The Use of Multi-Model Ensemble Clustering in WPC's Extended Range Forecast Experiment

Bill Lamberson[1,2], Mike Bodner[2], Sara Sienkiewicz[1,2], and Jim Nelson[2]

[1]I.M. Systems Group

[2]Weather Prediction Center

8th NCEP Ensemble Users Workshop

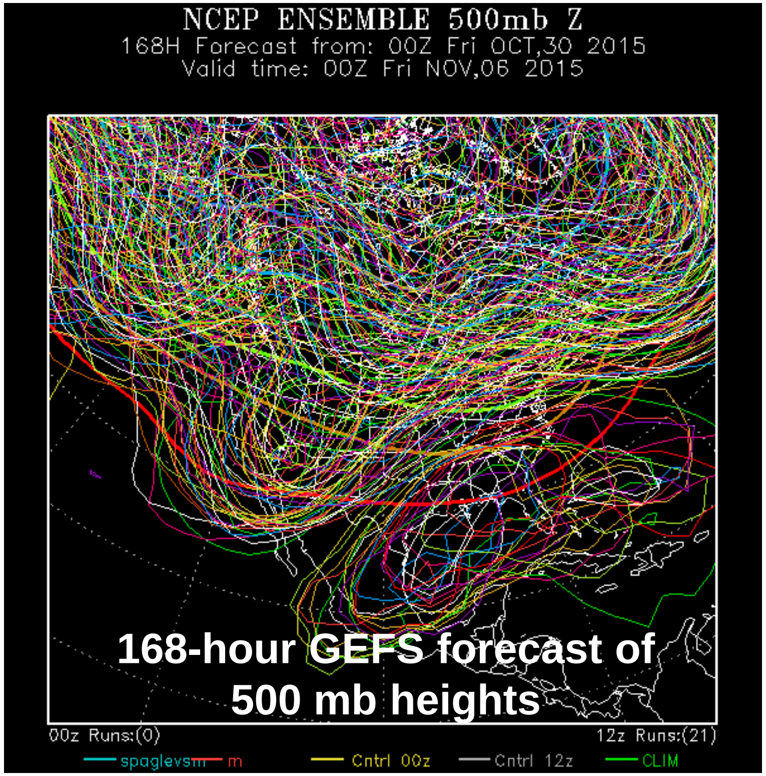
28 August 2019

The Meteorological Problem

- WPC's Extended Range Forecast Experiment (ERFE) tasked with investigating 8-10 day forecasts of maximum temperature, minimum temperature, and precipitation.

- Since the atmosphere is a chaotic system, forecasting at these longer lead times requires the use of ensembles.

- Ensembles show the uncertainty and the range of possible forecast scenarios.

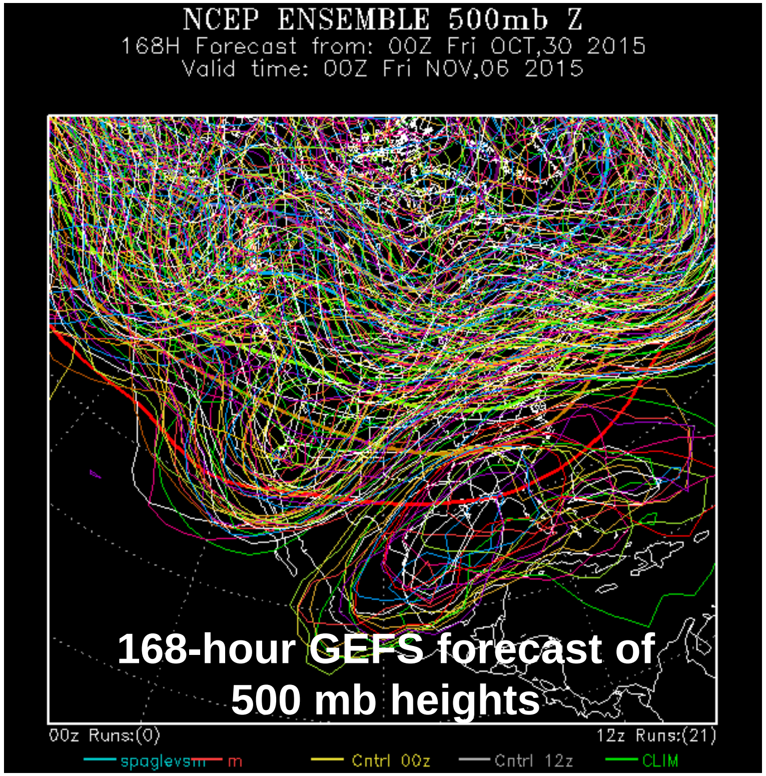


The Meteorological Problem

- Traditionally, extended range forecasts involve viewing and blending ensemble means from each global Ensemble Prediction System (EPS).

- This ignores a host of useful information.

- Forecasters and developers participating in ERFE decided to develop an ensemble clustering tool for days 8-10 to help understand and visualize the potential forecast scenarios contained in the ensemble.



What is Clustering?

- Clustering partitions the members of an ensemble into groups of members that have similar forecasts.

- It has two key benefits to the forecast process:

  1. It can help forecasters quickly assess where the greatest uncertainty in the forecast is.

  2. It can help solve the age old problem of so many ensemble members … so little time. Dont have time to look at all 90 ensemble members of the GEFS, ECMWF, and CMC? Clustering simplifies the onslought of data by distilling the ensemble down into the few forecast scenarios that predominate.

- Many ways to cluster but we use fuzzy clustering.

Fuzzy Clustering Recipe

1. Judiciously select field to cluster on.

2. Calculate and interpret the first two Empirical Orthogonal Functions (EOFs) of that field.

3. Use k-means clustering to create clusters of ensemble members based on their Principal Component (PCs) for the EOFs of the field you selected.

