



R202R in Action: Ensemble Clustering and Sensitivity Analysis

Austin Coleman, Jim Nelson, Brian Colle, & Travis Wilson

Motivation

- As technology improves and NWS responsibilities expand
 - Forecasters have access to **more data** with simultaneously **less time** to interrogate those data
- The National Blend of Models (NBM) is frequently used as a first-guess for said forecasts
 - Blends a large amount of forecast data, but can be seen as a black box
 - Forecasters desire more information about what makes up the NBM
- Ensemble mean of NBM's sub-ensemble systems (CMCE, GEFS, and ECMWF) is one way to quickly summarize solutions
 - **Problem:** it often washes out important nuance amongst ensemble membership

Motivation

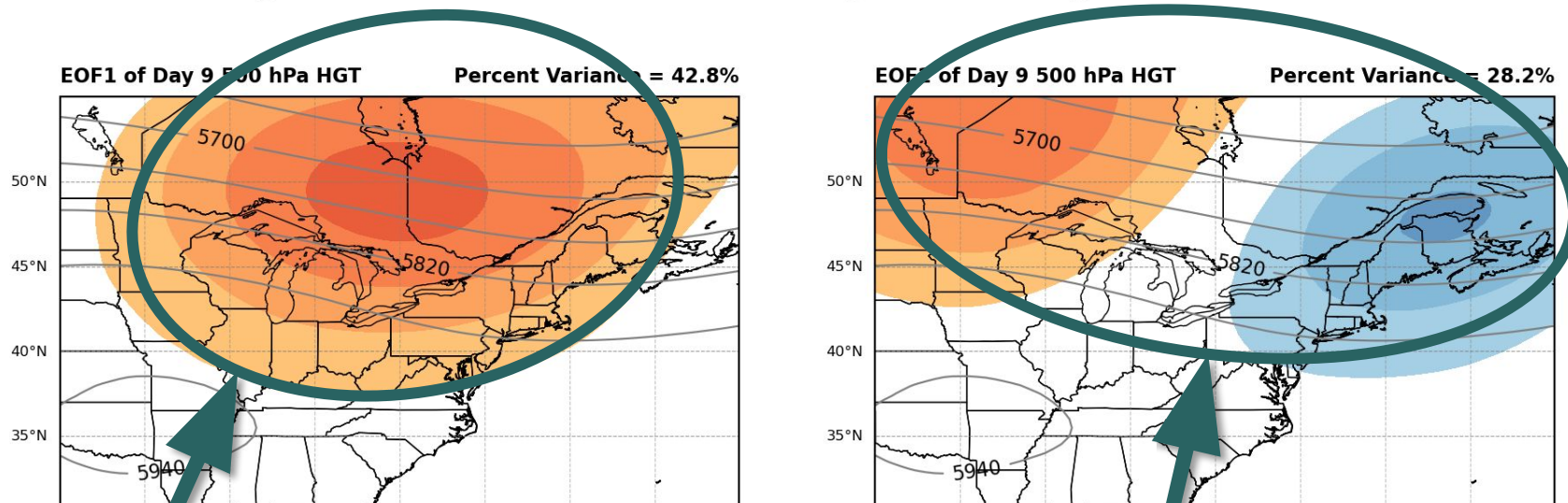
Solution? Develop a clustering approach to break down an ensemble forecast into its most prevalent scenarios!

Inspired by and in collaboration with fuzzy clustering work conducted at Stony Brook

How does it work?

