automated image capture
- Automated classification

www.epa.gov



Computer Vision Challenge: Fine-Grained Classification

- Challenges:
  - Many classes
  - Subtle differences between classes
  - Wide variety of poses
  - Substantial size and appearance variation within class



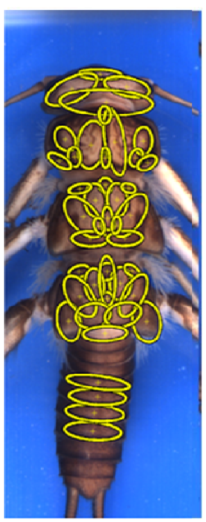
Hypotheses

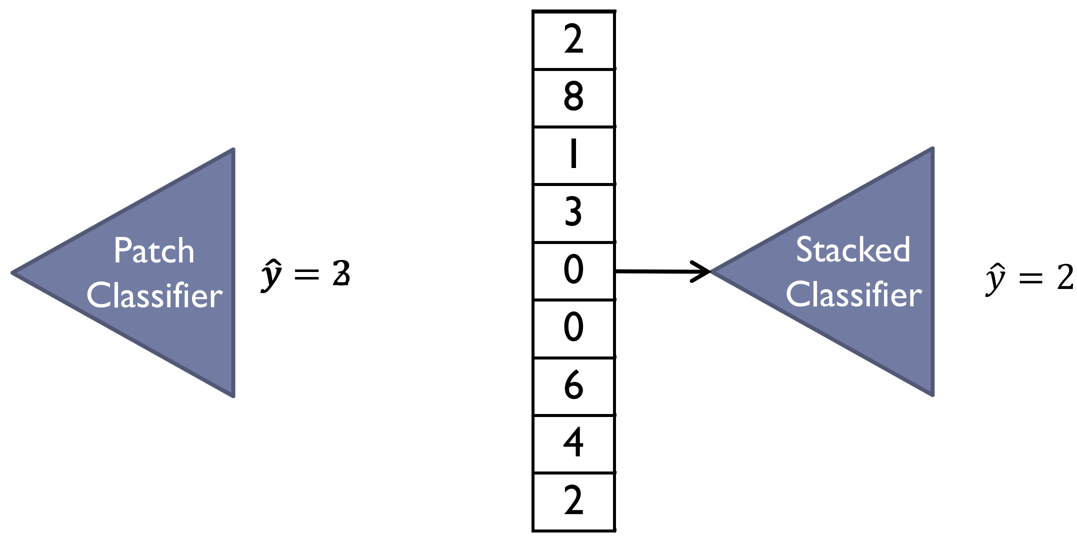
- Fine-grained classification requires
  - High-resolution images
  - Non-uniform extraction of information from the image

- Existing object recognition methods
  - Break image into set of patches
  - Extract a fixed number of bits from each patch
    - e.g., via vector quantization, filter banks, PCA, etc.
  - Classify image using extracted information

A “Variable Resolution” Method for Object Recognition

- **Stacked Patch Classifiers** [Martinez, et al, 2009]
  - Learn a classifier that <u>tries</u> to classify the whole image using detailed information from a single patch
  - Combine the single-patch classifications into a classification for the whole image





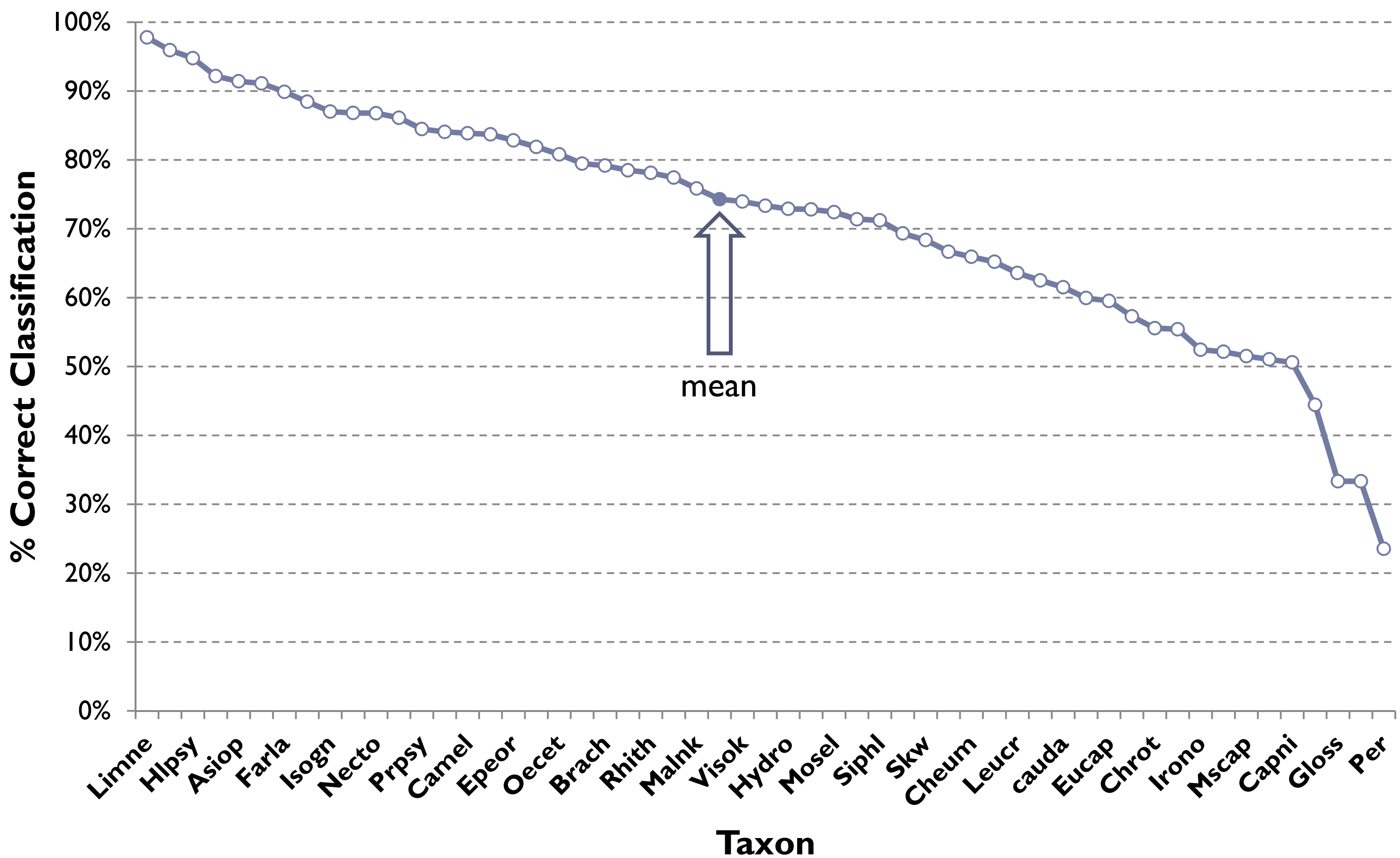
Results on STONEFLY9 Dataset

Configuration	Error Rate
Fixed resolution method	16.1%
Stacked Patch Classifier	5.6%

- Variable resolution method is much more accurate

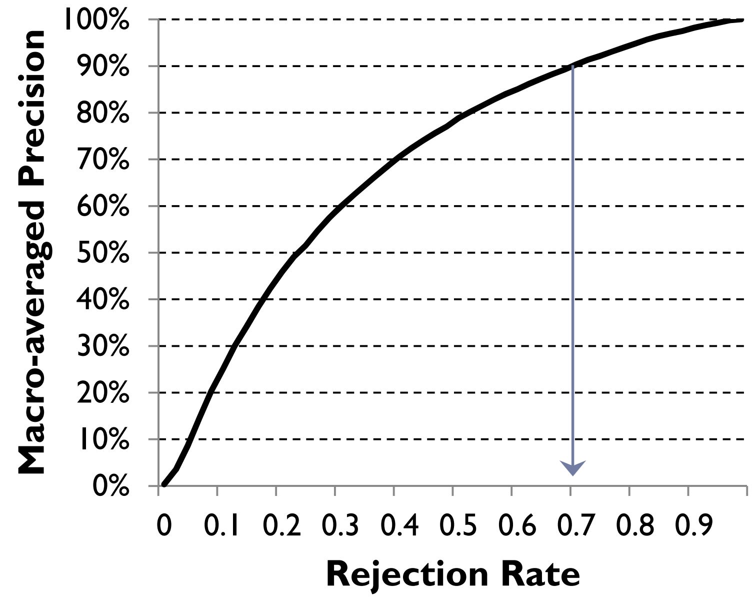
EPT54: 54 Species of Freshwater Macroinvertebrates

- Stacked Patch Classifier: 74.3% Correct



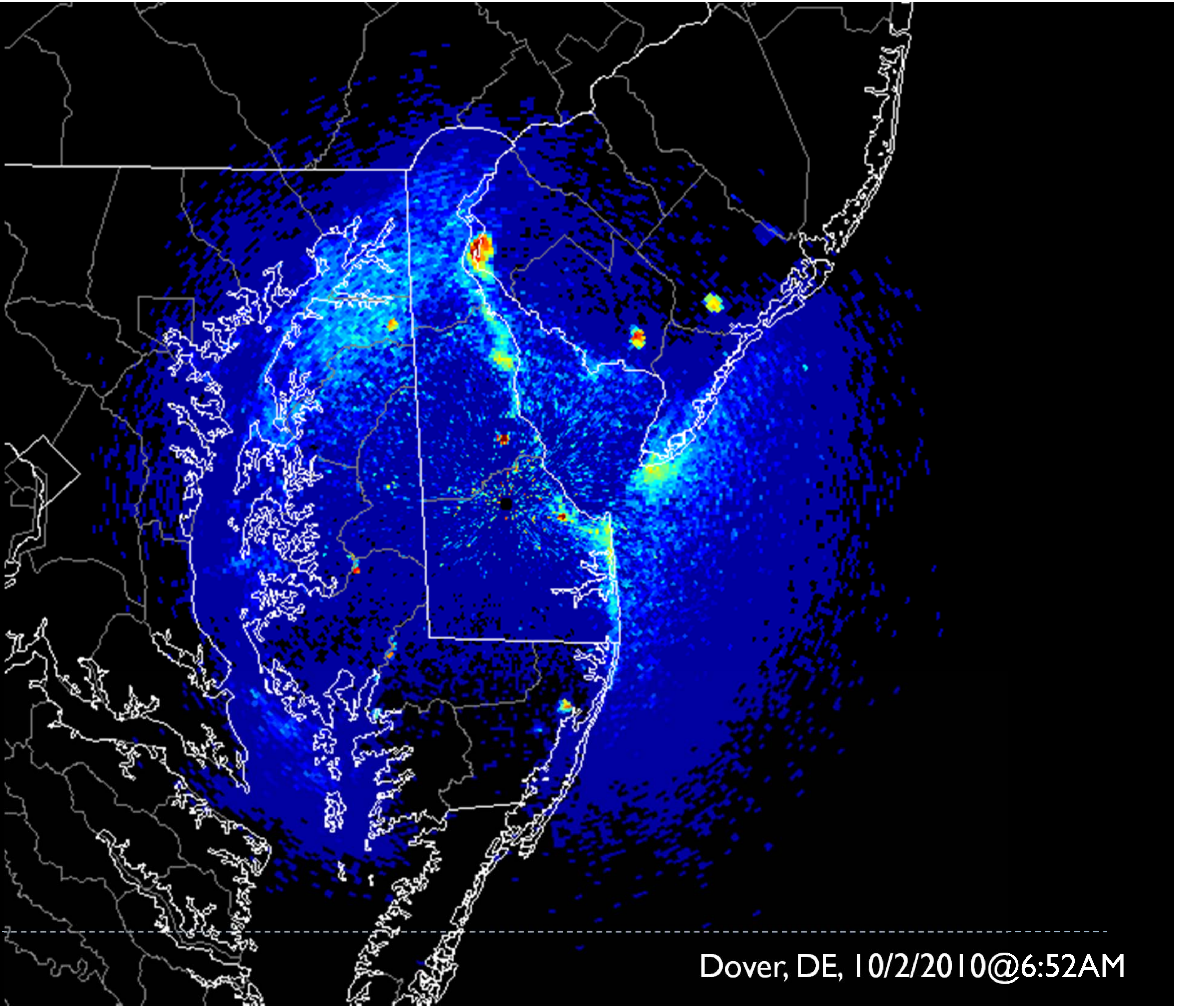
Open Problems

- Rejection:
  - Maximize recall subject to high precision
- Detect and reject novel (i.e., unknown to the classifier) species
- Scale to thousands of species
- Hierarchical loss functions
  - Order, Family, Genus, Species
  - Classify as finely as possible while bounding error rate



Tracking Tree Swallow Roosts Using NEXRAD Radar

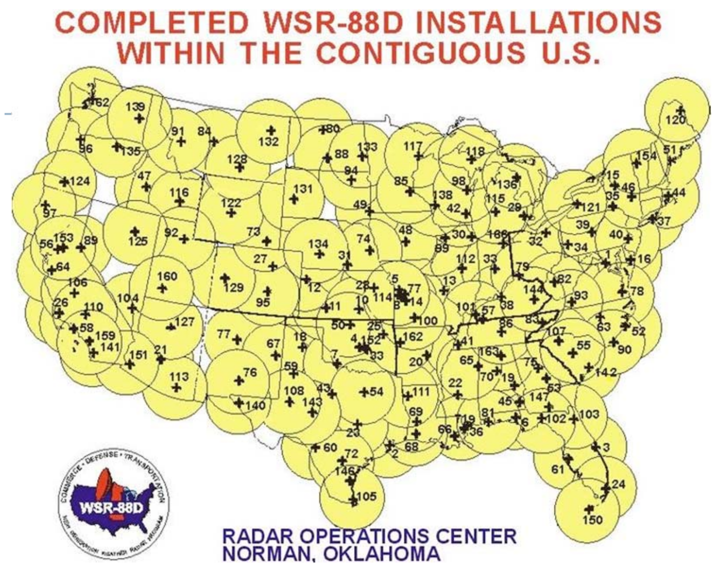
OSU: Sheldon & Dietterich; Cornell: Winkler



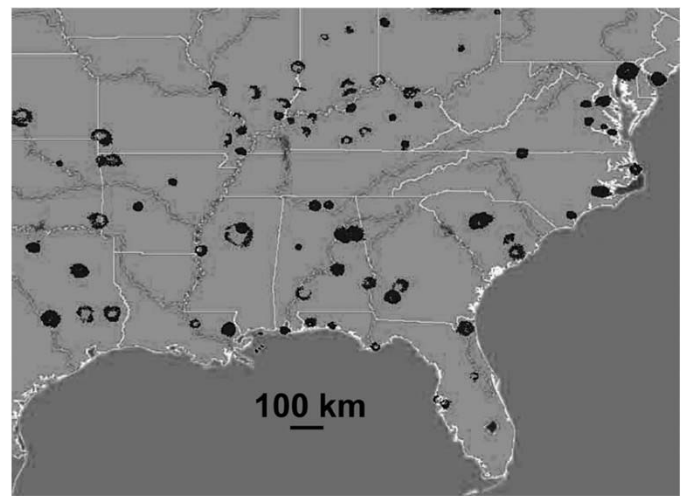
Dover, DE, 10/2/2010@6:52AM

The Dream

- Automatic detection of roosts at continent-scale on daily basis
  - Data gathering and repurposing
- Unprecedented view of species distribution
  - Spatial coverage
  - Temporal resolution
- Analyze results to learn about
  - Roost biology
  - Migration patterns
  - Climate change
    - Data archived since 1991



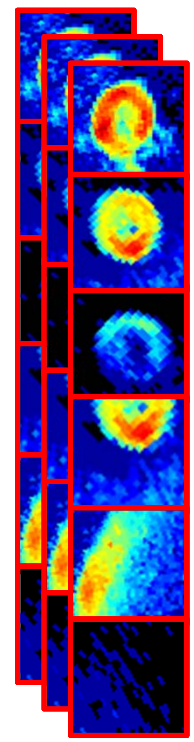
Source: NOAA



[Winkler, 2006]

Machine Learning Pipeline (1)

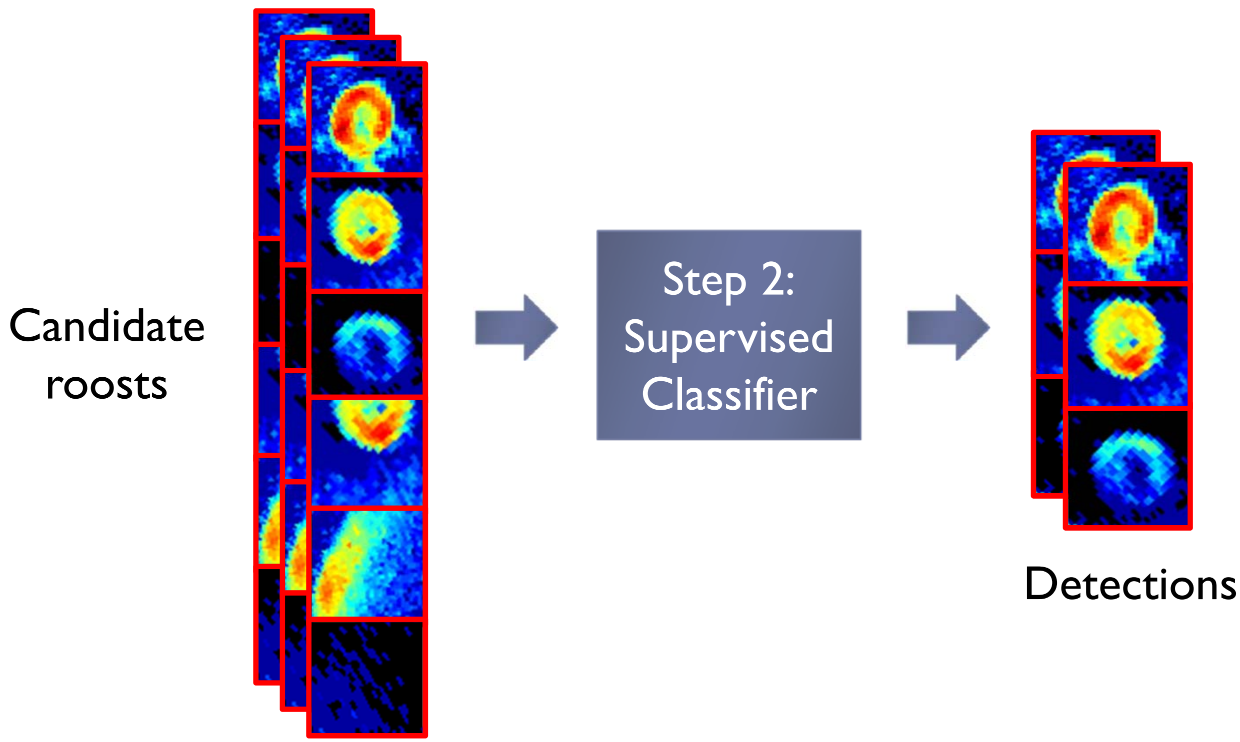
Radar scans (terabytes) → Step 1: Fast Unsupervised Detector →



Candidate roosts

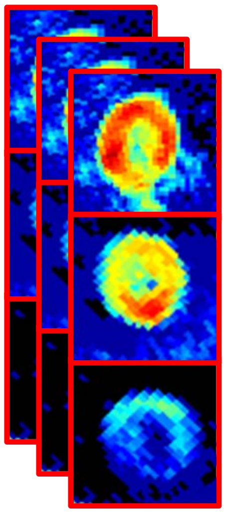
- Primary goal: data reduction
  - High recall
  - Many false positives

Machine Learning Pipeline (2)



- Shape features
- Biology features (velocity, habitat, weather, etc.)

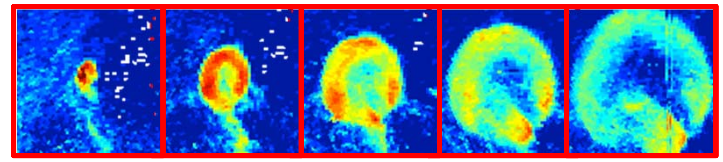
Machine Learning Pipeline (3)



Detections
(all frames)

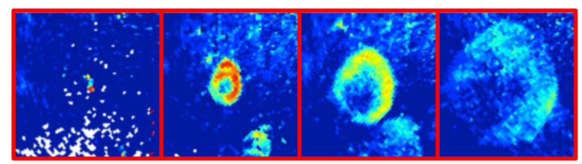
Step 3:
Sequence
assembly
(Multiple Target
tracking)

Roost 1



. . .

Roost n

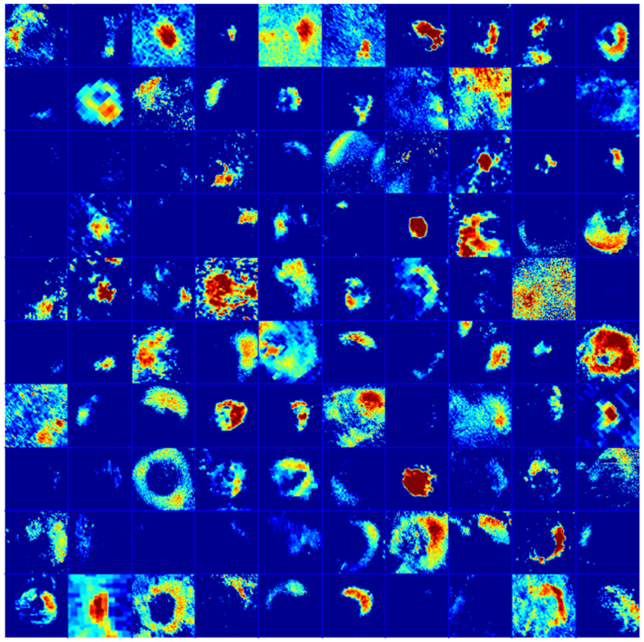


Sequences

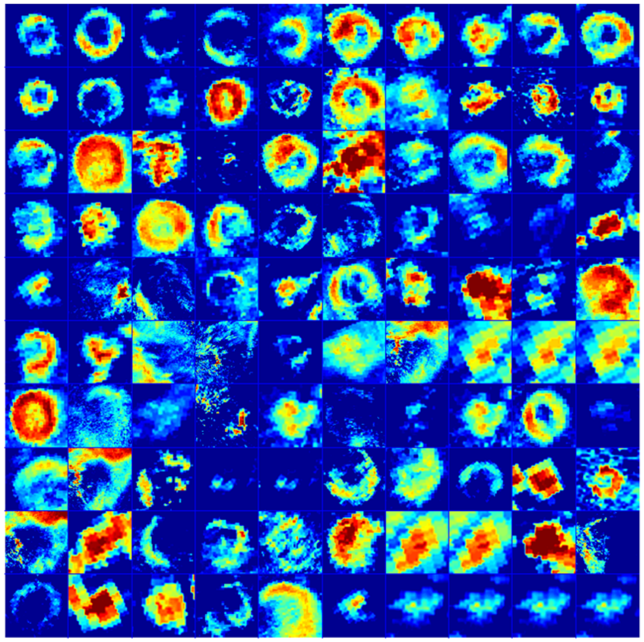
- Motivation:
  - Improve detection by using temporal context
  - Extract high-level information such as duration, maximum size, etc.

Progress: Machine Learning

- Steps 1 and 2
  - Primarily shape features to-date
  - High precision for roosts with "perfect appearance"
  - Variability in appearance is challenging → low recall



100 positive examples

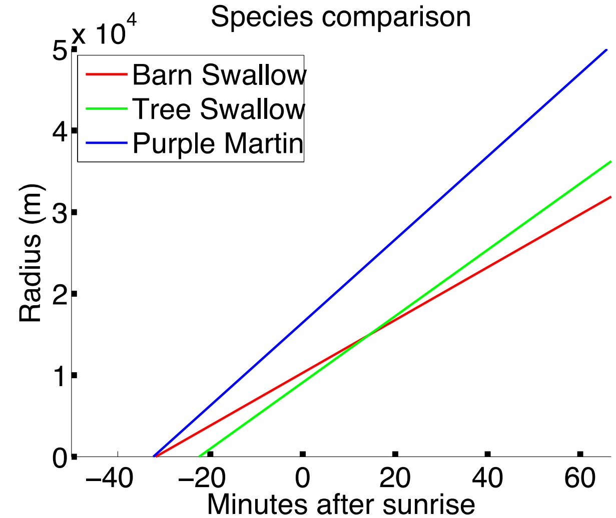


Top 100 predicted roosts
(shape features + SVM)

Progress: Ecology

- Locating roosts
  - Identifying roosts in radar images
    - Labeling efforts
  - Estimate ground location within a few km
    - Previously difficult task
    - 15+ roosts located in 2010-2011
      - Oregon, Florida, Louisiana

- Analysis of labeled data
  - Understand regional patterns
  - Roost growth dynamics
    - Very predictable
    - Potential species ID from radar!



Collaborators: Laughlin, Taylor (Tulane); Winker (Cornell); Burney (USAF)

Summary

- Ongoing project
  - A lot of work remains to reach "the dream"
- Significant opportunity for ML and ecology to develop in parallel

Questions on Part 1?

Part 2: Ecological Models

Outline

- Data Acquisition
  - Sensors: Physical sensors, human observers, repurposing data from other sources
  - Data interpretation: Extracting signals from data
- Ecological Models
  - Species Distribution Models
  - Dynamical Models: Dispersal, Migration, Invasion, Climate Change
- Policy Optimization
  - Conservation: Reserve design, Network design
  - Invasive species: Eradication, restoration, monitoring
  - Fisheries: Managing harvest levels

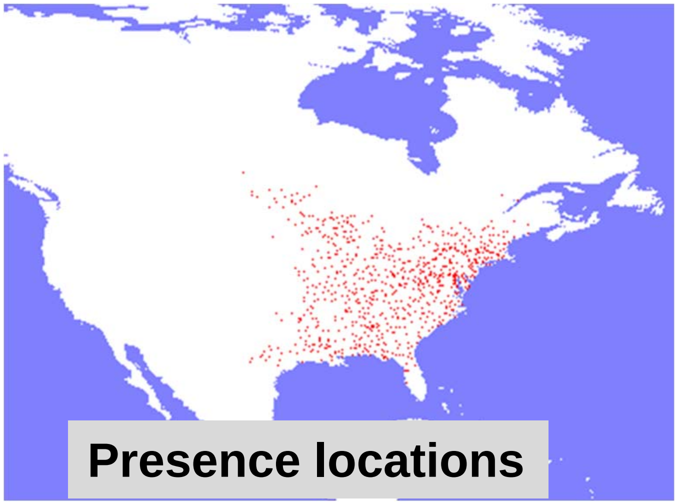
Ecological Models

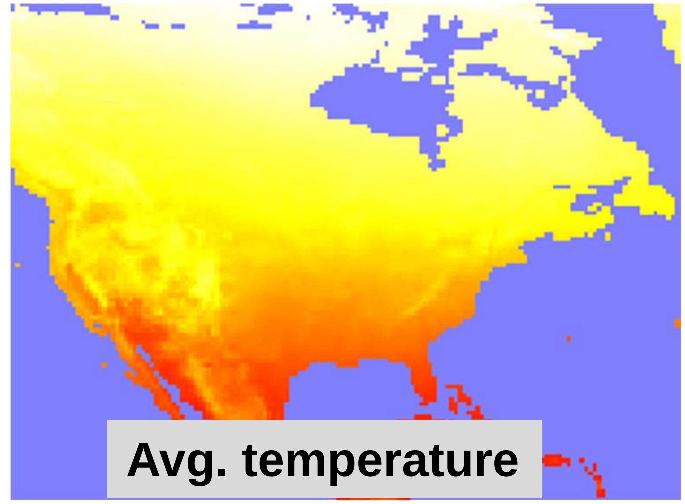
- Species Distribution Models
  - Static descriptions of the geographic distribution of a species.
  - Address the fundamental ecological question of why species are found where they are.

