Object-oriented Verification of Reflectivity Fields based on Cluster Analysis: A Report

Caren Marzban
Dept. of stat (UW), APL (UW), CAPS (OU)

Introduction

- Verification with meteorologically relevant objects.

Baldwin Bullock Brown Chapman Davis Du Ebert McBride Kain Lakshmivarahan Mullen Nachamkin.

- Verification with objects.

Marzban and Sandgathe.

- Objects defined with Cluster Analysis (CA).

- Agglomerative hierarchical = iterative:

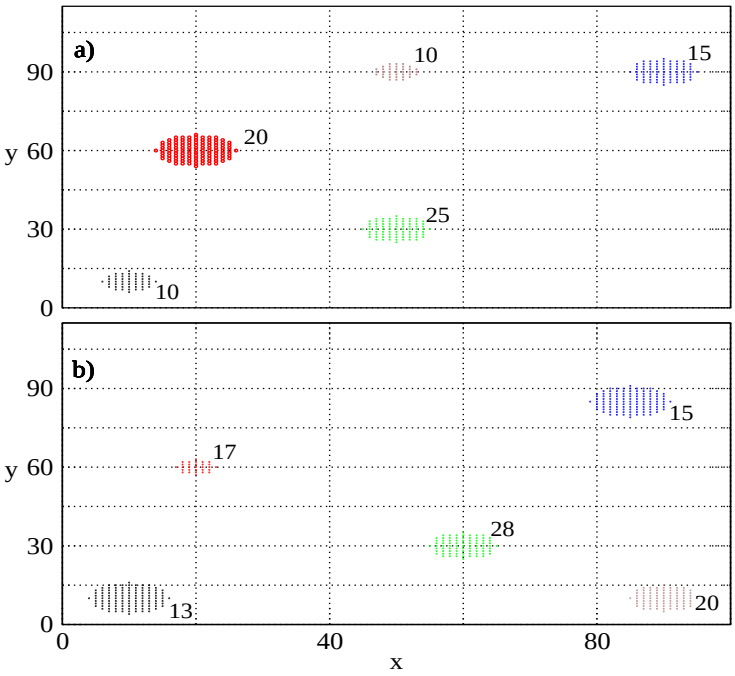
Start with N clusters
Identify nearest 2 clusters
Merge them.
Repeat.
End with 1 cluster.

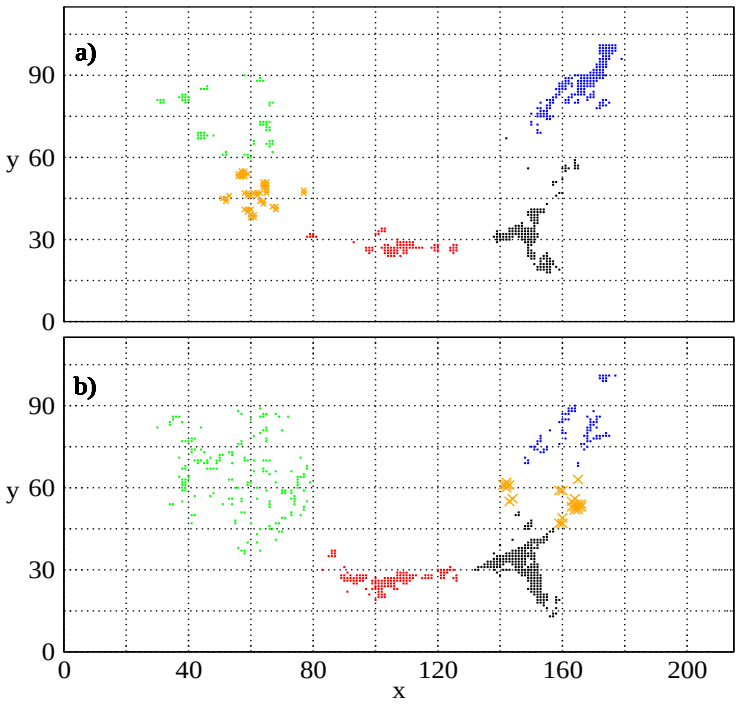
No. of clusters $\sim$ scale.

