Modeling bird migration by combining weather radar and citizen science data

Tom Dietterich '77

Oregon State University

In collaboration with

Postdocs: Dan Sheldon (now at UMass, Amherst)

Graduate Students: Liping Liu, Tao Sun, Akshat Kumar

Cornell Lab of Ornithology: Steve Kelling, Daniel Fink, Andrew Farnsworth, Wes Hochachka, Benjamin Van Doren, Kevin Webb

Bird Migration

- Many bird species are declining. Why?
  - Loss of summer and winter habitat
  - Loss of stop-over habitat during migration
  - Cats
  - Skyscrapers
  - Airplanes
  - Wind farms
  - Food asynchrony due to climate change

Understanding Bird Migration

- We need better models of
  - Required habitat for each species
  - Detailed dynamics of bird migration

- Bird decision making??
  - Absolute timing (e.g., based on day length)
  - Temperature
  - Wind speed and direction
  - Relative humidity
  - Food availability

Methodology

Mathematical Modeling → Model Fitting to Data → Policy Optimization

Step 1: Mathematical Modeling

$$s_1 \rightarrow s_2 \rightarrow s_3 \rightarrow s_4 \qquad \rightarrow s_T$$

$P(s_1)$: Initial State Distribution
$P(s_t|s_{t-1})$: State transition function

- Markov Process
  - The state at time $t+1$ depends only on the state at time t (and not the “history” of earlier states)
- Vector/Matrix representation

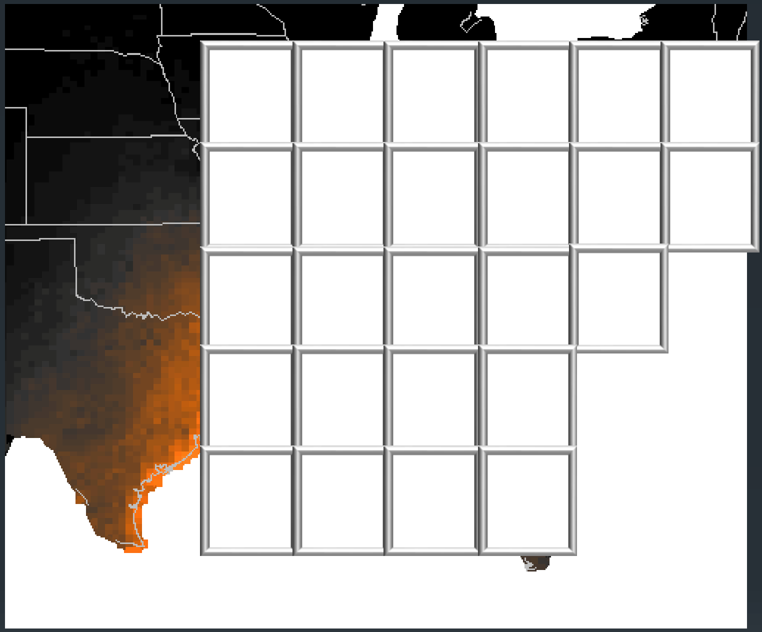
$$\begin{bmatrix} 0.25 \\ 0.50 \\ 0.25 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.50 & 0.0 & 0.0 & 0.0 \\ 0.50 & 0.50 & 0.0 & 0.0 \\ 0.0 & 0.50 & 0.50 & 0.0 \\ 0.0 & 0.0 & 0.50 & 1.0 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \\ 0 \\ 0 \end{bmatrix}$$

$$P(s_t = j) = \sum_x P(s_t = j|s_{t-1} = i)P(s_{t-1} = i)$$

States of our Markov Process = Grid Cells

- 36x28 grid of cells over Eastern US
- 1008 cells

- Problem 1: There are 1008 x 1008 = 1,000,064 transition probabilities to determine
- Problem 2: The transition probabilities are time-invariant, whereas we need them to change
  - Depending on the season
  - Depending on weather conditions

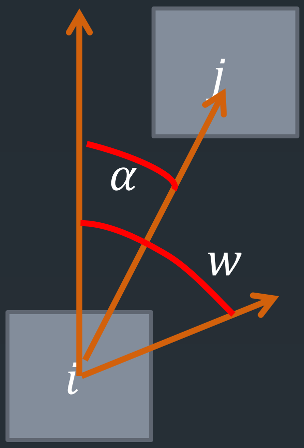


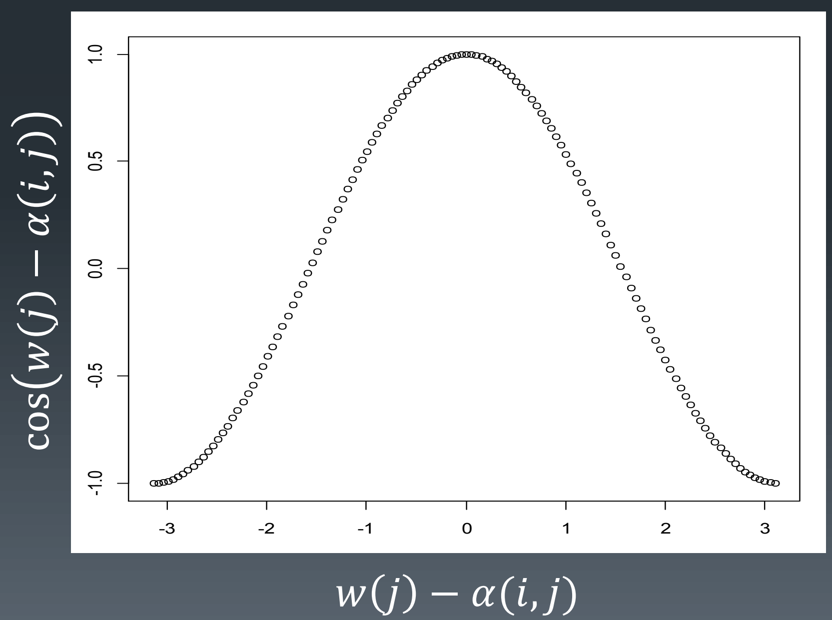
Solution: Make the transition probabilities depend on variables ("covariates")

- In each cell i on each night $(t, t+1)$, we will observe the following covariates $x_{t,t+1}(i)$
  - day of the year: t
  - wind speed: $v_t(i)$
  - wind direction: $w_t(i)$
  - temperature: $temp_t(i)$
  - relative humidity: $rh_t(i)$
- Between each pair of cells i and j we also know
  - distance: $dist(i, j)$
  - direction from i to j: $\alpha(i, j)$