4. View the clusters.

Harr, P. A., D. Anwender, and S. C. Jones, 2008: Predictability associated with the downstream impacts of the extratropical transition of tropical cyclones: Methodology and a case study of Typhoon Nabi (2005). *Mon. Wea. Rev.*, **136**, 3205–3225 https://doi.org/10.1175/2008MWR2248.1

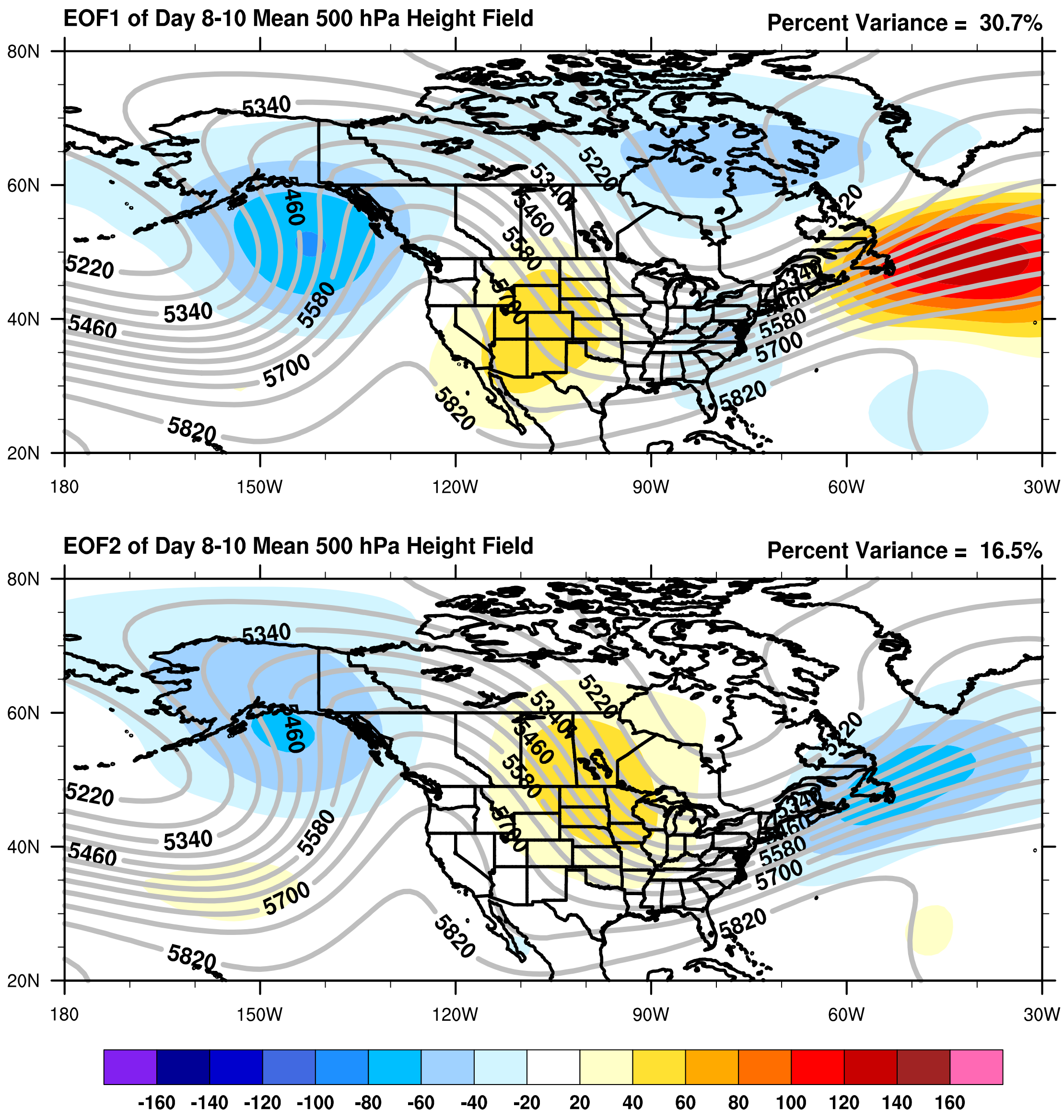
Zheng, M., E. K. Chang, B. A. Colle, Y. Luo, and Y. Zhu, 2017: Applying fuzzy clustering to a multimodel ensemble for U.S. East Coast winter storms: Scenario identification and forecast verification. *Wea. Forecasting*, **32**, 881–903, https://doi.org/10.1175/WAF-D-16-0112.1

1. Select Field to Cluster On

- We are interested in extended-range (8-10 day) forecasts.
- Long-wave pattern is more predictable at this time-range than anything else.
- We cluster on 500-hPa heights from the GEFS, ECMWF ENS, and GEPS during days 8-10 over North America.
- EOFs that are calculated across the model ensemble member dimension result in modes that show the dominant patterns of the differences between individual ensemble members and the ensemble mean.
- EOFs will show differences in the long-wave pattern for days 8-10 between the ensemble members.

2. Calculate and Interpret EOFs

Init: 0000 UTC 30 Nov 2017

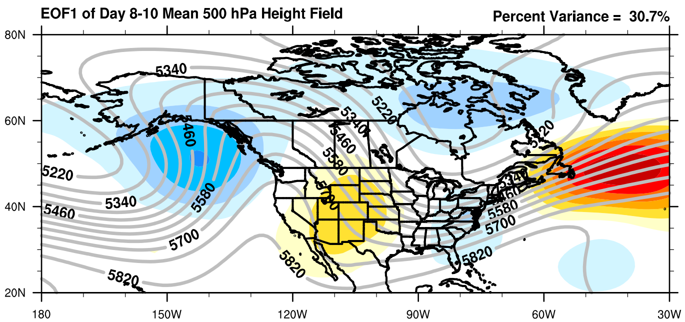


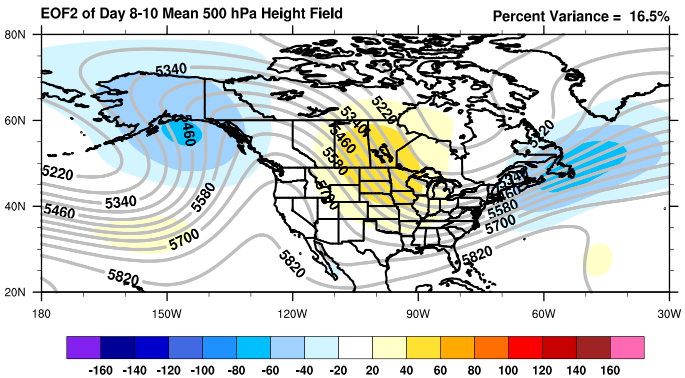
- EOF patterns are dipoles on either side of the trough and ridge axes.
- Amplitude is more certain
- What is uncertain is location (i.e., how progressive is the pattern) and orientation (i.e., how tilted will the trough and ridge be).

