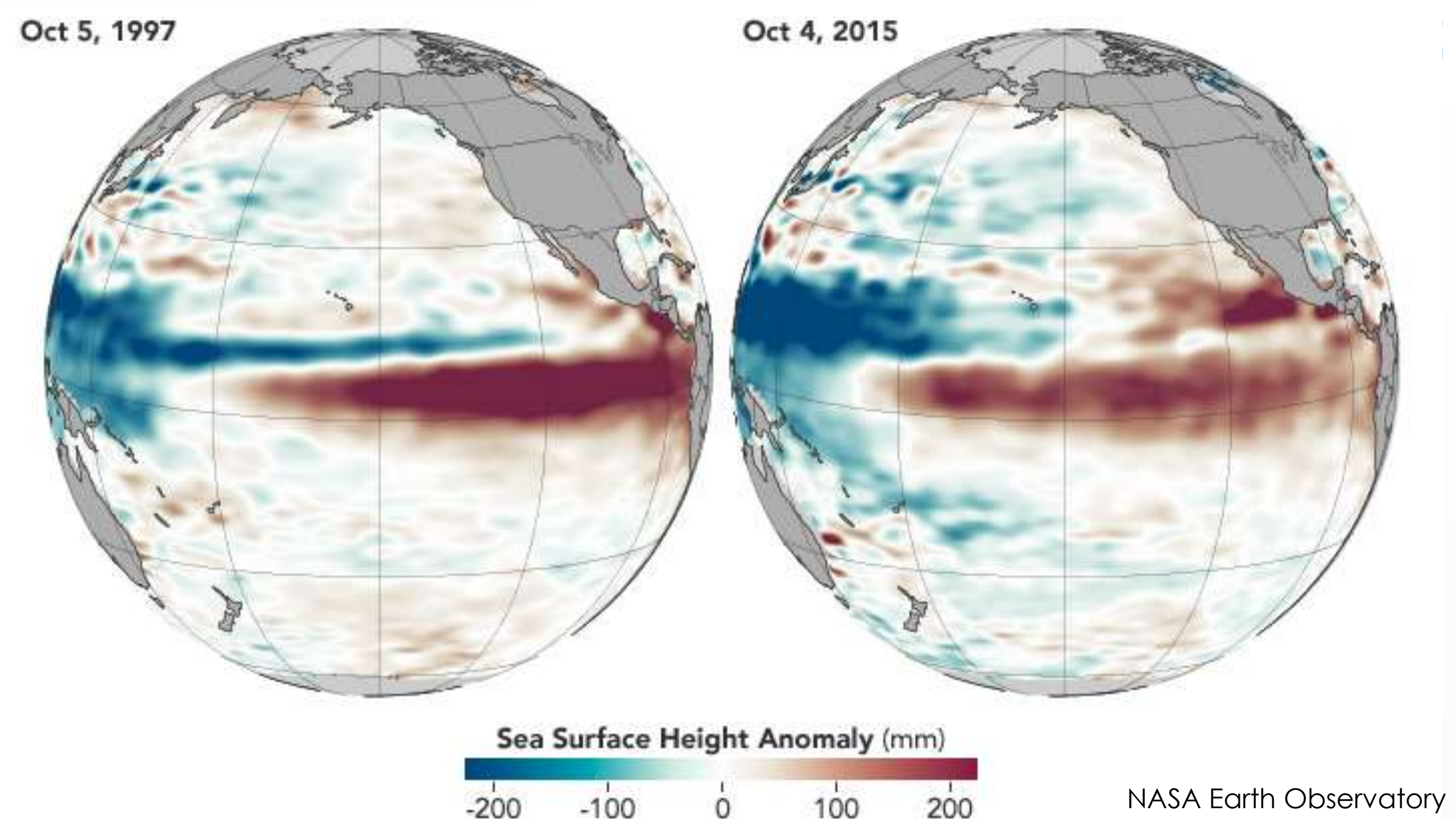
An Interactive System to Optimize Statistical Seasonal Forecasts and Climate Change Projections

Ben Zaitchik

Department of Earth & Planetary Sciences

Johns Hopkins University



NASA Earth Observatory

El Niño could leave 4 million people in Pacific without food or drinking water

Papua New Guinea drought has already claimed two dozen lives and looming El Niño weather pattern could be as severe as 1997-98, when 23,000 died globally

Children at Algi village in Papua New Guinea. PNG is in the midst of what could be its worst drought in close to

As El Niño gets stronger Australia gets hotter, drier and more ready to burn

Bureau of Meteorology says prospect of drier-than-normal October is about 70% in southern Australia which comes after third-driest September on record

The ruins of a house near Lancefield, north of Melbourne, or

When the Indian Ocean and El Niño join forces, things can get hot and dry

October 11, 2015 9.45pm EDT

El Niño: a global weather event that may save California – and destroy the tropics

WATER & DROUGHT OCTOBER 4, 2015

Will El Niño 'solve' drought? Not if the rain falls in Southern California

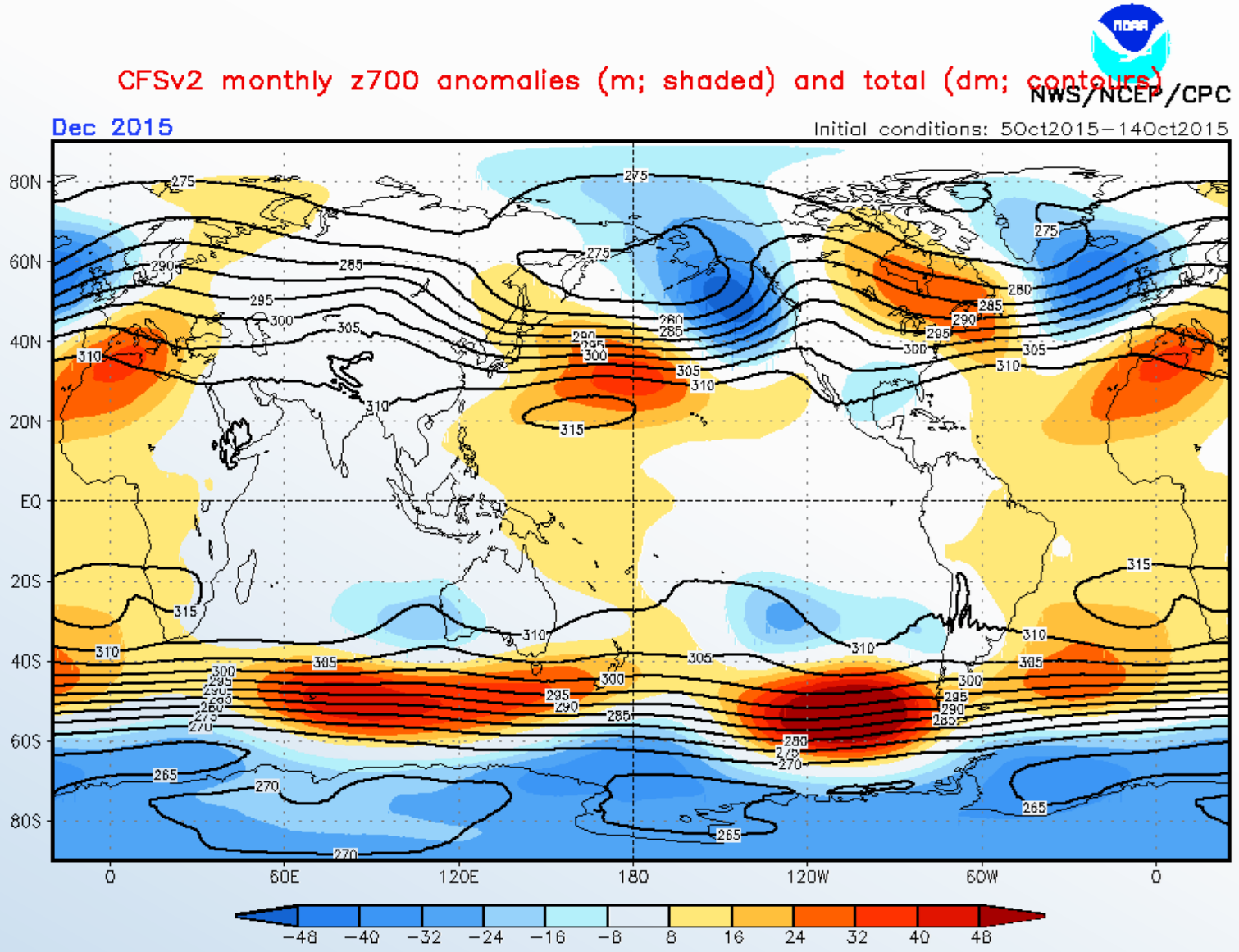
One region's weather win is another region's catastrop