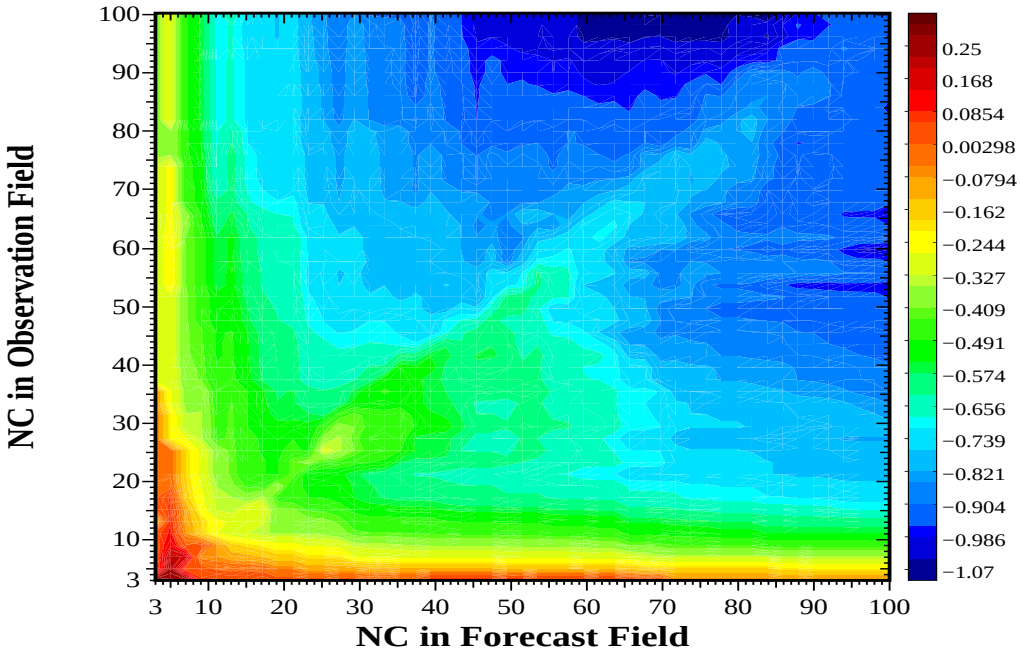
CA in (x,y), (x,y,z), (x,y,z,t), ...

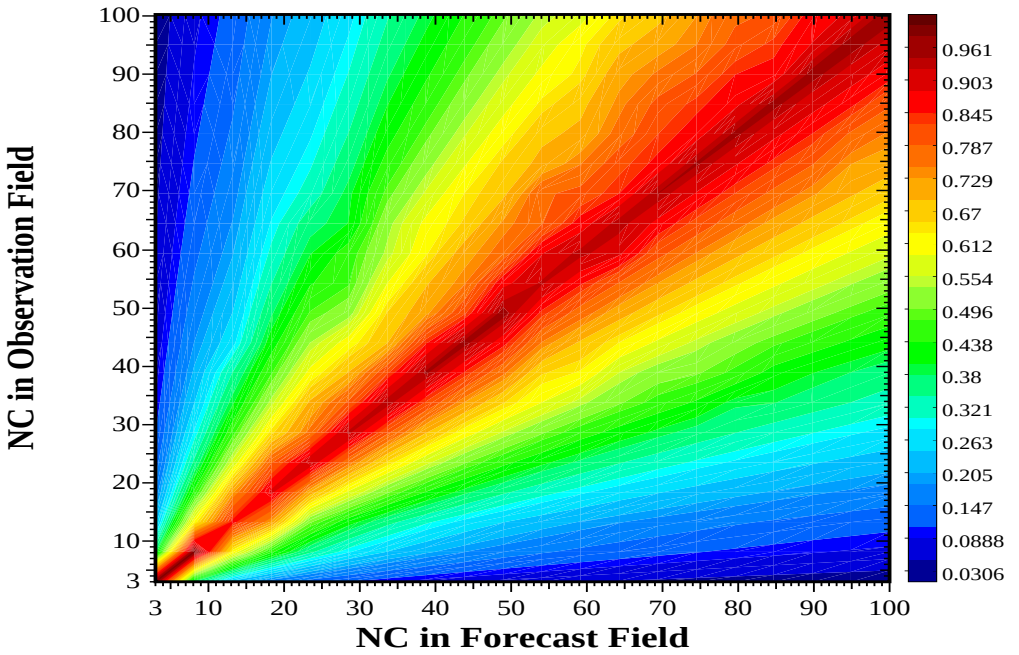
Theme:
- Do CA in obs field
- Do CA in forecast field
- At each iteration in each field, compute all inter-field inter-cluster distances.
- Drop all distances > median + sigma
. $\longrightarrow$ (false alarms & misses)
- Average the remaining distances
. $\longrightarrow$ (displacement error)
- Plot vs. N_f and N_o. ("error surface")

