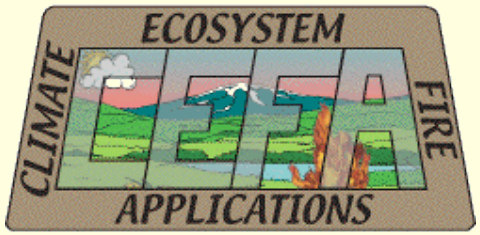
Quality Control of the National RAWS Database for FPA

Timothy J Brown
Beth L Hall

Desert Research Institute
Program for Climate, Ecosystem and Fire Applications
May 2005



DRI
Desert Research Institute

Project Objectives

- Run coarse QC on all RAWS sites
  - Remove any erroneous values
- Create a ‘complete’ dataset
  - Replace as much missing or erroneous values as possible
- Assess level of confidence in estimated values
  - Run validation tests

The Station List

- 1350 Total stations for over 120 FPUs
- 1075 potential RAWS matches
  - 1001 RAWS stations processed in Round 1
  - 74 not processed
    - no WIMS ID
    - Problems matching with WIMS data
- 275 non-RAWS matches
  - Manual station
  - METAR, AWOS, etc.

Issues with station lists

- Could not find matching WRCC RAWS data (275)
- Missing 6-digit NIFMID ID code (23)
- No NIFMID data to correlate to hourly RAWS (119)
- Periods of record between NIFMID and RAWS do not overlap (15)

Coarse QC

- RAWS parameters
  - Precipitation
  - Temperature
  - Relative Humidity
  - Wind Speed
  - Wind Direction

- Flag all values
  - Reasonable?
  - Missing?
  - Impossible?
    - Negative wind speed, humidity, direction, etc.
  - Questionable?
    - Wind speed > 120 mph; temperature < -40F
    - Unchanged values for too many consecutive hours

Create ‘Complete’ Datasets

- Correlate RAWS with Reanalysis
  - Reanalysis dataset
    - 2.5 degree spatial resolution
    - 6-hourly
    - Upper-air variables:
      - Temperature, humidity, u and v wind components
    - Surface variables:
      - Temperature, humidity, u and v wind components, downward long-wave radiation flux, downward short-wave radiation flux, precipitation rate, total cloud cover
    - Include 9 surrounding grid cells centered on RAWS site
- Use multiple regression output to estimate all missing, questionable, or erroneous data

Issues with Estimations

- Algorithms ‘blew up’ after many consecutive hours of estimation
  - Continuous trend for 20+ hours
  - All following hours made missing
- Estimation created bad value
  - Negative wind speeds, direction, precipitation
  - Values exceeded reasonable physical thresholds

Processing statistics

- 563 potential predictor variables
  - All reanalysis variables at…
    - 9 surrounding grid points
    - Pressure level variables at 8 levels of atmosphere
  - RAWS persistence variables for…
    - last hour
    - yesterday, same hour
  - Reduced to no more than 44

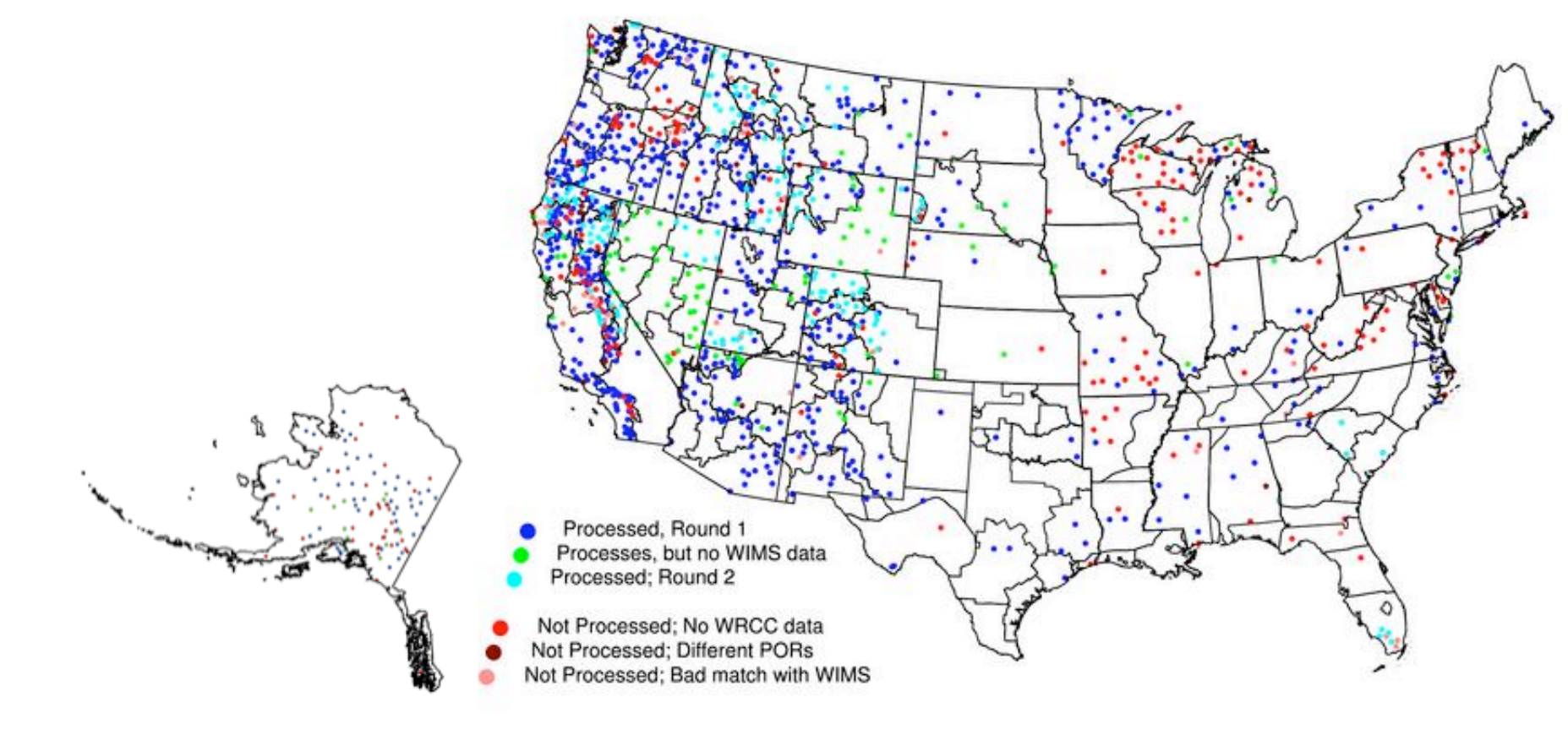
- 56 equations per stations
  - 7 variables, 4 time periods/day, 2 seasons
  - 56,056 equations

- 1.25 hours to process each station
  - Run consecutively on 1001 stations? Almost 2 months

Products

- Data sets from original station start date through 2004 (if possible)
  - Once-a-Day -- old 1972 NIFMID format (*.fwx)
  - Hourly -- new 1998 NIFMID format (*.fw9)
  - Comma delimited complete dataset with flags indicating value status (*.dat)
- Summary and Statistics
  - File that lists station status (could it be processed? Why or why not)
  - File that lists percentage of data that was estimated or had to be removed

