PRINCETON UNIVERSITY

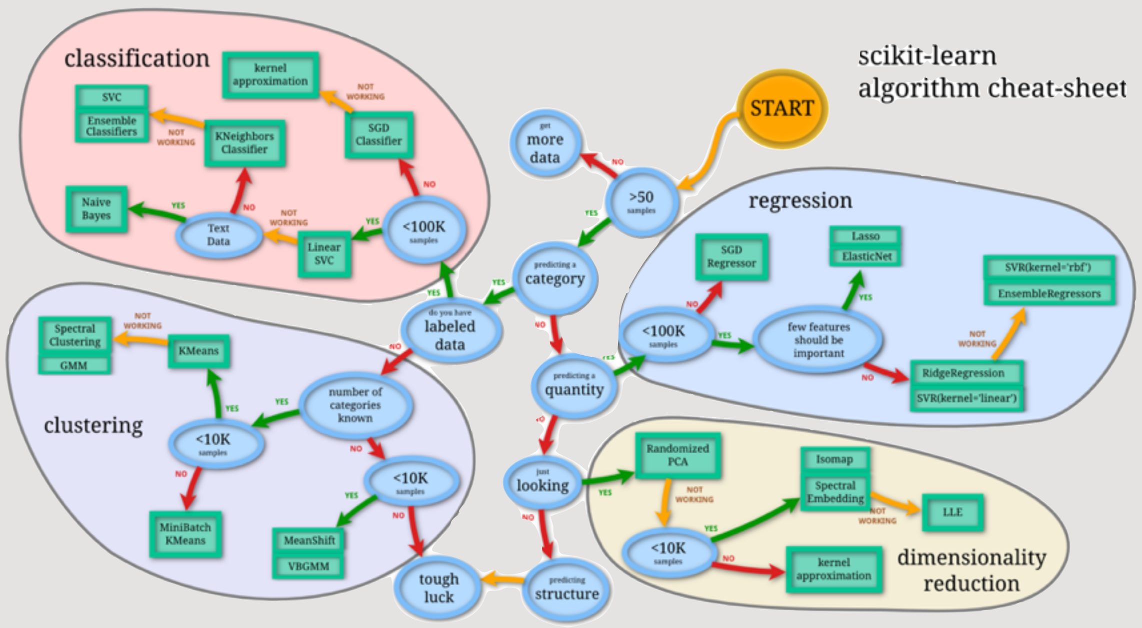
# Statistical Precipitation Downscaling Using Random Forests

*- Synthetic experiments over Southeast United States -*

**Xiaogang He**[1], Nathaniel W. Chaney[1,2], Justin Sheffield[1], Ming Pan[1], Eric F. Wood[1]

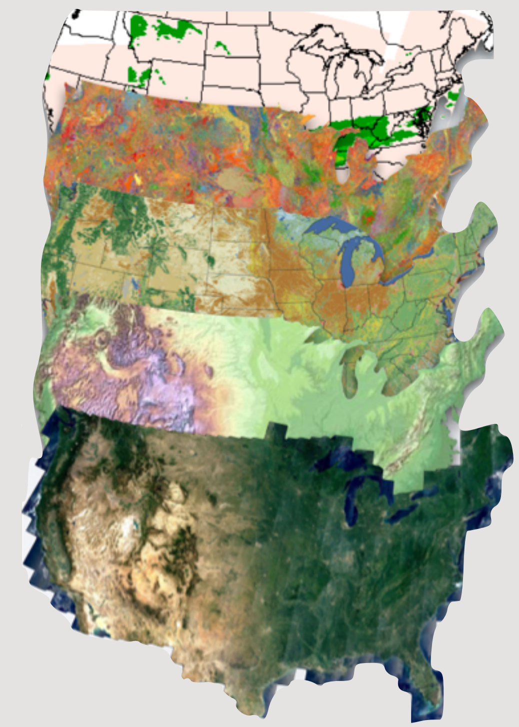
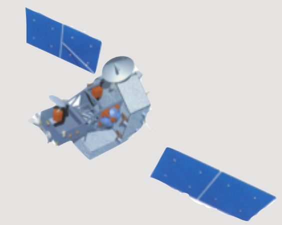
*[1] Department of Civil and Environmental Engineering, Princeton University*
*[2] Program of Atmospheric and Oceanic Sciences, Princeton University*

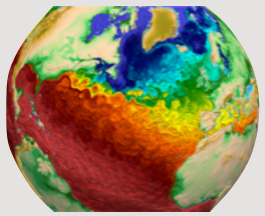
## Machine learning approach




Explore and construct complicated relationships

## Big data



$\Downarrow$


Precipitation spatial-temporal structures??

