Verification of the spatial structure of gridded forecasts

Corinne Jones, **Caren Marzban**, Ning Li, Scott Sandgathe

A subset of methods developed for spatial verification involve clustering methods for identifying meteorological objects within gridded observed and forecast fields. The problem of comparing two fields which have been clustered is closely related to the problem of comparing different clusterings. Several clustering comparison techniques have already been developed in the machine learning community. This work examines various ways in which the clustering methods can be combined with the clustering comparison methods (e.g., optimal matching) for the purpose of better assessing the quality of the forecasts. Many of the existing clustering comparison methods involve quantities for which statistical tests exist, but some of those developed here do not have simple tests, and so, null distributions are developed by comparison with random gaussian fields.

Introduction

- Precip fields (observed/analysis and forecast) are "lumpy."
- Many spatial verification methods have been proposed.
- Spatial Verification (e.g., special issue of *Wea. Forecasting*, **24(6)** 2008.)

- A subset involves clustering (of one or both fields).
- cluster = object = lump in the field.
- can be done in more than x-y (lat/lon) space.
- can assess spatial errors (displacement, size, intensity, orientation, etc.)

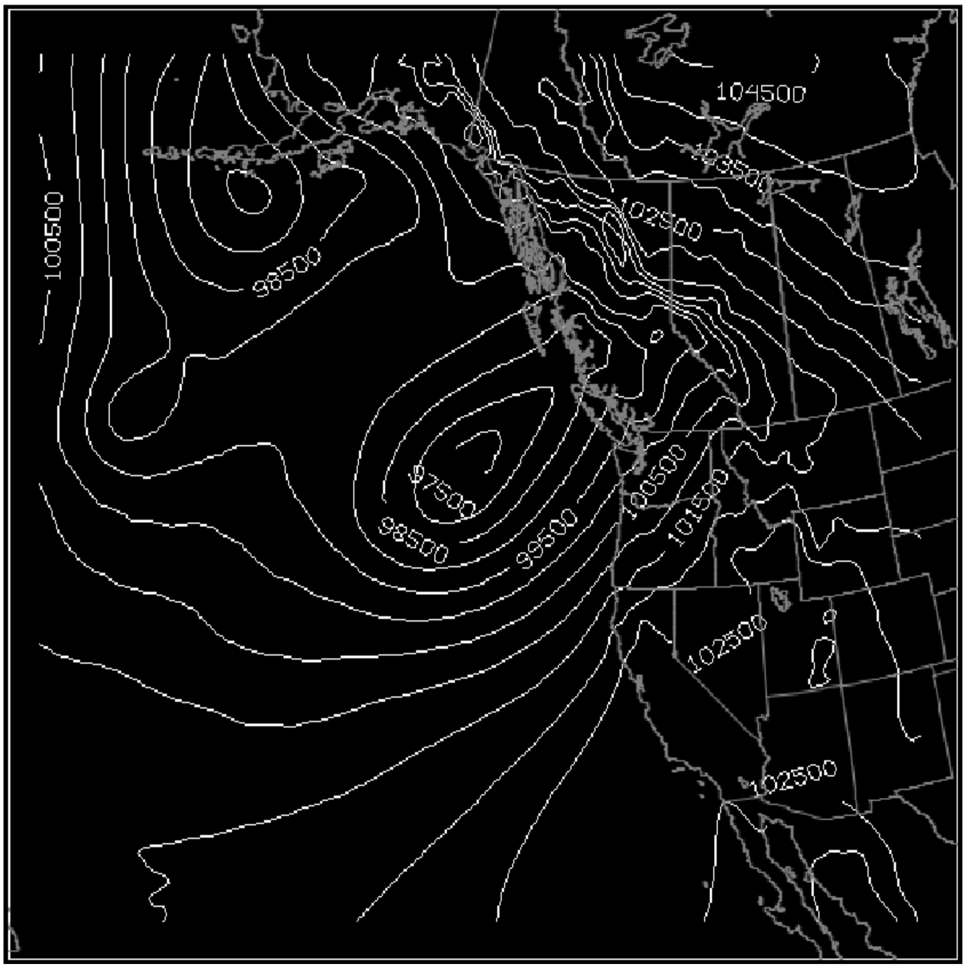
Here we examine two supplementary methods:
1) Clustering followed by pair counting
2) Optimal matching followed by clustering

The former does not (yet) provide estimates of spatial errors,
but it does allow for efficient comparison of NWP model forecasts.

The latter provides estimates of spatial errors, on a point-wise basis (for now),
assessing forecast quality and allowing for comparisons.

