Computational Sustainability: Applying Advanced Computing to Ecological Science and Ecosystem Management

Tom Dietterich

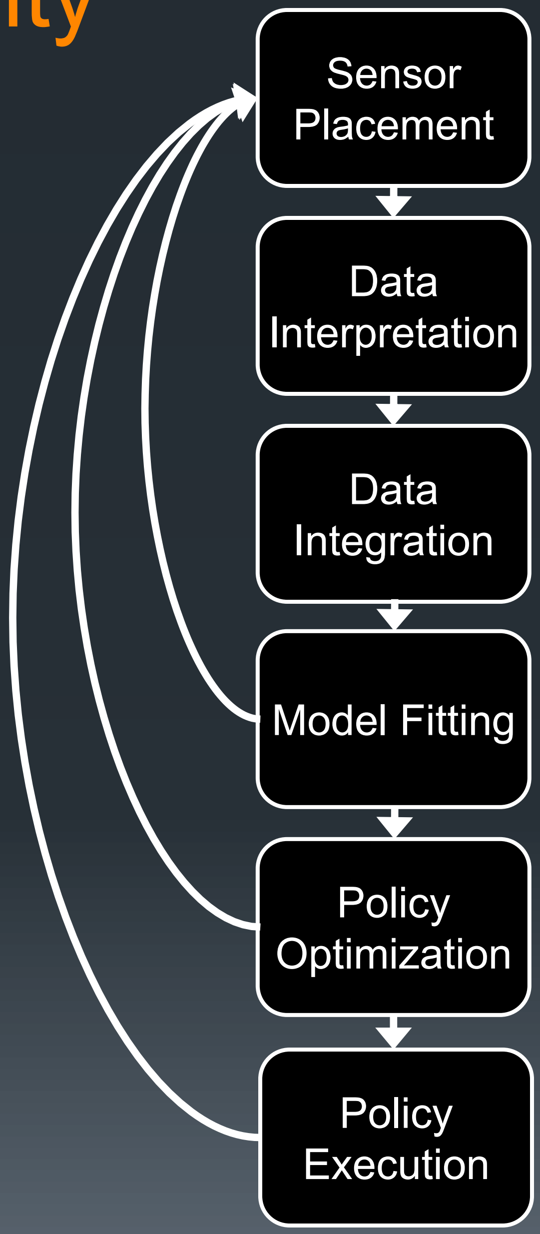
Oregon State University

In collaboration with Ethan Dereszynski, Rebecca Hutchinson, Dan Sheldon, Claire Montgomery and the Cornell Lab of Ornithology

Computational Sustainability

- The study of computational methods that can contribute to the sustainable management of the earth's ecosystems
  - biological
  - social
  - economic

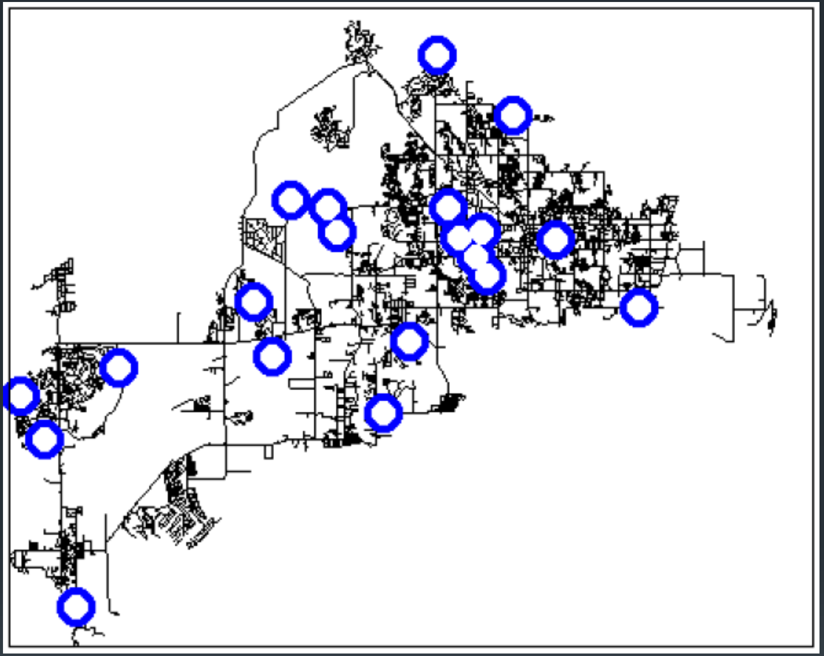
- Data → Models → Policies



Example Research Efforts

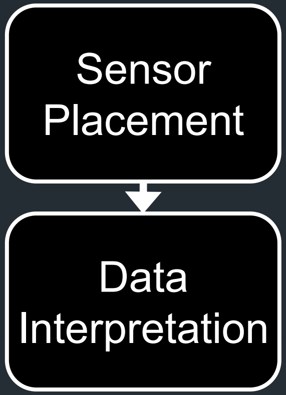
Sensor Placement

- Objectives
  - detection probability
  - improving model accuracy
  - improving causal understanding
  - improving policy effectiveness
- Key Tool: Submodular Functions
  - Formulate the problem in terms of a submodular objective
  - Greedy algorithm then works well and has provable performance



Leskovec et al, KDD2007

Data Interpretation



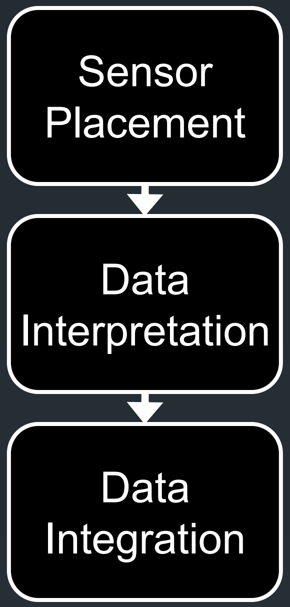
- Insect identification for population counting
- Raw data: image
- Interpreted data: Count by species
- Technology: Object recognition

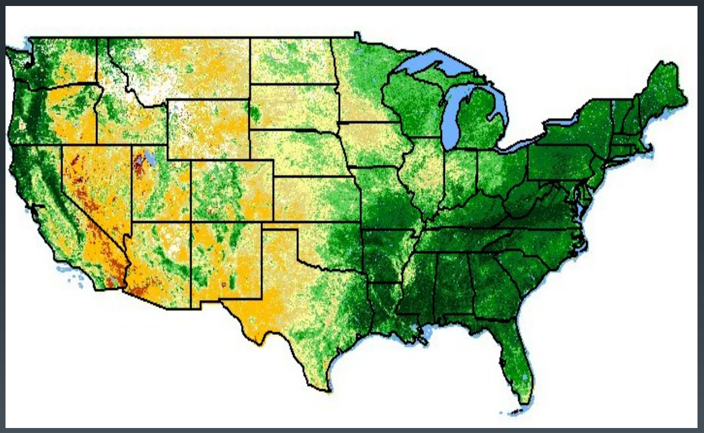
Species	Count
Nilaparvata lugens	12
Sogatella furcifera	8
Laodelphax striatellus	0
Cnaphalocrocis medinalis	0
Chilo suppressalis	45
Sesamia inferens	18

image: Qing Yao

Data Integration

- Integrating heterogeneous data sources to predict when migrating birds will arrive:
  - Landsat (30m; monthly)
    - land cover type
  - MODIS (500m; daily/weekly)
    - land cover type
    - “greening” index
  - Census (every 10 years)
    - human population density
    - housing density and occupation
  - Interpolated weather data (15 mins)
    - rain, snow, solar radiation, wind speed & direction, humidity
  - Integrated weather data (daily)
    - warming degree days
  - Digital elevation model (rarely changes)
    - elevation, slope, aspect