Parametric
State Transition Model

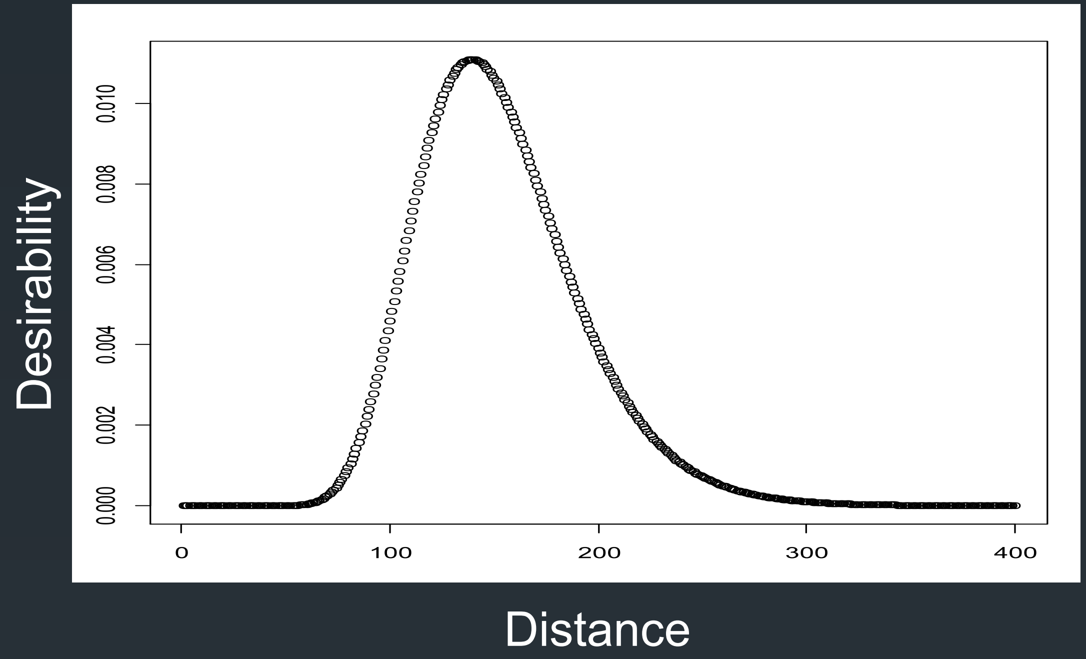
- Let $\alpha(i, j)$ be the heading from i to j
- Let $w(i)$ be the heading of the wind
- Let $v(i)$ be the speed of the wind
- Wind profit $v(i)\cos\big(w(i) - \alpha(i, j)\big)$
  - 1 if perfectly aligned
  - -1 if perfect headwind





Oberlin 2014

Distance Preferences



- $\text{Desirability}(dist) = \ Normal(\log dist\,; \mu, \sigma)$

Preferences for temperature, relative humidity, day of year, etc.

- $\left(temp - \theta_{temp}\right)^2$ ideal temperature
- $(rh - \theta_{rh})^2$ ideal relative humidity
- $t - \theta_{doy}(i)$ ahead/behind schedule

Combine into probability model

- $F(i,j) = \beta_0 + \beta_w v_t(i)\cos(w_t(i) - \alpha(i,j)) + \beta_d \mathrm{Normal}(\log dist(i,j); \mu_{dist}, \sigma_{dist}) + \beta_{temp}\left(temp_t - \theta_{temp}\right)^2 + \beta_{rh}(rh_t - \theta_{rh})^2 + \beta_{doy}\left(t - \theta_{doy}(i)\right) + \cdots$

- $P(s_t = j | s_{t-1} = i) = \frac{\exp F(i,j)}{\sum_{j'} \exp F(i,j')}$

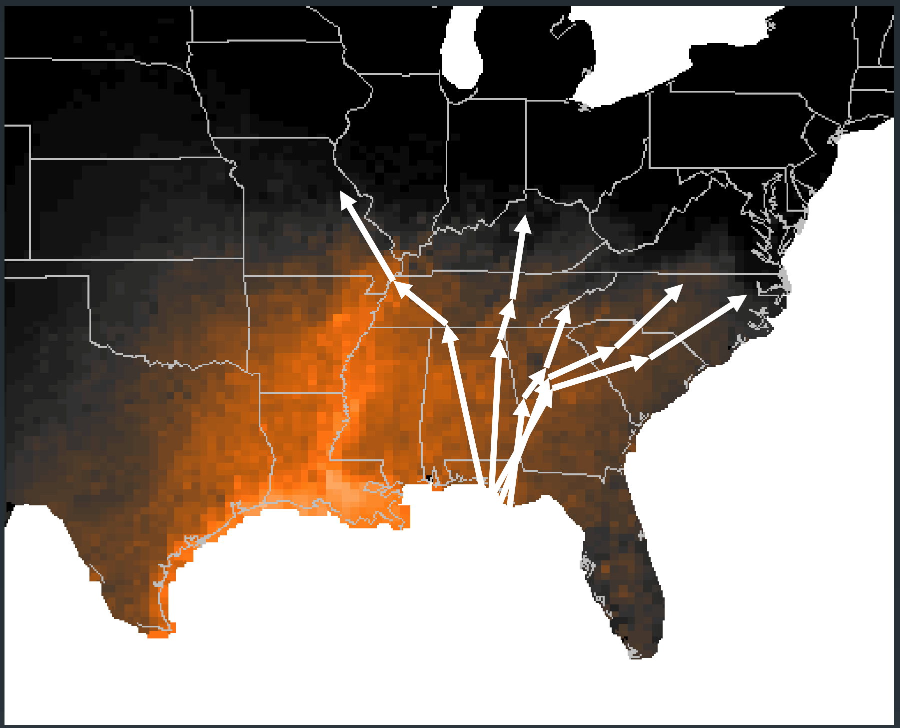
- Construct the transition matrix at time t by evaluating this function for each pair (i,j)

Step 2: Fitting the model to data
The data we wish we had:

- Tracks of individual birds over time
- Weather at every location

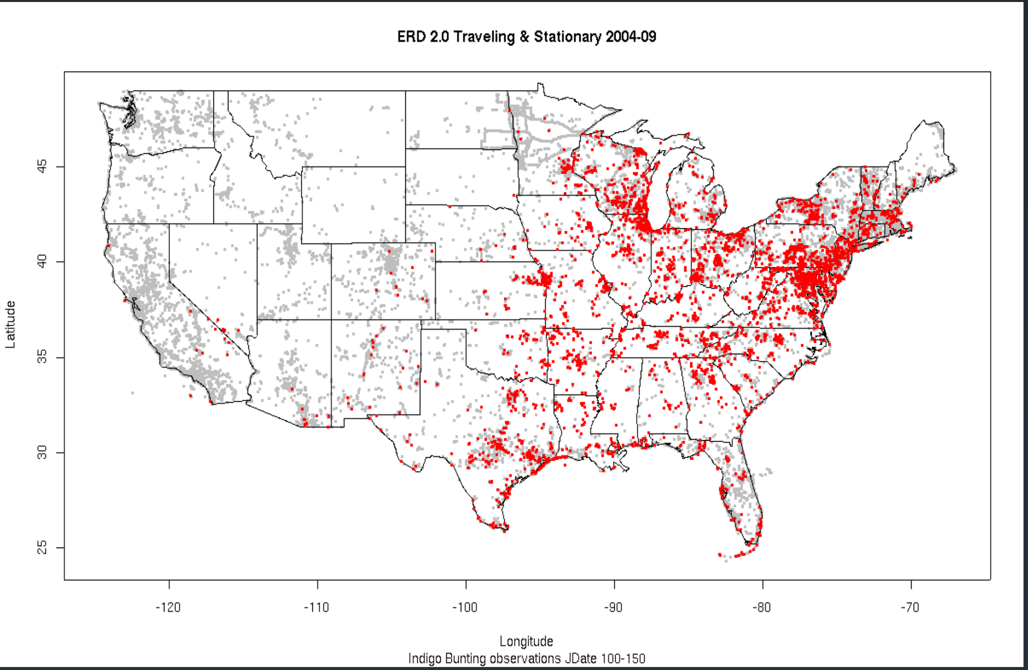

www.azoresbioportal.angra.uac.pt

macworld.com



This would give us points $(x_{t,t+1}(i), s_t(i), s_{t+1}(j))$ to which we could fit our model