First, we break down the forecast (super-ensemble of CMCE, GEFS, & ECMWF) into its leading modes of variability via EOF Analysis (traditionally known as PCA)

EOFs of 24-hour Mean 500-hPa Heights [meters]
Init: 00Z Fri Aug 11 2023 --> Valid: 24-hours Ending 00Z Mon Aug 21 2023



Leading mode of uncertainty:
Amplitude of pattern

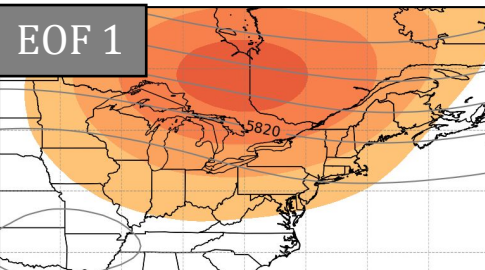
Secondary mode of uncertainty:
W-E positioning of pattern



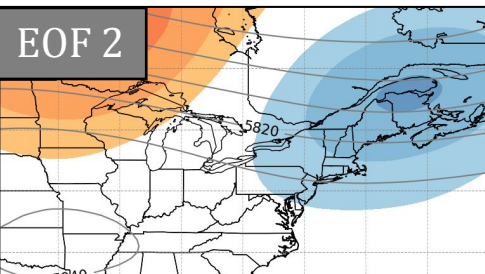
How does it work?

Next, we use k-means clustering to assign members to cluster scenarios

First two EOFs for reference



Leading uncertainty mode:
Amplitude of pattern



Secondary uncertainty mode:
W-E pattern position

Principal component (PC) phase space shows us the forecast scenario for each ensemble member (and system)

Members with positive PC1 will look more like EOF1 whereas members with negative PC1 will look opposite

K-means clustering groups members with similar scenarios

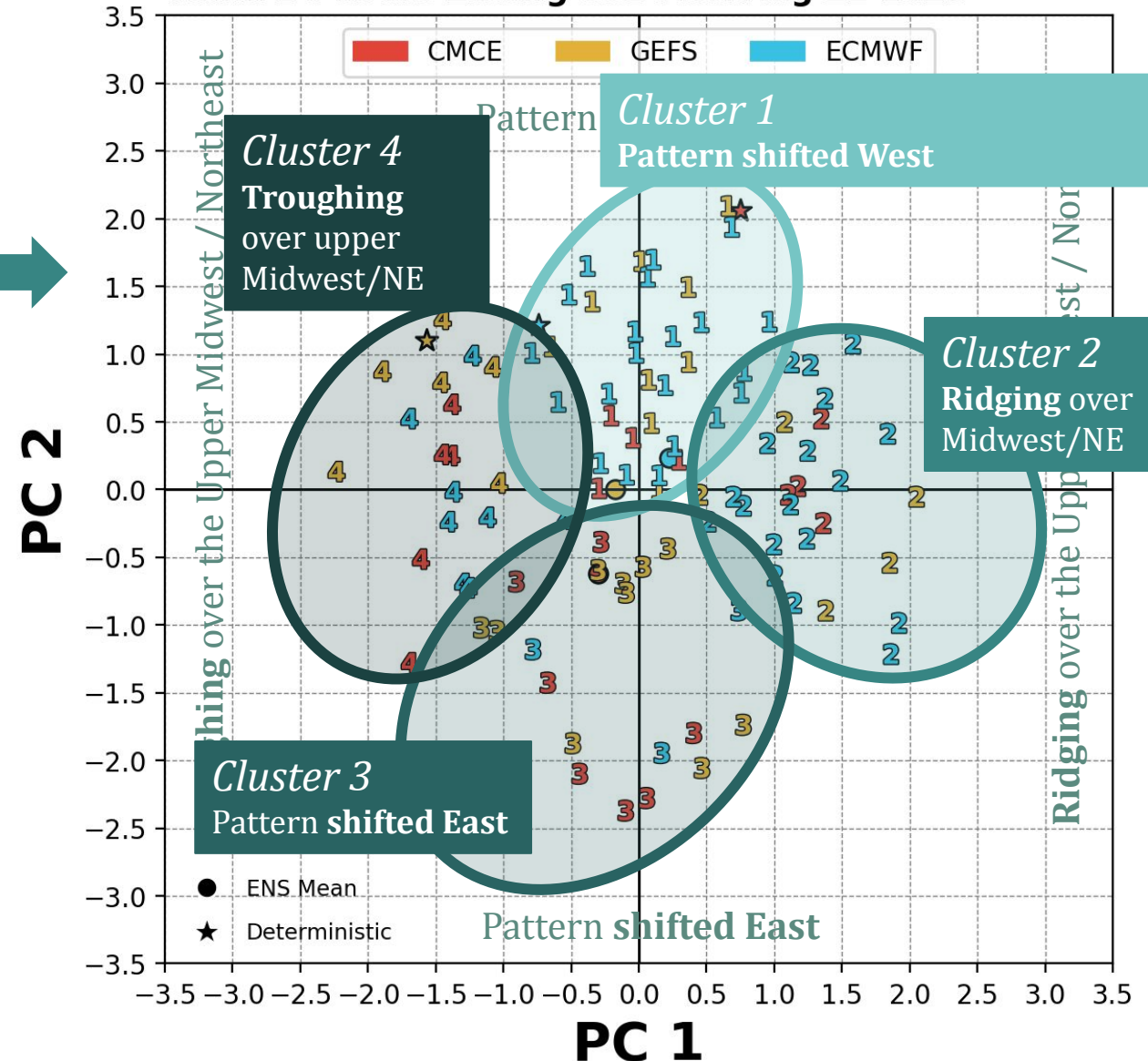
(WPC pg keeps # clusters fixed at 4)



