# A Unique Technique for Long Range Precipitation & Streamflow Forecasting and its Applications for Spring/Aquifer Flow Forecasting

*Mel Kunkel*

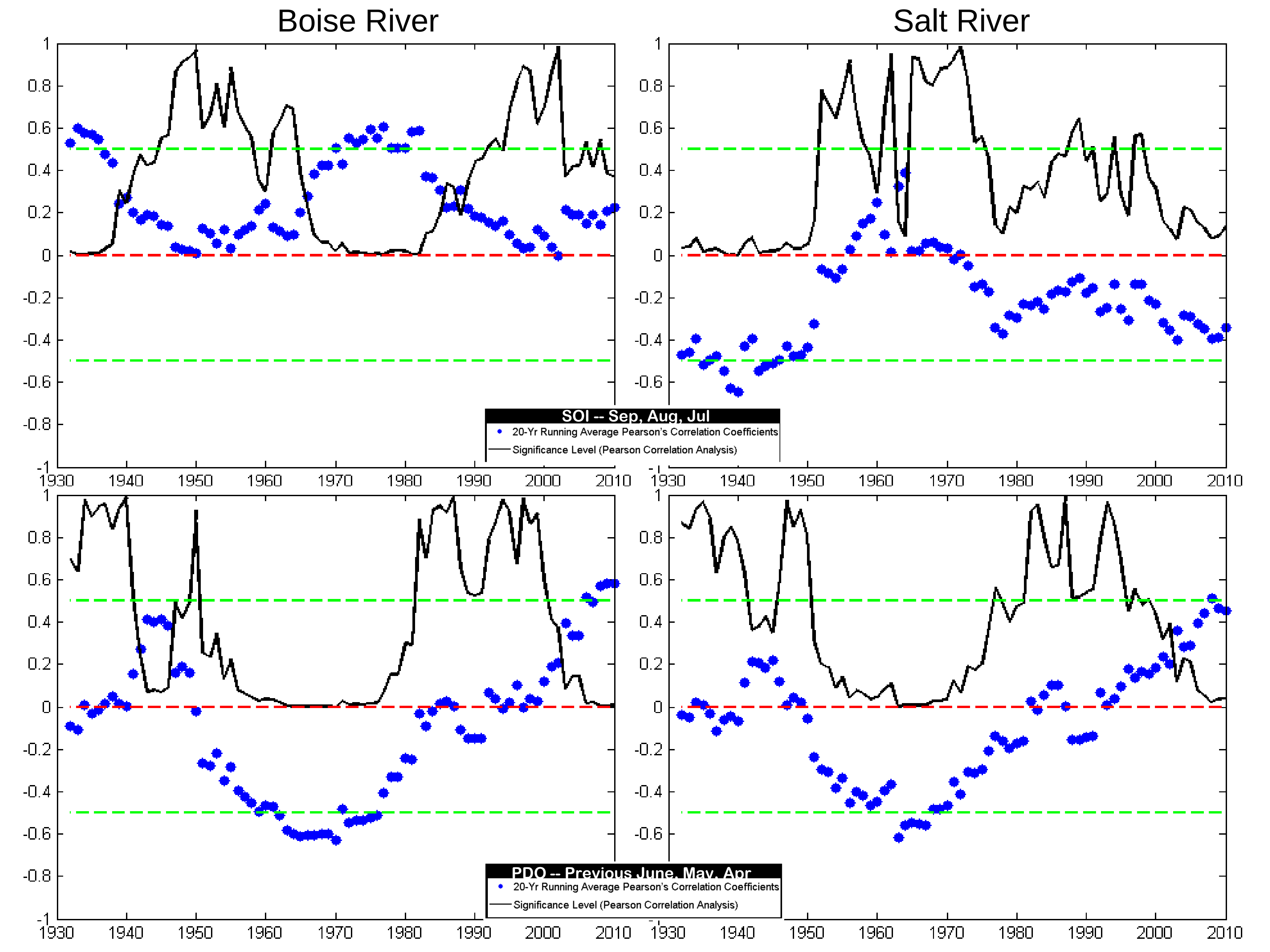
# Locations

Kettle River near Laurier, Wa
HCDN Stream data (1913-2010)

Boise River near Boise, Id
BOR Qu Data (1912-2010)

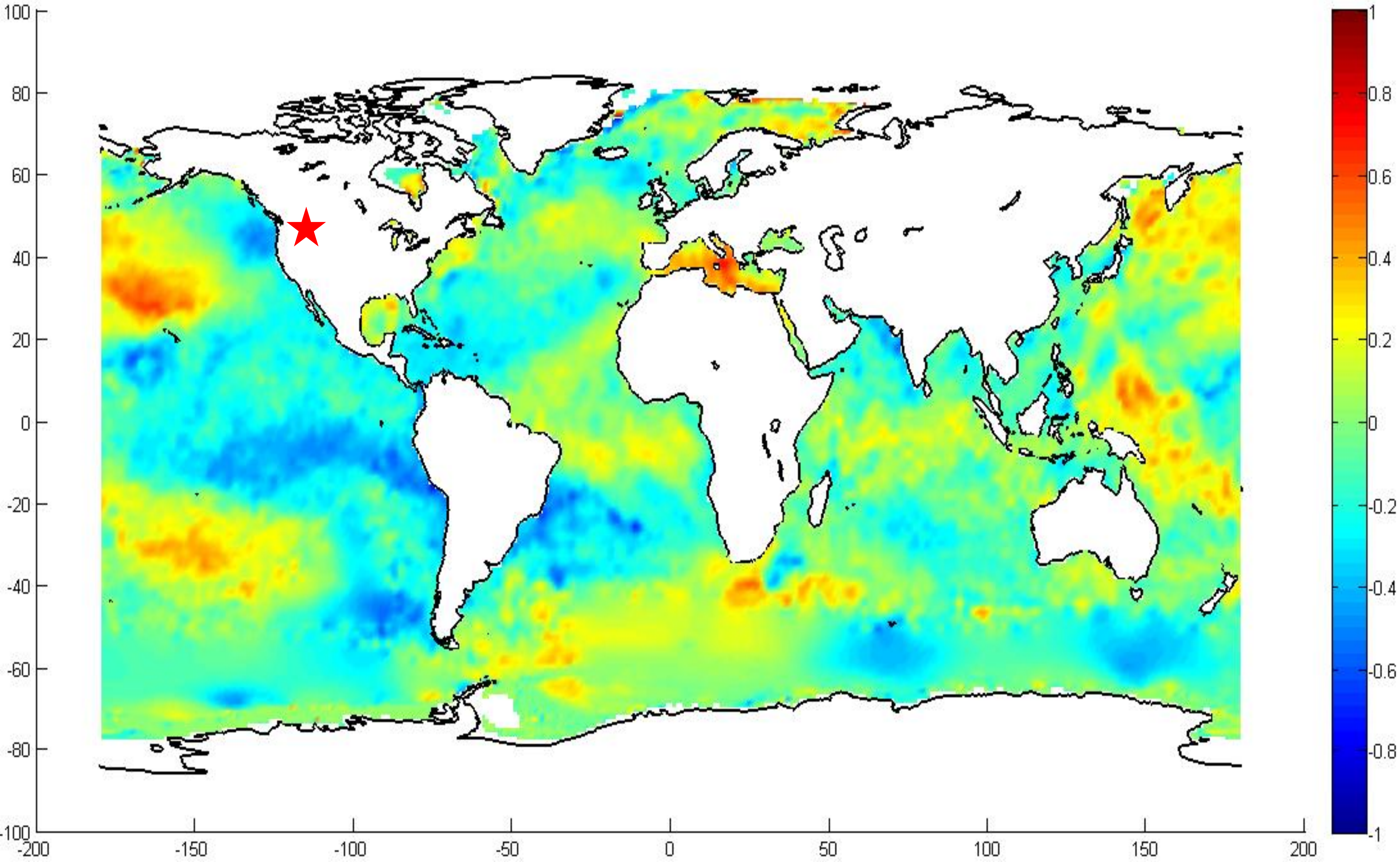
Green River near Green River, Wy
BOR Qu data (1912-2010)

Salt River near Roosevelt, Az
HCDN Stream (1912-2010)