El Niño: Wet winter likely across California, not just in south, new report says

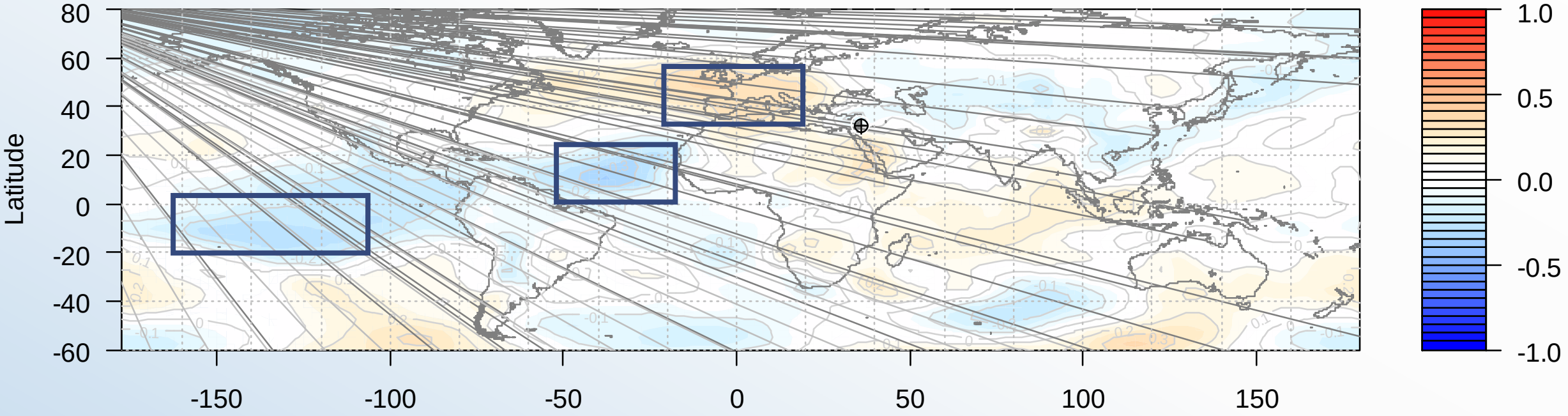
By Paul Rogers | progers@mercurynews.com

POSTED: 10/15/2015 11:44:06 AM PDT | UPDATED: ABOUT 6 HOURS AGO 21 COMMENTS

Dynamical vs. Statistical



Dynamical vs. Statistical



Correlation: Amman Nov rainfall and global Sea Level Pressure

Why use statistical methods?

- Performance
- Computational efficiency
- Local specificity
- Empirical foundation
- Ability to leverage independent datasets

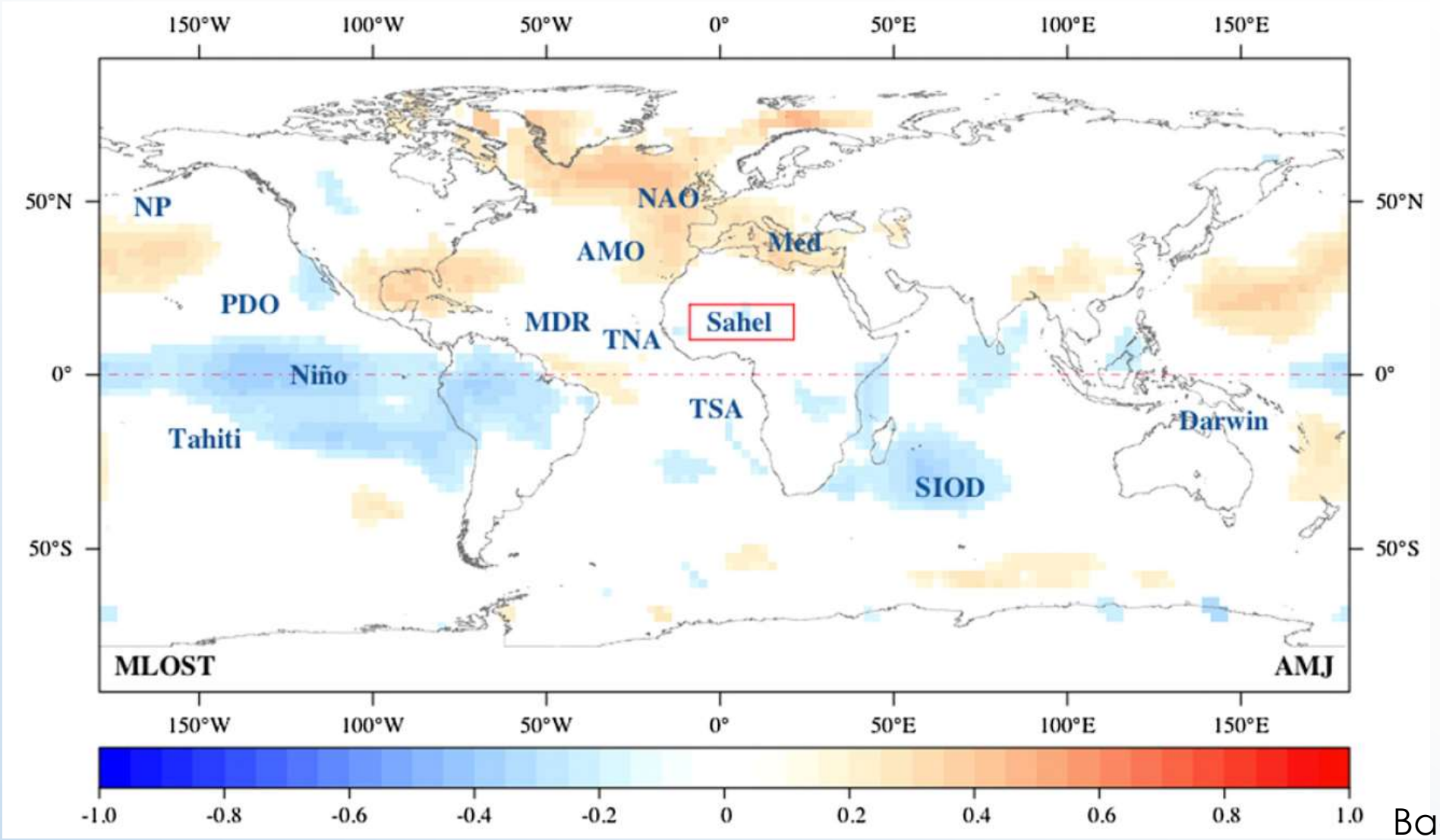
The Big Data Challenge

- Huge numbers of potential predictors
- Wide range of processing options
- Diverse analytic techniques

Our Goals:

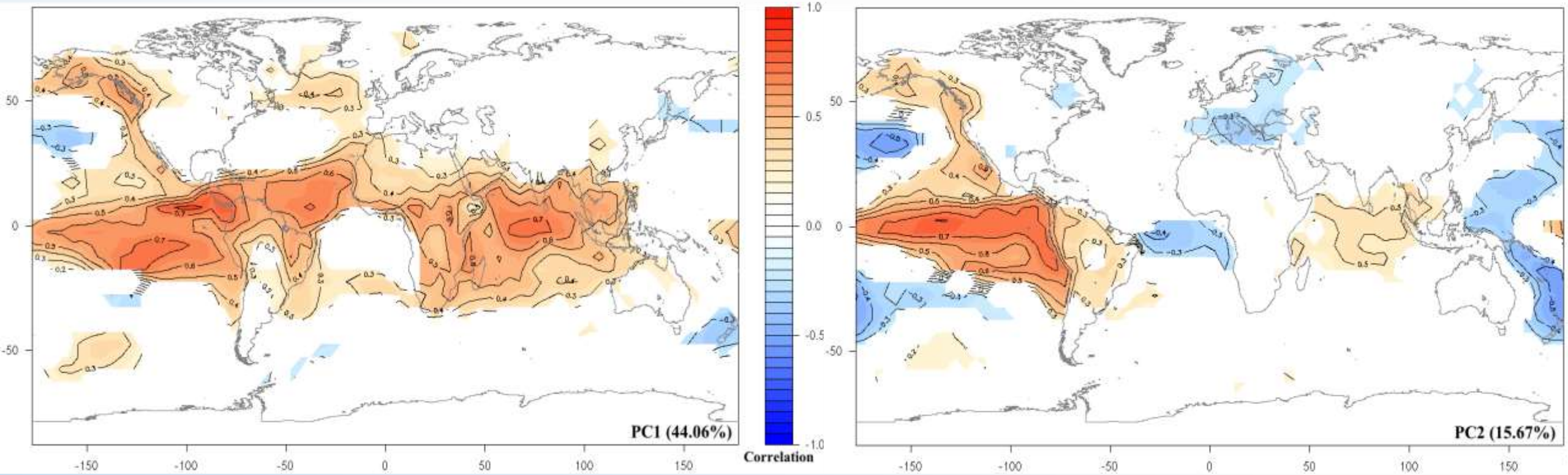
- Create a system for optimizing statistical forecasts
- Create an online tool for generating optimized forecasts
- Automate variable selection across datasets

Optimizing Forecasts: The Sahel



Badr et al. (2014)

Optimizing Forecasts: The Sahel



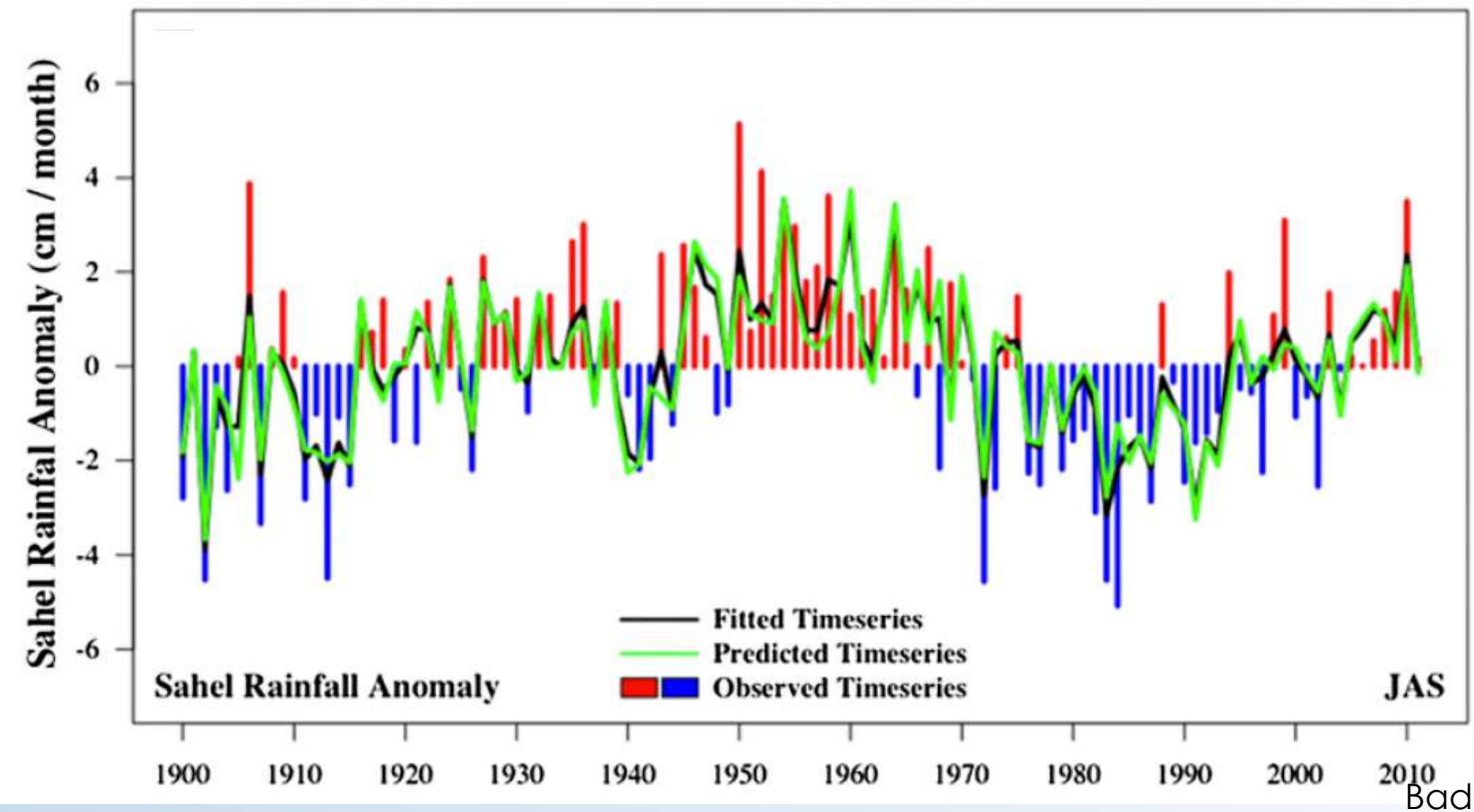
Badr et al. (2014)

Optimizing Forecasts: The Sahel

Model	Description
GLM	**Full-covariate generalized linear model**
SGLM	Selected generalized linear model based on stepwise selection
GAM	Full-covariate generalized additive model
SGAM	Selected generalized additive model based on penalized terms
MARS	Multivariate adaptive regression spline
CART	Classification and regression trees model
BCART	Bagged classification and regression trees model
BART	Bayesian additive regression trees model
RF	Random forest model
ANN	**Artificial neural network**
Avg	Prediction is the mean of the predictions of all models
Null	Prediction is the mean of the response variable in the training data
Memory	Model using last year's rainfall as this year's prediction