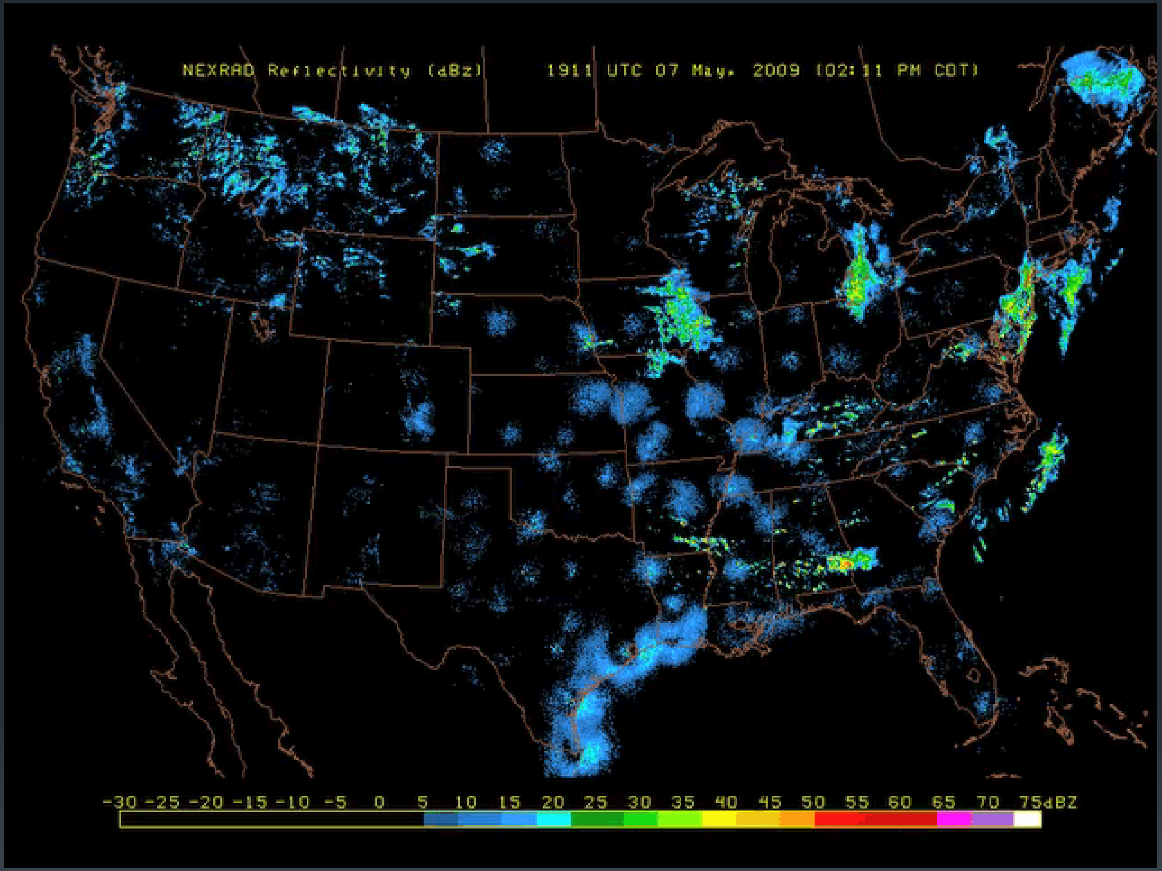
The data we have (1): Project eBird (www.ebird.org)



- Volunteer Bird Watchers
  - Stationary Count
  - Travelling Count
- Time, place, duration, distance travelled
- Species seen
  - Number of birds for each species or ‘X’ which means ≥ 1
- Checkbox: This is everything that I saw
- 8,000-12,000 checklists uploaded per day

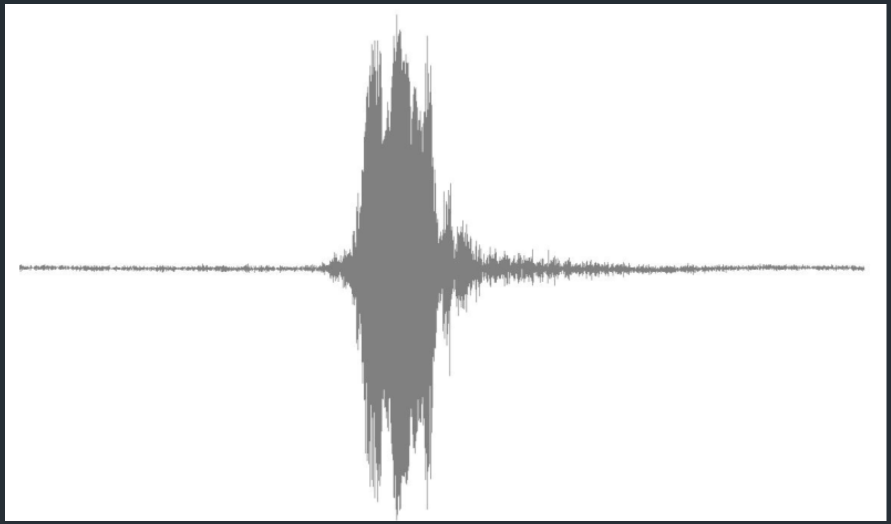
The data we have (2): Weather Radar

- Radar detects
  - weather (remove)
  - smoke, dust, and insects (remove)
  - birds and bats
- Removing weather
  - manual, using a web-based tool
- Removing smoke, dust & insects
  - estimate velocities
  - ignore pixels that are moving at same speed as wind



The data we (hope to) have (3): Acoustic monitoring

- Night flight calls
- People can identify species or species groups from these calls



The data we have (4): Weather data

- North American Regional Reanalysis
  - wind speed
  - wind direction
  - temperature
  - relative humidity

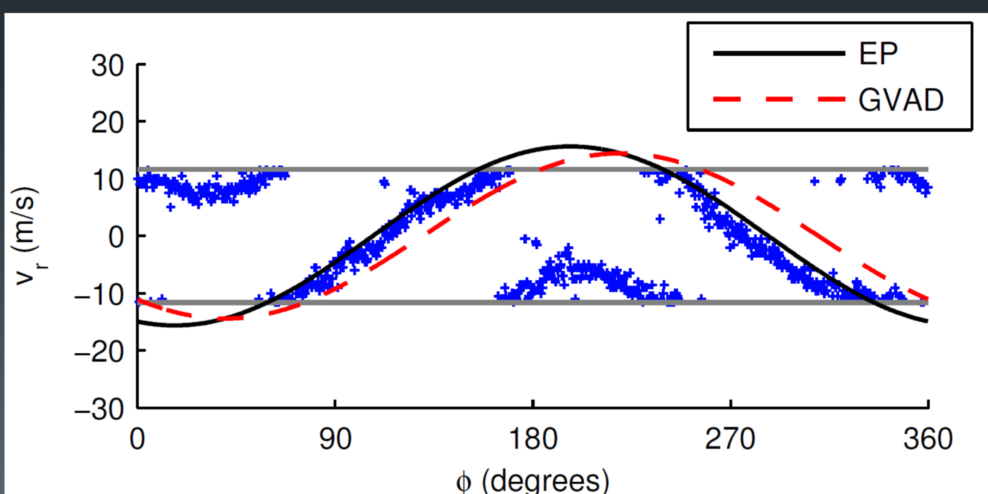
Modeling for each data source (1): eBird