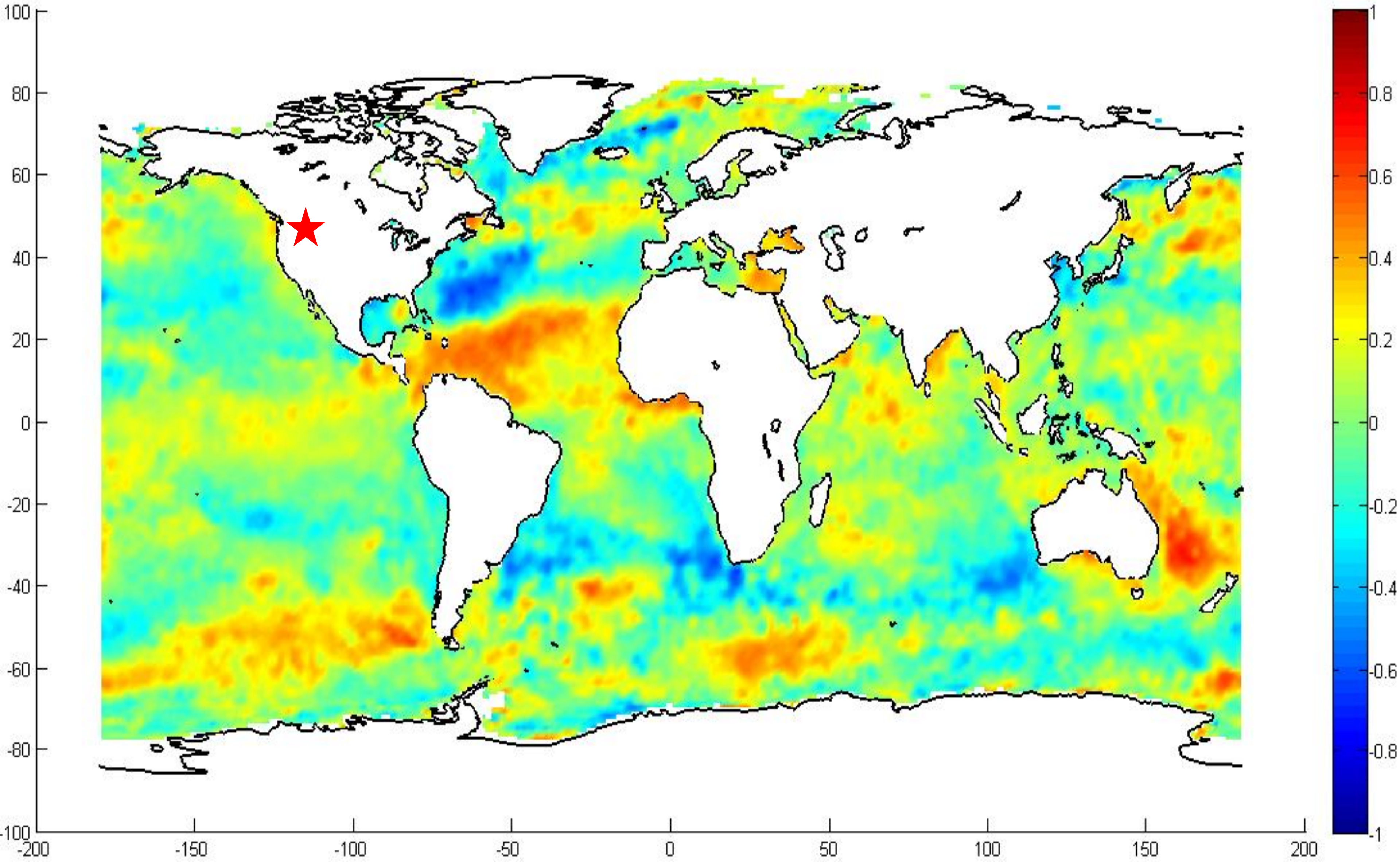

Boise River
Salt River
SOI -- Sep, Aug, Jul
20-Yr Running Average Pearson's Correlation Coefficients
Significance Level (Pearson Correlation Analysis)
PDO -- Previous June, May, Apr
20-Yr Running Average Pearson's Correlation Coefficients
Significance Level (Pearson Correlation Analysis)

- Boise flows verses the SOI and PDO index values.
  - Both records show significant variations in correlation over time with reversals in the sign of correlation and large changes in significance levels.
- Beyond the established teleconnections, I looked at the correlation between SSTs and Boise River annual Qu to see if similar variations occurred.

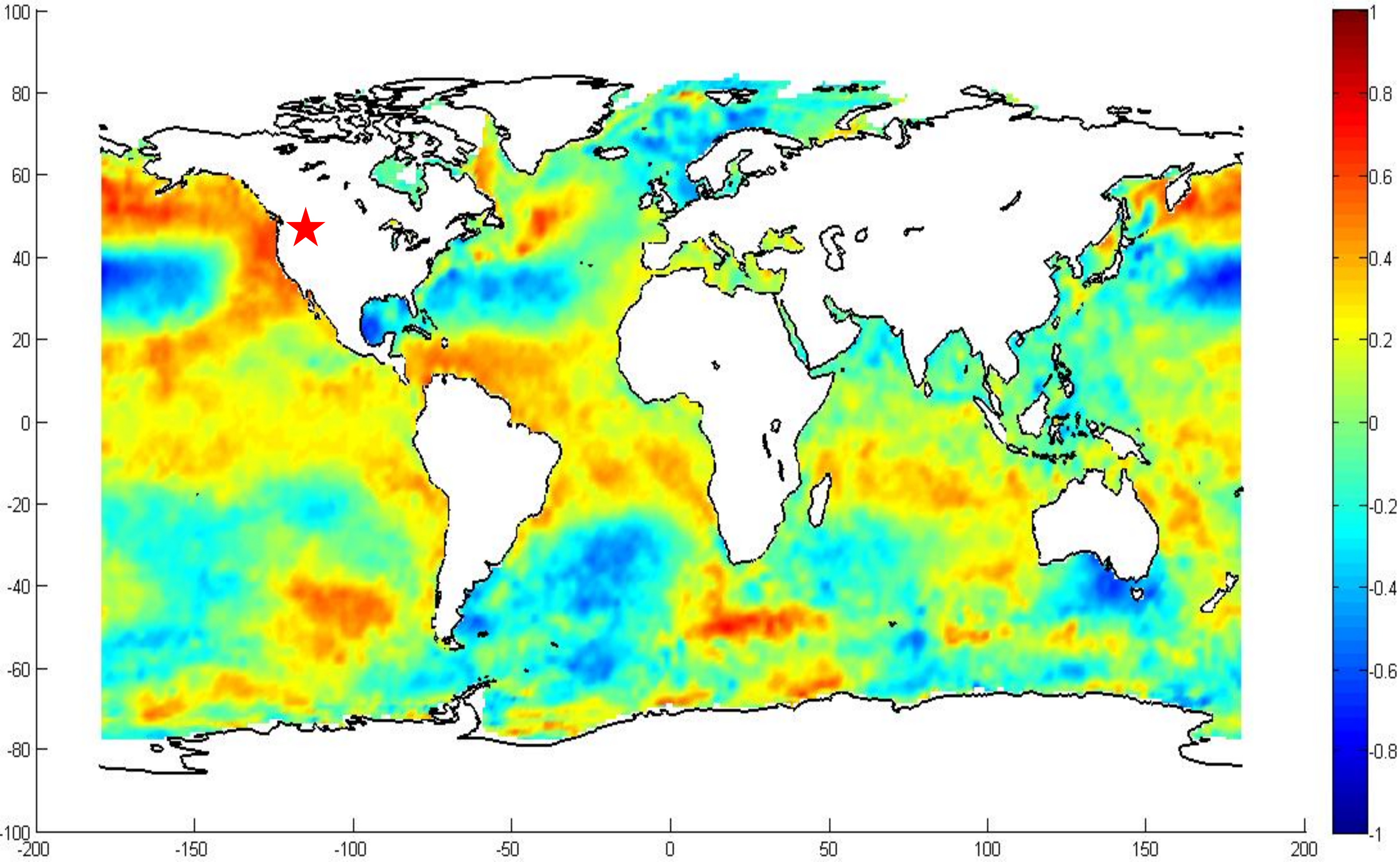
# Avg 20-year Pearson Correlation Coefficients Between Annual Boise River Streamflow and Previous Year April-June SST (1952-1971)



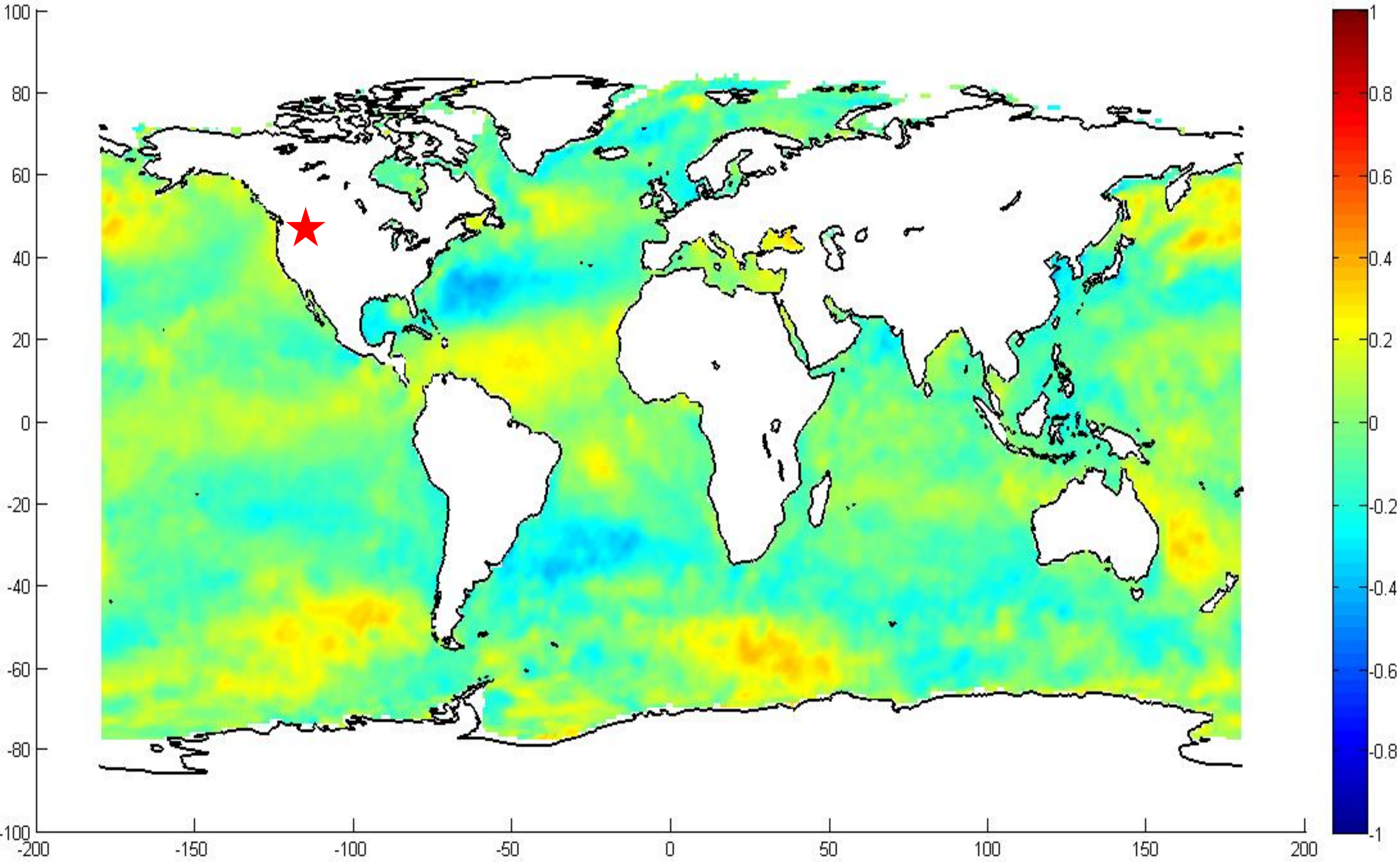
# Avg 20-year Pearson Correlation Coefficients Between Annual Boise River Streamflow and Previous Year April-June SST (1972-1991)