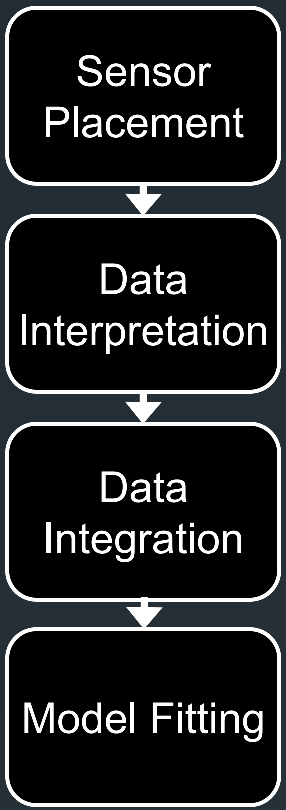




Landsat NDVI:
http://ivm.cr.usgs.gov/viewer/

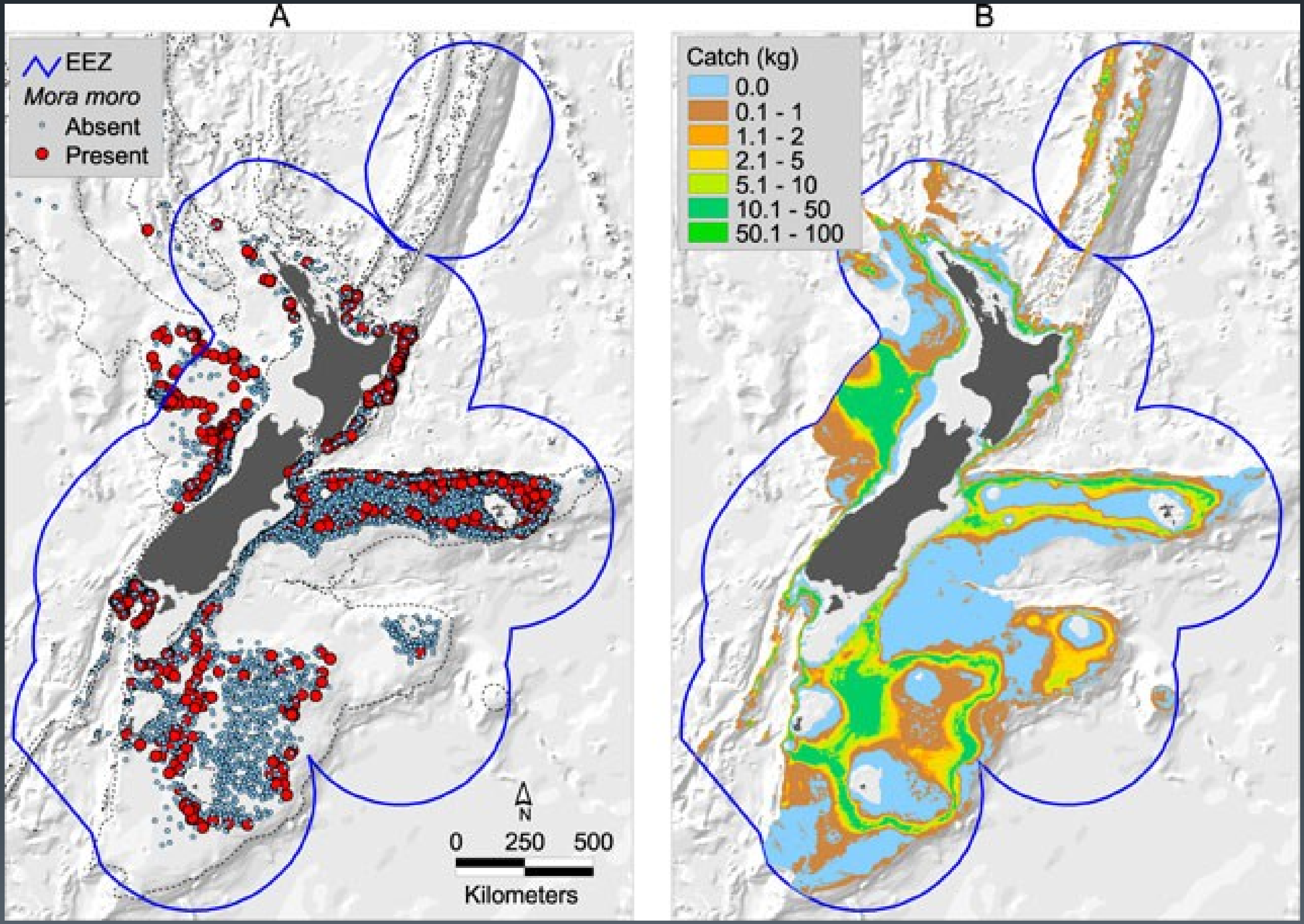
Model Fitting

- Species Distribution Models
  - create a map of the distribution of a species
- Meta-Population Models
  - model a set of patches with local extinction and colonization
- Migration and Dispersal Models
  - model the trajectory and timing of movement