Data processing status



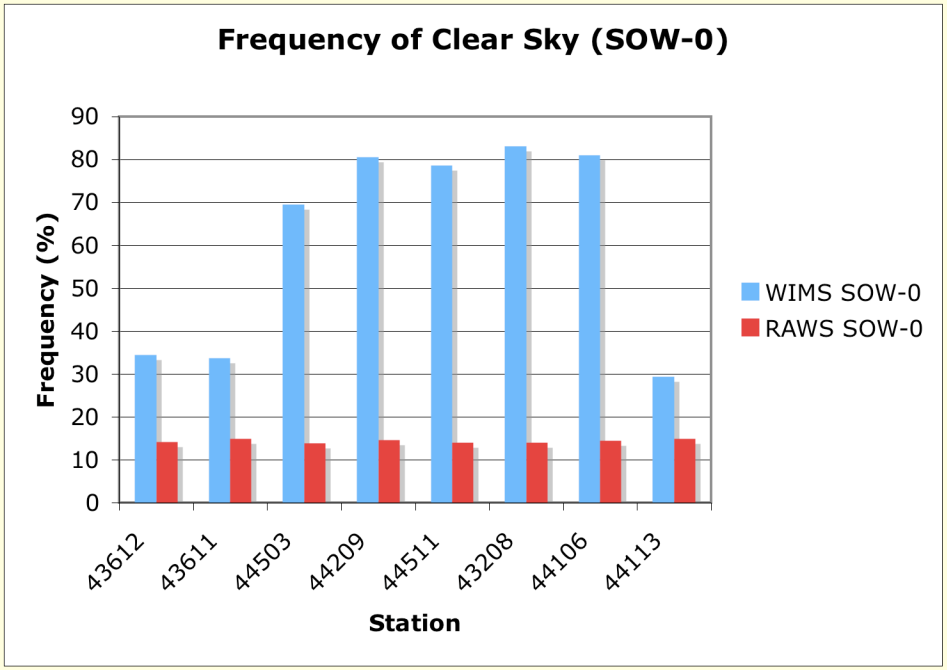
Validation of Estimations

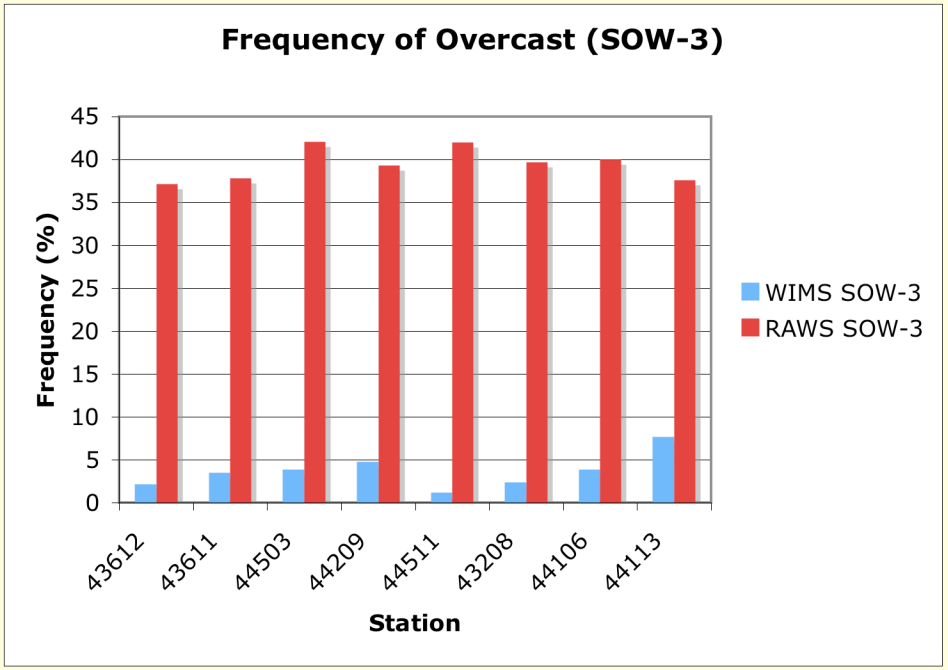
- State of the weather
- Other weather variables
  - Precipitation
  - Wind speed and direction
  - Temperature
  - Humidity
- Tested on several California stations

State of the Weather

- Did well for SOWs 4-9
  - Less than 10% of time
- Showed serious discrepancies for levels of cloudiness (SOW 0-3)
- Problem with current coarse reanalysis total cloud cover

