A Hybrid Spatio-Temporal Model for Wildlife Poaching Prediction Evaluated through a Controlled Field Test

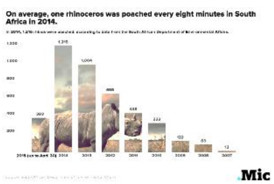
    

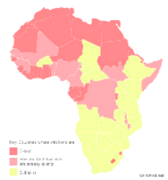
Shahrzad Gholami*, Benjamin Ford*, **Fei Fang**, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, and Joshua Mabonga

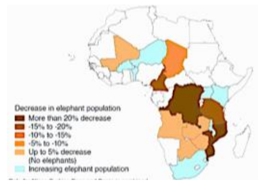
University of Southern California
Carnegie Mellon Univeristy
Wildlife Conservation Society
Uganda Wildlife Authority

* Equal contribution

Decline in Global Wildlife Population







NATIONAL GEOGRAPHIC

African Elephant Numbers Plummet 30 Percent, Landmark Survey Finds

PBS NEWSHOUR

ENDANGERED SPECIES

Report: Giraffe populations in Africa drop 40 percent in 15 years

treehugger

Wild Tiger Population Dropped by 96.8% in 20 Years

BBC

Rhino poaching: Another year, another grim record

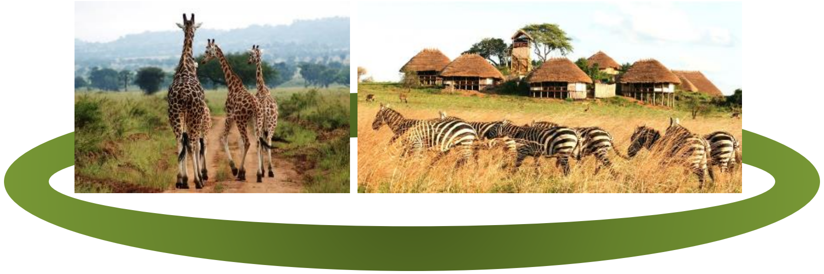
Research Goals

Real World Data
Machine Learning
Human Behavior Modeling

Perceive poachers behavior in-depth

Allocate park rangers strategically

Preserve protected areas efficiently



Problem Statement

How to predict poaching activity in real world?

How to conduct real field patrols to evaluate the predictions?

Contributions

- A Hybrid Spatio-temporal Model to Predict Poaching Activity based on Real World Historical data

- Conduct an extensive study of ML technique and field test in Queen Elizabeth National Park

Controlled field test shows

Predictive power of our model
Insight for organizing rangers efforts
Potential for saving wildlife

Outline

- Introduction
- Problem Statement
- Contributions
- Domain
- Dataset and Challenges
- Learn Model from Real-World Data
- Evaluation
  - Empirical
  - Field Tests
- Summary

Domain: Wildlife Protection in Uganda

Forest Area: QEPA

- Covers 2520 sq. km
- Divided into a grid of $1\text{km}\times 1\text{km}$

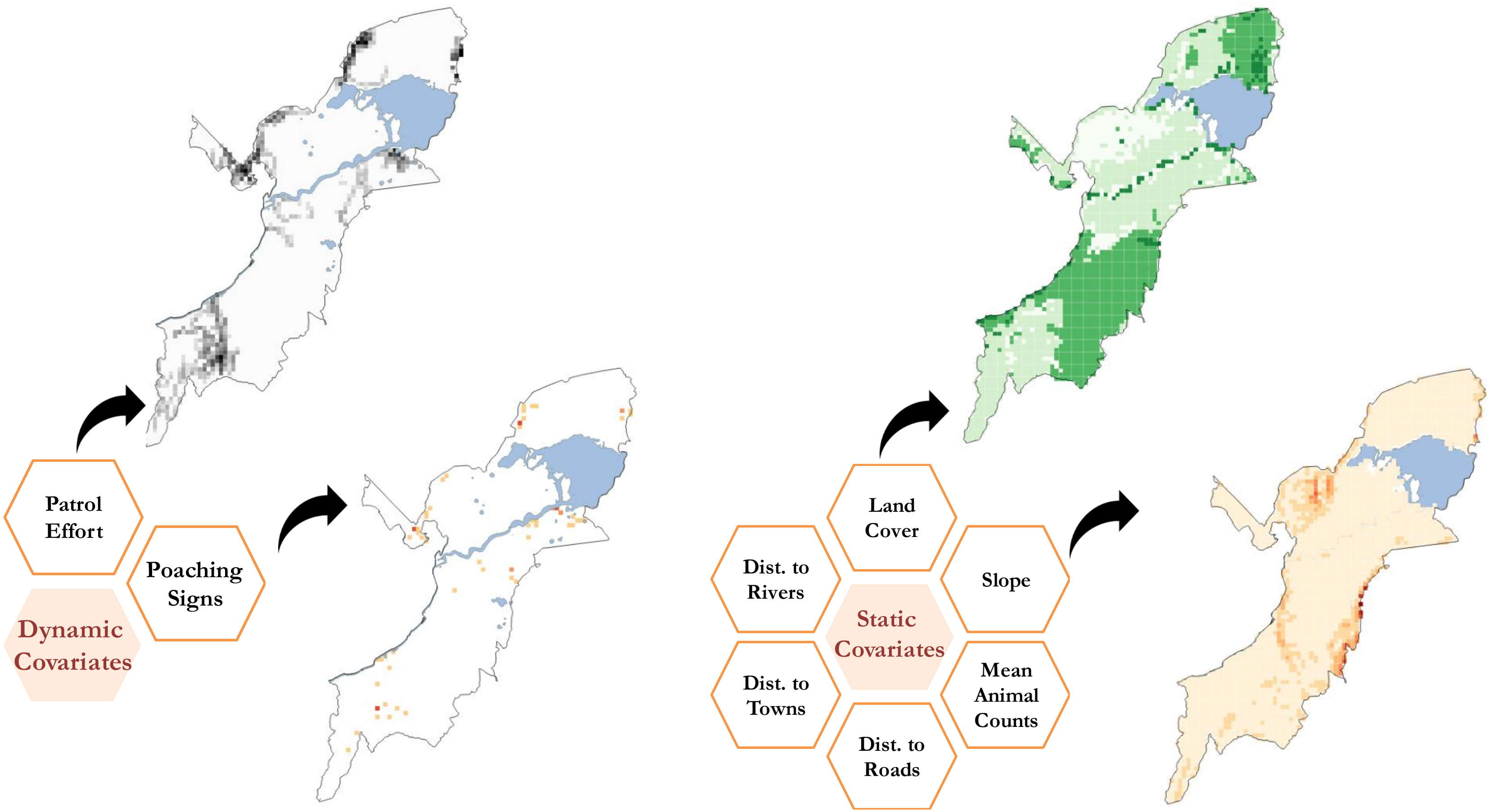
Poachers: Set trapping tools (e.g., snare)