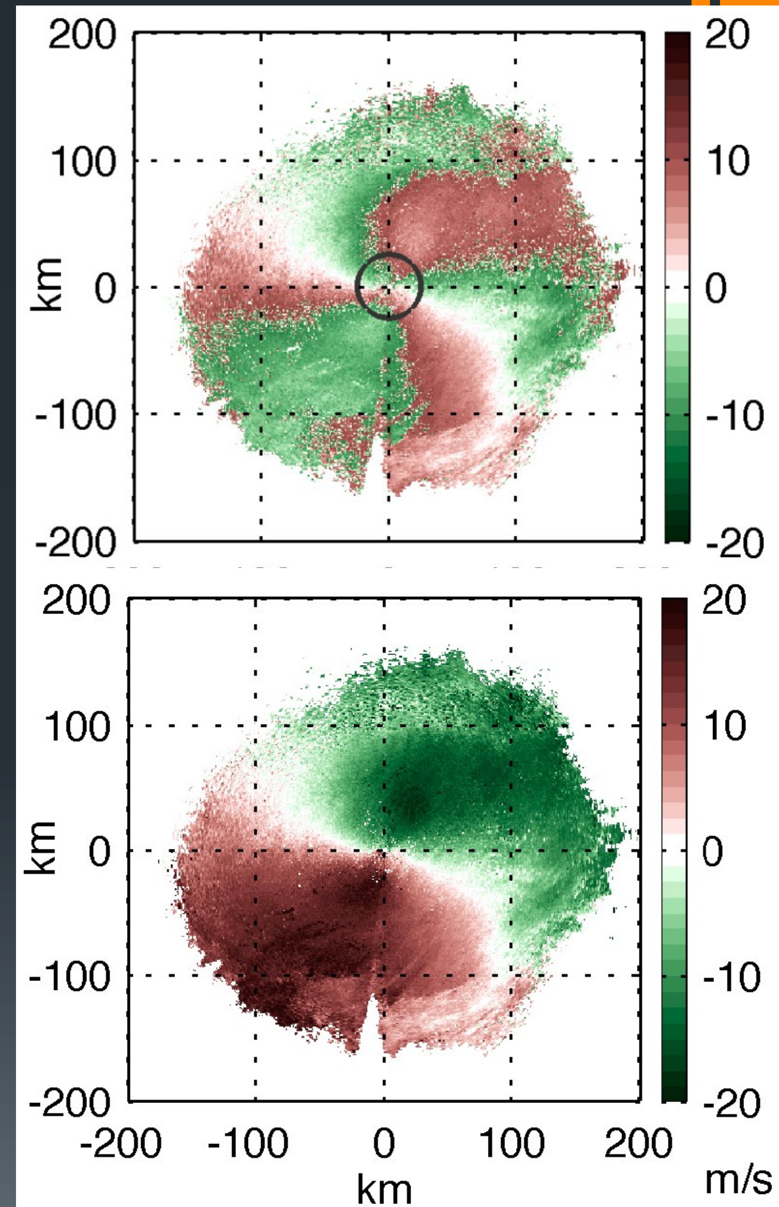
- Bird watchers do not detect *all* birds at a given location
  - detection probability
  - day of year
  - weather conditions
  - habitat (shoreline, meadow, dense forest)
  - expertise of the bird watcher
- Bird watchers may misidentify species

- Yu, J., Wong, W-K., and Hutchinson, R. (2010). Modeling Experts and Novices in Citizen Science Data for Species Distribution Modeling. Proceedings of the 2010 IEEE International Conference on Data Mining
- Yu, J., Wong, W-K. and Kelling, S. (2014). Clustering Species Accumulation Curves to Identify Skill Levels of Citizen Scientists Participating in the eBird Project. IAAI 2014
- Yu, J., Hutchinson, R. and Wong, W-K. (2014). A Latent Variable Model for Discovering Bird Species Commonly Misidentified by Citizen Scientists. AAAI 2014

Modeling for each data source (2): Weather radar

- Radar measures Doppler shift
  - Gives radial velocity r
  - Velocity is aliased: $r \ mod \ 2V_{max}$
- We developed a maximum likelihood model (EP) that includes the mod operator inside the likelihood function
  - “fix the model instead of the data”
  - Sheldon et al. (2013)
- Bird biomass per km^3

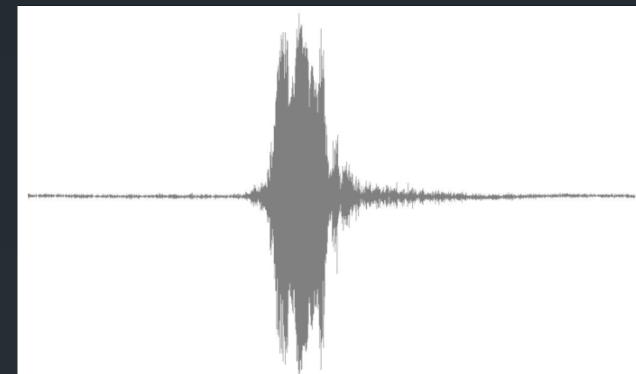


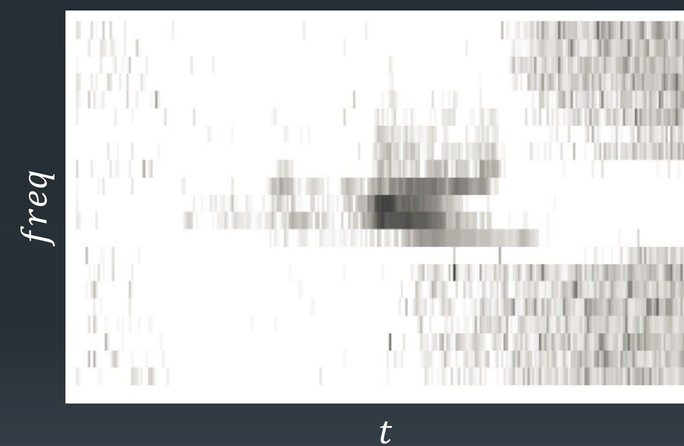


Radar Visualization

Modeling for each data source (3)
Night flight calls

- Fourier analysis over short time windows to obtain a spectrogram
- Dynamic time warping to match to spectrograms of known species
  - similar to DNA sequence alignment
  - allows time to stretch or shrink (with a penalty)
- Apply machine learning algorithm to predict the species
- Accuracy: 97% on 5 species (clean data using captive birds)





- Damoulas, Henry, Farnsworth, Lanzone, Gomes (2010). Bayesian classification of flight calls with a novel Dynamic Time Warping Kernel (ICDM 2010).