**Hyperresolution** precipitation downscaling??

# Regression model

$$P_d = f\,(P_c,\ A,\ T,\ S,\ V,\ L,\ D,\ \ldots) + \mathcal{E}$$

Downscaled Precipitation (Posterior)

Coarse Resolution Precipitation & Adjacent 4 Grid Cells

Atmosphere (Wind, Temperature, CAPE, Humidity, Surface Pressure)

Topography (Mean & Std of DEM, Slope, Aspect)

Soil texture

Vegetation Type

Location (Latitude, Longitude)

Day of Year

Errors

## Single learner

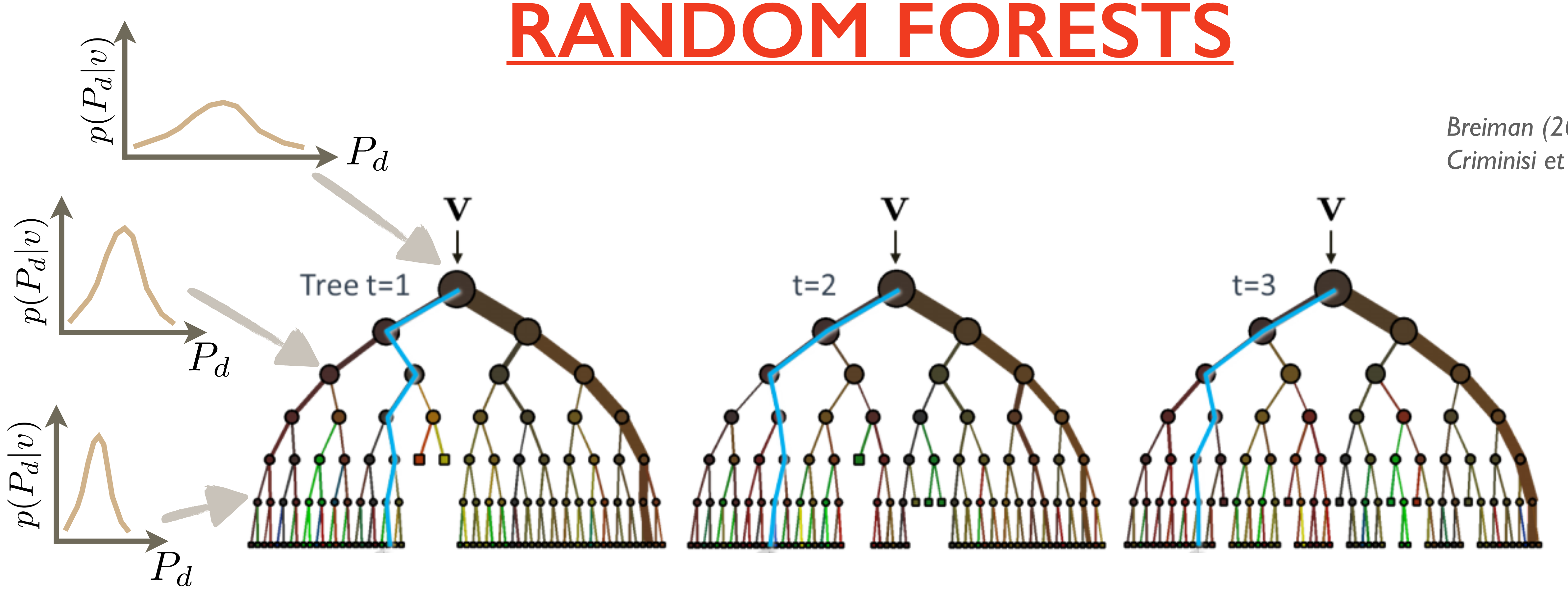
**vs**

## Group of learners

Ensemble methods → Reduce overfitting

# RANDOM FORESTS

Breiman (2001)
Criminisi et al., (2011)