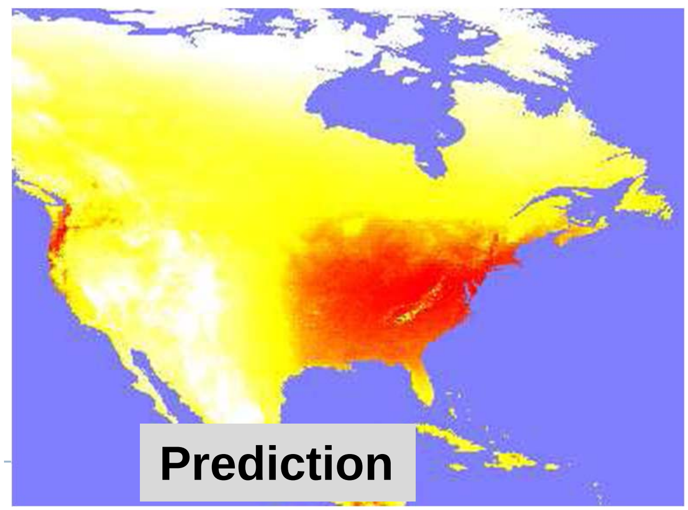
- Dynamical Models
  - Account for dynamic ecological processes like dispersal, migration, population growth, etc.

Example

- Yellow-throated vireo







Images are from Figure 1 in Phillips, et al., 2004.

Species Distribution Models (SDM)

- SDMs characterize the geographic distribution of a species in terms of environmental variables.

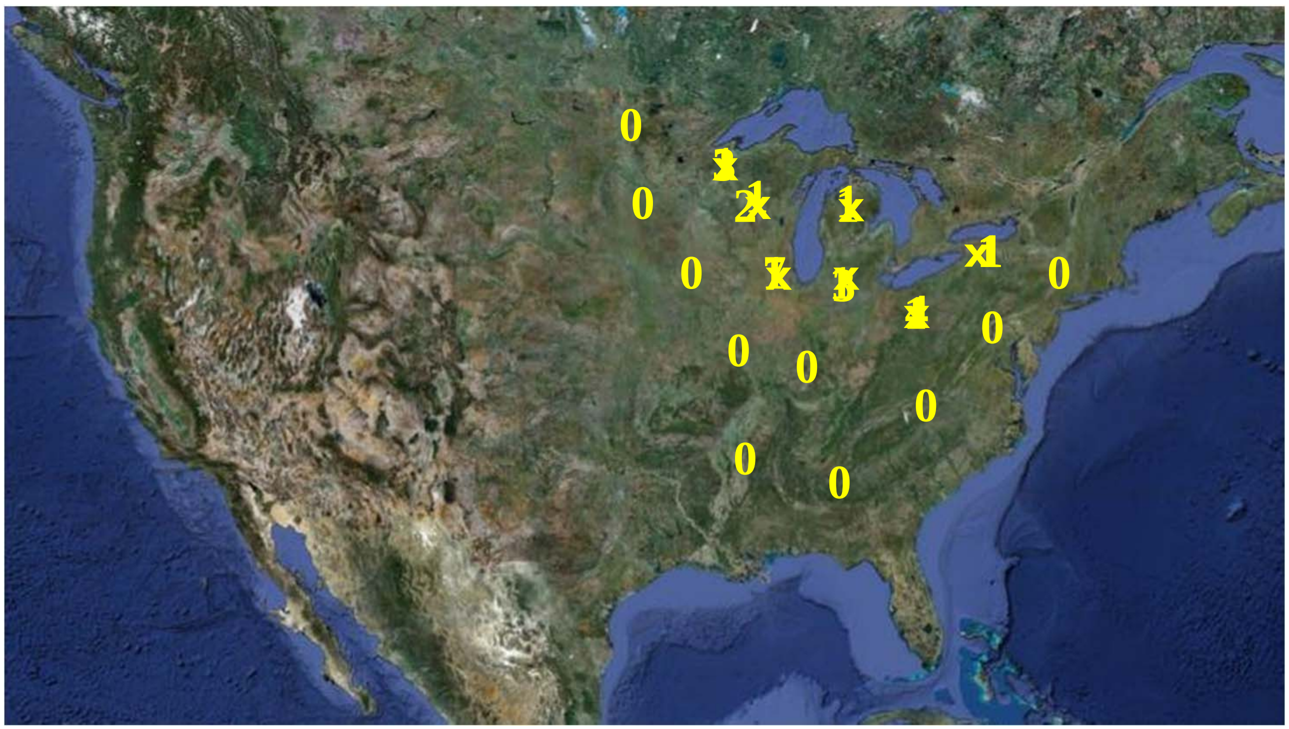
- Typically treated as a supervised classification problem with input variables **x** and species observations y.

$$\{(\boldsymbol{X}_i, y_i)\}_{i=\{1,\ldots,N\}} \rightarrow y = f(\boldsymbol{X})$$

- Goals:
  - Mapping current distribution
  - Understanding habitat requirements
  - Predicting distribution

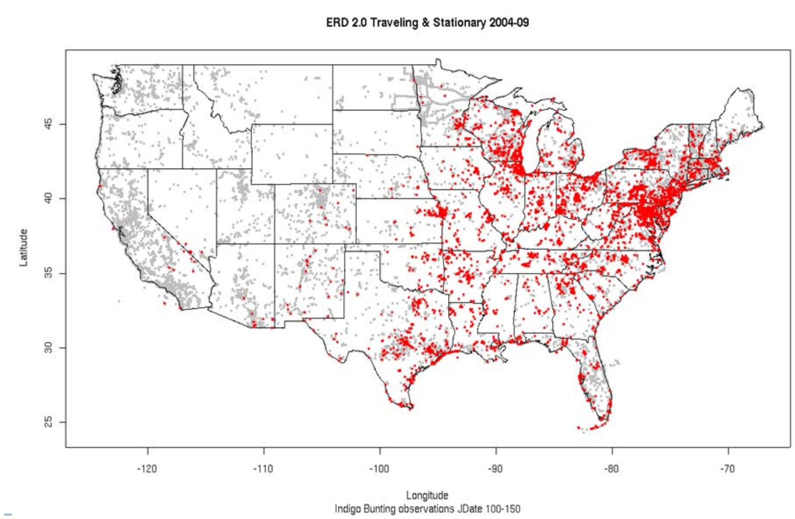
SDM: Data

- Types of y
  - Presence-only
  - Presence/absence
  - Abundance



eBird Data

- Bird watchers record their observations in a database through eBird.org.
  - "Citizen Science"
- Dataset available for analysis. (see references)
- Features
  - LOTS of data!
    - ~3 million observations reported this May
  - All bird species (~3,000)
  - Year-round
  - Continent-scale
- Challenges
  - Variable quality observations
  - No systematic sampling design



SDM: Methods

- Envelope Models
  - Bioclim
- Statistical and Machine Learning Models
  - Maxent
  - Generalized Linear Models
  - Generalized Additive Models
  - Multivariate Adaptive Regression Splines
  - Hierarchical Bayesian modeling
  - Boosted regression trees
  - Random forests
  - Genetic algorithms

…and more!

ML is already having an impact in SDM

- 16 methods
- 226 species
- 6 regions

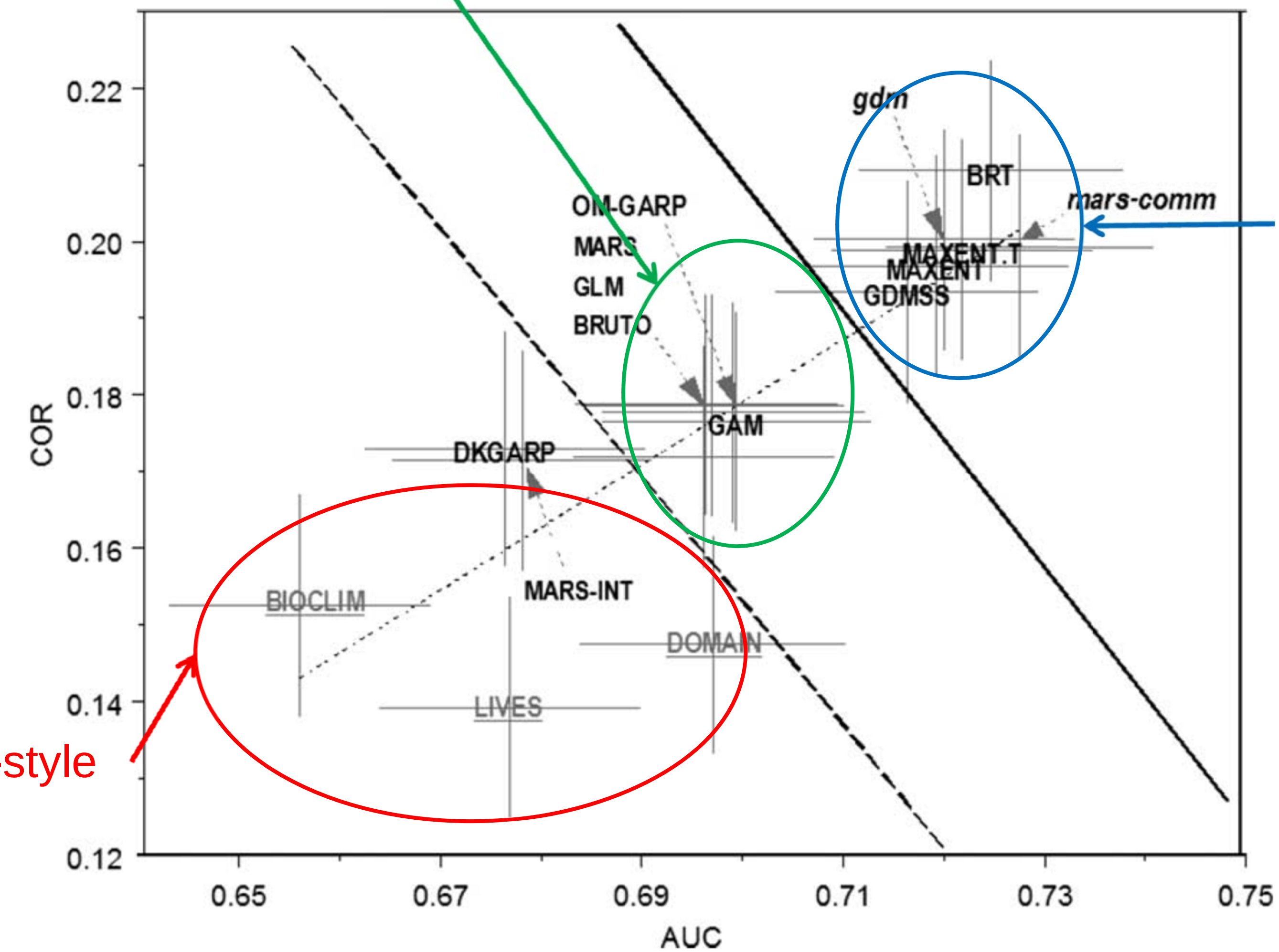
ECOGRAPHY 29: 129–151, 2006

Novel methods improve prediction of species' distributions from occurrence data

Jane Elith*, Catherine H. Graham*, Robert P. Anderson, Miroslav Dudík, Simon Ferrier, Antoine Guisan, Robert J. Hijmans, Falk Huettmann, John R. Leathwick, Anthony Lehmann, Jin Li, Lucia G. Lohmann, Bette A. Loiselle, Glenn Manion, Craig Moritz, Miguel Nakamura, Yoshinori Nakazawa, Jacob McC. Overton, A. Townsend Peterson, Steven J. Phillips, Karen Richardson, Ricardo Scachetti-Pereira, Robert E. Schapire, Jorge Soberón, Stephen Williams, Mary S. Wisz and Niklaus E. Zimmermann

- General result: new(er) statistical and/or machine learning methods outperformed older envelope/distance style models.

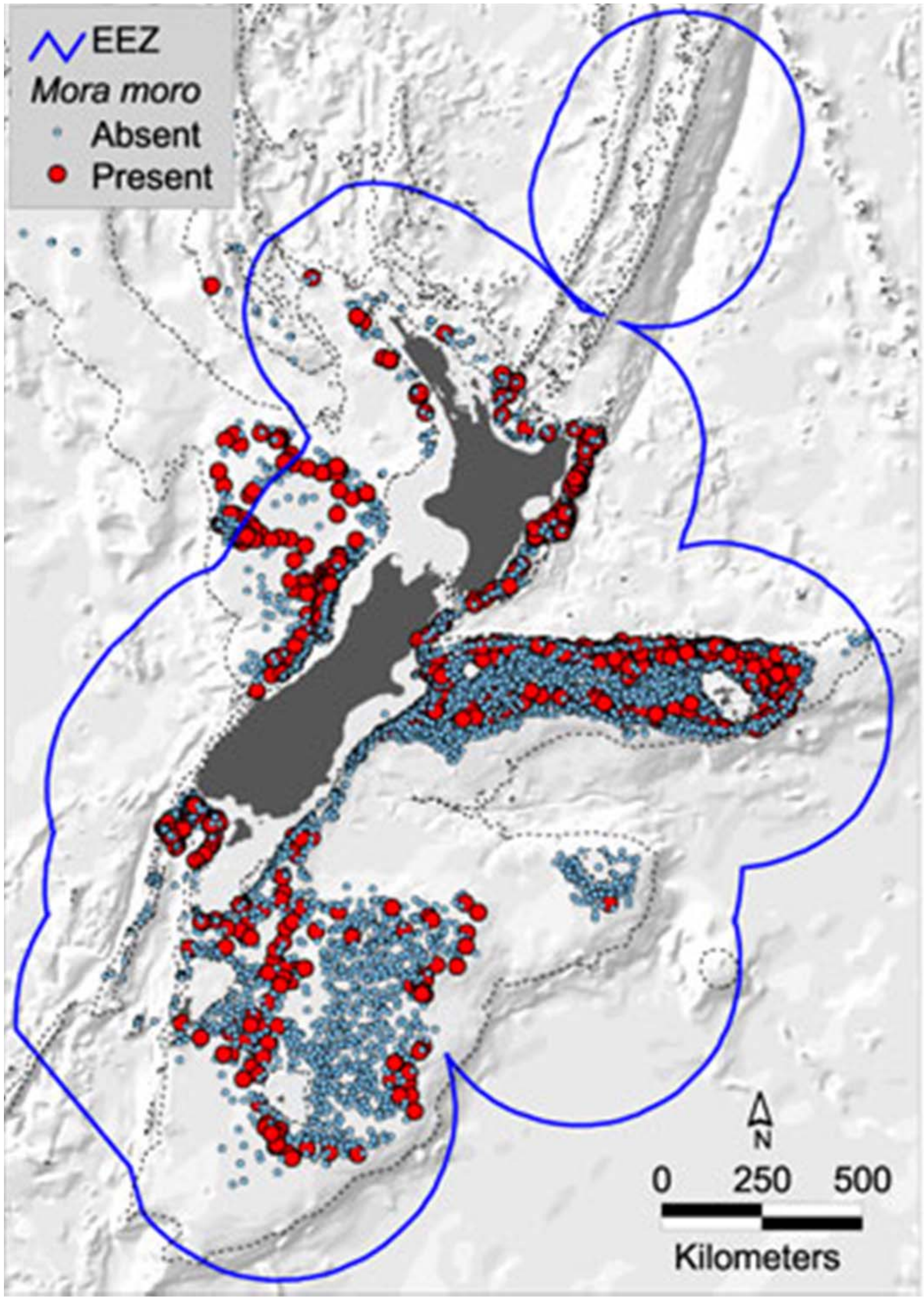
Statistical, regression-style models



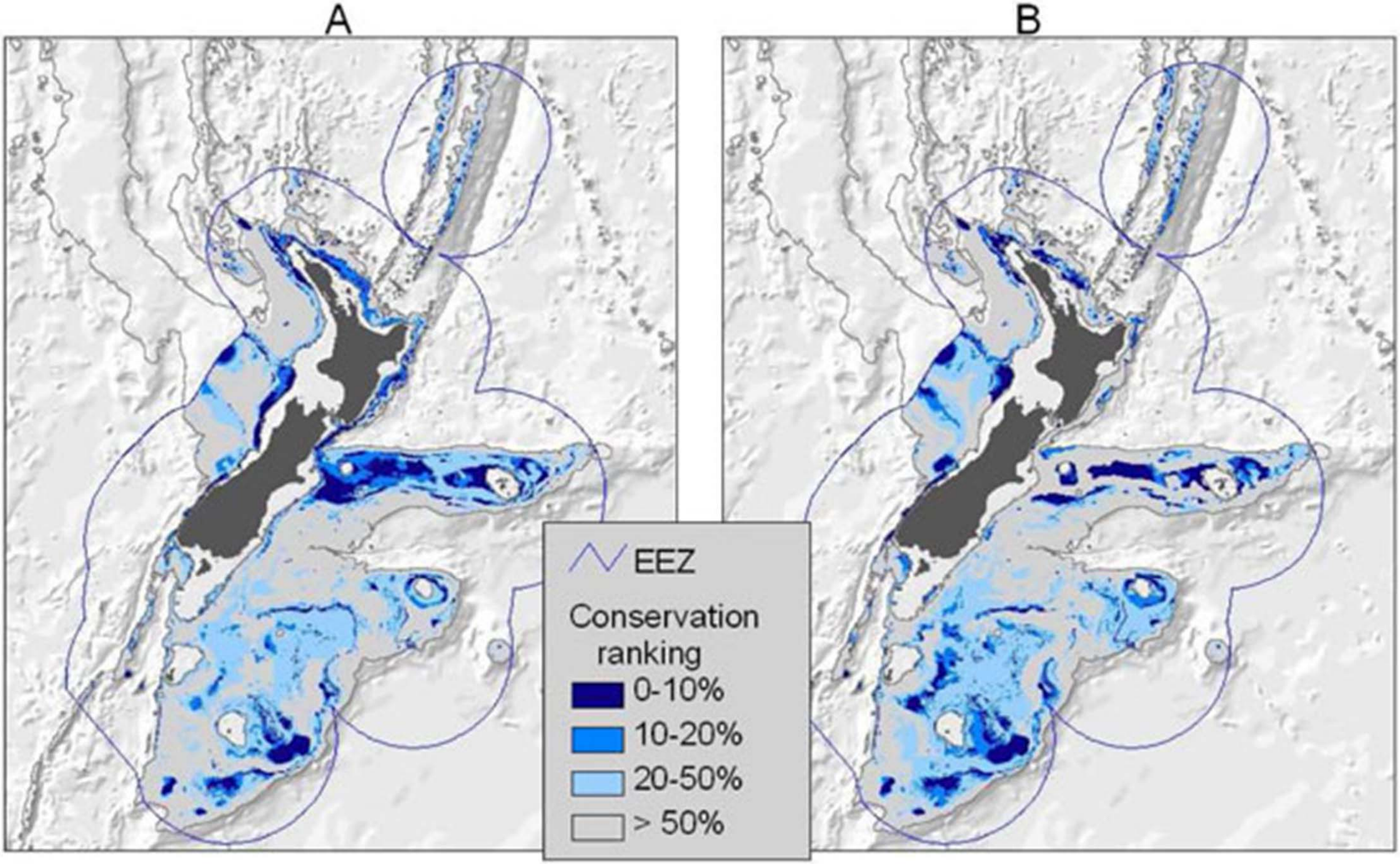
Newer, machine learning-style models

Older, envelope-style models

Elith et al, 2006



Leathwick et al, 2008



Disregarding costs to fishing industry

Full consideration of costs to fishing industry

Leathwick et al, 2008

A few SDM Challenges

- Presence-only data
- Predictor-response relationships are non-stationary
- Imperfect detection of the species on surveys
  - Often lack prior knowledge of the system for model building
  - Observers have variable expertise/biases

Challenge #1: Presence-only data

- Problem: some data sources only contain records of presence (like herbaria and museum collections)

- Solution: Maximum entropy modeling (Maxent)

Phillips, et al, 2006

Positive-Only Learning Problem

- Given:
  - Training examples $x_1, \dots, x_N$ where the species is present
  - These are assumed to be drawn from an unknown probability distribution: $\pi(x) = P(x|y = 1)$
  - A set of feature functions $\phi_1, \dots, \phi_J$ such that $\phi_j(x)$ computes the value of the jth feature of x. Let $\Phi(x) = \big(\phi_1(x), \dots, \phi_J(x)\big)$.
- Find:
  - A good approximation $\hat{\pi}$ to π
- Method:
  - *Maximum entropy principle:* Among all distributions consistent with the data, prefer the distribution of maximum entropy
  - Find the maximum entropy distribution subject to expectation constraints:

$$\hat{\pi} = \operatorname*{argmax}_q H(q) \text{ subject to } \mathrm{E}_q\big[\phi_j(x)\big] = \frac{1}{N}\sum_i \phi_j(x_i) \ \forall j$$

Dudik, Phillips, Schapire, 2004

Solving the Maxent Optimization

- Step 1: Relax the constraints:

$$\hat{\pi} = \underset{q}{\operatorname{argmax}}\, H(q) \ \text{ subject to}$$

$$\left| \mathrm{E_q}\left[\phi_j(x)\right] - \frac{1}{N}\sum_i \phi_j(x_i) \right| \leq \beta_j \ \ \forall j$$

- Step 2: Assume a parametric form for $\hat{\pi}$:

$$\hat{\pi}(x) = \frac{1}{Z(\boldsymbol{w})}\exp[\boldsymbol{w} \cdot \Phi(\boldsymbol{x})]$$

- Step 3: Apply duality methods to show this is equivalent to an L_1-regularized linear optimization

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}}{\operatorname{argmax}} \sum_i \boldsymbol{w} \cdot \Phi(x_i) - \sum_j \beta_j |w_j|$$

Dudik, Phillips, Schapire, 2004

Obtaining an SDM

- Problem: We have a model $\hat{\pi}$ of $P(x|y=1)$ but we want a model of $P(y=1|x)$.
- Solution: Apply Bayes' Rule

$$P(y=1|x) = \frac{P(x|y=1)P(y=1)}{P(x)}$$

  - $P(y=1)$ is the "abundance". It is a constant that is *not* identifiable from presence-only data.
  - $P(x)$ is the "background distribution" of the study area (often assumed uniform).
  - Therefore,

$$P(y=1|x) \propto \hat{\pi}(x)$$

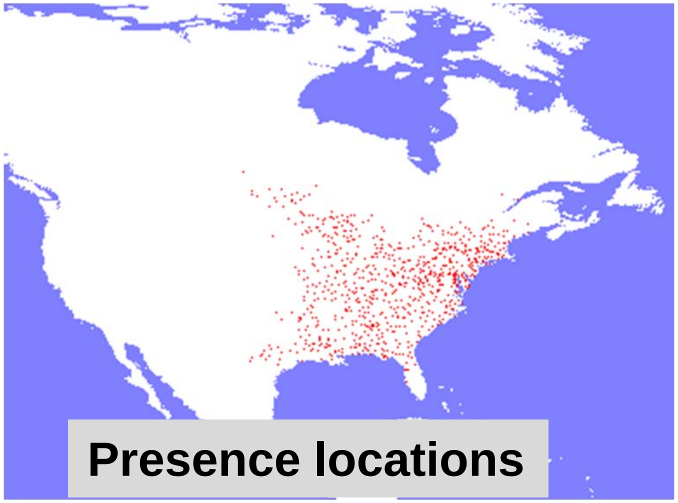
Phillips & Dudik, 2008

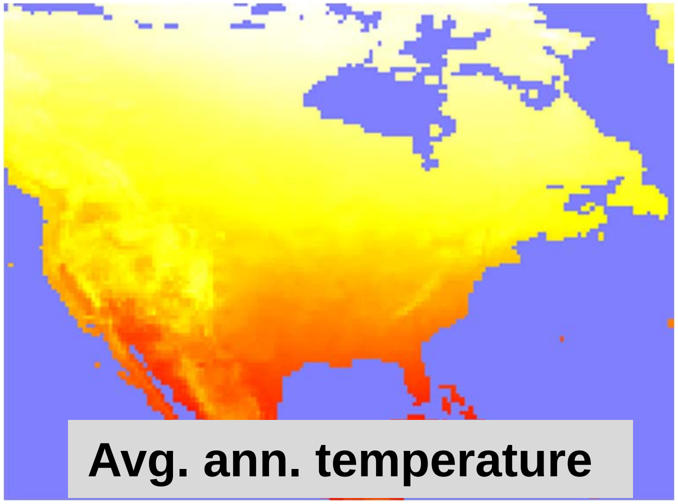
Creating a Usable Tool

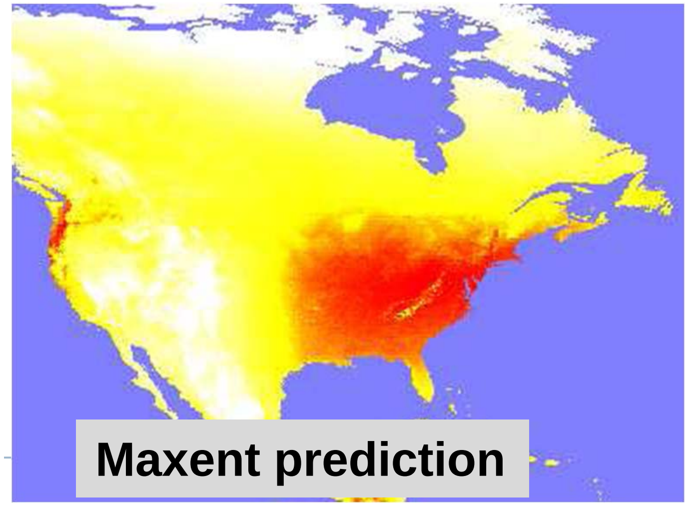
- Free software package for SDM
  - http://www.cs.princeton.edu/~schapire/maxent/
  - Has had a huge impact in the ecology literature
- Provides a rich set of feature types ϕ^t
  - linear
  - quadratic
  - thresholds
  - ramps
  - pairwise products of these
- Provides default settings for the βs
  - The method requires tuning a separate β_j for each feature, which is hard to do via cross-validation.
  - Defaults are based on tuning for 6 datasets from Elith, et al. [2006]

Example

- Yellow-throated vireo







Images are from Figure 1 in Phillips, et al., 2004.

