Where the Wildfire Things are

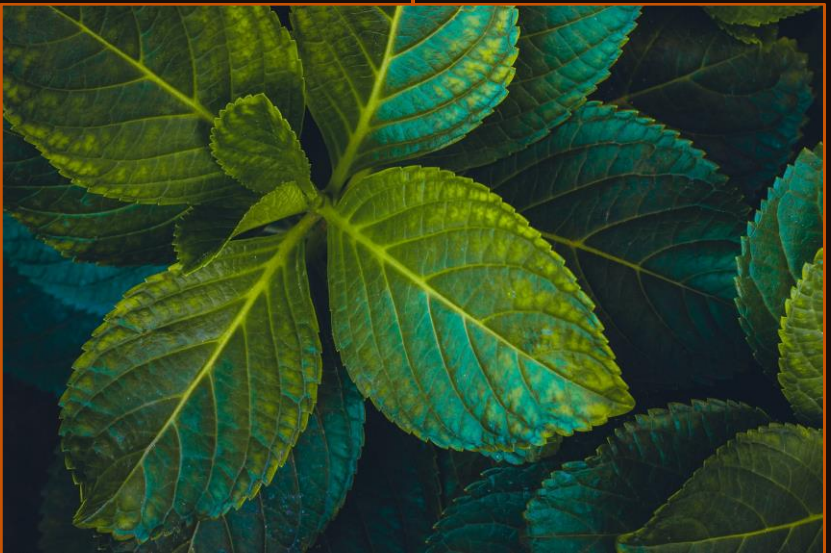
MAX MAUERMAN, MUGHIL PARI,
PEI YIN TEO, PRIYANKA SETHY

Agenda

- **Introduction**
- Data Description
- Analysis
- Conclusion

Our Research Question

What effects do
ecological patterns,
climate, and
land-use
have on the incidence and severity of wildfires?



Policy Relevance

Background

- **Extreme wildfire events** in California and Australia have highlighted the growing importance of fire management
- Globally, **climate change** is projected to increase wildfire risk and making the consequences less predictable
- We need **better statistical analysis** to identify long-term tendencies in wildfire incidence and extent

Implications

- This can inform changes in the following areas to better manage fire risk
  - **Agriculture** and **land use** policy
  - **Emergency response**
  - **Energy** and **power policy**

Agenda

Introduction

Data Description

Analysis

Conclusion

Data Description

Wildfire Incidence

- USDA **Fire Occurrence Database** (FOD) – all recorded wildfires in the US, 1992-2015
- **1.88 million** fires
- Contains data on
  - Discovery **date**
  - **Coordinates** of ignition point
  - Rough **extent** (in km^2/acres)
- Preprocessed California data for **comparison**