Rangers: Conduct patrols

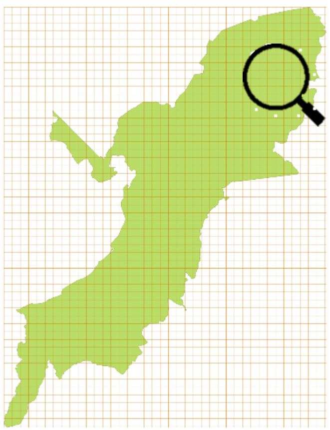
- On foot or by ground vehicles
- From 2003-2017

Collaborators: Wildlife Conservation Society, Uganda Wildlife Authority,
Rangers Pictures: Trip to Indonesia with World Wide Fund for Nature

Dataset Covariates: Queen Elizabeth Park

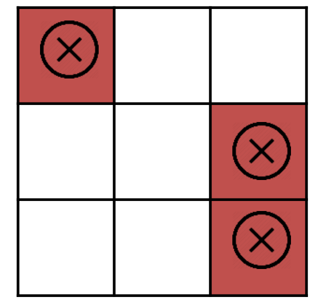


Challenges: Data Uncertainty

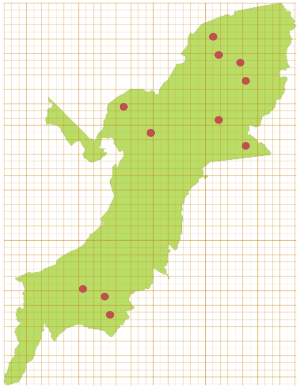


Attacked

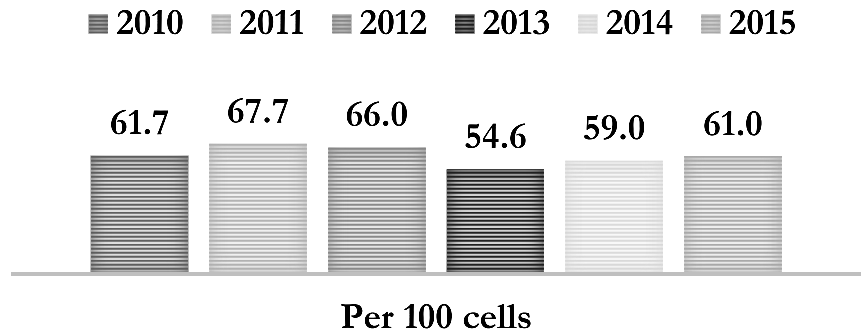
Not Attacked



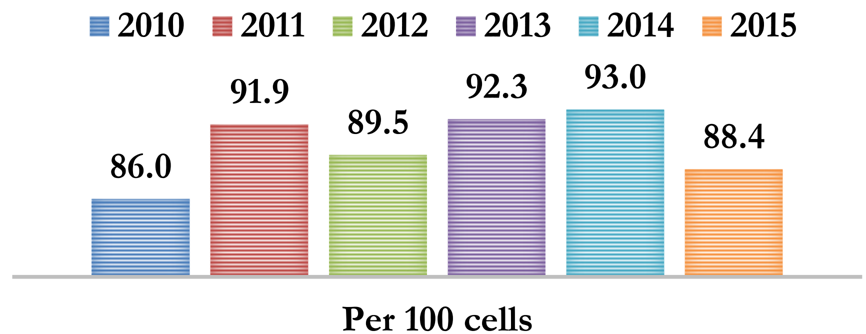
Challenges: Small Number of Recorded Attacks



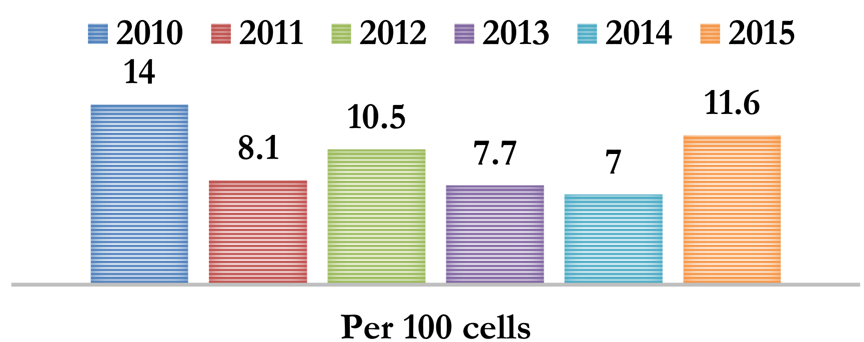
Patrolled Cells
(Year)