**Regression forest posterior:** $p(P_d|v) = \frac{1}{T}\sum_{t=1}^{T} p_t(P_d|v)$

**v**: covariates
***P_d***: predictand (downscaled prec)
***p_t(P_d|v)***: individual tree posterior

# Step 1:
## Generate synthetic covariates

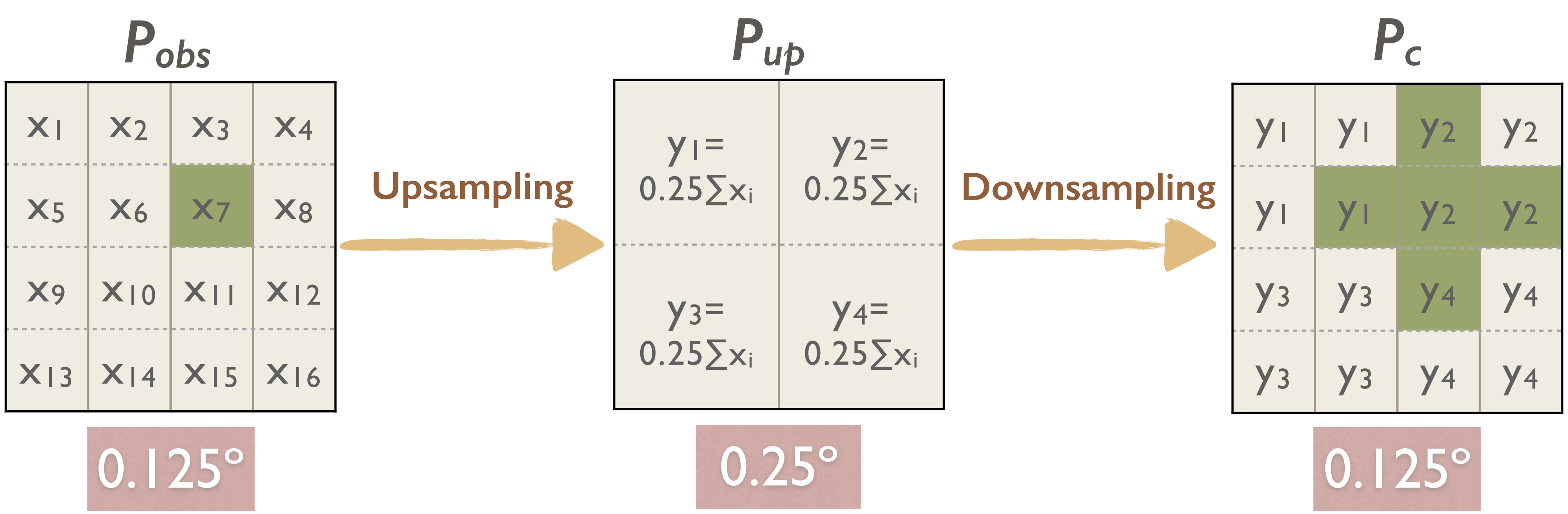
# Synthetic Experiment Design

| EXP | $P_C$ | A |
|---|---|---|
| $P_{0.25}A_{0.125}$ | 0.25° | 0.125° |
| $P_{0.25}A_{0.25}$ | 0.25° | 0.25° |
| $P_{0.5}A_{0.125}$ | 0.5° | 0.125° |
| $P_{0.5}A_{0.5}$ | 0.5° | 0.5° |
| $P_{1}A_{0.125}$ | 1° | 0.125° |
| $P_{1}A_{1}$ | 1° | 1° |