# Avg 20-year Pearson Correlation Coefficients Between Annual Boise River Streamflow and Previous Year April-June SST (1992-2011)



# Avg 60-year Pearson Correlation Coefficients BetweenAnnual Boise River Streamflow and Previous Year April-June SST (1952-2011)

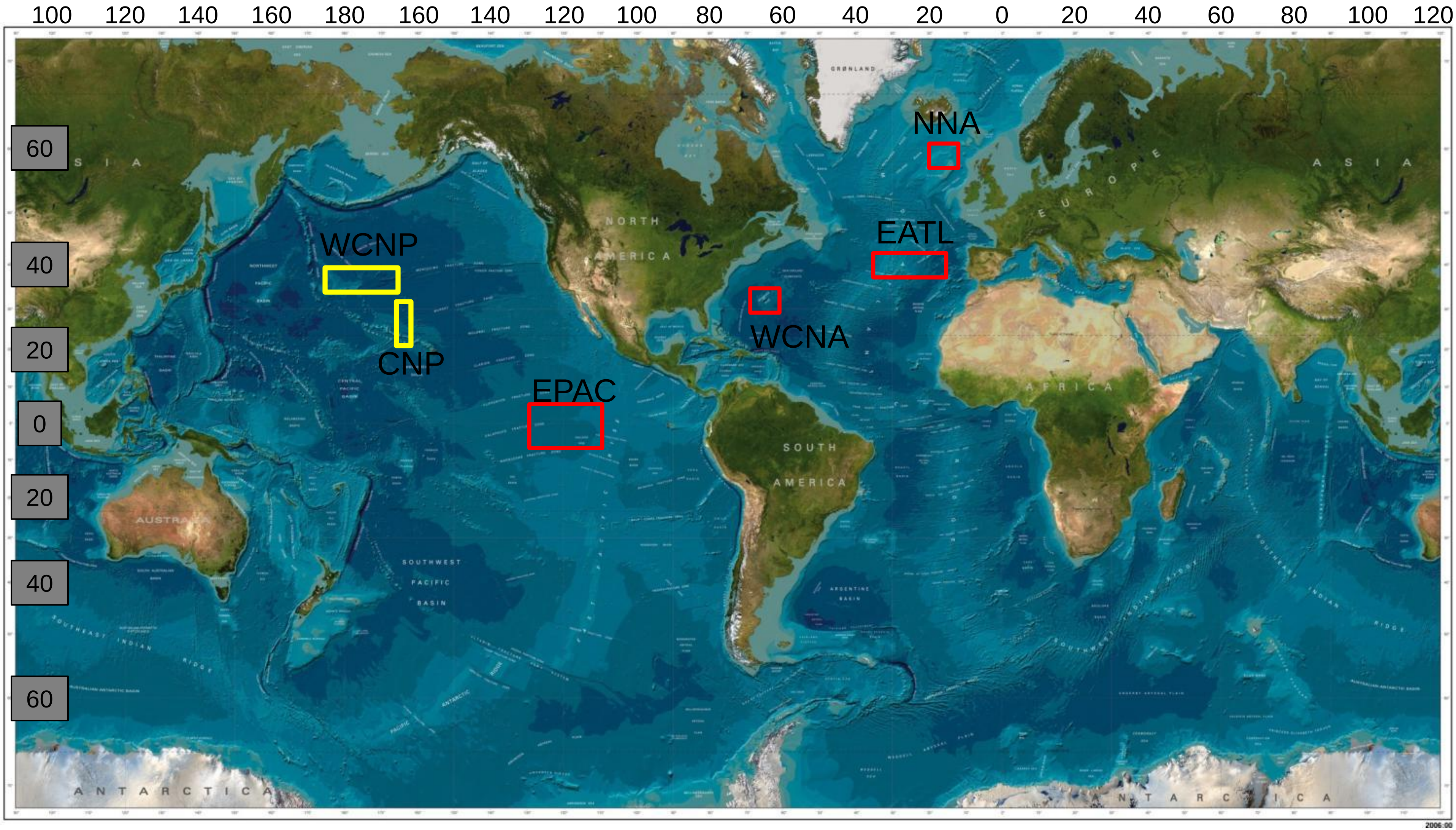


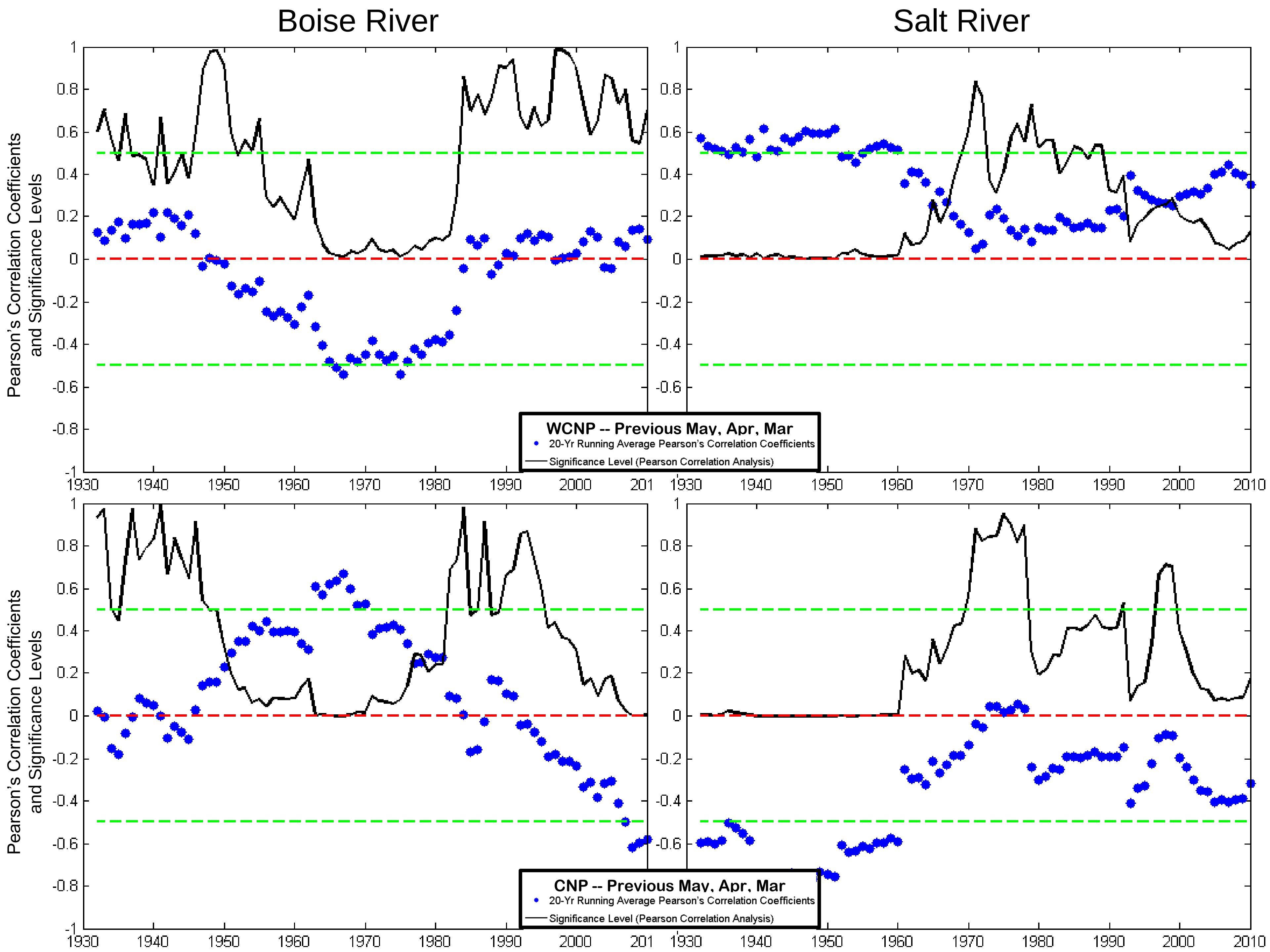
- Significant variations appear through out the record
  - Numerous changes in both sign and strength occur throughout the record.
  - Areas of strongest influence (upon streamflow into Lucky Peak) appear to be transitory, moving locations and even ocean basins over the record.

- Significant variations appear through out the record
  - Numerous changes in both sign and strength occur throughout the record.
  - Areas of strongest influence (upon streamflow into Lucky Peak) appear to be transitory, moving locations and even ocean basins over the record.