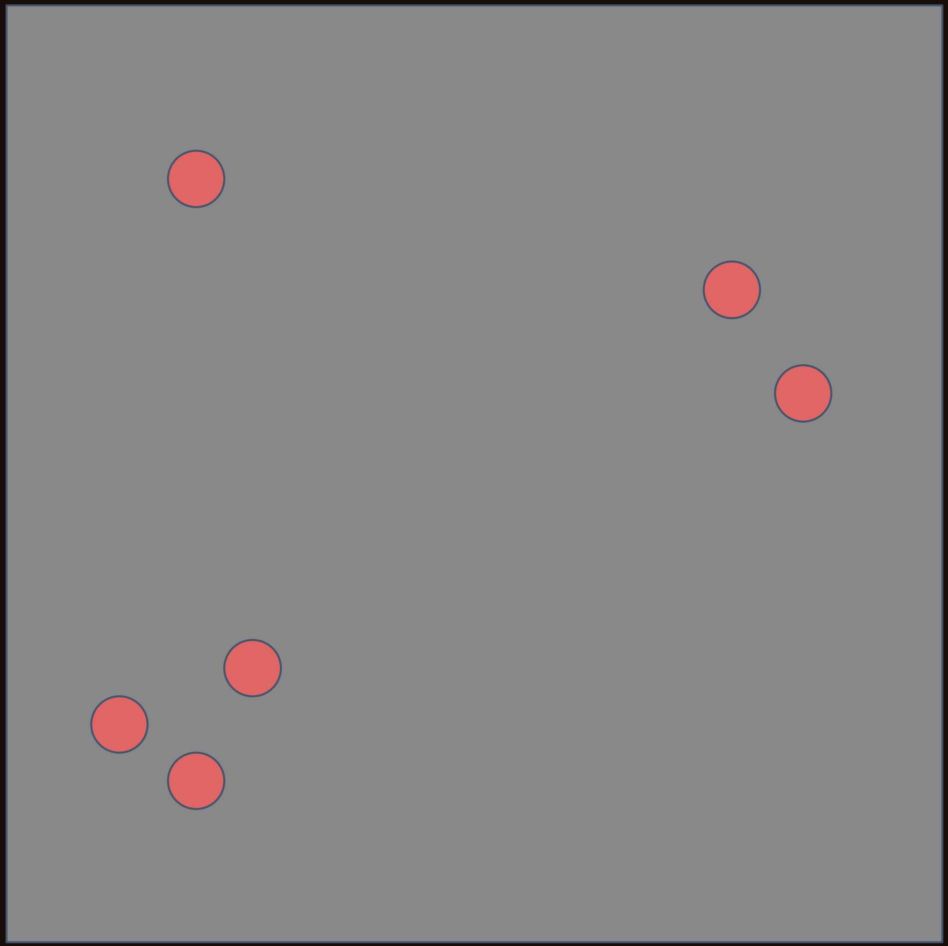
Additional Predictors

- **Climatology** – long-term monthly climate patterns (WorldClim) at a **1/12-degree resolution**
  - Temperature
  - Precipitation & humidity
  - Wind speed
- **Land use** classification (USGS National Land Cover Dataset)
  - **16** different biome and human land use types

Agenda

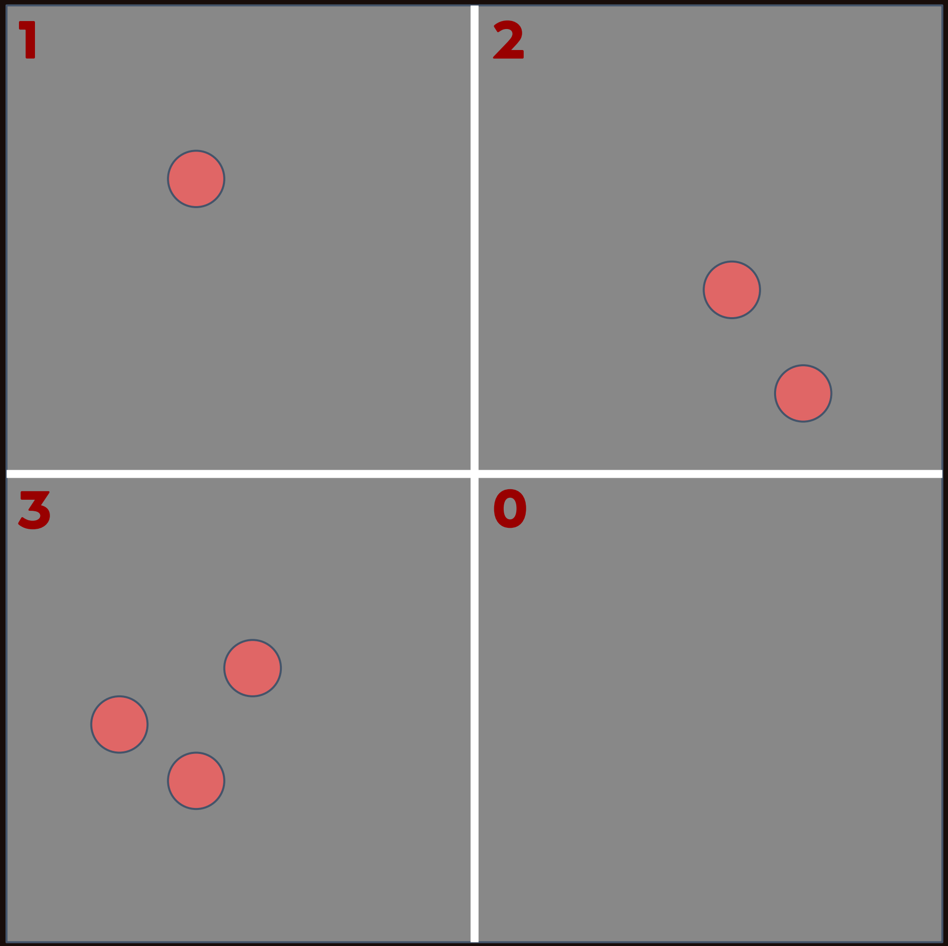
- Introduction
- Data Description
- **Analysis**
- Conclusion