Not Attacked Patrolled Cells



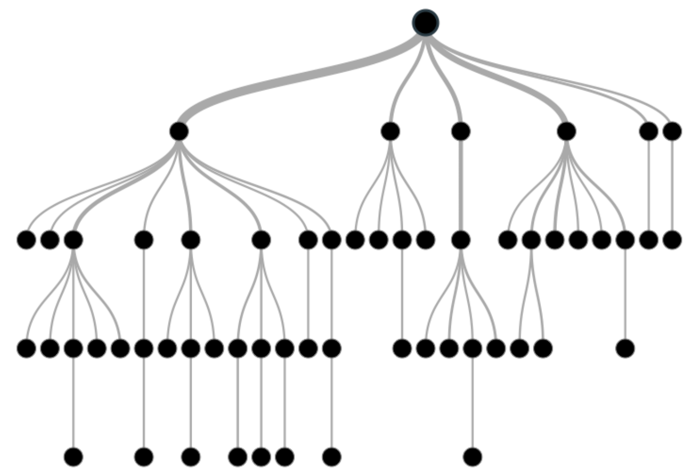
Attacked Patrolled Cells



Outline

- Introduction
- Problem Statement
- Contributions
- Domain
- Dataset and Challenges
- Learn Model from Real-World Data
- Evaluation
    - Empirical
    - Field Tests
- Summary

Classifier: Decision Tree



PROS

- High speed
- Learn global poachers behavior
- Learn nonlinearity in geo-spatial predictor

CONS

- No explicit temporal dimension
- No aspect for label uncertainty

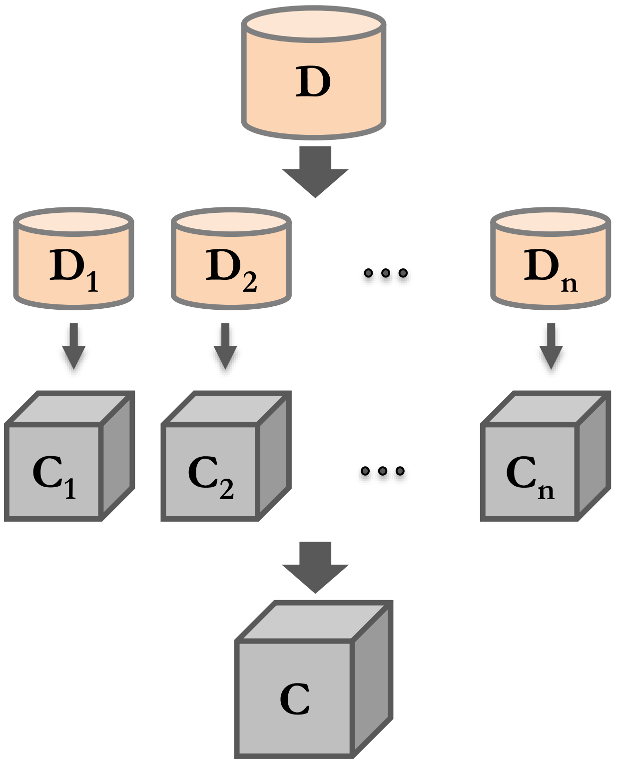
Bagging Ensemble: More Stable, Less Noisy due to Diversification

Original Training Data

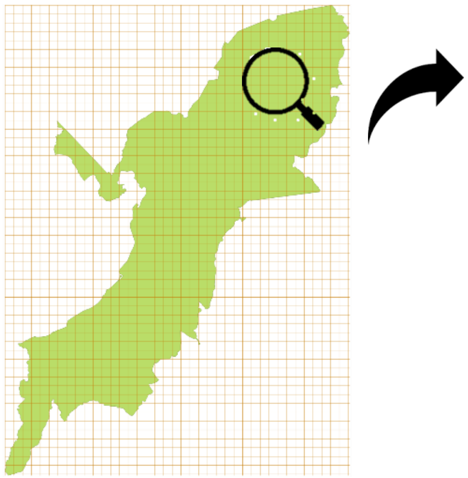
Create Multiple Datasets

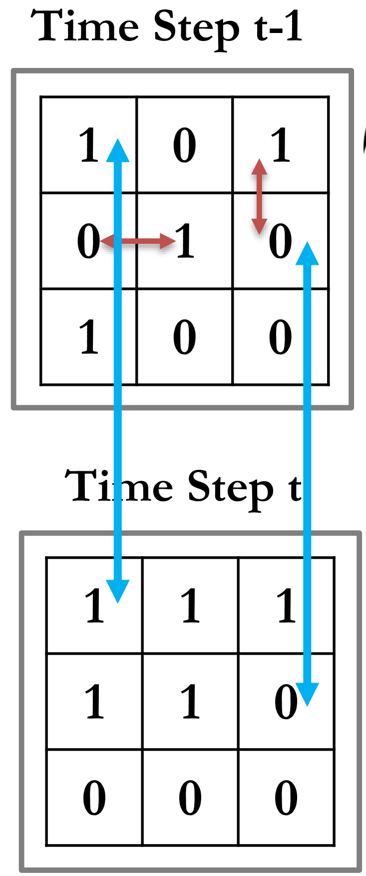
Build Multiple Classifier

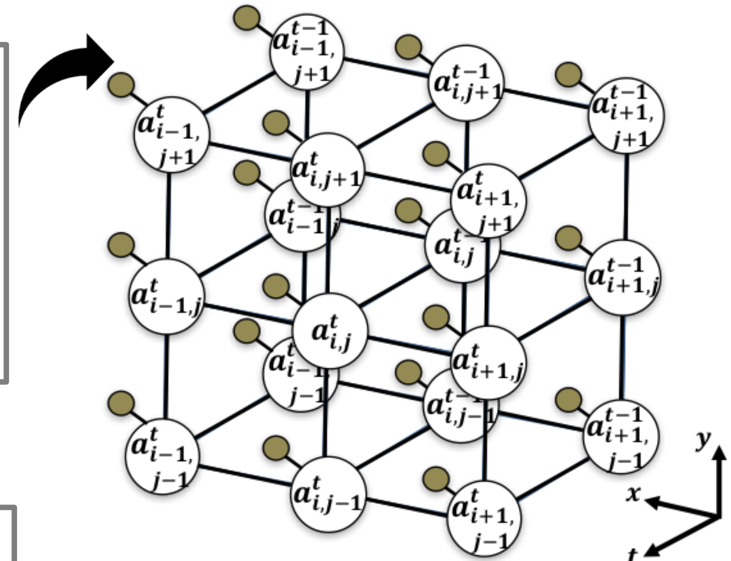
Combine Classifiers



Markov Random Field for Poaching







PROS

- Explicit spatial dimension
- Explicit temporal dimension
- Addresses label uncertainty