PC1-PC2 Phase Space

Init: 00Z Fri Aug 11 2023 -->

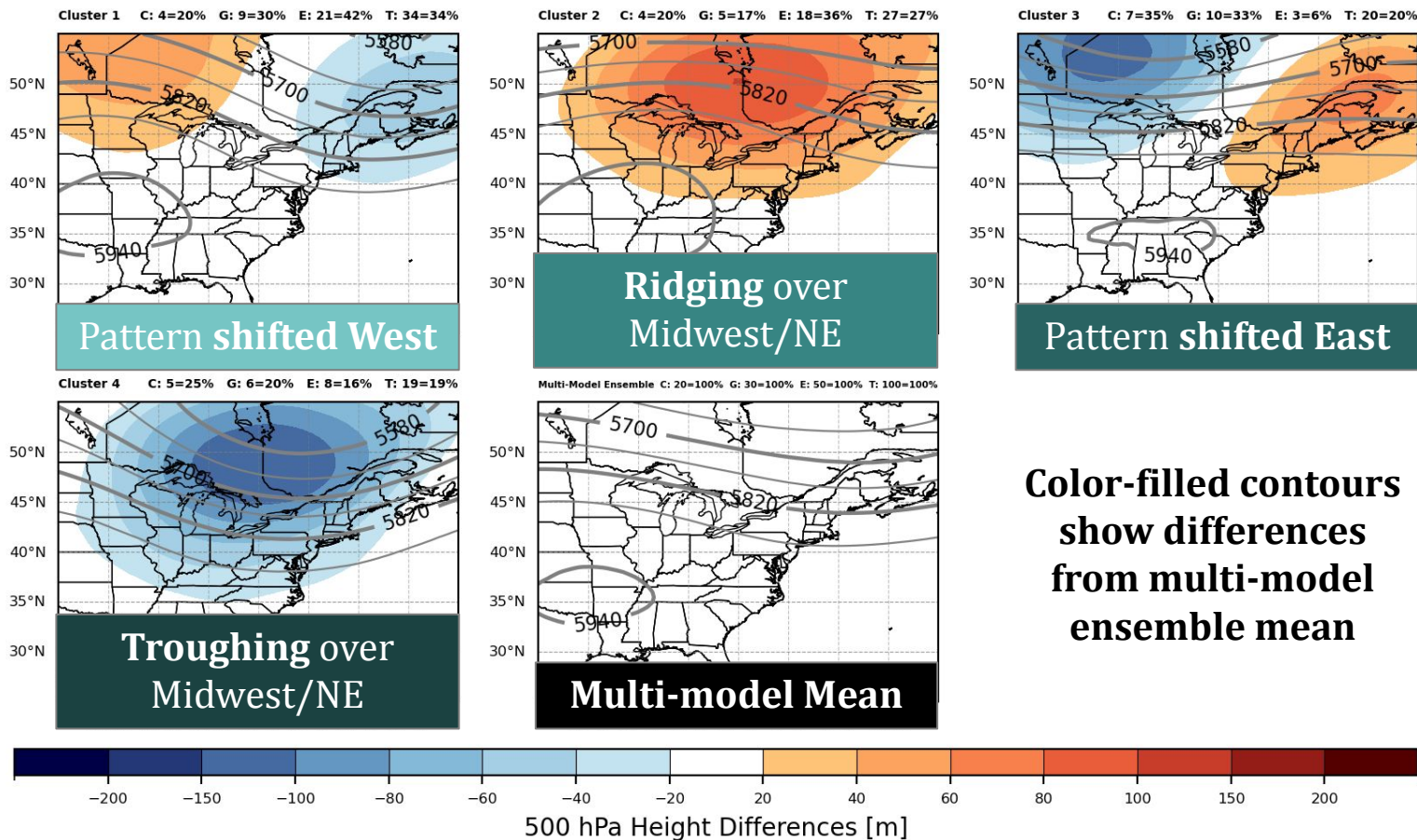
Valid: 24-hours Ending 00Z Mon Aug 21 2023



