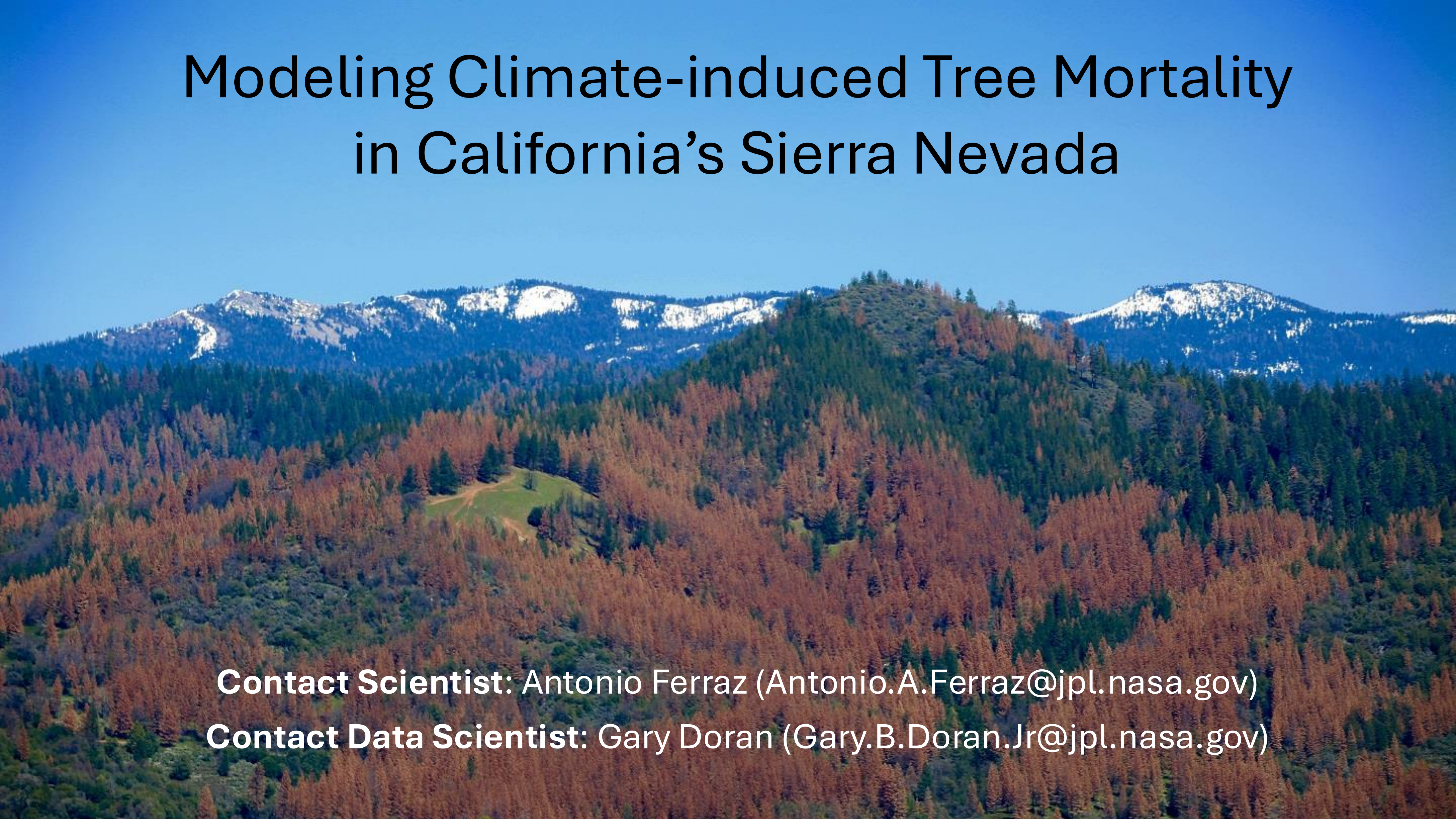
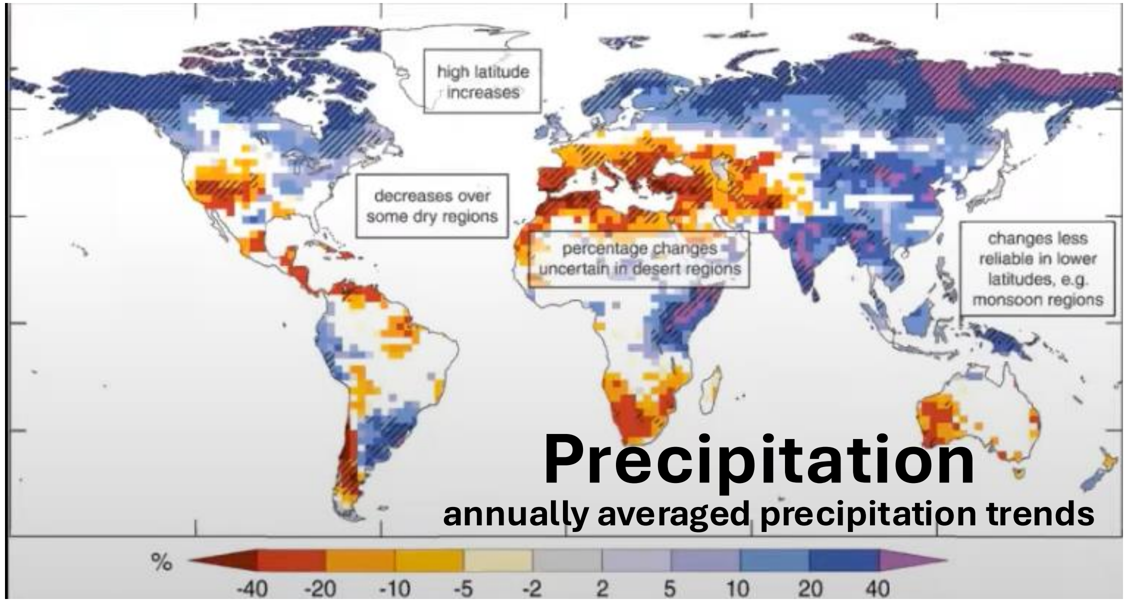
Modeling Climate-induced Tree Mortality in California's Sierra Nevada