- To look at specific changes to correlations to the SSTs, an analysis similar to that done with the SOI and PDO was conducted with the following areas.
    - Results shown for areas highlighted in bright yellow
    - Shown as previously displayed.

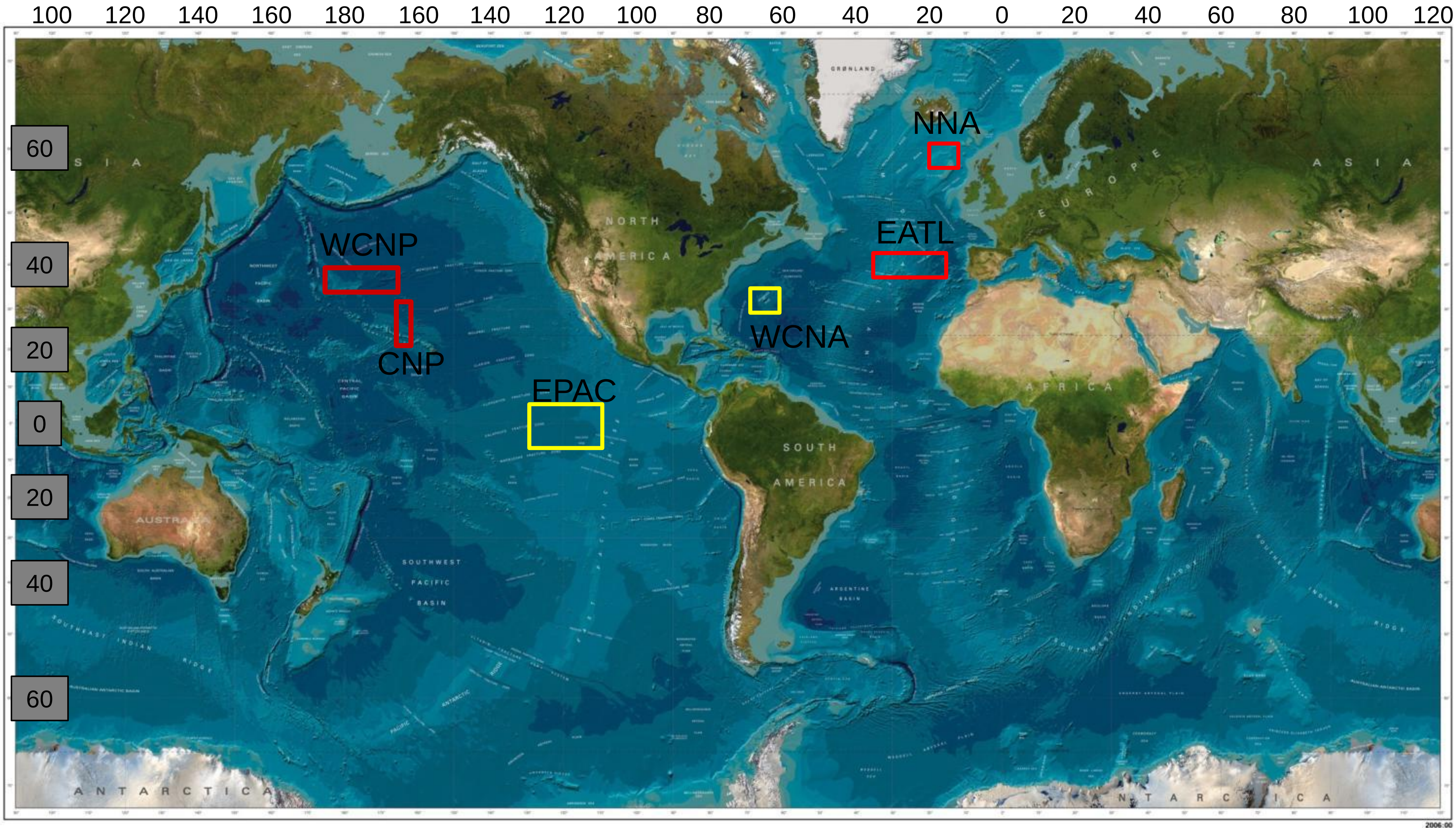
# Sea Surface Temperature Locations

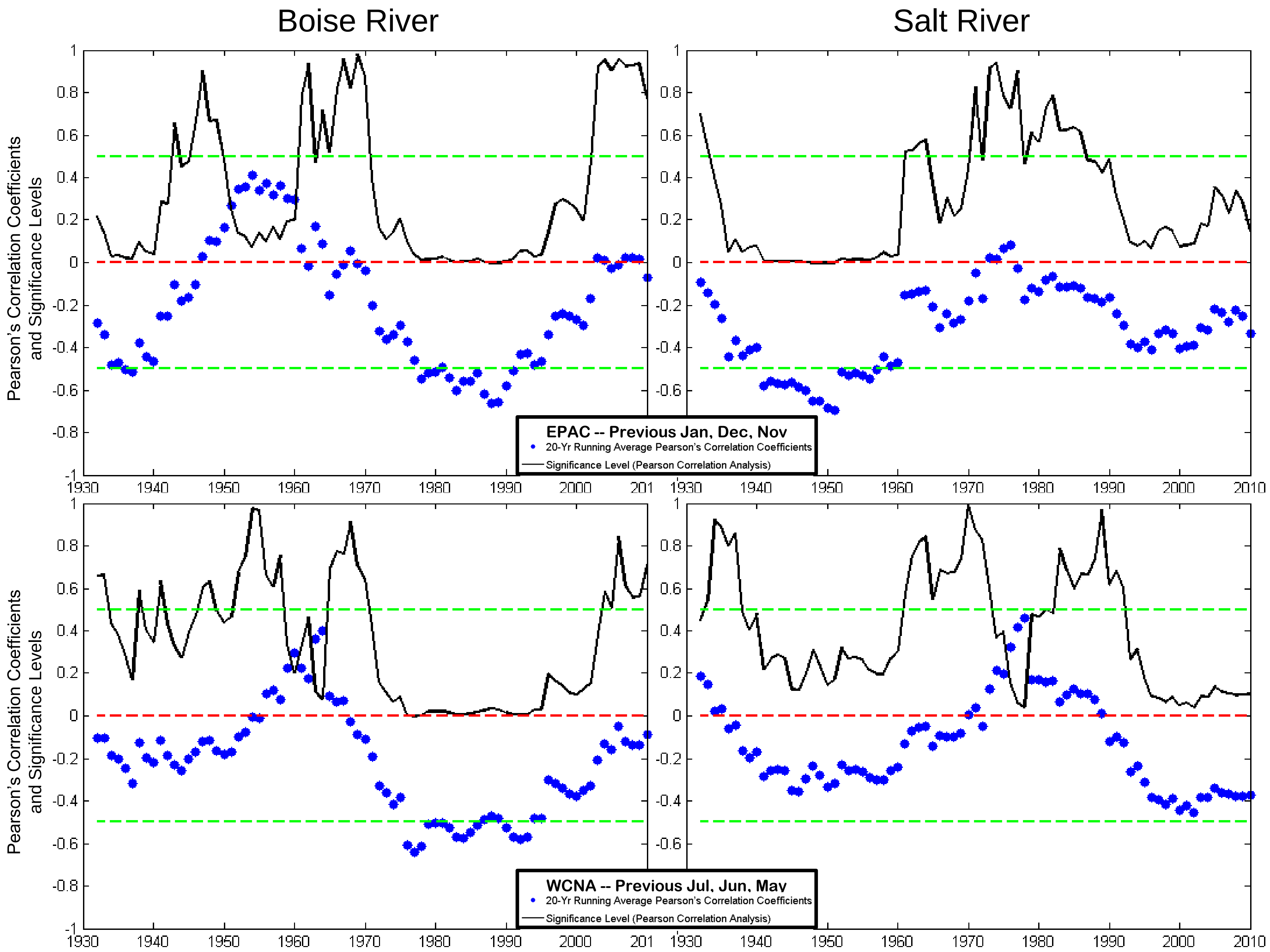




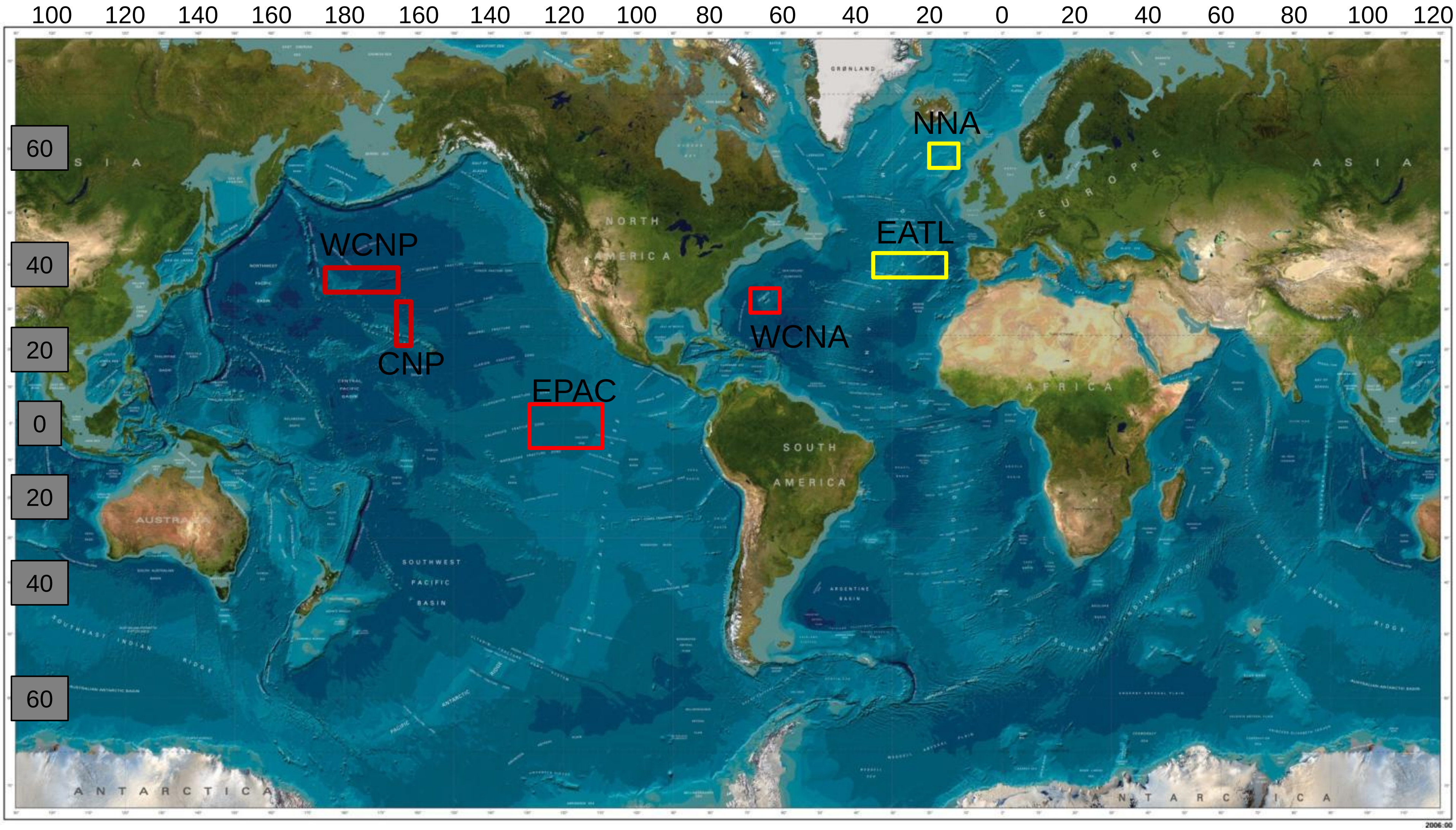
Boise River
Salt River
Pearson's Correlation Coefficients and Significance Levels
WCNP -- Previous May, Apr, Mar
20-Yr Running Average Pearson's Correlation Coefficients
Significance Level (Pearson Correlation Analysis)
CNP -- Previous May, Apr, Mar
20-Yr Running Average Pearson's Correlation Coefficients
Significance Level (Pearson Correlation Analysis)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
1930
1940
1950
1960
1970
1980
1990
2000
201
2010

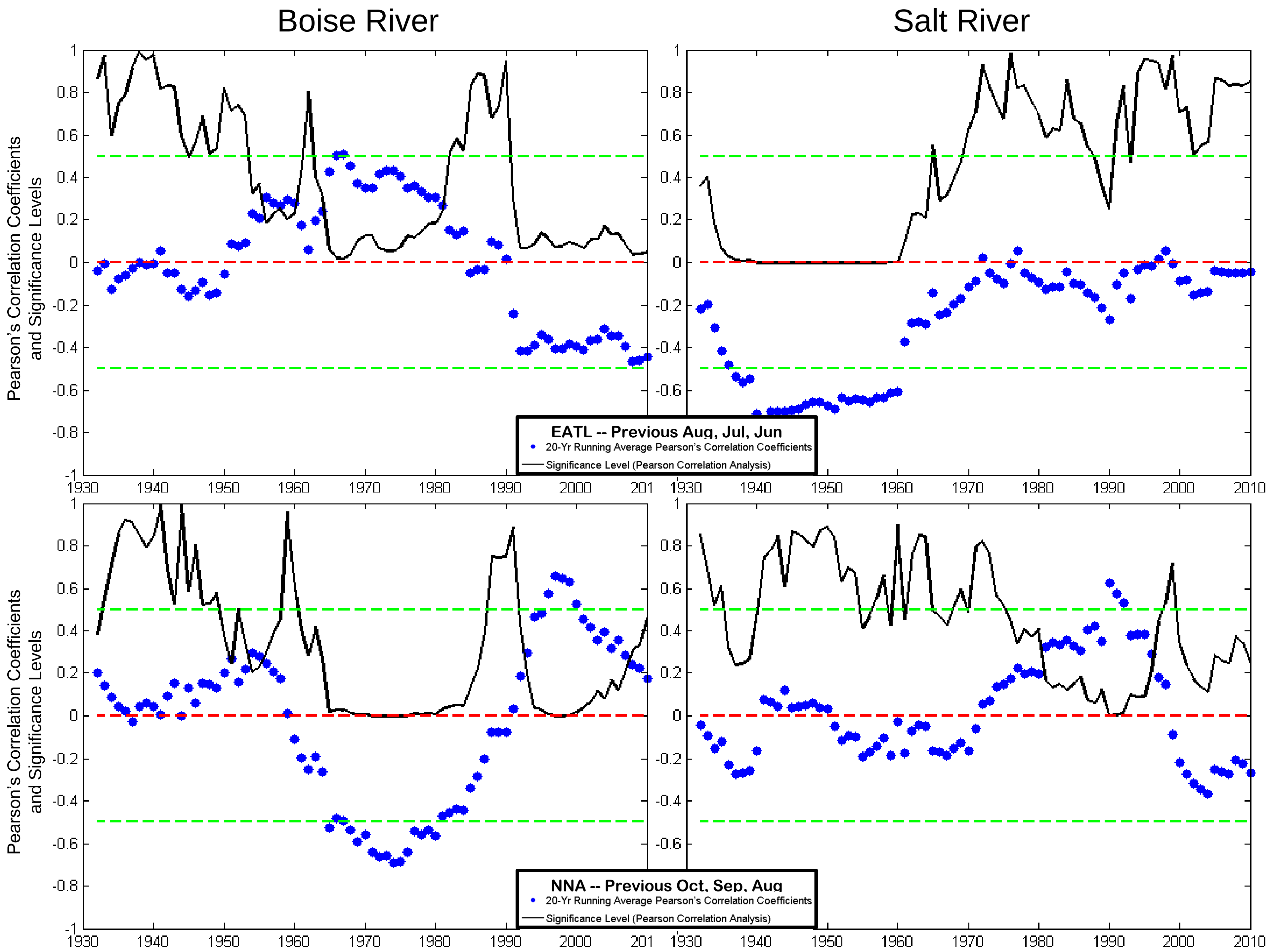
# Sea Surface Temperature Locations




Boise River
Salt River
Pearson's Correlation Coefficients and Significance Levels
EPAC -- Previous Jan, Dec, Nov
20-Yr Running Average Pearson's Correlation Coefficients
Significance Level (Pearson Correlation Analysis)
WCNA -- Previous Jul, Jun, May
20-Yr Running Average Pearson's Correlation Coefficients
Significance Level (Pearson Correlation Analysis)
1930
1940
1950
1960
1970
1980
1990
2000
2010

# Sea Surface Temperature Locations





Boise River
Salt River
Pearson's Correlation Coefficients and Significance Levels
EATL -- Previous Aug, Jul, Jun
20-Yr Running Average Pearson's Correlation Coefficients
Significance Level (Pearson Correlation Analysis)
NNA -- Previous Oct, Sep, Aug
20-Yr Running Average Pearson's Correlation Coefficients