**Data:** NLDAS2
North American Land Data Assimilation System 2

**Period:** 2011.06.01—2011.08.31 (Hourly)

**Domain:** Region: Southeast United States
Grid: 80×72 (0.125°)

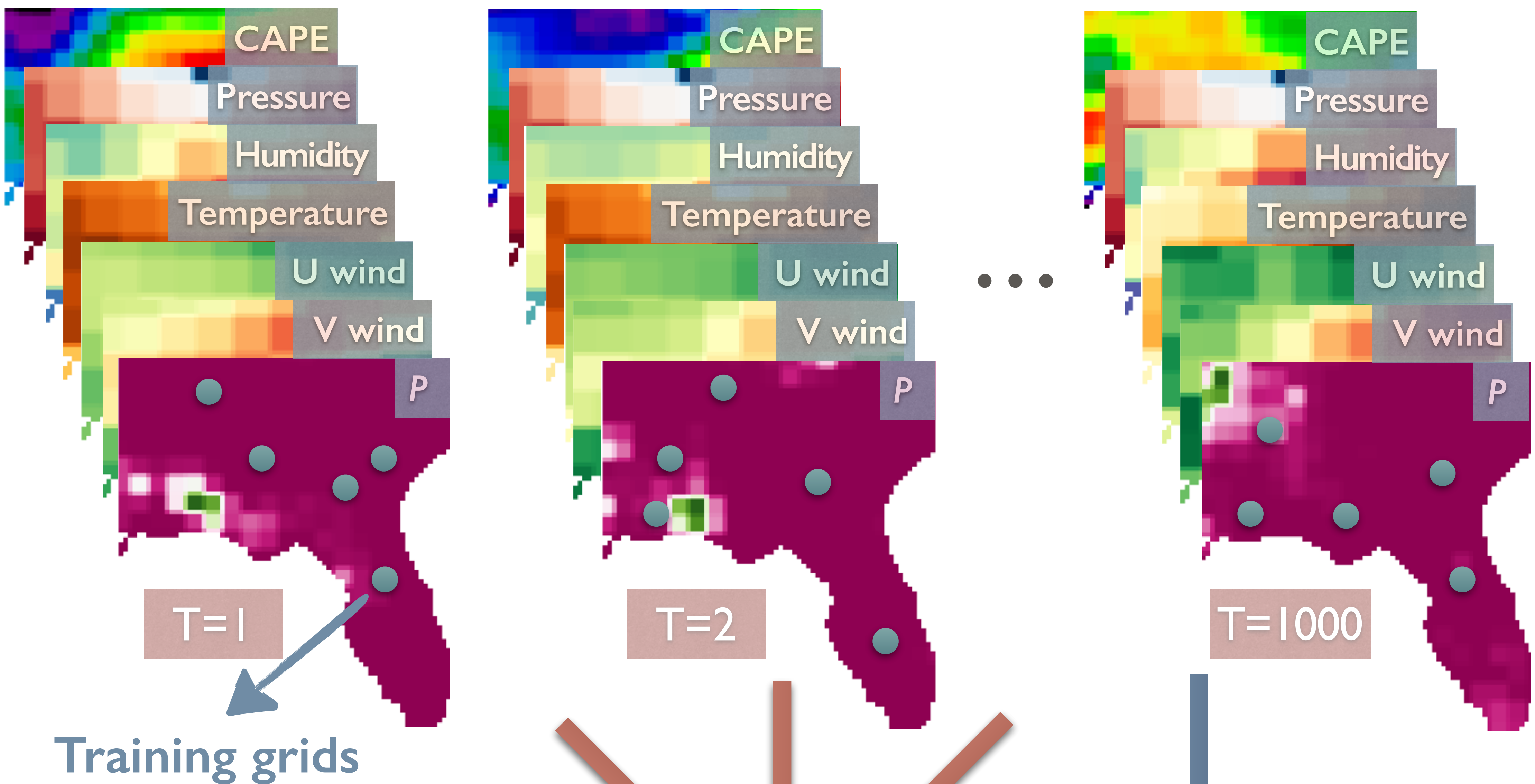



Adjacent grid cells are also considered
Same procedures for atmospheric covariates