CONS

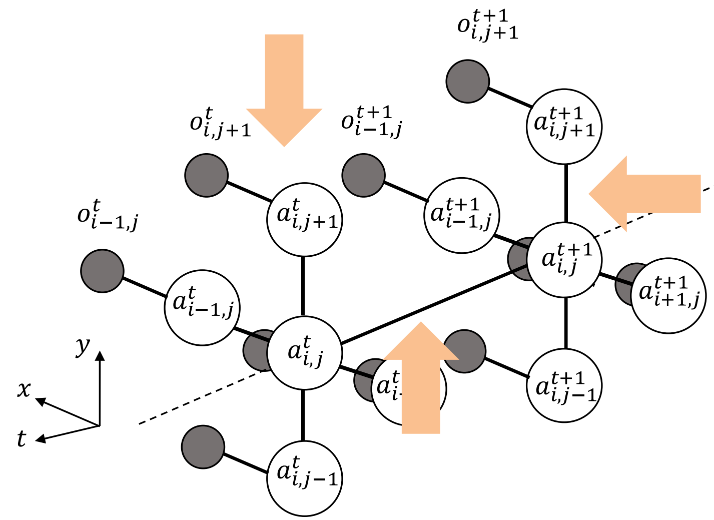
- Low speed
- Data greedy

Markov Random Field for Poaching

Observed Data Cliques

Spatial Cliques

Backward Temporal Cliques

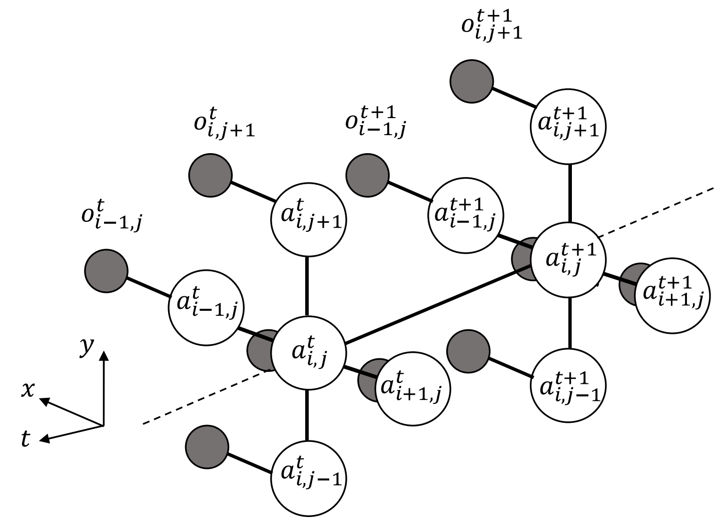


Markov Random Field for Poaching

Observed Data Cliques

$$\phi = \begin{bmatrix} P(o_i = 0 | a_i = 0) & P(o_i = 0 | a_i = 1) \\ P(o_i = 1 | a_i = 0) & P(o_i = 1 | a_i = 1) \end{bmatrix}$$

$$\phi = \begin{bmatrix} 1 & \frac{1}{1+e^{-\boldsymbol{\beta}.[c_i, p_i, 1]^T}} \\ 0 & \frac{e^{-\boldsymbol{\beta}.[c_i, p_i]^T}}{1+e^{-\boldsymbol{\beta}.[c_i, p_i, 1]^T}} \end{bmatrix}$$



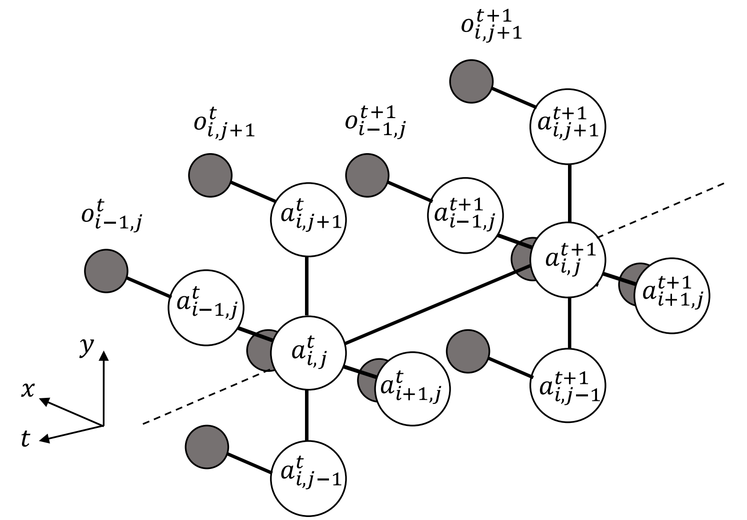
- Coverage, c_i
- Distance from patrol post, p_i