Preprocessing



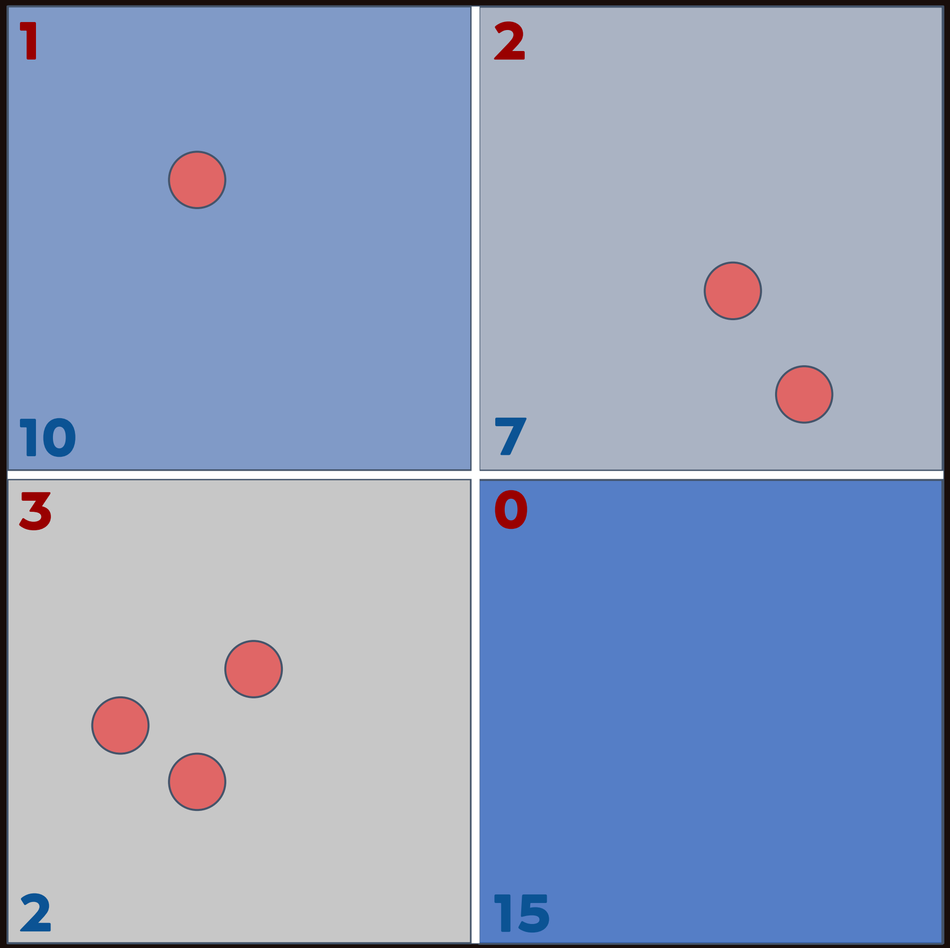
- Begin with **fire ignition** coordinates, 1992-2015

Preprocessing



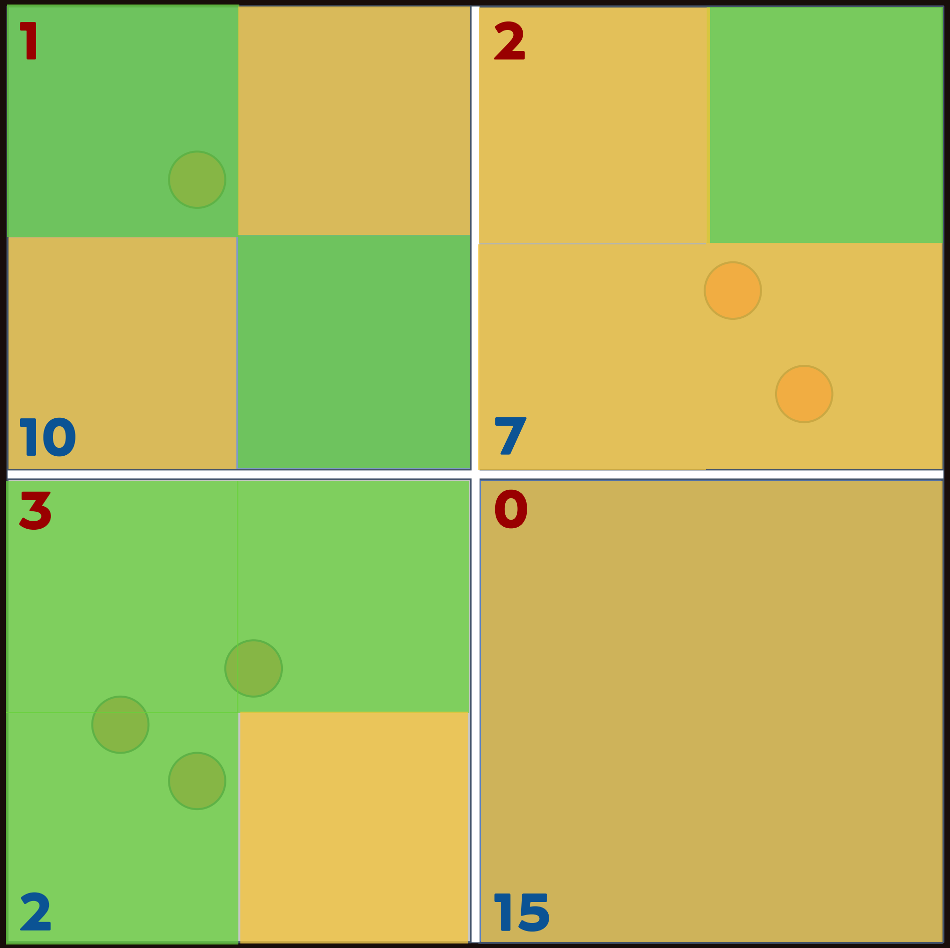
- Begin with **fire ignition** coordinates, 1992-2015
- Aggregate to **counts** in each grid cell