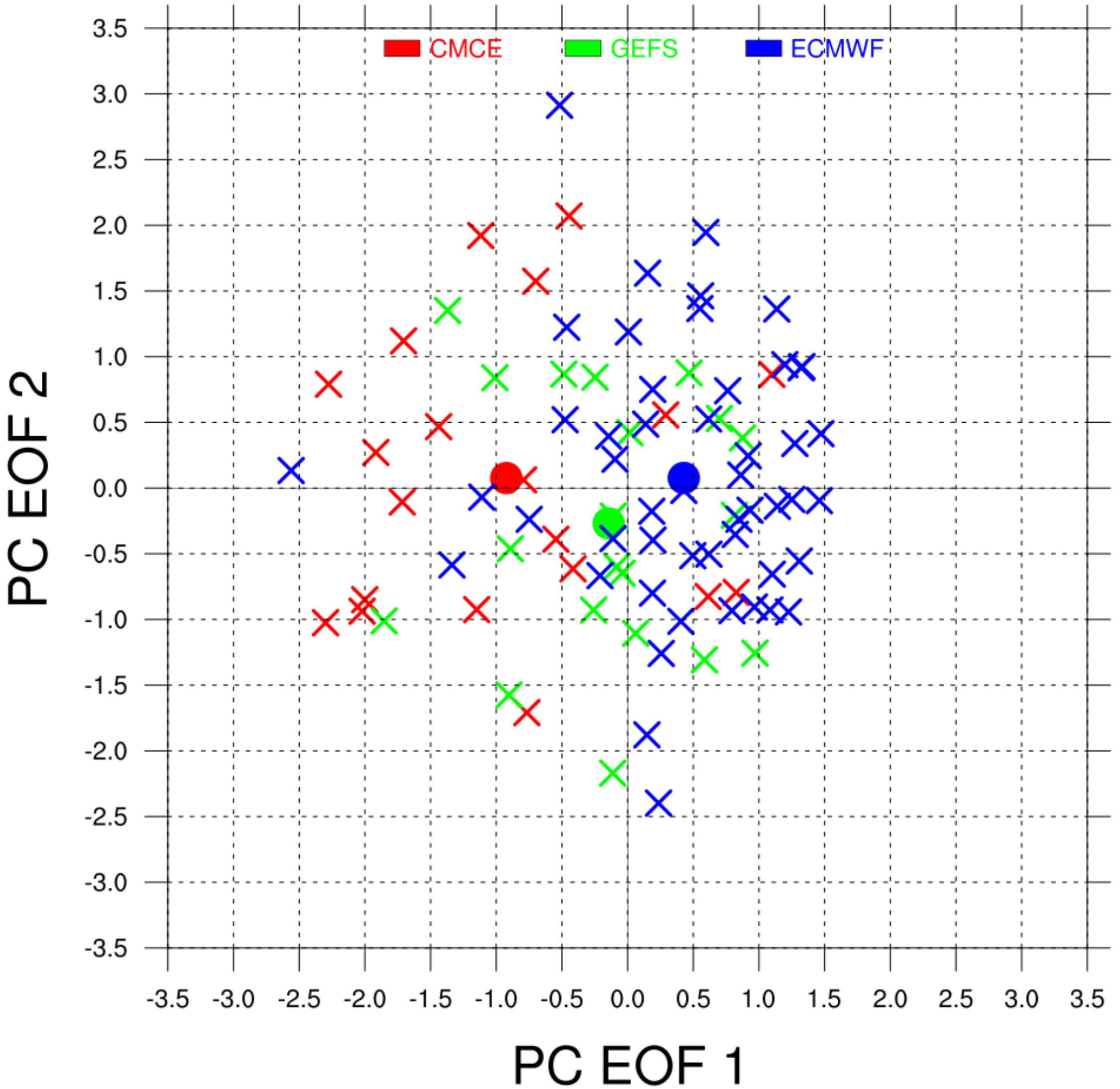
2. Calculate and Interpret EOFs

- **EOF1: East-west differences in long-wave pattern placement and how tilted the trough and ridge are.**