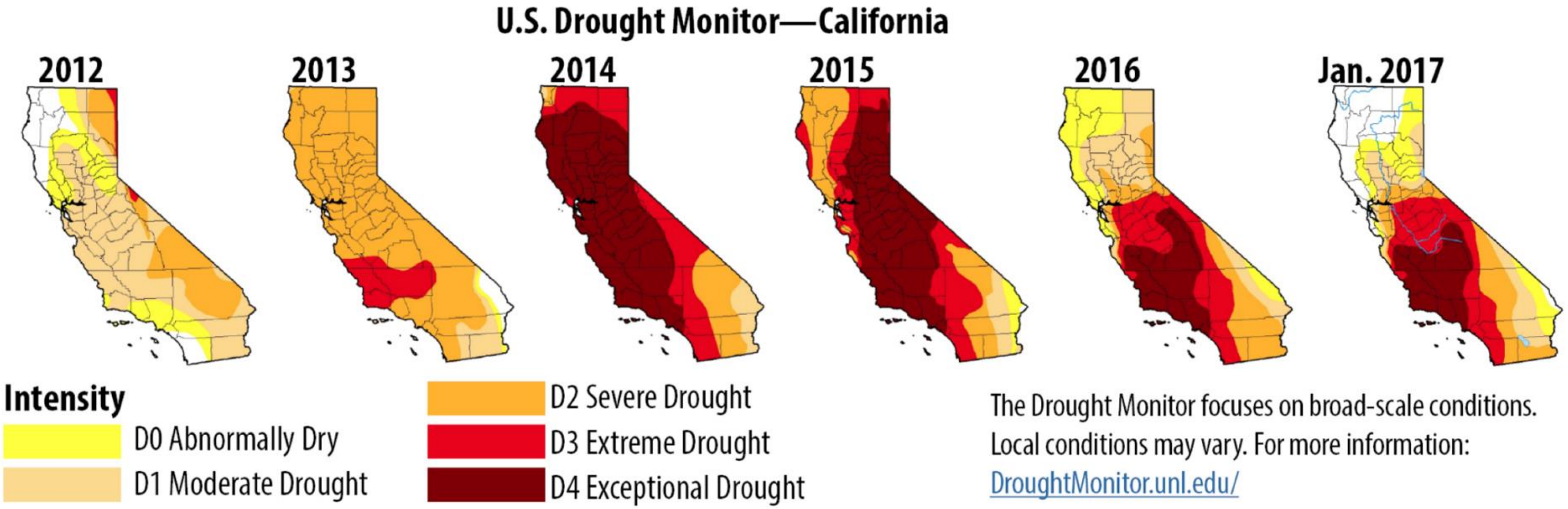
Contact Scientist: Antonio Ferraz (Antonio.A.Ferraz@jpl.nasa.gov)

Contact Data Scientist: Gary Doran (Gary.B.Doran.Jr@jpl.nasa.gov)