Data

- 24hr accumulated precip
- “Heavy” precip (e.g. exceeding 10 mm per 24 hour)
- $36km$ grid spacing
- 416 Days, observed and forecast, April 2, 2008 - Nov. 2, 2009.
- WRF, MM5, COAMPS (all old versions)
- Domain:

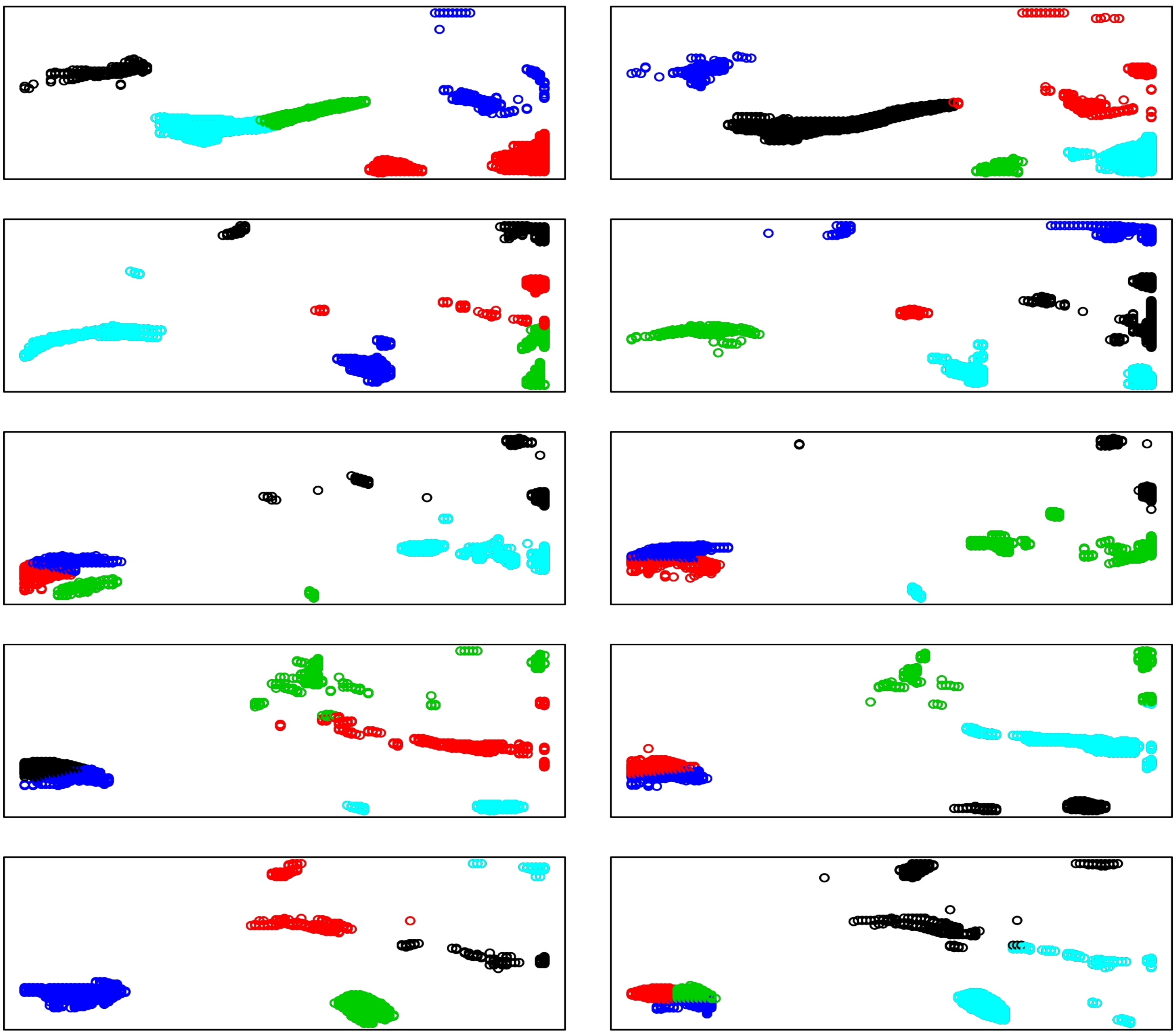


Clustering

Clustering method can make a big difference.

Five examples/days of “bad” (physically unreasonable) clustering; k-means(NC=5)