Species Distribution Modeling

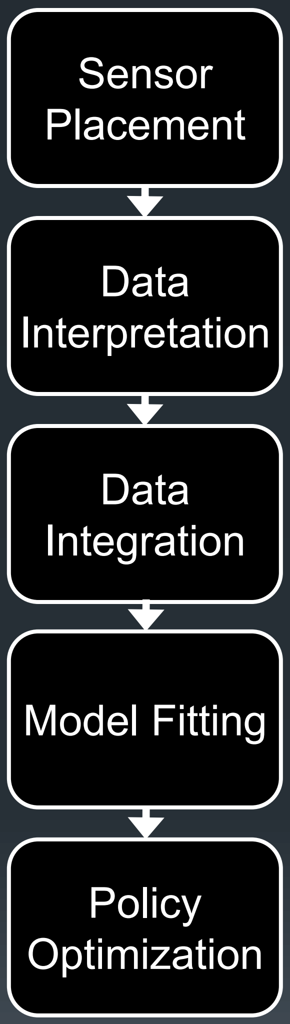
Observations | Fitted Model

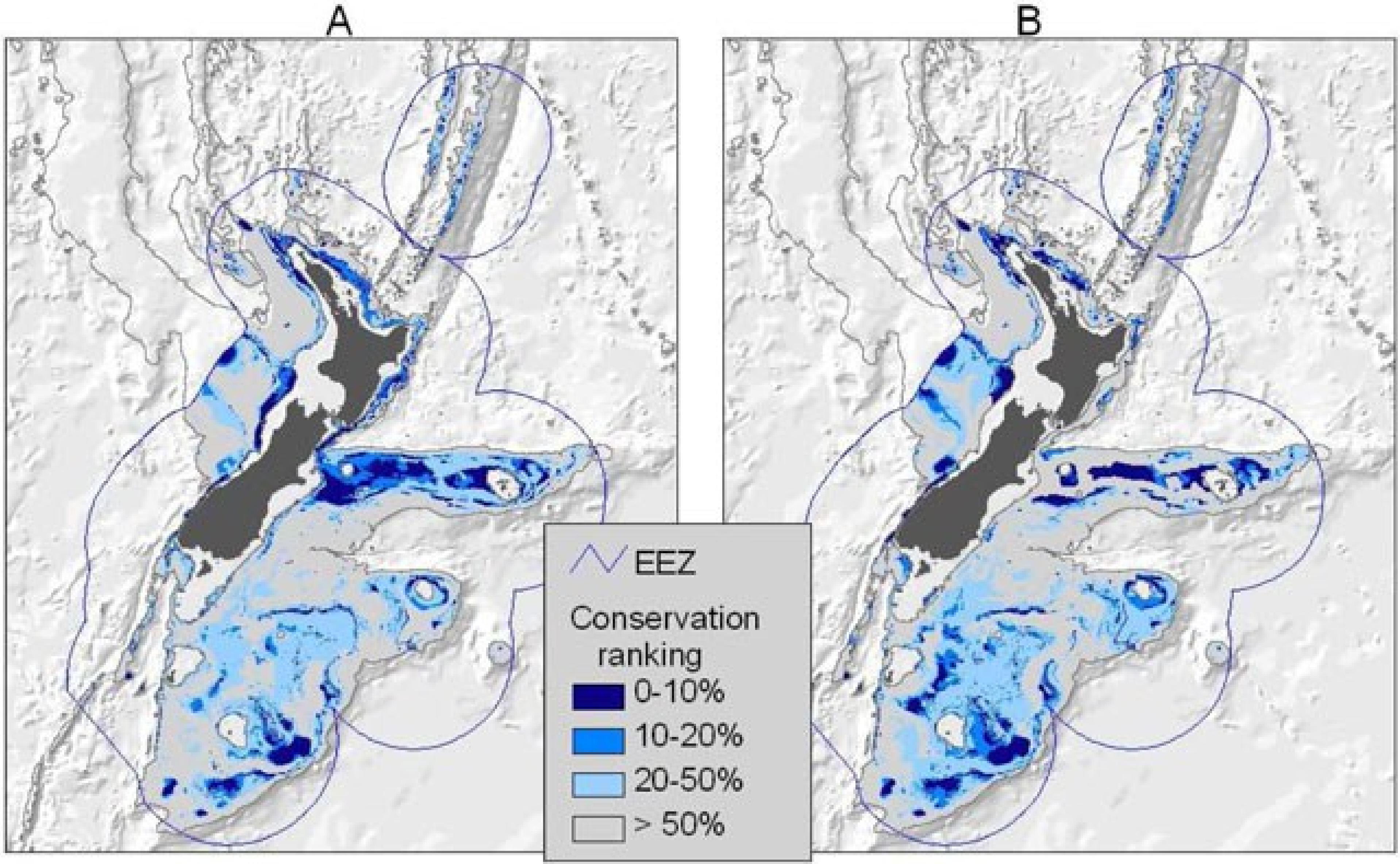


Leathwick et al, 2008

Policy Optimization

- Design of Marine Reserves
  - Fishing is excluded
  - Fish species population is able to recover and remain healthy
- Requires multiple models
  - Fish population behavior and dispersal
  - Fishing effort
  - Fishing effectiveness





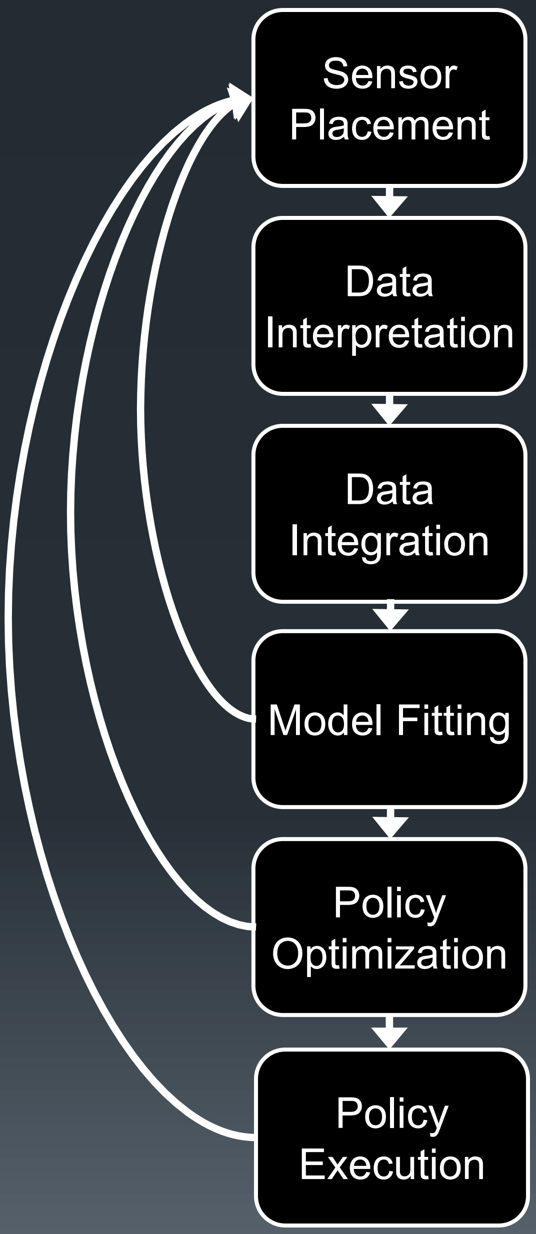
Disregarding costs to fishing industry

Full consideration of costs to fishing industry

Leathwick et al, 2008

Policy Execution

- Repeat
  - Observe Current State
  - Choose and Execute Action
- Challenge: We must start taking actions while our models are still very poor.
  - Need to choose actions to learn more about the system in addition to moving the system toward desired outcomes

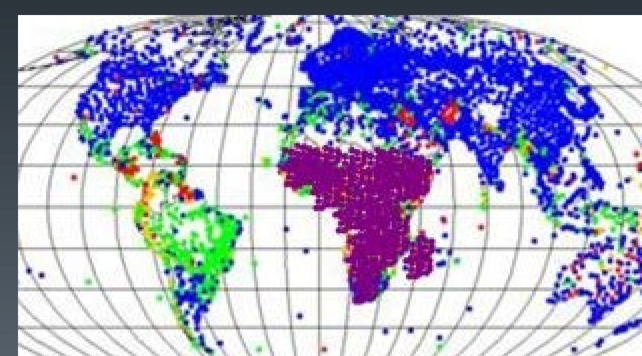


Three Projects at Oregon State

- TAHMO: Continent-Scale Hydro-Meteorological Network
- BirdCast: Continent-Scale Bird Migration Forecasting
- Wildfire Management: Spatial optimization of fire management

Project TAHMO:
20,000 hydro-met stations for Africa

- Africa is very poorly sensed
  - Only a few dozen weather stations reliably report data to WMO (blue points in map)
- Project TAHMO (tahmo.org)
  - TU-DELFT & Oregon State University
  - Design a complete hydrology/meteorology sensor station at a cost of EUR 200
  - Deploy 20,000 such stations across Africa



Why TAHMO?

- Agricultural Production
  - Better water management
  - Improved crop yields
- Crop Insurance Programs
  - Insurers are not willing to write insurance because of lack of data
- Public Health
  - Predict disease outbreaks and respond rapidly