Significance Level (Pearson Correlation Analysis)

# I found that teleconnections are frequently ephemeral through time!!

- ➢ SSTs being significantly and highly correlated with streamflow during some time periods and not during others.

# Key Questions

- Can we exploit correlations between large scale ocean/atmospheric indexes and Idaho and western US Watershed(s)?

- Can techniques be developed for forecasting basin-wide climate characteristics that can be employed throughout the watersheds of the Western United States **at long lead times**?

# Assumptions

- Development of statistical streamflow (and precipitation) models using teleconnections <u>typically</u> assumes:
  - Teleconnections are stationary through time
  - Developed models will maintain efficacy independent of any climate changes that may occur.

# Data Used

## Predefined indexes:

ENSO
PDO
AMO
PNA
AO
NAO

## Physically Defined Parameters:

SST
Atmospheric Conditions
Historic Flow Data
Historic Precipitation Data
Historic Temperature Data

# Correlation Analysis

- Series of correlation selection/techniques applied
  - Critical correlation coefficient selection
  - Summed annual/seasonal correlation coefficient selection
  - Superior interaction correlation coefficient (SICC) selection
    - Based upon
      - Maximum Adjusted $r^2$ and/or minimum RMSE with minimum number of predictors
      - Mallow's Cp to act as a stopping rule (reduce possible multicollinearity)
      - Desired Adjusted $r^2$
- SICC provides best end results

# Model Development

- Determine skill level desired
  - Selected 95% as target based upon desire to see if a useable model could be developed at that skill level.