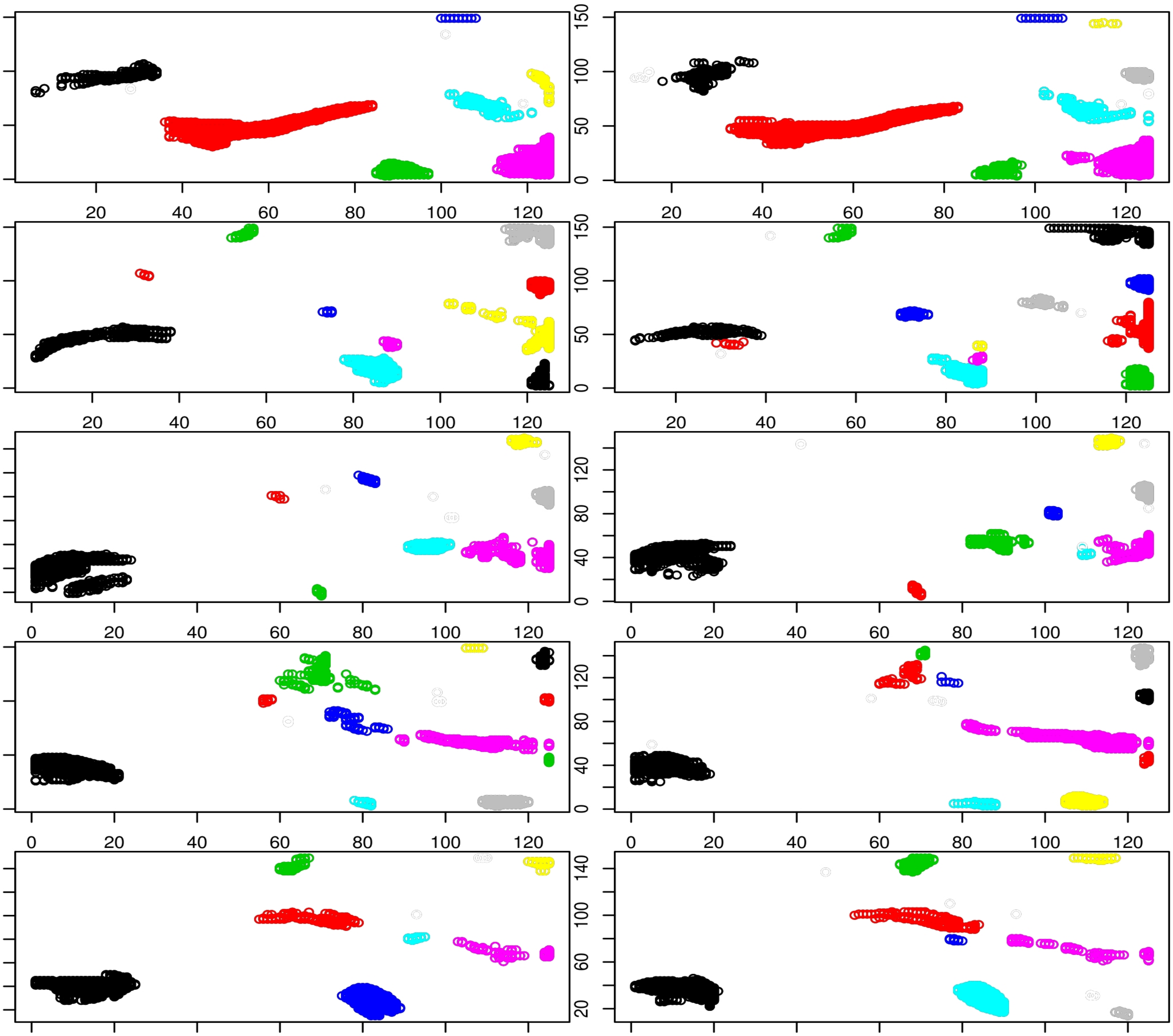
Observed prcp 24hr forecast prcp



Five examples/days of "good" (sensible) clustering; DBSCAN(eps=5)

Observed prcp

24hr forecast prcp



Comparing Clustering

There are (at least) three types of clustering comparison methods:
(Marina Meila, Comparing Clusterings: *J. of Multivariate Analysis*, 2007)

1) Pair Counting, 2) Set Matching, 3) Variation of Information

1) Consider two clusterings C and C', and the contingency table:

N_{11} = no. of point pairs that are in the same cluster under both C and C'
N_{00} = no. of point pairs in different clusters under both C and C'
N_{10} = no. of point pairs in the same cluster under C but not under C'
N_{01} = no. of point pairs in the same cluster under C but not under C'

Wallace proposed:

$$W_I(C, C') = \frac{N_{11}}{\sum_k n_k(n_k - 1)/2} \quad W_{II}(C, C') = \frac{N_{11}}{\sum_{k'} n'_{k'}(n'_{k'} - 1)/2}$$