Markov Random Field for Poaching

Spatial Cliques

$$\psi = \begin{bmatrix} P(a_i = 0 | u_{N_i}^{t-1}) \\ P(a_i = 1 | u_{N_i}^{t-1}) \end{bmatrix}$$

$$\psi = \begin{bmatrix} \dfrac{1}{1 + e^{-\boldsymbol{\alpha}.[\boldsymbol{X}, u_{N_i}^{t-1}, c_i, 1]^T}} \\ \dfrac{e^{-\boldsymbol{\alpha}.[\boldsymbol{X}, u_{N_i}^{t-1}, c_i, 1]^T}}{1 + e^{-\boldsymbol{\alpha}.[\boldsymbol{X}, u_{N_i}^{t-1}, c_i, 1]^T}} \end{bmatrix}$$



- Coverage, c_i
- Fraction of neighbors which are attacked, $u_{N_i}^{t-1}$
- All static features including distance from patrol posts, $\boldsymbol{X}$

Learn Parameters: EM

- **Goal:** $\boldsymbol{\theta}^*$ =argmax $P(o|\theta)$
- **E-step,** $\theta = \{\alpha, \beta\}$**:**

$$Q(\theta|\theta^{(k)}) = \mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{o},\theta^{(k)}}[logP(\boldsymbol{a},\boldsymbol{o}|\theta)]$$

$$= \sum_{a\epsilon\mathcal{A}} P\left(\boldsymbol{a} \,\middle|\boldsymbol{o},\theta^{(t)}\right). logP(\boldsymbol{a},\boldsymbol{o}|\theta)$$

- **M-step:**

$$\theta^{(k+1)} = \boldsymbol{argmax}_{\theta} Q(\theta|\theta^{(k)})$$

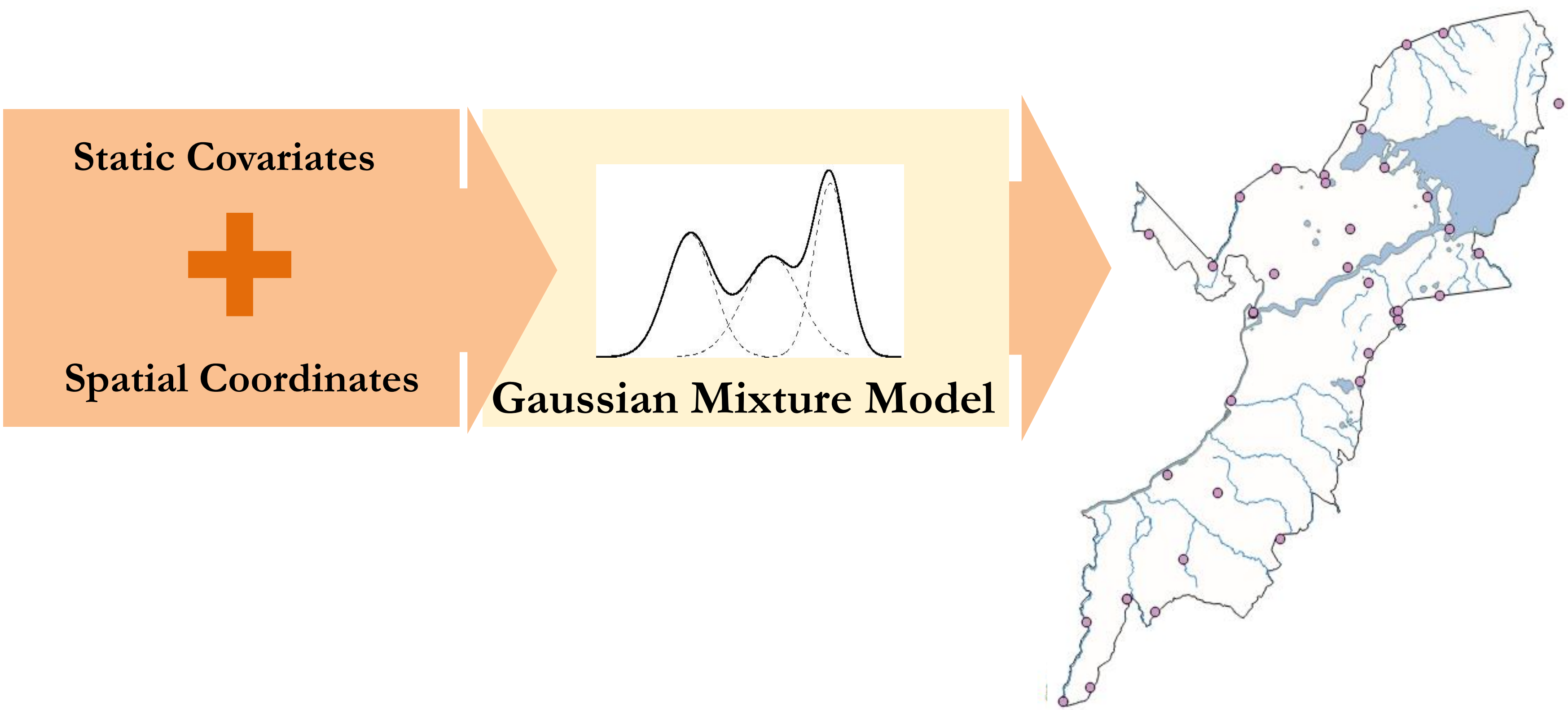
- **Update $\boldsymbol{\theta}$ until convergence:**

$$\theta^{(k)} \leftarrow \theta^{(k+1)}$$

MRF for Geo-clusters

Geo-clusters around patrol posts to learn:

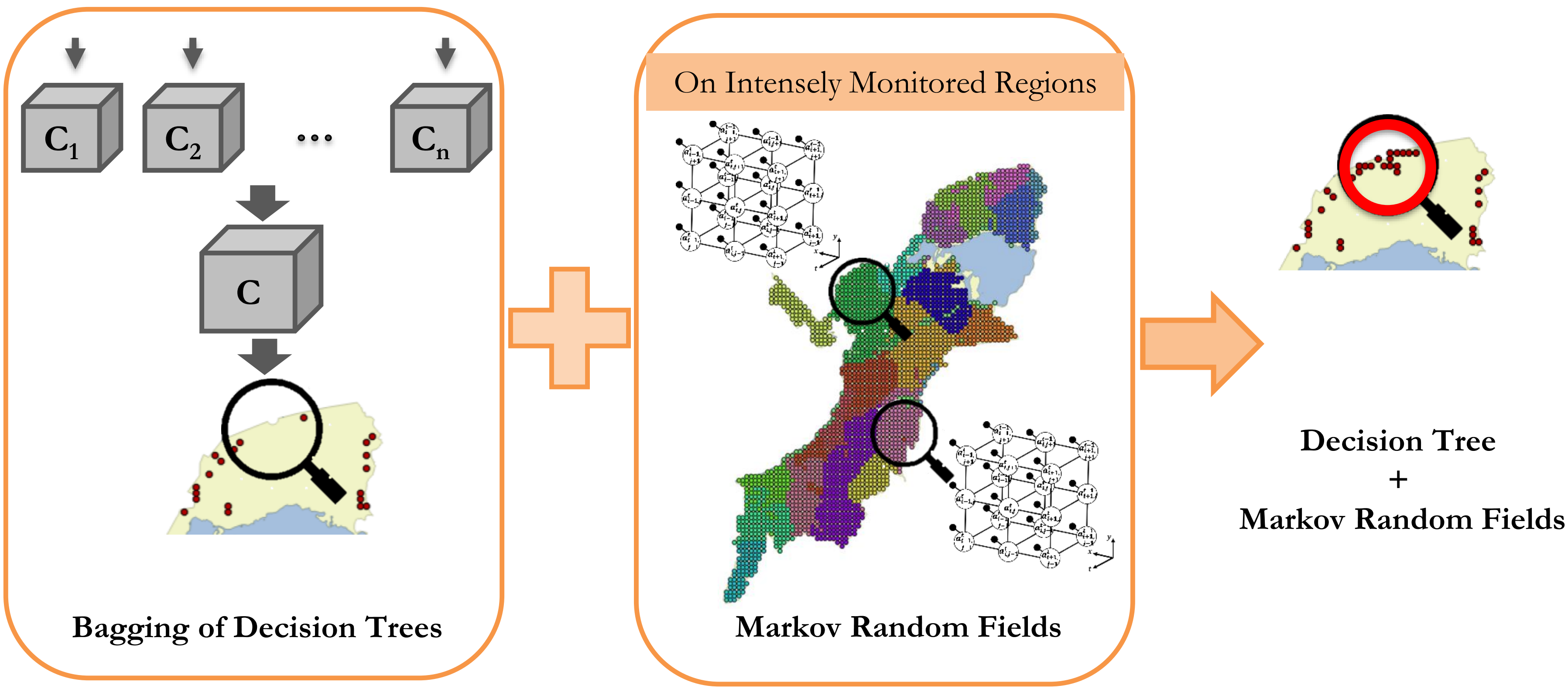
- local poachers' behavior
- Distinct parameters to expedite the local training of MRF



Hybridizing Bagging Model with Markov Random Fields

Boost by geo-clustered behaviorally inspired models:

- Improve the accuracy
- Learn local poachers' behavior; distinct parameters

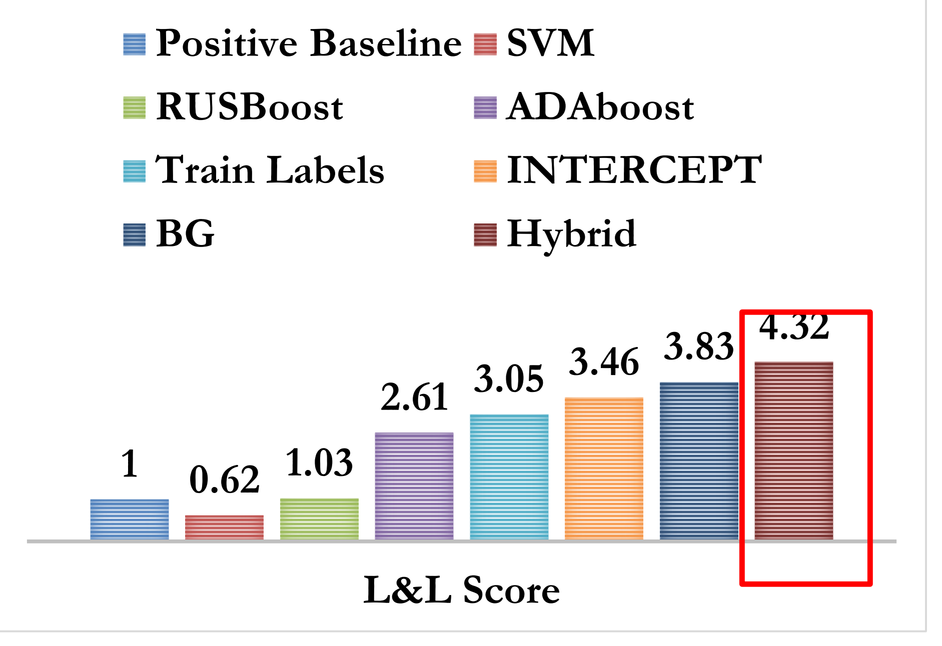


Outline

- Introduction
- Problem Statement
- Contributions
- Domain
- Dataset and Challenges
- Learn Model from Real-World Data
- Evaluation
  - Empirical
  - Field Tests
- Summary