Challenge #2: Non-stationarity

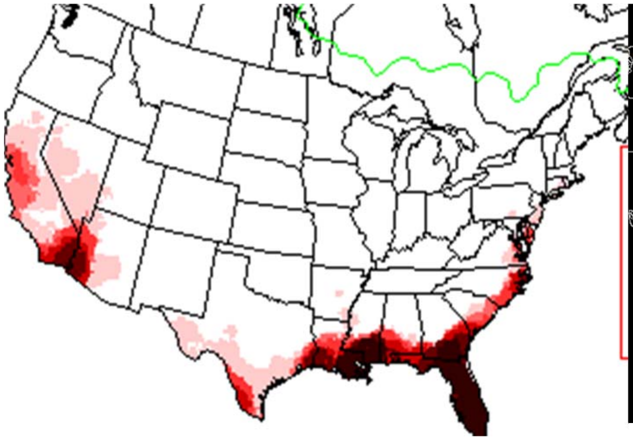
- Problem: predictor-response relationships change over space and time

- Solution: Spatial-Temporal Exploratory Models (STEM)
  - Create ensembles with local spatial/temporal support
  - Base learner = classification trees

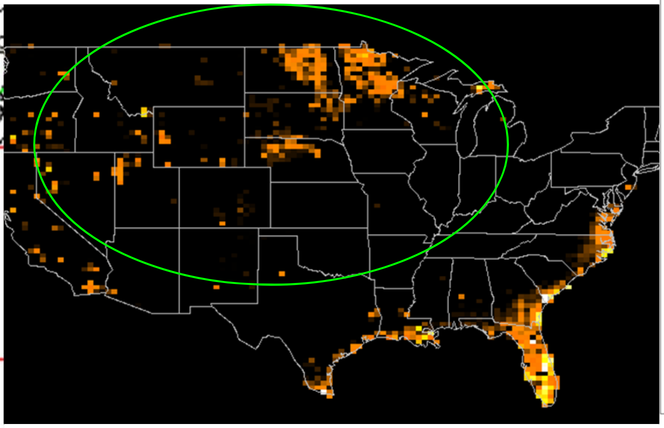
Fink, et al, 2010

(*Tachycineta bicolor*)

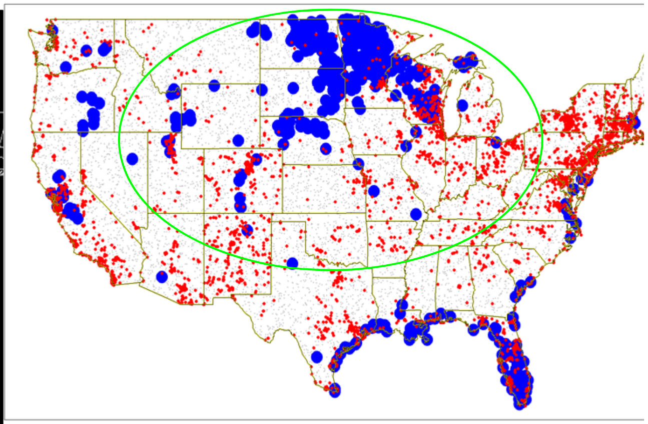
Tree Swallow
Winter Distribution Analysis
Bagged Decision Tree



Winter Distribution



eBird Bagged Decision Trees



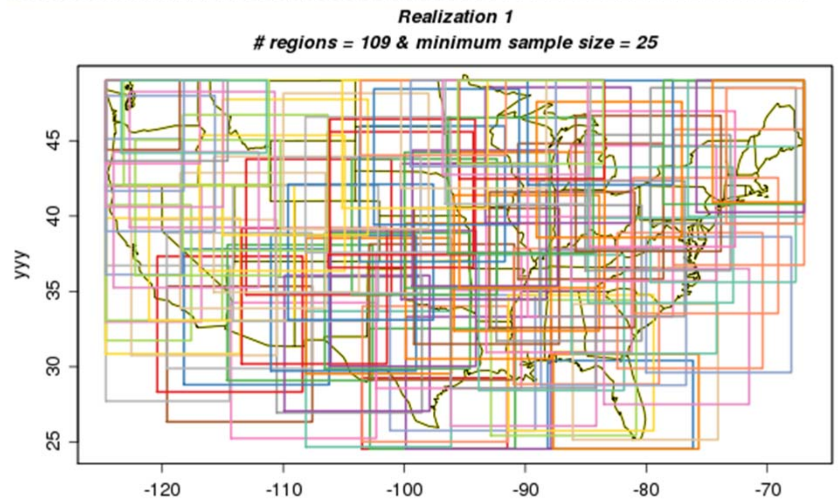
Wetland Coverage > 5%

- Nonparametric model with global support may aggregate data in ways that are ecologically impossible
- Risk Factors – complexity of process, model flexibility, predictor variation, and data sparsity

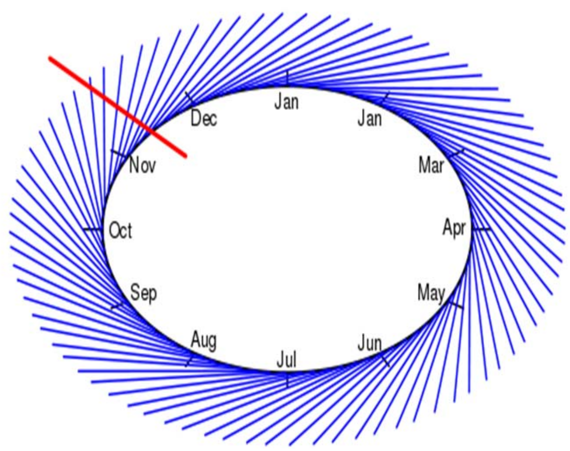
slide courtesy of Daniel Fink

STEM

- Idea:
  - Slice space and time into hyperrectangles: lat x long x time
  - Train a decision tree on the data inside each hyperrectangle
  - To predict at a new point x, vote the predictions of all trees whose hyperrectangle contains x
- Hyperrectangles:
  - Space: random rectangles of fixed size
  - Time: 40-day overlapping intervals spaced evenly throughout the year
  - Discard hyperrectangles that contain fewer than 25 training locations

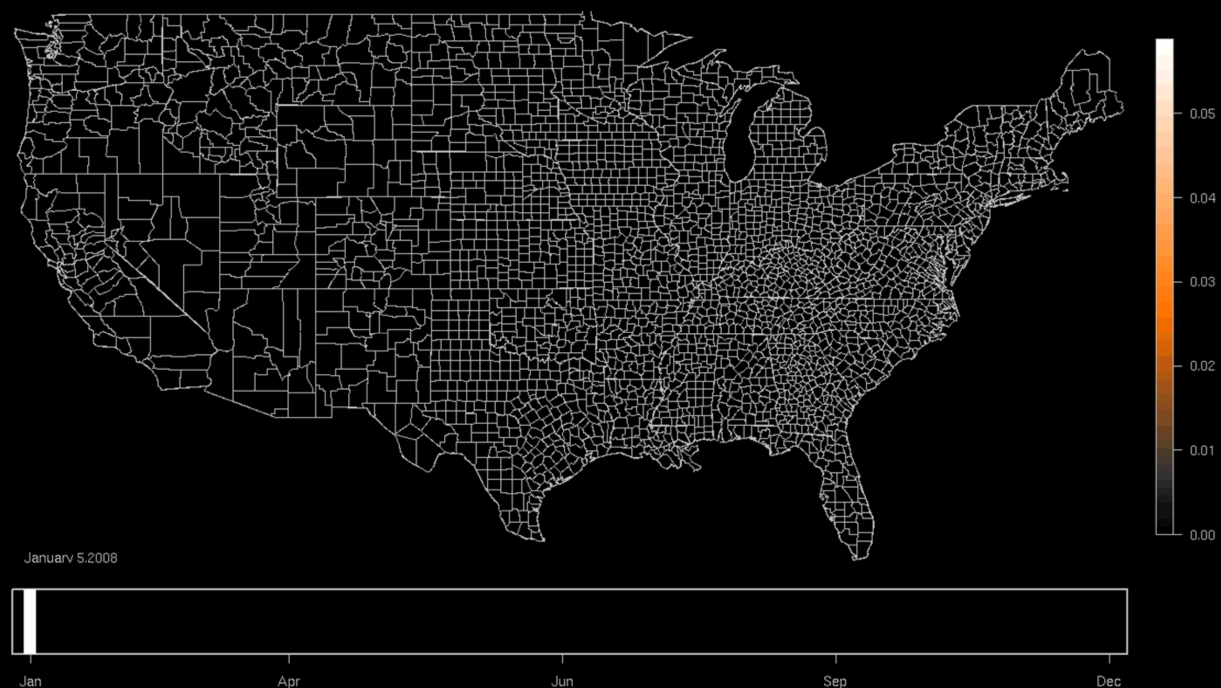


STEM Temporal Design



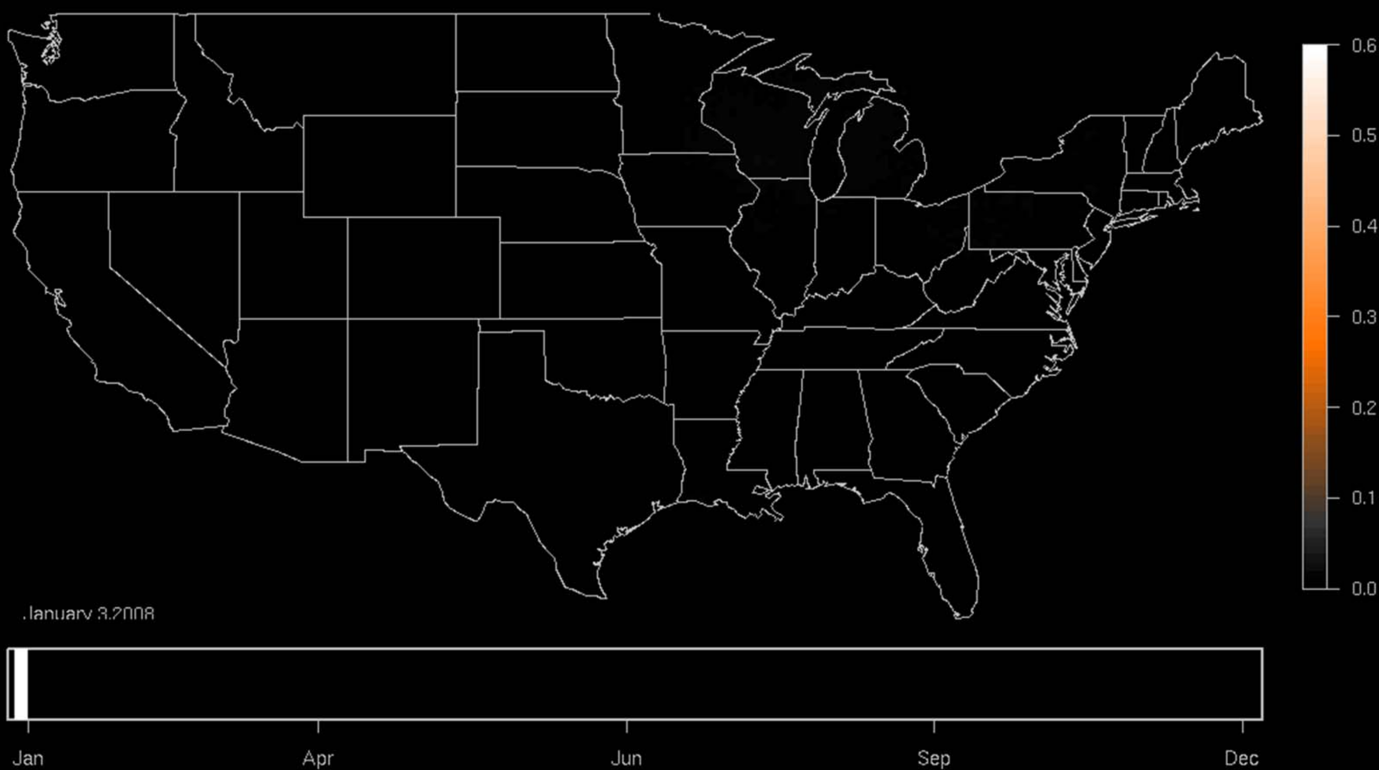
Fink, et al, 2010

STEM SDM: Solitary Sandpiper



slide courtesy of Daniel Fink

STEM SDM: Indigo Bunting



slide courtesy of Daniel Fink

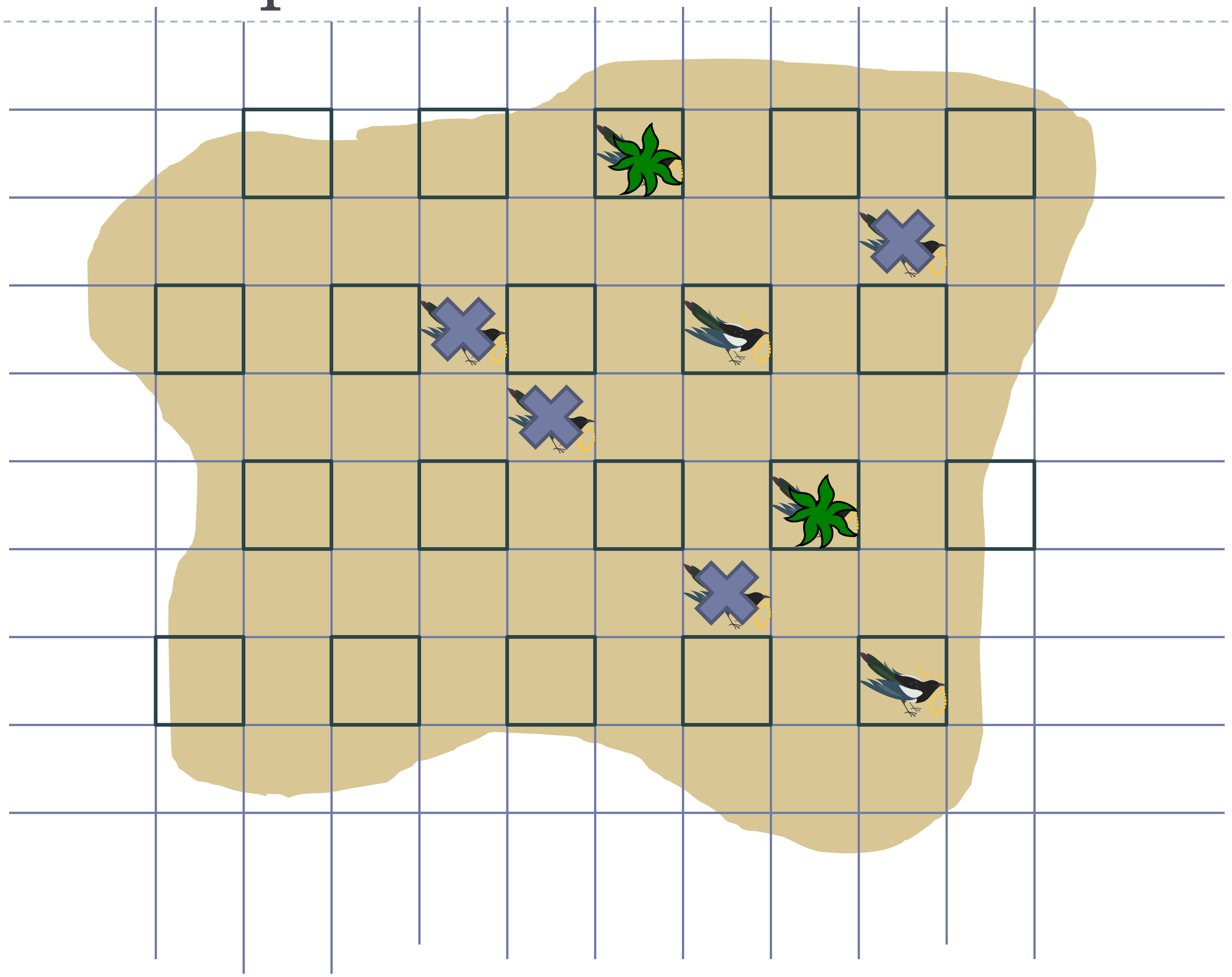
Challenge #3: Imperfect Detection

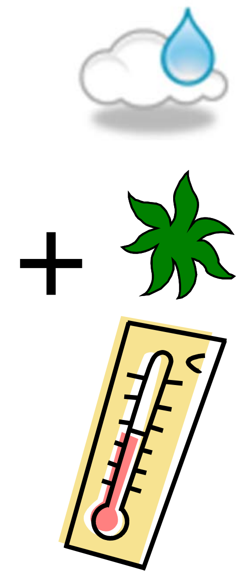
- Problem: many species are hard to detect even when present, so their data contain false negatives

- Solution:
  - survey sites several times
  - use a hierarchical model to describe the data collection process explicitly and correct for false zeros

MacKenzie, et al, 2006

Toy Example



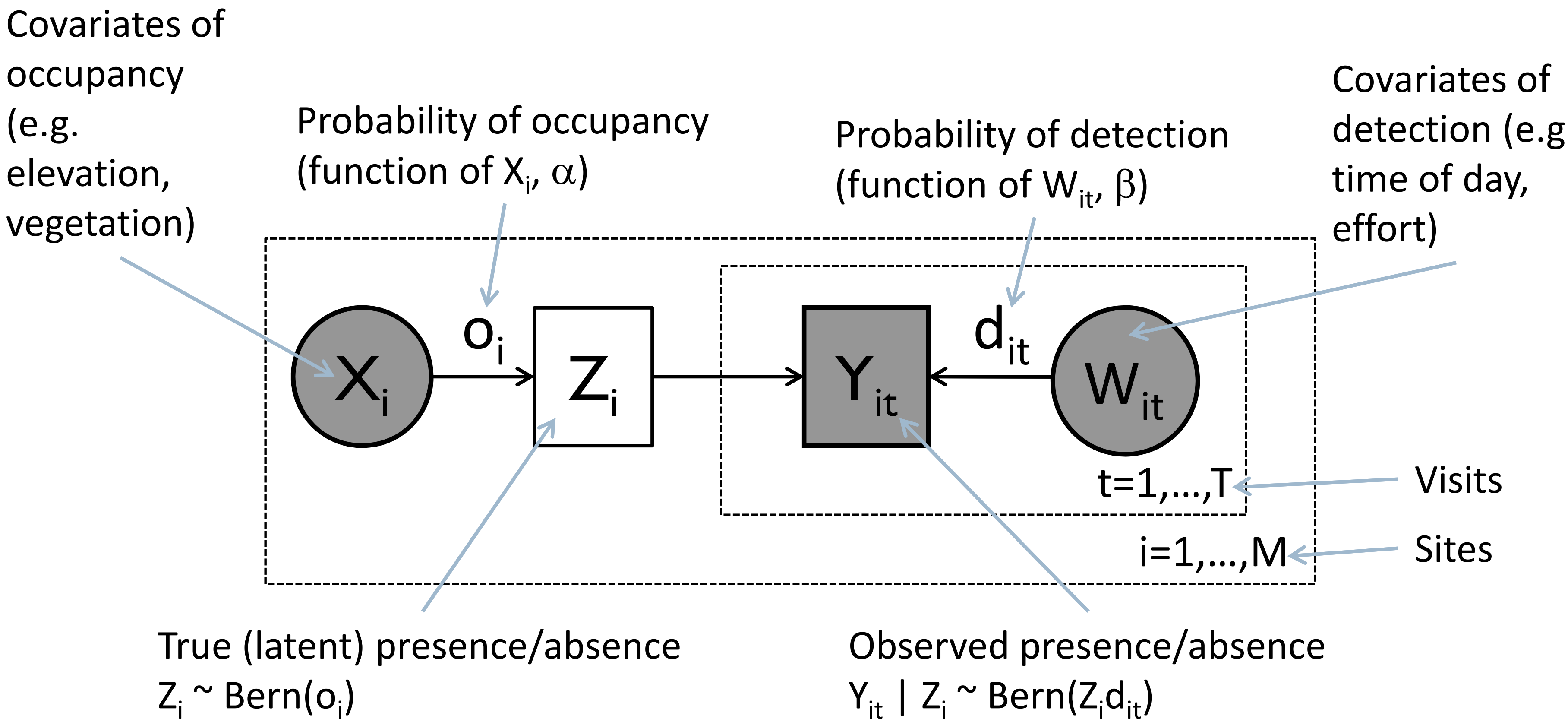


Data: detection histories

		Detection History		
Site	*True occupancy (latent)*	Visit 1 (rainy day, 12pm)	Visit 2 (clear day, 6am)	Visit 3 (clear day, 9am)
A (forest, elev=400m)	*1*	0	1	1
B (forest, elev=500m)	*1*	0	1	0
C (forest, elev=300m)	*1*	0	0	0
D (grassland, elev=200m)	*0*	0	0	0

* Assumption: experts never report false positives.

Occupancy Model



MacKenzie, et al, 2006

Typical Usage

$$\text{logit}(o_i) = F(X_i) = \alpha \cdot X_i$$

$$\text{logit}(d_{it}) = G(W_{it}) = \beta \cdot W_{it}$$

- Model selection:
  - construct models including different sets of occupancy and detection covariates
  - evaluate fit with AIC
  - hypothesis tests/confidence intervals

MacKenzie, et al, 2006

Imperfect Detection + Lack of Prior Knowledge

- Problem: occupancy models require parametric assumptions too rigid for exploratory modeling

- Solution: incorporate flexible models into the covariate-probability functions while maintaining hierarchical structure to account for imperfect detection

Hutchinson, et al, 2011

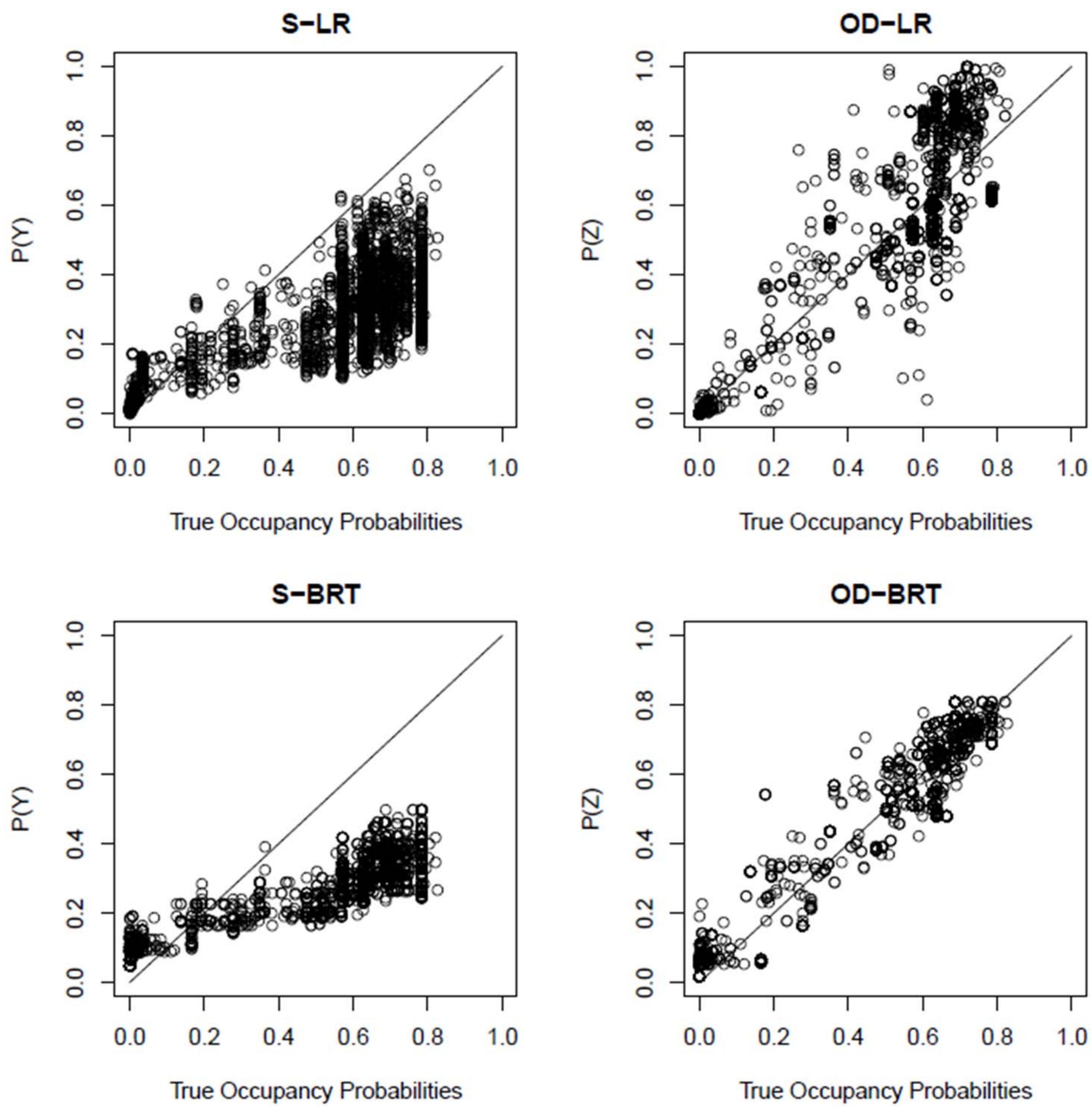
Integrating regression trees

$$\text{logit}(o_i) = F(X_i) = \sum_{j=1}^{J} \rho_j^{(o)} tree_j^{(o)}(X_j)$$

$$\text{logit}(d_{it}) = G(W_{it}) = \sum_{j=1}^{J} \rho_j^{(d)} tree_j^{(d)}(W_{it})$$

- Fit with *functional gradient descent* [Friedman, 2001]
  - On each iteration:
    - compute pseudo-targets (gradient of loss at each data point)
    - grow another tree to predict pseudo-targets
    - compute a weight for the tree and add to ensemble
  - Maximizes log-likelihood of occupancy model

Hutchinson, et al, 2011



Synthetic Species built from eBird covariates (with non-linearities)

S = supervised, with no latent structure (left column)

OD = occupancy model structure (right column)

LR = linear (top row)

BRT = tree-based (bottom row)

Hutchinson, et al, 2011

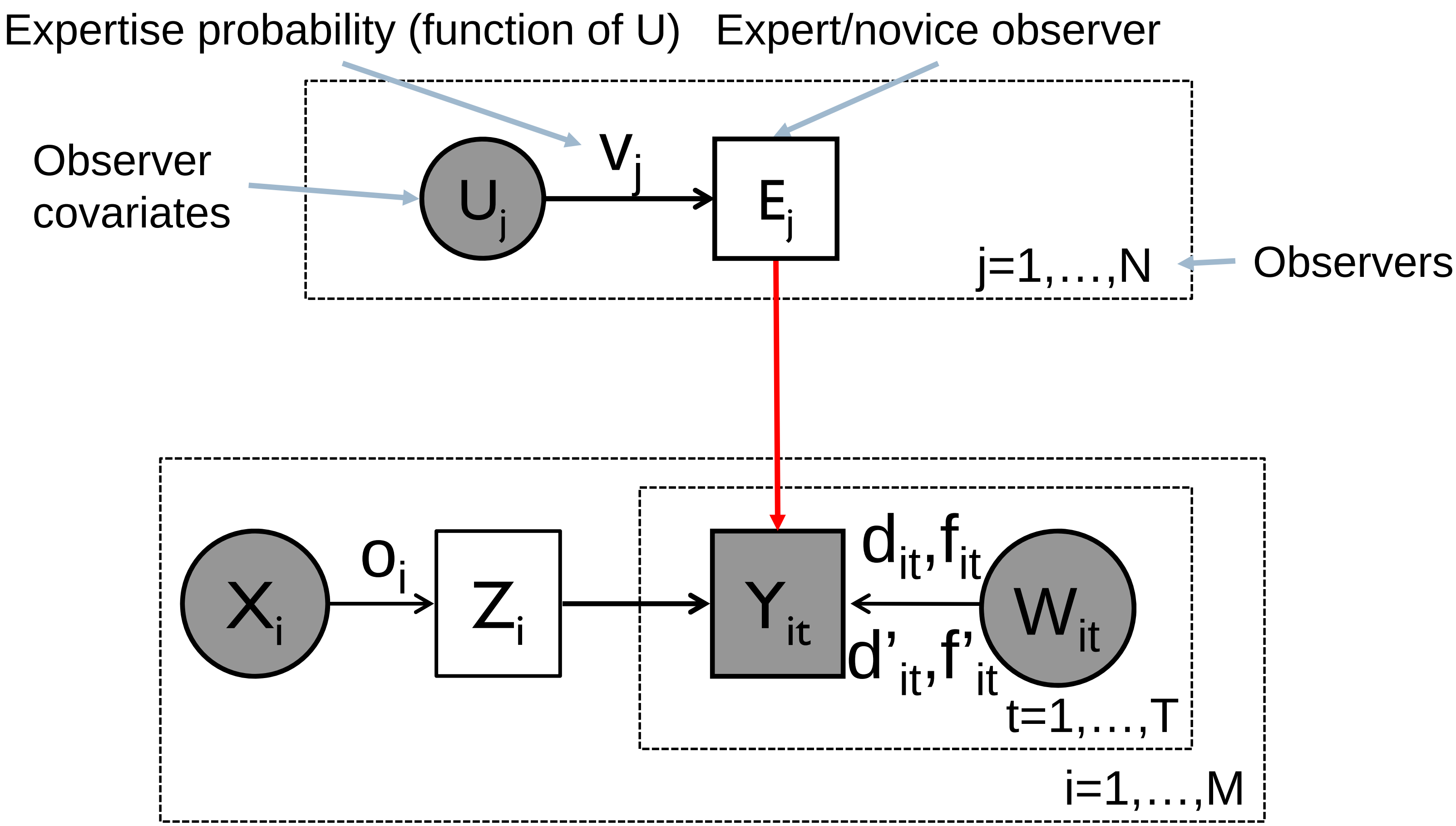
Imperfect Detection + Variable Expertise