The result:


a)
10
15
20
25
10
b)
15
17
28
13
20
y
x
0
30
60
90
0
40
80

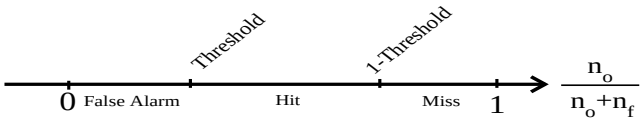

a)
b)
y
x
0
30
60
90
0
40
80
120
160
200


NC in Observation Field
NC in Forecast Field
100
90
80
70
60
50
40
30
20
10
3
3 10 20 30 40 50 60 70 80 90 100
0.25
0.168
0.0854
0.00298
−0.0794
−0.162
−0.244
−0.327
−0.409
−0.491
−0.574
−0.656
−0.739
−0.821
−0.904
−0.986
−1.07


NC in Observation Field
NC in Forecast Field
100
90
80
70
60
50
40
30
20
10
3
3 10 20 30 40 50 60 70 80 90 100
0.961
0.903
0.845
0.787
0.729
0.67
0.612
0.554
0.496
0.438
0.38
0.321
0.263
0.205
0.147
0.0888
0.0306

Variation:

- Combine obs and forecasts into 1 field.
- Do CA.
- At each iteration (in 1 field), compute fraction obs.
- If fraction $<$ threshold (e.g. 0.1) $\rightarrow$ false alarm.
- If fraction $>$ 1-threshold (e.g. 0.9) $\rightarrow$ miss.
- Else $\rightarrow$ hit.
- Plot CSI vs. no. of clusters. ("CSI curve")



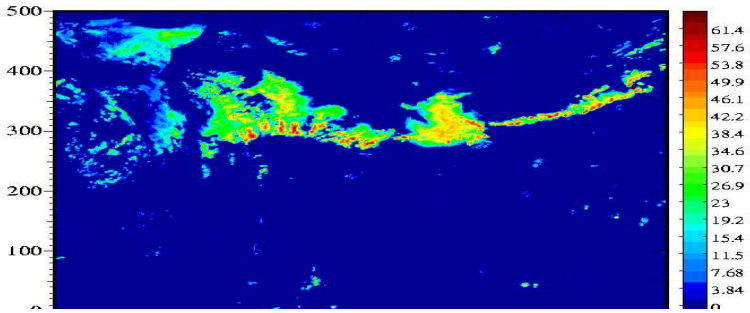
Advantages:

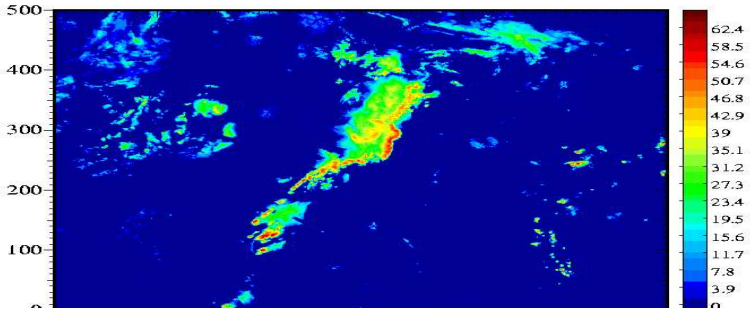
- Better suited for CSI
- Cleaner.

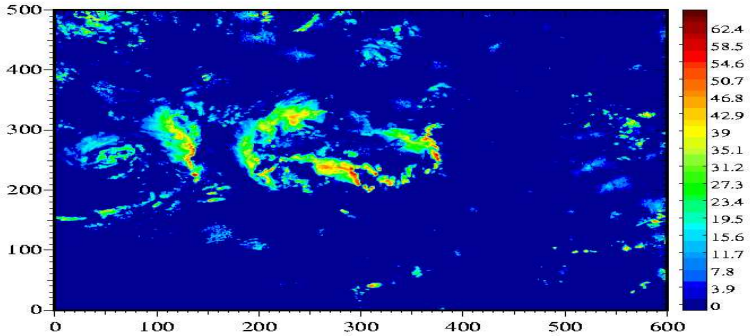
Disadvantages:

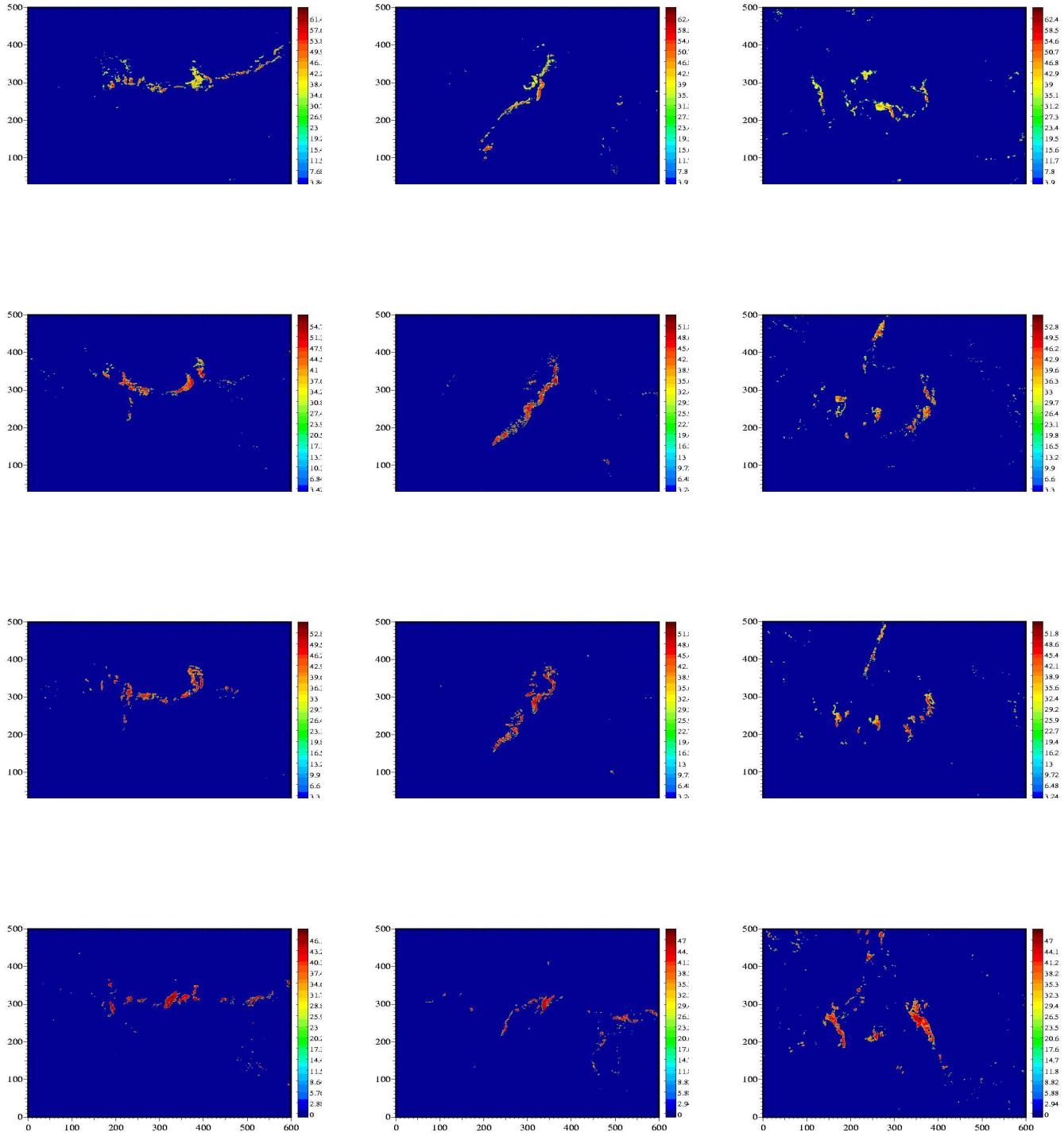
- ??