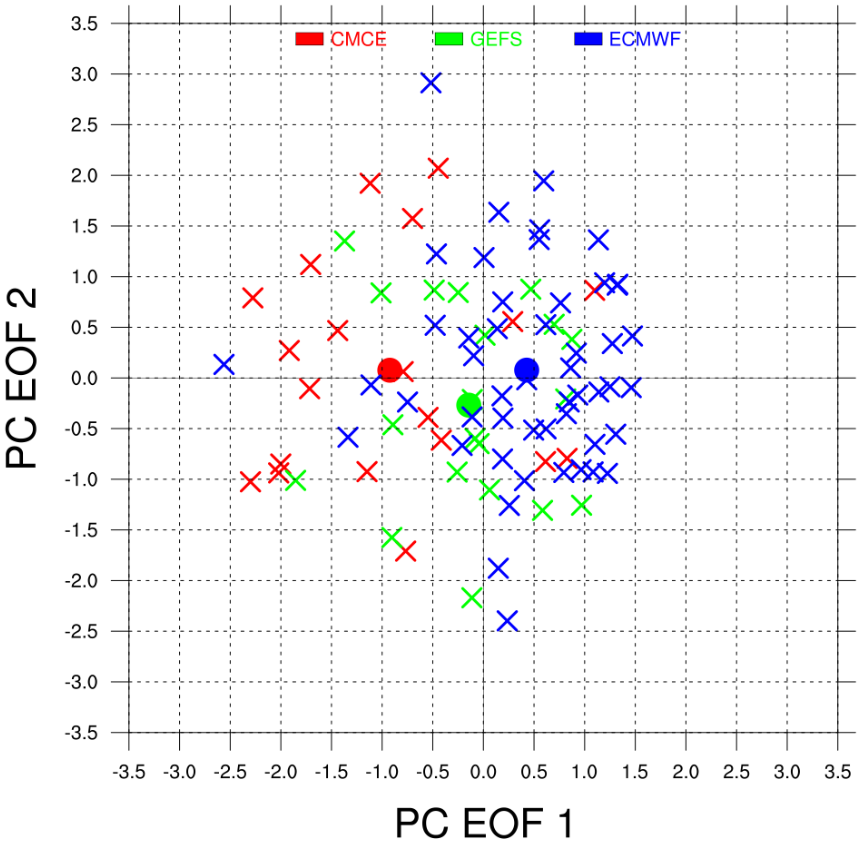
- **EOF2: Same**



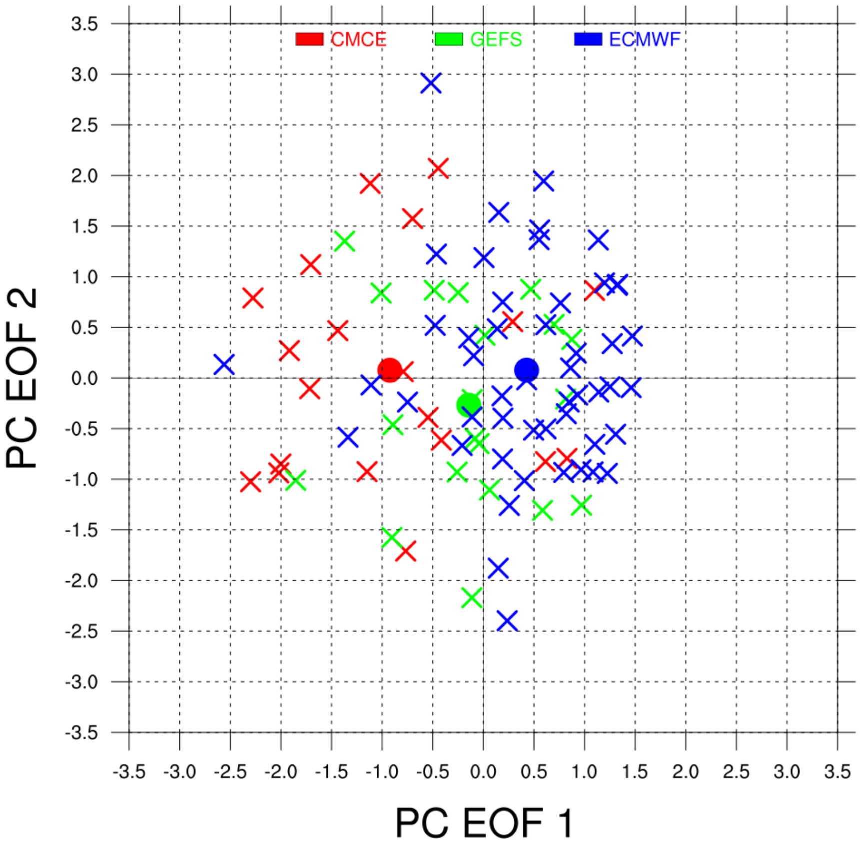


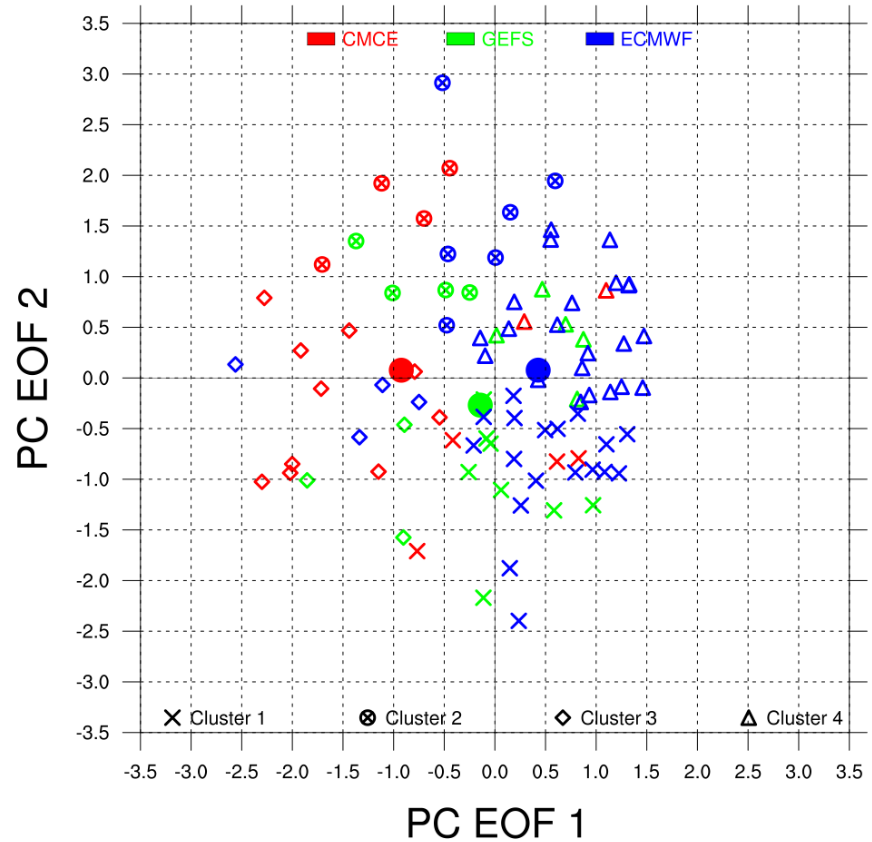


3. Use k-means Clustering