- Problem: expert and novice observers contributing observations to citizen science data generate different mistakes/biases

- Solution: extend occupancy models so that observer expertise affects the detection model

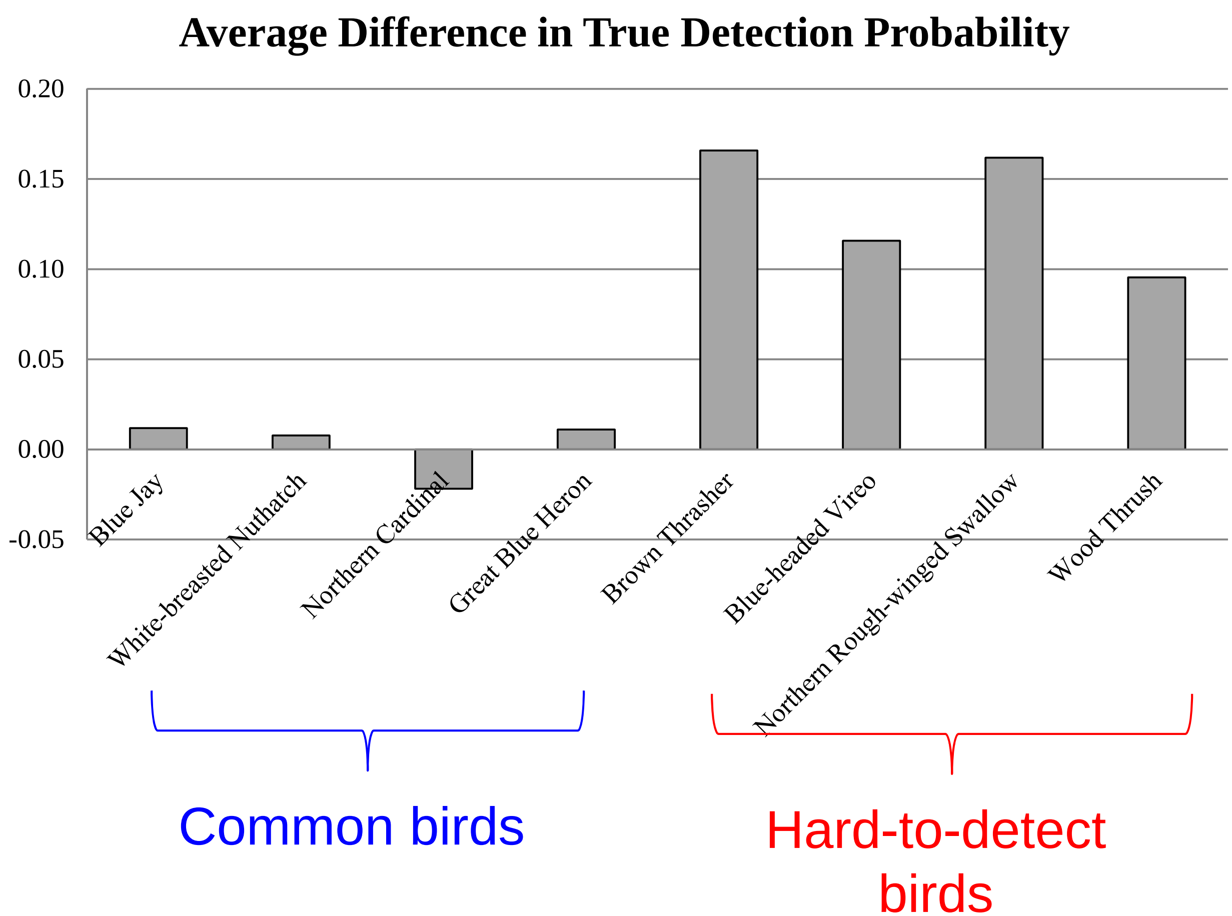
Yu, et al, 2010

Extending Occupancy Models



Yu, et al, 2010

Expert vs Novice Differences

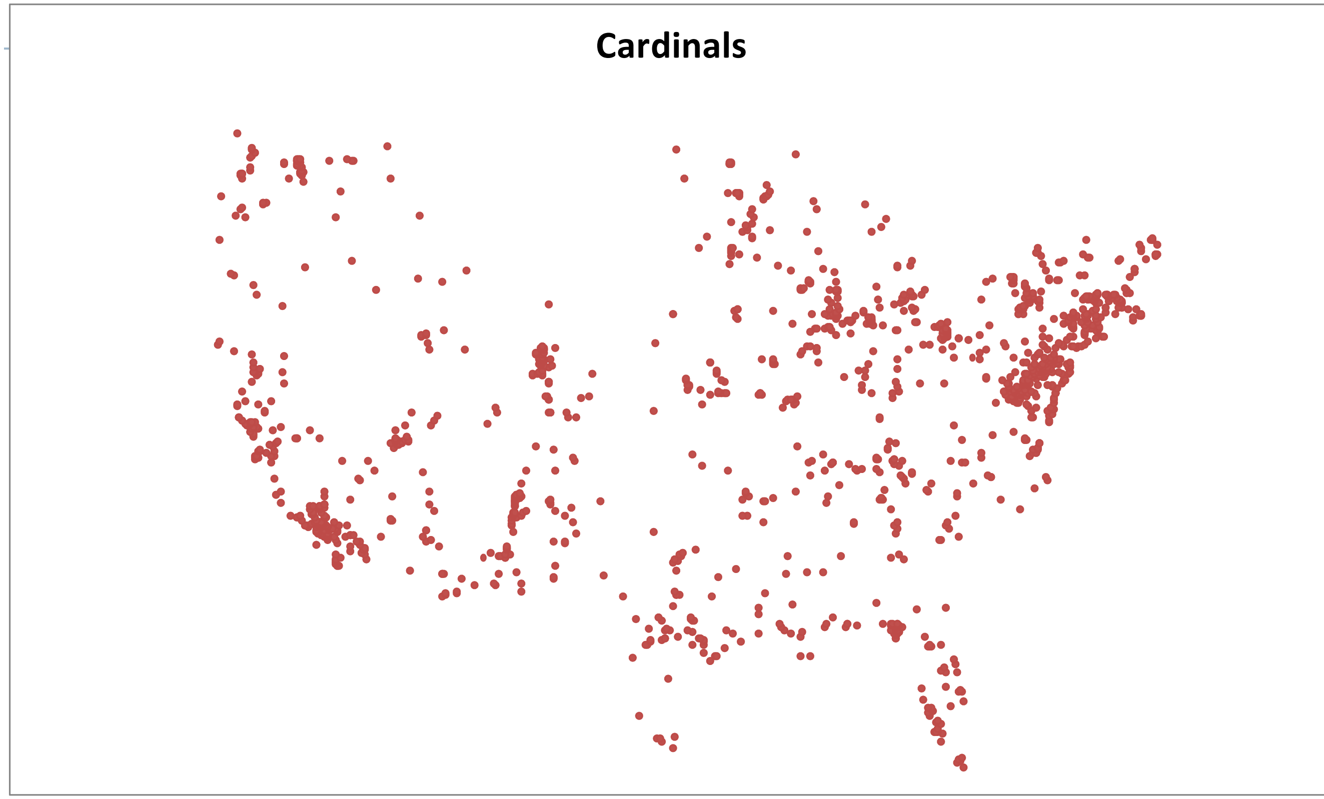


Yu, et al, 2010

A few SDM Challenges

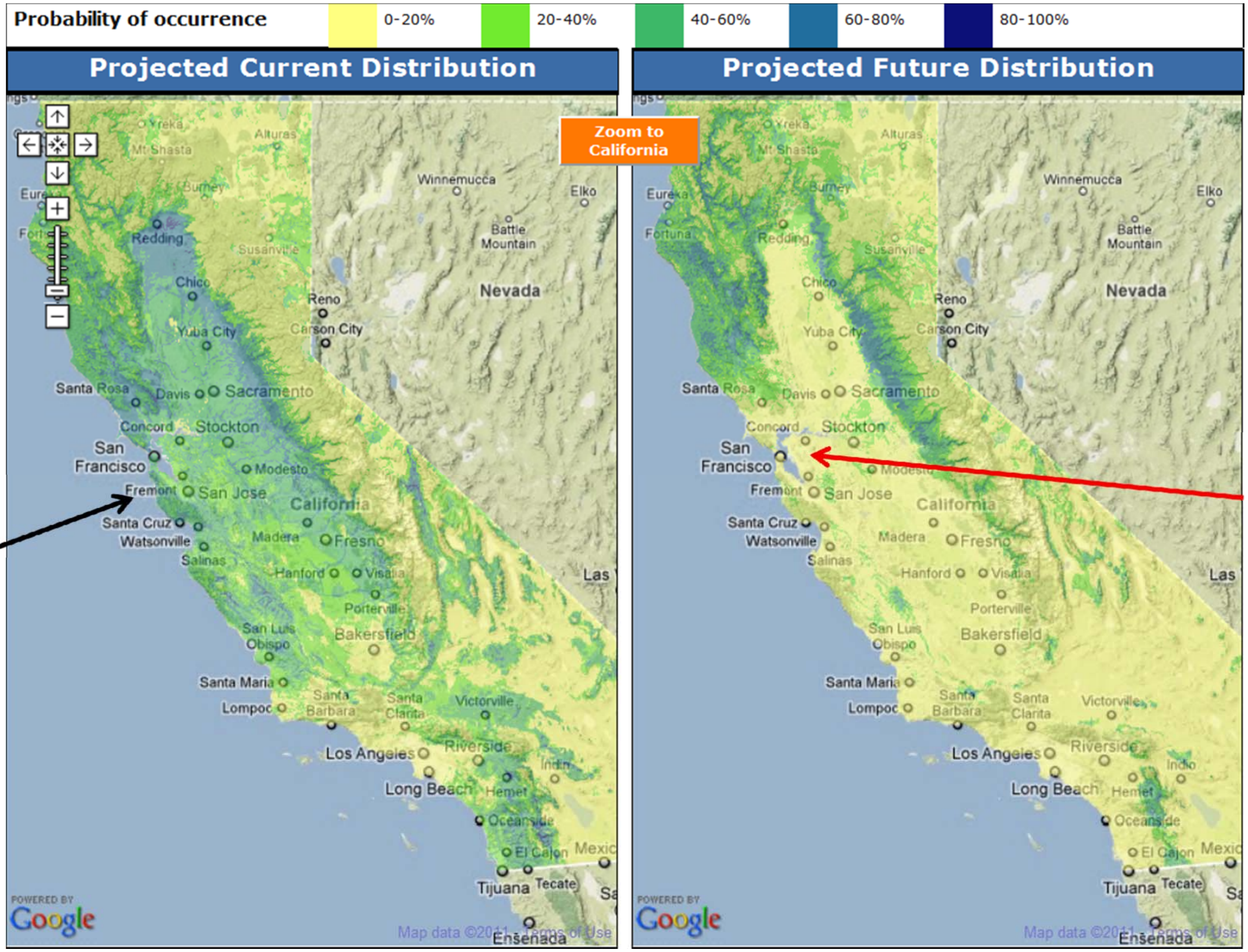
- Presence-only data
- Predictor-response relationships are non-stationary
- Imperfect detection of the species on surveys
  - Often lack prior knowledge of the system for model building
  - Observers have variable expertise/biases
- Sampling bias
- Extrapolation (e.g. under climate change)
- Evaluation strategies
- Estimating temporal trends directly
- More biologically-realistic models
- Multi-species models
- Models of abundance (instead of presence/absence)

Sampling Bias



- eBird participants tend to stay close to home.
- How can we make good predictions uniformly across the U.S.?

Inappropriate Extrapolation



Model learned with data from 1992-2007

Applied to conditions projected for 2070, according to IPCC scenarios

http://data.prbo.org/cadc2/index.php?page=climate-change-distribution

A few SDM Challenges

- Presence-only data
- Predictor-response relationships are non-stationary
- Imperfect detection of the species on surveys
  - Often lack prior knowledge of the system for model building
  - Observers have variable expertise/biases
- Sampling bias
- Extrapolation (e.g. under climate change)
- Evaluation strategies
- Estimating temporal trends directly
- More biologically-realistic models
- Multi-species models
- Models of abundance (instead of presence/absence)

Coffee Break

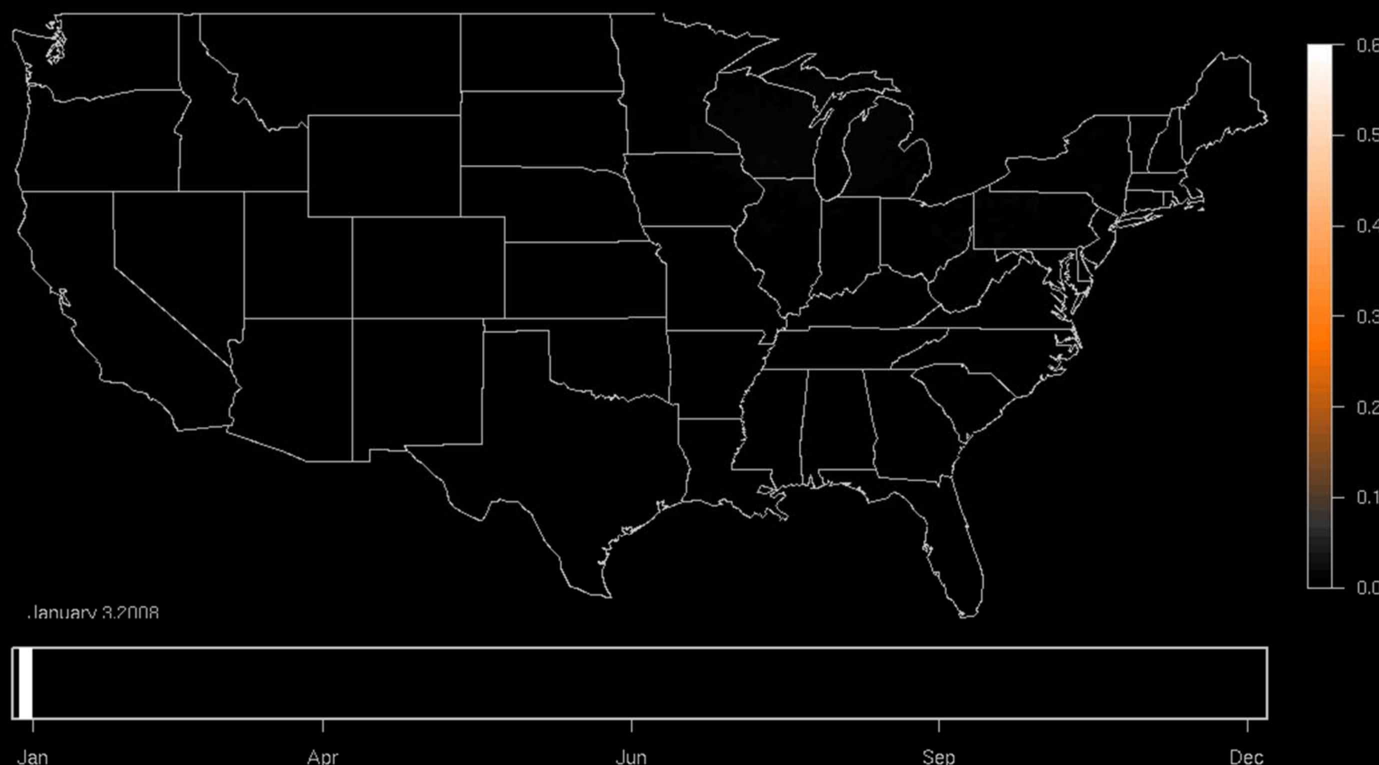
Ecological Models (part 2): Dynamical Models

Dynamical Models

- Ecology is about dynamics!
  - Population growth, animal movement, predator/prey interaction, evolutionary game theory, etc.
- We will look at two particular models of broad-scale population dynamics
  - Bird migration
  - Metapopulations
- Primary motivation: treat species distributions explicitly as spatiotemporal processes
  - Foundation for prediction about future outcomes
  - In contrast with SDMs

Dynamical Model #1: Bird Migration

- Motivation: eBird demonstrates clear migration patterns (but without a dynamical model)



- Unique opportunity to extract quantitative knowledge about migration

Challenges Extracting Migration Knowledge

- Migration is a latent process
  - eBird data and SDM predictions are *static*
    - Each observation/prediction for particular place and time
    - We see a sequence of snapshots

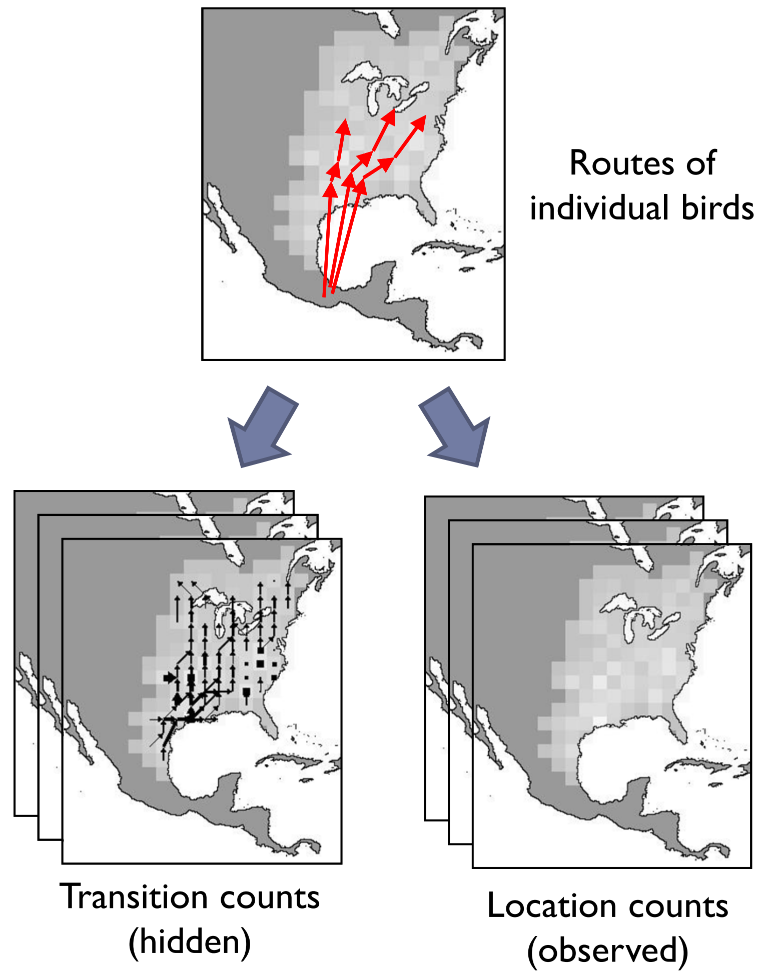
- Observations are noisy and incomplete

- Migration most naturally described at level of individual behavior, but we can only observe population-level statistics
  - Lack of modeling techniques to link the two

Overview: Collective Hidden Markov Models

- Generative model for population data from individual behavior

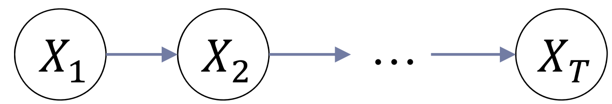
1. Define migration model for individual bird
   - Markov chain on grid cells
2. Generate routes for each individual in population
   - Assume birds are iid
3. Derive population statistics at each time step
   - Transition counts: # birds that fly between each pair of grid cells
   - Location counts: # birds in each grid cell



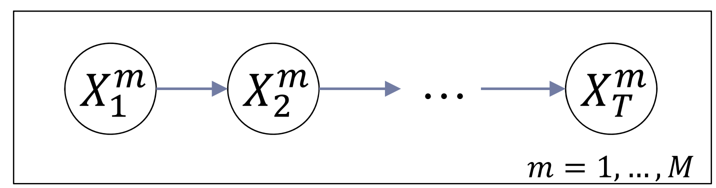
Sheldon et al. 2007; Sheldon 2010

Overview: Collective Hidden Markov Models

Individual model:
Markov chain on grid cells

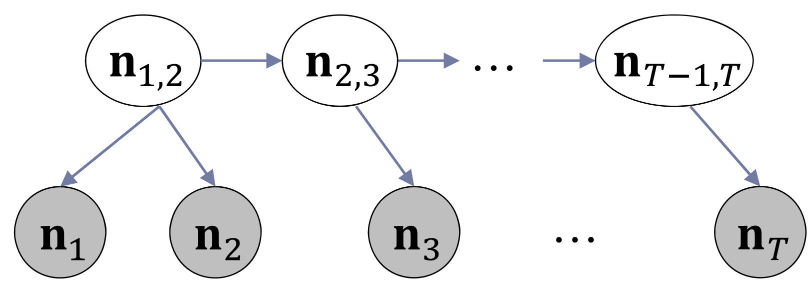


Population model:
iid copies of individual model



Marginalize out individuals:
chain-structured model on
sufficient statistics

Transition counts

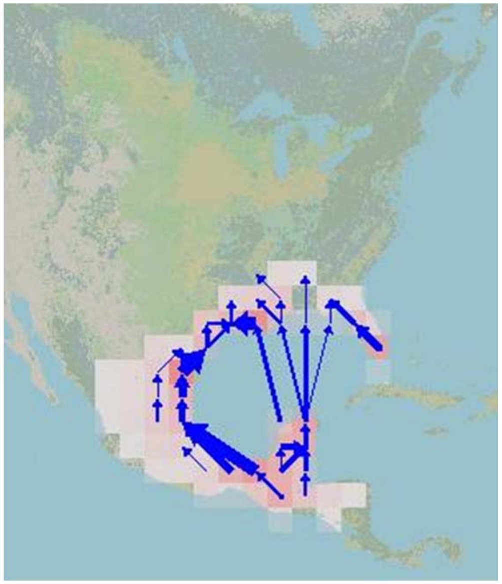


Add observations: location counts

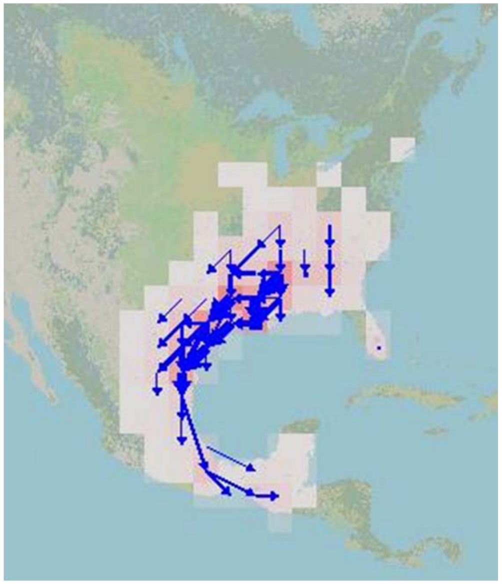
Sheldon et al. 2007; Sheldon 2010

Results

- Reconstruction by network flow techniques
- Use to visualize bird migration
  - E.g. Ruby-throated Hummingbird



Northbound
March 5



Southbound
October 1

Sheldon et al. 2007; Sheldon 2010

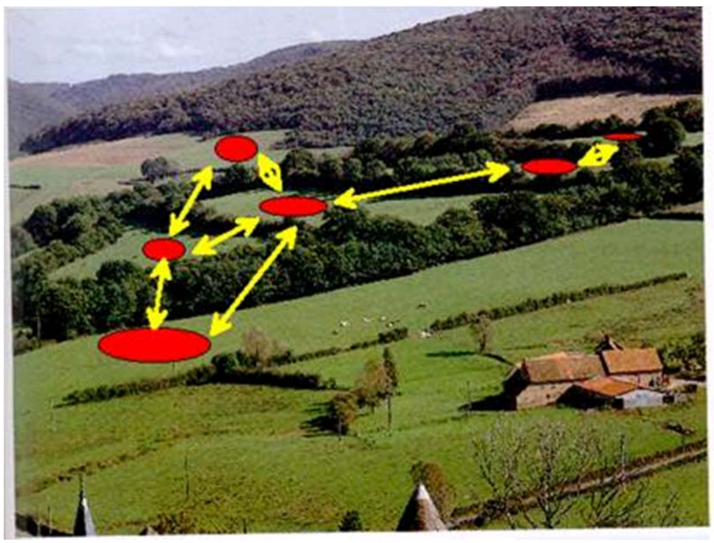
Extensions

- Collective Graphical Models
  - Substantial generalization of modeling ideas
  - Parameter learning
- Future: *BirdCast*
  - Joint project with Cornell Lab of Ornithology
  - Apply these ideas to forecast bird migration at continent-scale
    - Data: eBird + radar + acoustic + weather
  - Funding available!
    - Seek outstanding candidate to begin Ph.D. at OSU
    - Interest in ML and ecology

Dynamical Model #2: Metapopulations

- Dynamics of spatially disjoint populations
  - Butterflies in alpine meadows
  - Birds in a fragmented forest

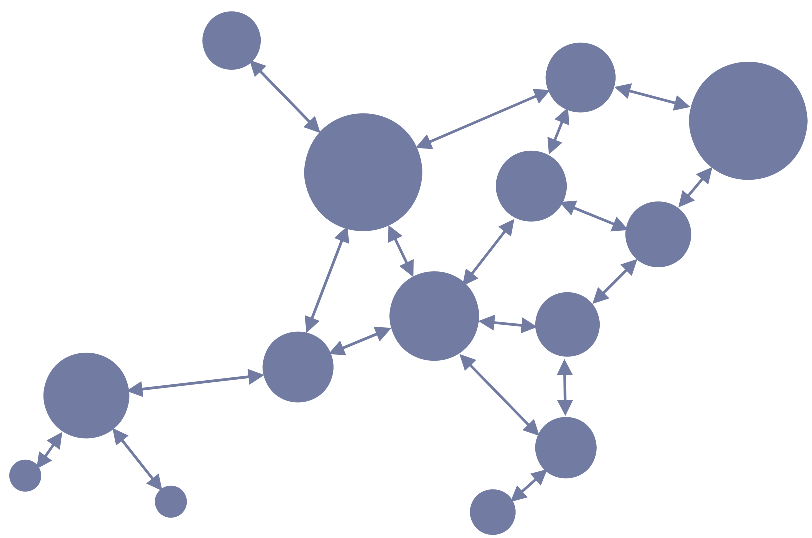
http://www.bio.uni-potsdam.de/



Metapopulation = population of populations

Basic Components

- A network of *habitat patches*
- Dynamics models
  - *Local population dynamics in each patch*
  - *Interaction between patches (dispersal/colonization)*