Challenges

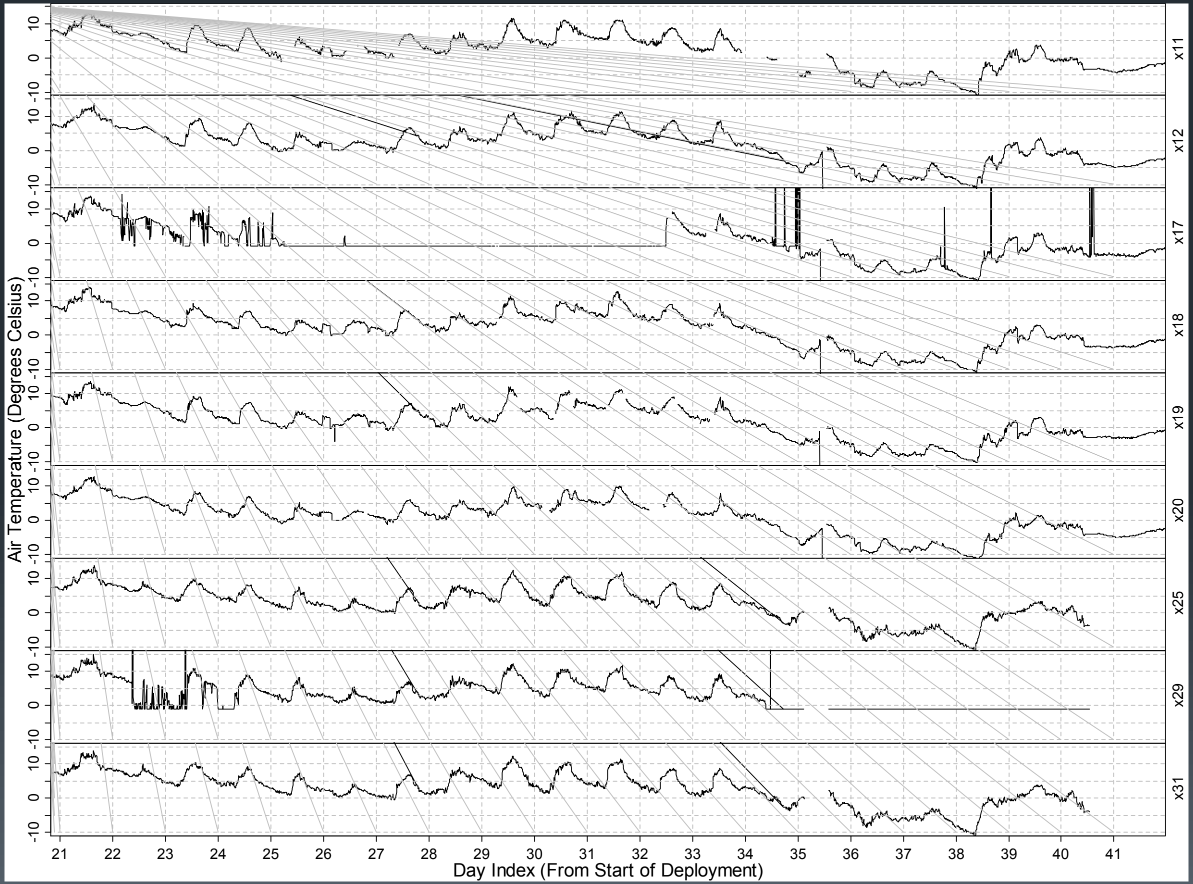
- Sensor Design
  - $200 station with no moving parts

- Sensor Placement
  - Multiple criteria:
    - accuracy of reconstructing maps of
      - temperature, precipitation, solar radiation, wind speed and direction, relative humidity
    - accuracy of estimates of composite variables
      - Evapo-transpiration
    - robustness to sensor failure
    - accessibility and safety

- Continent-scale Data Quality Control
  - Sensors fail for infinitely many reasons
  - Detect failures and impute missing data

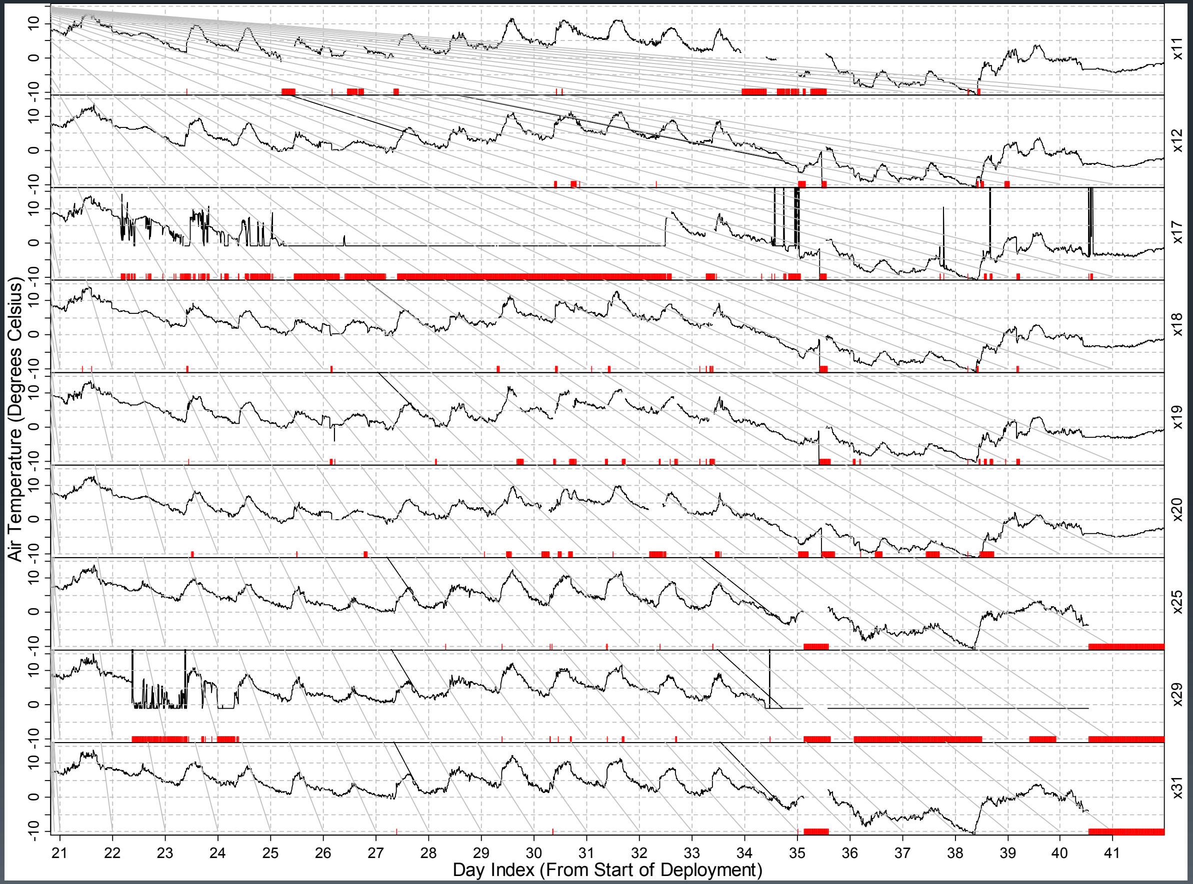
Functions of a Data Cleaning Method

- An ideal method should produce two things given raw data:



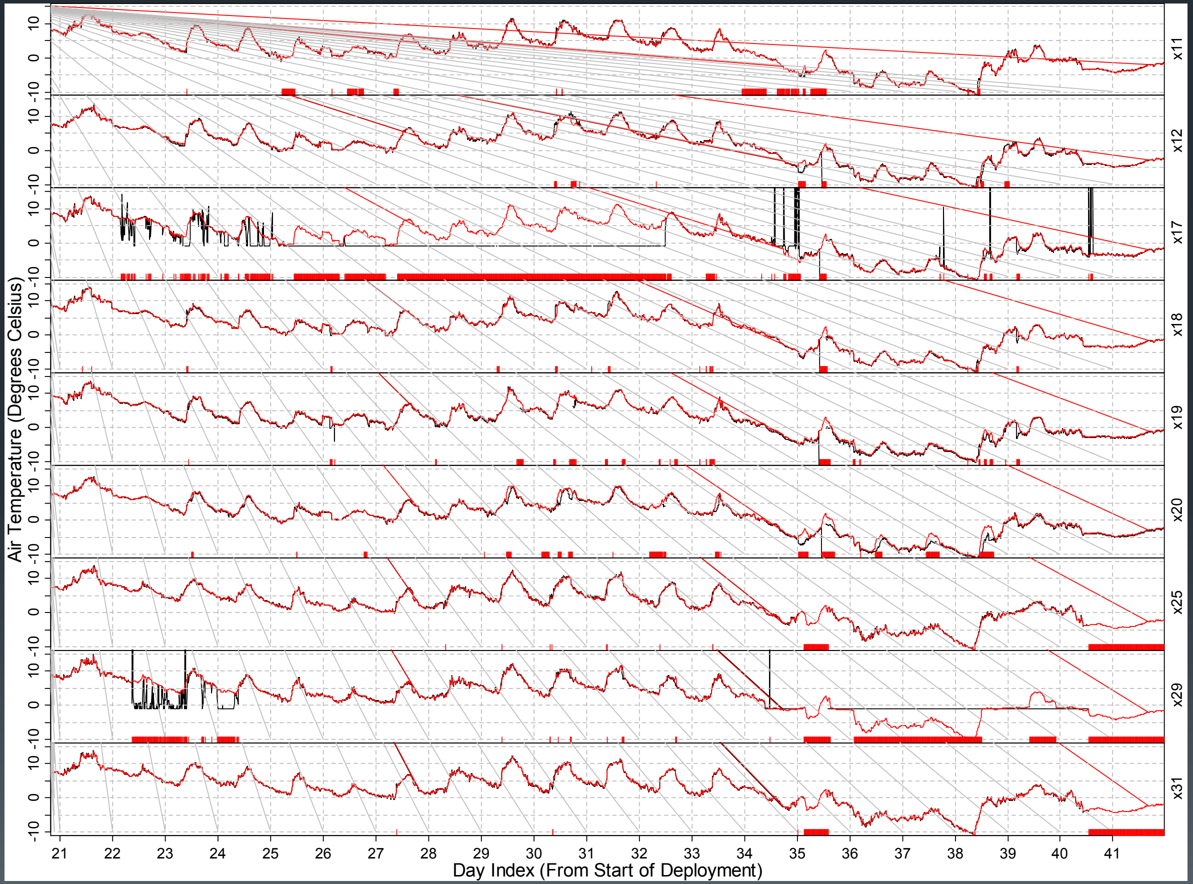
Functions of a Data Cleaning Method

- An ideal method should produce two things given raw data:
  - A label that marks anomalies



Functions of a Data Cleaning Method

- An ideal method should produce two things given raw data:
  - A label that marks anomalies
  - An imputation of the true value (with some confidence measure)



QC State of the Art

- Joint Probabilistic Modeling with Approximate Inference (Dereszynski & Dietterich, 2011)
  - scales as N^3 where $N =$ the number of sensors
  - operates at a single temporal resolution
  - assumes fixed correlation relationships between sensors

- What is Needed
  - methods that scale as N or $N \log N$
  - multi-scale modeling in time and space
  - dynamic correlation structure

Three Projects at Oregon State

- TAHMO: Continent-Scale Hydro-Meteorological Network
- BirdCast: Continent-Scale Bird Migration Forecasting
- Wildfire Management: Spatial optimization of fire management

BirdCast: Continent-Scale Bird Migration Forecasting

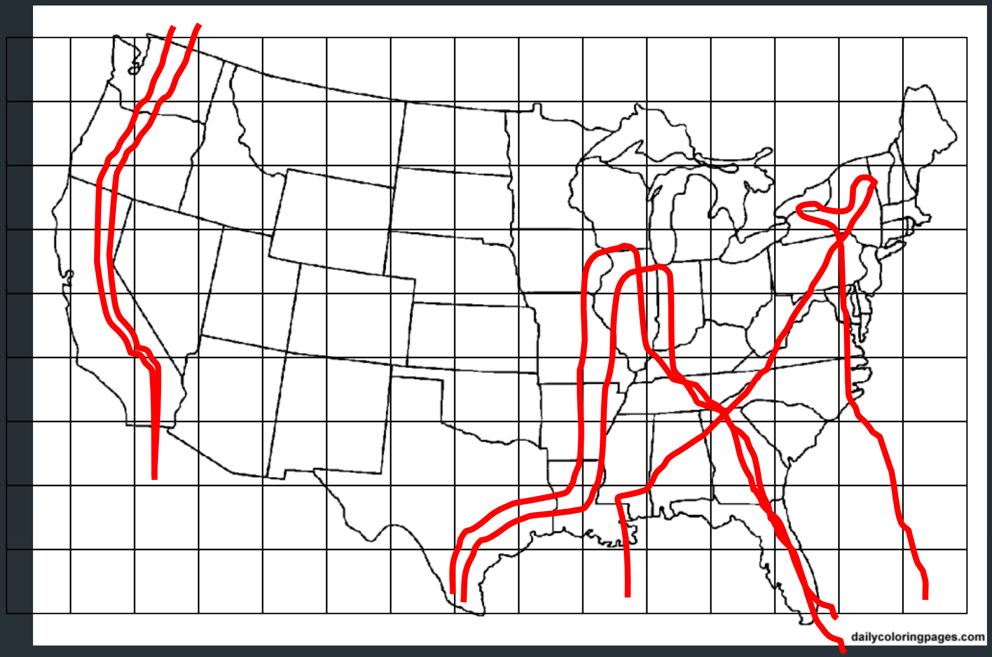
- Why?
  - Science: Bird Migration is poorly understood
  - Conservation: Bird species populations are declining
  - Management:
    - Military training flights occur exactly where birds are migrating...
      - at night
      - at low altitudes
    - Wind farm siting and management
    - Management of skyscraper lighting

OSU Oregon State University

The Cornell Lab of Ornithology
Exploring and Conserving Nature

Data We Wish We Had

- If every bird carried a cell phone...
  - Migration track for every bird
- We could compute
  - n_i^t = # of birds in cell i at time t
  - $n_{ij}^{t,t+1}$ = # of birds moving from cell i to cell j from time t to $t+1$
- We could fit a model to predict probability of moving from one cell to another as a function of
  - weather
  - date
  - location