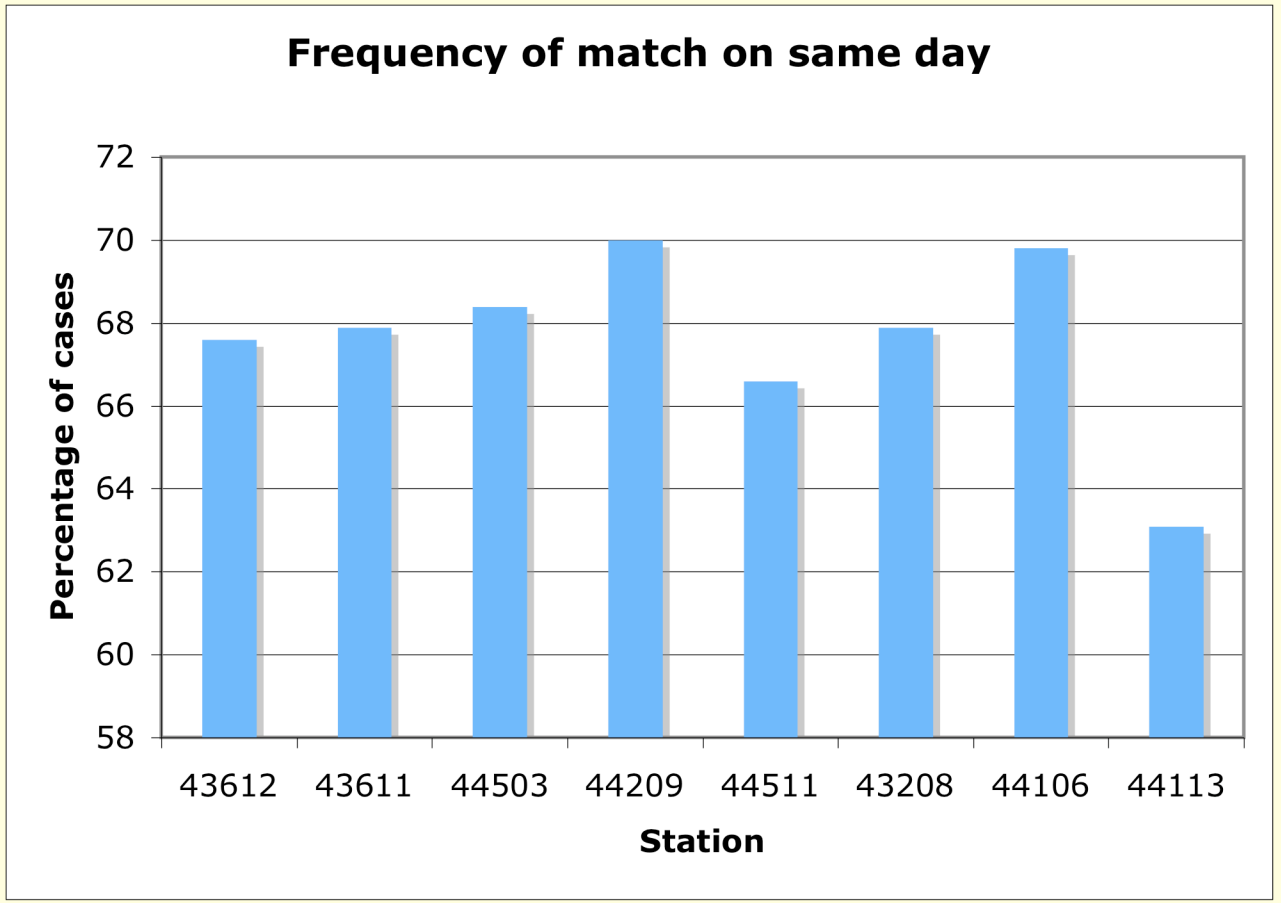
State of the Weather

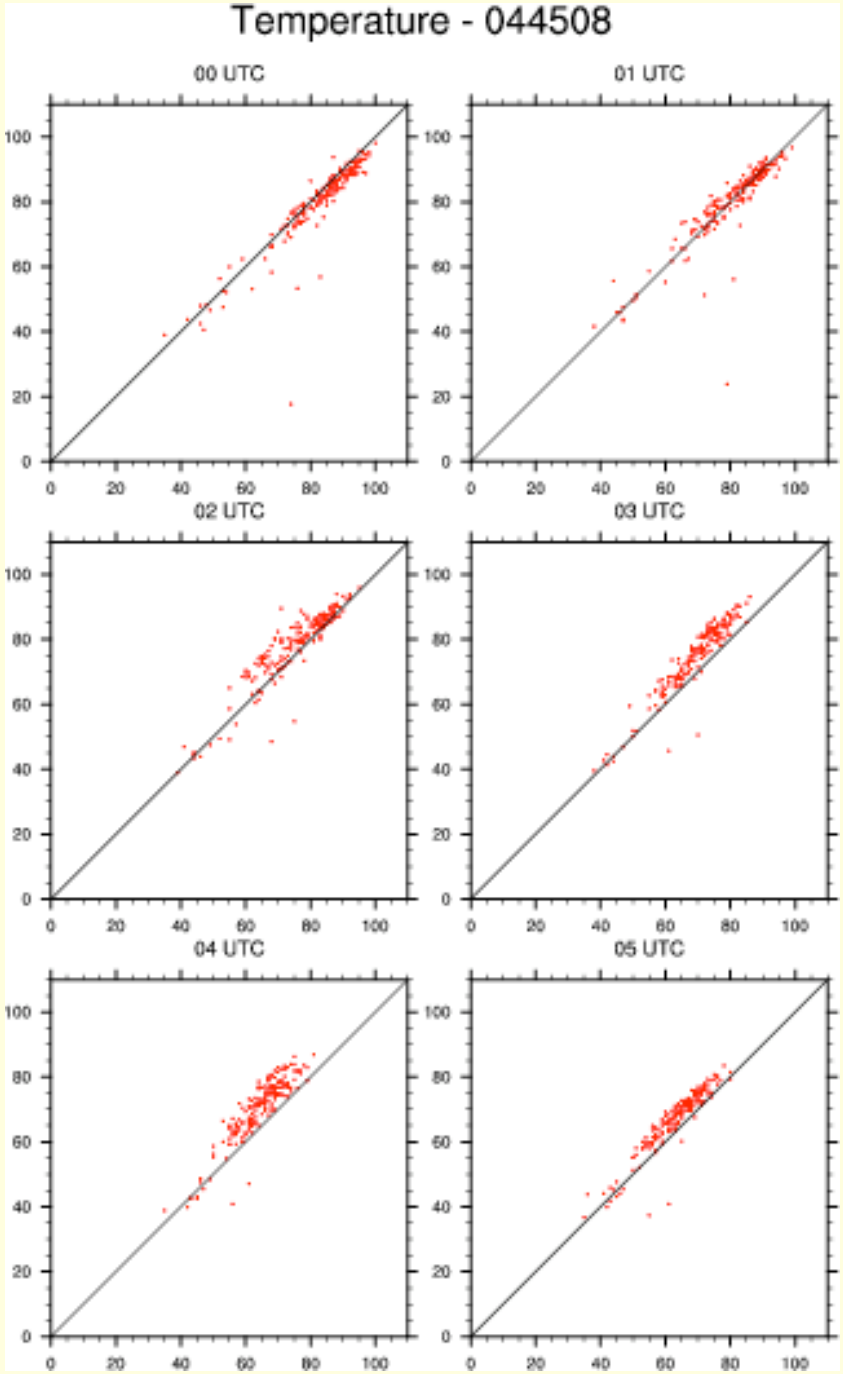




State of the Weather

- Did an estimated SOW between 10 AM and 5PM ever match with WIMS on same day?