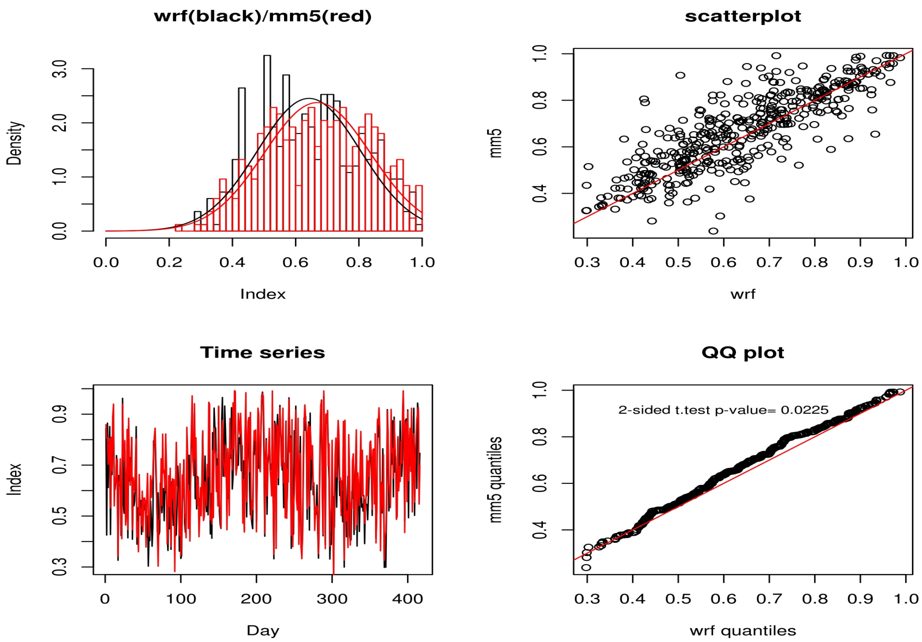
n_k = number of points in cluster C_k, etc.

Fowlkes-Mallows Index (Skill score): $F(C, C') = \sqrt{W_I(C, C') W_{II}(C, C')}$

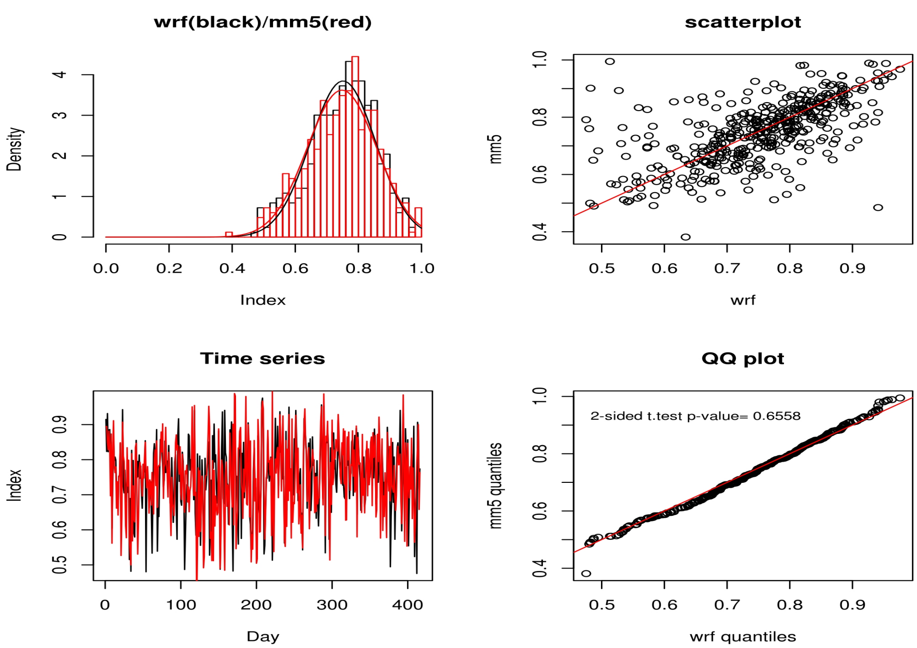
Adjusted Rand Index: $R(C, C') = \frac{N_{11}+N_{00}}{N_{00}+N_{01}+N_{10}+N_{11}}$