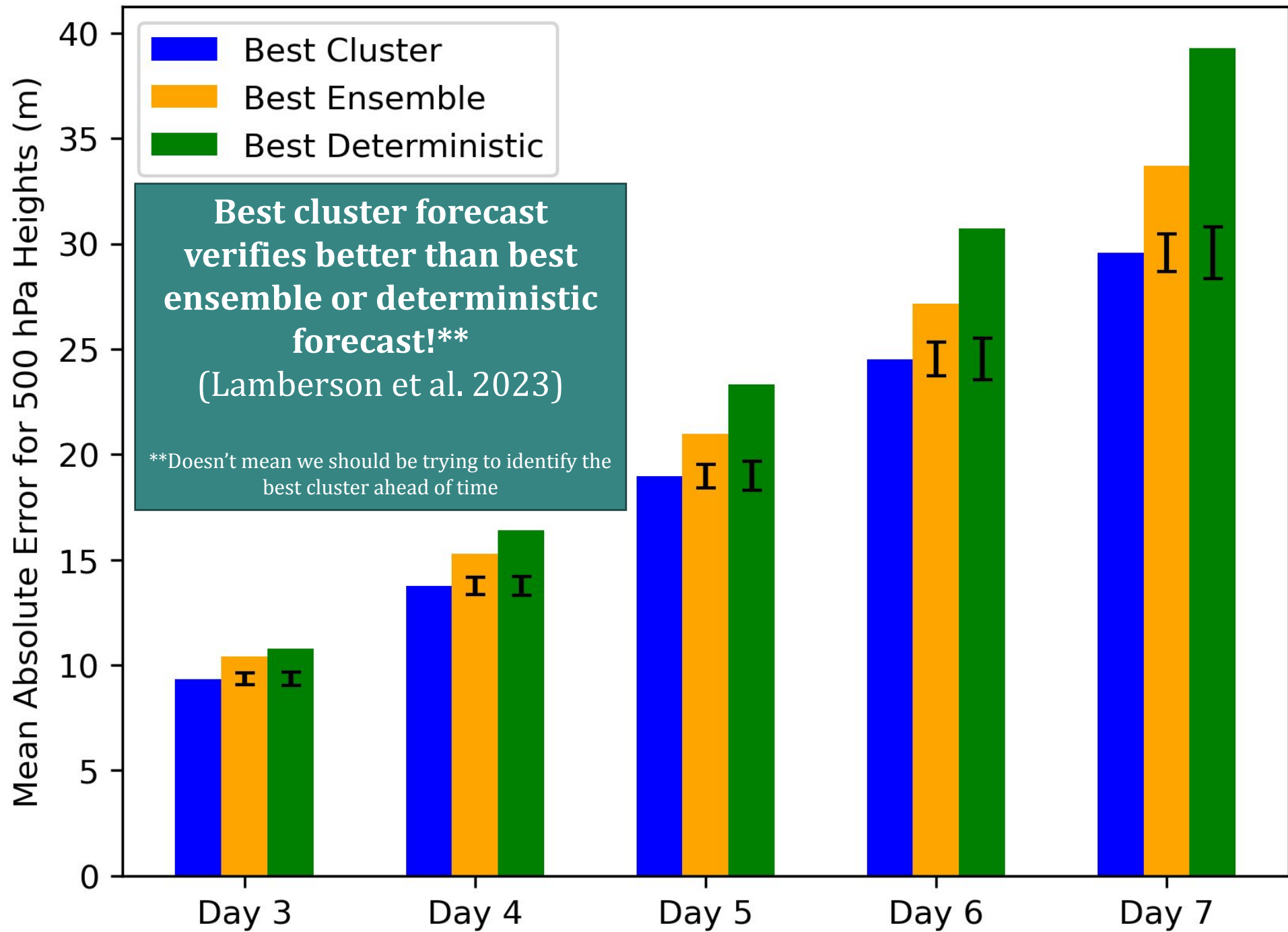
Don't even need to look at EOFs or PCs to use!