Preprocessing



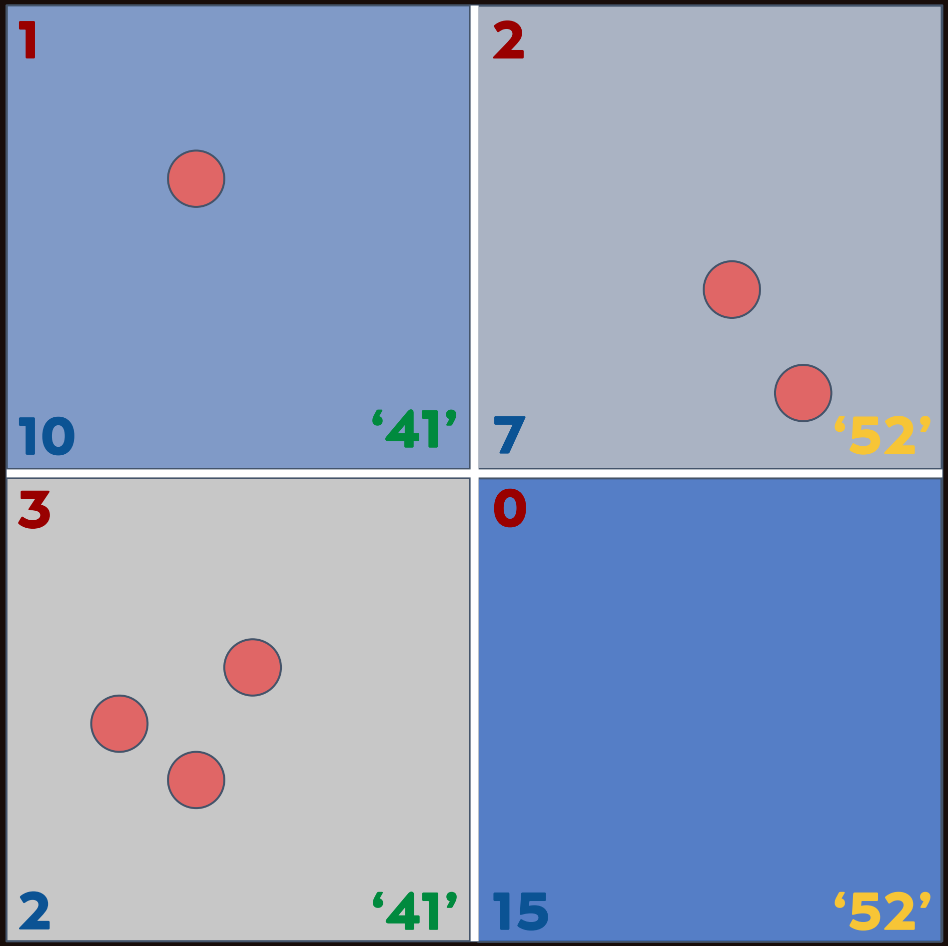
- Begin with **fire ignition** coordinates, 1992-2015
- Aggregate to **counts** in each grid cell
- Add gridded **climatology** variables (one shown)

Preprocessing



- Begin with **fire ignition** coordinates, 1992-2015
- Aggregate to **counts** in each grid cell
- Add gridded **climatology** variables (one shown)
- Add **land cover** with majority in each grid cell (2 types shown, **'41'** and **'52'**)