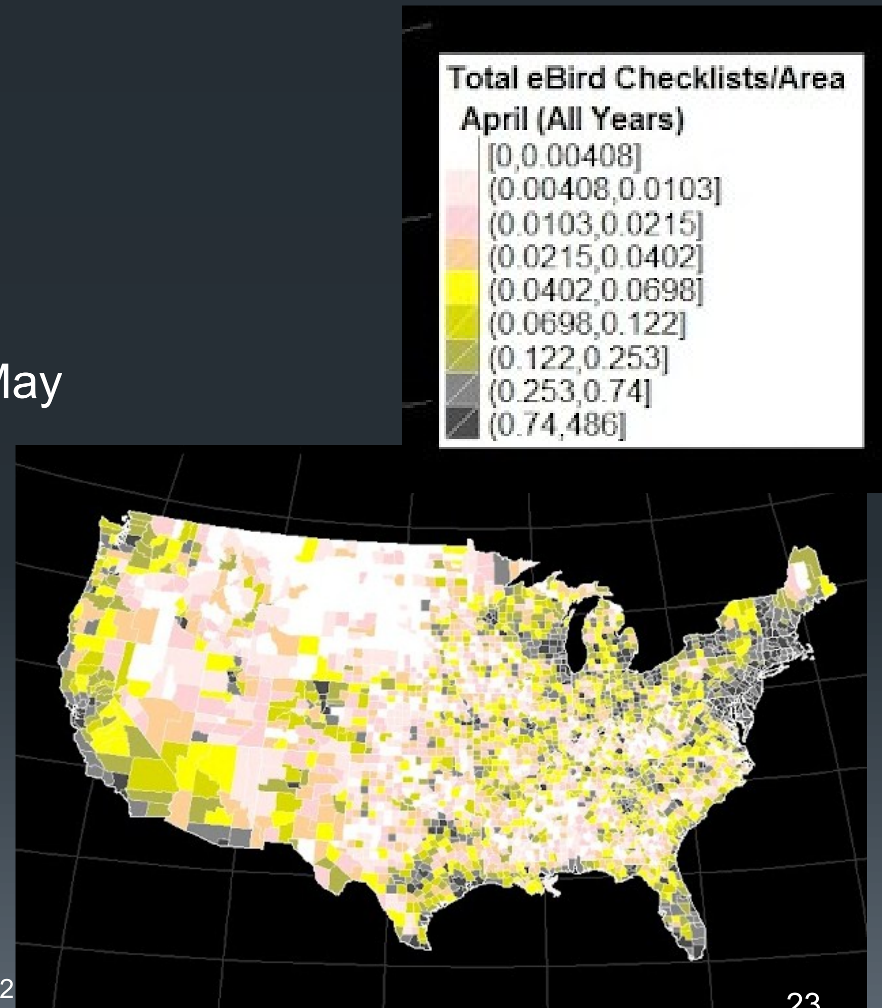


Data We Have

- Birdwatcher count data: eBird.org
- Doppler weather radar
- Night flight calls

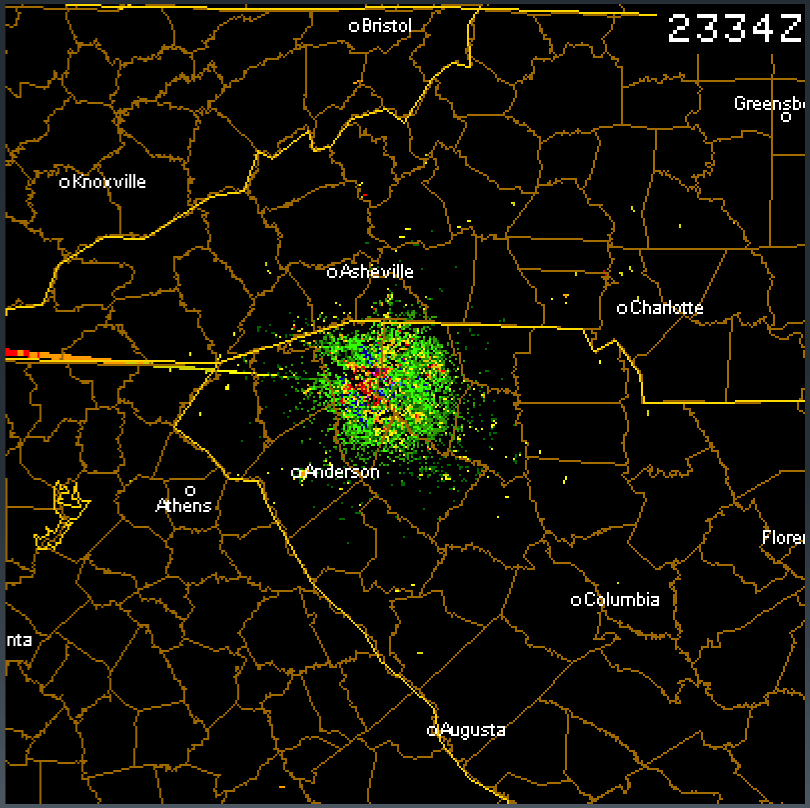
eBird Data

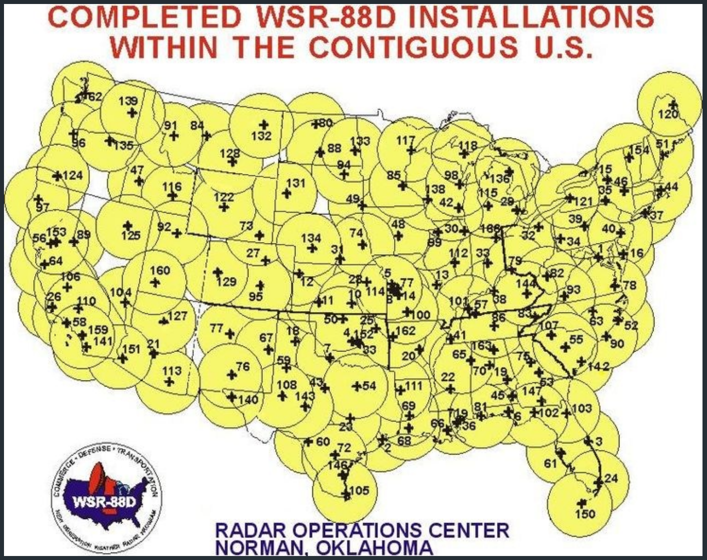
- Bird watchers record their observations in a database through eBird.org.
  - “Citizen Science”
- Dataset available for analysis
- Features
  - LOTS of data!
    - ~3 million observations reported last May
  - All bird species (~3,000)
  - Year-round
  - Continent-scale
- Challenges
  - Variable quality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



Doppler Weather Radar

- Weather radar detects migrating birds

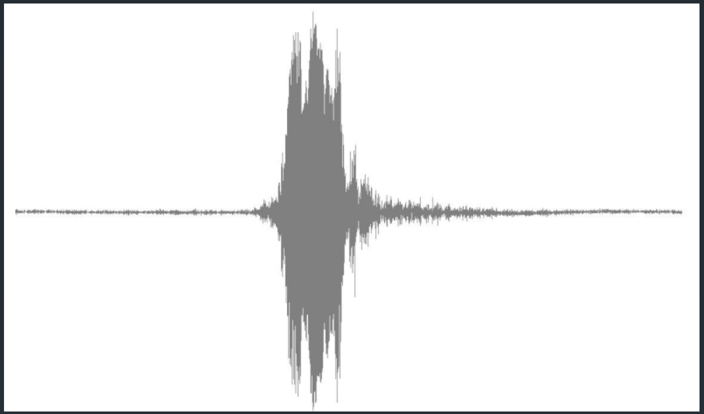




- Can estimate total biomass
- No species information
- Archived data available back to 1995

Night Flight Calls

- Many species of migrating birds emit flight calls that can be identified to species or species group
- New project at Cornell to roll out a large network of recording stations
- Automated detection and classification
- DTW kernel
  - Damoulas, et al, 2010
  - Results on 5 species
  - Clean recordings

