Use of Cross Validation in Forecast Verification

Tressa L. Fowler

tressa@ucar.edu

Quality Assessment Group

Research Applications Program

National Center for Atmospheric Research

September 15, 2004

What is cross-validation?

- Ideally, collect data to create a forecast then collect new data to verify forecast.
- When not possible, split collected data into two sets.
  - Training set used to create forecast.
  - Testing set used to verify forecast.

Assumptions for cross-validation

- The testing and training sets are independent.*
- Observations in the testing and training sets come from the same distribution.
- Testing set observations are accurate and unbiased.

*Not the same as IID.

When might cross-validation be used?

- Field project data (e.g. Juneau Wind Hazard Alert System).
- For nowcast products that are developed using current observations (e.g. Ceiling and Visibility product).
- Others?

Some Methods of Selecting Data

- Randomly.
- Create pairs, place one member in each set.
- Divide at some time (not recommended when time trends may be present, such as seasonal or increasing).
- Coverage design for spatial data.
- Systematically (i.e. jackknife).

SUGGESTED METHOD FOR SMALL SETS OF CASES – Repeated Cross Validation

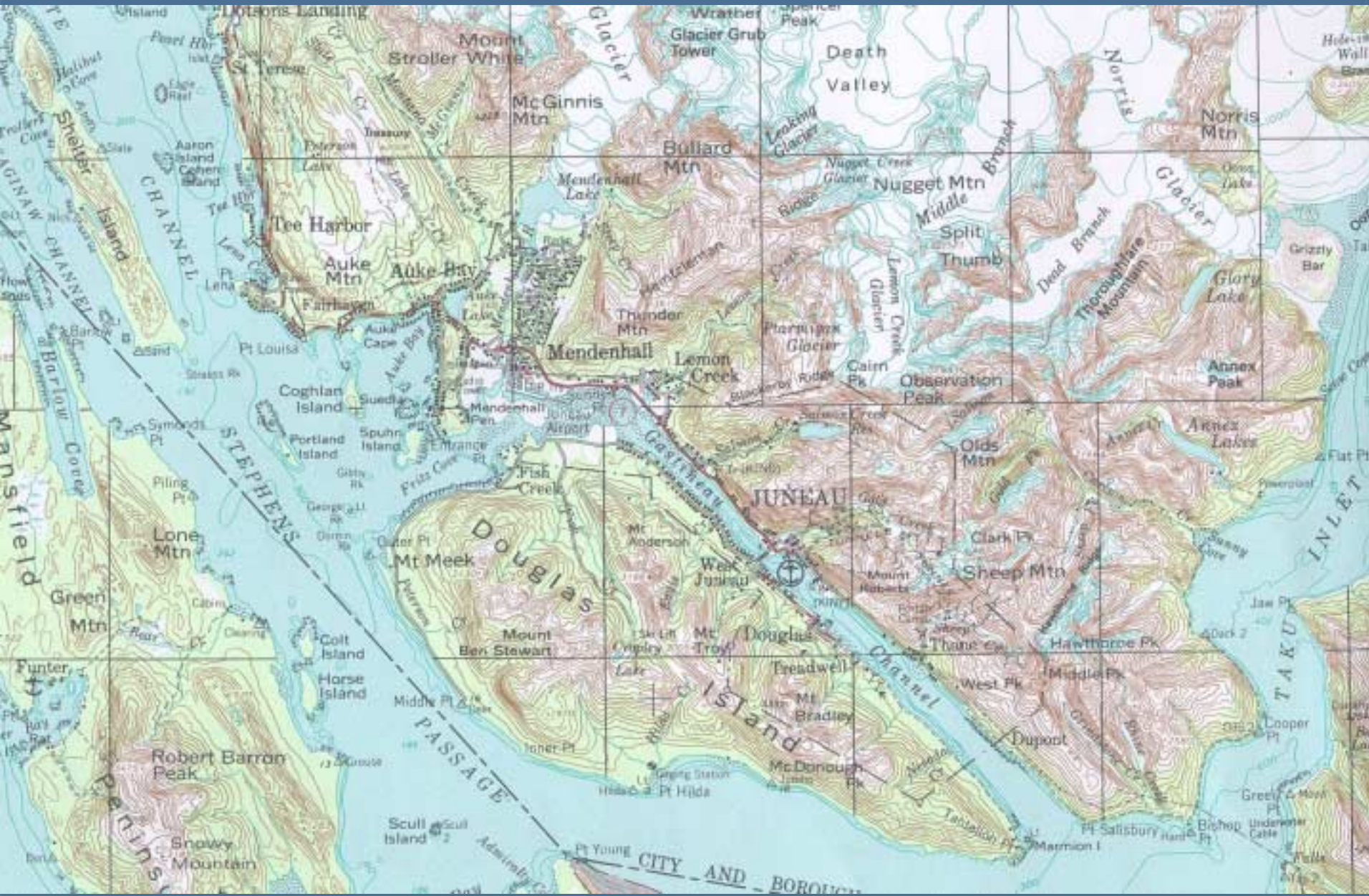
- Select a very small sample of cases to leave out of the training set (e.g. 1 or 5). Create forecast using remaining cases.
- Verify on withheld cases.
- Repeat many times (e.g. 100).
- Summarize results from errors computed for each verification set (e.g. POD or MAE).