Climate is changing fast



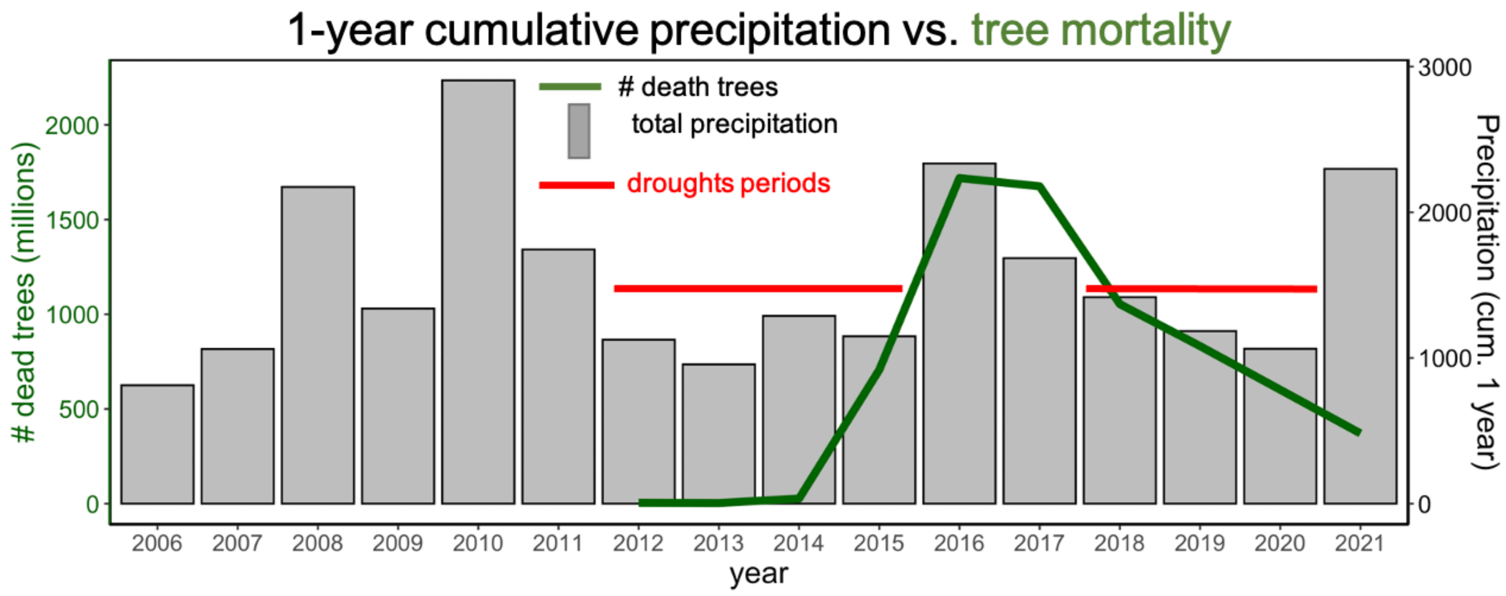
Precipitation trends indicate some drier regions are getting even drier, including the **Western US**



Climate is changing fast … so are forests

California: ~237 million trees died between 2010-2023

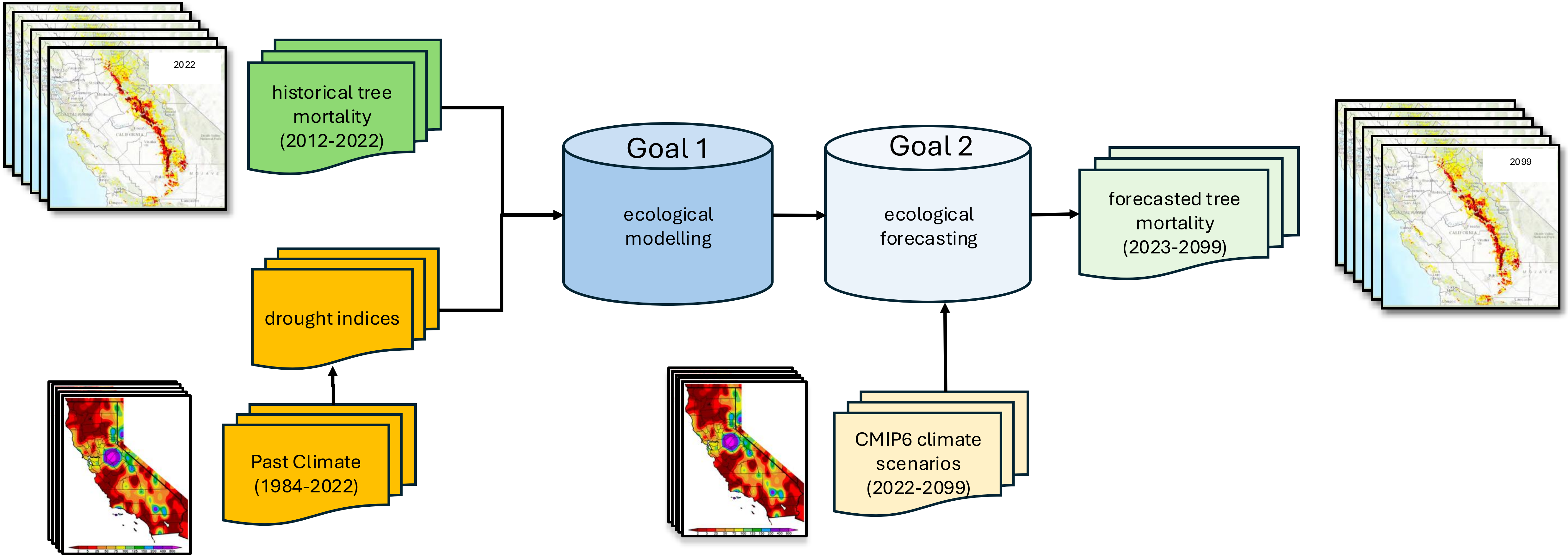
Is there a relationship between tree mortality and changes in precipitation?