# Step 2: Train and test Random Forests

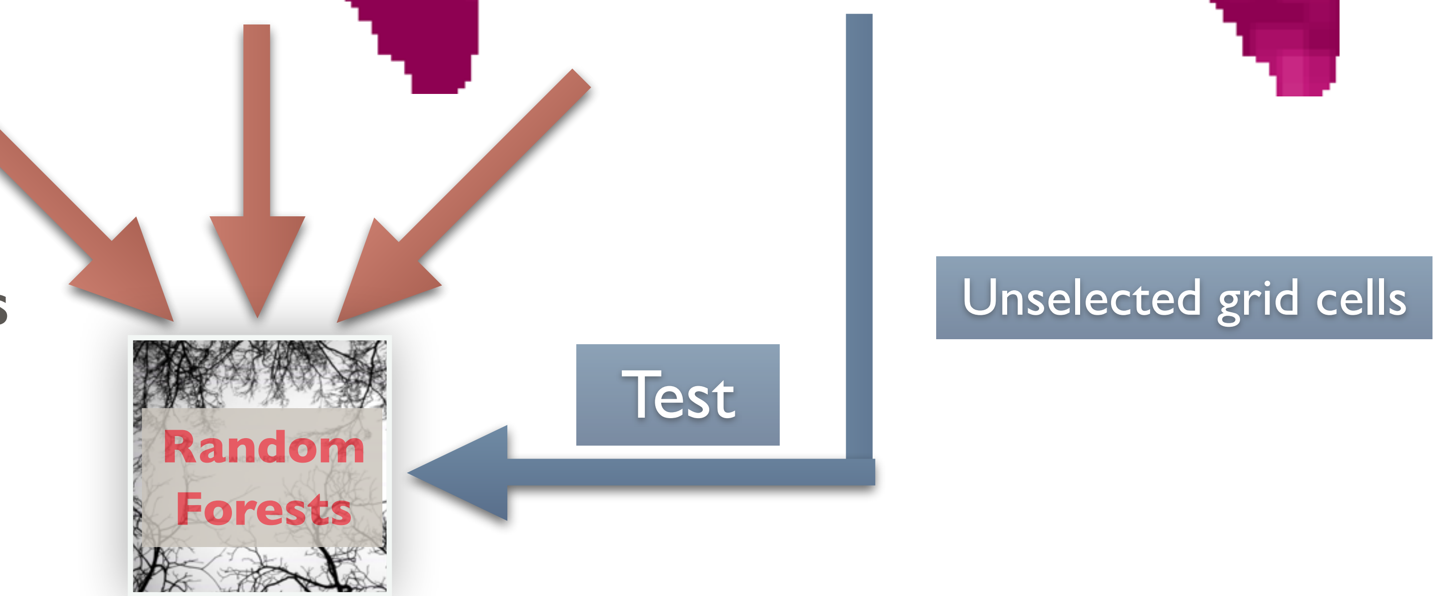
# Synthetic Experiment Design

Coarse $P$ + Dynamic/Fixed Covariates ⇛ Downscaled $P$




Random sample 10% of total grid cells from 1000 images for 17 covariates

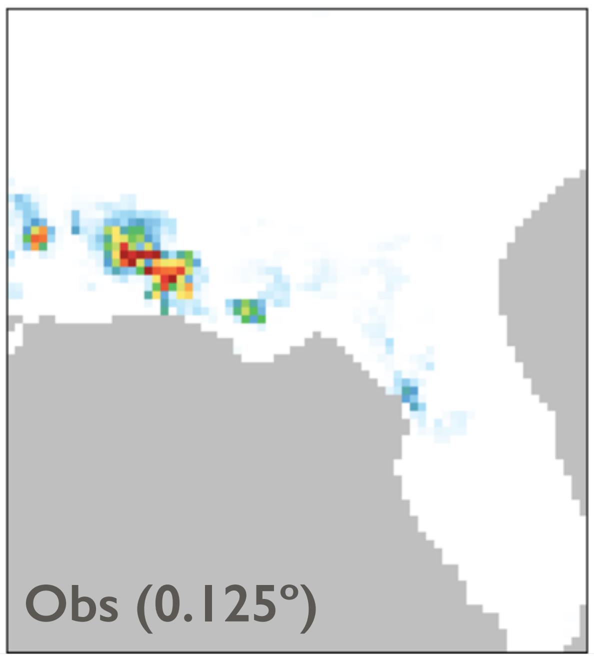
Training samples: 625000 grid cells