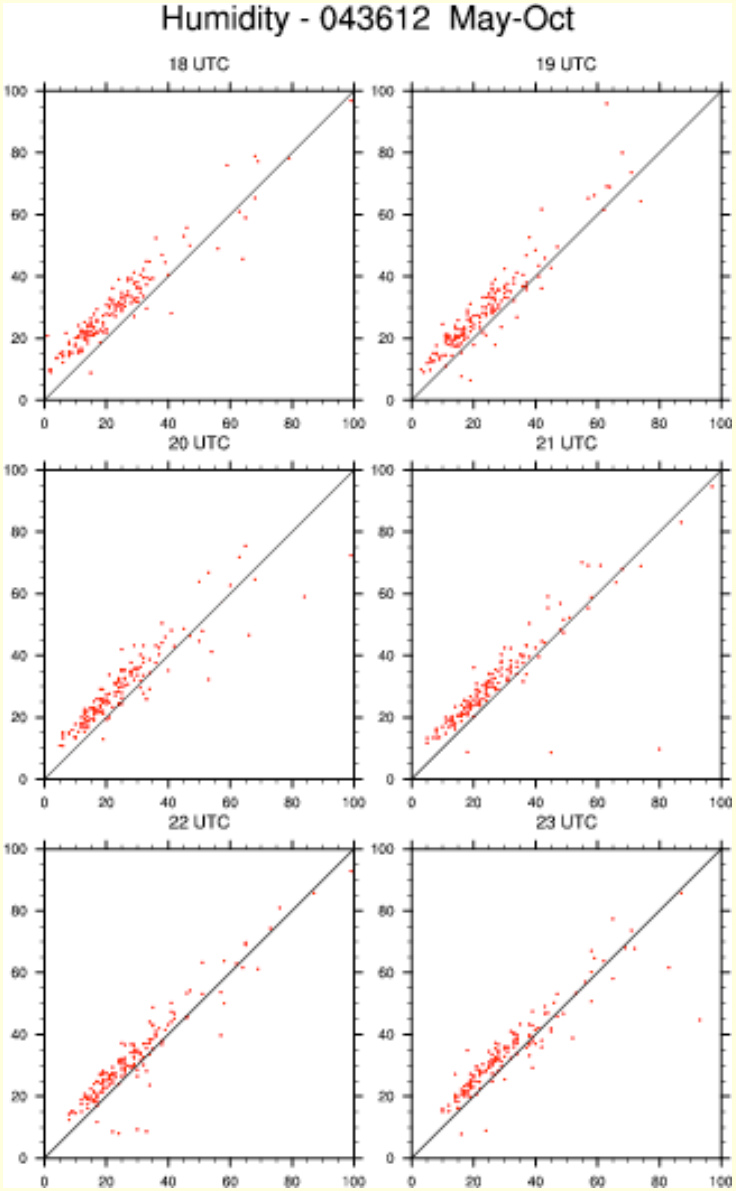


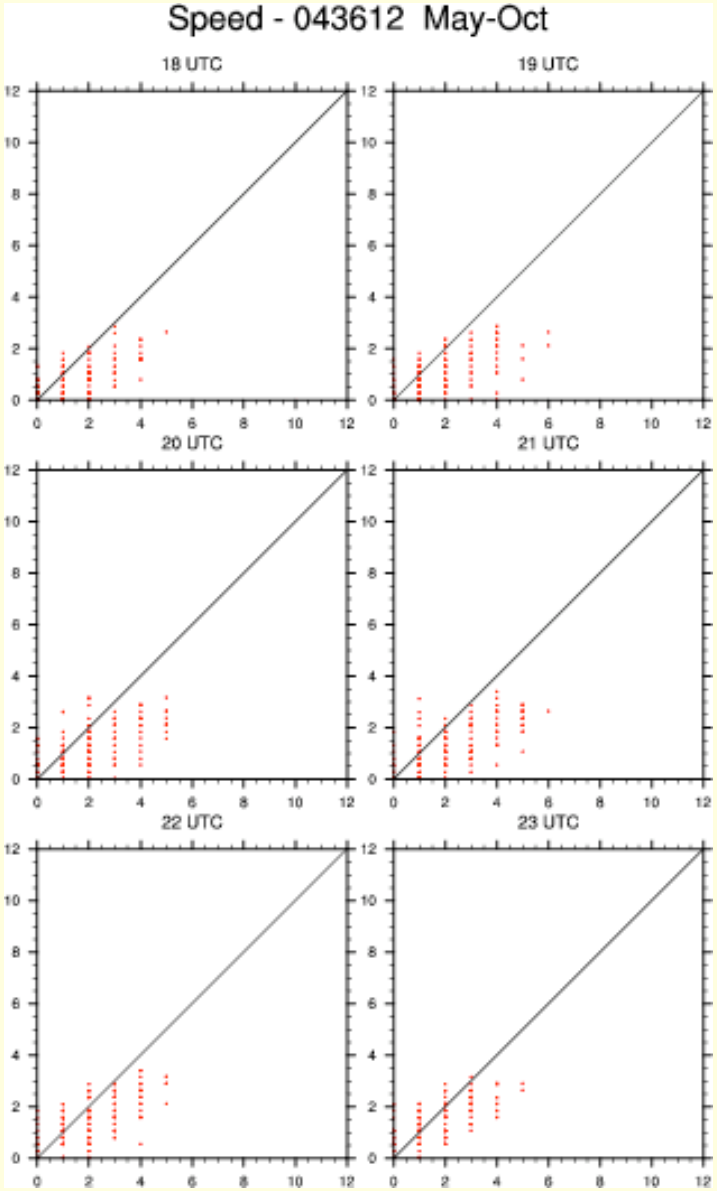
Weather Variables

Validation method

- Compute an estimate for days in 2003
- Compare to observations

Weather Variables

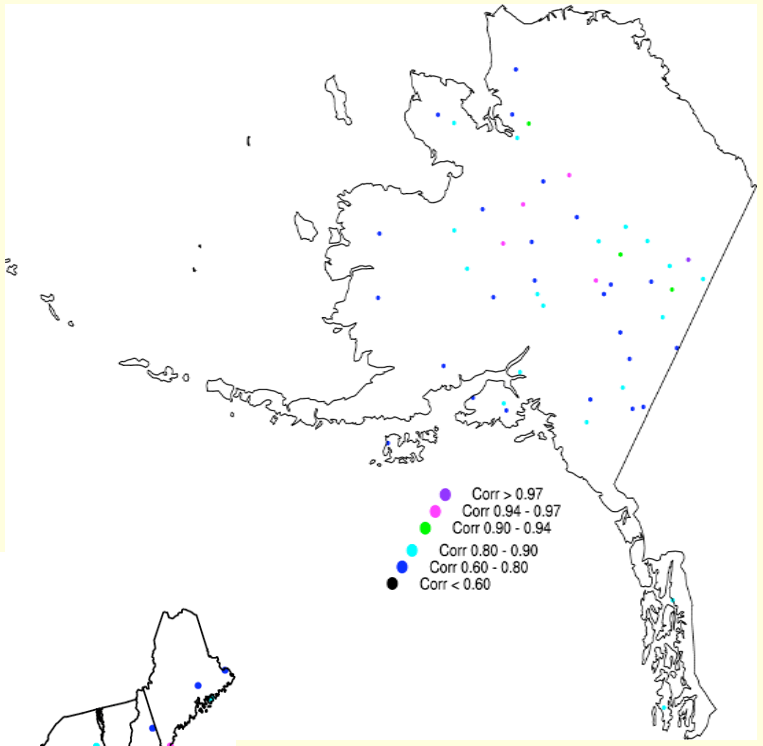


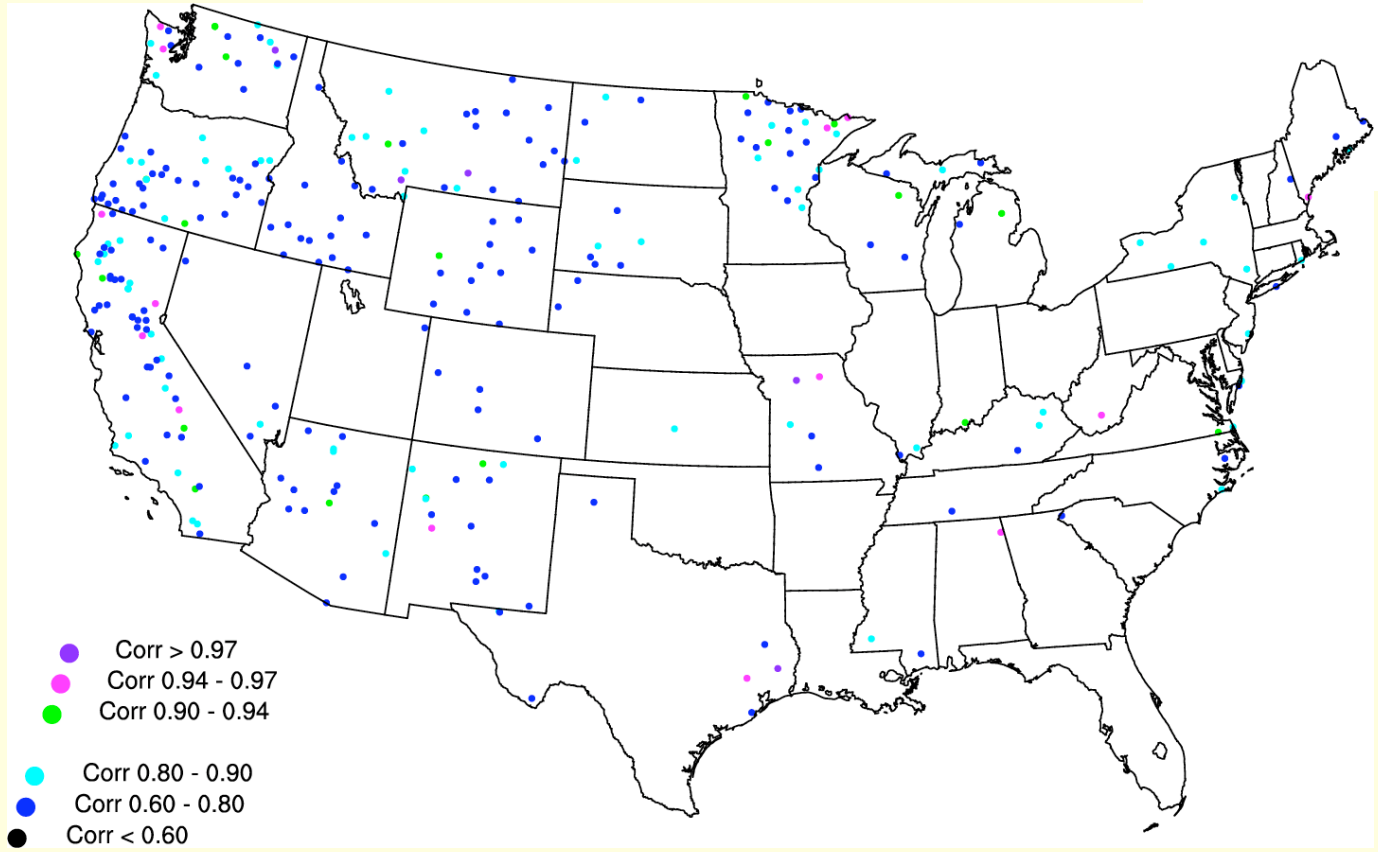


Weather Variables

Wind direction

1300 LT; May-Oct