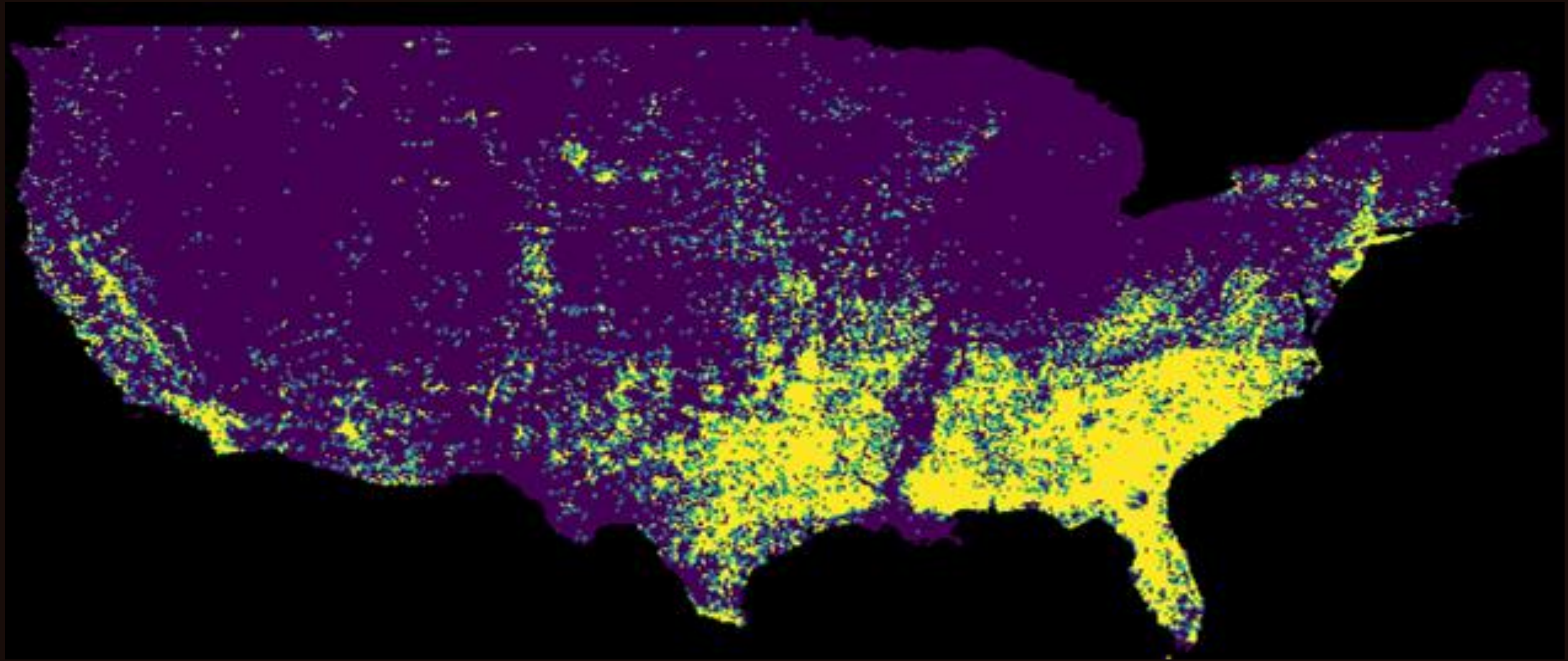
Preprocessing



- Begin with **fire ignition** coordinates, 1992-2015
- Aggregate to **counts** in each grid cell
- Add gridded **climatology** variables (one shown)
- Add **land cover** with majority in each grid cell (2 types shown, **'41'** and **'52'**)