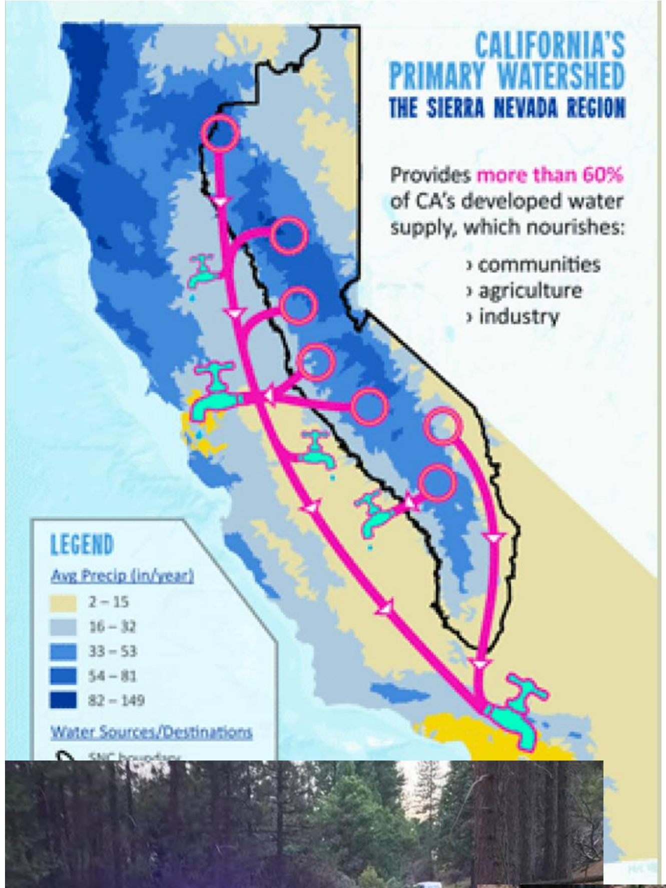
- The pulse of tree mortality is associated with consecutive drought periods
- Analyses suggest that tree mortality is delayed compared to drought periods, indicating some level of tree resilience

Goal #1: model how droughts contribute to observed conifer tree mortality in the Sierra Nevada

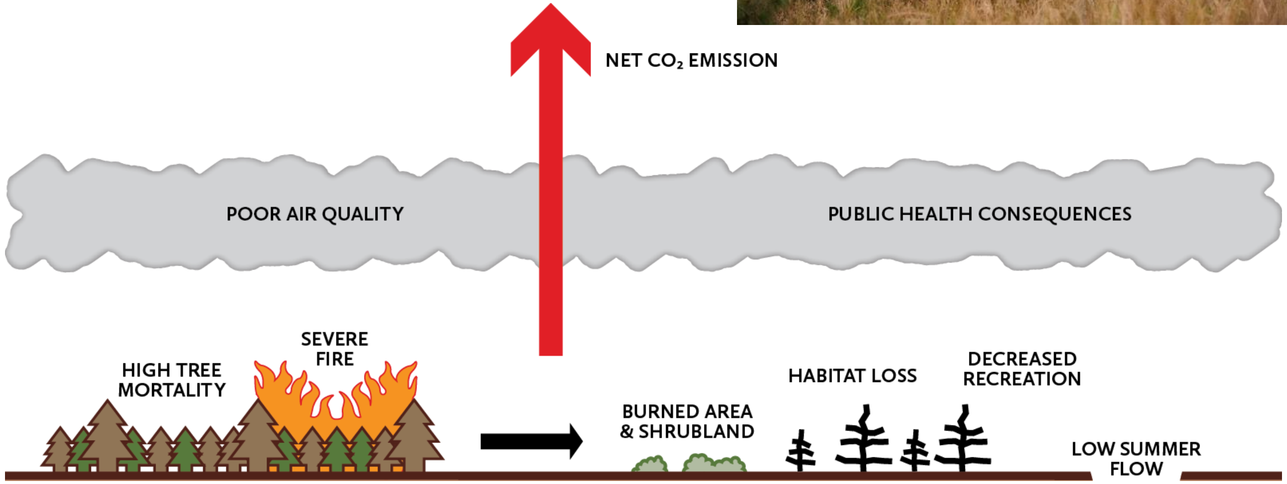
Goal #2: apply ecological models to existing climate projections to forecast tree mortality rates



Forecasting tree mortality can inform strategies to mitigate:



- Carbon emissions and climate change
- Risk if wildfires
- Poor air quality
- Water quality and water shortage
- Habitats and biodiversity loss
- Loss of valuable recreational areas
- Hazards to humans and infrastructure



Datasets

- Pre-generated, gridded drought indexes and tree mortality values
- Mortality (United States Forest Service)
  - TPA (Trees per Acre), Severity
- Drought Indexes (Basin Climate Model v8):
  - PR (Cumulative Precipitation)
  - SPI (Cumulative Precipitation, Standardized)
  - PR-ET (Cumulative Water Balance)
  - SPEI (Cumulative Water Balance, Standardized)

Data Processing Overview

Climate Data, Historical Period:
1980-2005

Training Period:
2006-2021

Projection Period:
2025-2100

Mortality Data:
2012-2021

Data Processing Overview

Climate Data, Historical Period:
1980-2005

Training Period:
2006-2021

Projection Period:
2025-2100

Mortality Data:
2012-2021

Step 1:

Climate Data, Historical Period:
1980-2005

Training Period:
2006-2021

Projection Period:
2025-2100

Historical climate data used to compute standardized precipitation indexes for future training and projection periods

Data Processing Overview

Climate Data, Historical Period:
1980-2005

Training Period:
2006-2021

Projection Period:
2025-2100

Mortality Data:
2012-2021

Step 2:

Climate and mortality data from the training period is used to train an ecological model for predicting tree mortality using climate indexes.