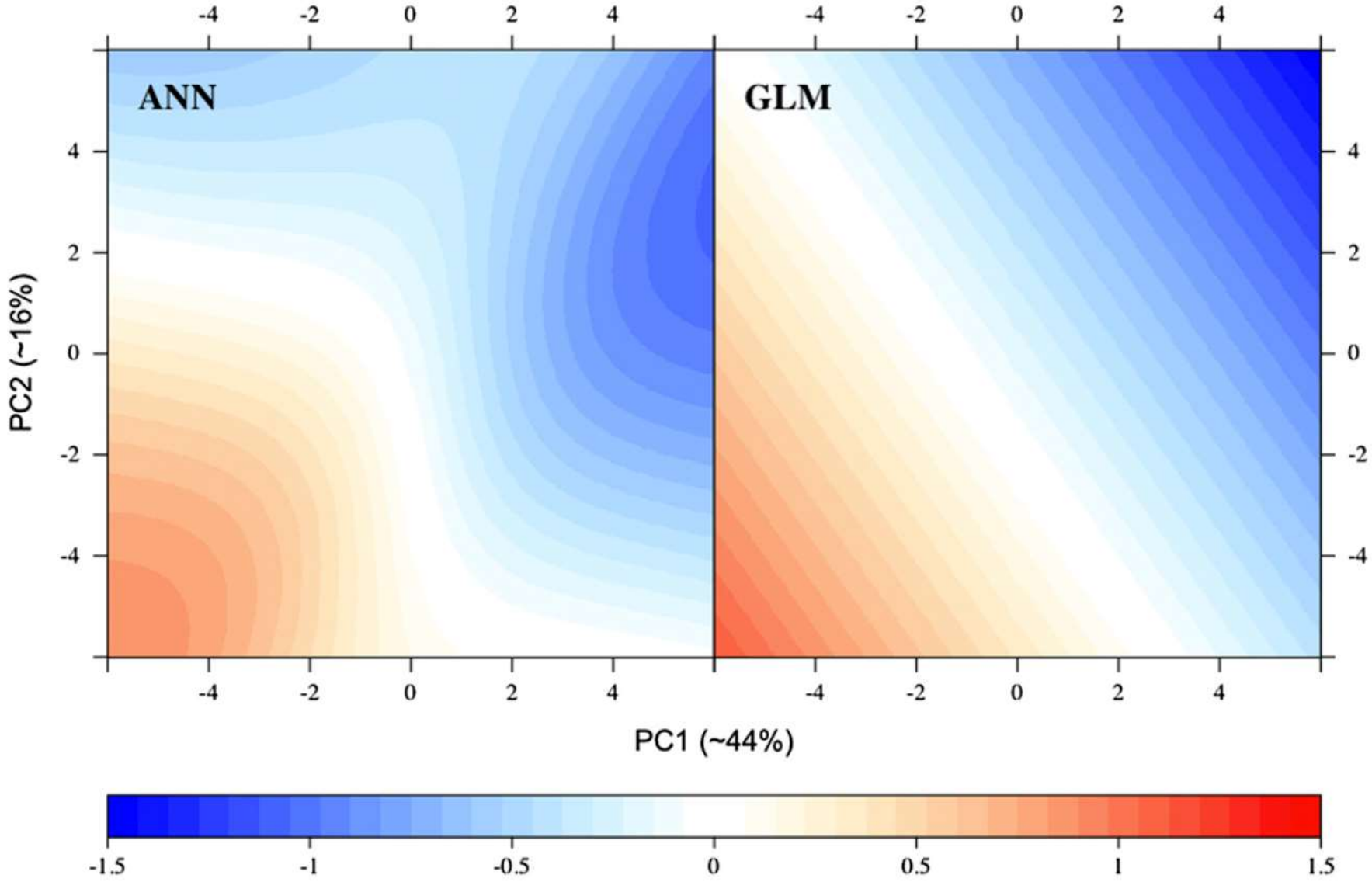
Badr et al. (2014)

Optimizing Forecasts: The Sahel



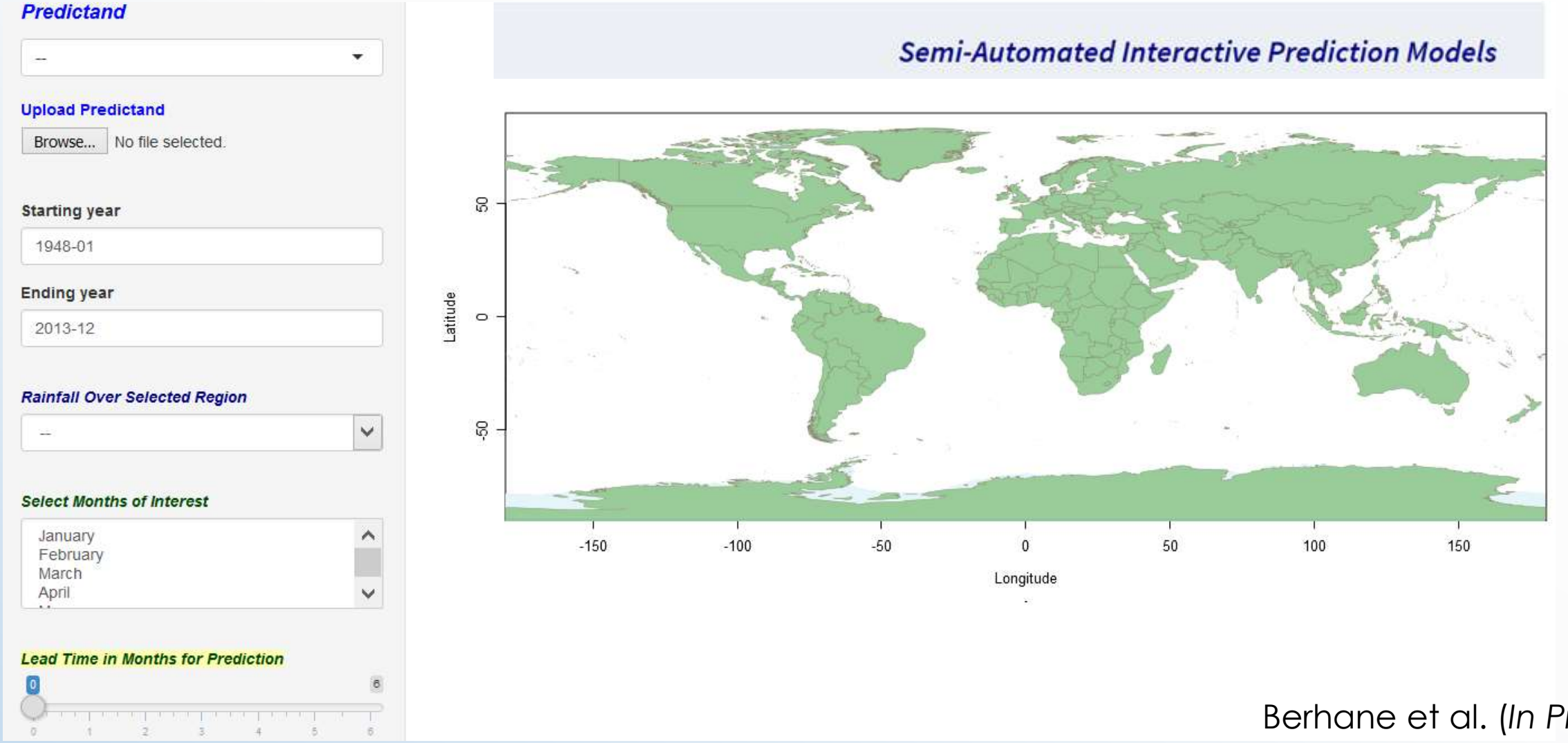
Badr et al. (2014)