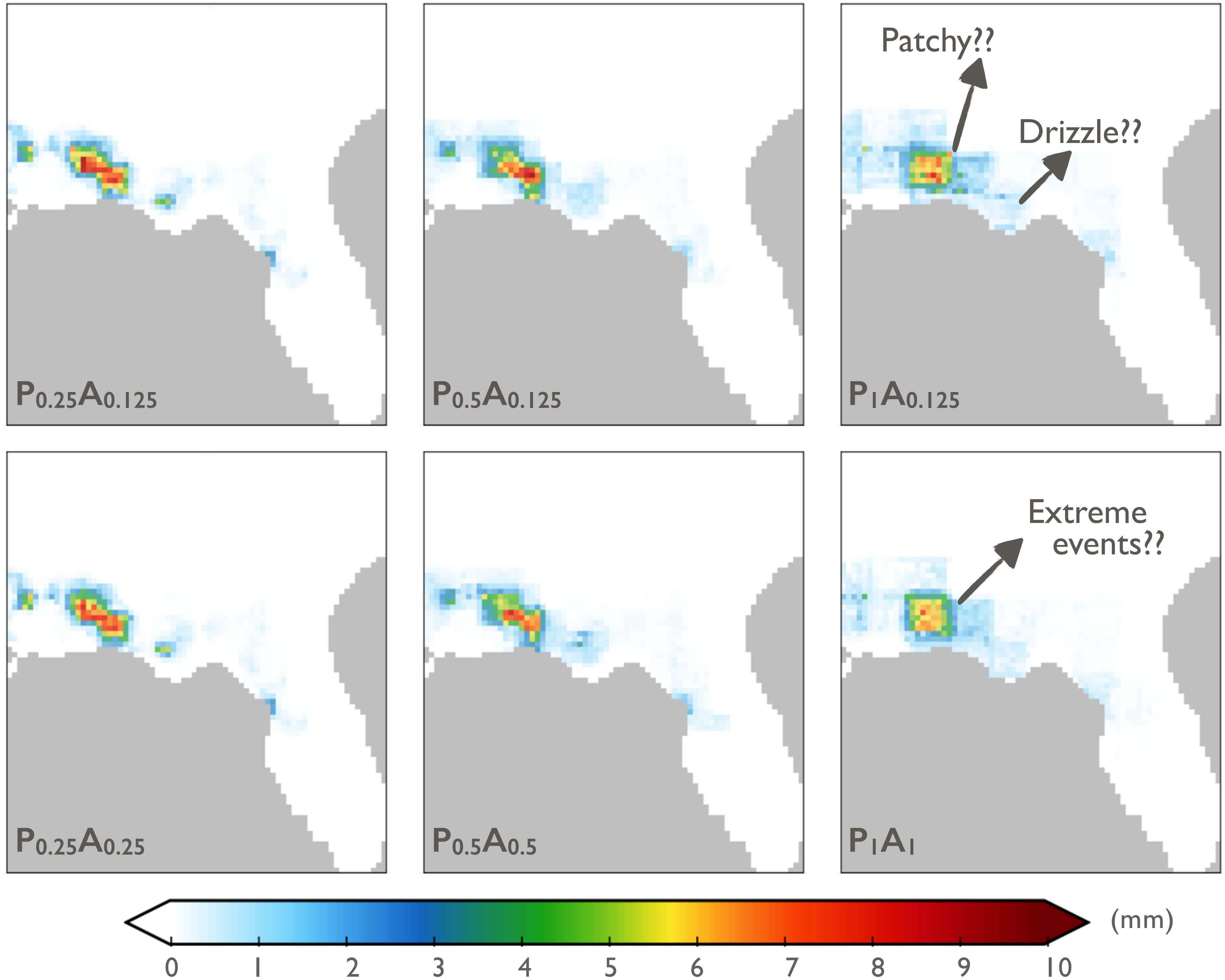
# Spatial pattern

## Results

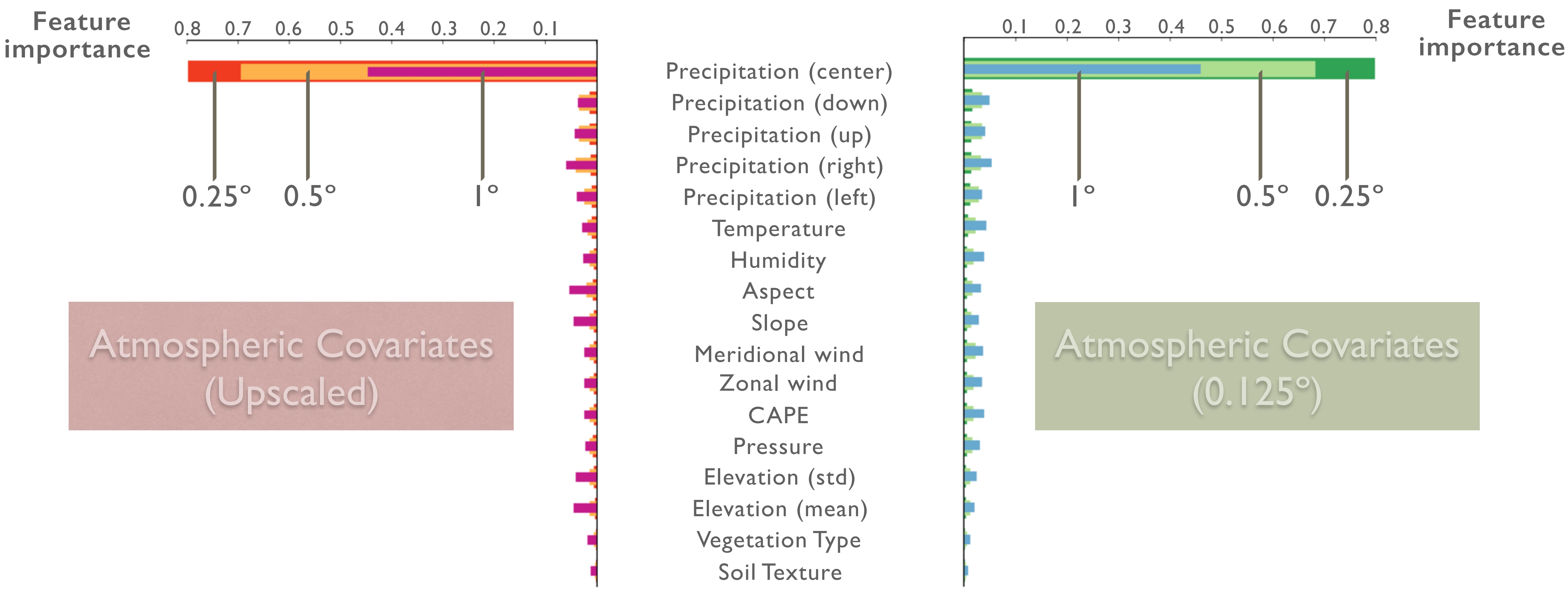
**Scaling Experiments**




**Example:**
**2011.06.06.23:00**



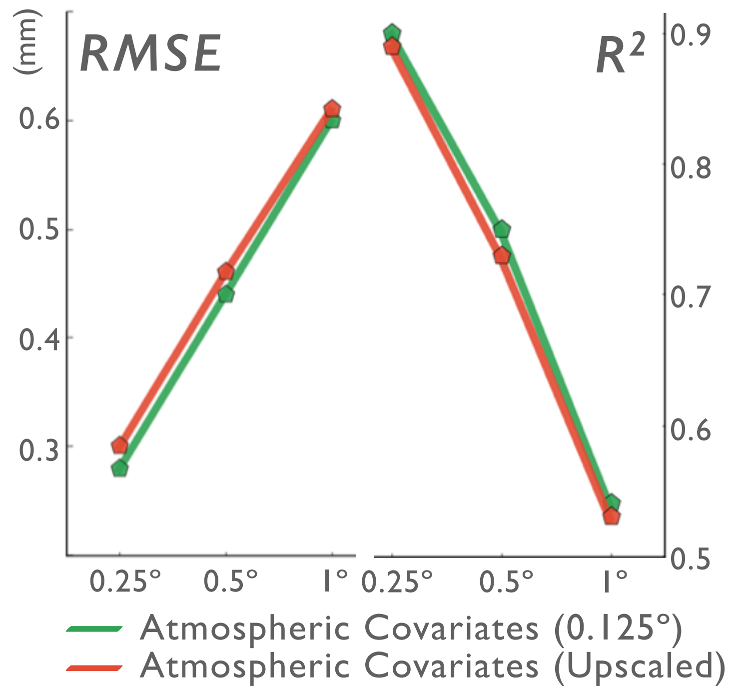
***Feature importance*** measures the prediction strength for each covariate