Metapopulation Background

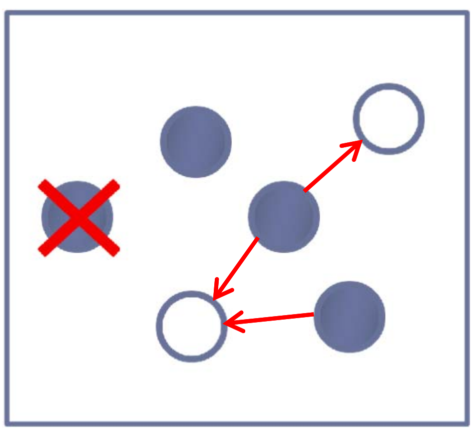
- Extremely important models in ecology

- Thousands of articles dating to 1960s with many modeling variations
  - Originally mathematical models for idealized landscapes
    - E.g. equidistant patches
  - Move to applied models, real landscapes

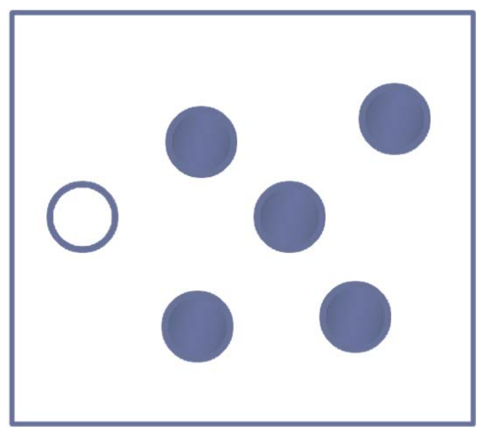
- Importance: formal basis for reasoning about the effects of habitat configuration on species persistence

SPOM: Stochastic Patch Occupancy Model

- Patches are occupied or unoccupied
- Two types of stochastic events:
  - *Local extinction:* occupied → unoccupied
  - *Colonization:* unoccupied → occupied (from neighbor)
- Independence among all events



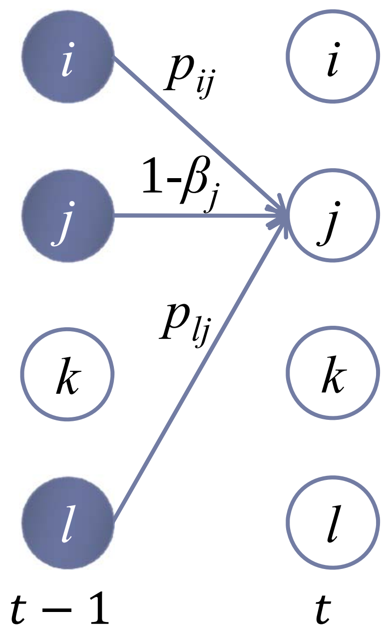
Time 1



Time 2

Hanski, 1994

SPOM Probability Model

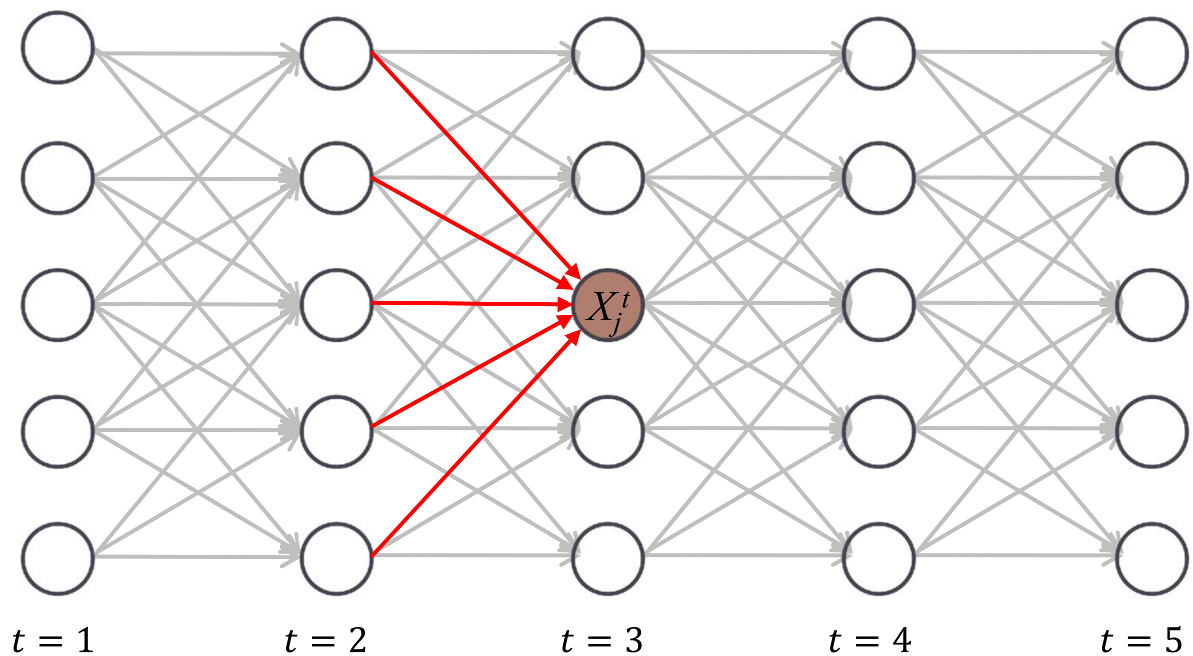


- To determine occupancy of patch j at time t
  - For each *occupied* patch $i \neq j$ from time $t-1$, flip coin with probability p_{ij} to see if i colonizes j
  - If j is occupied at time $t-1$, flip a coin with probability $1-\beta_j$ to determine survival (non-extinction)
  - If any of these events occurs, j is occupied

- Parameters:
  - p_{ij}: colonization probability
  - β_j: extinction probability
  - Simple parametric functions of patch-size, inter-patch distance, etc.

SPOM as Dynamic Bayes Net (DBN)

- Let $X_j^t = 0$ or 1 be occupancy of patch j at time t



$$\Pr\left(X_j^t = 1 \,\middle|\, \mathbf{X}^1, \dots, \mathbf{X}^{t-1}\right) = \Pr(X_j^t = 1 | \mathbf{X}^{t-1})$$

SPOM Fitting

- Major advance in practical utility of SPOMs was ability to fit to survey data
  - Given: Occupancy vectors $\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^T$
  - Find: Parameters Θ for colonization and extinction models

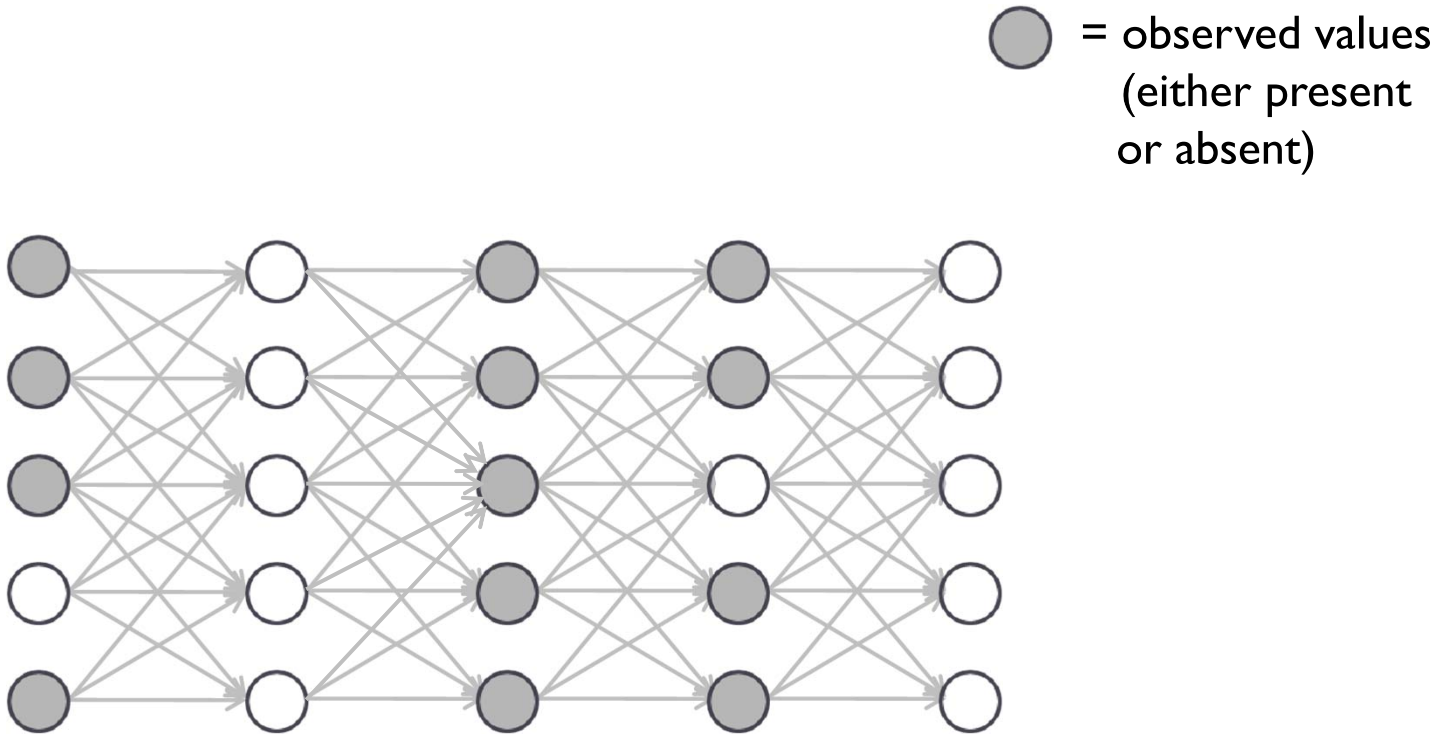
- Hanski [1994] gave heuristic approach based on equilibrium properties of metapopulation

- Moilanen [1999]
  - Maximum likelihood approach

$$L(\Theta) = p(\mathbf{X}^1; \Theta) \prod_{t=2}^{T} p(\mathbf{X}^t | \mathbf{X}^{t-1}; \Theta)$$

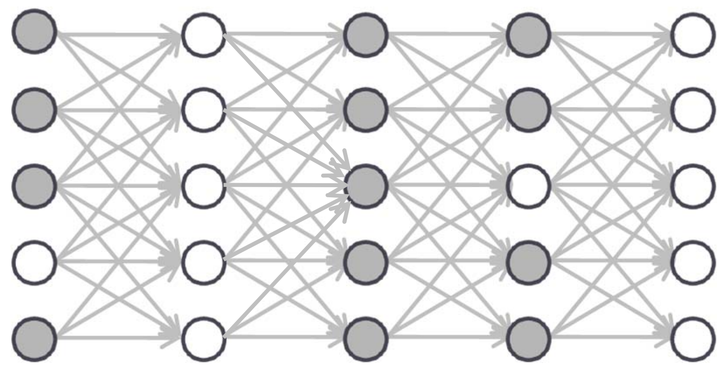
  - Easy in principle
    - Likelihood easy to evaluate
    - Small parameter space

Challenge: Missing Data



- Field data is sparse and messy
  - Surveys conducted in non-consecutive time steps
  - Some patches are not surveyed

Fitting by Data Augmentation



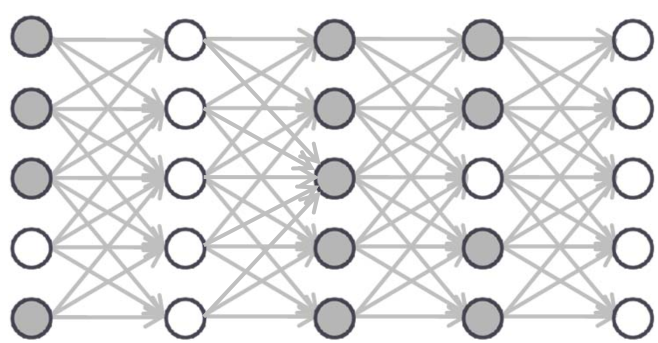
- Key step: fill in missing data by sampling from distribution of missing data given observed data
  - Maximum-likelihood approach of Moilanen [1999]
  - Bayesian approach of Ter Braak and Etienne [2003]

ML Opportunity

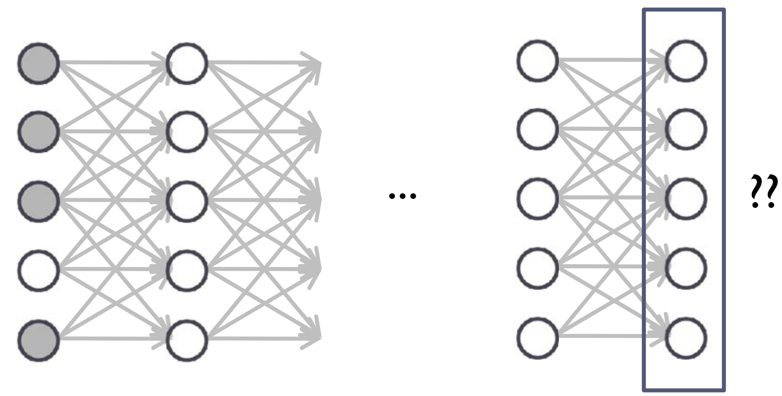
- Improved methods for fitting?

- Key step is inference in p(missing | observed)
  - I.e., inference in DBN with metapopulation structure
  - Approximate inference techniques

- Importance of inference:



Fitting

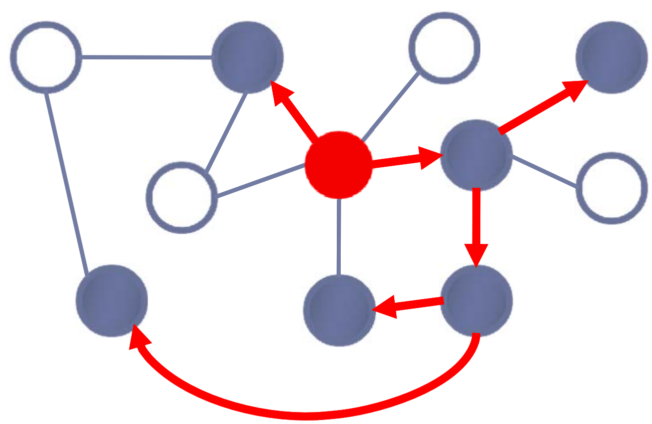


Prediction

Connections to Network Cascades

- Models for diffusion in (social) networks
  - Spread of information, behavior, disease, etc.

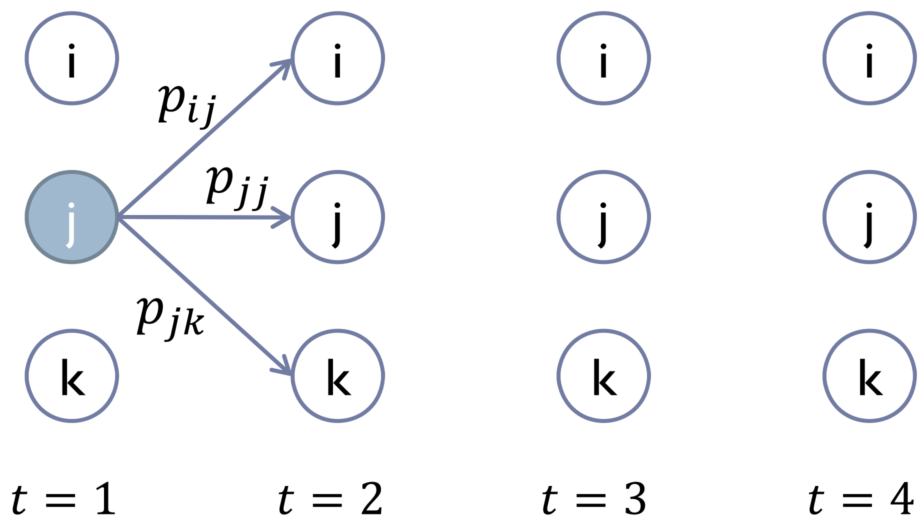
- *Independent cascade model*
  - Each individual passes information to friends independently with specified probability



[Goldenberg, Libai, Muller 2001][Kempe, Kleinberg, Tardos 2003]

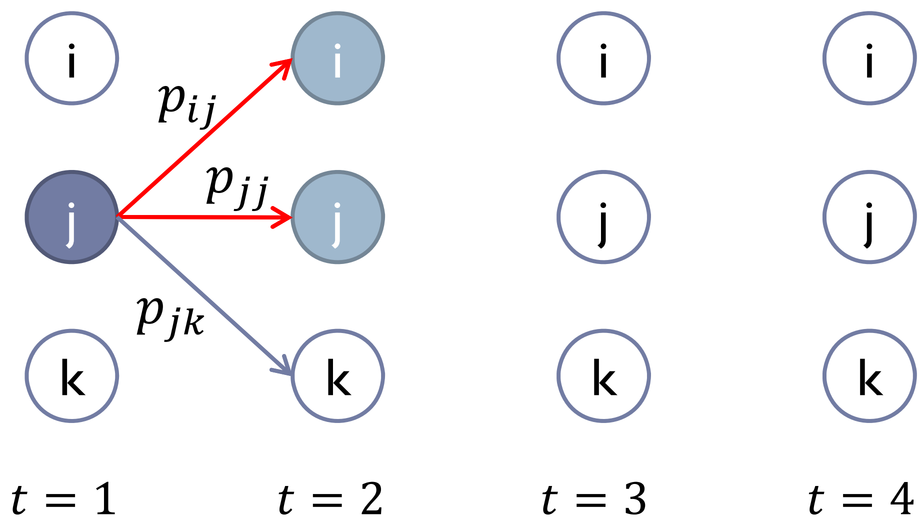
Metapopulation = Cascade

- SPOM dynamics are a cascade in the layered graph representing patches over time



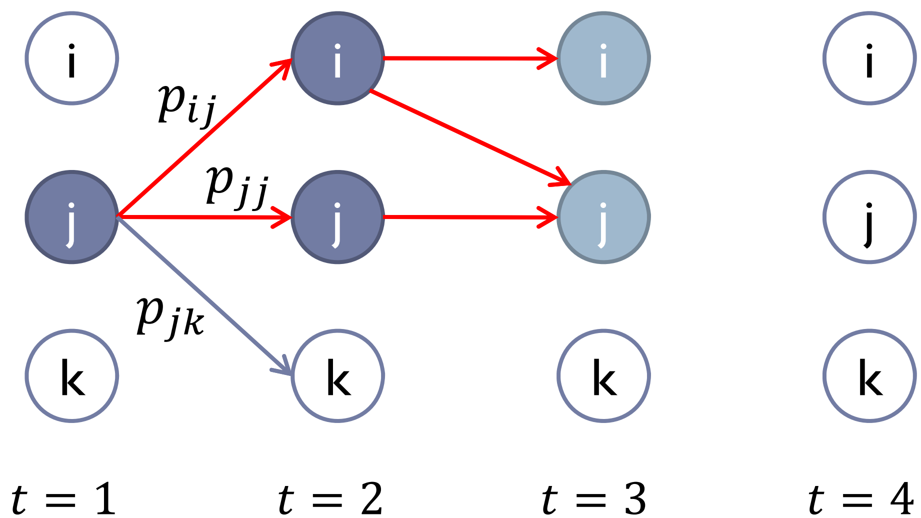
Metapopulation = Cascade

- SPOM dynamics are a cascade in the layered graph representing patches over time



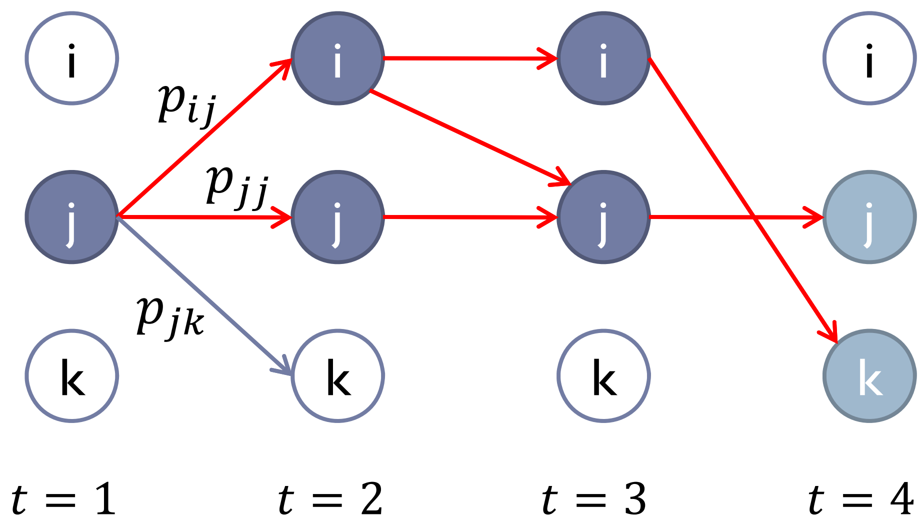
Metapopulation = Cascade

- SPOM dynamics are a cascade in the layered graph representing patches over time



Metapopulation = Cascade

- SPOM dynamics are a cascade in the layered graph representing patches over time



ML connection: Social Network Inference

- Recent work in ML community to learn cascade models
  - Network is hidden
  - Observe *infection times* of nodes
- Maximum-likelihood estimation by convex optimization
  - [Myers and Leskovec, 2010]
  - [Gomez-Rodriguez et al., 2011] *(this conference)*
- Applicability to SPOM fitting?
  - Model differences
    - Layered vs. non-layered graph
    - Time model
  - Much different parameterization

Part 3: Policy Optimization

Outline

- Data Acquisition
  - Sensors: Physical sensors, human observers, repurposing data from other sources
  - Data interpretation: Extracting signals from data
- Ecological Models
  - Species Distribution Models
  - Dynamical Models: Dispersal, Migration, Invasion, Climate Change
- Policy Optimization
  - Conservation: Reserve design, Network design
  - Invasive species: Eradication, restoration, monitoring
  - Fisheries: Managing harvest levels

Optimal Policies for Environmental Management

- One-shot problems
  - Network design
  - Reserve design
- Sequential decision-making problems (known as "Active Management")
  - Fisheries management
  - Fire management
  - Invasive species management
  - Reserve design and conservation easements over time
- Most problems are really sequential decision-making problems

Distinctive Aspects

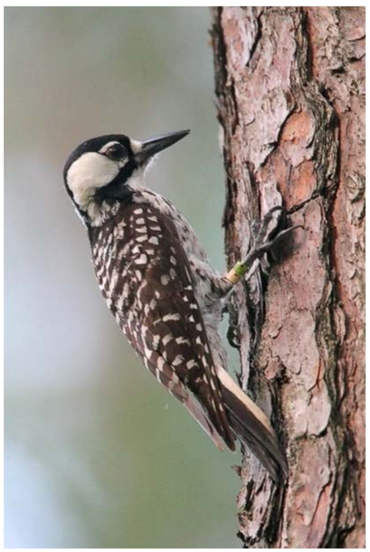
- Optimizing an objective computed using a *learned* model of the system
  - Generalization of reinforcement learning
- Models are typically very bad
  - Doak, et al. 2008: Ecological Surprises
    - "Surprises are common and extreme"
  - Costs and benefits may be highly uncertain and non-stationary
  - Multiple objectives: Harvest + Species Viability
  - Need solutions that are robust to misspecified models
- Large state and action spaces
  - Spatial models

Plan

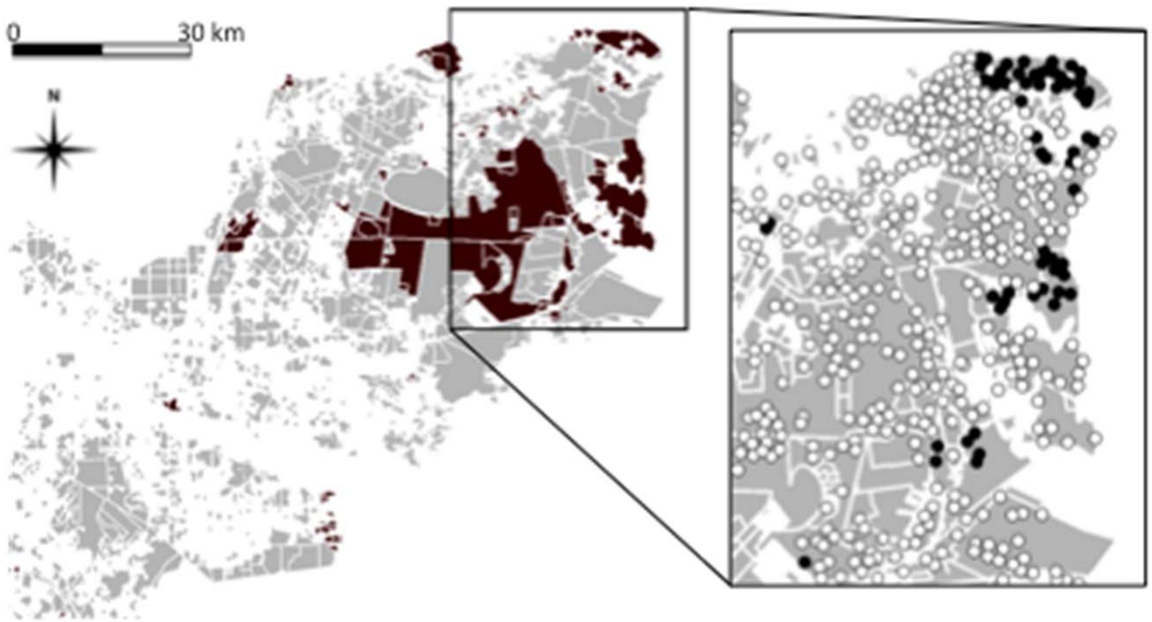
- Reserve Design for the Endangered Red Cockaded Woodpecker
  - One-shot design problem
- Optimal Policies for Managing Fisheries
  - Markov Decision Problem with analytical characterization of the optimal policy
- Managing Wildfire in Eastern Oregon
  - Large Spatial Markov Decision Problem (MDP)
- Optimal Management of Difficult-to-Observe Invasive Species
  - Small Partially-Observable MDP (POMDP)

SPOM Optimization: Reserve Design for Endangered Species

- Given a limited budget to purchase additional patches, which should you buy?