Can skip straight to the cluster forecasts (of 500-hPa heights in this case)

Cluster Mean 24-hour Mean 500-hPa Heights and Difference from Multi-Model Mean [m]
Init: 00Z Fri Aug 11 2023 --> Valid: 24-hours Ending 00Z Mon Aug 21 2023



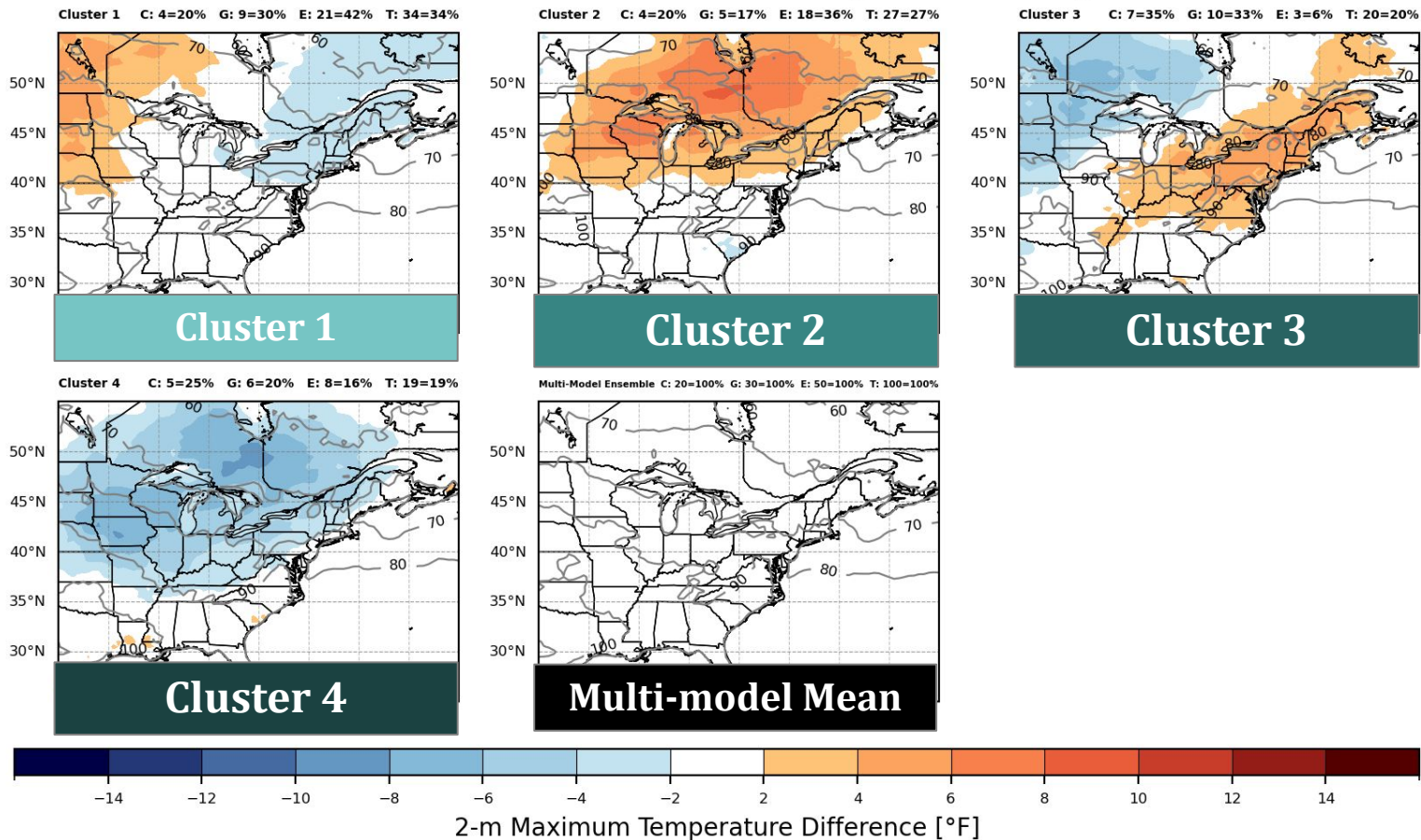
Do
Can



Can use 500-hPa height clusters to predict other fields

Maximum Temperatures

2-m Maximum Temperature Difference from Multi-Model Mean [°F]
Init: 00Z Fri Aug 11 2023 --> Valid: 24-hours Ending 00Z Mon Aug 21 2023