Data: Mike Baldwin (May 12 , May 13, and June 4, 2005)

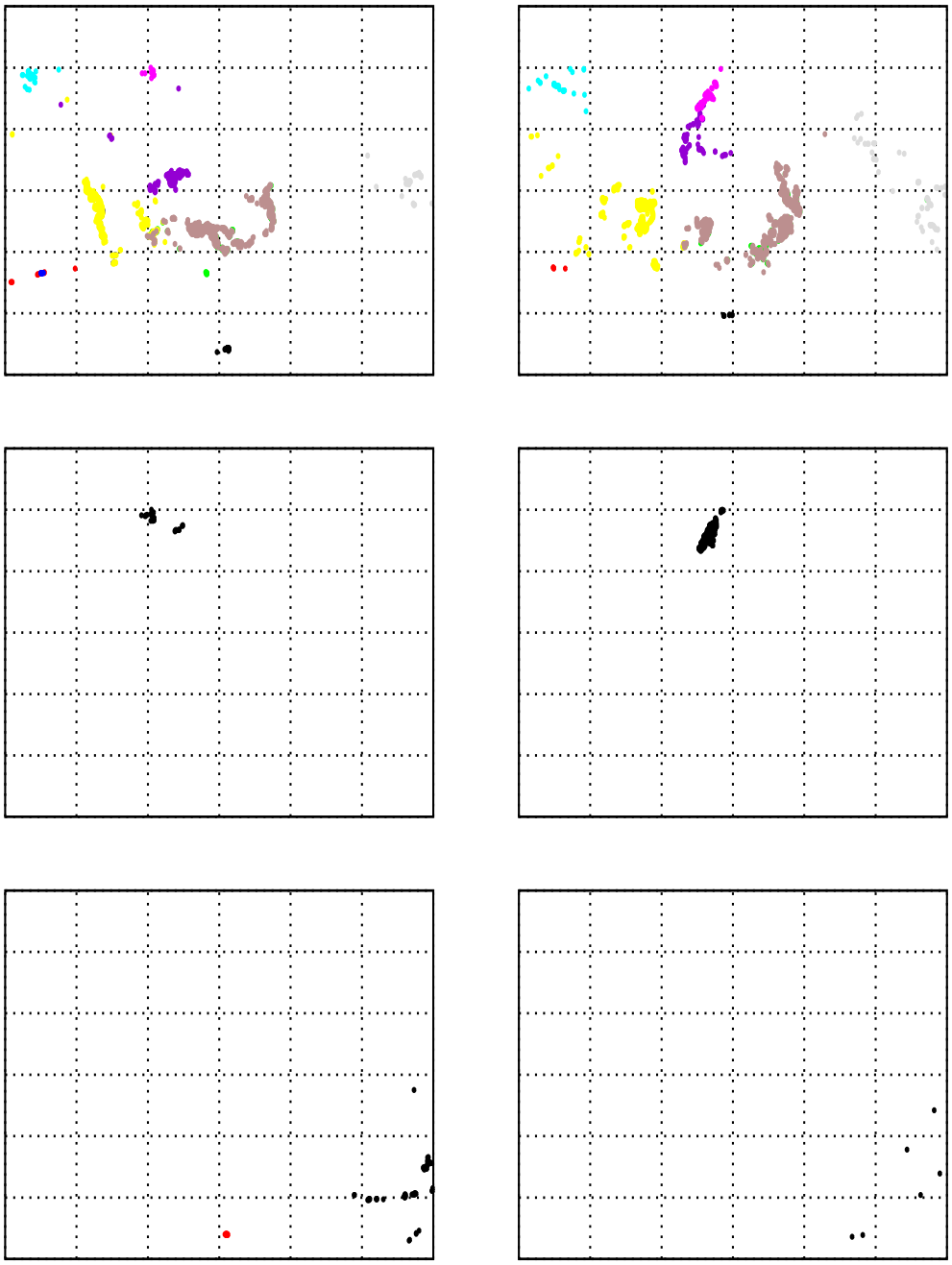




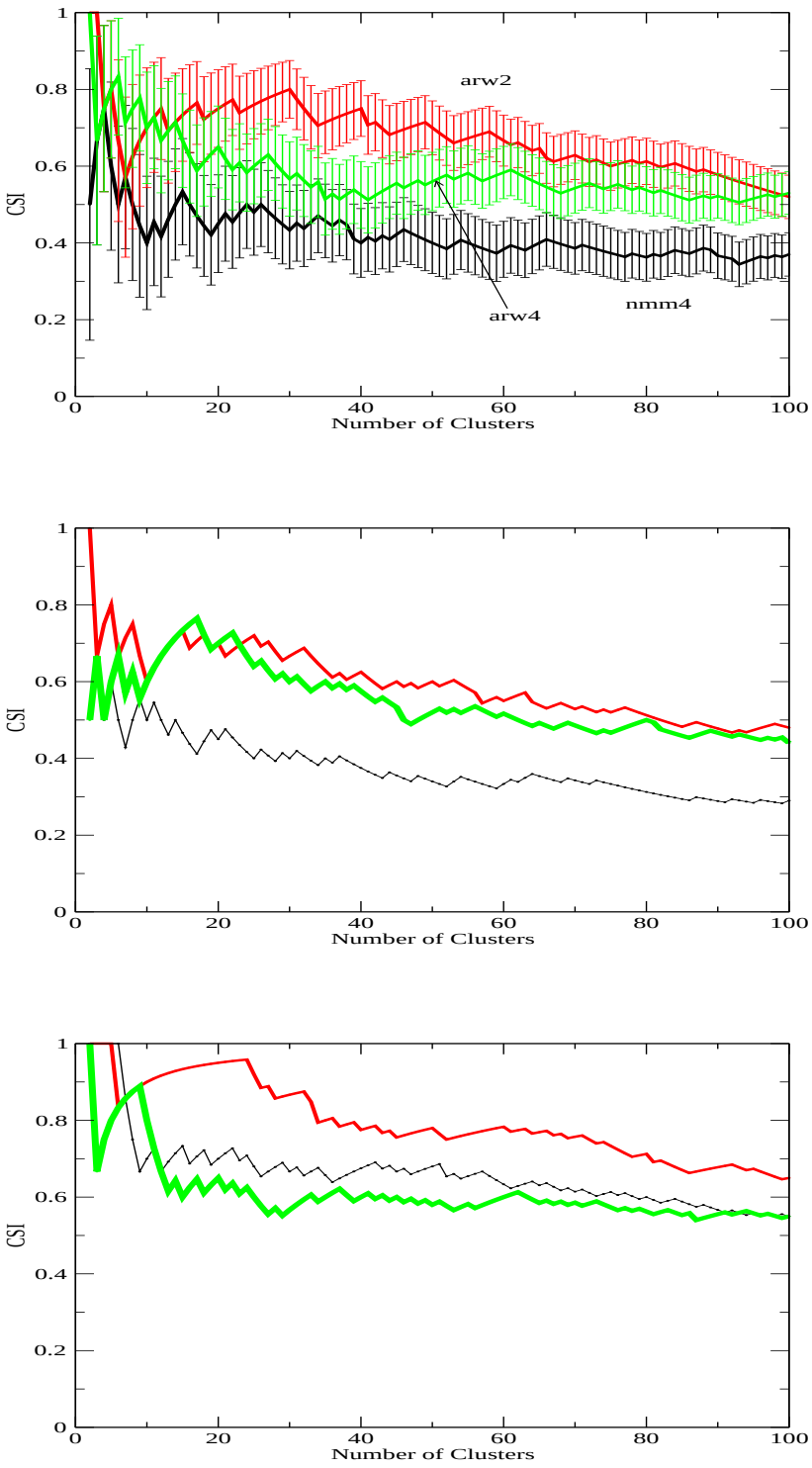