Training Period:
2006-2021

Mortality Data:
2012-2021

Ecological Model

Data Processing Overview

Climate Data, Historical Period:
1980-2005

Training Period:
2006-2021

Projection Period:
2025-2100

Mortality Data:
2012-2021

Step 3:

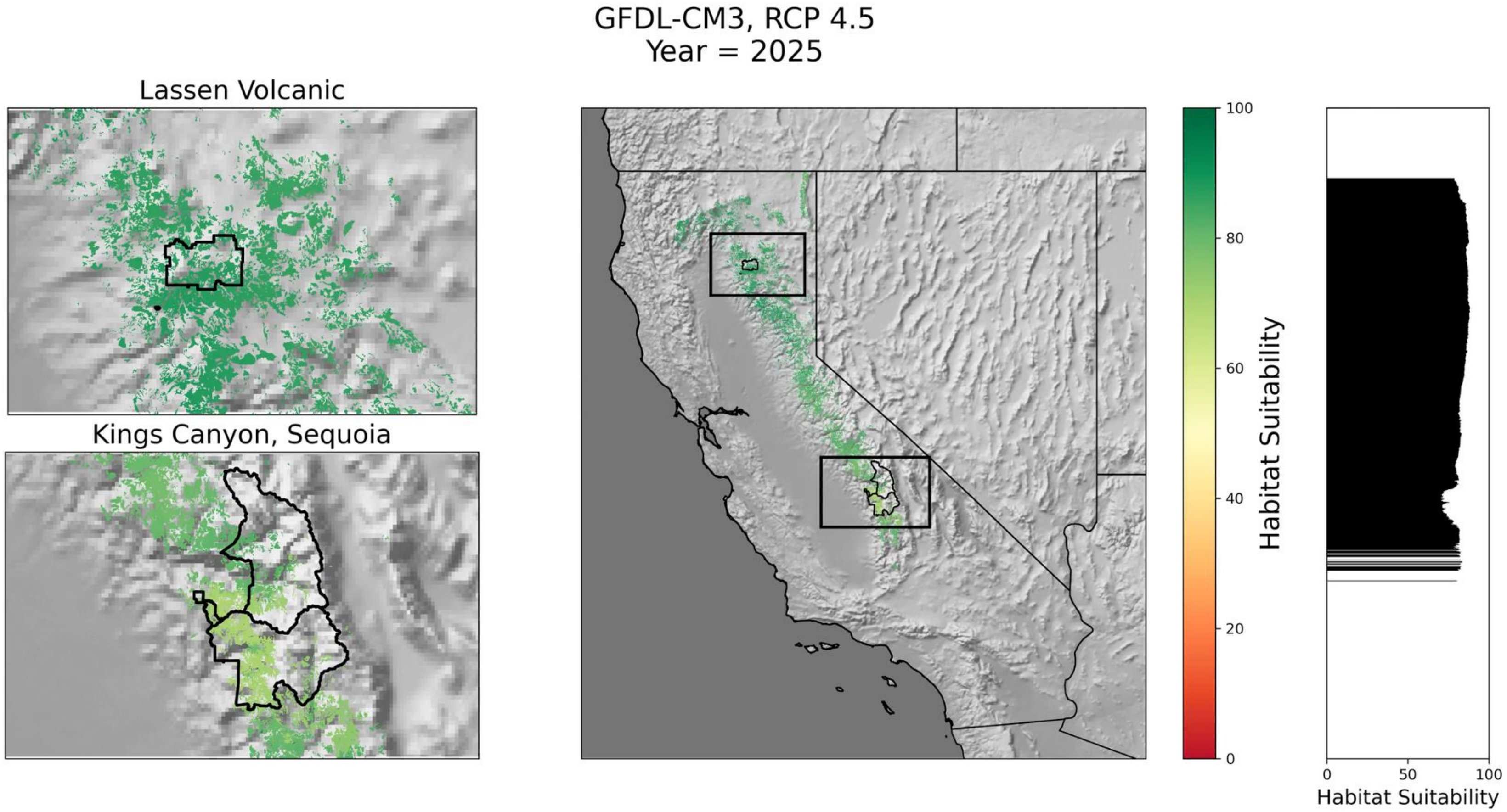
Deploy ecological model to the projected climate data to estimate future mortality.