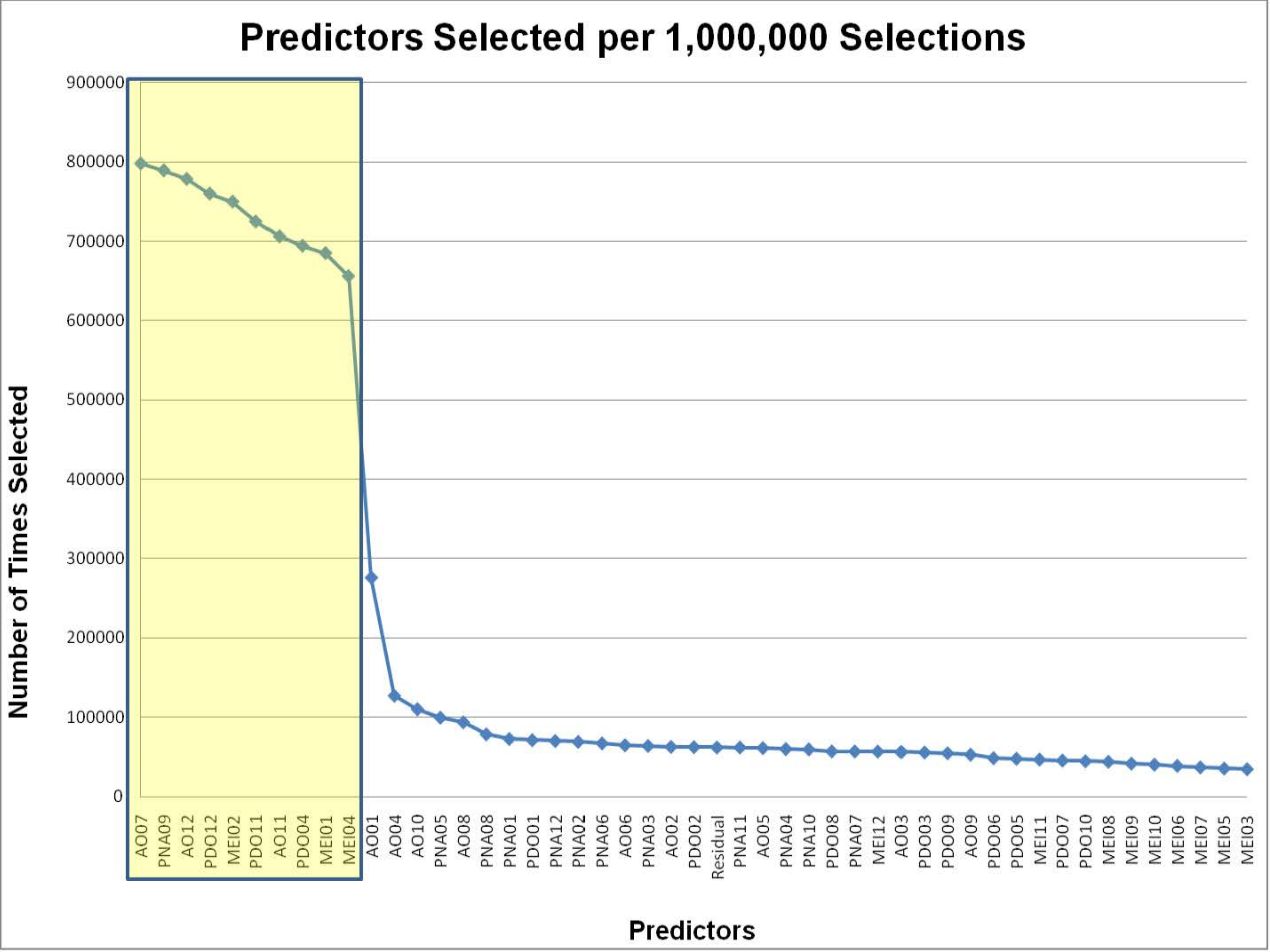
# Model Development

- Determine skill level desired
  - ➢ Selected 95% as target based upon desire to see if a useable model could be developed at that skill level.

| Number in Model | Adjusted R-Square | Variables in Model |
|---|---|---|
| 1 | 0.3255 | AO01 |
| 2 | 0.5093 | AO01 AO04 |
| 3 | 0.6287 | AO01 AO04 AO10 |
| 4 | 0.7271 | PDO12 AO01 AO04 AO10 |
| 5 | 0.7682 | Residual PDO12 AO01 AO04 AO10 |
| 6 | 0.8274 | MEI09 MEI10 MEI11 PDO08 PNA07 AO03 |
| 7 | 0.8728 | MEI09 MEI10 MEI11 PDO08 PDO12 PNA07 AO03 |
| 8 | 0.9241 | MEI09 MEI10 MEI11 PDO01 PDO12 PNA07 AO03 AO07 |
| 9 | 0.9496 | Residual MEI04 MEI08 PDO02 PDO05 PDO08 PNA07 AO04 AO06 |
| 10 | 0.9822 | Residual MEI01 MEI05 MEI08 PDO03 PDO05 PDO08 PNA07 AO04 AO06 |

# Model Development

- Predictor selection
  - Using the predictor number (N) identified by SICC
    - Run Monte Carlo series to identify the “N” predictors selected most often
    - ~1,000,000 renditions
    - Computationally very expensive



Predictors Selected per 1,000,000 Selections
Number of Times Selected
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
AO07
PNA09
AO12
PDO12
MEI02
PDO11
AO11
PDO04
MEI01
MEI04
AO01
AO04
AO10
PNA05
AO08
PNA08
PNA01
PDO01
PNA12
PNA02
PNA06
AO06
PNA03
AO02
PDO02
Residual
PNA11
AO05
PNA04
PNA10
PDO08
PNA07
MEI12
AO03
PDO03
PDO09
AO09
PDO06
PDO05
MEI11
PDO07
PDO10
MEI08
MEI09
MEI10
MEI06
MEI07
MEI05
MEI03
Predictors

# Model Development

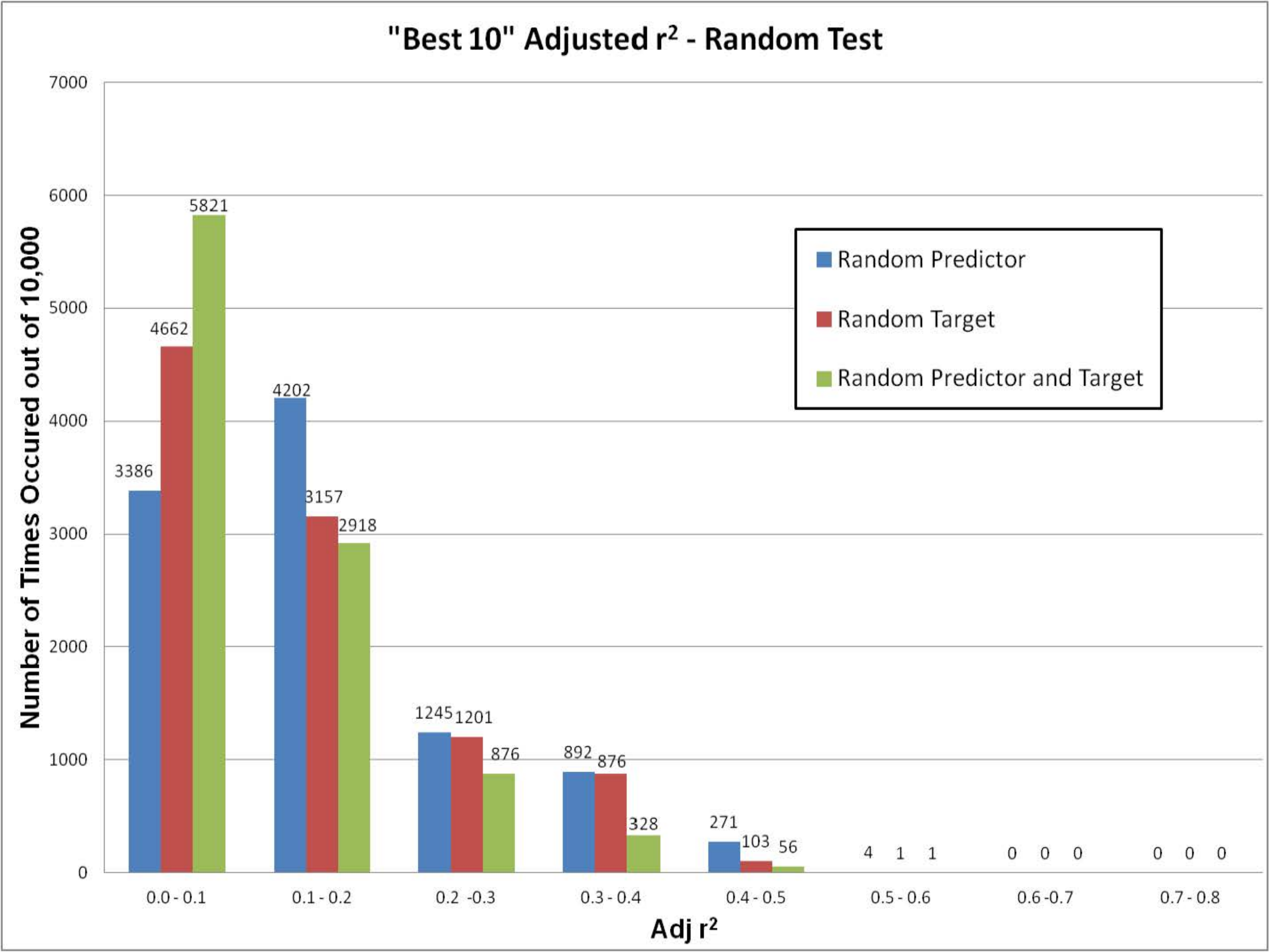
- Predictor selection
  - ➢ Using the predictor number (N) identified by SICC
    - Run Monte Carlo series to identify the N predictors selected most often
    - ~1,000,000 renditions
    - Computationally very expensive
- Develop regression equations
  - ➢ Take “Best N” predictors, run MC routine
    - subsets from 0.85 to 0.45 (~1,000,000)
    - Record averages/median for each run parameter estimates
- Apply to calibration and validation sets
  - ➢ Compute differences
  - ➢ Select parameter estimate (average vs. median) with best overall results

# Model Validation

- Cross validation
  - ➢ Leave-one-out cross validation

- Random data/development validation
  - ➢ Developed random “predictor” and “target” sets
    - Based upon original predictor/target sets observed averages, variances and co-variances

  - ➢ Redevelop models with random data sets (random predictor set vs. original target set; original predictor set vs. random target set; and random predictor set vs. random target set).
    - Apply new models with and record adjusted $R^2$
    - Calculate number/percent of times in $R^2$ zones (i.e. <0.10, 0.10 – 0.20, etc...)