Summary of Processed Data

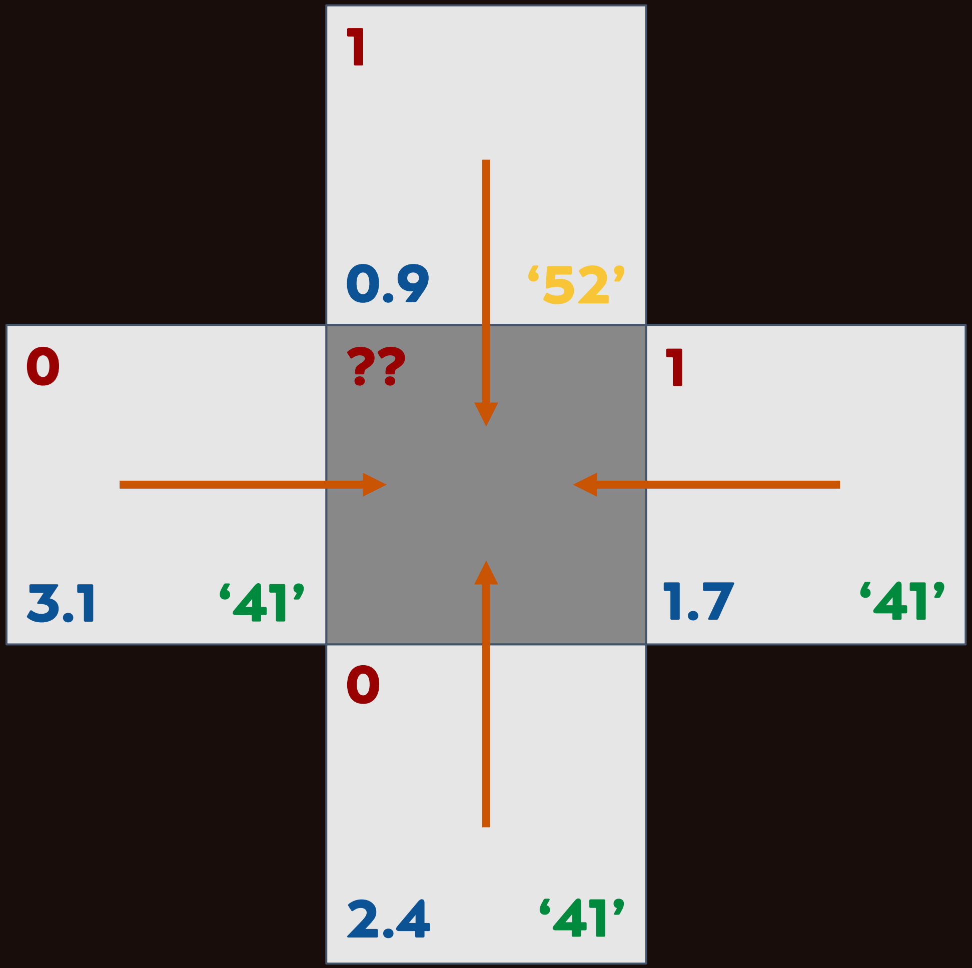
~30% of pixels have at least one fire



Model Specification

- Focusing on **fire incidence** → binary classification
  - All cells > 1 change to 1 = "a fire happened here"
- **Logistic regression** → probability that a fire happened
- Includes **spatial "lags"** → features in cells that are 1, 2, etc. cells away from center
- Features (at each lag)
  - % of cells that had fire
  - Average of 5 metrics
  - % of cells that had each of 15 land cover categories
    - Multicollinearity
  - **Total – 21 features for each lag**

Example for Lag of 1



- **Goal – P(?? = 1)**
- Fire Incidence
  - **50%** had fire
- Metrics
  - Avg. precip. = **2.025**
  - Repeat for other metrics…
- Land Use
  - 25% had **'52'**
  - 75% had **'41'**
  - Others get 0%

Findings – National

Regression Results

- Logistic regression up to lag 5 yielded similar but **mixed results**
  - 85% **accuracy** (70% base)
  - 77% **precision**
  - 69% **recall**
- **Neural nets** similar performance
- **Standardizing** had no effect
- Real estate, pastures → **higher** risk
- Open space → **lower** risk

Interpretation

- Data is **not well-separable** by any function
  - Split by acres burned?
- Data coverage → Is our FOD data **comprehensive**?
- Predictors → **Time factors**?

Findings – California

Summary

- Tested **logistic regression model** on separate California dataset (Mann et al. 2016)
- Different predictors – **moisture balance**, **topography**, and **human activity**
- Timeframe – **25 years**
  - Data for 1951 – 1975 and 1976 – 2000
  - Forecast data for 2001 – 2025
- **8%** of pixels have at least one fire

Black: Actual burned area 1976 - 2000

Findings – California

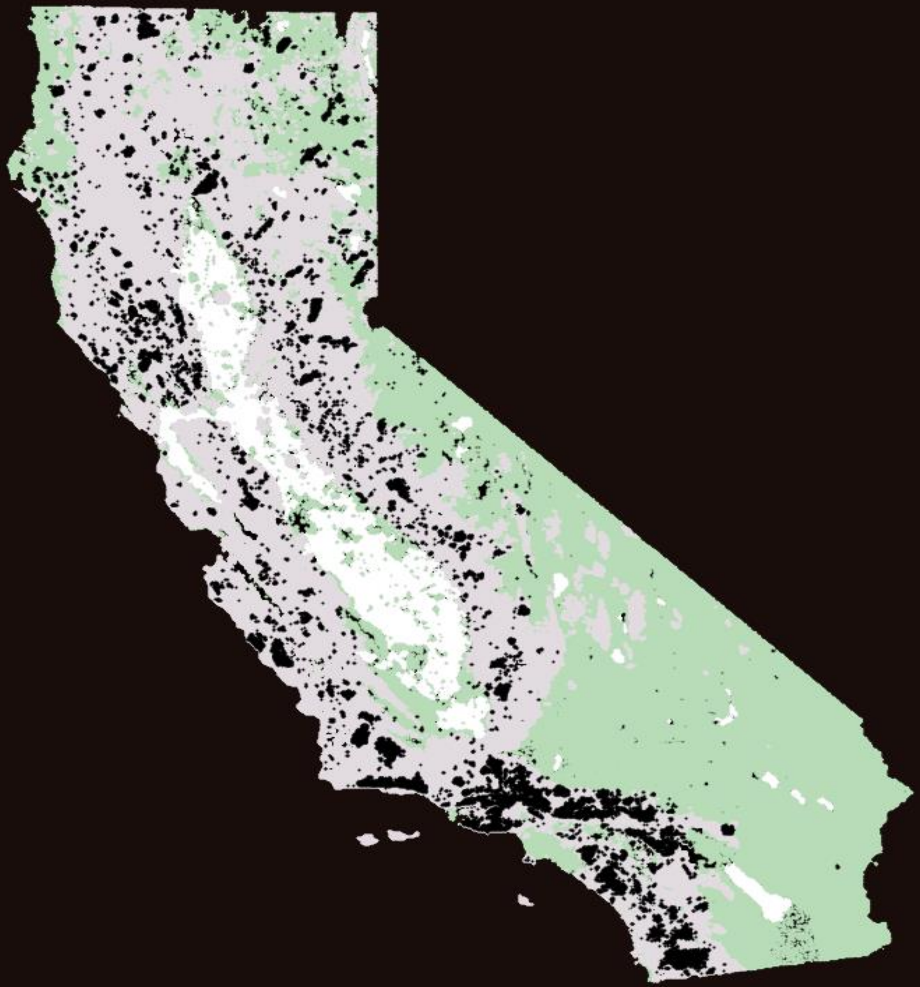
Model Performance

- **Train/test** only on 1976 – 2000: Logistic regression model yields **mixed results**
  - 62% **accuracy** (92% base)
  - 14% **precision**
  - 78% **recall**