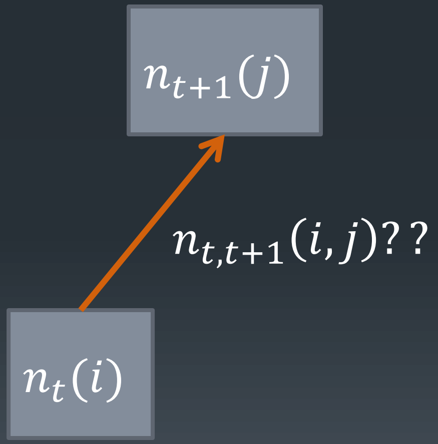
Modeling for each data source (4) NARR data

- NARR data product is the result of performing “data assimilation”
  - Observed variables from radiosonde balloons
  - Update a physics-based model of the atmosphere via Bayes theorem

www.ncdc.noaa.gov

Challenge: Aggregate anonymous counts

- We do not observe the behavior of individual birds
- We only obtain information about aggregated counts of birds

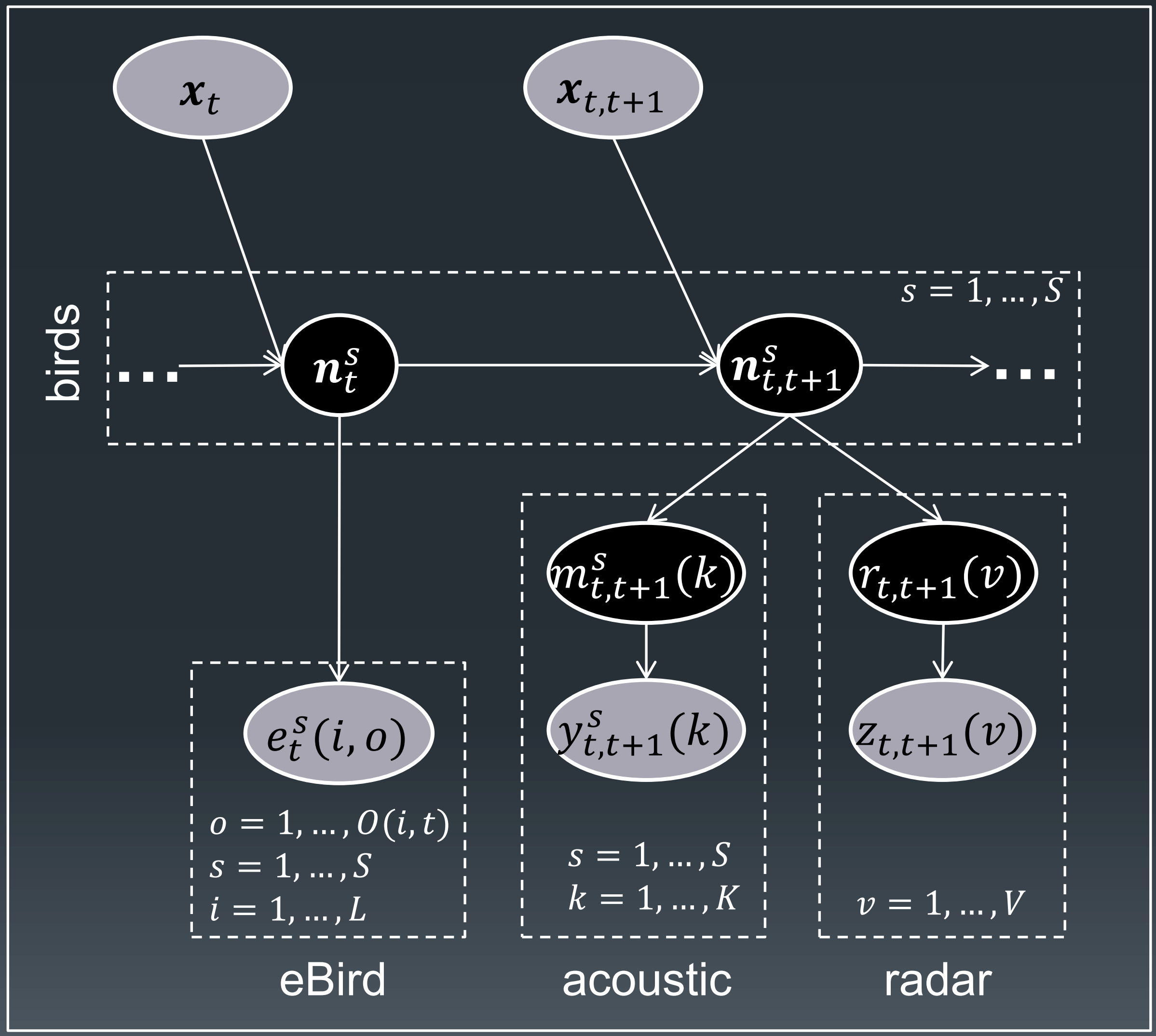


Solution: Collective Graphical Models

- New method for fitting models of individual behavior from noisy aggregate counts

- Assumes all birds make their decisions independently according to the same $P\left(s_{t+1} = j \middle| s_t = i, x_{t,t+1}(i,j)\right)$

Full Migration Model



Fitting Latent Variable Models

Expectation Maximization (EM; MAP version)

1. Make initial guess about the parameter values
   $\Theta = \beta_0, \beta_w, \beta_{temp}, \theta_{temp}, \beta_{rh}, \theta_{rh}, \beta_{doy}, \theta_{doy}(i), \beta_{dist}, \mu$ [illegible]
2. Compute the most likely number of birds flying from cell i to cell j each night (for all i, j). $\boldsymbol{n}^{s}_{t,t+1}(i \rightarrow j)$.

   "Maximum Aposteriori Probability (MAP) estimate"
3. Pretend these are the true values of the latent variables and adjust the parameters Θ to maximize the likelihood of the $\boldsymbol{n}^{s}_{t,t+1}(i \rightarrow j)$ values:

   $$\operatorname{argmax}_{\Theta} P\left(\boldsymbol{n}^{s}_{t,t+1} \middle| \Theta\right)$$
4. Repeat 2-3 until convergence

"E-Step"
Very Difficult

"M-step"
Easy: Can be solved with gradient descent

Intractability of the E step in the Collective Graphical Model

- Let M be the population size
- Let L the number of grid cells
- Theorem: Unless $P = NP$, there is no exact inference algorithm with runtime that is simultaneously polynomial in both M and L