"Best 10" Adjusted r2 - Random Test
Number of Times Occured out of 10,000
7000
6000
5000
4000
3000
2000
1000
0
5821
4662
3386
4202
3157
2918
1245
1201
876
892
876
328
271
103
56
4 1 1
0 0 0
0 0 0
0.0 - 0.1
0.1 - 0.2
0.2 -0.3
0.3 - 0.4
0.4 - 0.5
0.5 - 0.6
0.6 -0.7
0.7 - 0.8
Adj r2
Random Predictor
Random Target
Random Predictor and Target

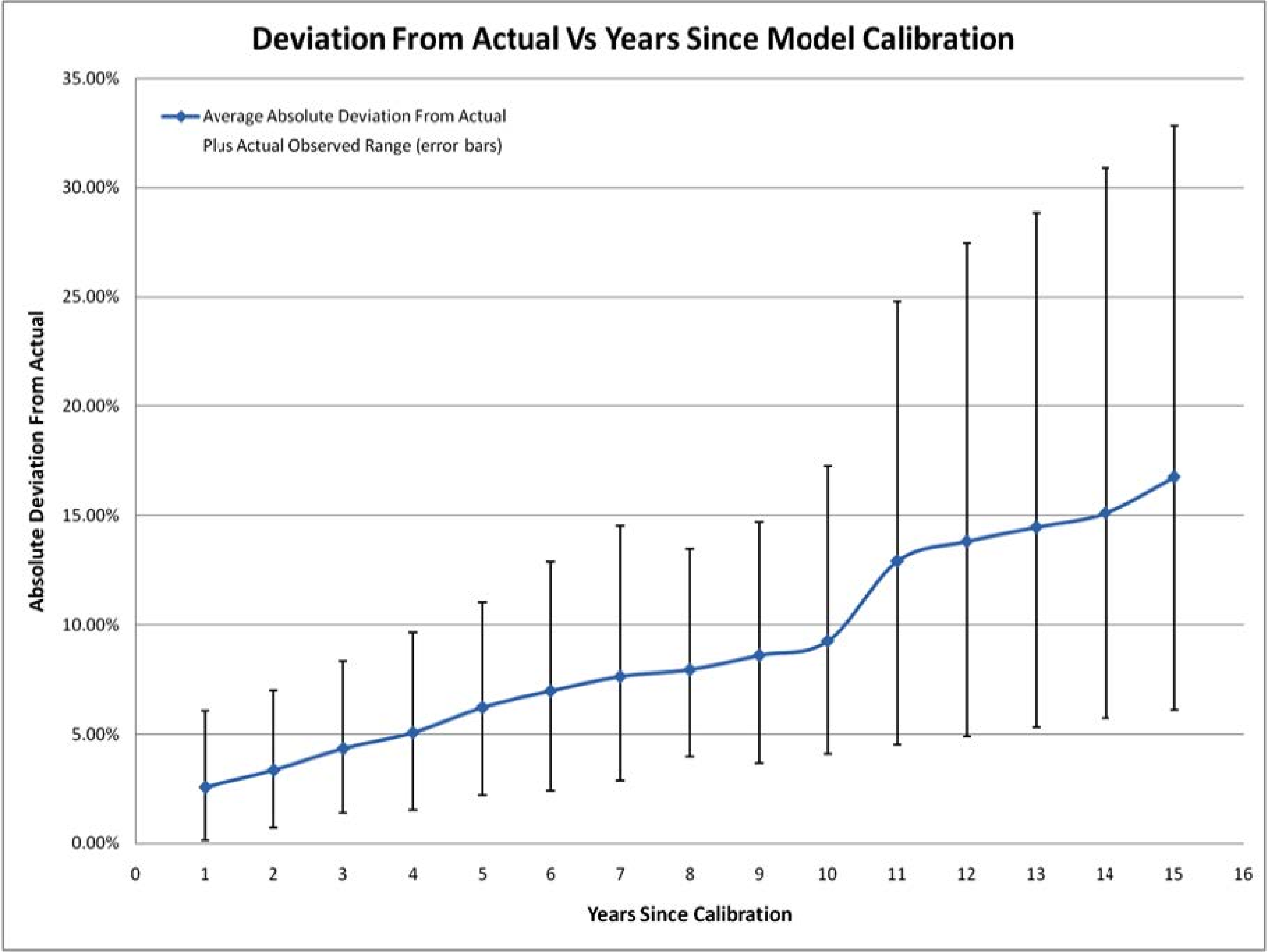
# How well does the method work in the long-term?

- In an effort to establish how effective the developed modeling technique may be through time:

    - **<u>100</u>** forecast models developed (based upon varying lengths of calibration years and starting years).

    - Models applied out 15 years from the end of the calibration period.

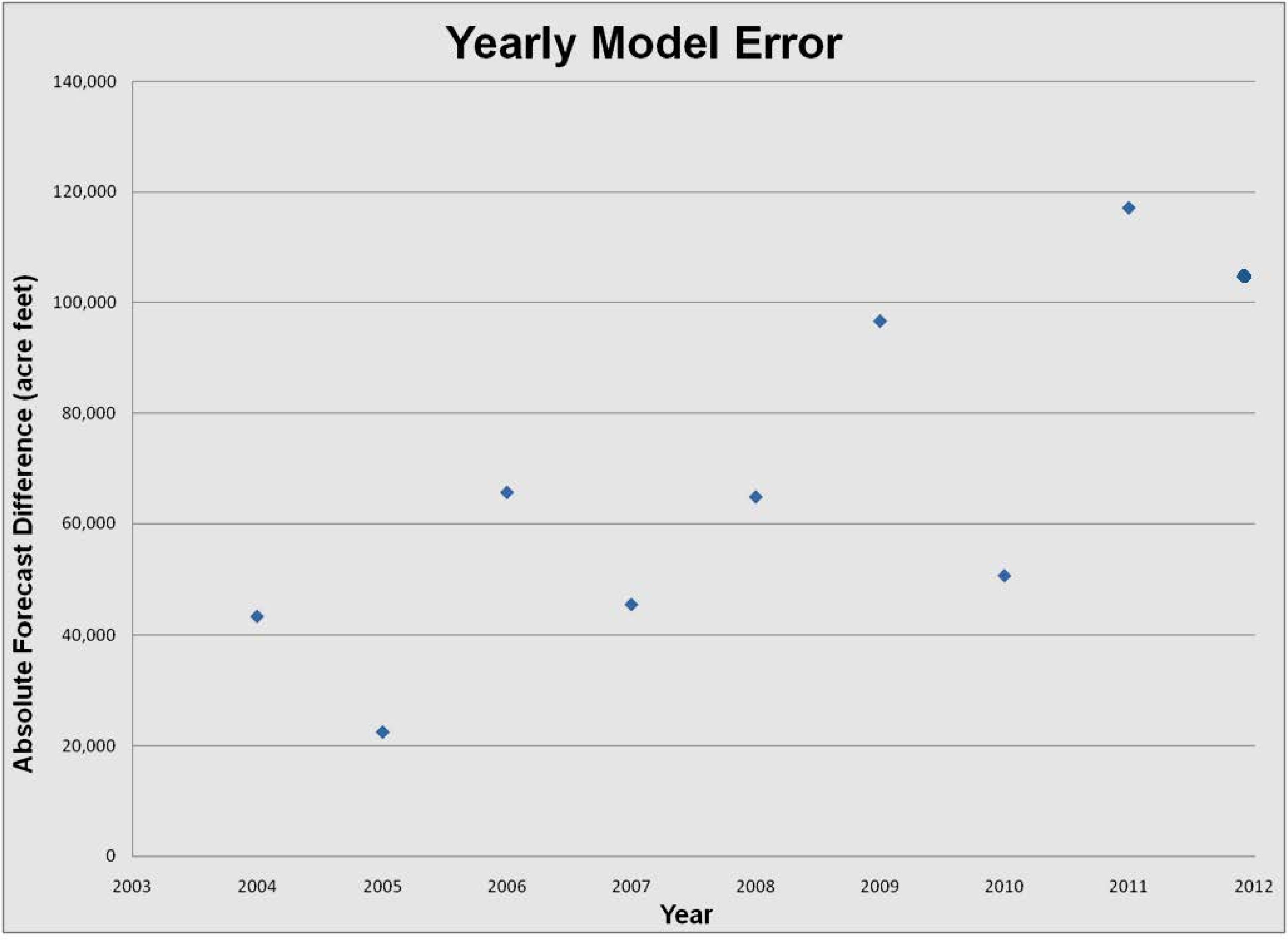
    - Results indicate the models generally perform well during the first few years after calibration and then decline in performance quickly.



Deviation From Actual Vs Years Since Model Calibration
Average Absolute Deviation From Actual
Plus Actual Observed Range (error bars)
Absolute Deviation From Actual
Years Since Calibration
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

- Average model performance
  - ➢ Between 2.5 and 6.5% of actual flow values for the first five years after calibration
  - ➢ Decreasing in performance to an average of 16.8% by the $15^{th}$ year.

- Average values do not show the complete picture!!
  - ➢ Some models performance was less than 6% of actual flow in the first year and decreases to more than 11% by the $5^{th}$ year.
  - ➢ For the forecast period 11 to 15 years after calibration, some models missed the forecast by >25%.