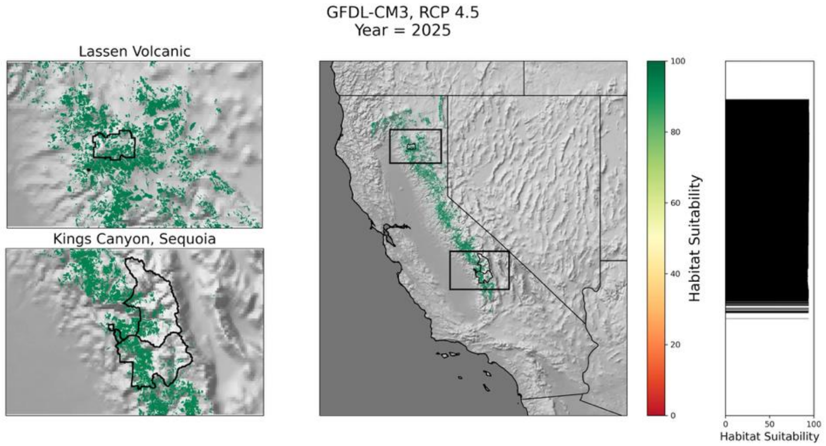
Ecological Model

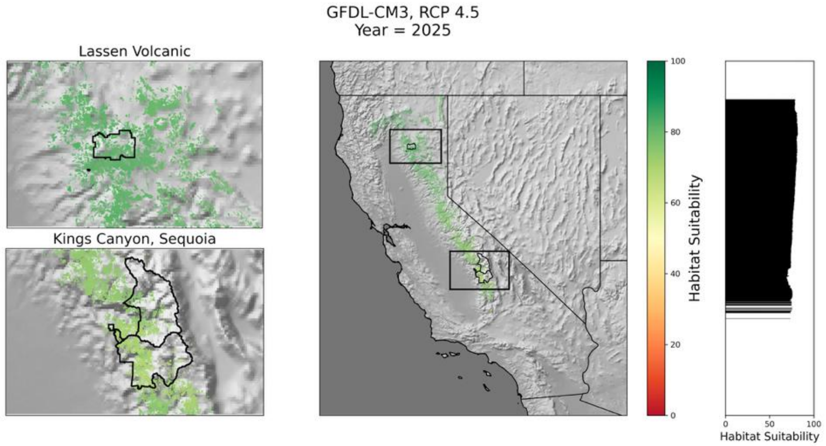
Projection Period:
2025-2100

Projected Mortality:
2025-2100

Example Mortality Projections



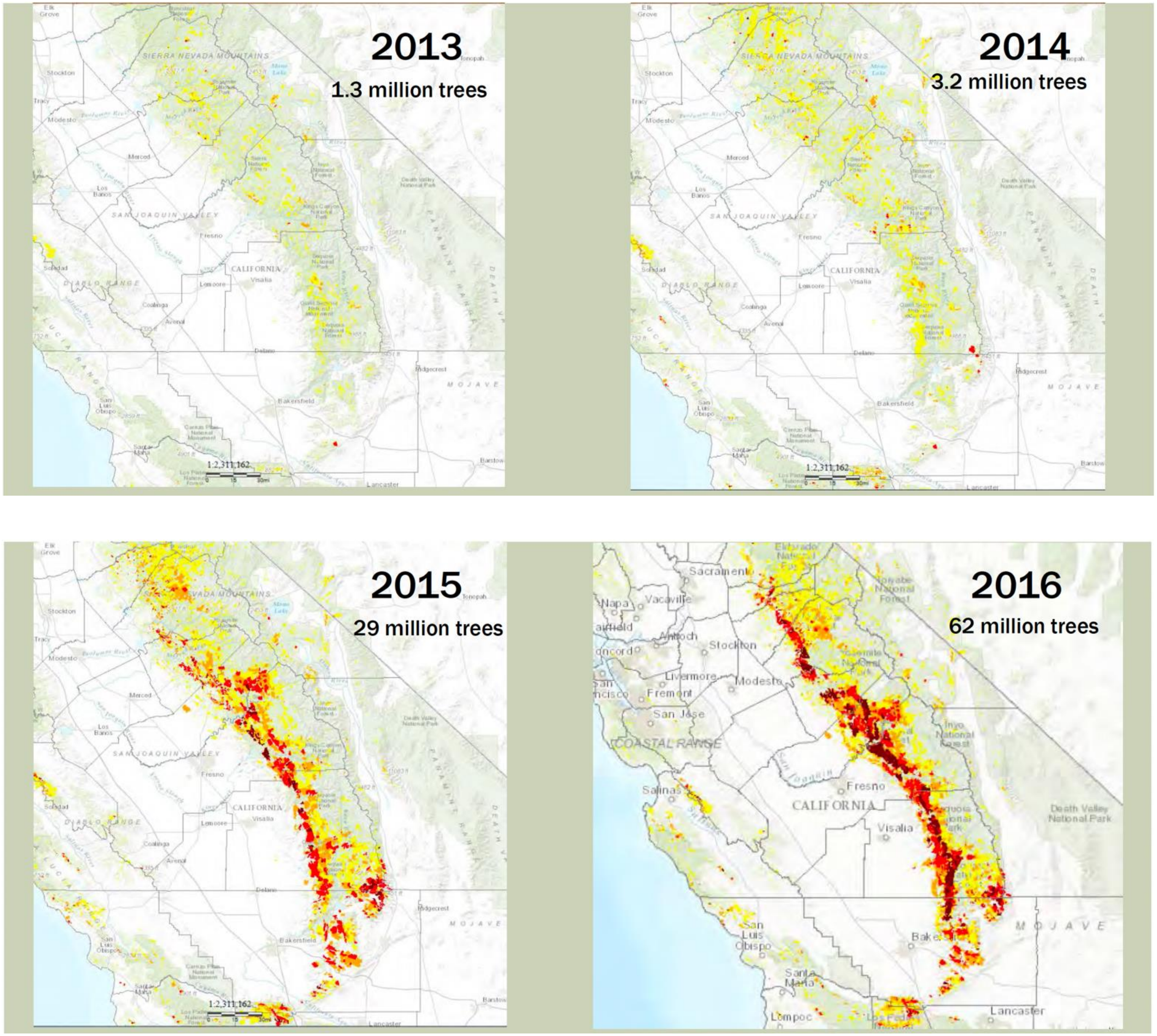




Details and Exercises

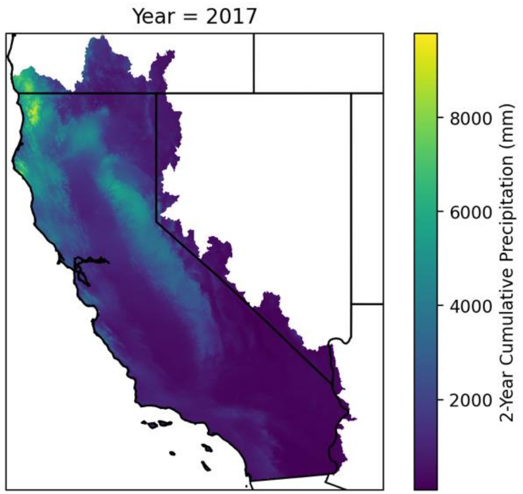
Tree mortality survey

The United States Forest Service (USFS) conducts annual tree mortality