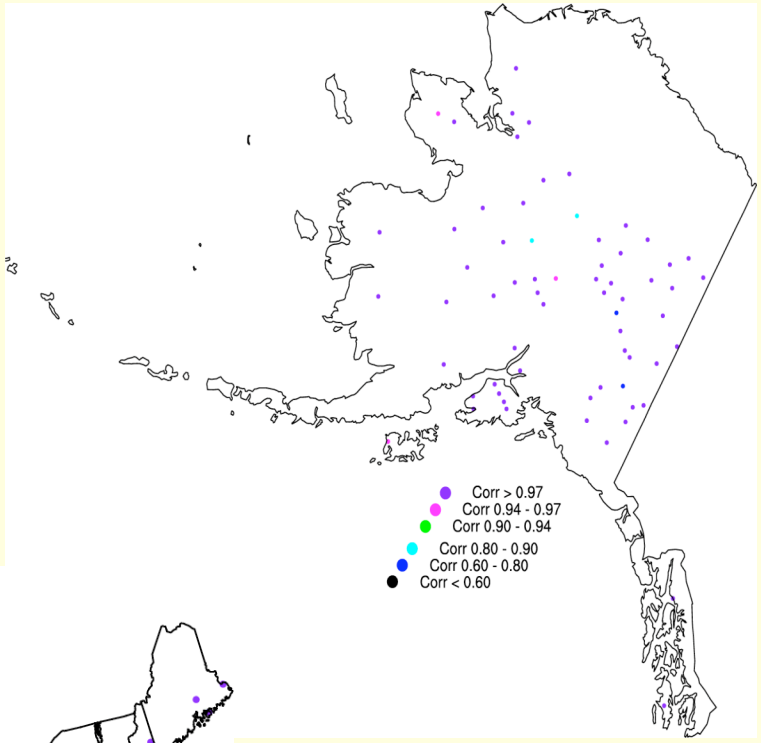


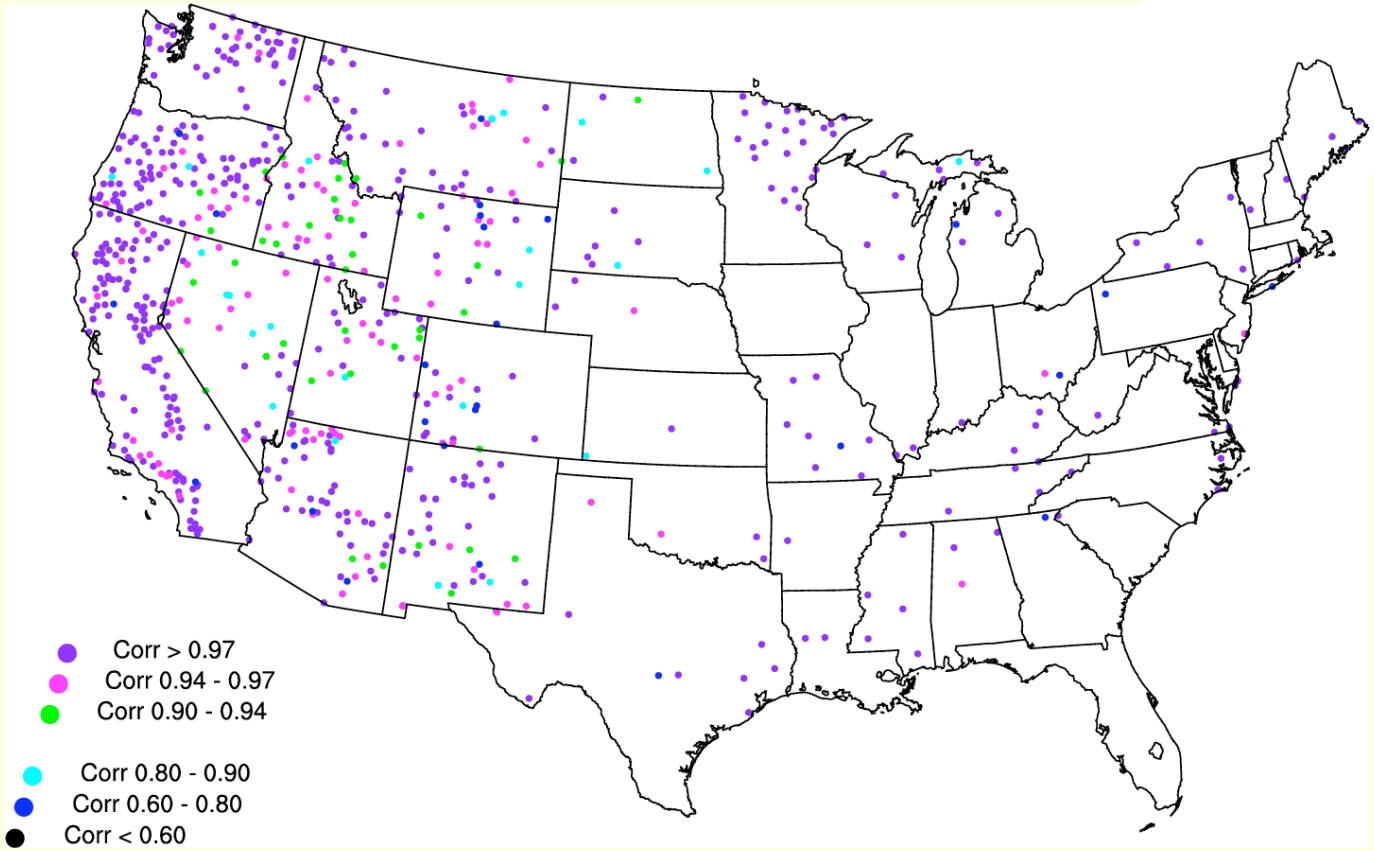


Weather Variables

Wind speed

1300 LT; May-Oct

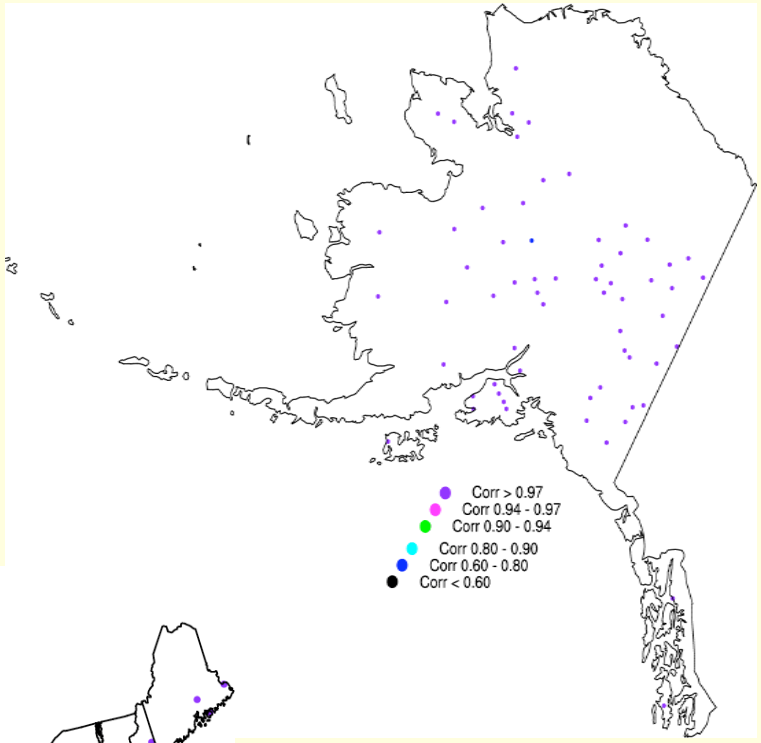


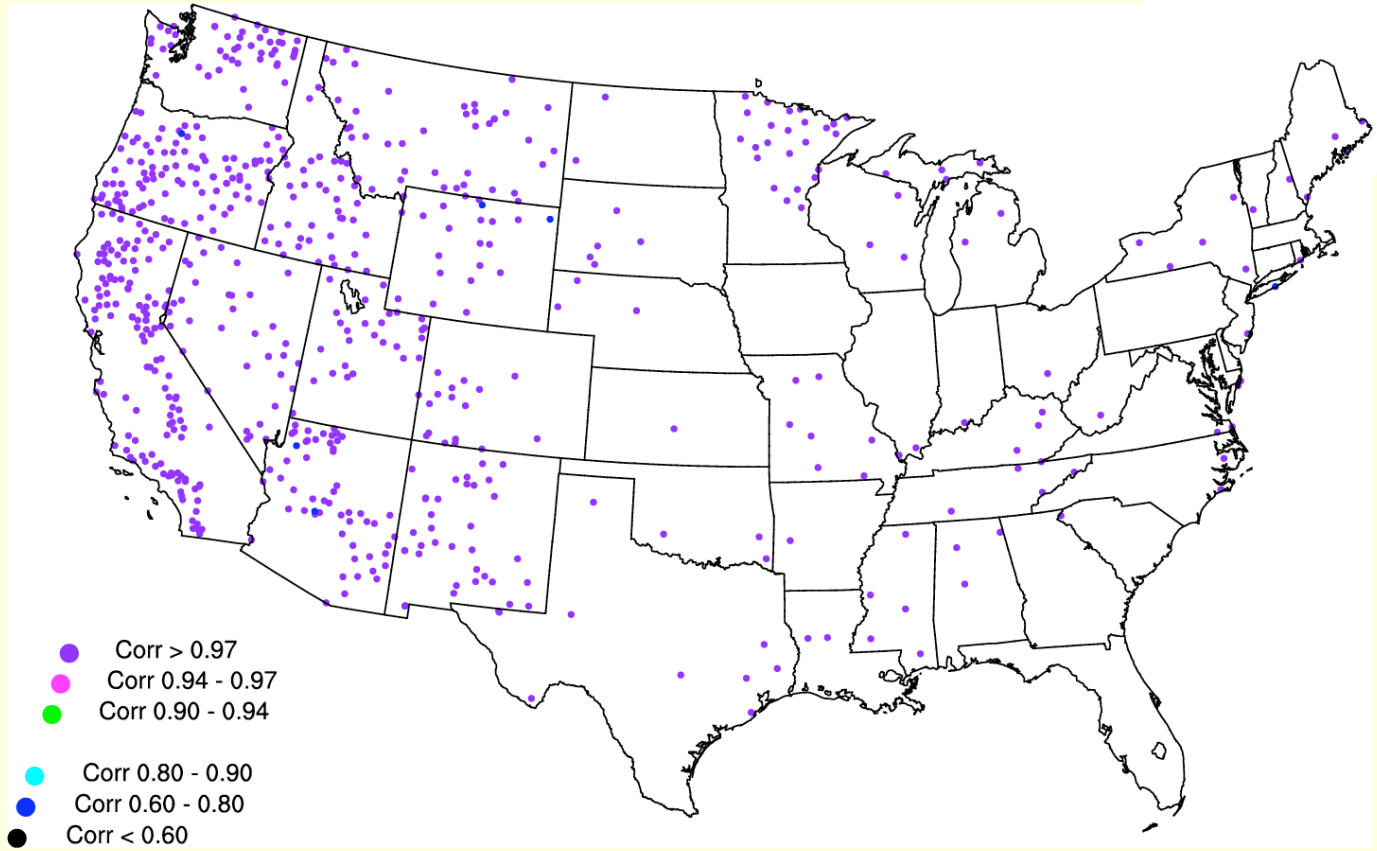


Weather Variables

Precipitation

1300 LT; May-Oct

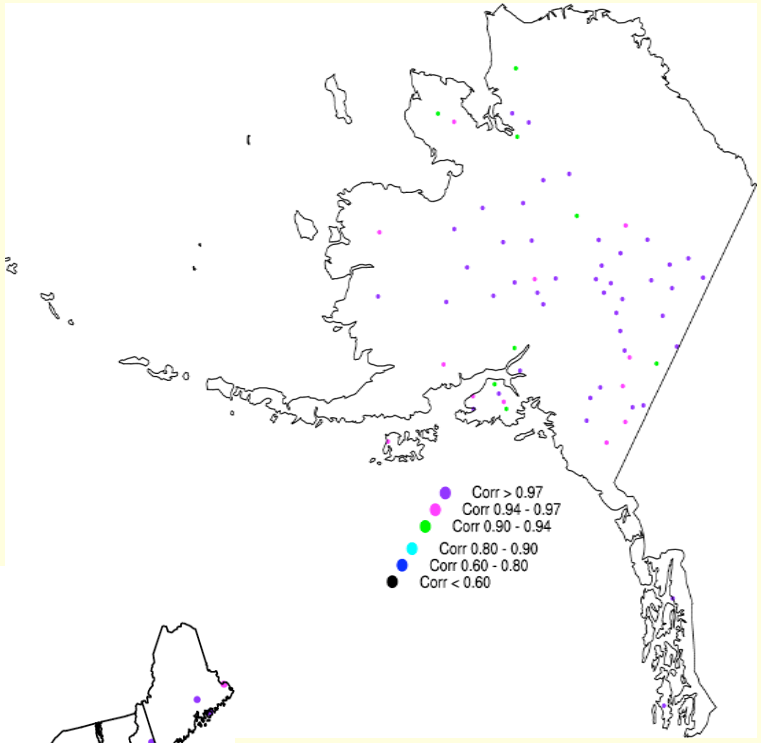


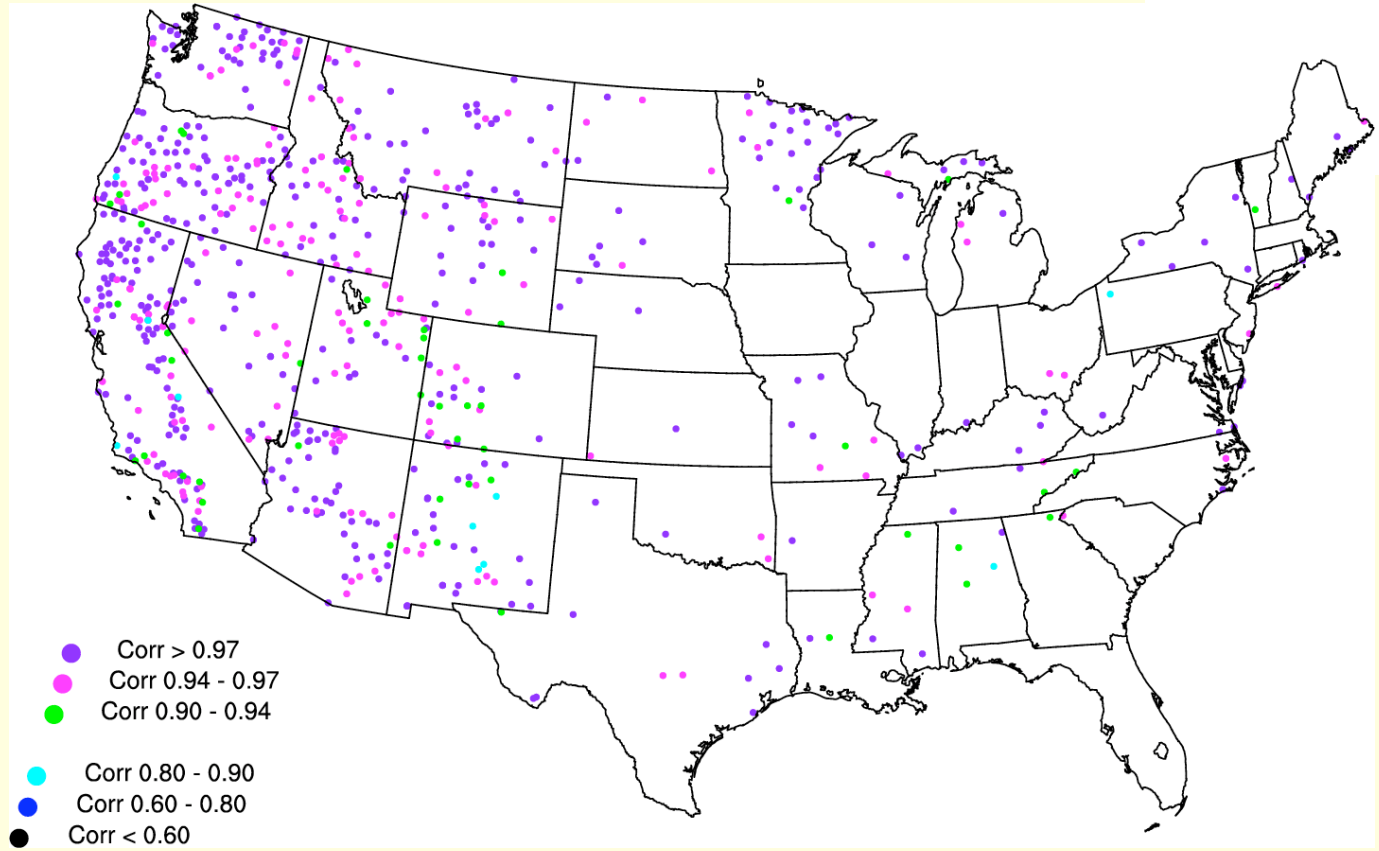


Weather Variables