- **Train** on 1976 – 2000, **test** on 1951 – 1975: Similar results, **better recall rate**
  - 60% **accuracy** (94% base)
  - 11% **precision**
  - 86% **recall**

- Actual – **6%** of pixels have at least one fire

- Predicted – **44%** of pixels have at least one fire



Black: Actual burned area 1951 - 1975
Grey: Predicted burned area 1951 - 1975

Findings – California

Model Performance

- **Train/test** only on 1976 – 2000: Logistic regression model yields **mixed results**
  - 62% **accuracy** (92% base)
  - 14% **precision**
  - 78% **recall**
- **Train** on 1976 – 2000, **test** on 1951 – 1975: Similar results, **better recall rate**
  - 60% **accuracy** (94% base)
  - 11% **precision**
  - 86% **recall**
- Actual – **6%** of pixels have at least one fire
- Predicted – **44%** of pixels have at least one fire

Black: Actual burned area 1951 - 1975
Grey: Predicted burned area 1951 - 1975

Findings – California

Model Performance

- 60% **accuracy**, 11% **precision**, and 86% **recall**

Modeling with different predictor sets

- Natural variables
  - **Evapotranspiration** and **climatic water deficit**
  - Slope and elevation
- Human variables
  - **Housing density** and **distance to campgrounds**
- **Performance** – Model results are similar
  - 84% - 86% recall rate

Interpretation

- Data is **not well-separable** by functions tested (including decision tree and random forest)
- High number of **false positives** – Data only for major fires (>10 acres)
- In this case, **human variables** do not improve performance
- **Future analysis** - Use lag approach tested with national dataset