Red-cockaded woodpecker (endangered)



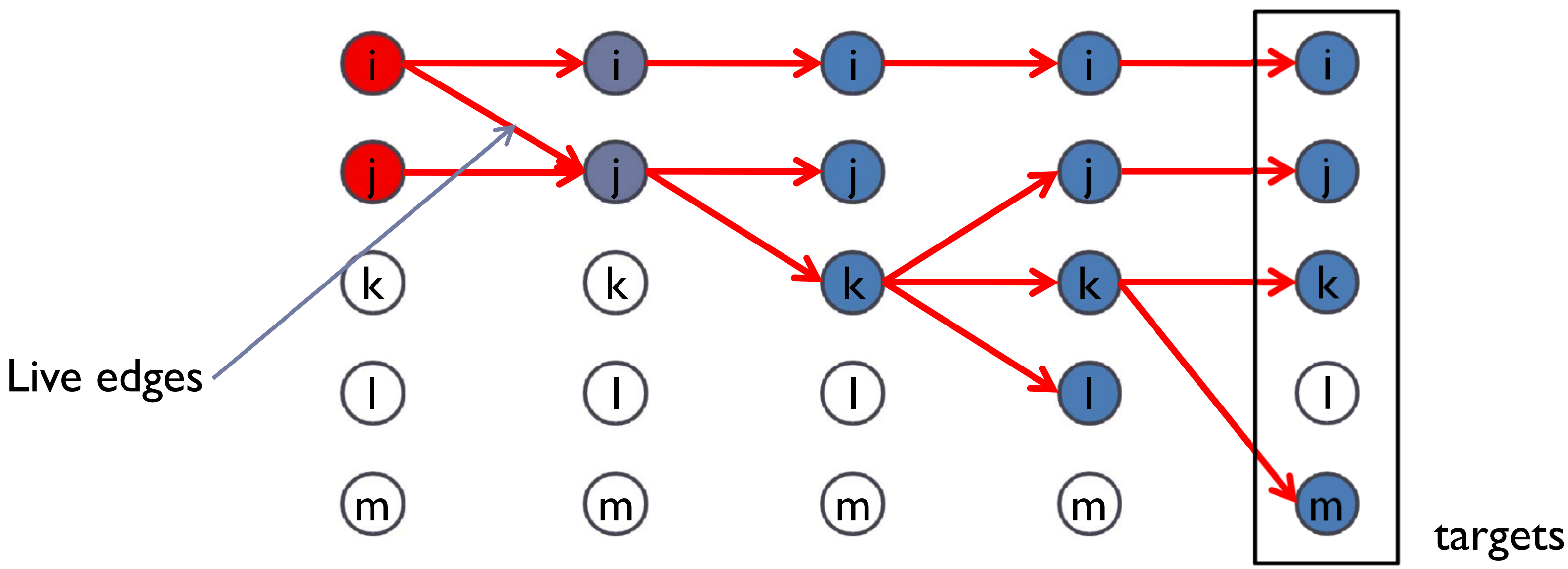
Sheldon et al., 2010

Key Observation

- By viewing SPOM dynamics as a network cascade in the layered graph, we can formulate the conservation problem as a cascade optimization problem

Insight #1: Objective as Network Connectivity

- Conservation objective: maximize expected # occupied patches at time T

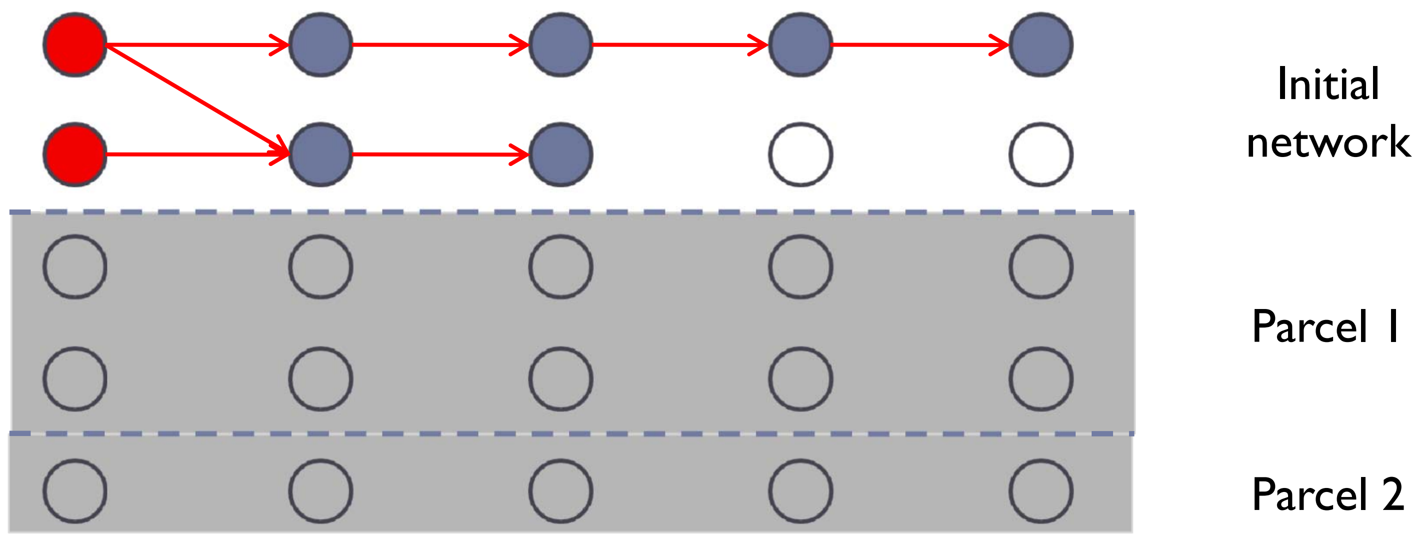


Occupied patches = nodes reachable by live edges

Sheldon et al., 2010

Insight #2: Management as Network Building

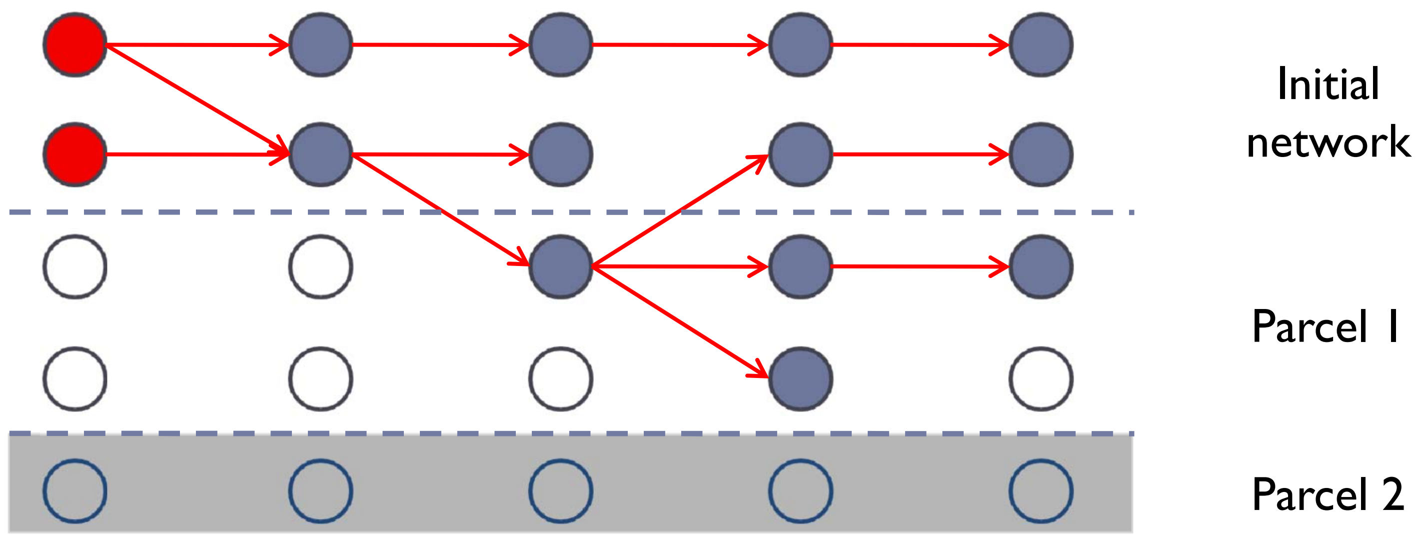
- Conserving parcels adds nodes to the network



Sheldon et al., 2010

Insight #2: Management as Network Building

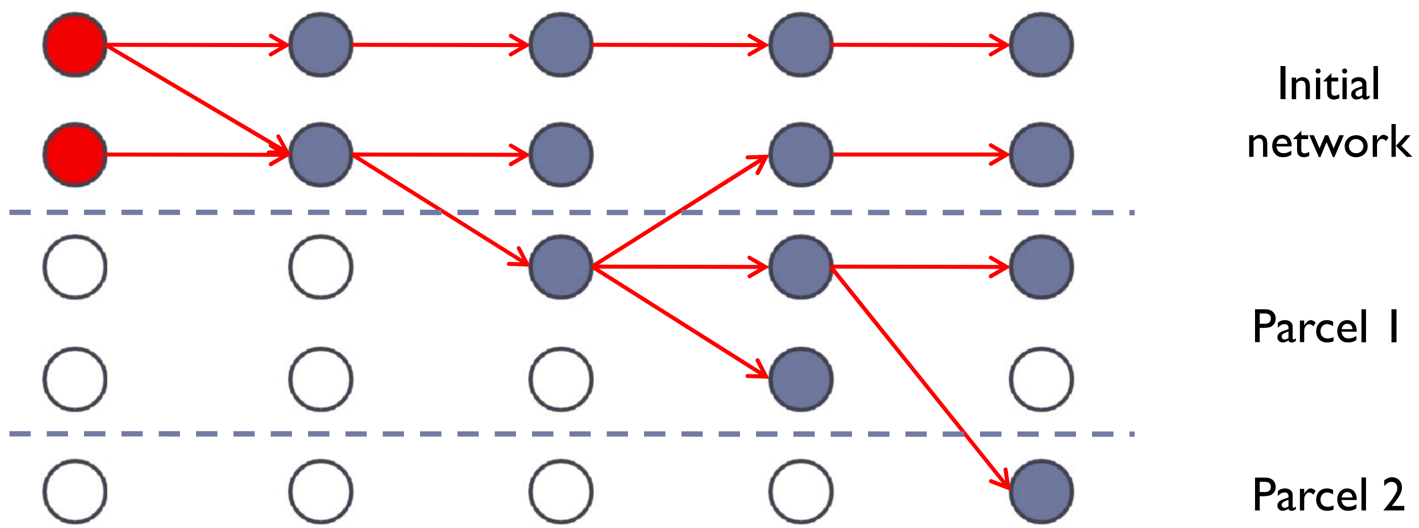
- Conserving parcels adds nodes to the network



Sheldon et al., 2010

Insight #2: Management as Network Building

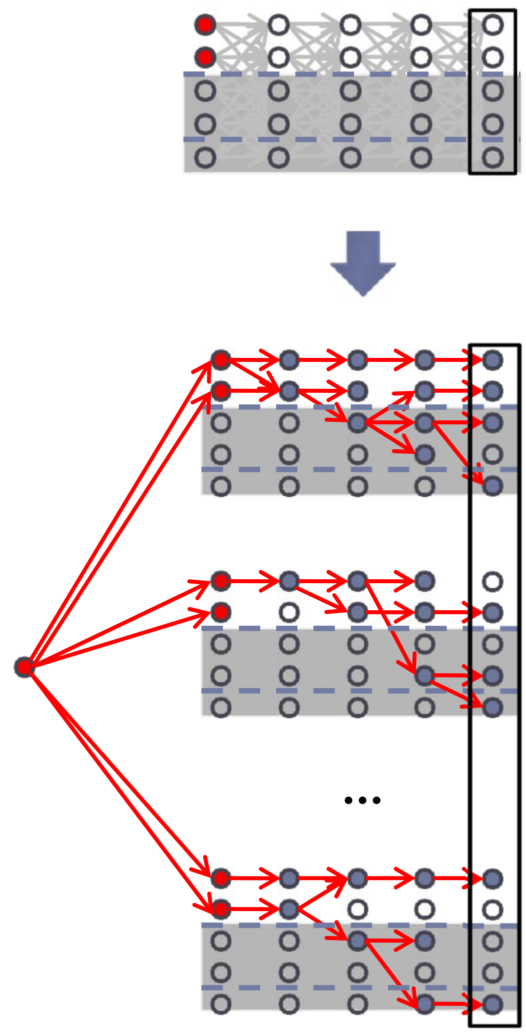
- Conserving parcels adds nodes to the network



Sheldon et al., 2010

Solution Strategy

1. Formulate as *stochastic network design problem* with limited budget
2. Convert stochastic problem to deterministic network design problem
   - Sample Average Approximation (SAA)
3. Solve as a mixed integer program



Sheldon et al., 2010

Sample Average Approximation (SAA)

- Generic approach to convert stochastic problem to deterministic problem:

$$\max_X E_Y[f(X,Y)] \quad \longrightarrow \quad \max_X \frac{1}{N}\sum_{i=1}^{N} f(X,Y_i)$$

  - X: decision variable
  - Y: random variable
  - $Y_1, \ldots, Y_N$: realizations of Y

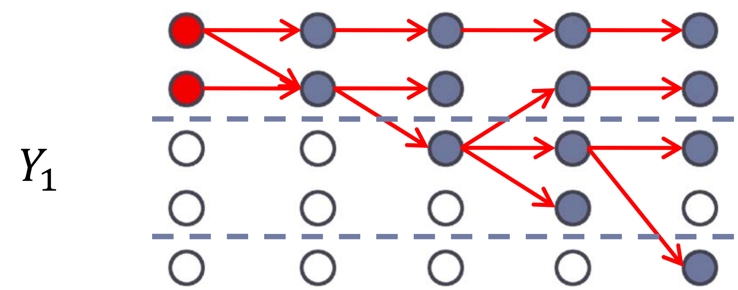
- Nice properties
  - Converges to true optimum as $N \to \infty$
  - Error bounds

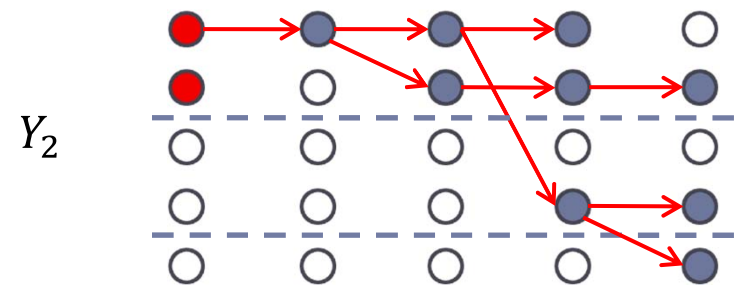
- Can we solve the sample average problem?

Sheldon et al., 2010

SAA and Network Design

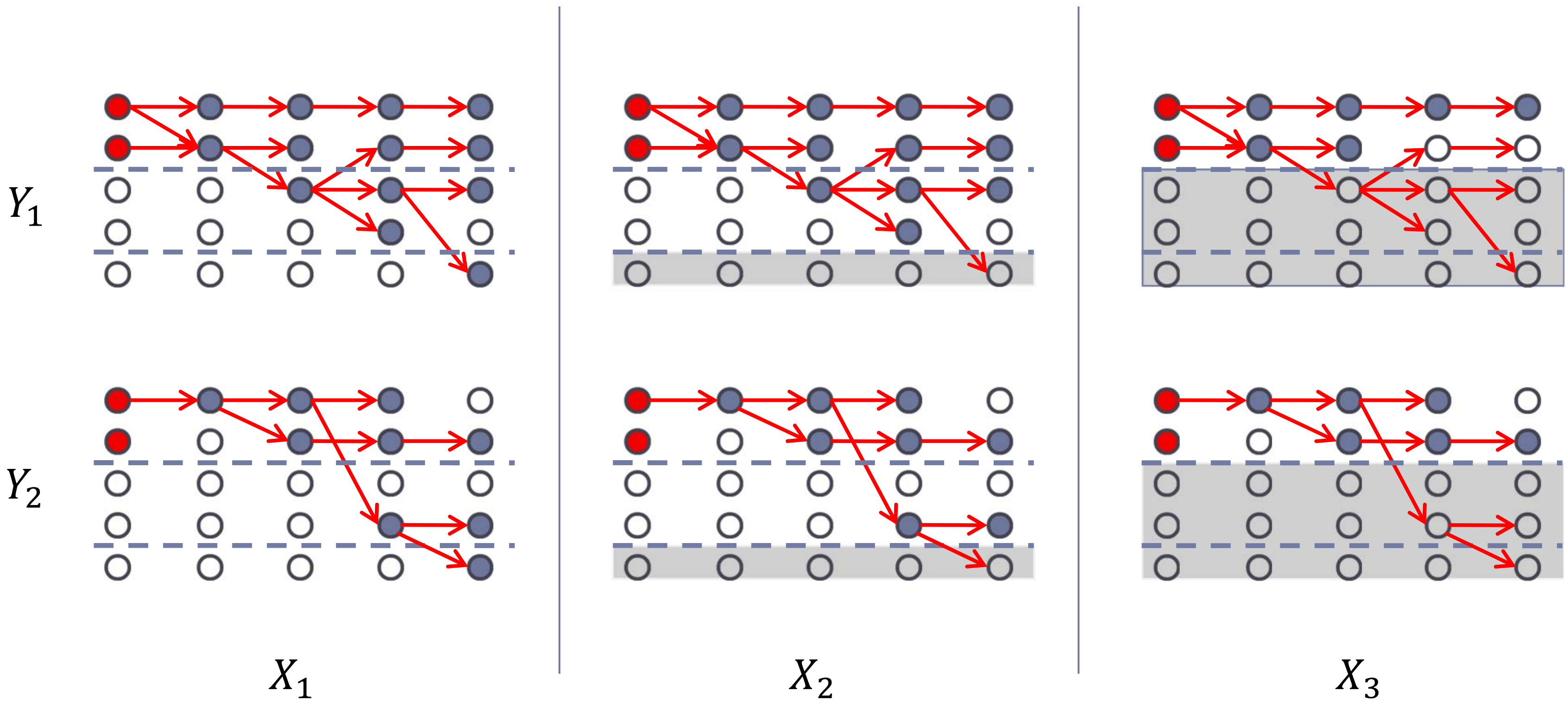
- In our problem
  - X: purchased network components
  - Y: edges of cascade
- Natural way to reuse random variables Y for many different X
  - Y: Simulate cascade in advance in *full network*
  - $f(X,Y)$: measure connectivity in *subgraph* purchased by X



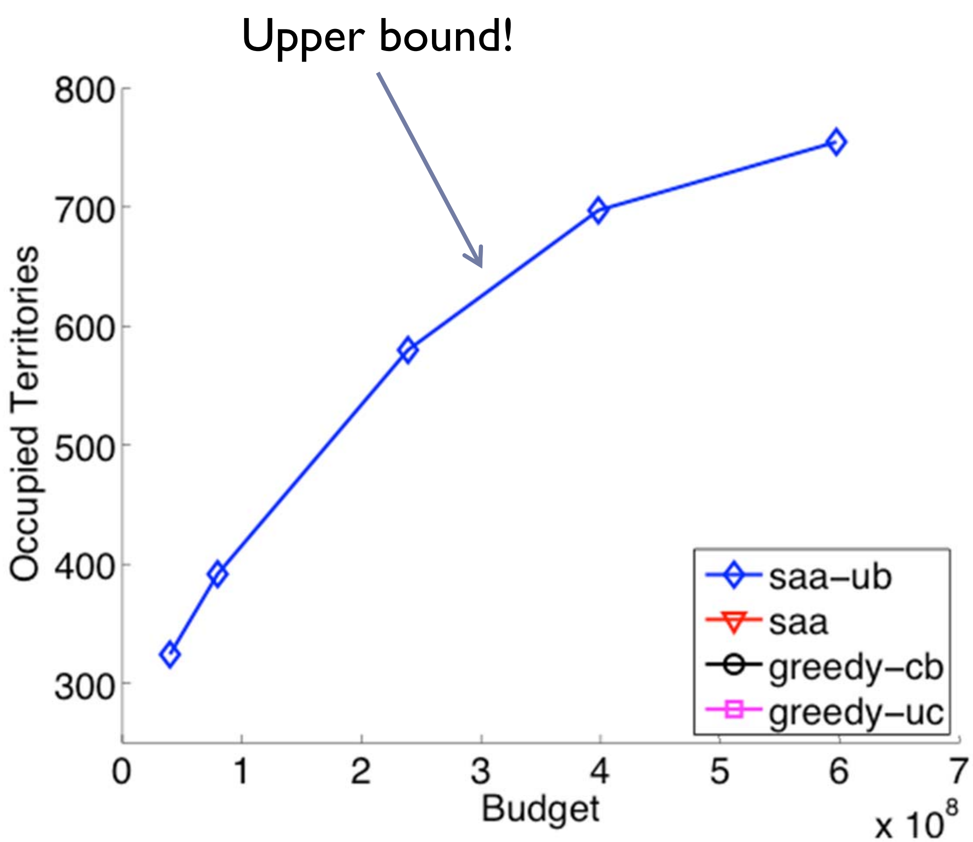


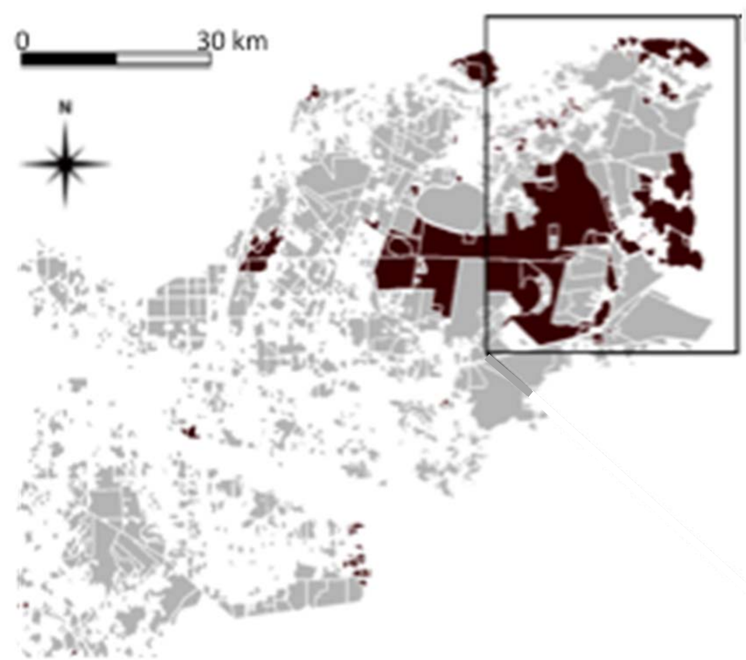
Sheldon et al., 2010

Example



Results

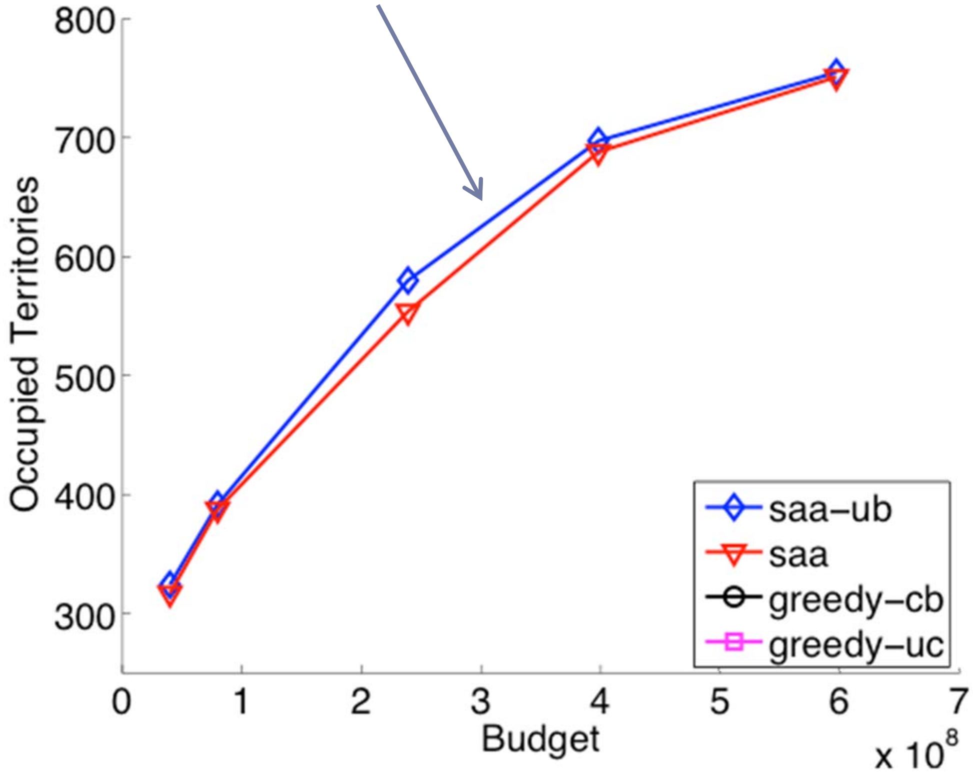


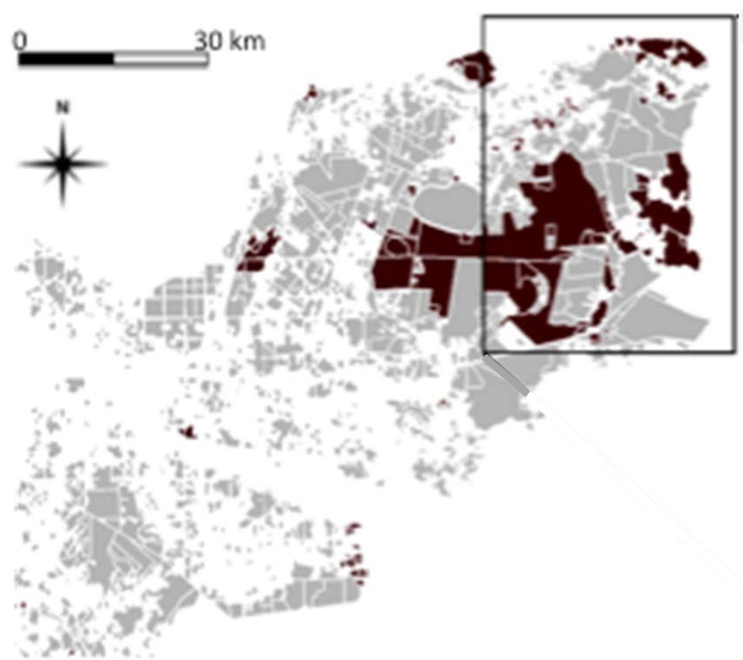


Sheldon et al., 2010

Results

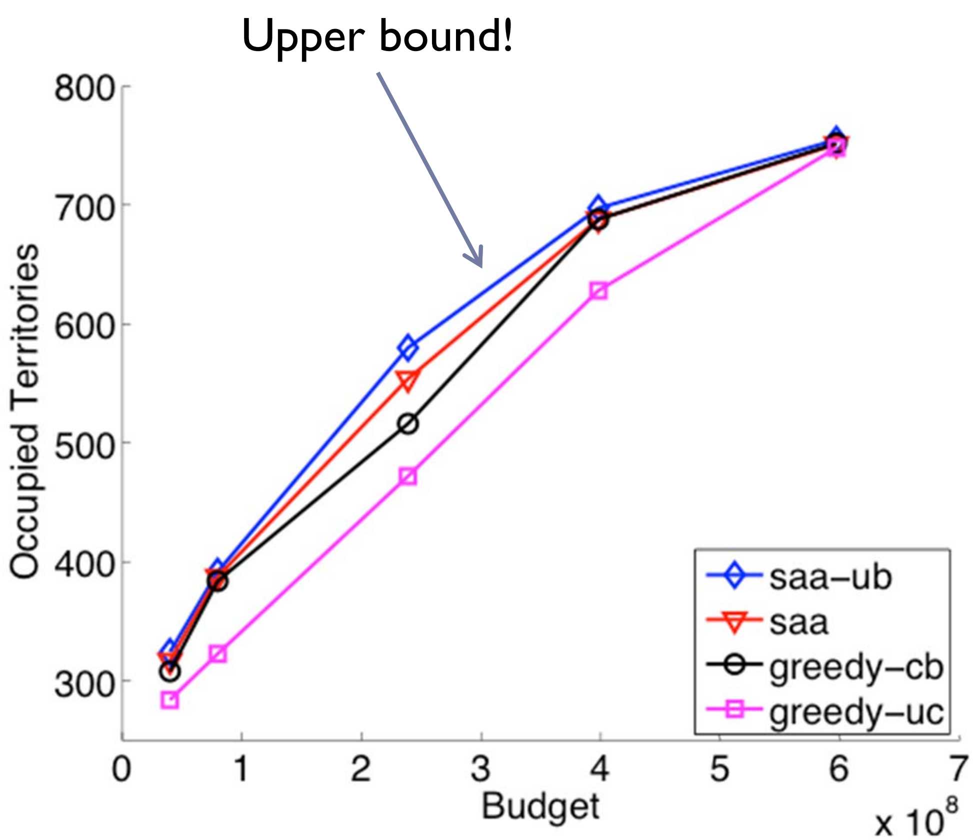
Upper bound!