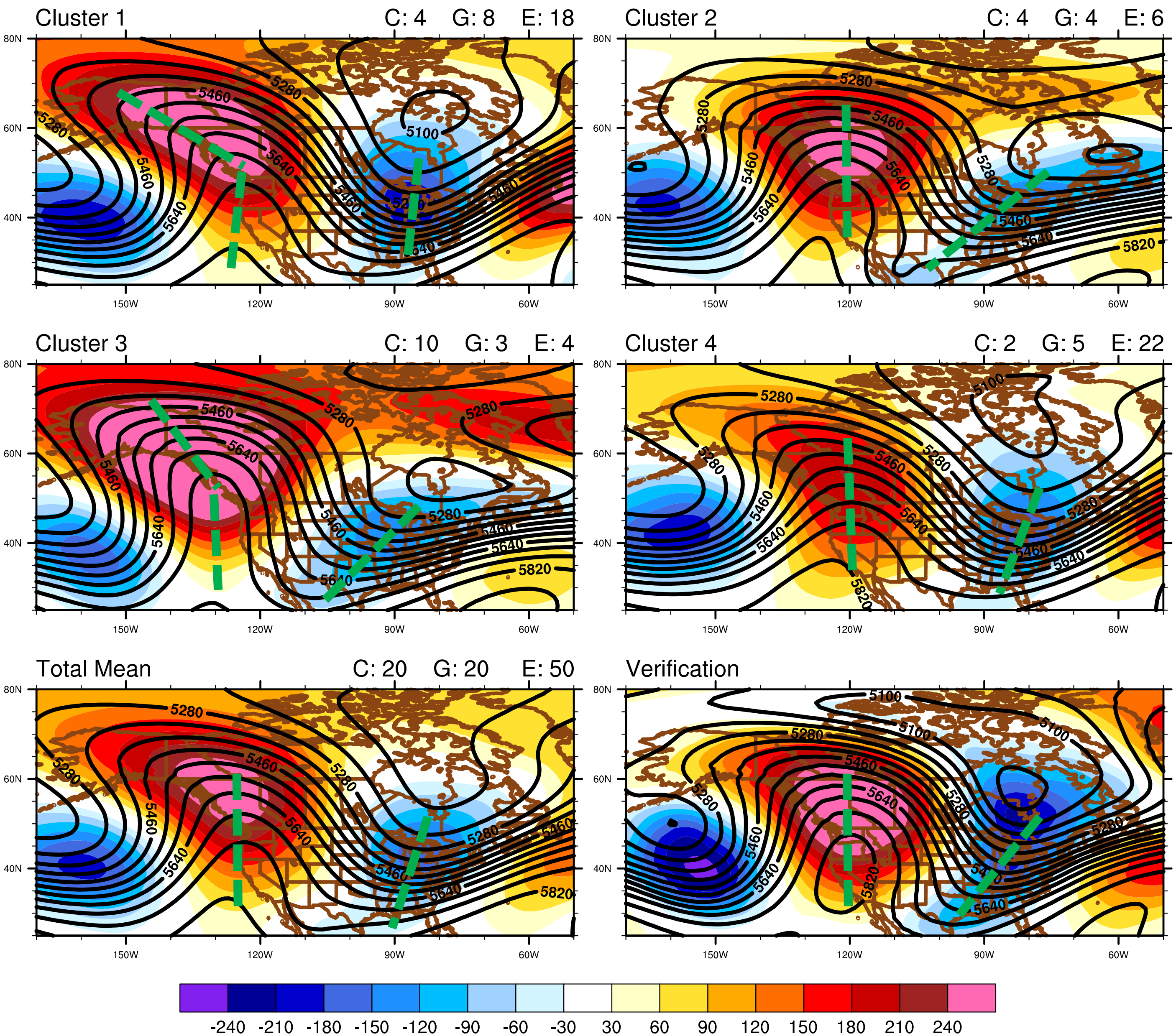


3. Use k-means Clustering





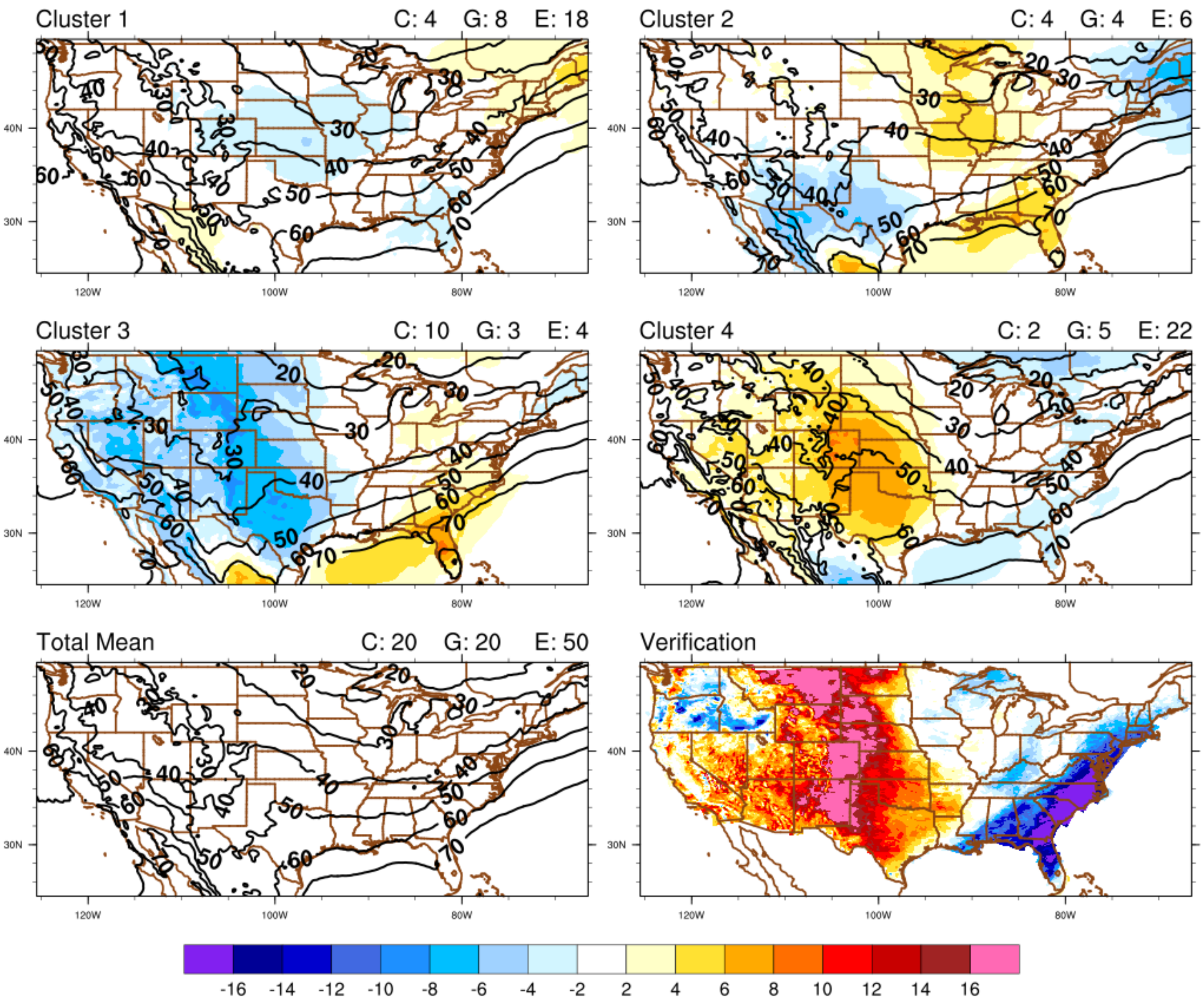
4. View Clusters (Forecast Scenarios)