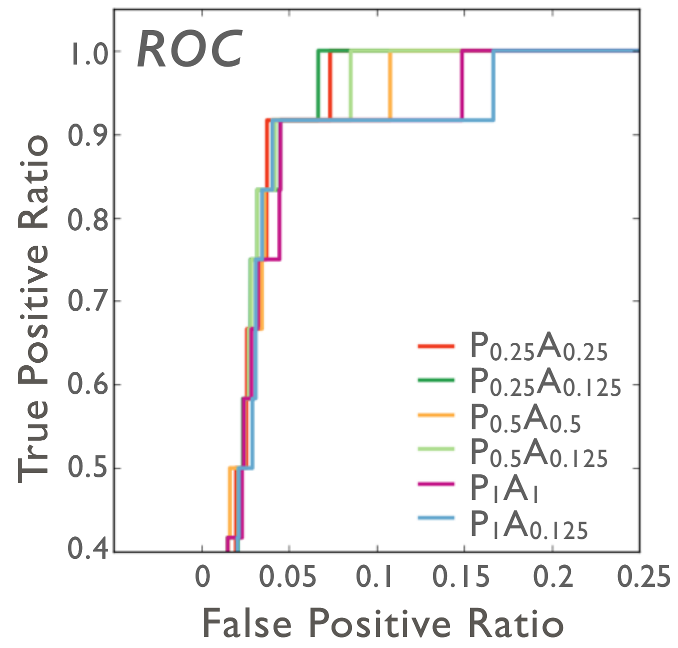

★ Central coarse precipitation dominates
★ Dynamic fields matter
★ Topography matters (e.g., $P_1A_1$)