Optimizing Forecasts: The Sahel



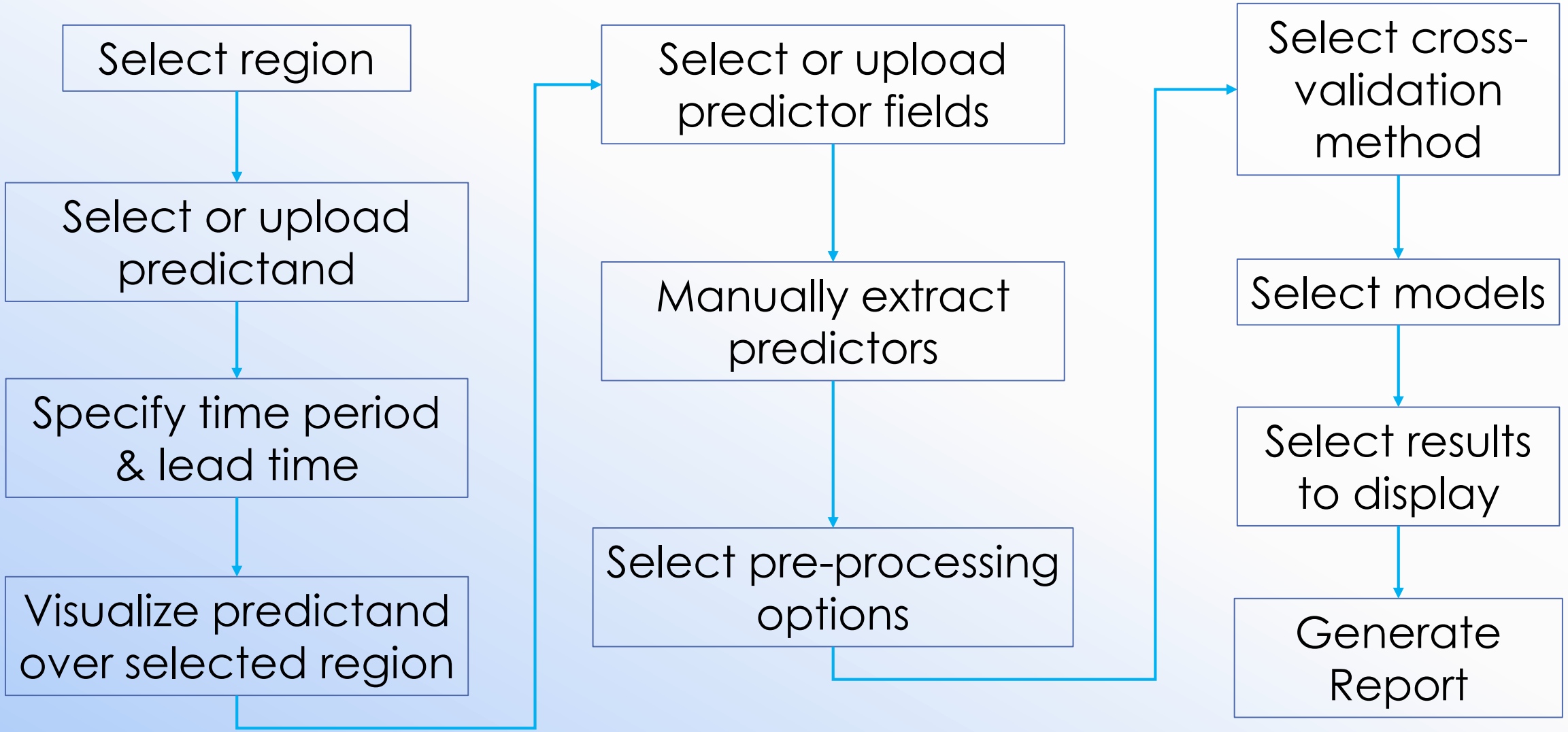
Badr et al. (2014)

Online Prediction Tool



Berhane et al. (*In Prep*)

Online Prediction Tool



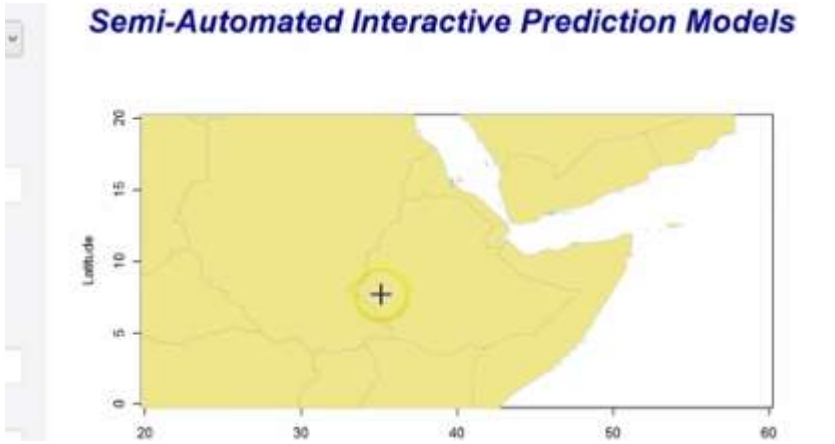
Online Prediction Tool

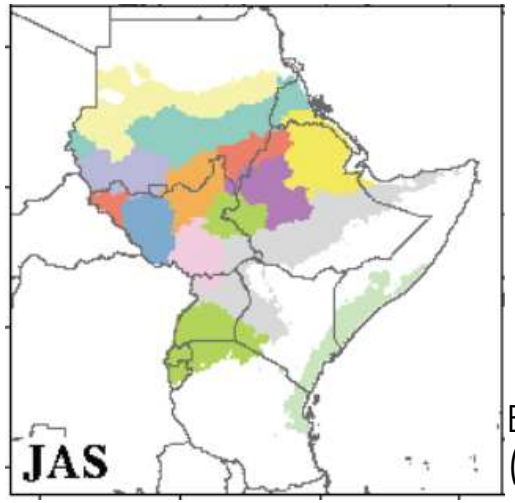
Select region

Select or upload predictand

Specify time period & lead time

Visualize predictand over selected region





Badr et al. (2015)