Cluster mean day 8-10 mean 500-hPa height field (contoured) and anomalies when compared to the CFSR climatology (color fill) for each cluster, the ensemble mean of all 90 members and the verifying 8-10 day mean 500-hPa height field.

4. View Clusters (Forecast Scenarios)

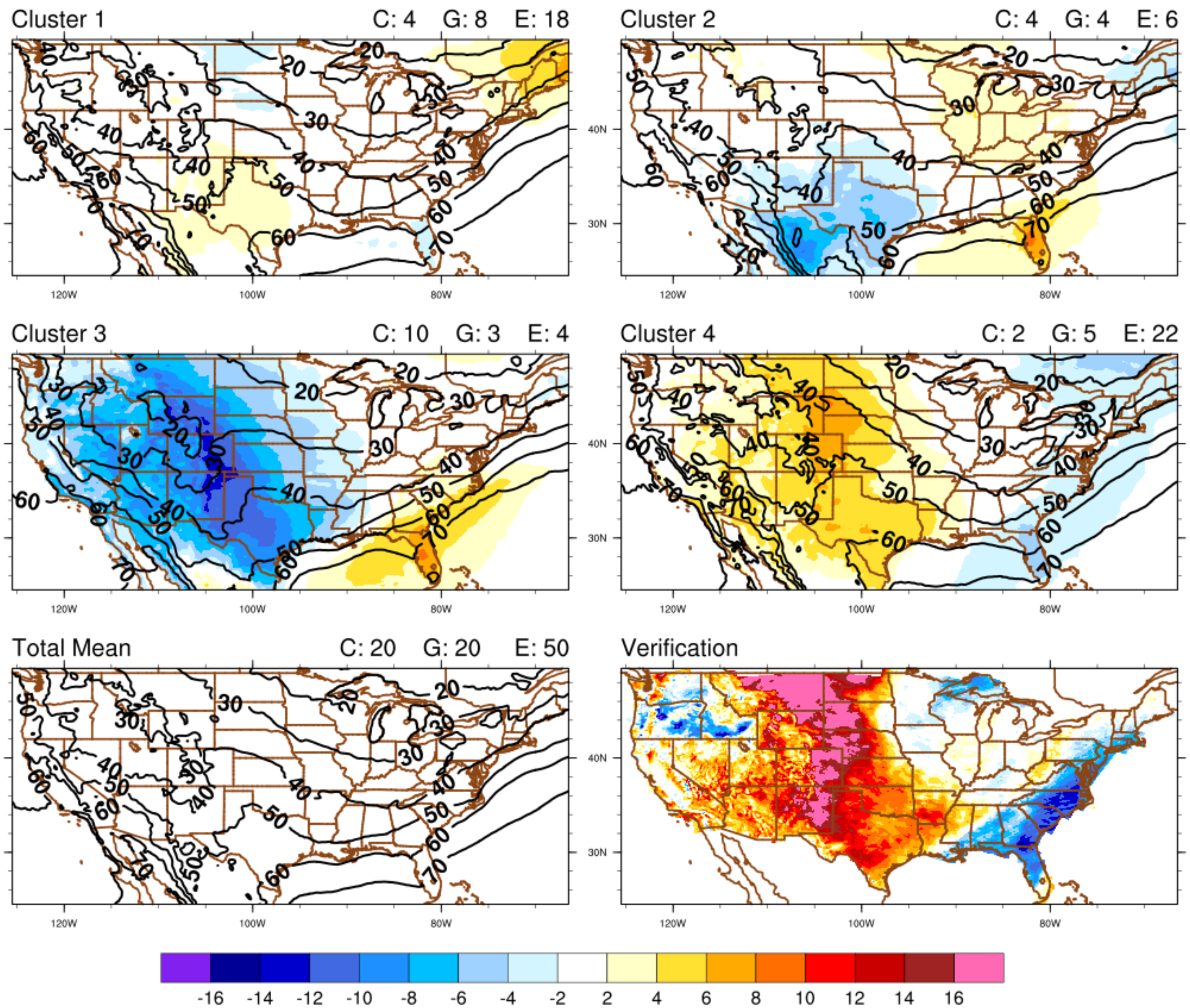
Day 8 Maximum Temperatures



Cluster mean day 8 maximum temperatures (contoured) and difference from the ensemble mean of all 90 members (color fill) for each cluster, the ensemble mean of all 90 members, and the verifying day 10 maximum temperatures