Pair Matching Results

Fowlkes' Index



Rand's Index

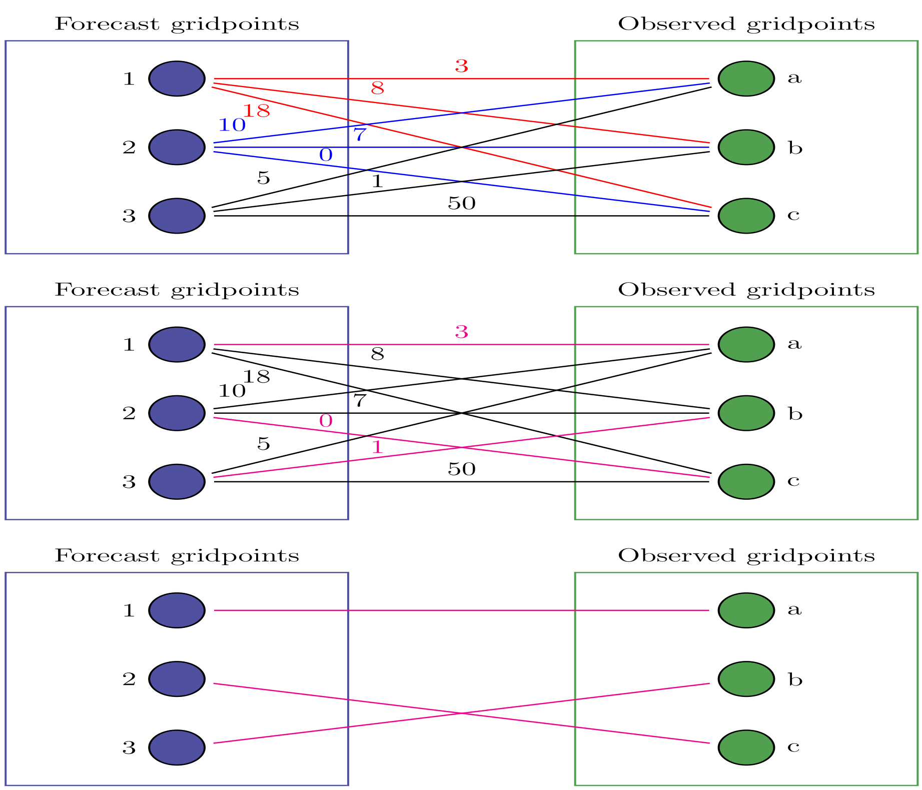


MM5 appears to be marginally "better" than WRF in terms of Fowlkes' index, but quite comparable in terms of Rand's index.

Optimal Matching

- Assignment Problem (big in machine learning)
- Optimal (Mass) Transportation (Fields Medal)
- Optical Flow
- Hungarian Method (Balanced)
- Minimum-Cost bipartite graph
- Bertsekas' auction algorithm

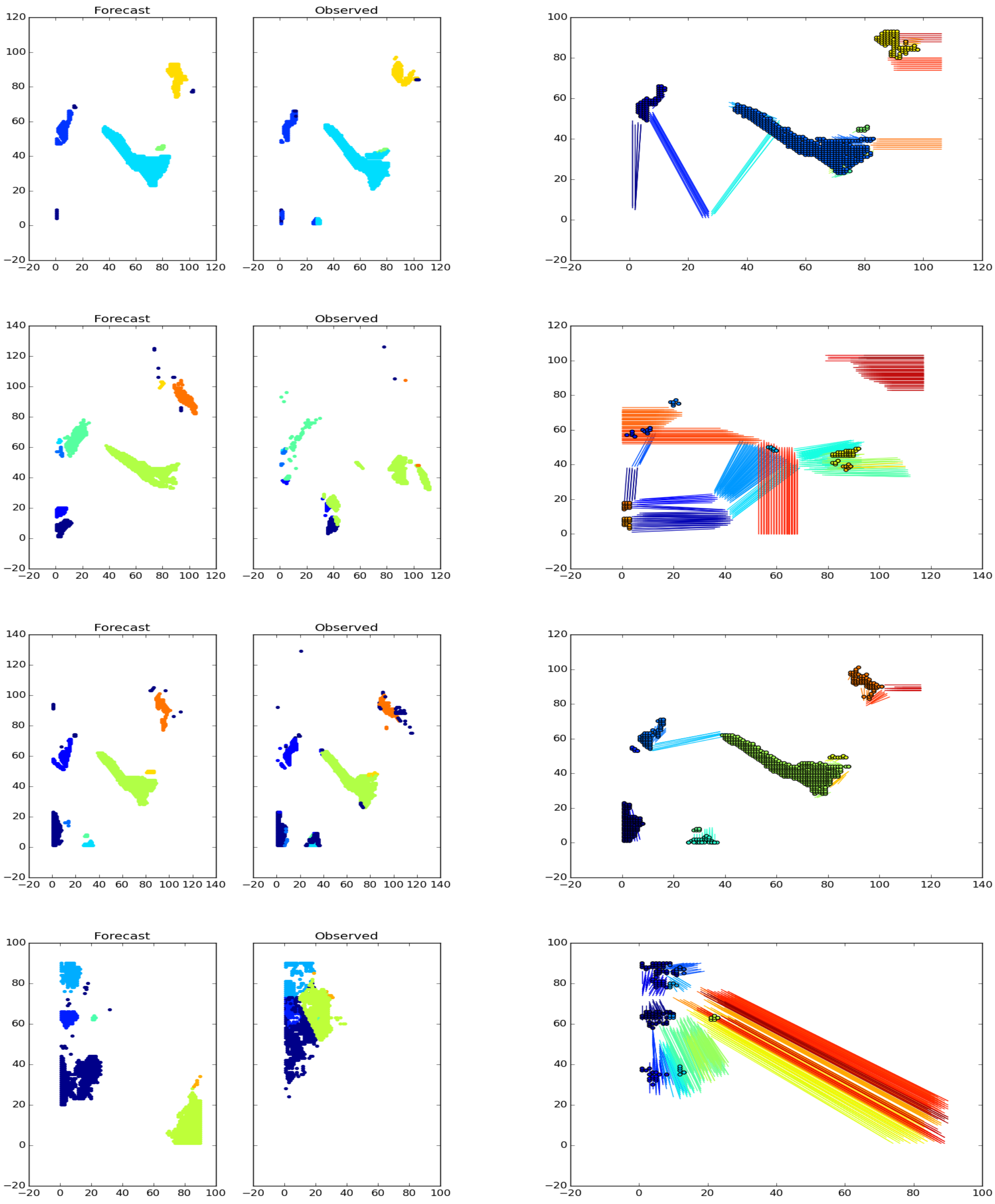
Bertsekas, D.P. 1998.
Bertsekas, D. P., & Castanon, D. A. (1992).
Review: http://www.mit.edu/ dimitrib/Auction_Encycl.pdf