Select cross-validation method

Select models

Select results to display

Generate Report

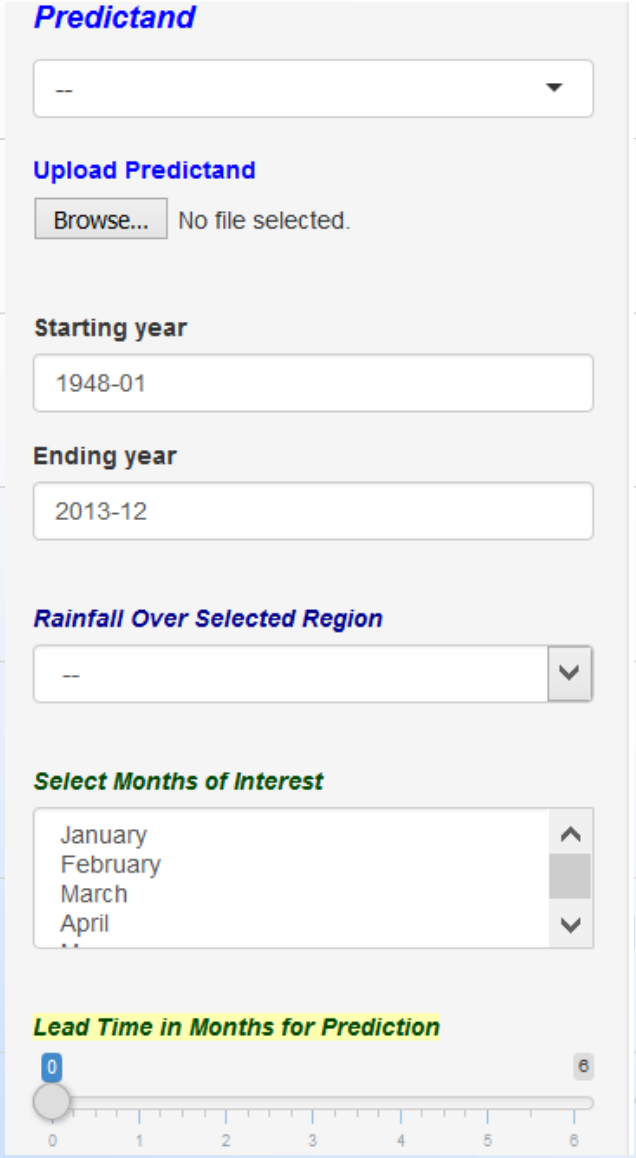
Online Prediction Tool

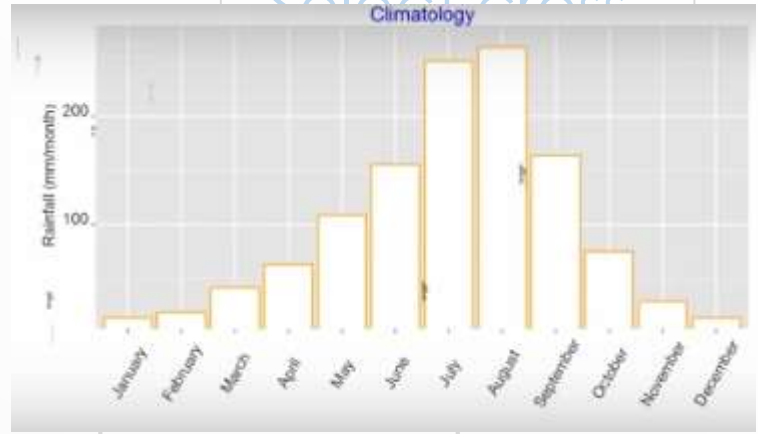
Select region

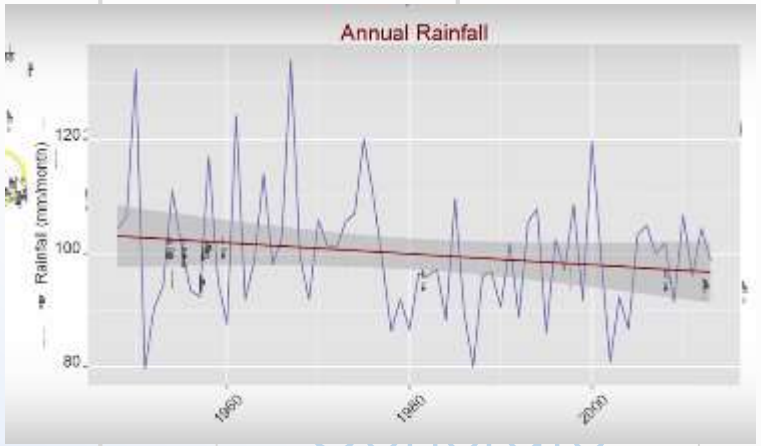
Select or upload predictand

Specify time period & lead time

Visualize predictand over selected region







Online Prediction Tool

Select region

Selec
pre

Specify
& l

Visualiz
over se

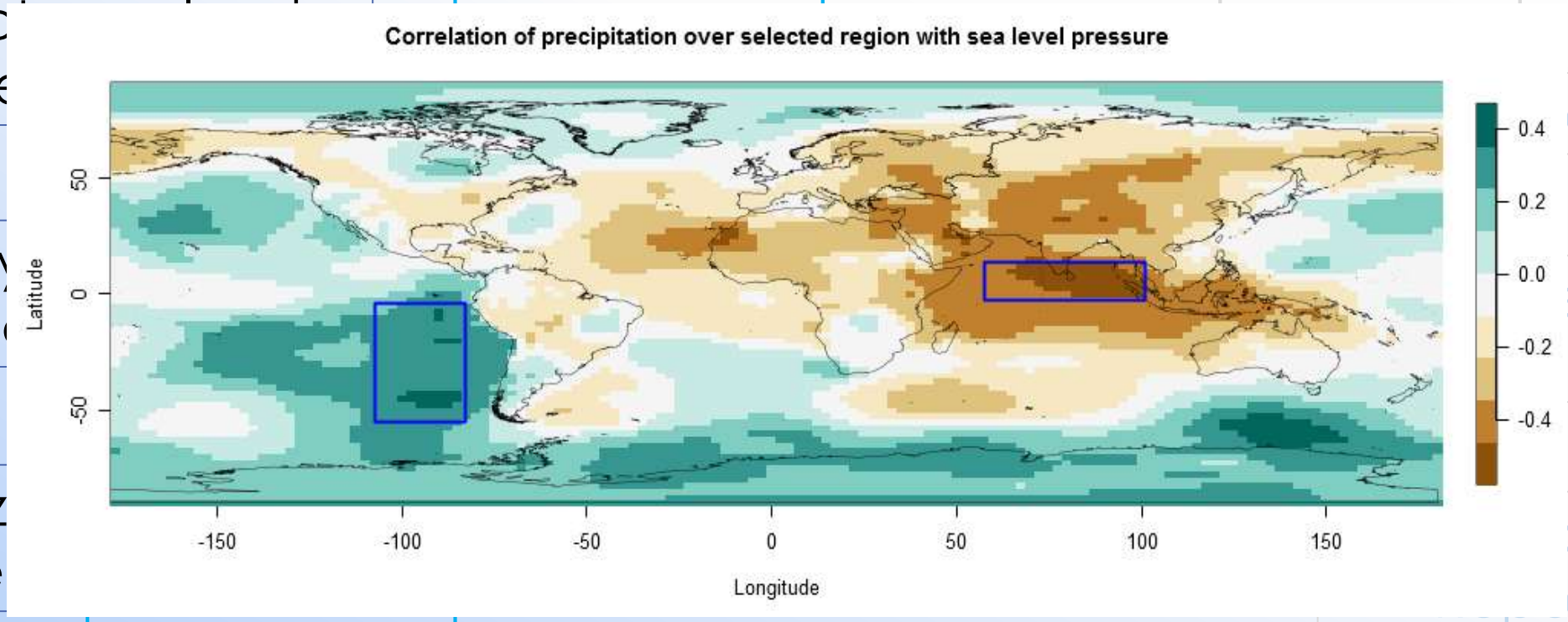
Select or upload predictor fields

Select cross-validation method

odels

esults
lay

ate
rt



ction Tool

Models

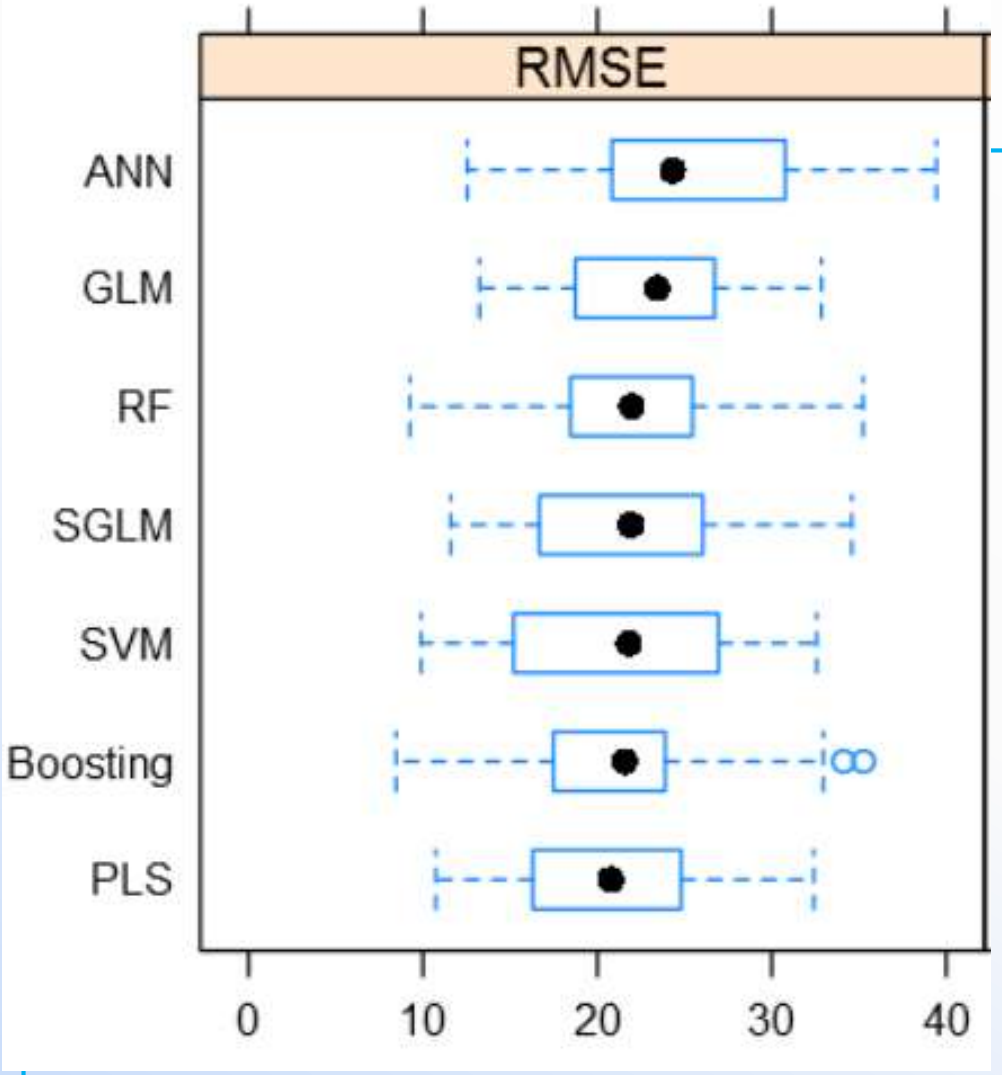
GLM
SGLM
PLS
LASSO
RIDGE
ENET
GAM
SGAM
CART
BCART
RF
Boosting
SVM
ANN
MARS
ALL