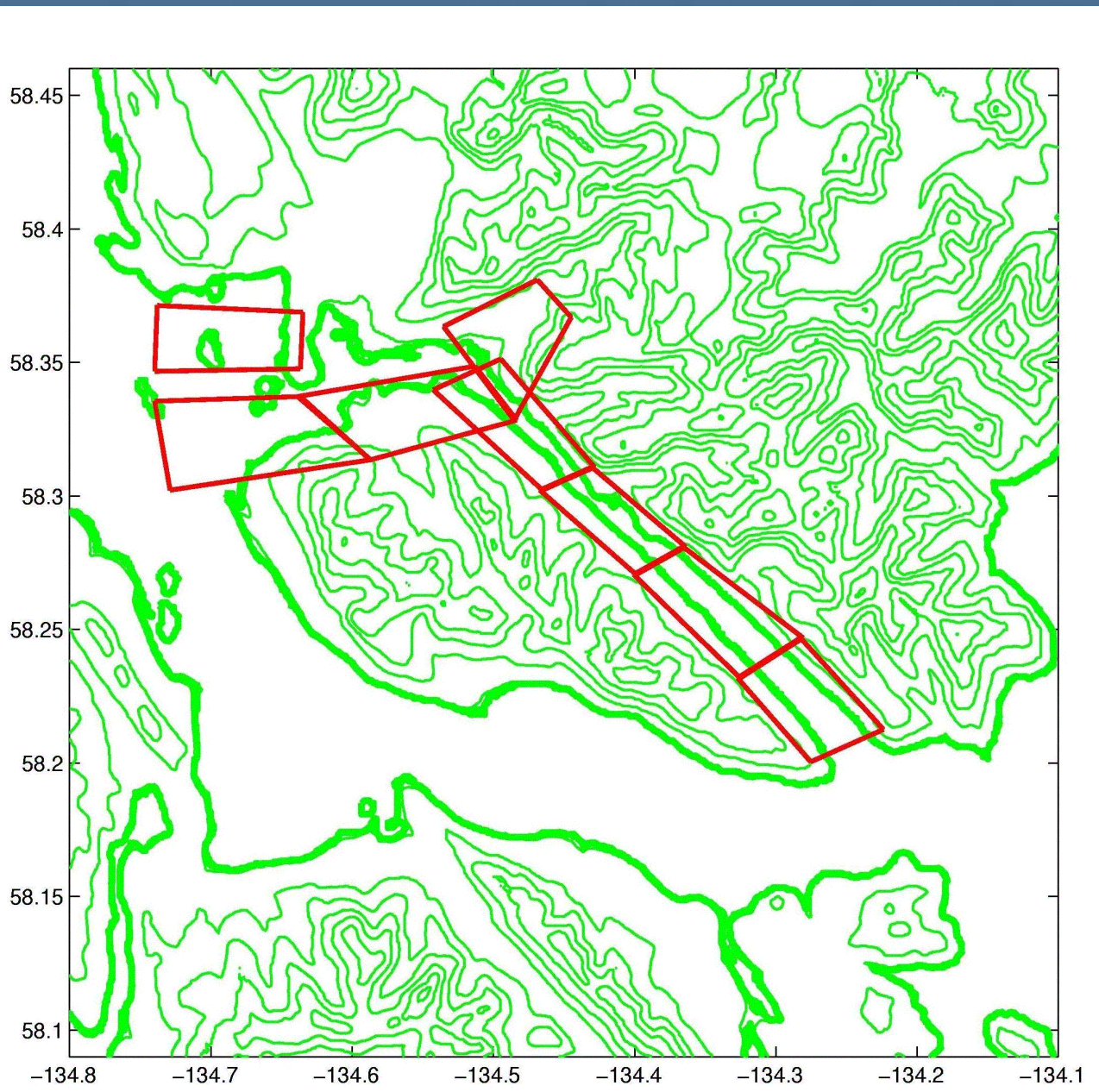
JWHAS example

- Need to verify JWHAS forecast EDR value.
- Aircraft measurements of EDR from the 2003 field project form basis for product.

$$Aircraft\ measured\ \ ZEDR = \sum_{i=0}^{4} \beta_i \cdot winds_i$$

- These observations are also the only available verification data.


Lotsons Landing
Mount Stroller White
St Terese
McGinnis Mtn
Bullard Mtn
Glacier Grub Tower
Death Valley
Norris Mtn
Nugget Mtn
Split Thumb
Tee Harbor
Auke Mtn
Auke Bay
Fairhaven
Thunder Mtn
Mendenhall
Lemon Creek
Coghlan Island
Spuhn Island
Portland Island
Fish Creek
JUNEAU
Olds Mtn
Sheep Mtn
Clark Pk
Mt Meek
Douglas
Douglas Island
Lone Mtn
Green Mtn
Colt Island
Horse Island
Mount Ben Stewart
Mt Troy
Treadwell
Mt Bradley
Dupont
Hawthorne Pk
Robert Barron Peak
Snowy Mountain
Scull Island
Pt Young
Pt Hilda
Pt Salisbury
Marmion I
Shelter Island
Barlow Cove
Mansfield Peninsula
STEPHENS PASSAGE
GASTINEAU CHANNEL
TAKU INLET
CITY AND BOROUGH


58.45
58.4
58.35
58.3
58.25
58.2
58.15
58.1
−134.8
−134.7
−134.6
−134.5
−134.4
−134.3
−134.2
−134.1

JWHAS Data Issues

- Each flight has time correlation.
- Spatial correlation is unlikely as each box is treated separately.
- One instrument takes all measurements, thus systematic bias is possible.

JWHAS Cross Validation Methods

- Cases < 21: Don't bother.
- Cases < 25: leave each case out 1 time, run n times (Jackknife)
- Cases < 35: leave 2 cases out each time at random, run 25 times
- Cases < 60: leave 5 cases out at random, run 10 times
- Else: leave 10 cases out at random, run 5 times