Top: Cost of transportation for all combinations.
Bottom: Optimal assignment

Optimal Matching Results

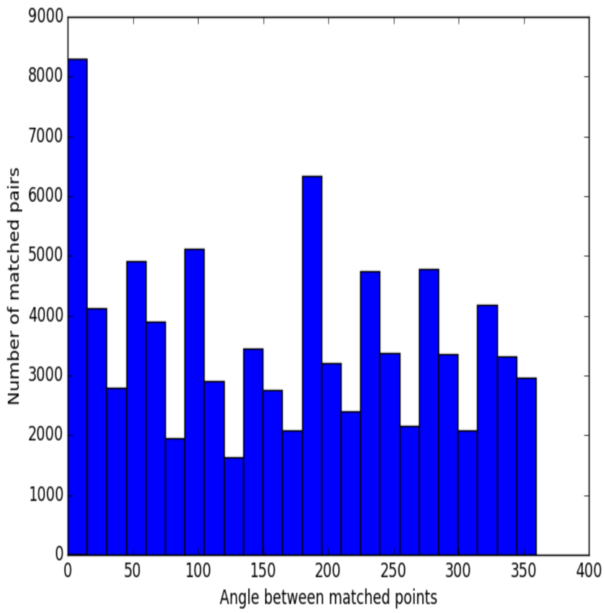
DBSCAN($\epsilon = 2\sqrt{2}, min_n = 5$), for visualization.



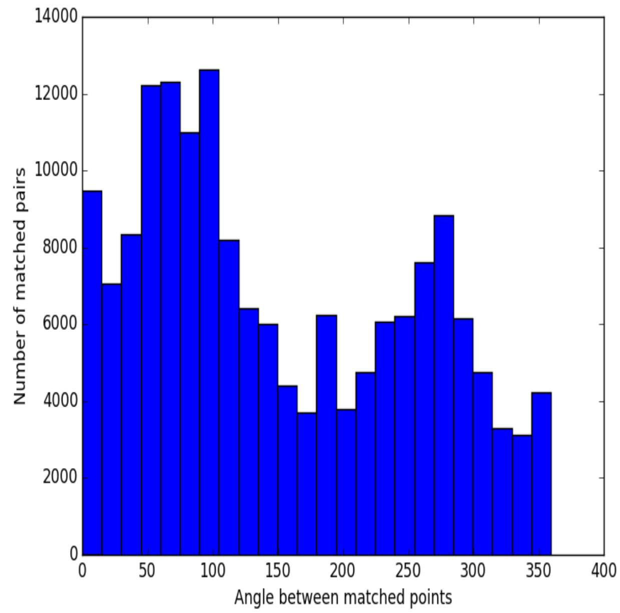
Top to bottom: MM5, COAMPS, WRF, Random Gaussian Field (RGF)