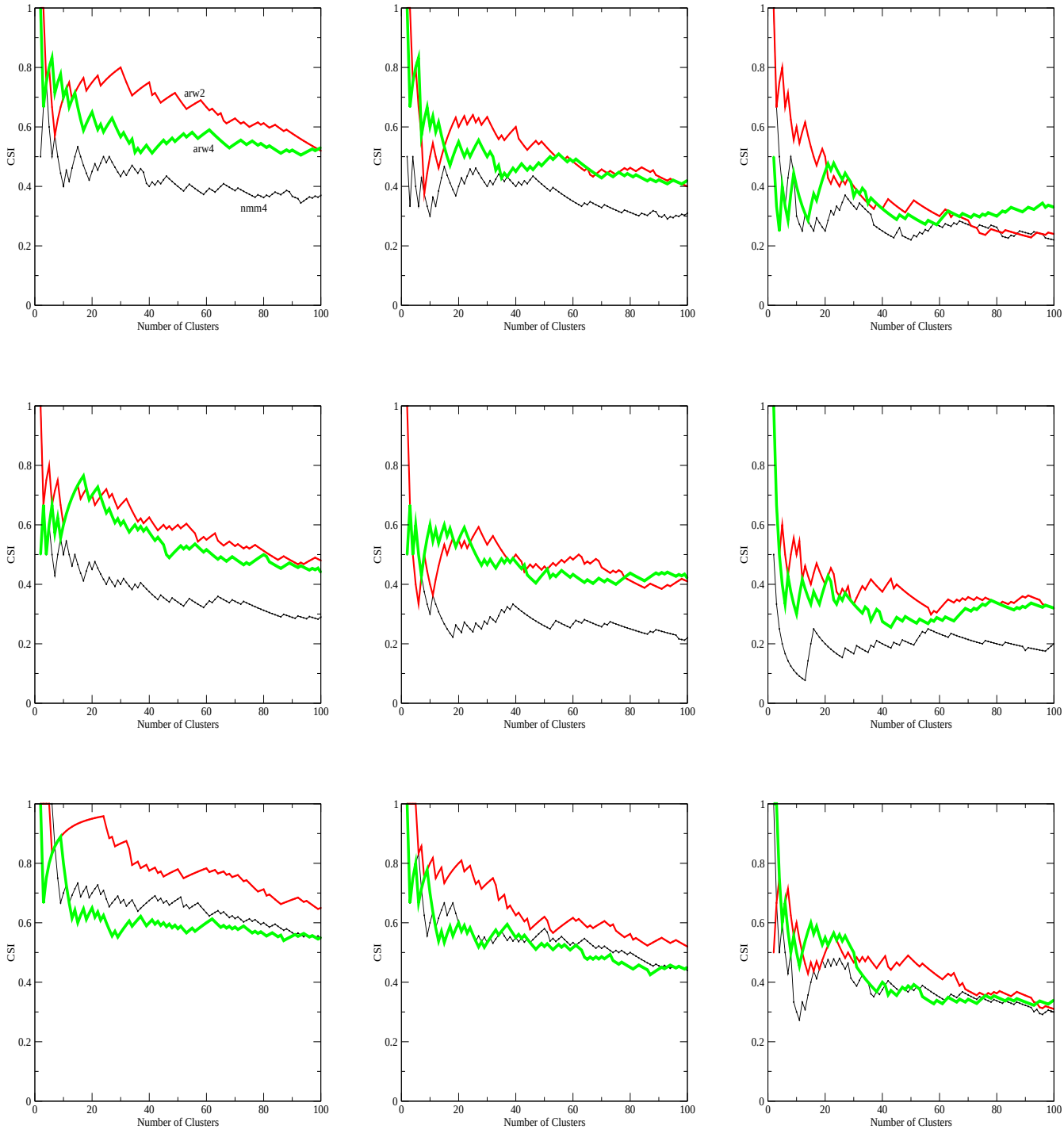
Reflectivity obs (top row), and the corresponding 24hr forecasts according to arw2, arw4, and nmm4 (from top to bottom). Columns: May 12, 13, June 4. US East of the Mississippi.