Quick Overview of Model Reults

--

RMSE
Rsquared
Vimportance
--

Download results



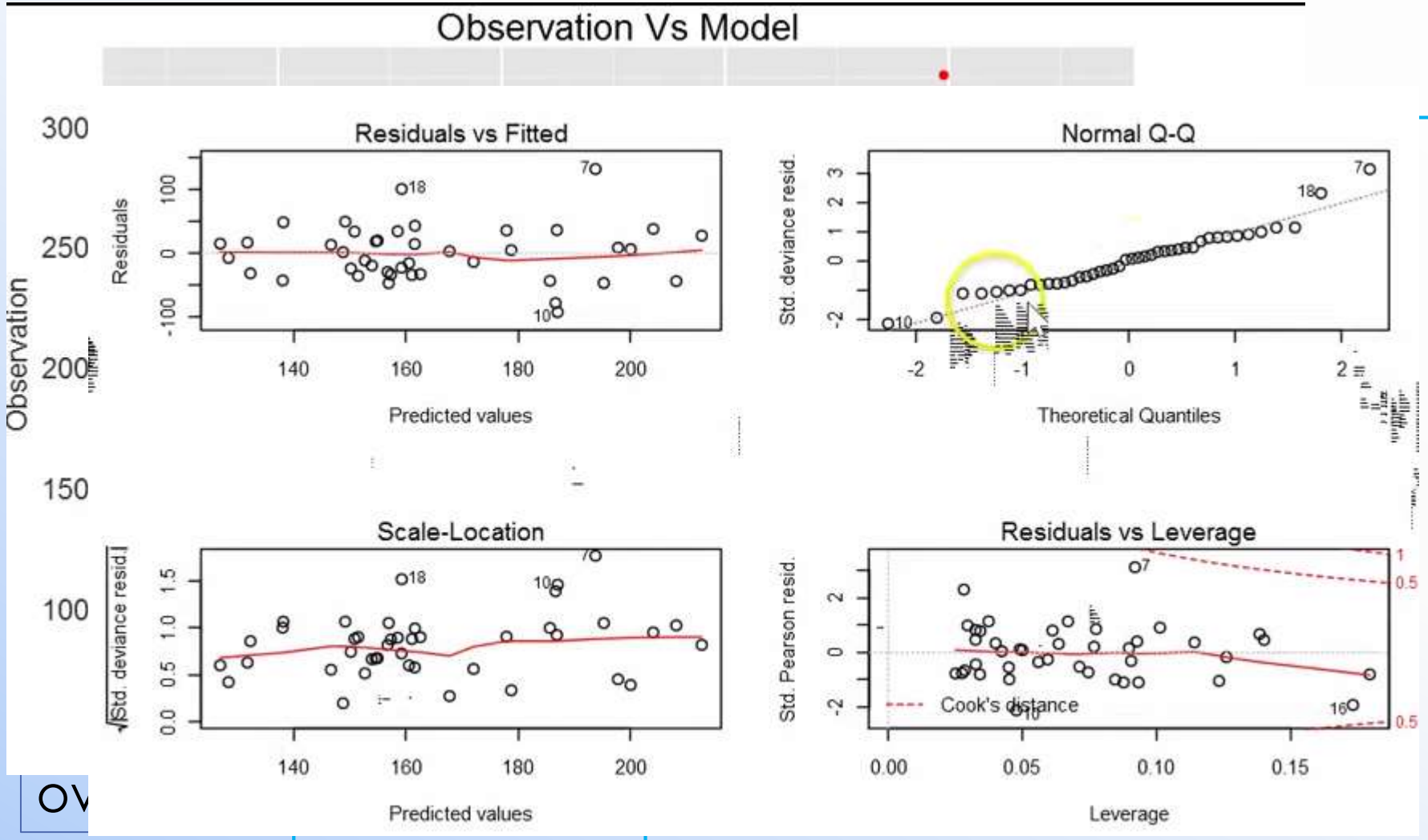
Select cross-validation method

Select models

Select results to display

Generate Report

Online Prediction Tool



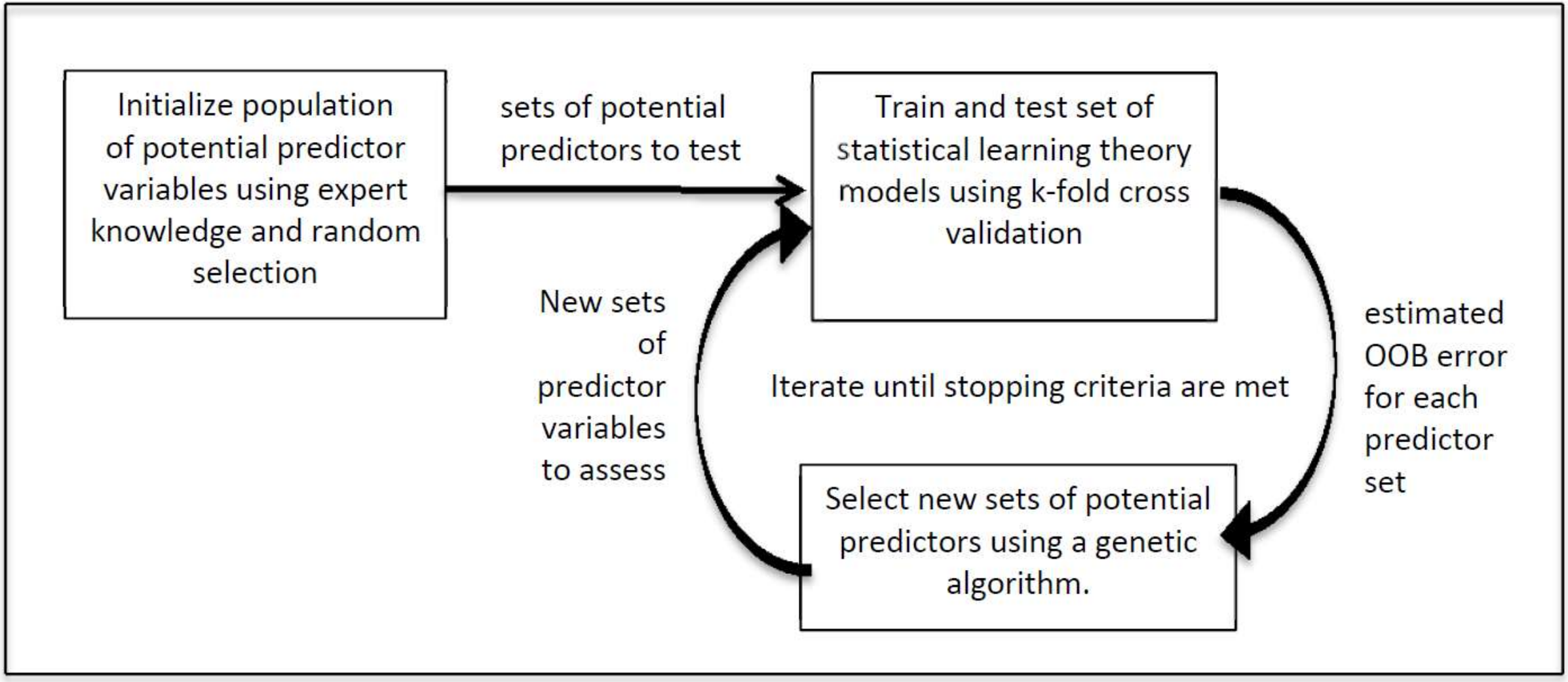
Select cross-validation method

Select models

Select results to display

Generate Report

Next Step: full automation



Thank you

Select region (from global map)

Select or upload predictand (Provided: CRU and CHIRPS)

Specify time period & lead time

Visualize and explore predictand over selected region (distribution, outliers, trend, etc.)