surveys, providing spatially explicit maps of tree mortality intensity given by the number of dead trees per acre.

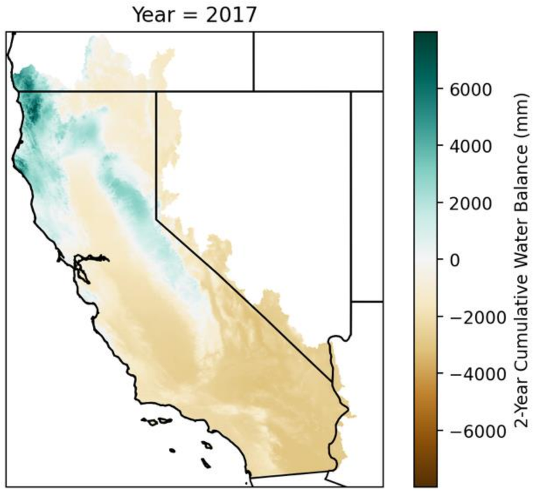


Climate Indexes

PR(n) & SPI(n):
n-Year Cumulative Precipitation
(Raw & Standardized)

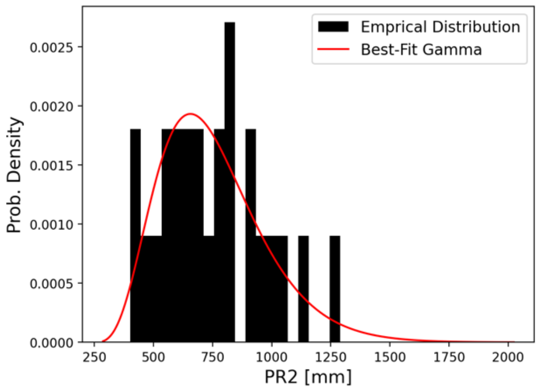


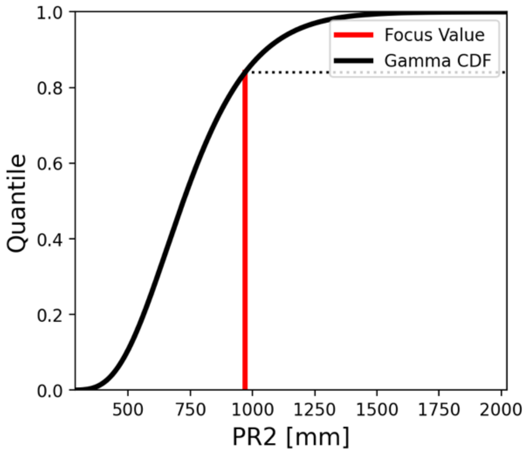
PR-ET(n) & SPEI(n):
n-Year Cumulative Water Balance
(Raw & Standardized)