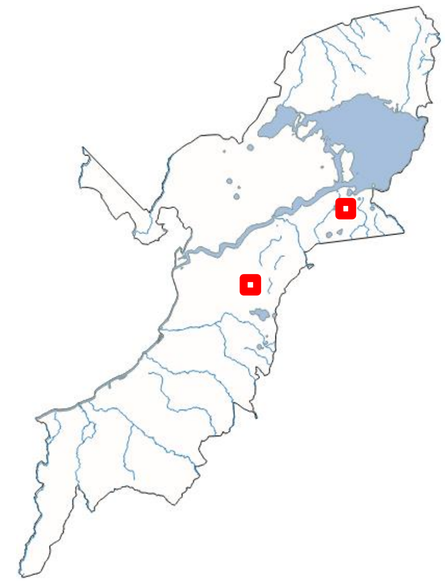
Empirical Evaluation

$$\text{L\&L Score} = \frac{Recall^2}{Probability\ of\ making\ a\ Positive\ Prediction}$$



Real-world Deployment: 1 Months of Field Tests

- Two 9-sq. km patrol areas
  - Infrequent patrols
  - Predicted hotspot
- Trespassing
  - 19 signs of litter, ashes, etc.
- Poached animals
  - 1 poached elephant
- Snaring
  - 1 active snare
  - 1 cache of 10 antelope snares
  - 1 roll of elephant snares
- Snaring hit rates
  - Outperform 91% of months

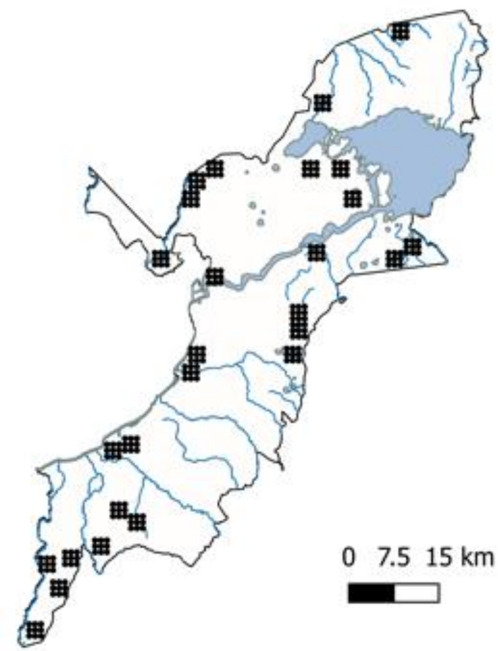


Historical Base Hit Rate	Our Hit Rate
Average: 0.73	3

Real-world Deployment: 1 Months of Field Tests

Real-world Deployment: 8 Months of Field Tests

- 27 areas, 9-sq km each
- 2 experiment groups
  HIGH: 5 areas
  LOW: 22 areas

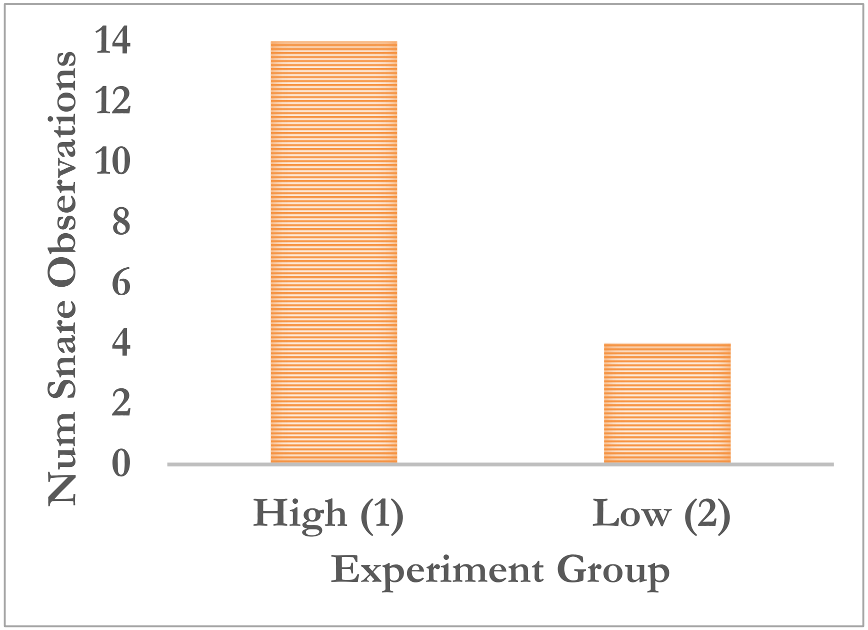


Real-world Deployment: 8 Months of Field Tests

- 27 areas, 9-sq km each

- 2 experiment groups
  HIGH: 5 areas
  LOW: 22 areas

- 8 month, 452 km patrolled in total



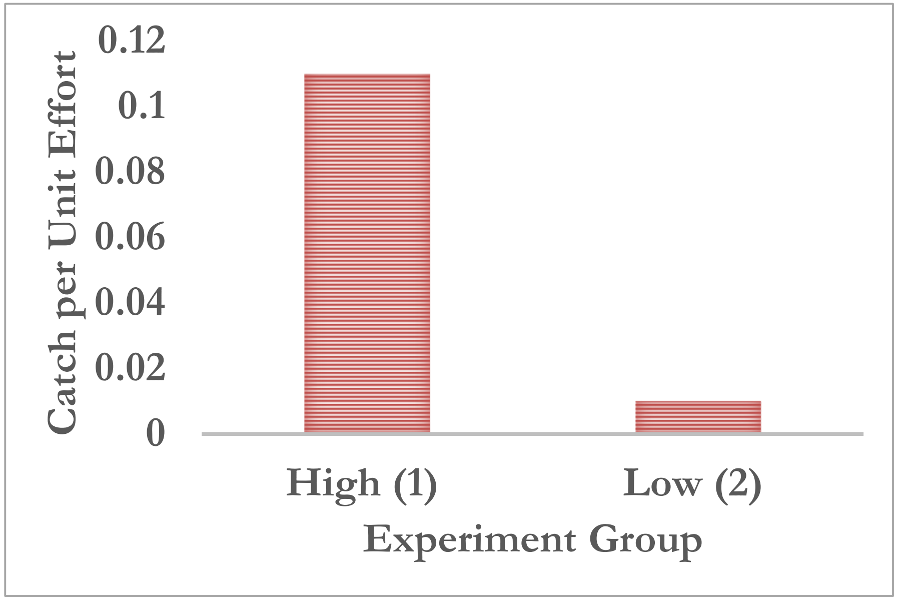
Real-world Deployment: 8 Months of Field Tests