Orientation Error

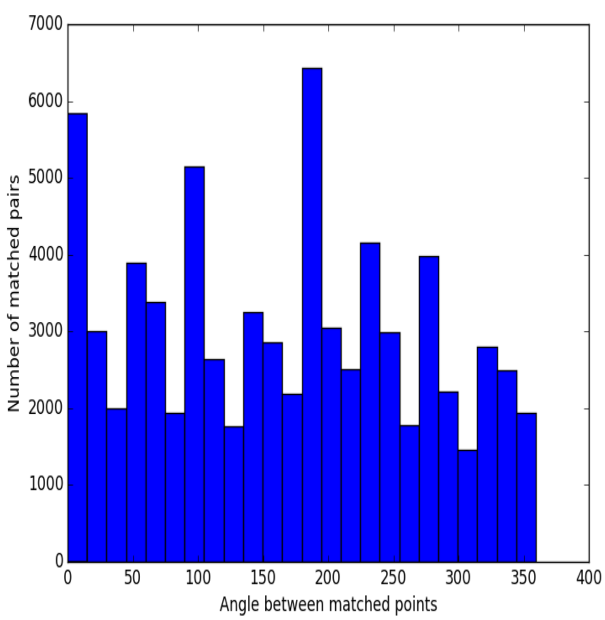
MM5



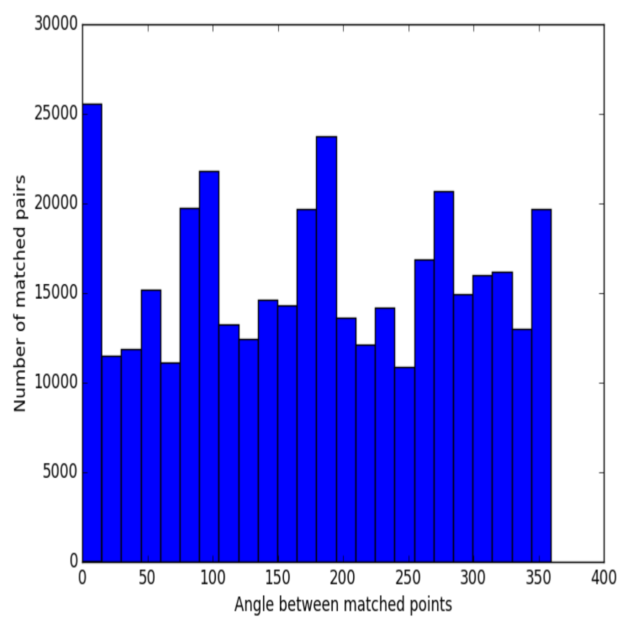
COAMPS



WRF



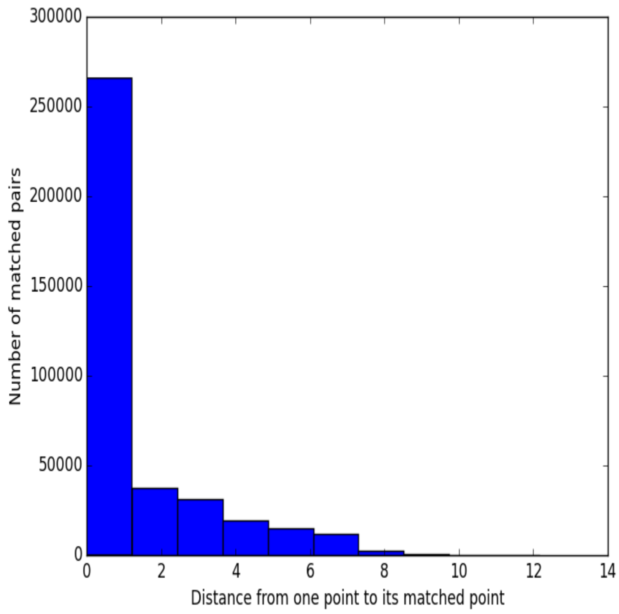
RGF

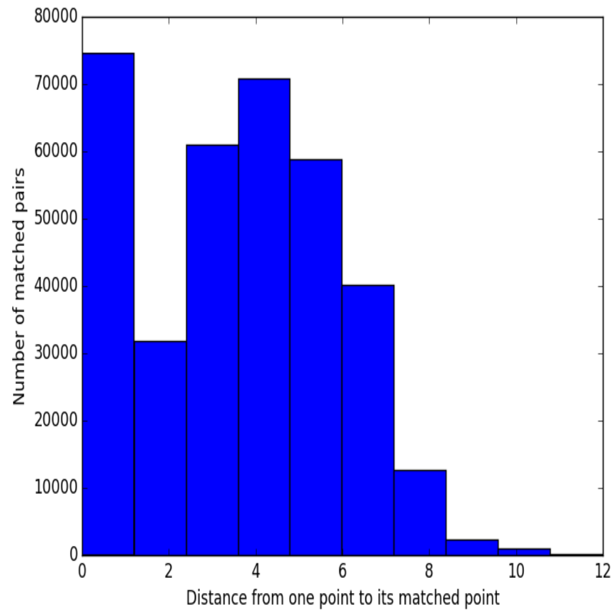


Displacement Error

MM5

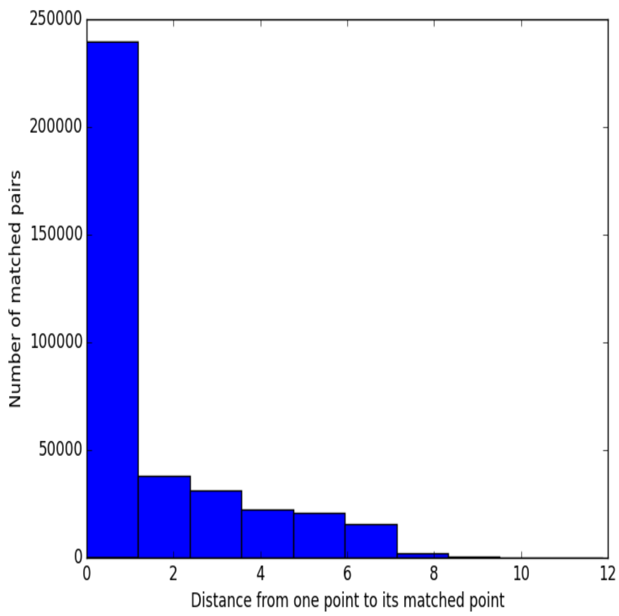
COAMPS

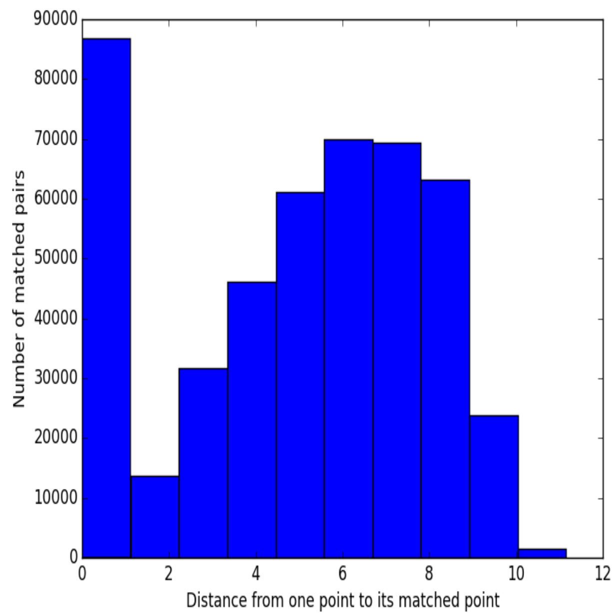




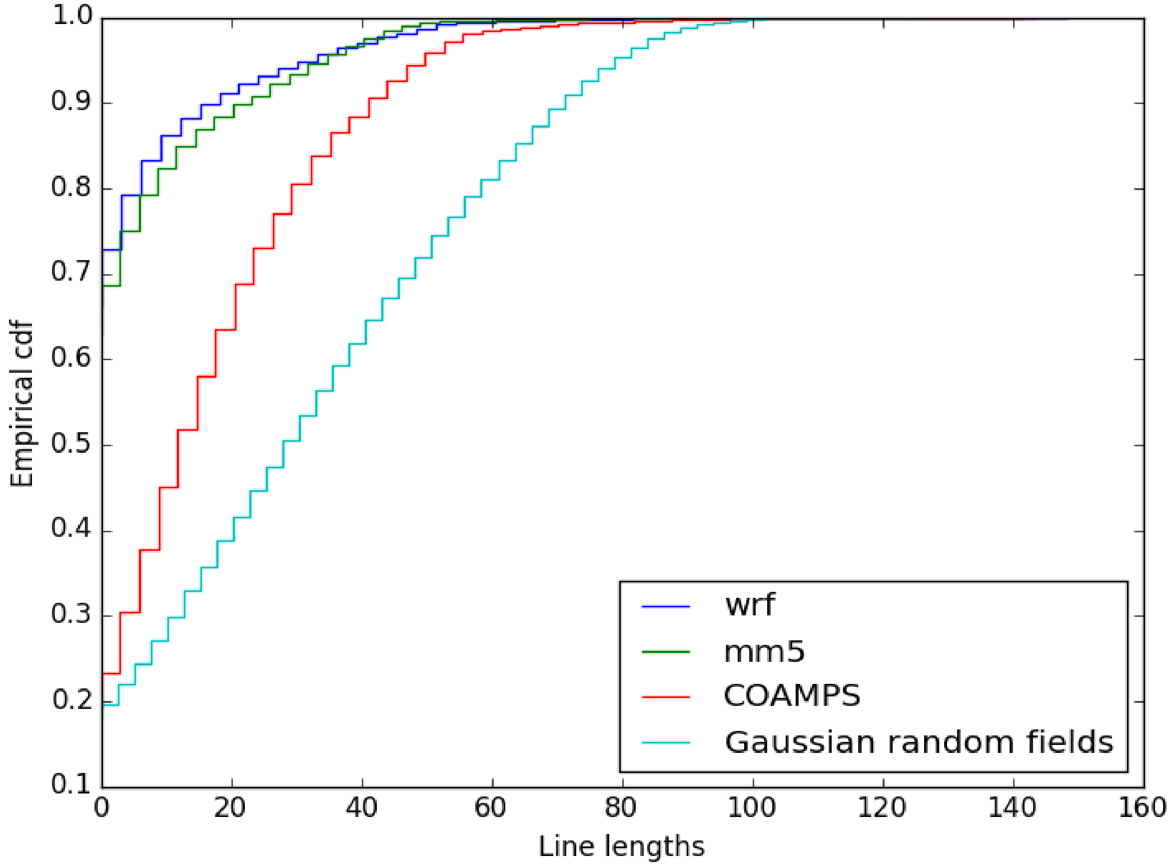
WRF

RGF





Comparing Distributions



Summary and Conclusions

Precip forecasts for MM5 and WRF are comparable
COAMPS not as good.
But our data is old!

The developed methodology for clustering, in conjunction with methods for comparing clusterings, and optimal matching, appears to be useful for assessing forecast quality and for comparing different NWP models. RGFs can provide the necessary null distribution for performing statistical tests.