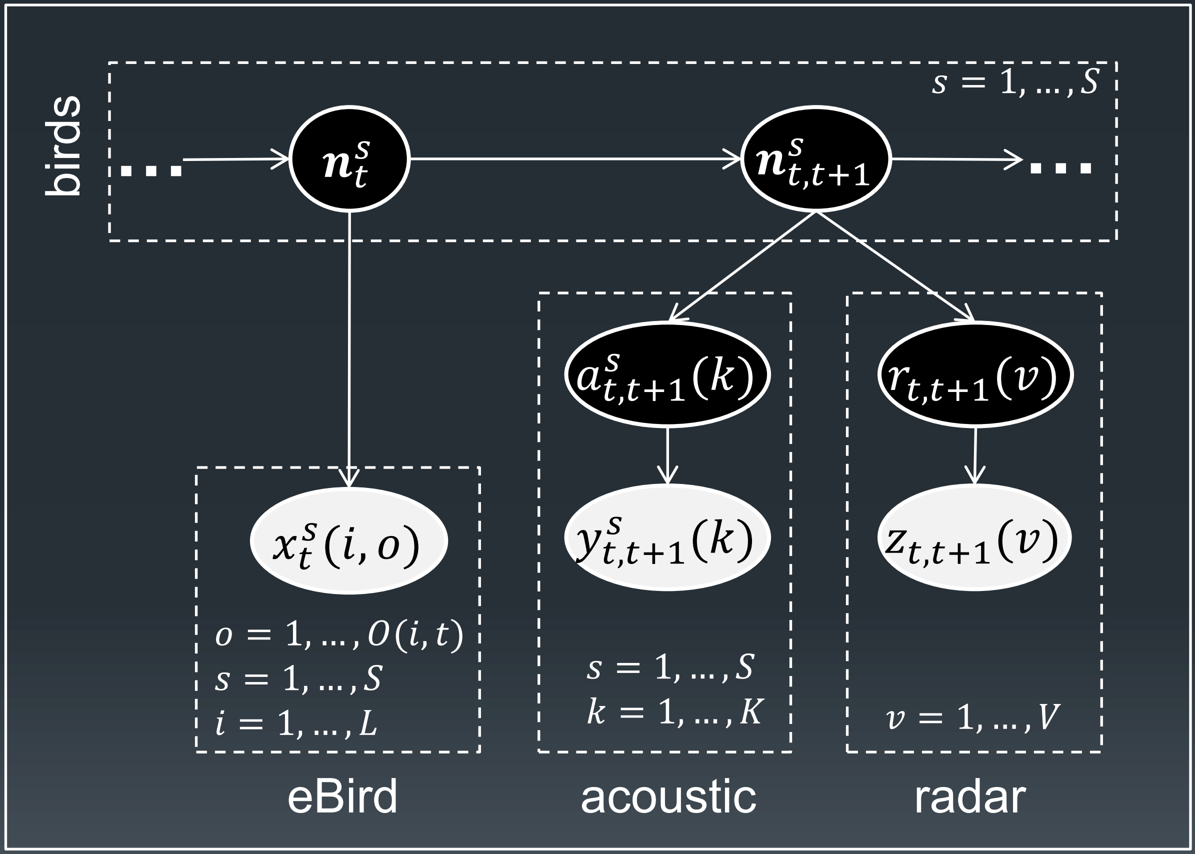


Classifier	*Feature Extraction Method*	10×10CV %
J48	DTW_{global}	87.1 ± 1.14
Kstar	DTW_{global}	$\mathbf{96.6 \pm 0.65}$
BayesNet	DTW_{global}	93.2 ± 0.27
Simple Logistic	DTW_{global}	94.9 ± 0.55
Decision Table	DTW_{global}	72.8 ± 3.82
Random Forest	DTW_{global}	93.2 ± 0.84
Logit Boost	DTW_{global}	91.7 ± 1.64
Rotation Forest	DTW_{global}	94.5 ± 1.06
$SVM^{multiclass}$	DTW_{global} Kernel	95 ± 0.43
VBpMKL	DTW_{global} Kernel	$\mathbf{97.6 \pm 0.68}$

Fusing eBird, radar, and acoustic data via Observation Models

- Species s
- Observers o
- Sites i
- Acoustic stations k
- Radar sites v

- Observation model for eBird (detection, expertise, etc.)
- Observation model for night flight calls (distance to ground, ambient noise)
- Observation model for radar (signal cone, weather, radar "plankton")



Current Status:
Animation of Static Distribution Models

Indigo Bunting

slide courtesy of Daniel Fink

Three Projects at Oregon State

- TAHMO: Continent-Scale Hydro-Meteorological Network
- BirdCast: Continent-Scale Bird Migration Forecasting
- Wildfire Management: Spatial optimization of fire management

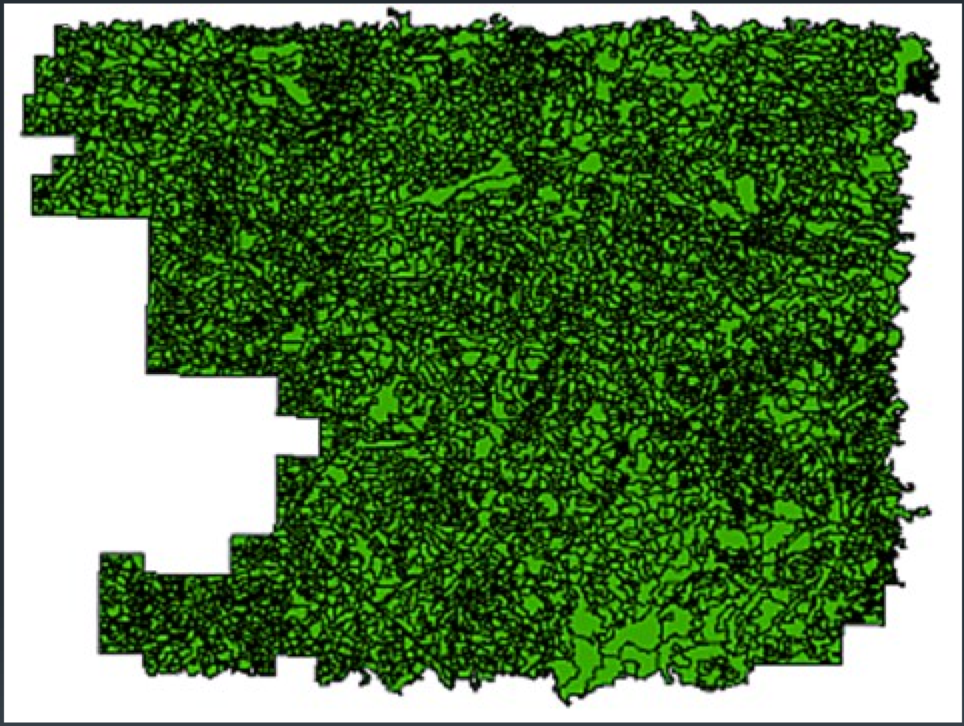
Managing Wildfire in Eastern Oregon

- 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

Study Area: Deschutes National Forest

- Goal: Return the landscape to its "natural" fire regime
- Management Questions:
  - LET-BURN: When lightning ignites a fire, should we let it burn?
  - FUEL TREATMENT: Where should we perform mechanical fuel reductions?
- ~4000 management units



Study area in Deschutes National Forest

LET BURN

- Relevant factors:
  - Weather forecast
  - Is there likely to be rain that will extinguish the fire?
  - State of the landscape
    - Fuel conditions at the point of ignition
    - Fuel conditions downwind, up slope, etc.
  - Benefits that this fire would have on future fires
    - Would this fire create a barrier that would prevent other fires from growing too large?