Sheldon et al., 2010

Results

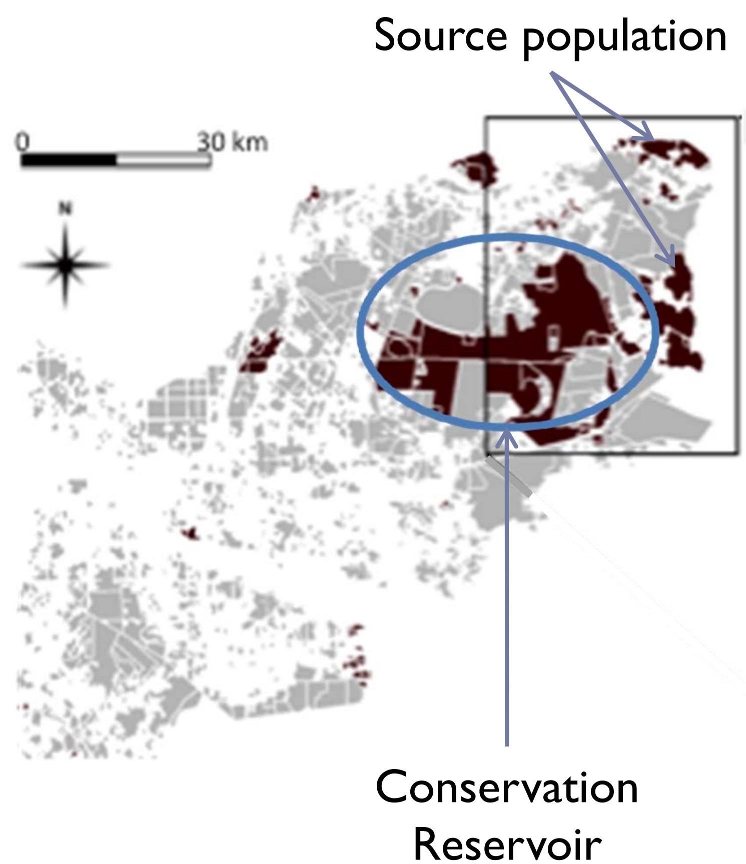


- Greedy baselines from related submodular cascade optimization problems
  - [Kempe et al. 2003]
  - [Leskovec et al. 2007]
- Our problem is *not* submodular
  - Why is greedy performing well?

Sheldon et al., 2010

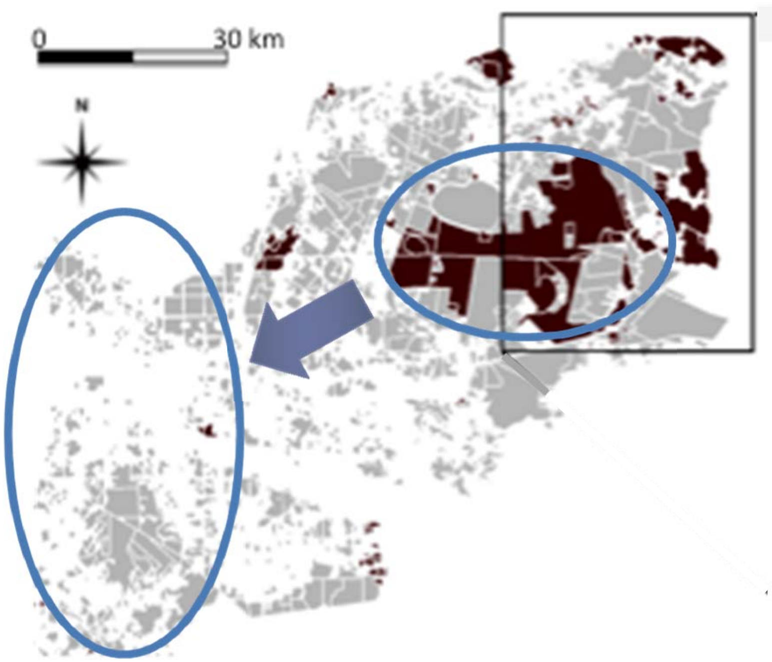
Conservation Strategies

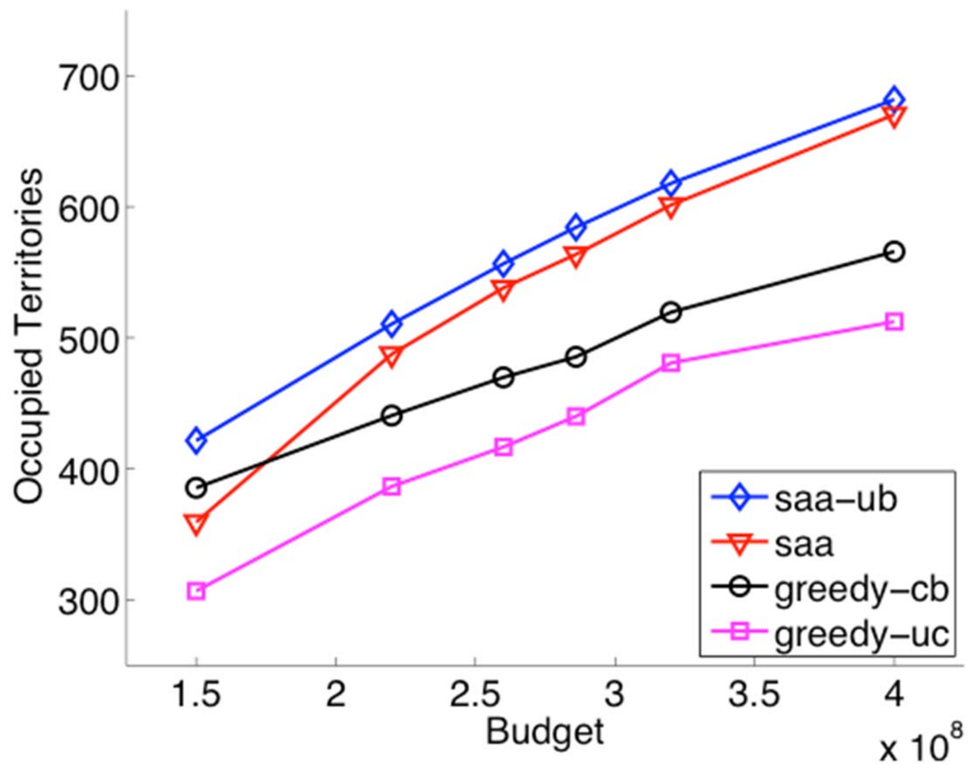
- Both approaches build outward from source
  - Greedy buys best patches next to currently owned patches
  - Optimal solution builds toward areas of high conservation potential
- In this case, the two strategies coincide



Sheldon et al., 2010

A Harder Instance





Move the conservation reservoir so it is more remote.

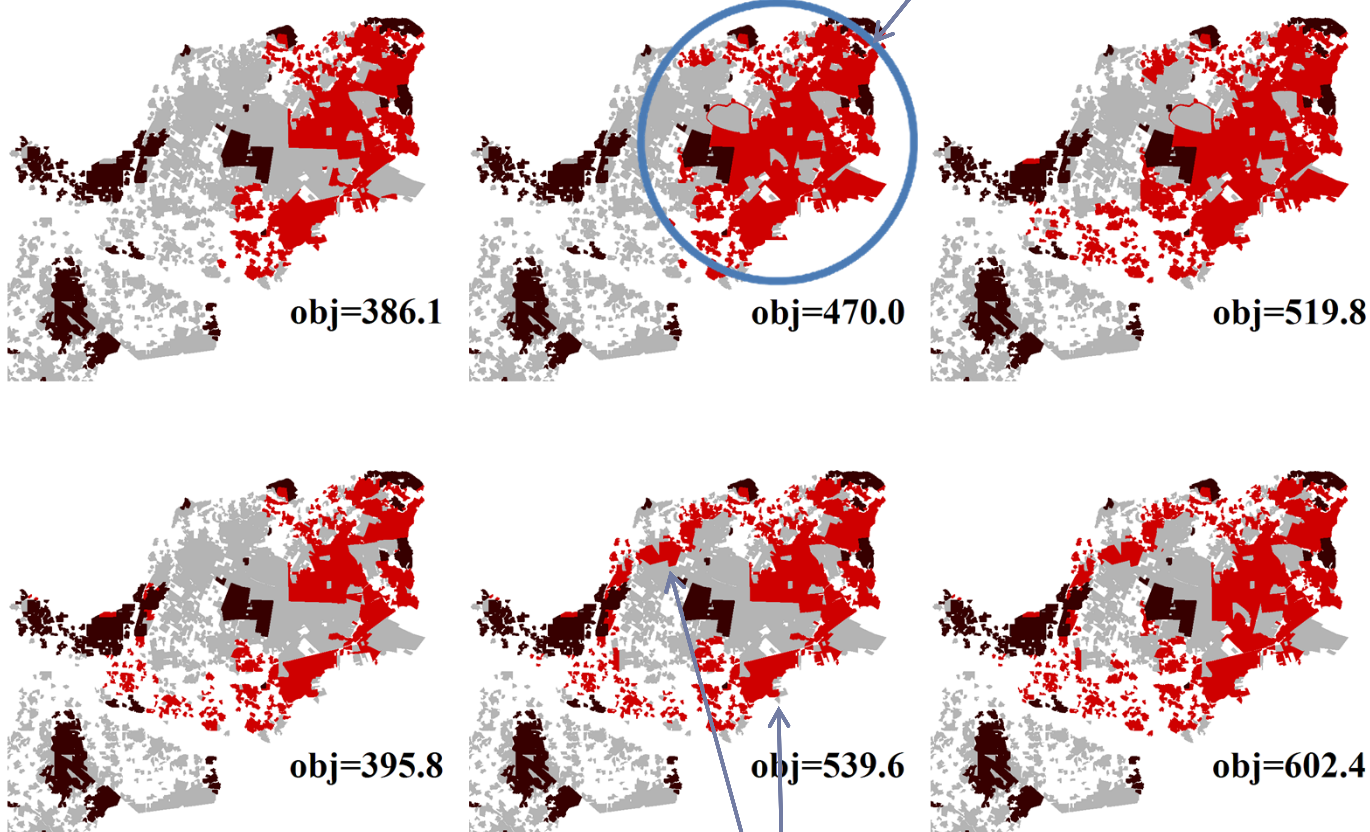
Sheldon et al., 2010

Conservation Strategies

Build outward from sources

Greedy Baseline

SAA Optimum (our approach)



$150M

$260M

$320M

Path-building (goal-setting)

Summary and Future Challenges

- Spatial optimization with underlying dynamics model is a very promising direction

- The real world is complex
  - Competing objectives
    - Multiple species
    - Resource economics
  - Model dynamics
    - Species interactions
    - Uncertainty about dispersal behavior
      - Combine learning and optimization
  - Adaptive management
    - Budget comes in installments [Golovin et al. 2011]

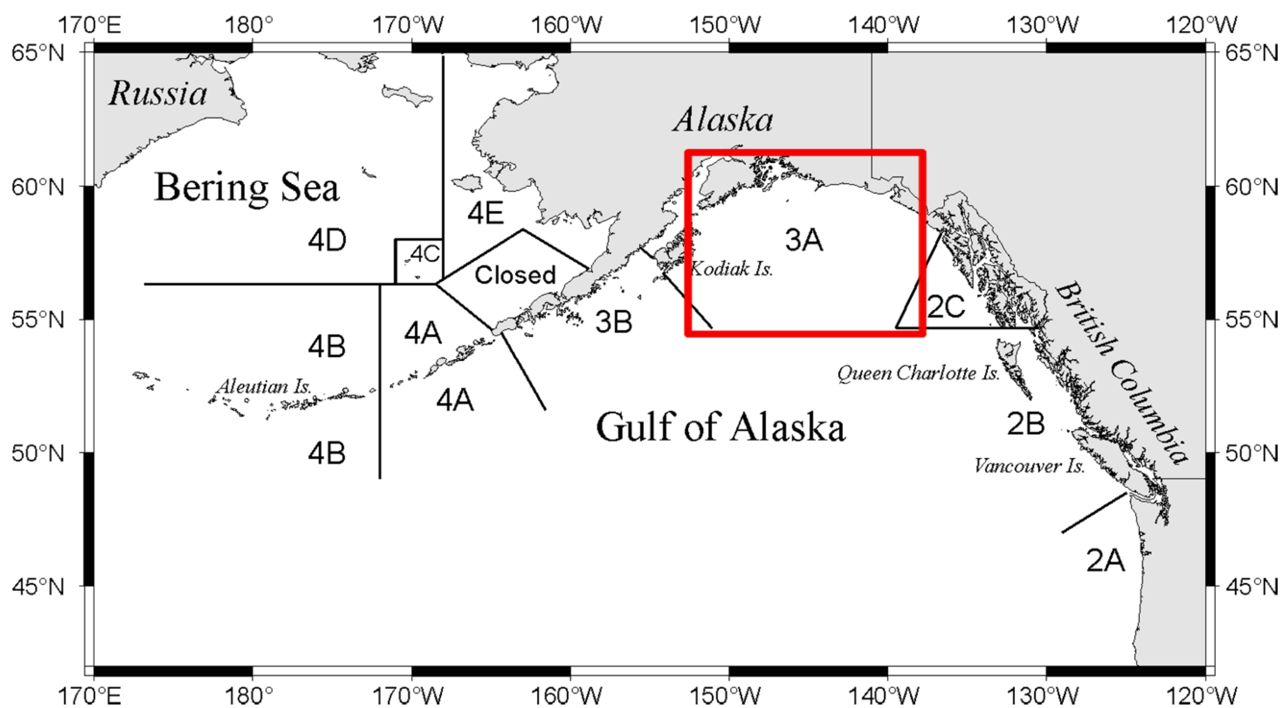
Sheldon et al., 2010

Fishery Management [Ermon et al. 2010]

- How to sustainably exploit a renewable and economically valuable resource such as forest or fishery?

Pacific Halibut Fishery

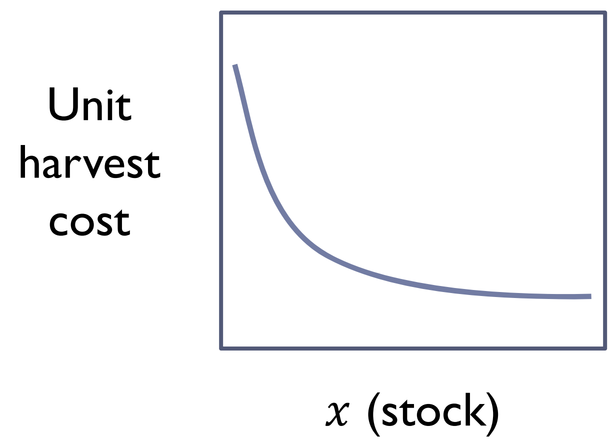
International commission decides each year's harvest (total allowable catch)



Ermon et al., 2010

MDP Formulation

- State variable
  - x: stock (population size)
- Actions
  - Harvest amount h in each year
- Reward model
  - Fixed cost K when $h > 0$
  - Per-unit harvest cost
    - More $$ when fish are scarce
  - Per-unit market price p
  - Discount rate for future reward



Ermon et al., 2010

MDP Formulation (2)

- Dynamics
  - Growth function f (post-harvest)

$$x_{t+1} = f(x_t - h_t, w_t)$$

stock harvest input from nature

  - Idea: w_t captures stochasticity or modeling uncertainty

- State transition model
  - $x \rightarrow f(x - h, w)$ with probability $p(w)$

Population Dynamics

- Beverton-Holt Model

Growth rate

$$x_{t+1} = f(s_t, w_t) = (1-m)s_t + w_t \frac{r_0 s_t}{1 + s_t/M}$$

Post-harvest stock

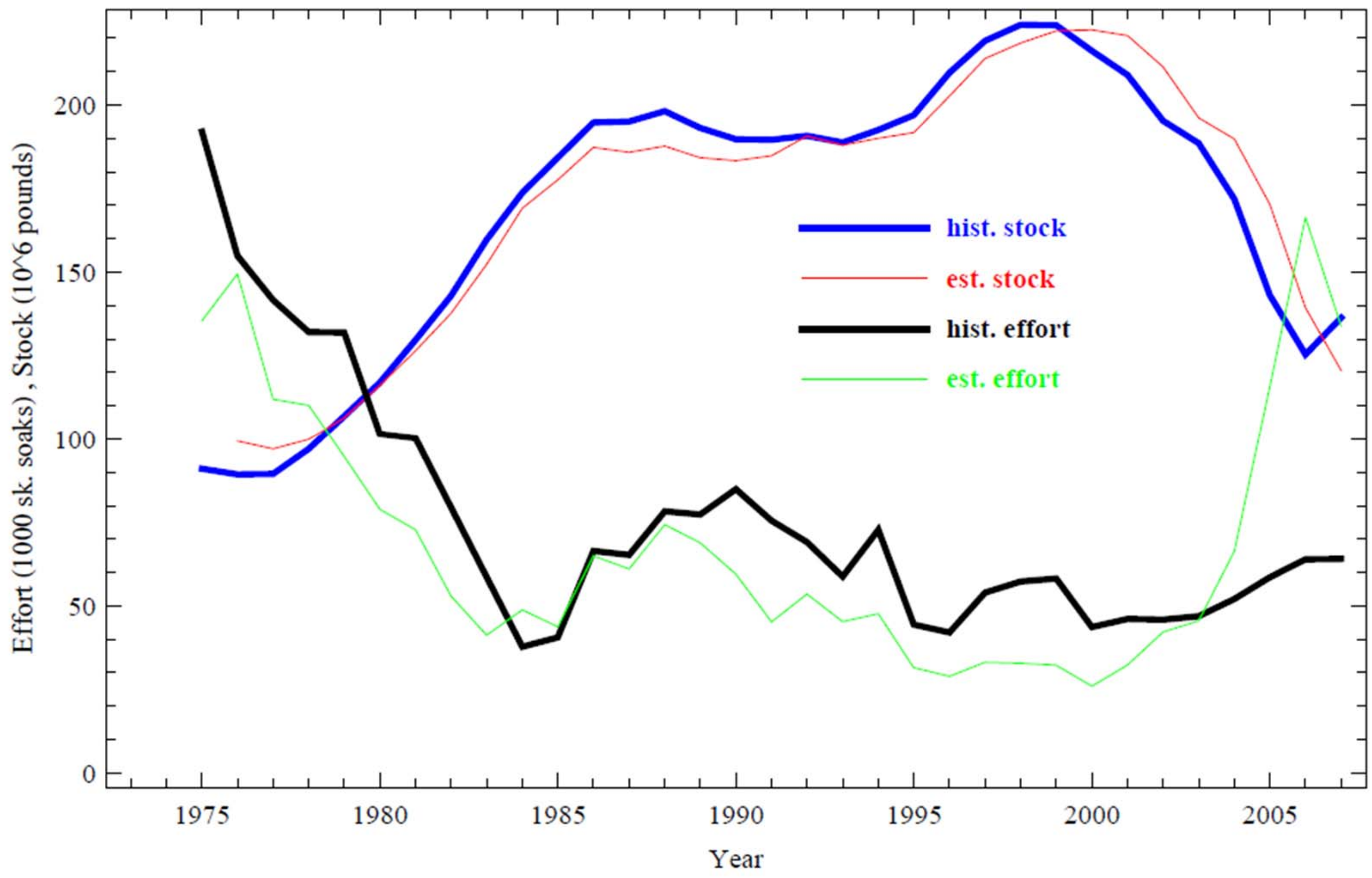
Mortality

"Shock" from nature

Capacity limitation

Population Dynamics

- Fit to historical data ($w = 1$)



Ermon et al., 2010

Robust Optimization

- Traditional MDP approach
  - Maximize expected total discounted reward

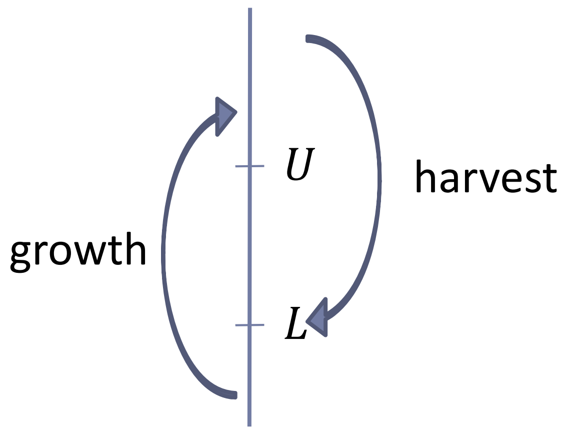
- Their approach: “Game against Nature”
  - Nature chooses w adversarially
  - Maximize *worst-case* total discounted reward

- Advantages:
  - Avoid catastrophic outcomes such as collapse of fishery
  - Don’t need fine-grained model for $p(w)$
    - Only specify allowable range of w

Ermon et al., 2010

Main Result

- Analytical characterization of optimal policy
  - For a general class of growth and economic models, optimal policy is of S-s type



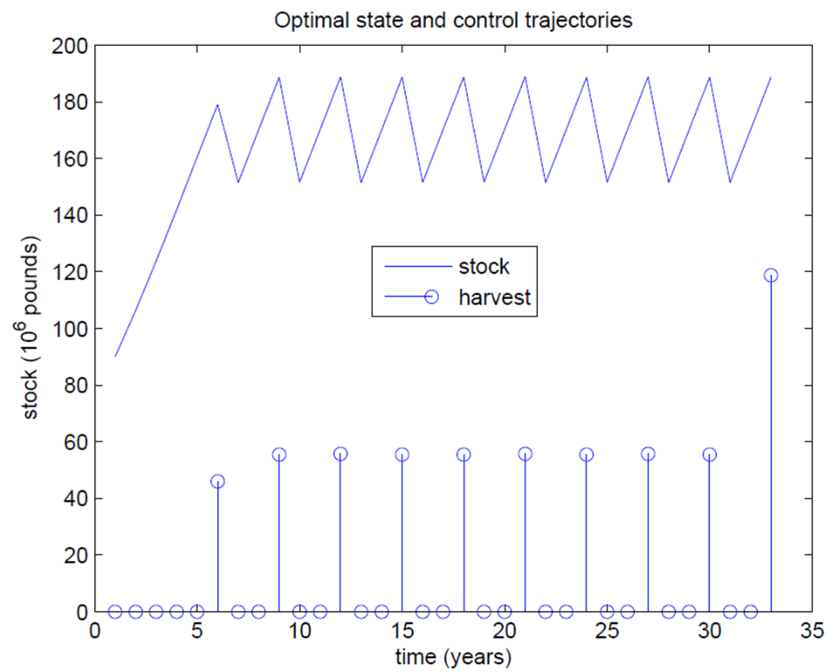
Stock $> U$ → Harvest down to L

Stock $\leq U$ → Let grow until U

- Proof based on mathematical notion of *K*-concavity [Scarf 1960]
  - From inventory control problems in econmics and operations research

Pacific Halibut Results

- Reanalysis of 1975-2007 data
  - Fitted growth model
  - *Worst-case* environmental inputs
- Optimal policy involves *periodic closures* of fishery
  - Maintain supply by rotating closures
- More revenue than baselines
  - Historical revenue
  - Current IPHC policy



Policy	**Disc. revenue (\$)**	**Loss (\$)**
Optimal $S-s$	9.05141×10^8	–
Historical rates	7.06866×10^8	1.98275×10^8
Average CPP	6.51849×10^8	2.53292×10^8

Ermon et al., 2010

Important Themes for Environmental Policy

- Synergy between economic reward and ecosystem stability
  - Why no over-exploitation?
    - Protect future value of fishery
    - Cost to harvest scarce stock
  - Cautionary notes
    - Barriers to over-exploitation are not intrinsic
      - High discount rate → prioritize present reward too much
      - Technology improvements → cheaper to harvest
    - Models often wrong or missing important side-effects

- Robust optimization
  - Prevents catastrophic outcomes (within modeling framework)
  - Is worst-case too severe?
    - Extension to broader class of risk-sensitive objectives [Ermon et al. IJCAI, 2011]

Ermon et al., 2010

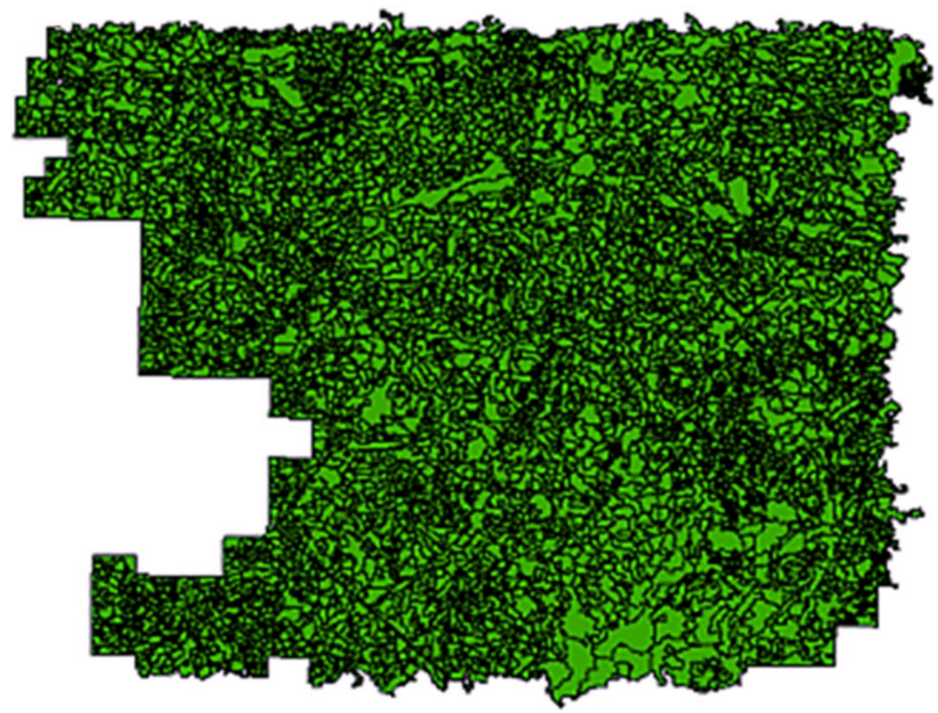
Managing Wildfire in Eastern Oregon

- Natural state (hypothesized):
  - Large Ponderosa Pine trees with open understory
  - Frequent "ground fires" that remove understory plants (grasses, shrubs) but do not damage trees
- Fires have been suppressed since 1920s
  - Large stands of Lodgepole Pine
  - Heavy accumulation of fuels in understory
  - Large catastrophic fires that kill all trees and damage soils
  - Huge firefighting costs and lives lost

Formulation as an MDP

- Divide landscape into 4000 Management Units
- 10-Year time step
- State of each MU:
  - Age of trees
    - {0-9, 10-19, 20-29, 30-39, 40-49}
  - Amount of fuel
    - {none, low, medium, high, very high}
  - 25 possible combinations
  - 25^{4000} possible states for the landscape

- Actions in each MU each decade
  - Do nothing
  - Fuel treatment (costs money)
  - Harvest trees (makes money, but increases fuel)
  - Harvest + Fuel
  - 4^{4000} possible actions over landscape



Study area in Deschutes National Forest

Game Against "Neutral" Nature

- For each time step t
  - Our turn:
    - Observe current state s_t (i.e., state of all MUs)
    - Choose action vector a_t
    - Execute the actions in the MUs
  - Nature's turn:
    - Stochastically ignite and burn fires on the landscape (Implemented by ignition model + fire spread model)
    - Grow trees and fuel (Implemented by forest growth model)