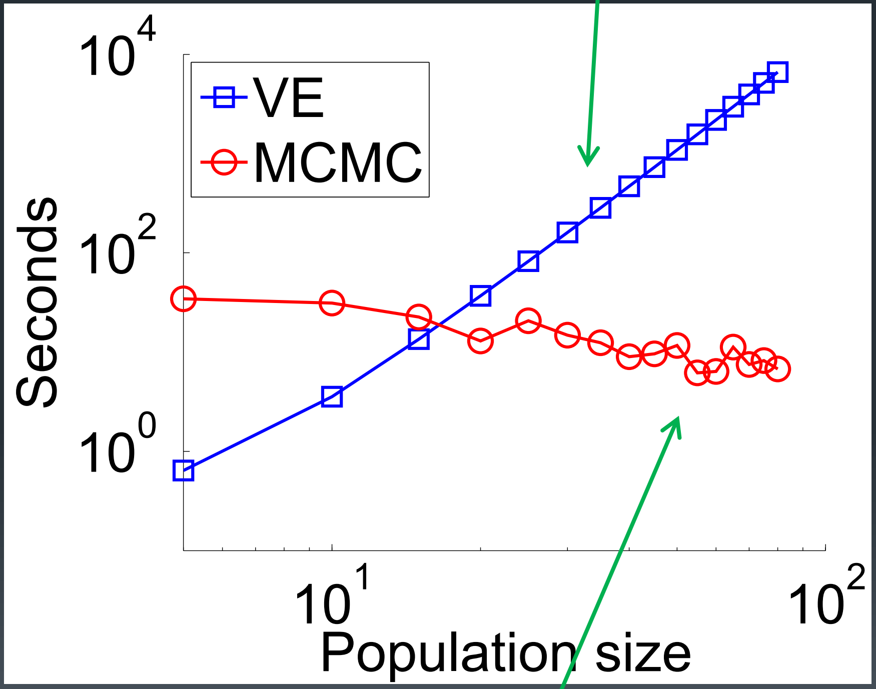
- Bird migration has $M \approx 10^9$ and $L = 1008$
- We must approximate!!

Approximation #1: Markov Chain Monte Carlo (MCMC) Algorithm

(Sheldon & Dietterich, NIPS 2011)

- Samples from $P(\boldsymbol{n}_{t,t+1} | \boldsymbol{n}_1, \ldots, \boldsymbol{n}_T)$
  - posterior distribution of “flows” from cell to cell
  - respects Kirchoff’s laws
  - running time is independent of population size
  - converges (slowly) to the correct distribution

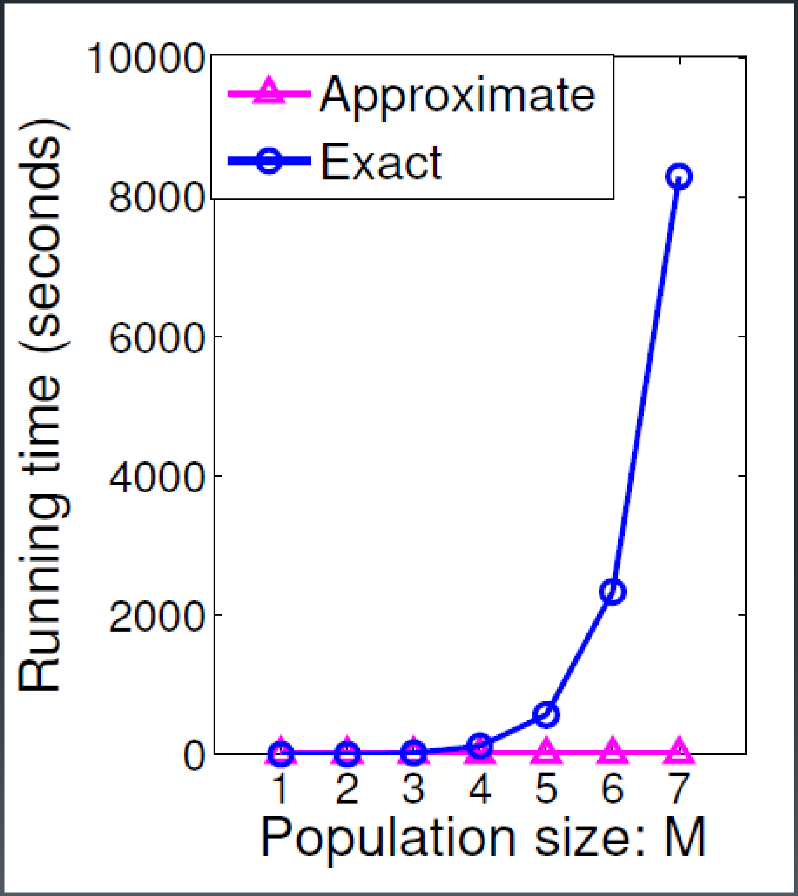
Best exact method



Our method
(to 2% relative error)

Approximation #2: MAP approximation
(Sheldon, Sun, Kumar, Dietterich, ICML 2013)

- Approximate MAP inference
  - Continuous relaxation (allow counts to be real numbers)
  - Sterling's approximation: $\log n! \approx n \log n - n$
- Theorem: With these two approximations, the CGM log likelihood is convex
- Solve using Matlab interior point solver