Yearly Model Error
Absolute Forecast Difference (acre feet)
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
Year

# The Model Development and Results

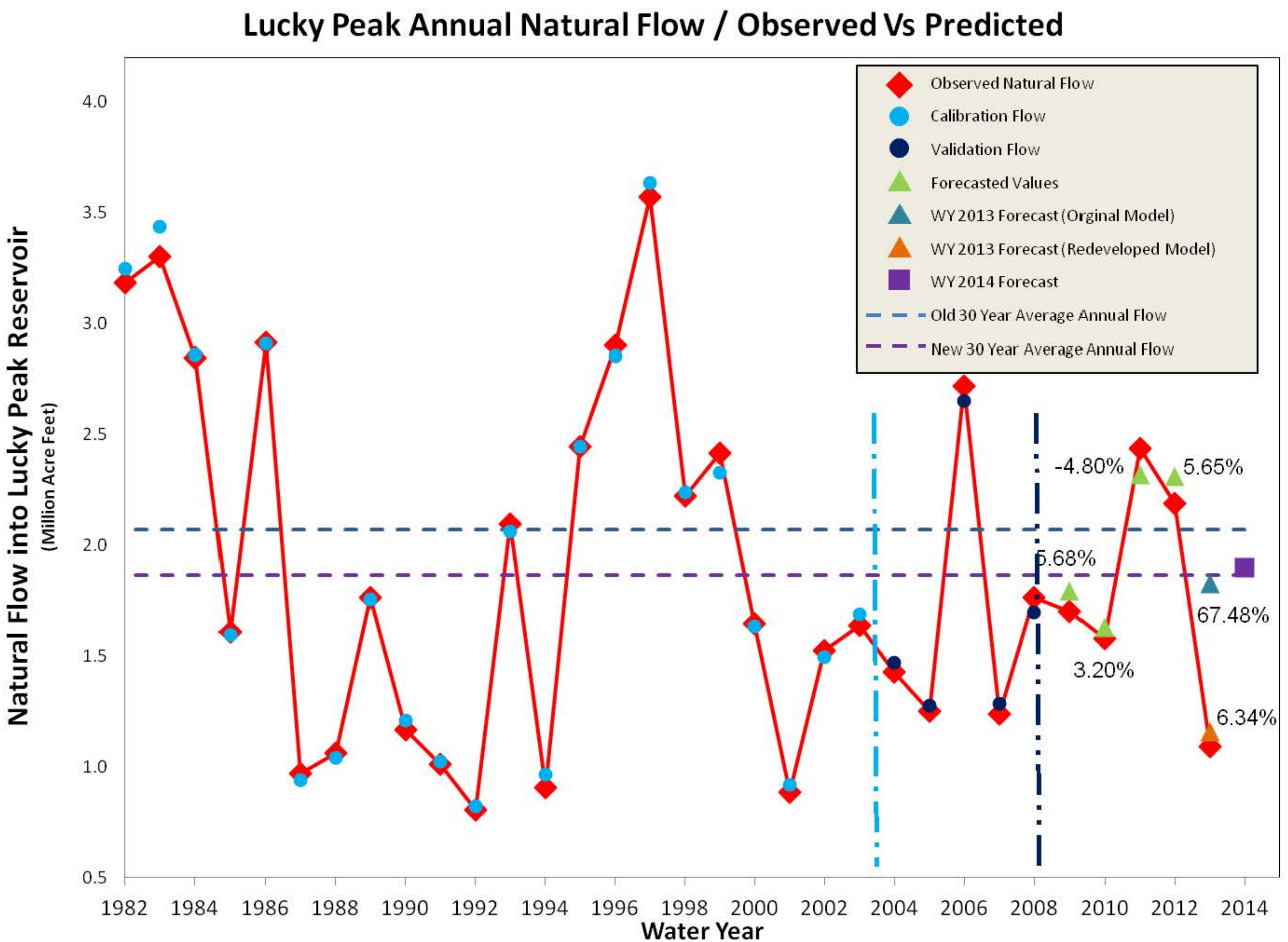
- Model developed with data from 1982 – 2003 (Calibration)
  - Model was validated using data from 2004 – 2008 (Validation)
  - Forecast issued from this model 2009 – 2013
  - Model redeveloped in 2012 for the 2013 forecast

- All forecasts issued near the end of October when all teleconnection data available and covers the entire water year (October – September)

- Covers
  - WY Annual (Oct-Sep)

- Similar results for annual precipitation and temperature forecasts.

- Working on development of Seasonal Forecasts (Oct-Dec, Jan-Mar, Apr-Jun, Jul-Sep)



Lucky Peak Annual Natural Flow / Observed Vs Predicted
Observed Natural Flow
Calibration Flow
Validation Flow
Forecasted Values
WY 2013 Forecast (Orginal Model)
WY 2013 Forecast (Redeveloped Model)
WY 2014 Forecast
Old 30 Year Average Annual Flow
New 30 Year Average Annual Flow
-4.80%
5.65%
5.68%
67.48%
3.20%
6.34%
Natural Flow into Lucky Peak Reservoir
(Million Acre Feet)
Water Year
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014

**Suggests periodic redevelopment of statistical models emphasizing "current" teleconnections may be the key to continued success for statistical streamflow model use!!!**

# Discussion

- Predictors
  - ➢ Concerns of over fitting the model
  - ➢ Concerns and problems with too many potential predictor sets
- Development of statistical streamflow models using teleconnections typically assumes:
  - ➢ Teleconnections are stationary through time
  - ➢ Models will maintain efficacy independent of any climate changes that may occur.
- Associated uncertainties

# Current Work

# Data

## Predefined indexes:

ENSO
PDO
AMO
PNA
AO
NAO
TNI

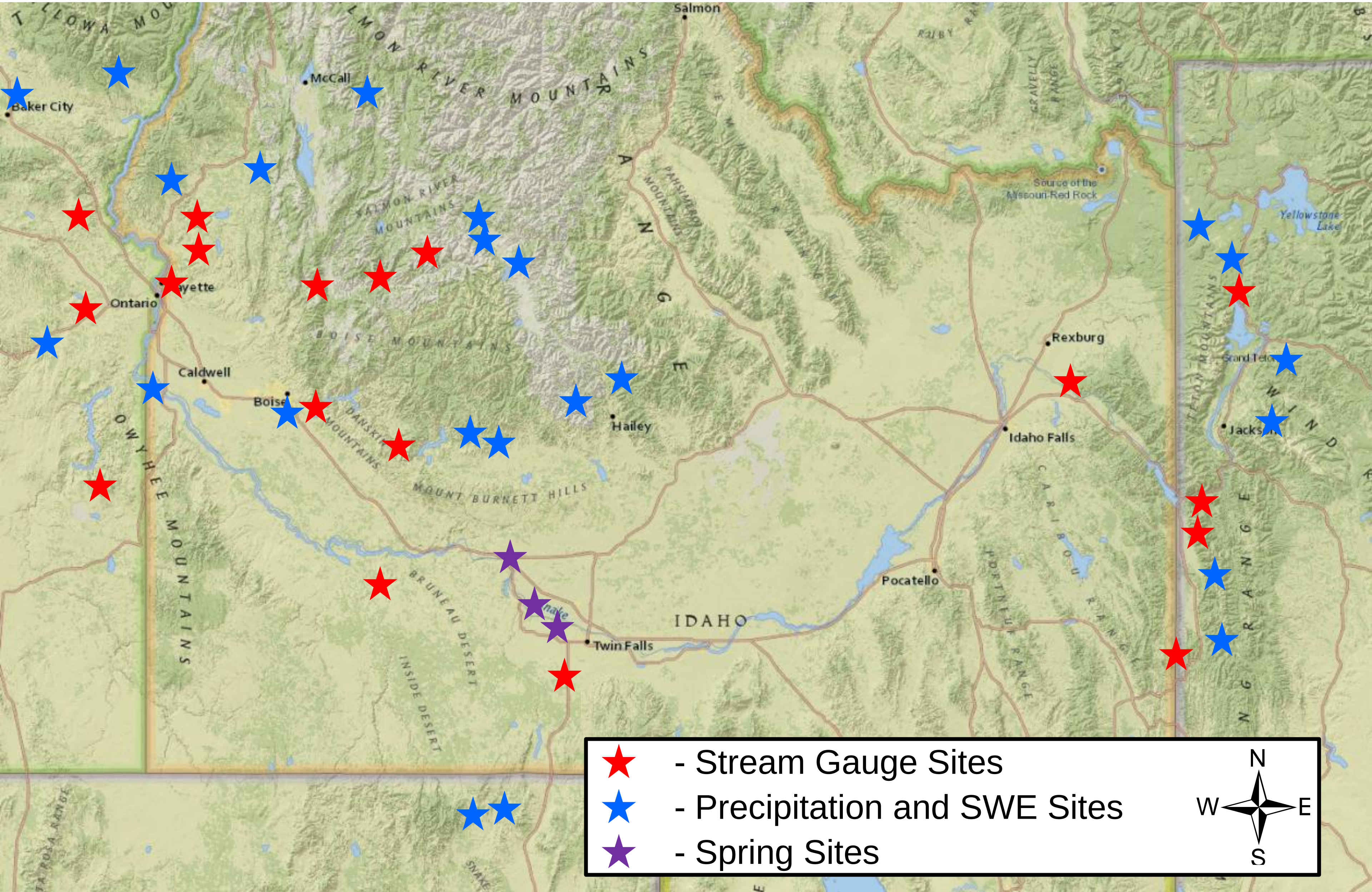
## Physically Defined Parameters:

SST
Atmospheric Conditions
Historic Flow/Reservoir Data
Historic Precipitation Data
Historic Temperature Data
Historic Well Data

## Would Be Nice to Have:

Historic Pumping Data
Historic Recharge Data

# Current Work



# Thanks!!