# Downscaling skill

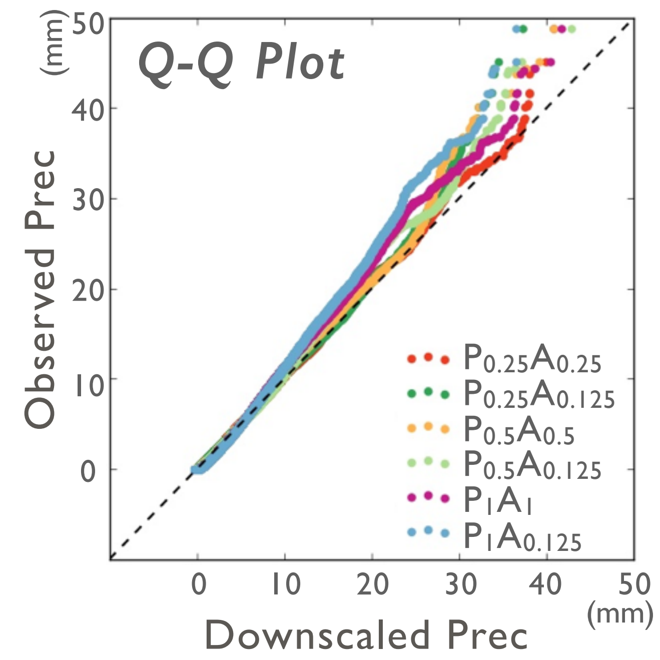
## Goodness-of-fit



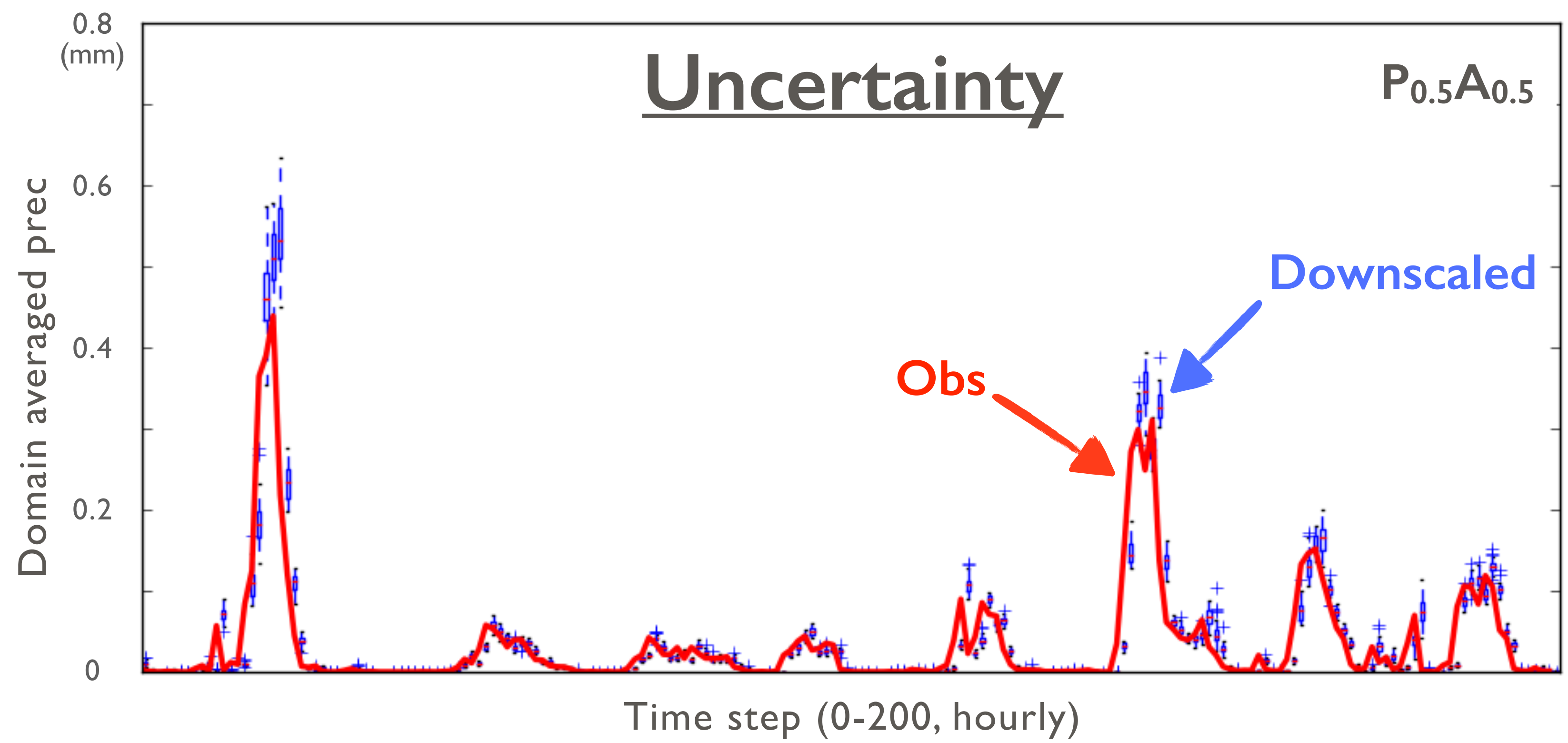

## Classification




## Distribution

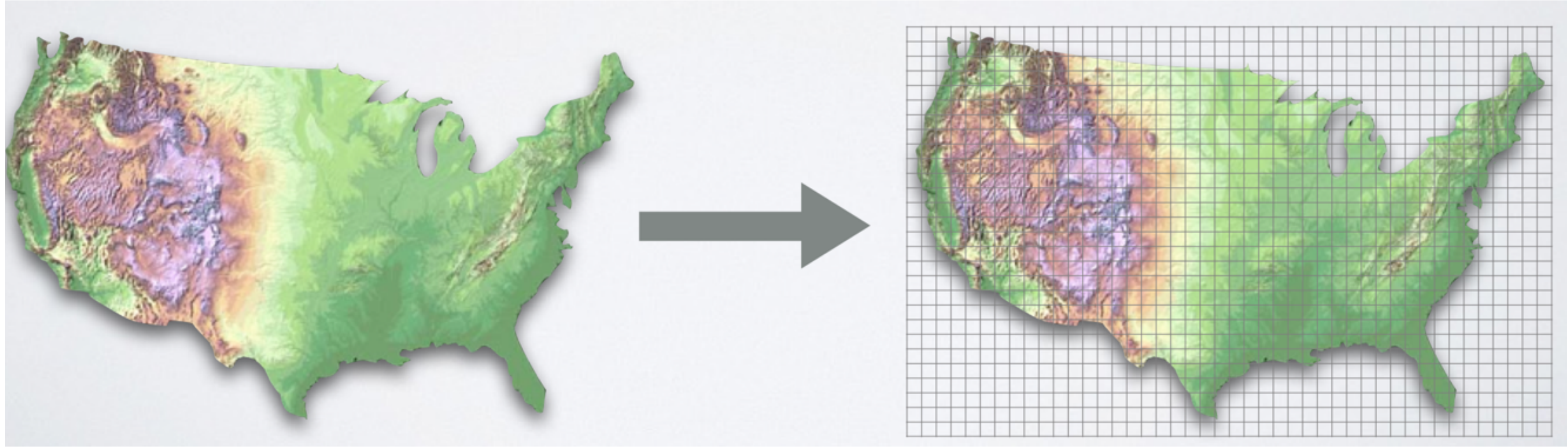



## Uncertainty



# Future work

- Synthetic experiments → Try other regions
  - Western US (California and mountainous regions)
  - Northeastern climate division
  - Central US
- Real experiments
  - Train StageIV, downscale satellite/reanalysis
- CONUS/Global scale → Moving window approach



*Thanks!!*

*Questions??*