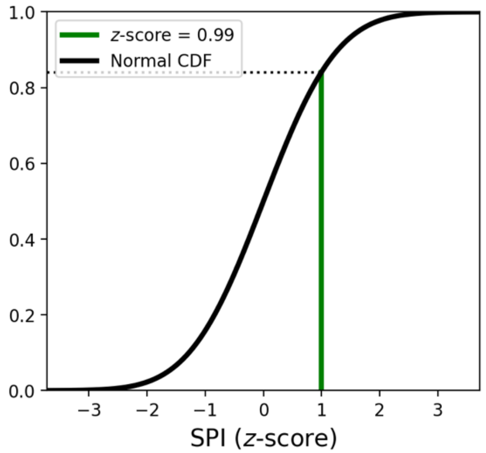


n = {1, ..., 6}

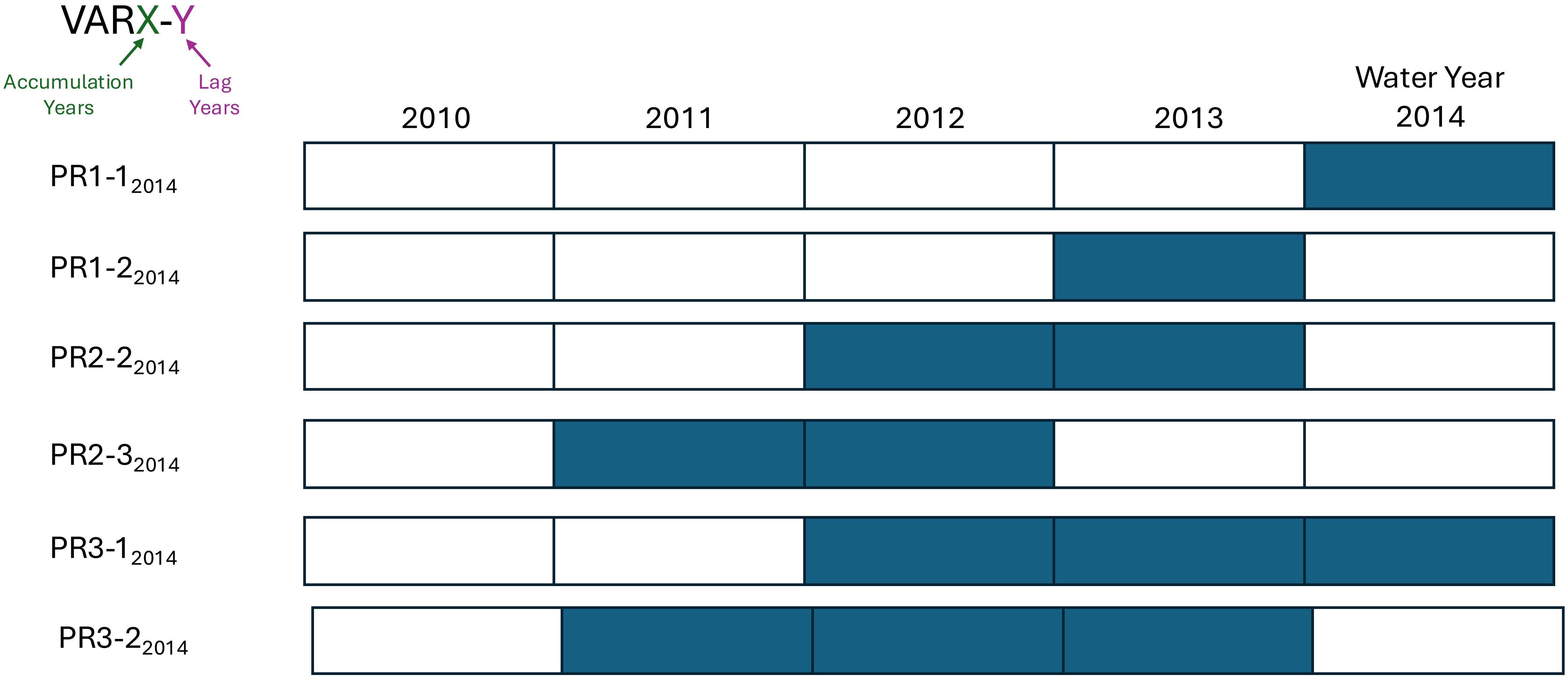
Standardization:
Find quantile of raw value in historical distribution, then map to corresponding z-score







Variable Naming Convention



Exercise 1: Comparing Gridded and Random Folds

Given the comparison between the results using two different ways of splitting data into folds, reflect on the following questions:

1. Do you observe a difference between the mean squared error (MSE) of the two approaches? Which approach performs better? Is the difference statistically significant at an $\alpha = 0.05$ significance level (i.e., is the *p*-value less than 0.05)? Is the difference consistent across multiple years?
2. A key assumption made by statistical machine learning models that the training and test sets are statistically independent and identically distributed. If this assumption is violated (e.g., if examples in the training and testing set are correlated, or if the training and testing sets are drawn from different underlying probability distributions), then performance of the model on the test set may not be reflective of the ability of the model to generalize to new, independently sampled data. What might cause the assumption of independent to be violated in this case, and how might it affect each approach for splitting data into folds? If the assumption is violated, is it more likely to lead to over-estimates or under-estimates of the ability of the model to generalize?
3. Given the considerations above, is there a recommended approach to use for model evaluation for this particular problem?

Exercise 2: Regression Model Comparison

Using the provided code as an example, implement and evaluate at least 3 different regression models from scikit-learn, and at least 1 additional "baseline" model (e.g., simple linear regression). Which models exhibit the most success in predicting mortality in the held-out year?

1. What factors might explain the difference in model performance across years?
2. There is a phenomenon in statistical machine learning called the "bias-variance tradeoff", in which more complex models are better able to fit arbitrary relationships between the input variables and target values (i.e., they exhibit less "bias" towards a particular, for example linear, relationship), but this also leads to more "variance" in performance due to over-fitting to the particular random sample of data in the training set. Some models have parameters that control the complexity of the relationships that can be learned, such as the “max_depth” parameter of the random forest. Do you observe a bias-variance tradeoff as you explore different models and parameters for this problem? Provide some examples of instances where you observe the effect and how you mitigated it.

Exercise 3: Exploring Spatial Biases in Predictions

1. Do you observe any spatial bias in prediction errors?
2. If so, does the degree or pattern of bias change with model type and parameterization?
3. Does the pattern of bias change across years?
4. What factors might induce spatial bias in this dataset? How can these be mitigated?

Exercise 4: Generating and Analyzing Projections

1. What are the general patterns you observe in the spatial and temporal distribution of habitat suitability loss?
2. Are these patterns consistent across different model types? If not, what are some differences you observe?
3. If you were advising forest managers planning reforestation efforts in the Sierra Nevada, how might you use this analysis to inform their efforts?