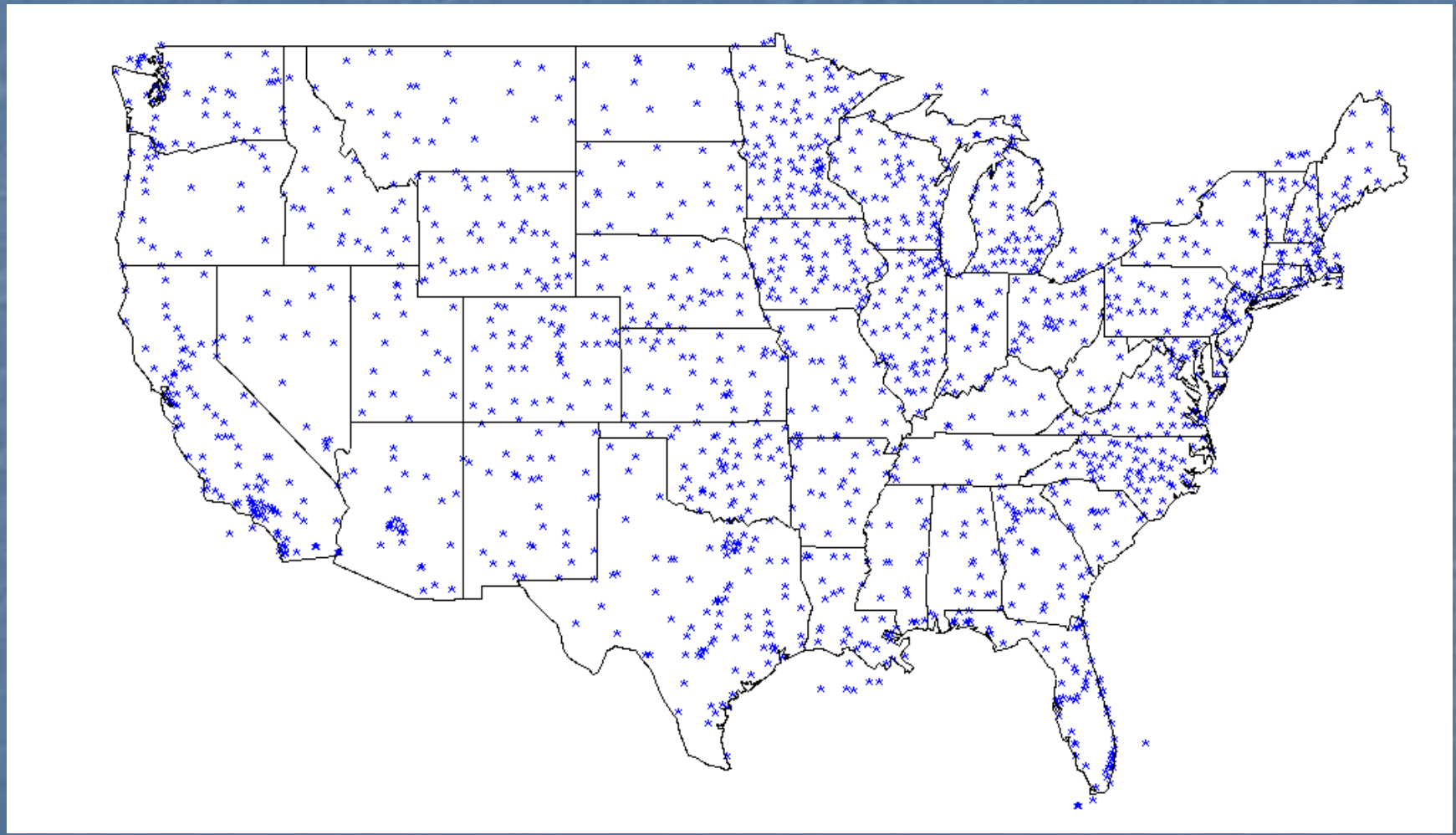
Example Results

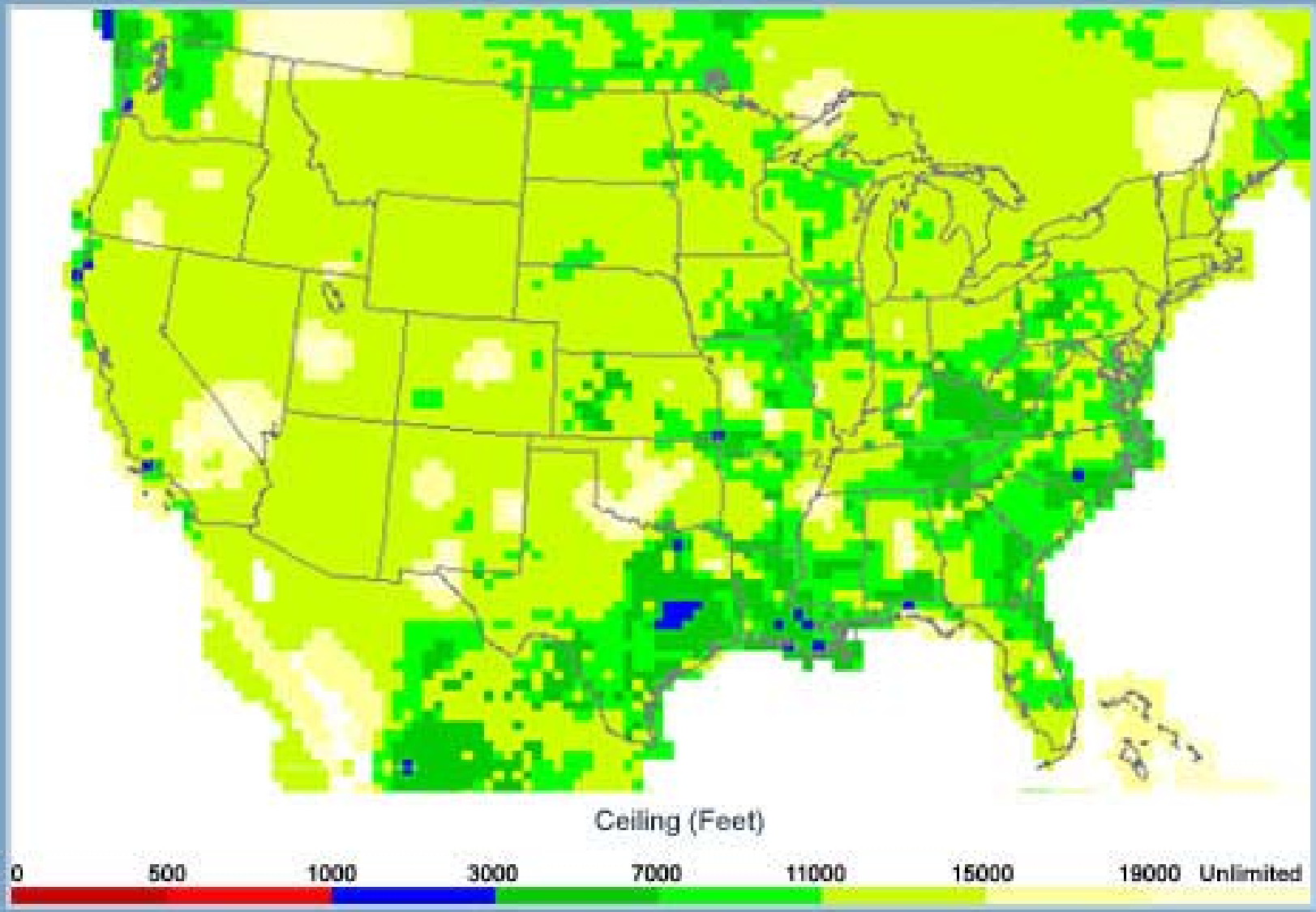
	PODy	PODn	FAR	TSS
JWHAS Cross-validated	0.772	0.883	0.292	0.655
JWHAS “Cheating”	0.832	0.881	0.241	0.713

Ceiling and Visibility example

- Need to verify Ceiling and Visibility analysis product, i.e. nowcast.
- METAR stations measurements of C and V form basis for analysis product.
- The observations at METAR stations are also the only available verification data.