Relative Humidity

1300 LT; May - Oct

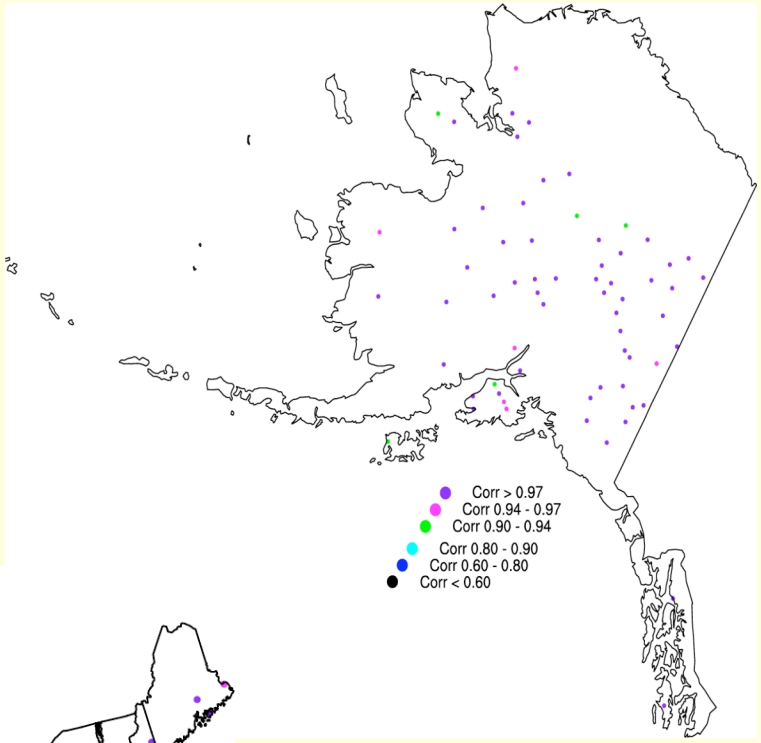


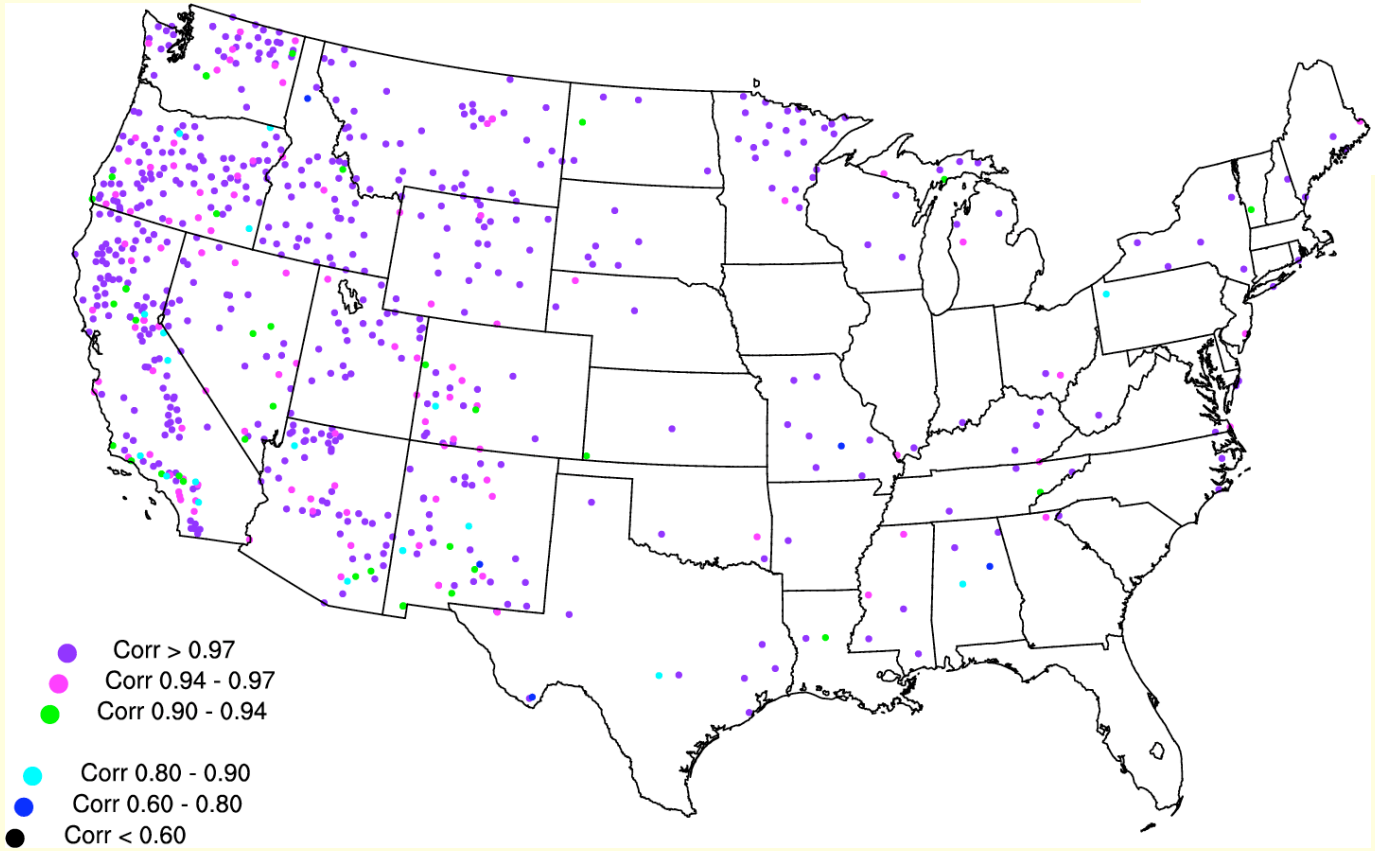


Weather Variables

Temperature

1300 LT; May-Oct





Options for improvement

- Modify equation scheme
  - 2-month seasons
  - Hourly equations
  - Different predictor variables
- Regional reanalysis

Options for improvement
2-month seasons

Correlations for 1300 LT

044195 – Temperature			
May - Jun	.834	May - Oct	.982
Jul - Aug	.948		
Sep - Oct	.999		

044195 – Relative Humidity			
May - Jun	.933	May - Oct	.977
Jul - Aug	.937		
Sep - Oct	.994		

044195 – Wind Speed			
May - Jun	.675	May - Oct	.993
Jul - Aug	.616		
Sep - Oct	.998		

Regional Reanalysis

3-hourly frequency (1979-2003)