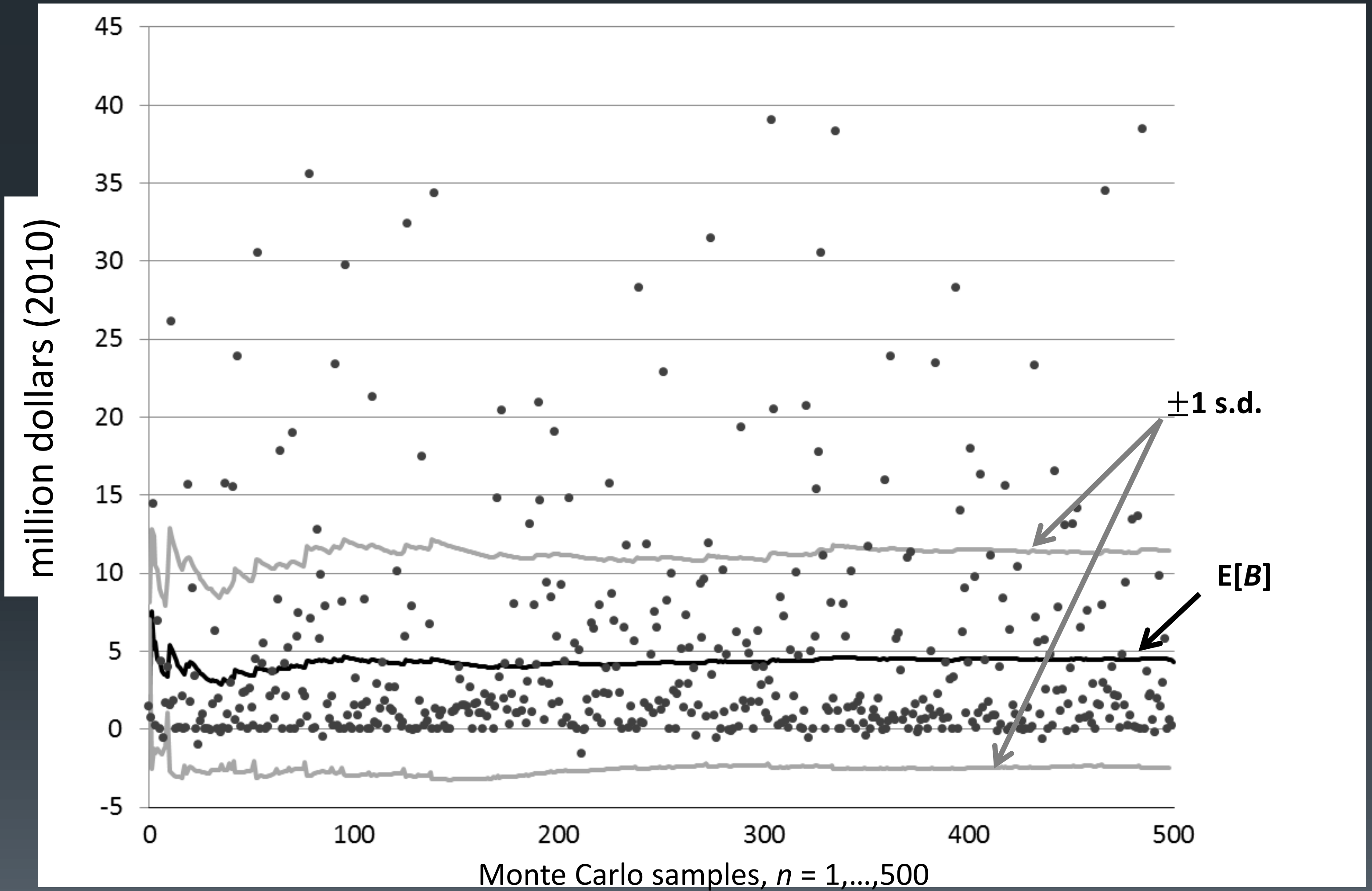
- Central Challenge:
    - 4000 management units
    - Each unit in one of 25 states
    - 25^{4000} possible states!

Two Studies

- Study 1: LET BURN only in year 1, then SUPPRESS for 99 years
  - Solved via Monte Carlo simulations

- Study 2: For 100 years, for each ignition, decide whether to LET BURN or SUPPRESS
  - Requires solving 100-year stochastic dynamic program
  - We don’t know how to do this (yet)

STUDY 1 Monte Carlo Results



FUEL TREATMENT: Game Against 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)

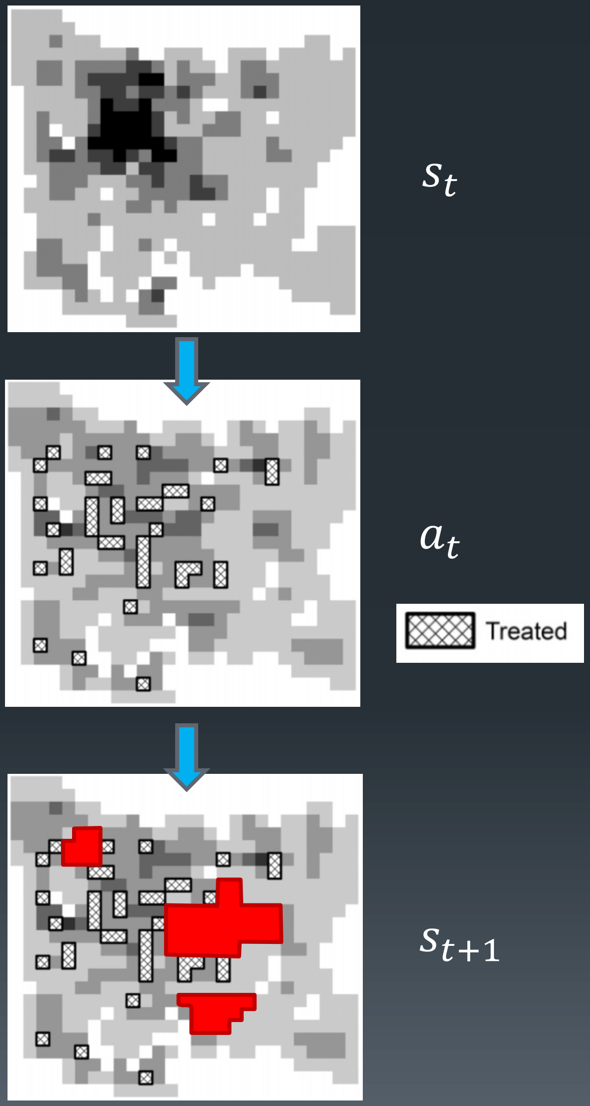
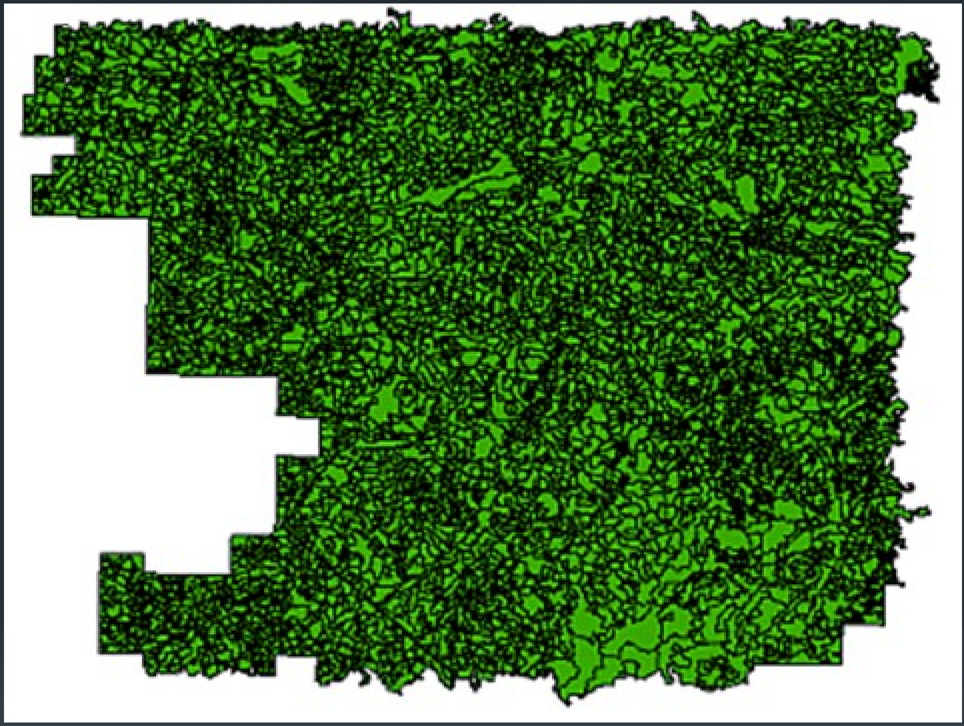


Image: Wei et al, 2008

Formulation as a Markov Decision Process

- 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



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

Summary: Computational Sustainability

- Data → Models → Policies

- Three projects
  - TAHMO
  - BirdCast
  - Fire Management

Computational Sustainability

- There are many opportunities for ICT to contribute to a sustainable planet
- There are many challenging research problems to be solved
- Institute for Computational Sustainability: http://www.computational-sustainability.org/

Thank-you

- John Selker (OSU), Nick van de Giesen (TUDelft): TAHMO
- Steve Kelling (Cornell), Dan Sheldon (UMass): BirdCast
- Claire Montgomery, Rachel Houtman, and Sean McGregor (OSU): Wildfires

- National Science Foundation Grants 0705765, 0832804, and 0905885