Findings – California

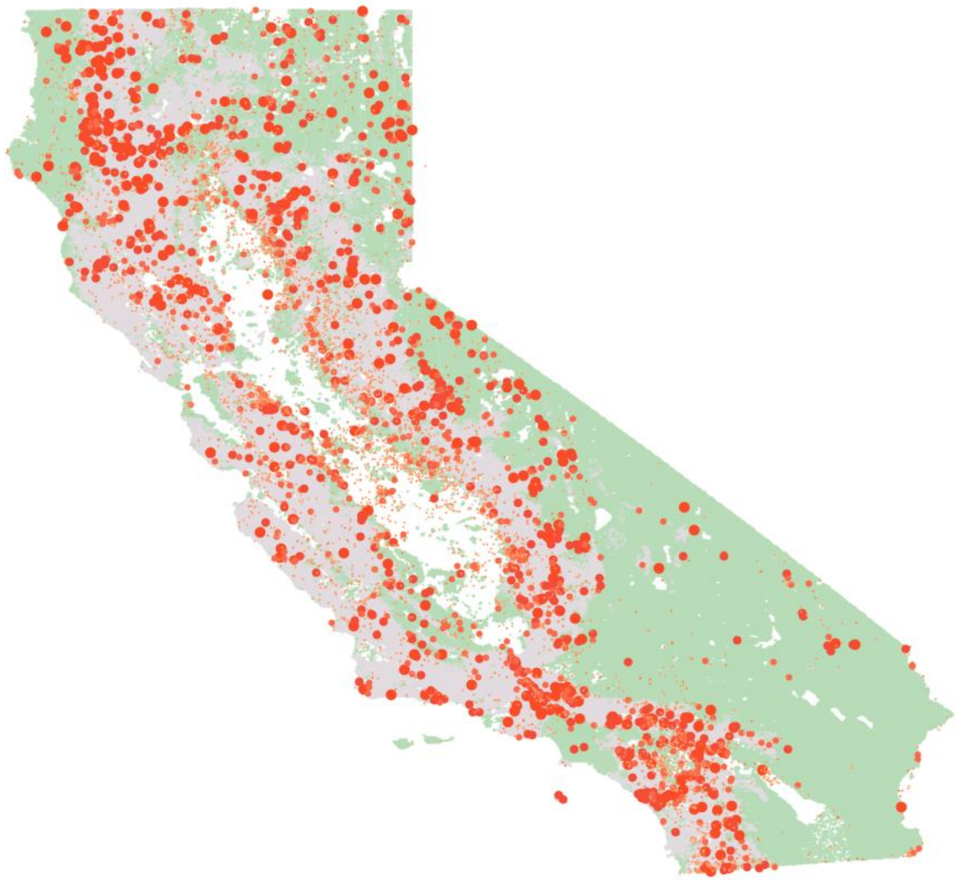
Model Forecast

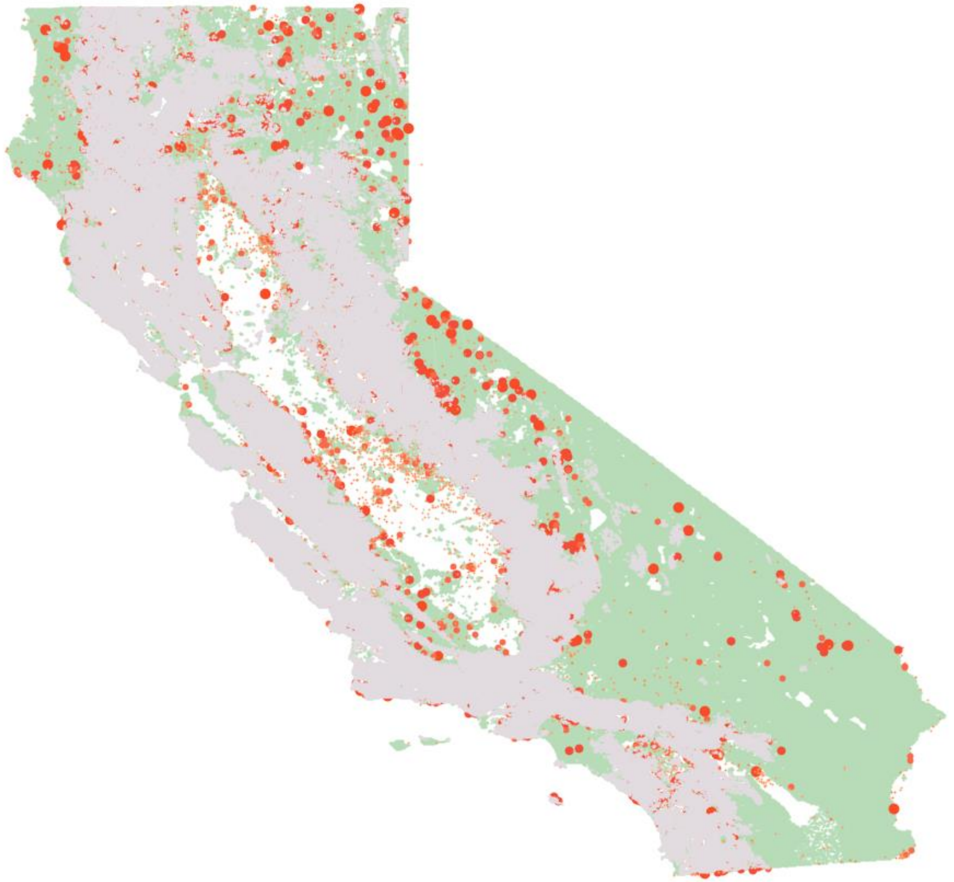
- **Forecast** data for 2001 – 2025 based on the CMIP3 Parallel Climate Model
- Compare **actual FOD fire center** with **predicted burned area**
- Matching previous trends
  - Many **actual fire ignition points** overlap with **black burned areas**
- Predicted high risk areas
  - Many **actual fires** lie within the **grey predicted areas**

Black: Actual burned area 1976 - 2000
Orange: Actual FOD fire center 2001 - 2015
Grey: Predicted burned area 2001 - 2025

Findings – California

Visualizing Actual vs. Predicted Fires





Orange: Actual FOD fire center 2001 - 2015
Grey: Predicted burned area 2001 - 2025

Agenda

Introduction

Data Description

Analysis

Conclusion

Summary of Findings

National

- **Long-term fire** risk based on WorldClim variables for 1980 – 2000
- **Logistic regression** using spatially **lagged** variables
- 85% **accuracy**, 77% **precision**, 69% **recall**

California

- **25-year** periods 1951 – 2025
- **Different** climate predictors
  - Evapotranspiration
  - Climatic water deficit
- **Human factors**
  - Housing density
- Similar performances on **actual** and **forecasted** data

Conclusion

Policy Implications

- **Human development** appears to be driving risk
- **Defensible space** planning policies?
- **Climate change** is likely to increase wildfire risk – but magnitude of estimates **vary greatly**
- Important to model **interannual variability**, not just long-term climate trends

Future Work

- **Refining data** to include only major fires (> 10 acres)
- **Presence-only models** e.g. negative binomial or maximum entropy
- Better **predictors** e.g. evapotranspiration
- **Higher frequency** climatology


Thanks for listening!
Questions?
Max Mauerman, Mughil Pari,
Pei Yin Teo, Priyanka Sethy