Observed and arw2 forecast fields on June 4, parsed into hits (top), false alarms (middle), and misses (bottom).



CSI curves for the 3 models on May 12 (top), May 13 (middle) and June 4 (bottom).



CSI curves from an (x, y, z) analysis for the 3 dates (rows) and three different thresholds - 0.01, 0.1, and 0.2 (from left to right).

Conclusions:

0) CA-based object-oriented verification seems to work.

1) arw2 > arw4 > nmm4, marginally.

2) across scales, except for extremes.

3) on smaller scales (larger number of clusters), the differences between the models appear to diminish.

4) performance falls off with increasing cluster number (smaller scale). (Appendix B)

Future Plan (next week):

Extend to more dates, now automatically.
Show-down.

Acknowledgements

Above names+Scott Sandgathe+Bob Gall+DTC staff.

Appendix A: Error-bar for CSI

Consider the contingency table

$$\begin{pmatrix} a & b & \\ c & d & N_1 \\ & F_1 & \end{pmatrix},$$

$$b = binomial(F_1, b/F_1).$$

$$c = binomial(N_1, c/N_1).$$

$$\begin{aligned} E[b] &= F_1 \frac{b}{F_1} = b \\ Var[b] &= F_! \frac{b}{F_1}(1 - \frac{b}{F_1}) = \frac{bd}{F_1} \\ E[c] &= N_1 \frac{c}{N_1} = c \\ Var[c] &= N_1 \frac{c}{N_1}(1 - \frac{c}{N_1}) = \frac{cd}{N_1}. \end{aligned}$$

$$\sigma_{CSI} = \text{CSI} \sqrt{\frac{1}{d}\left(\frac{b}{b+d} + \frac{c}{c+d}\right)}\,.$$

Appendix B: CSI vs. no. of clusters

At each split of CA, either b, or c, or d, is incremented.

If b or c are incremented then $\mathrm{CSI}_n = \frac{1}{n}\mathrm{CSI}_1$.

If d is incremented, then CSI may increase.

Result, CSI falls off with n, but with periods of increase.