4. View Clusters (Forecast Scenarios)

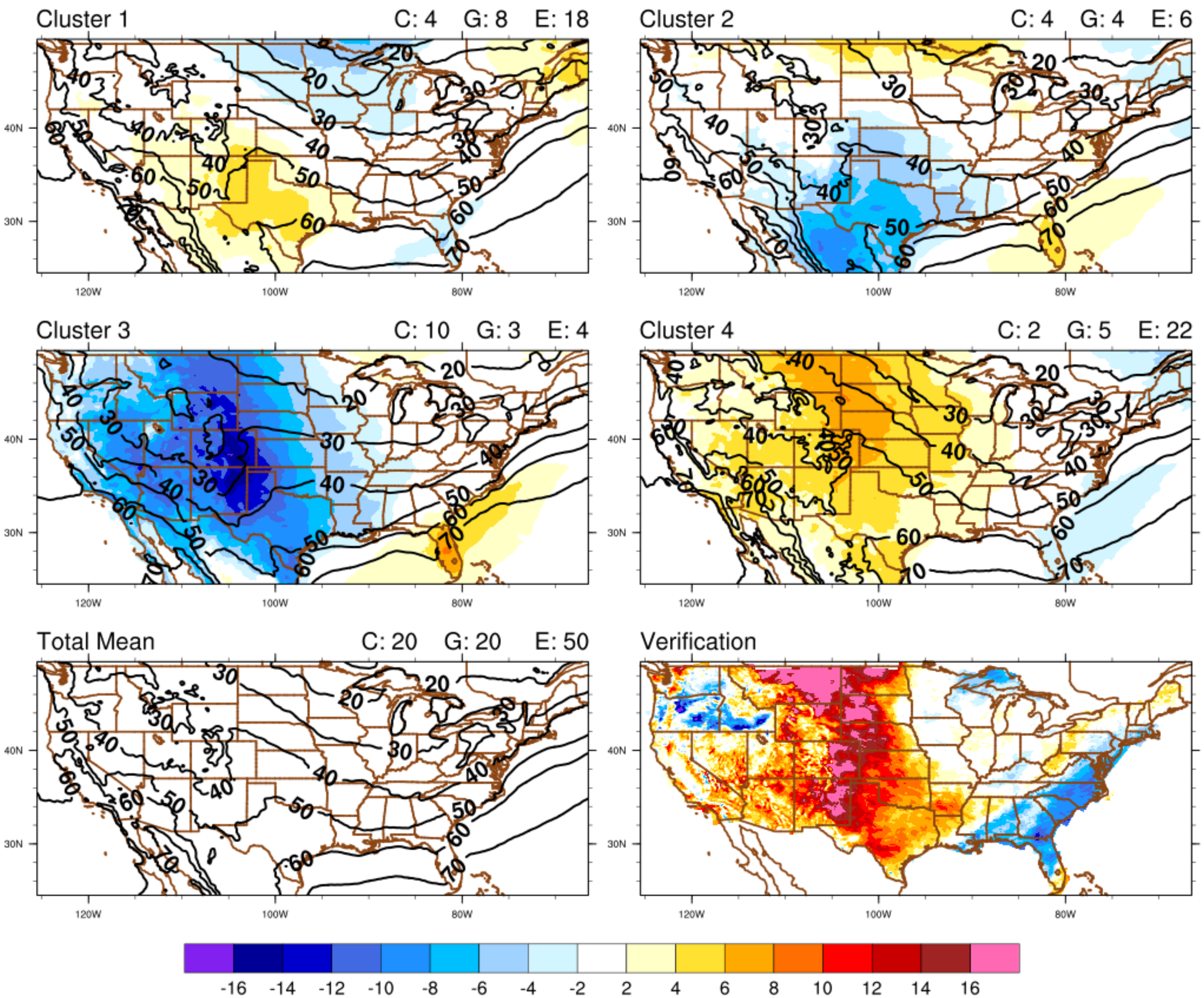
Day 9 Maximum Temperatures



Cluster mean day 9 maximum temperatures (contoured) and difference from the ensemble mean of all 90 members (color fill) for each cluster, the ensemble mean of all 90 members, and the verifying day 10 maximum temperatures