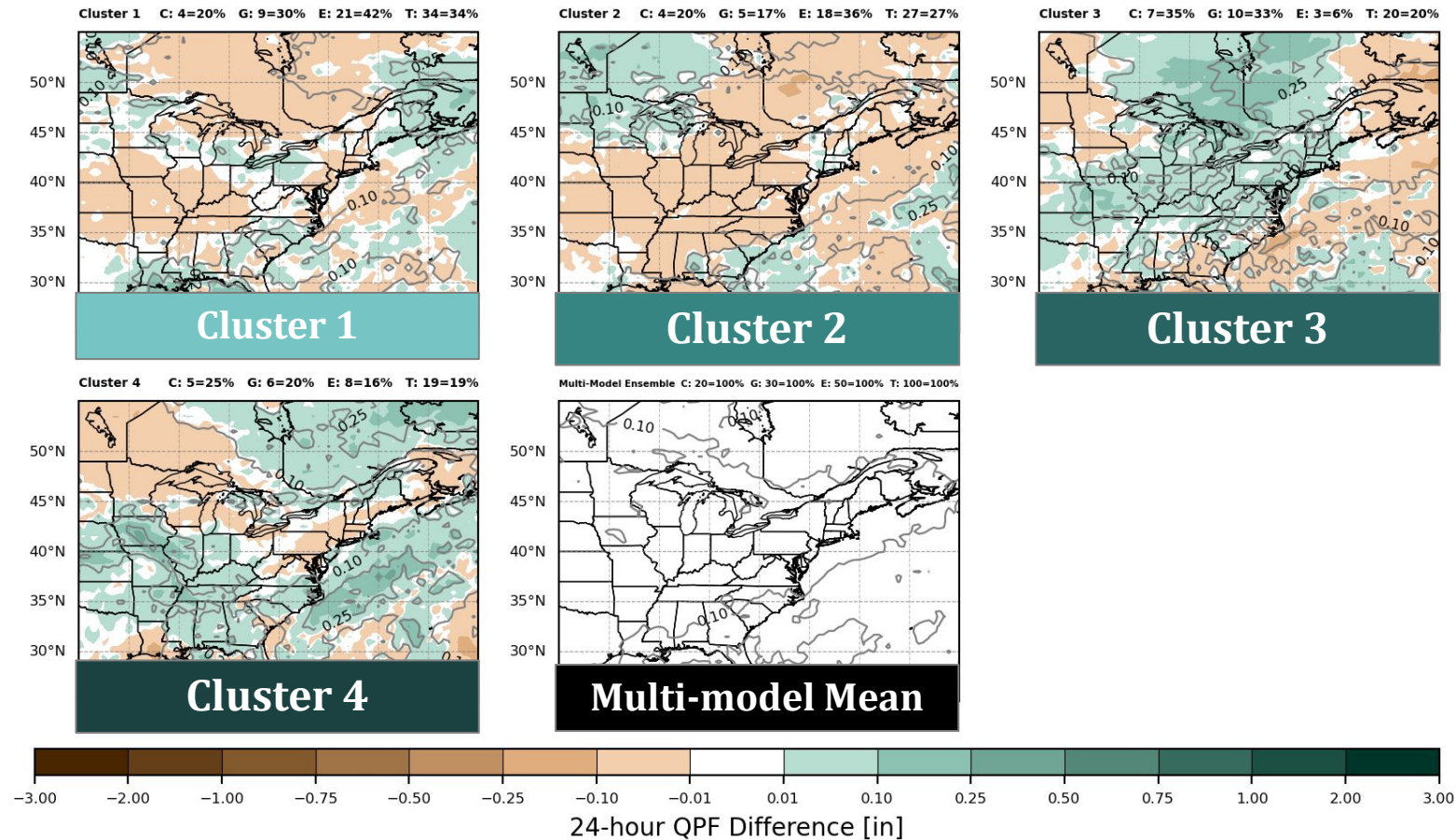


Can use 500-hPa height clusters to predict other fields

24-hr QPF

24-hour QPF Difference from