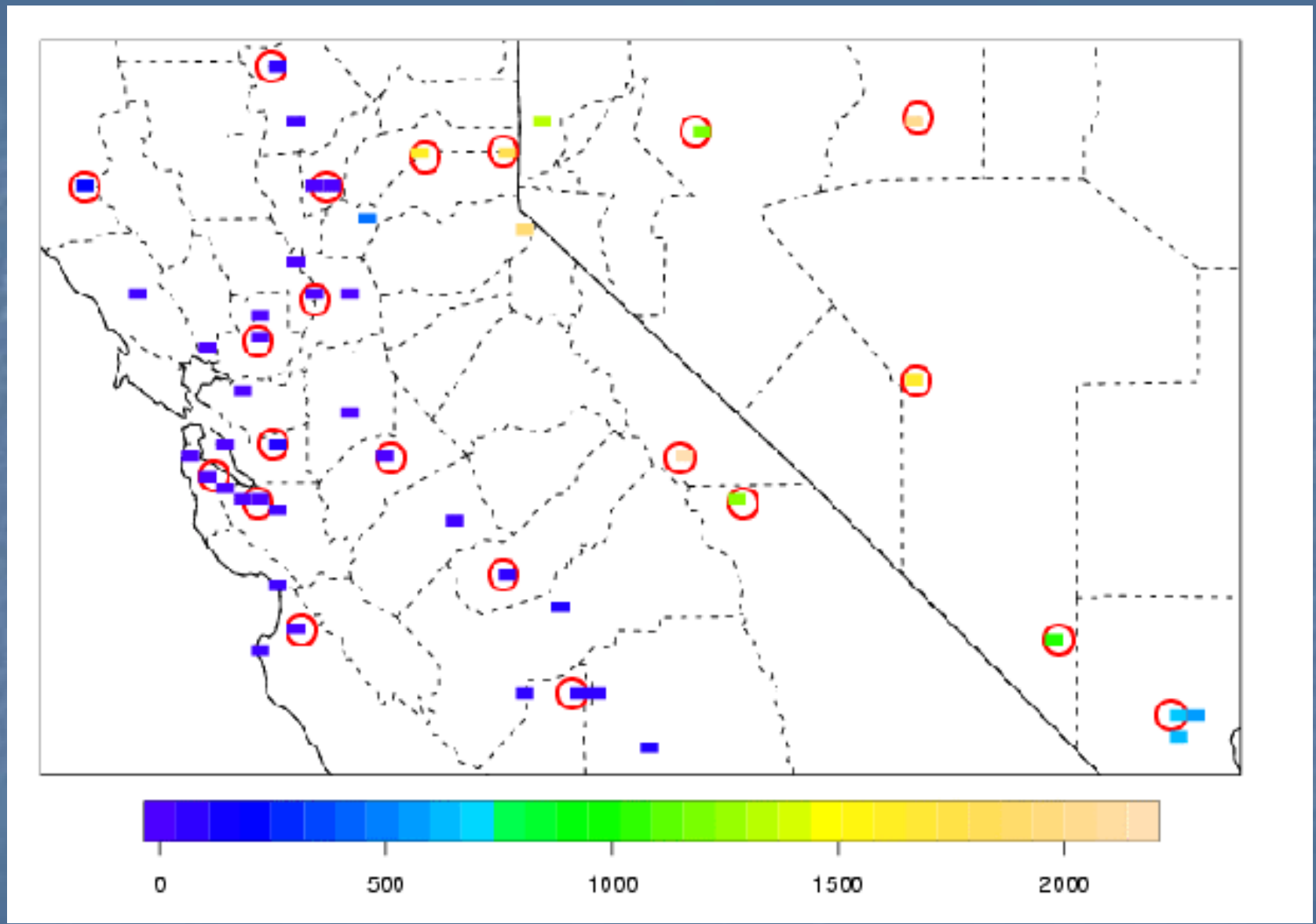
LOCATIONS OF CONUS METAR STATIONS USED





Ceiling forecast algorithm output VT: 2100 UTC 08 Jun 01

UCAR/RAP


0
500
1000
1500
2000

Ceiling and Visibility Data Issues

- Data have spatial correlation.
- All observations are taken from a single hour, so there is no time-series.
- The instruments at each location are distinct, thus less likely to have systematic bias.

Conclusions

- Cross-validation can be used in forecast verification when better methods are not possible.
- Forecast verification data most likely violate some assumptions, thus care should be used in application of cross-validation methods.