4. View Clusters (Forecast Scenarios)

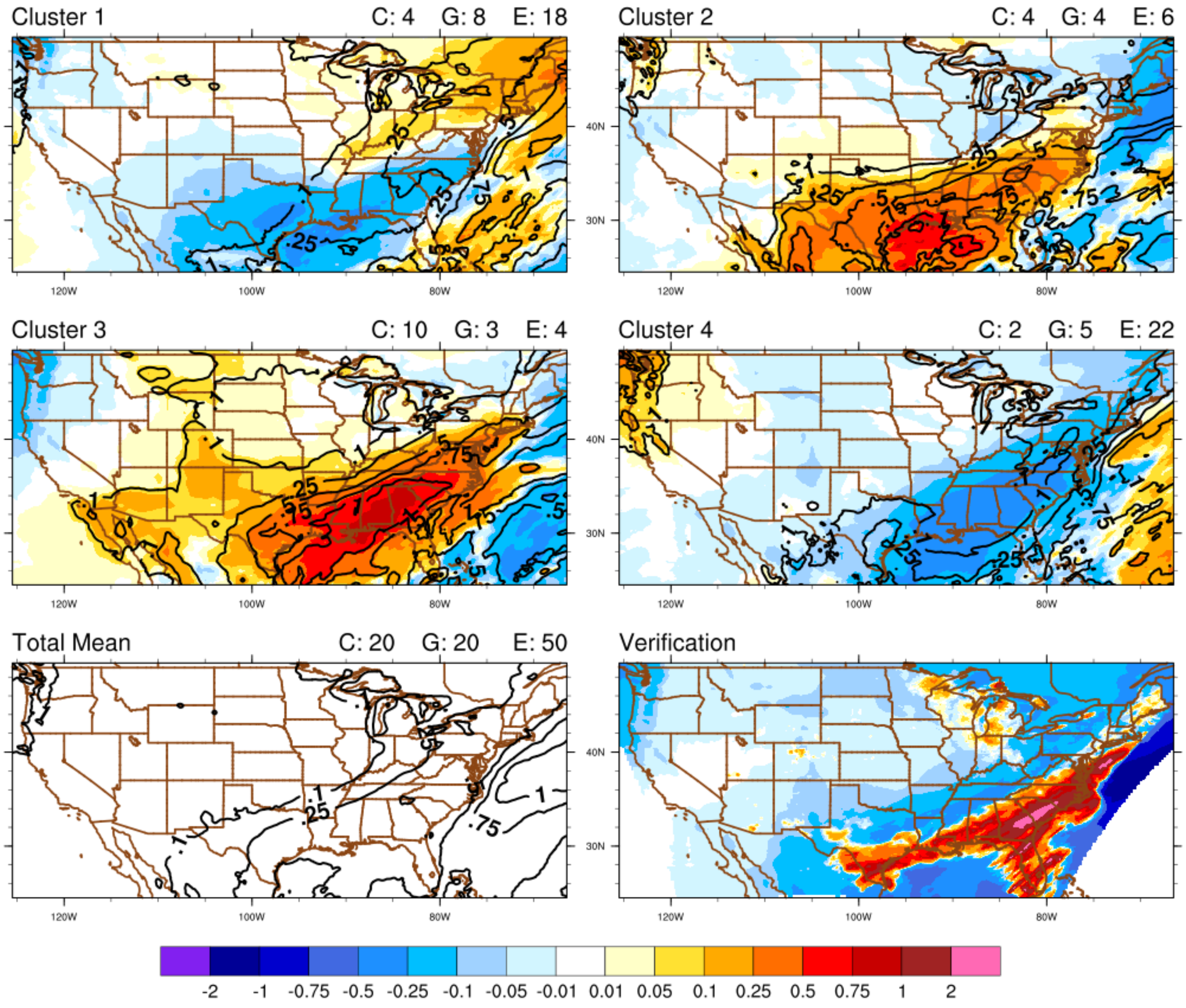
Day 10 Maximum Temperatures



Cluster mean day 10 maximum temperatures (contoured) and difference from the ensemble mean of all 90 members (color fill) for each cluster, the ensemble mean of all 90 members, and the verifying day 10 maximum temperatures

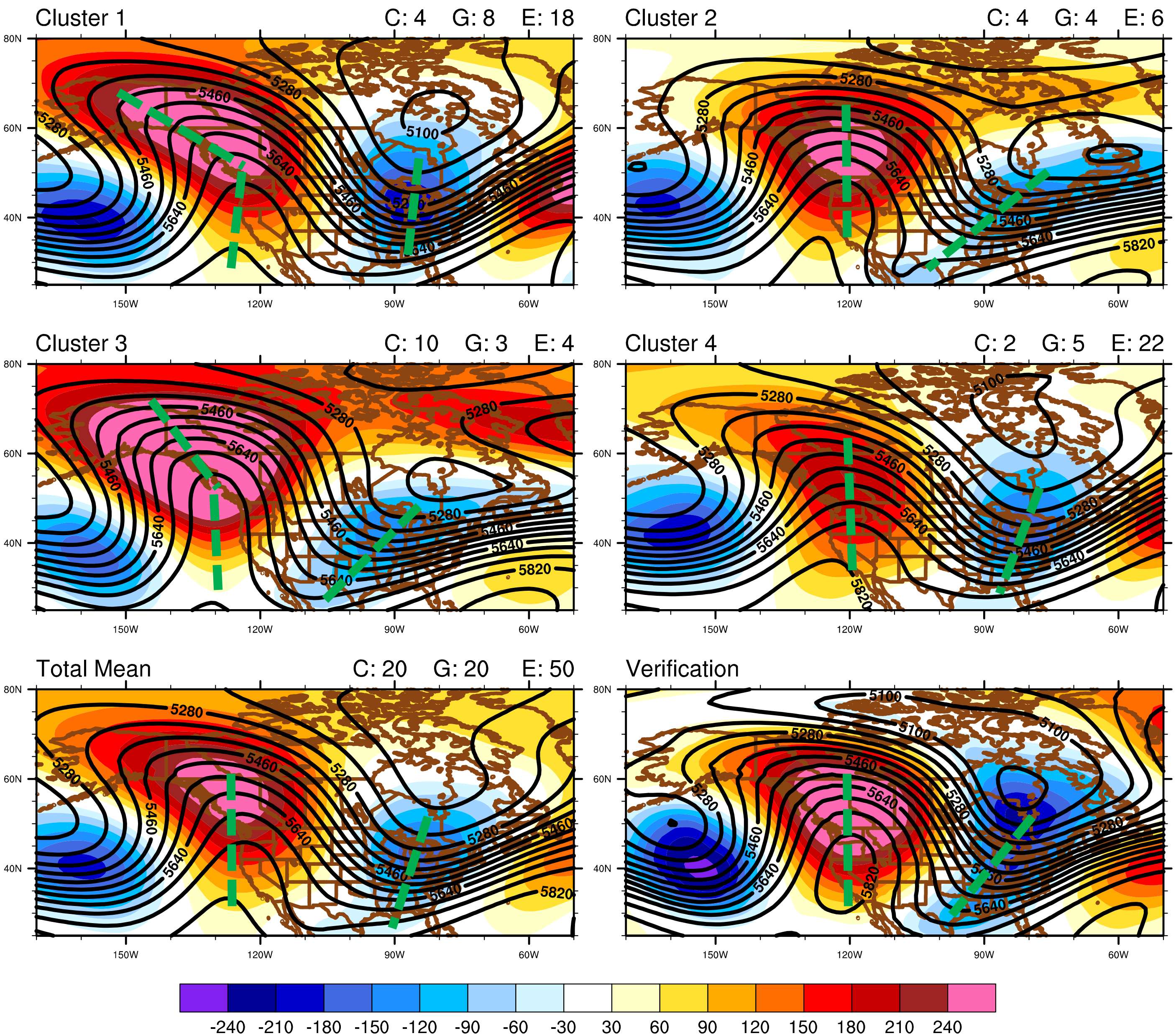
4. View Clusters (Forecast Scenarios)

Day 8-10 Precipitation



Cluster mean day 8-10 precipitation (contoured) and difference from the ensemble mean of all 90 members (color fill) for each cluster, the ensemble mean of all 90 members, and the verifying day 10 maximum temperatures

4. View Clusters (Forecast Scenarios)



Cluster mean day 8-10 mean 500-hPa height field (contoured) and anomalies when compared to the CFSR climatology (color fill) for each cluster, the ensemble mean of all 90 members and the verifying 8-10 day mean 500-hPa height field.

Summary and Future Work

- Applied fuzzy clustering to multi-model 500 hPa height forecasts for days 8-10 in order to see the range of possible forecast scenarios for this time period.

- This clustering method does a good job of breaking the ensemble forecast down to a few key and dominant forecast scenarios.

- It greatly enhances forecaster situational awareness.

- Also plan to apply this methodology to other forecasting challenges faced by WPC.

Thanks!

Questions or Comments?