- 27 areas, 9-sq km each

- 2 experiment groups
  HIGH: 5 areas
  LOW: 22 areas

- 8 month, 452 km patrolled in total

- Catch Per Unit Effort (CPUE)
  Unit Effort = km walked
  Historical CPUE: **0.03**



Real-world Deployment: 8 Months of Field Tests

- Statistical Significance

- Cohen's D

  Effect size: A standardized measure of the difference between two Means

$$d = \frac{experimental\ mean\ - control\ mean}{pooled\ standard\ deviation}$$

Interpretation*

0.2: Small

0.5: Medium (Visible to naked eye)

0.8: Large (Grossly perceptible)

High Group Mean (std)	Low Group Mean (std)	p-value	Cohen's d
0.12 (0.44)	0.01 (0.13)	p<0.0001	0.52

Summary

- Hybrid spatio-temporal model that outperforms other models

- First Field Test (1 months)
  - Demonstrated potential for predictive analytics in the field

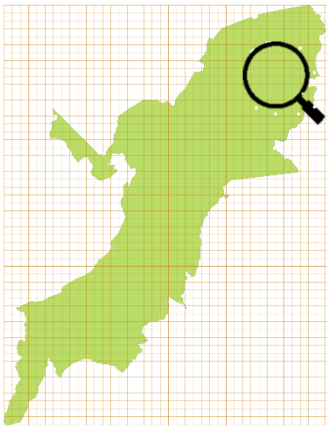
- Second Field Test (8 months)
  - First-of-its-kind field test of an ML model in this domain
  - Approximately 452 km patrolled
  - Demonstrated selectiveness of model's predictions w/ statistical significance

- Saved more animals!

Thank you

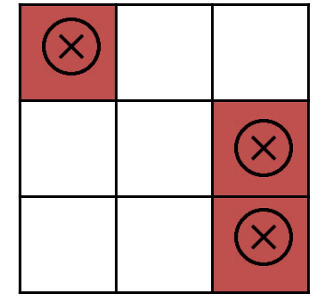
Email:
feifang@cmu.edu
sgholami@usc.edu

Challenges: Data Uncertainty

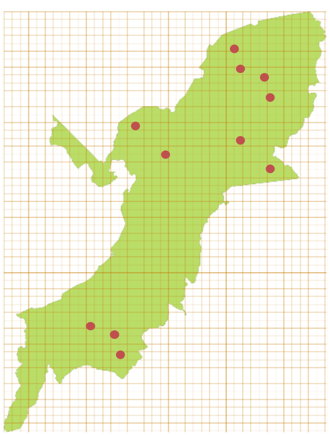


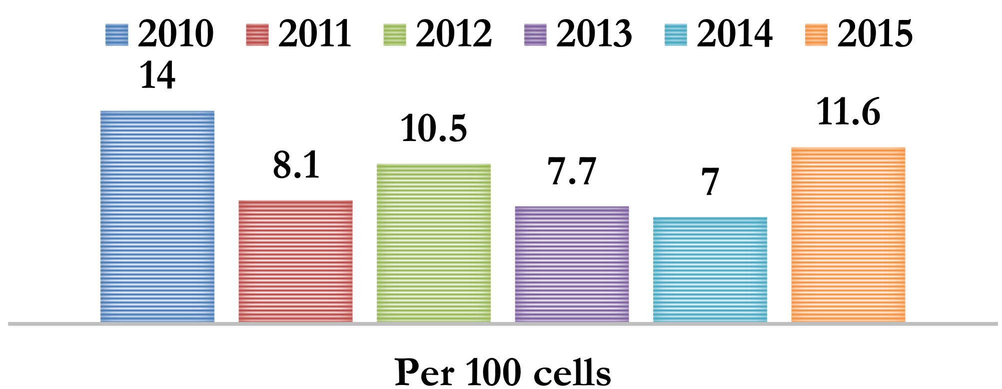
Attacked

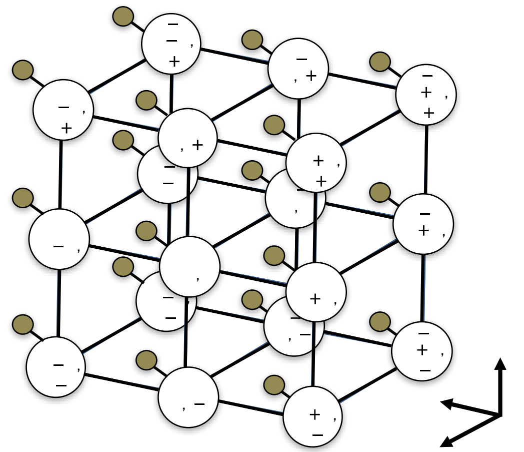
Not Attacked



Challenges: Small Number of Recorded Attacks







	Actual numbers					
	2010	**2011**	**2012**	**2013**	**2014**	**2015**
patrolled cells	1555	1708	1664	1372	1376	1539
not attacked	1338	1569	1489	1266	1280	1360
attacked	217	139	175	106	96	179
	Percentage					
patrolled cells	61.7	67.7	66.0	54.4	54.6	61.0
not attacked	86.0	91.9	89.5	92.3	93.0	88.4
attacked	14.0	8.1	10.5	7.7	7.0	11.6