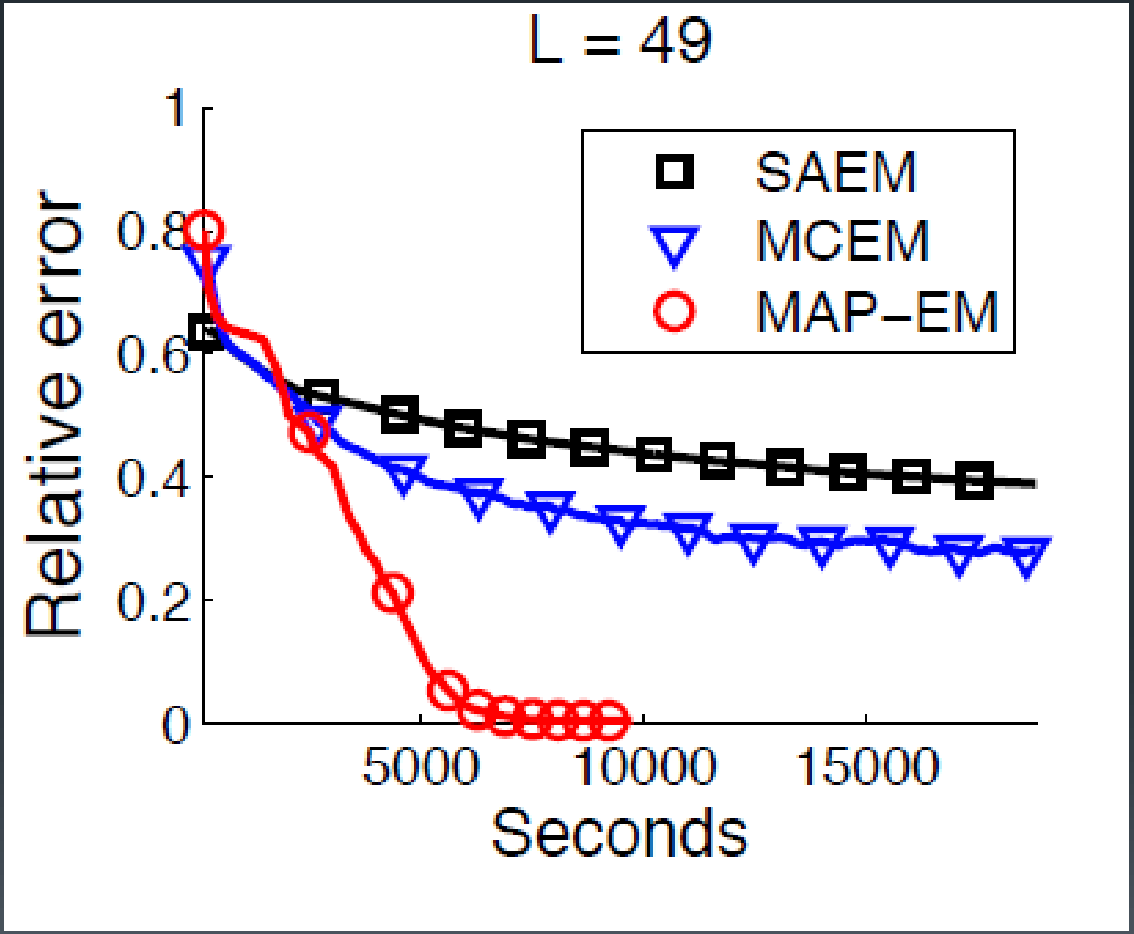


Comparison of #1 and #2: Accuracy and speed of parameter fitting

SAEM: Stochastic approximation EM

MCEM: MCMC + EM

MAP-EM: MAP approximation + EM



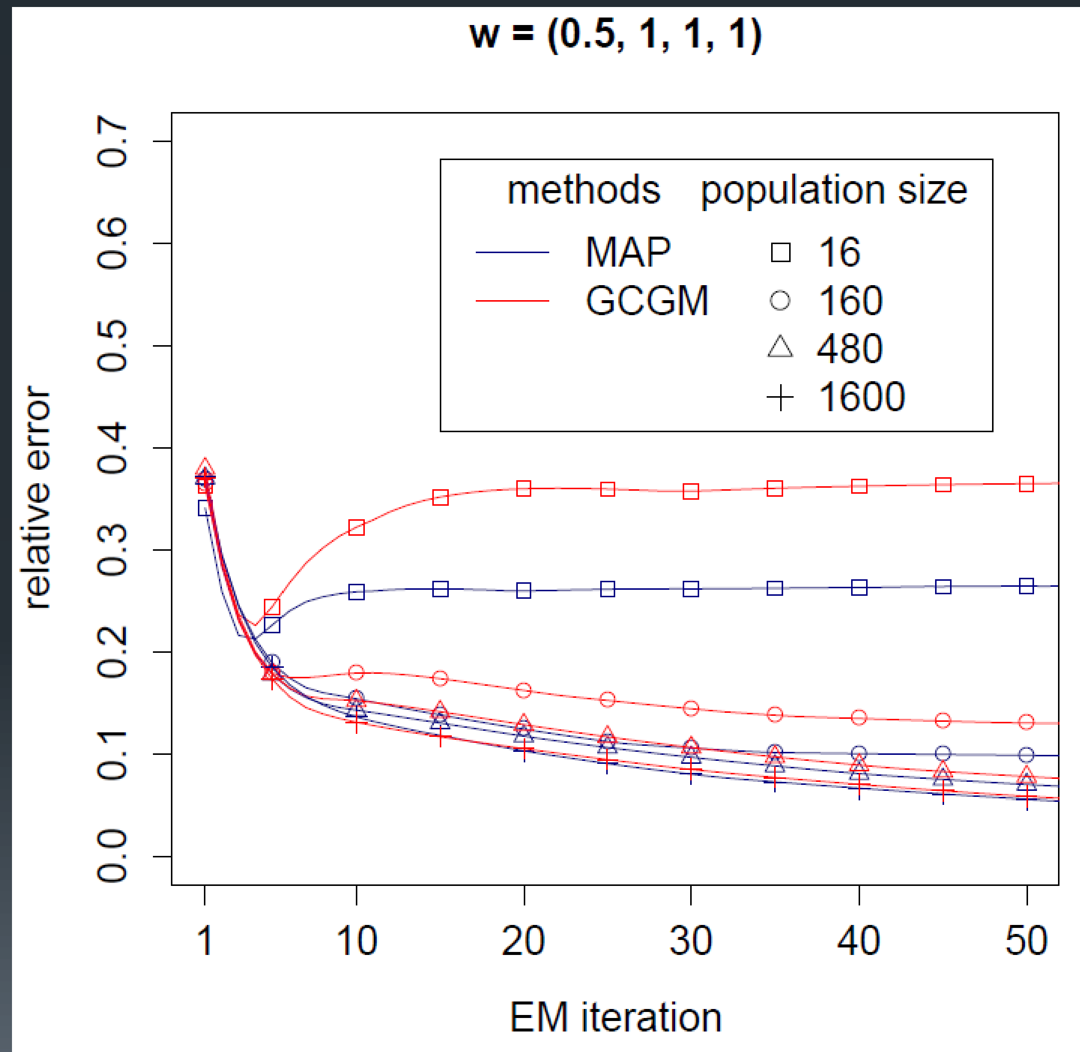
Approximation #3: Gaussian Approximation
(Liu, Sheldon, Dietterich, 2014)

- The statistics in the CGM are combinations of multinomial distributions
- The multinomial distribution can be approximated well by a multivariate Gaussian distribution once the counts are large enough

- Theorem:
  - The Gaussian CGM converges in distribution to the exact CGM as $M \to \infty$
  - The Gaussian CGM has the same sparsity structure as the CGM

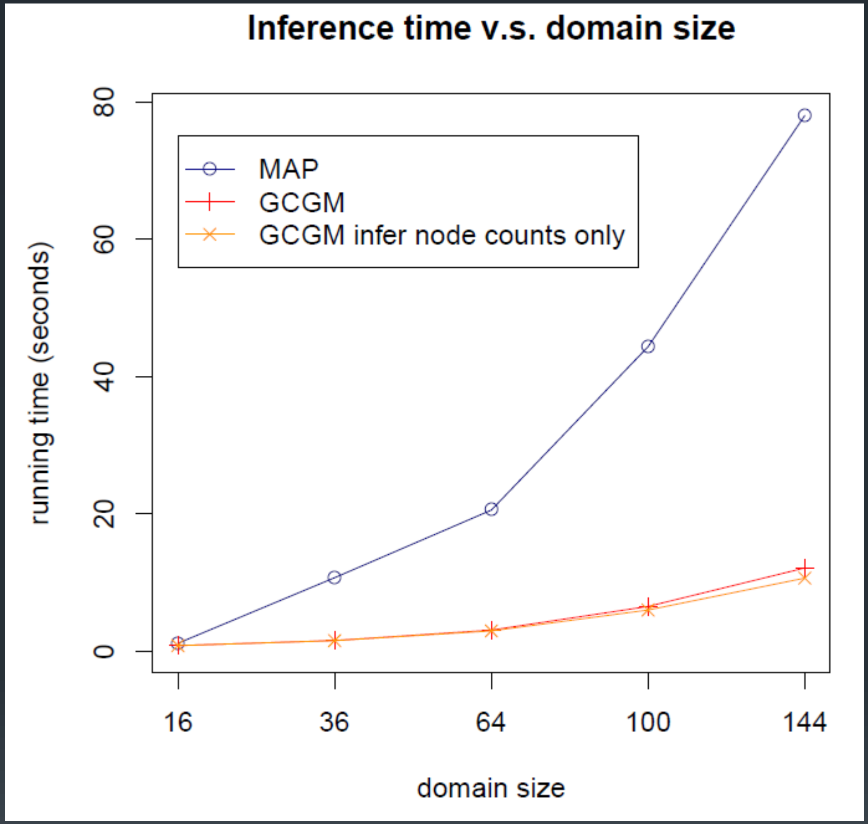
Comparison of #2 and #3: Fitting the parameters