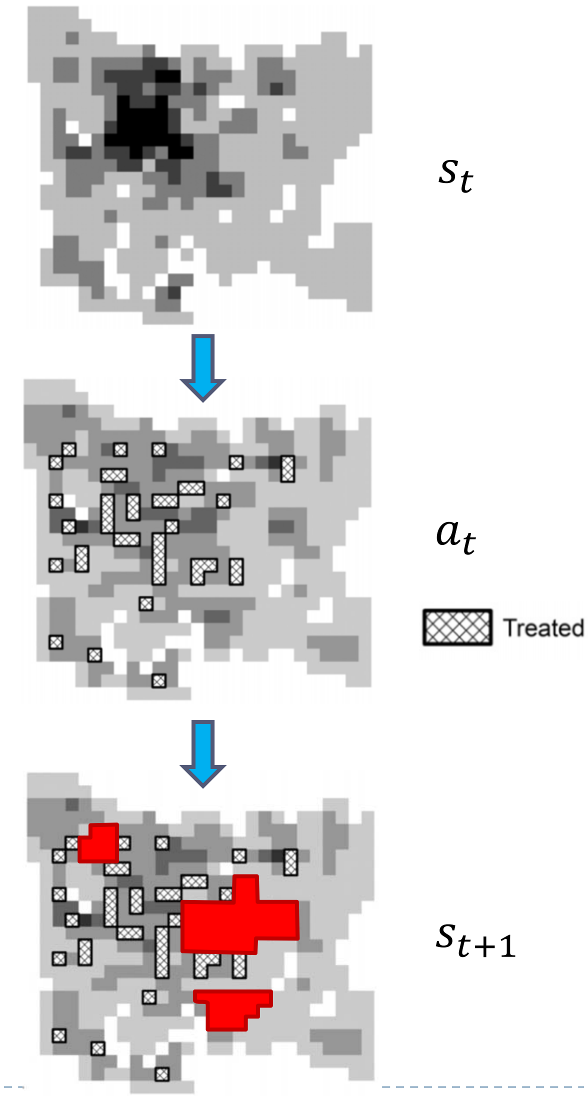


Image: Wei et al, 2008

Open Problem: Solving This MDP

- One-shot Method [Wei, et al., 2008]
  - Run 1000s of simulated fires to generate fire risk map and fire propagation graph
  - Formulate and solve Mixed Integer Program to compute optimal one-shot solution
- Challenge:
  - Develop methods that can solve the MDP over long time horizons

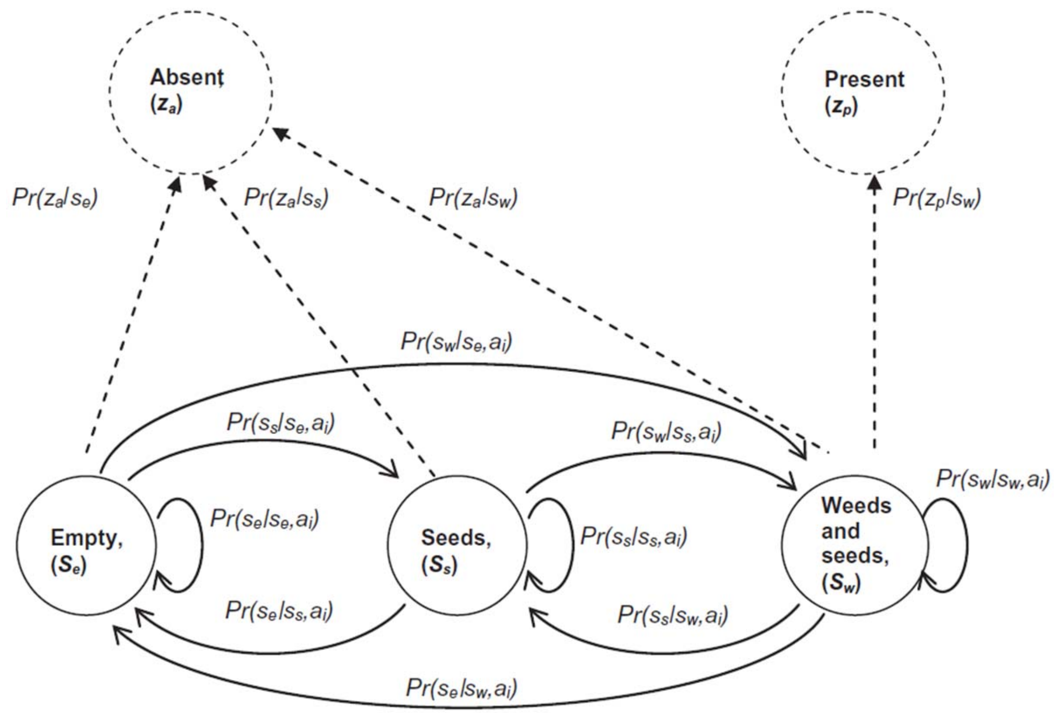
Optimal Management of Difficult-to-Observe Invasive Species [Regan et al., 2011]

- Branched Broomrape (*Orobanche ramosa*)
  - Annual parasitic plant
  - Attaches to root system of host plant
  - Results in 75-90% reduction in host biomass
  - Each plant makes ~50,000 seeds
  - Viable for 12 years
- Quarantine Area in S. Australia
  - 375 farms; 70km x 70km area

Regan, Chades & Possingham, 2011

Formulation as a POMDP: Single Farm

- States:
  - {Empty, Seeds, Plants & Seeds}
- Actions:
  - {Nothing, Host Denial, Fumigation}
- Observations:
  - {Absent, Present}
  - Detection probability d
- Rewards:
  - Cost(Nothing) < Cost(Host Denial) ≪ Cost(Fumigation)
- Objective:
  - 20-year discounted reward (discount = 0.96)



State Diagram

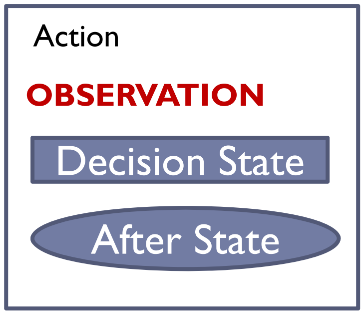
Regan, Chades & Possingham, 2011

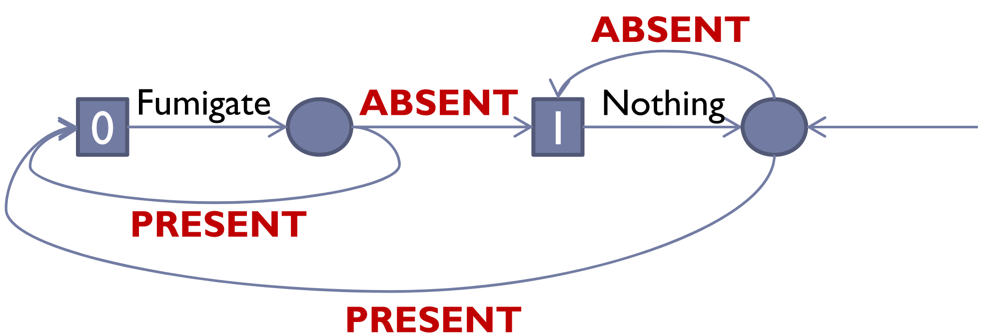
Optimal MDP Policy

- If plant is detected, Fumigate; Else Do Nothing
  - Assumes perfect detection

www.grdc.com.au

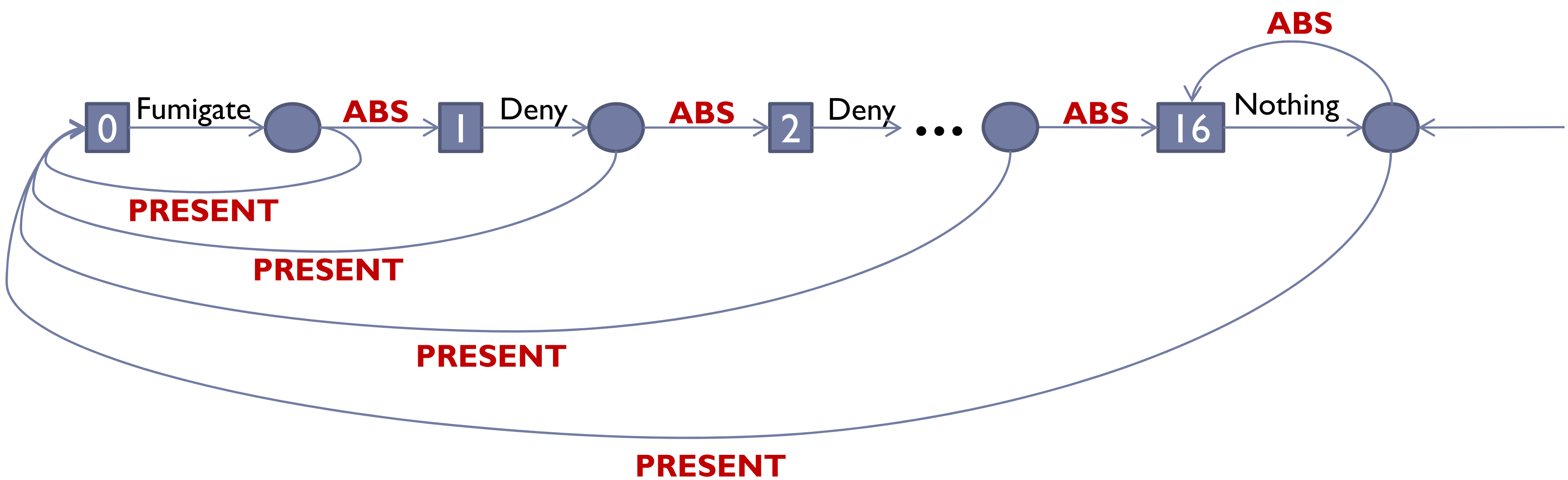
Optimal Policy for $d \geq 0.5$





- Same as the Optimal MDP Policy

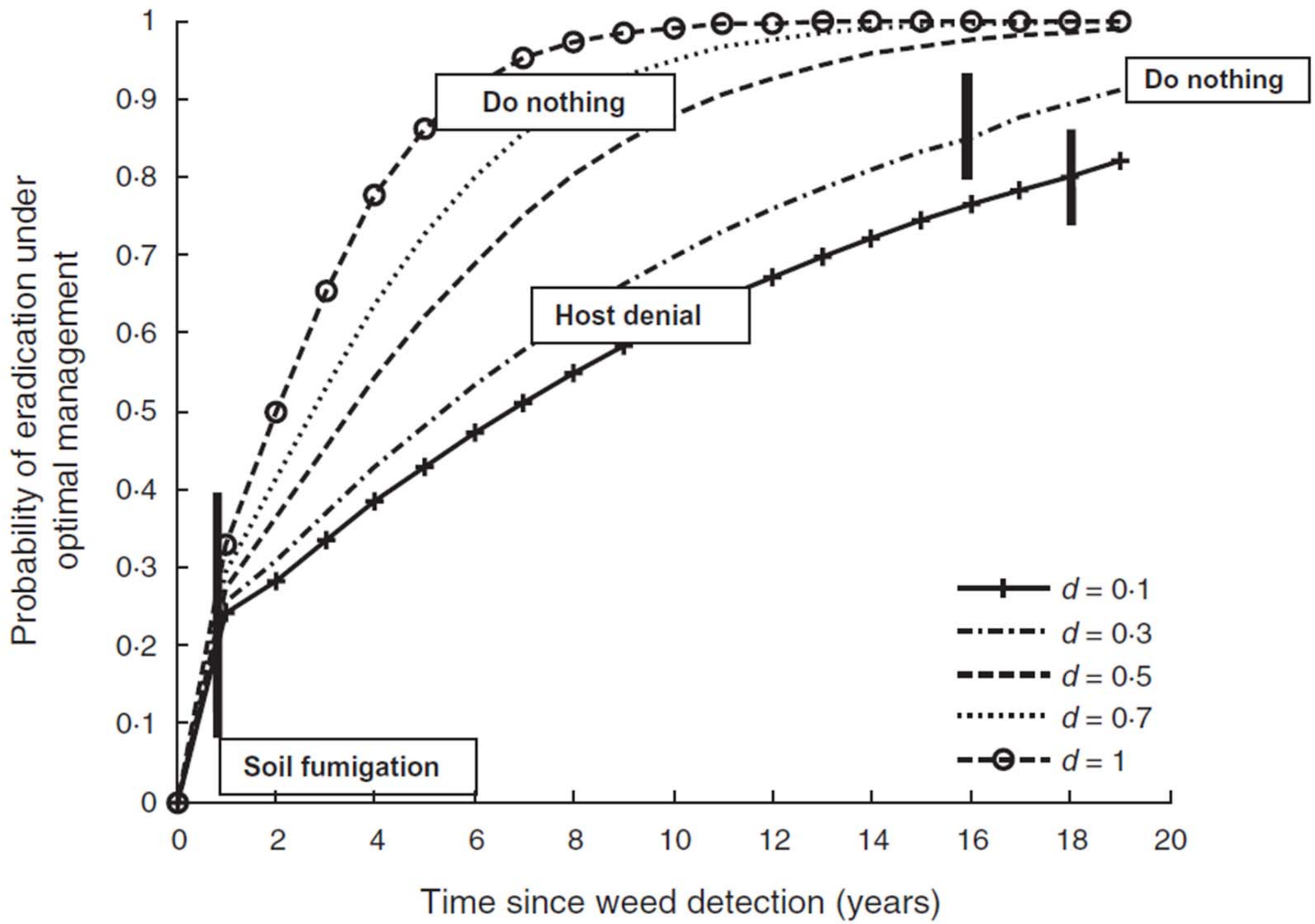
Optimal Policy for $d = 0.3$



- Deny Host for 15 years before switching to Nothing
- For $d = 0.1$, Deny Host for 17 years before switching to Nothing

Regan, Chades & Possingham, 2011

Probability of Eradication



Regan, Chades & Possingham, 2011

Discussion

- POMDP is exactly solvable because the state space is very small
- Real problem is a spatial meta-population at two scales
  - Within a single farm
  - Among the 375 farms in the quarantine area
    - 3^{375} states
  - Exact solution of large POMDPs is beyond the state of the art

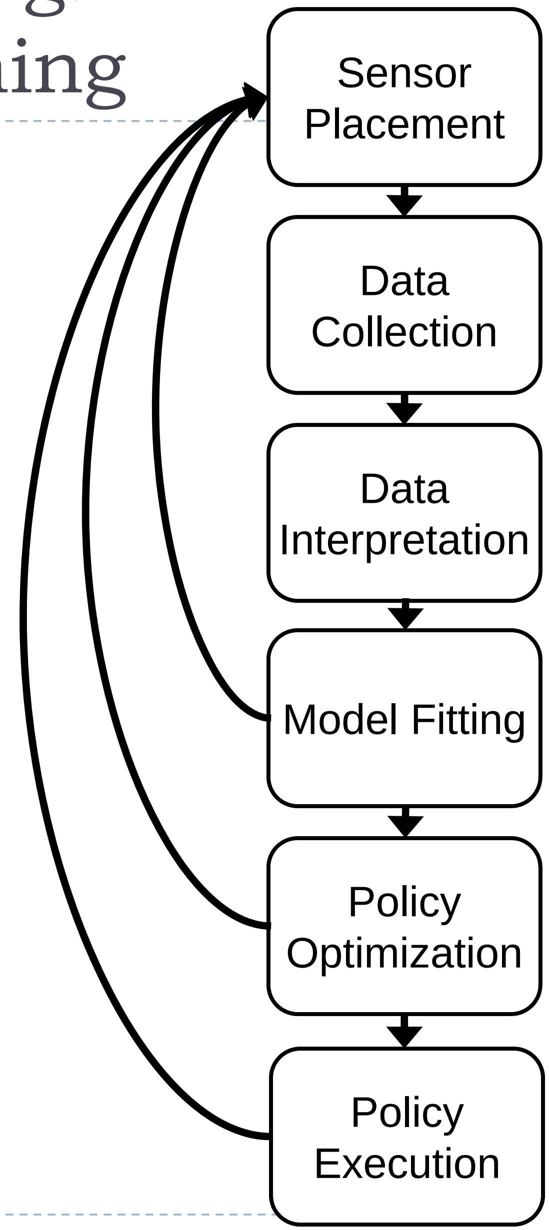
Regan, Chades & Possingham, 2011

Outline

- Data Acquisition
  - Sensors: Physical sensors, human observers, repurposing data from other sources
  - Data interpretation: Extracting signals from data
- Ecological Models
  - Species Distribution Models
  - Dynamical Models: Dispersal, Migration, Invasion, Climate Change
- Policy Optimization
  - Conservation: Reserve design, Network design
  - Invasive species: Eradication, restoration, monitoring
  - Fisheries: Managing harvest levels

Challenges for Machine Learning: Sensor Placement/Active Learning

- We have...
  - Algorithms for one real-valued quantity
    - assuming stationary correlations, perfect observations
- We need...
  - Algorithms for multiple quantities
    - real-valued: nutrients, temperature, precipitation
    - counts: species abundance for multiple species
    - discrete: species presence/absence for multiple species
  - Algorithms that consider dynamics, detectability, patchiness (meta-populations)



Challenges for Data Interpretation

- We have…
  - Algorithms for individual modalities at single scales
    - object recognition
    - bioacoustics
    - RFID tags
- We need…
  - Methods for integrating sensor modalities at vastly different scales in space and time
    - data integration at multiple scales
    - joint interpretation (sensor fusion) of multiple sensors to improve accuracy of data interpretation
  - Better tools for data management, feature definition, visualization, synthetic data generation (for debugging and testing)

Challenges for Learning Algorithms

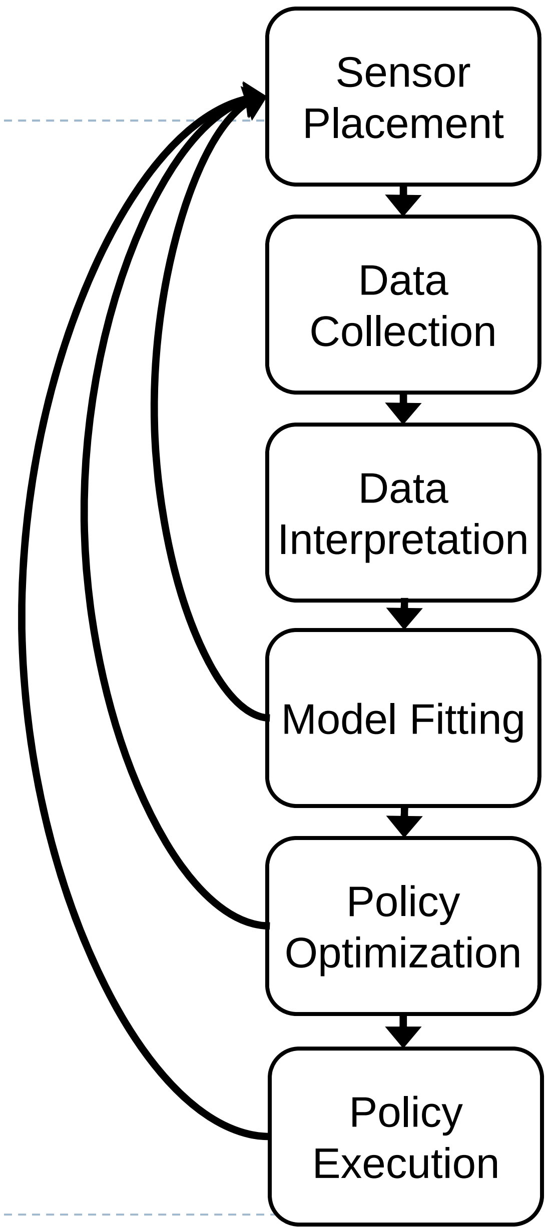
- We have…
  - Species Distribution Models for single species with partial detectability
    - stationary, non-spatial
- We need…
  - Species Distribution Models for thousands of species
    - model competition, predation, dispersal
    - explicitly spatial
  - Meta-Population Models for multiple species
  - Models that link abiotic quantities (nutrients, temperature, precipitation) and biotic quantities (species, populations)

Challenges for Optimization

- We have…
  - One-shot algorithms for meta-populations and fires
  - Exact algorithms for modest-sized MDPs
  - Exact algorithms for tiny POMDPs
  - Algorithms that optimize a scalar reward in expectation
- We need…
  - Algorithms for MDP/POMDP planning for meta-populations and spatial processes
  - That consider multiple criteria
  - That are robust to misspecified dynamics and rewards

Data → Models → Policies: Overall Challenges

- It isn't a pipeline
  - We need algorithms that integrate/couple all parts of the process
  - Learning algorithms should be integrated with policy optimization
  - Sensor placement should be sensitive to all goals



Closing

- Links to data, software, and papers available in the electronic version of these slides
- Thank-you's
  - BugID team, especially Wei Zhang, Natalia Larios, Junyuan Lin, Gonzalo Martinez
  - David Winkler
  - Jane Elith and Steven Phillips
  - Lab of Ornithology collaborators: Daniel Fink, Steve Kelling and the thousands of eBirders
  - National Science Foundation support under grants 0705765, 0832804, and 0905885

Questions?

Data Resources

- Species Distribution Models
  - eBird Reference Dataset 2.0
    - http://www.avianknowledge.net/content/features/archive/ebird-reference-dataset-2-0-released
    - eBird checklist data along with an excellent set of covariates
    - set of suggested analysis problems
- Fine-Grained Image Classification
  - Oregon State STONEFLY9 dataset
    - http://web.engr.oregonstate.edu/~tgd/bugid/stonefly9/
    - EPT 54 coming soon...
  - Caltech/UCSD CUB-200 bird dataset
    - http://www.vision.caltech.edu/visipedia/CUB-200.html
  - Oxford Flower dataset (102 classes)
    - http://www.robots.ox.ac.uk/~vgg/data/flowers/102/index.html

Model Resources

- Meta-Population Models
  - SPOMSIM
    - http://www.helsinki.fi/bioscience/consplan/software/SPOMSIM.html
  - Synthetic Red-Cockaded Woodpecker instances
    - http://www.cs.cornell.edu/~kiyan/rcw/generator.htm

Machine Learning Algorithms

- Phillips' Maxent Package
  - http://www.cs.princeton.edu/~schapire/maxent/

References

Crowley M, Nelson J, Poole D. **Seeing the Forest Despite the Trees: Large Scale Spatial-Temporal Decision Making**. In: *AAAI 2011*.; 2011.

Doak DF, Estes JA, Halpern BS, et al. **Understanding and predicting ecological dynamics: are major surprises inevitable?** *Ecology*. 2008; 89(4):952-961.

Elith J, Graham CH, Anderson RP, et al. **Novel methods improve prediction of species' distributions from occurrence data.** *Ecography*. 2006;29(2):129-151.

Elith J, Phillips SJ, Hastie T, et al. **A statistical explanation of MaxEnt for ecologists.** *Diversity and Distributions*. 2011;17:43-57.

Ermon, S., Conrad, J., Gomes, C., & Selman, B. (2010). **Playing games against nature: optimal policies for renewable resource allocation**. *Proc. of The 26th Conference on Uncertainty in Artificial Intelligence.*

Ermon, S., Conrad, J., Gomes, C., & Selman, B. (2011). **Risk-sensitive Policies for Sustainable Renewable Resource Allocation**. *To appear in IJCAI 2011*.

Fink D, Hochachka WM, Zuckerberg B, et al. **Spatiotemporal exploratory models for broad-scale survey data.** *Ecological applications : a publication of the Ecological Society of America*. 2010;20(8):2131-47

Friedman J. **Greedy function approximation: a gradient boosting machine.** *Annals of Statistics*. 2001;29(5):1189-1232.

References (2)

Gomez-Rodriguez, M., Balduzzi, D., & Schölkopf, B. (2011). **Uncovering the Temporal Dynamics of Diffusion Networks.** *ICML 2011*.

Hanski, I. (1994**). A practical model of metapopulation dynamics**. *Journal of Animal Ecology*, *63*(1), 151–162.

Hutchinson RA, Liu L-P, Dietterich TG. **Incorporating Boosted Regression Trees into Ecological Latent Variable Models.** In: *Proceedings of the Twenty-fifth Conference on Artificial Intelligence*.; 2011.

Kempe, D., Kleinberg, J., & Tardos, É. (2003). **Maximizing the spread of influence through a social network**. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (p. 137–146).

Krause A, Singh A, Guestrin C**. Near-Optimal Sensor Placements in Gaussian Processes: Theory , Efficient Algorithms and Empirical Studies**. *Journal of Machine Learning Research*. 2008;9:235-284.

Leathwick JR, Moilanen A, Francis M, et al. **Novel methods for the design and evaluation of marine protected areas in offshore waters.** *Conservation Letters*. 2008;1:91-102.

Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., & Glance, N. (2007). **Cost-effective outbreak detection in networks**. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (p. 420–429).

MacKenzie DI, Nichols JD, Royle JA, et al. ***Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence.*** Elsevier, San Diego, USA; 2006.

References (3)

Martinez G, Zhang W, Payet N, et al. **Dictionary-Free Categorization of Very Similar Objects via Stacked Evidence Trees.** In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2009)*. IEEE; 2009: 1-8.

Moilanen, A. (1999). **Patch occupancy models of metapopulation dynamics: efficient parameter estimation using implicit statistical inference**. *Ecology*, *80*(3), 1031–1043.

Myers, S. A., & Leskovec, J. (2010). **On the convexity of latent social network inference**. *NIPS 2010*.

Phillips SJ, Anderson R, Schapire R. **Maximum entropy modeling of species geographic distributions.** *Ecological Modelling*. 2006;190(3-4):231-259.

Phillips SJ, Dudik M, Schapire RE. **A maximum entropy approach to species distribution modeling.** *Twenty-first international conference on Machine learning - ICML '04*. 2004:83.

Regan TJ, Chadès I, Possingham HP. **Optimally managing under imperfect detection: a method for plant invasions**. *Journal of Applied Ecology*. 2011;48(1):76-85.

Scarf, H. (1960). **The Optimality of (S,s) policies in the dynamic inventory problem.** *Stanford mathematical studies in the social sciences, p. 196.*

Sheldon, D., Elmohamed, M. A. S., & Kozen, D. (2008). **Collective inference on Markov models for modeling bird migration**. *Advances in Neural Information Processing Systems 2007*, *20*, 1321–1328.

References (4)

Sheldon, D. (2010). **Manipulation of PageRank and Collective Hidden Markov Models**. *Ph.D. Thesis, Cornell University.*

Sheldon, D., Dilkina, B., Elmachtoub, A., Finseth, R., Sabharwal, A., Conrad, J., Gomes, C., Shmoys, D., Allen, W., Amundsen, O., & Vaughaun, B. (2010.). **Maximizing the Spread of Cascades Using Network Design**. *UAI-2010: 26th Conference on Uncertainty in Artificial Intelligence* (p. 517–526).

Ter Braak, C. J. F., & Etienne, R. S. (2003). **Improved Bayesian analysis of metapopulation data with an application to a tree frog metapopulation**. *Ecology*, *84*(1), 231–241.

Wei Y, Rideout D, Kirsch A. **An optimization model for locating fuel treatments across a landscape to reduce expected fire losses**. *Canadian Journal of Forest Research*. 2008;38(4):868-877.

Winkler, D. W. (2006). **Roosts and migrations of swallows**. *El hornero*, *21*, 85–97.

Yu J, Wong W-K, Hutchinson RA. **Modeling Experts and Novices in Citizen Science data for Species Distribution Modeling.** In: *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM 2010)*.; 2010.

Zhang W, Surve A, Fern X, Dietterich T. **Learning Non-Redundant Codebooks for Classifying Complex Objects.** In: *Proceedings of the International Conference on Machine Learning, ICML-2009*.; 2009:1241-1248.