180 km

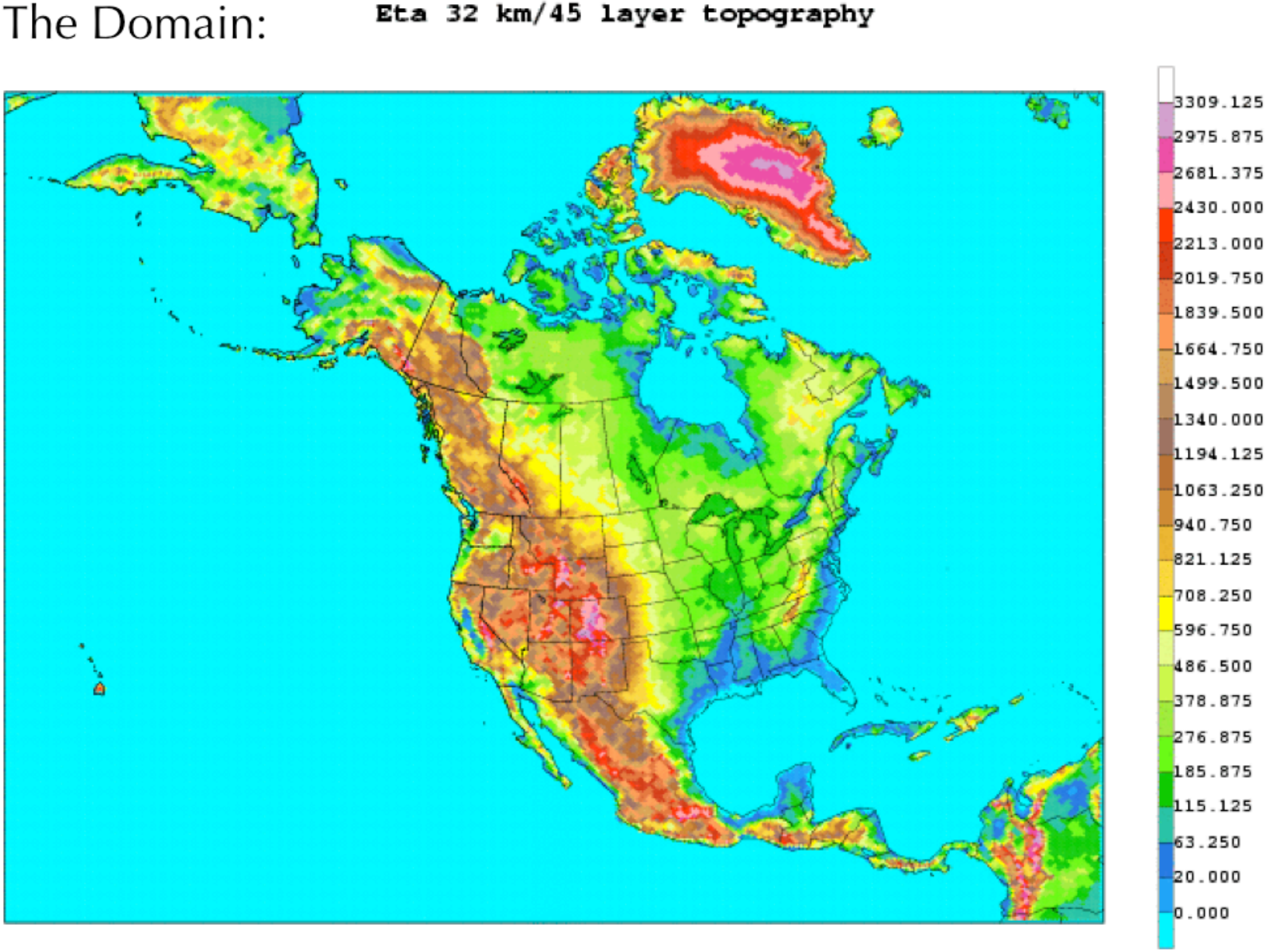
Global Reanalysis

32 km

N.A. Regional Reanalysis

Regional Reanalysis

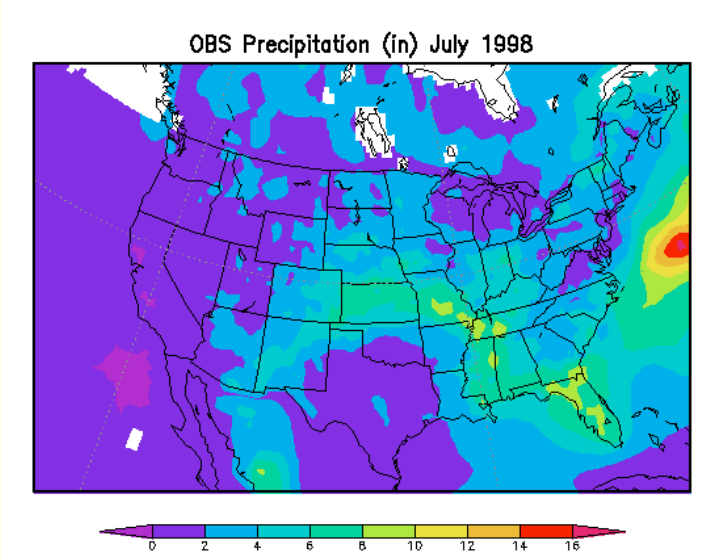
The Domain:

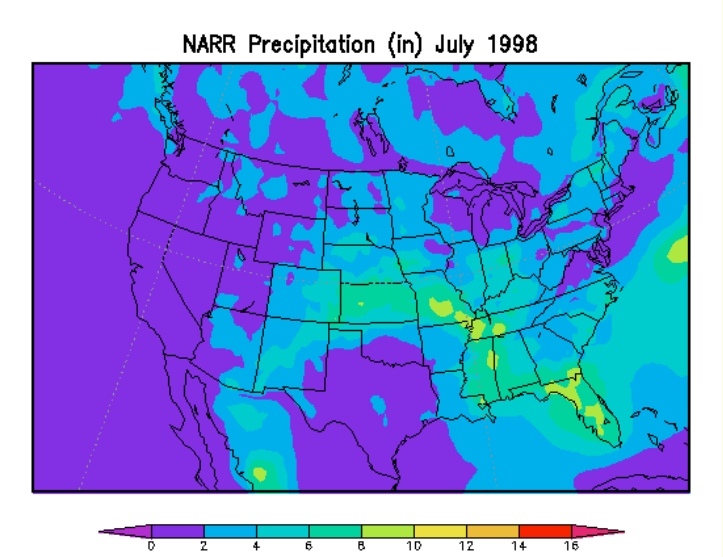


Data Added or Improved Upon for Regional Reanalysis

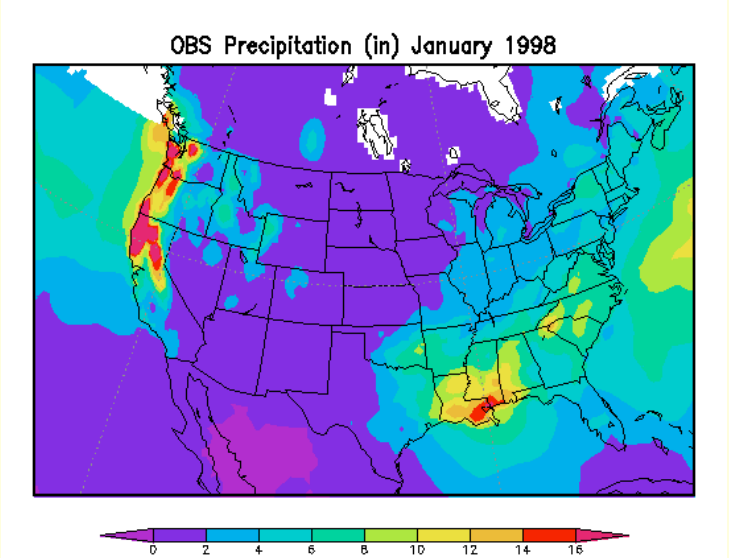
Dataset	Details	Source
Precipitation	CONUS (with PRISM), Mexico, Canada, CMAP over oceans (<42.5°N)	NCEP/CPC,Canada, Mexico
TOVS-1B radiances	Temperature, precipitable water over ocean	NESDIS
NCEP Surface	Wind, moisture	GR
TDL Surface	Pressure, wind, moisture	NCAR
COADS (ships/buoys)	Pressure, wind, moisture	NCEP/EMC
Air Force Snow	Snow depth	COLA and NCEP/EMC
SST	1-degree Reynolds, with Great Lakes SSTs	NCEP/EMC, GLERL
Sea and lake ice	Contains data on Canadian lakes, Great Lakes	NCEP/EMC, GLERL, Ice Services Canada
Tropical cyclones	Locations used for blocking of CMAP Precipitation	Lawrence Livermore National Laboratory

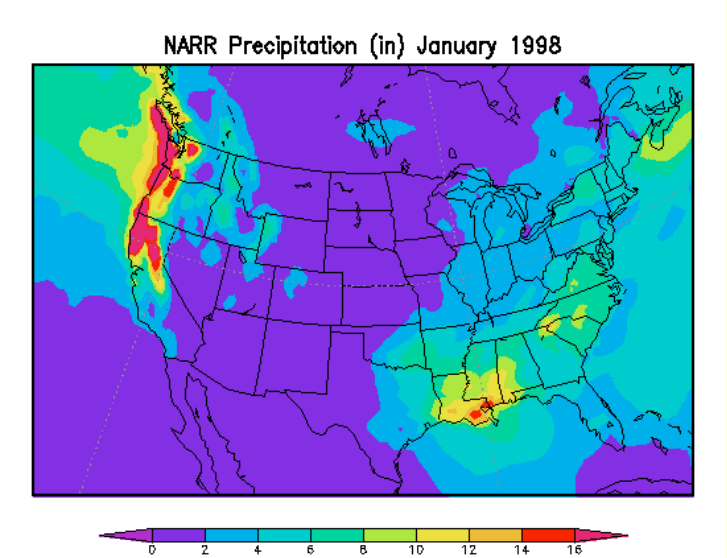
July 1998 Precipitation Results





January 1998 Precipitation Results





Recommendations

- Try to include the 75 ‘bad’ stations
  - Find the missing NWS ID
  - Determine why bad match with WIMS
- North American Regional Reanalysis (NARR)
- Start preparing for technology transfer and real-time operation