- If M is too small, both the MAP approximation and the GCGM lose badly, but GCGM is much worse
- For $M \geq 480$, GCGM gives answers identical to those of the MAP approximation

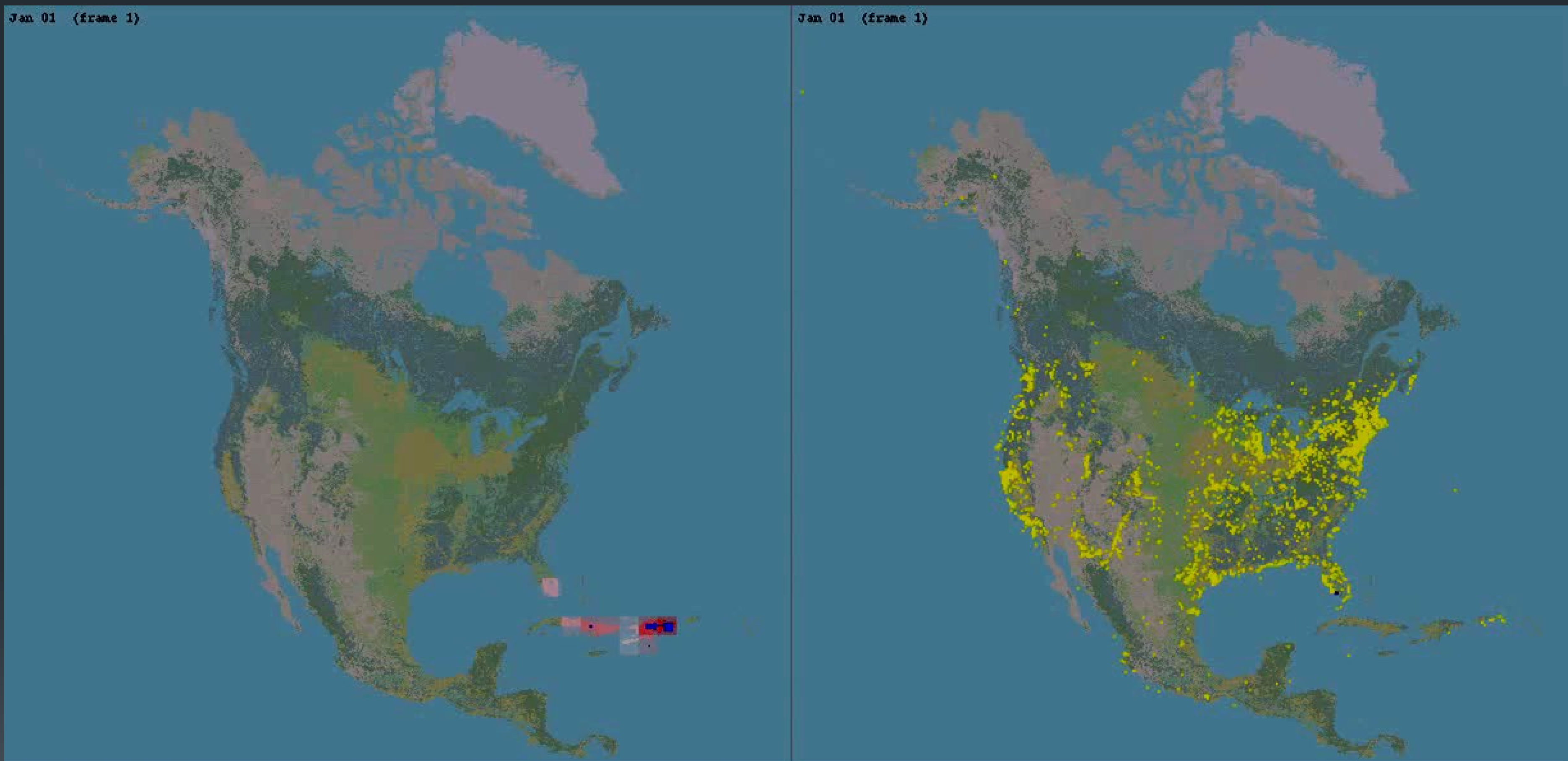


Comparison of #2 and #3: Computation Speed

- We expect a 100-fold speedup on a 1008-cell grid



Initial Results: Movement Reconstruction [Sheldon, 2009]



Fitted Migration Model

Observations (eBird volunteers)

Black-throated Blue Warbler

Current Status

- We have developed a faster algorithm for the MAP approximation (approximation #4)
- We are currently fitting both the MAP (#2) and GCGM (#3) methods to the eBird data

Step 3: Policy Optimization

Policy Questions:

1. Where should conservation reserves and habitat restoration be performed?
   - Examine which cells are being used by the birds
   - We have also developed habitat models directly from eBird data
2. Where should wind farms be located?
3. When and where should low-altitude flight training be allowed?
4. When should wind turbines be operated?
5. When should lights in skyscrapers be turned off?
6. Where should I go bird watching if I want to see species s?

Summary

Mathematical Modeling → Model Fitting to Data → Policy Optimization

- Modeling:
  - Non-linear probabilistic model of the behavior of individual birds
  - Collective graphical model (in order to work with aggregate data)
- Fitting to Data:
  - EM algorithm
  - Computational complexity requires developing algorithms for approximate inference
- Policy Optimization:
  - Straightforward in this application

Open Problems: Uncertainty and Robustness

Mathematical Modeling → Model Fitting to Data → Policy Optimization

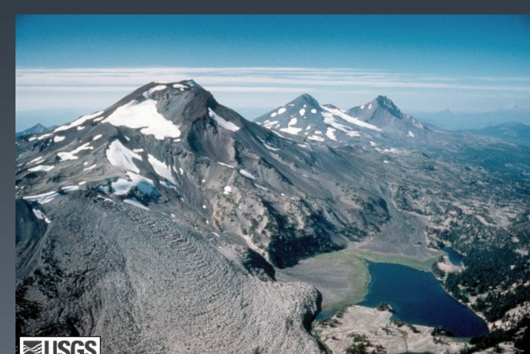
- Uncertainty:
  - Errors in our model
  - Errors in the models of each data source
  - Errors resulting from noisy and insufficient data
  - Errors from computational approximations
- Robustness:
  - How can we make our policies robust to both the known and unknown errors in our models?

Opportunities at Oregon State

- "Spring Break Class in Monte Carlo AI"
  http://web.engr.oregonstate.edu/mcai

- Summer REU program: Eco-Informatics Summer Institute
  http://eco-informatics.engr.oregonstate.edu/

- PhD and Postdoc Research Projects
  - Fundamental research in machine learning and AI with applications in sustainability



Thank-you

- Dan Sheldon, Akshat Kumar, Liping Liu, Tao Sun: Collective Graphical Models
- Steve Kelling, Andrew Farnsworth, Wes Hochachka, Daniel Fink: BirdCast
- Carla Gomes for spearheading the Institute for Computational Sustainability

- National Science Foundation Grants 0705765, 0832804, 0905885, 1331932

Questions?

Mathematical Modeling → Model Fitting to Data → Policy Optimization