Select or upload fields (Provided: SST, SLP, Hgt, wind, airT)

Extract features by drawing polygons over regions of maximum correlation

Select pre-processing options (Provided: remove outliers, PCA, standardize,

Select cross-validation method

Select models

Select results to display

Generate & Download Report (html, Pdf)

Forecast vs. Projection

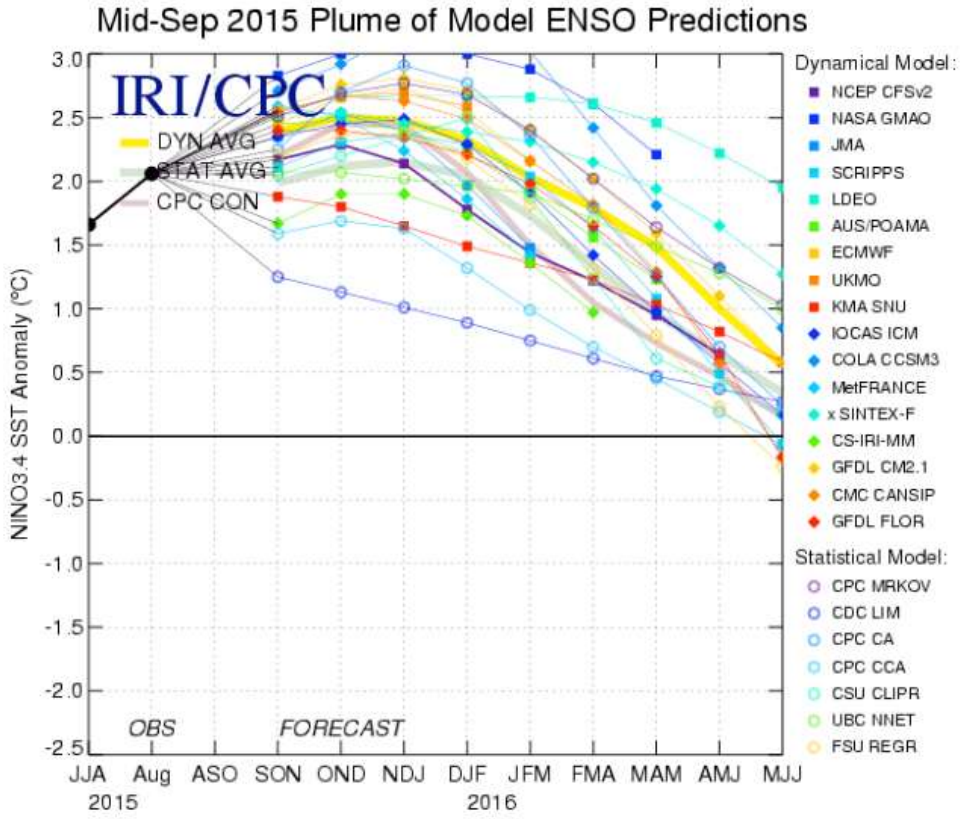
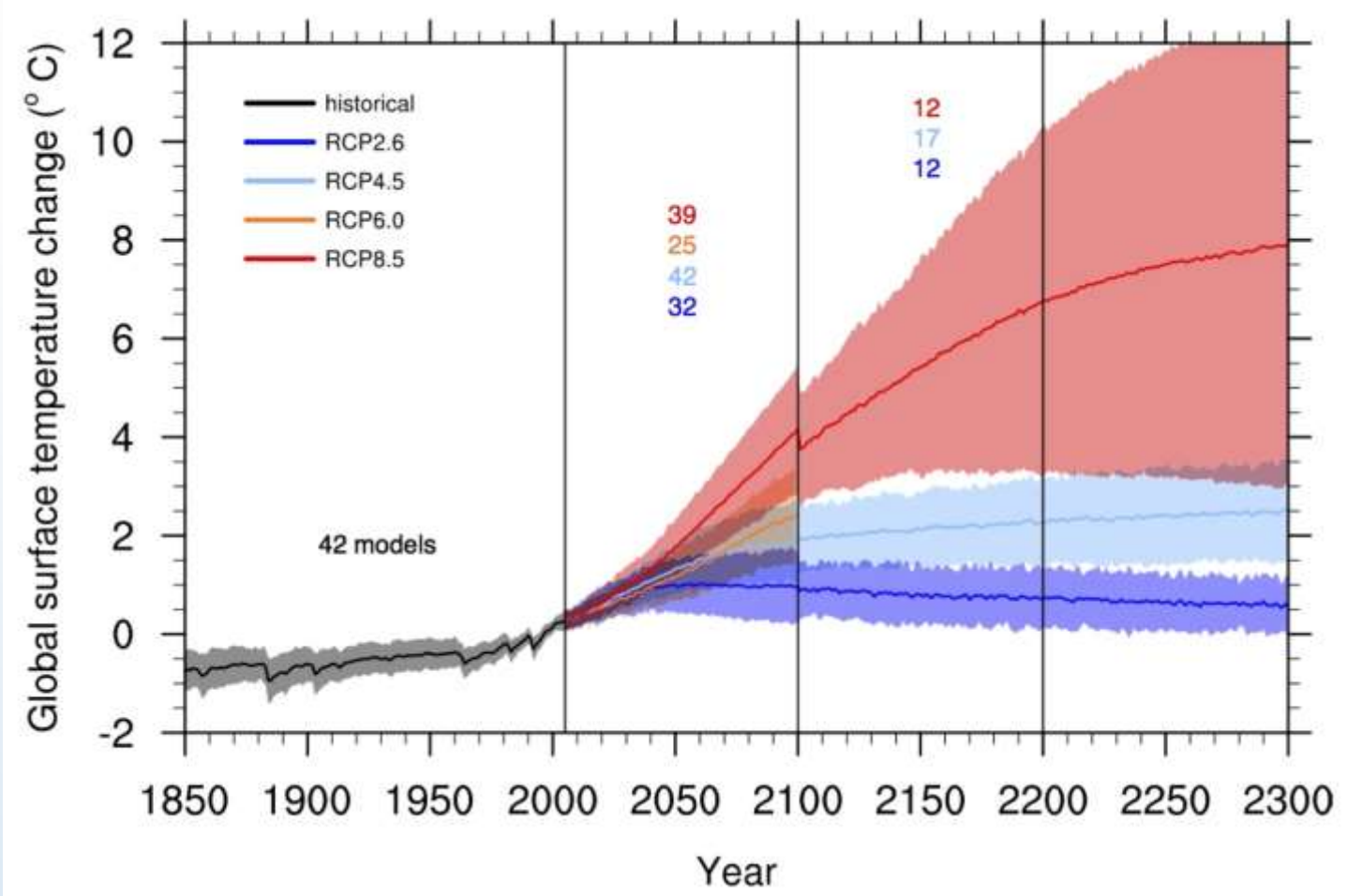



Figure 6. Forecasts of sea surface temperature (SST) anomalies for the Niño 3.4 region (5°N-5°S, 120°W-170°W). Figure updated 15 September 2015.