Multi-Model Mean [in]

Init: 00Z Fri Aug 11 2023 --> Valid: 24-hours Ending 00Z Mon Aug 21 2023

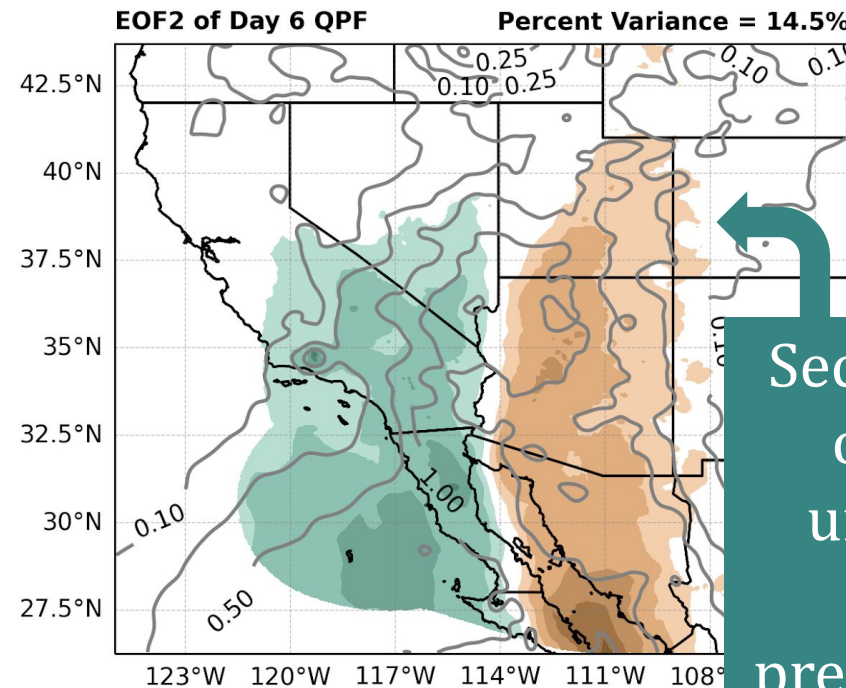
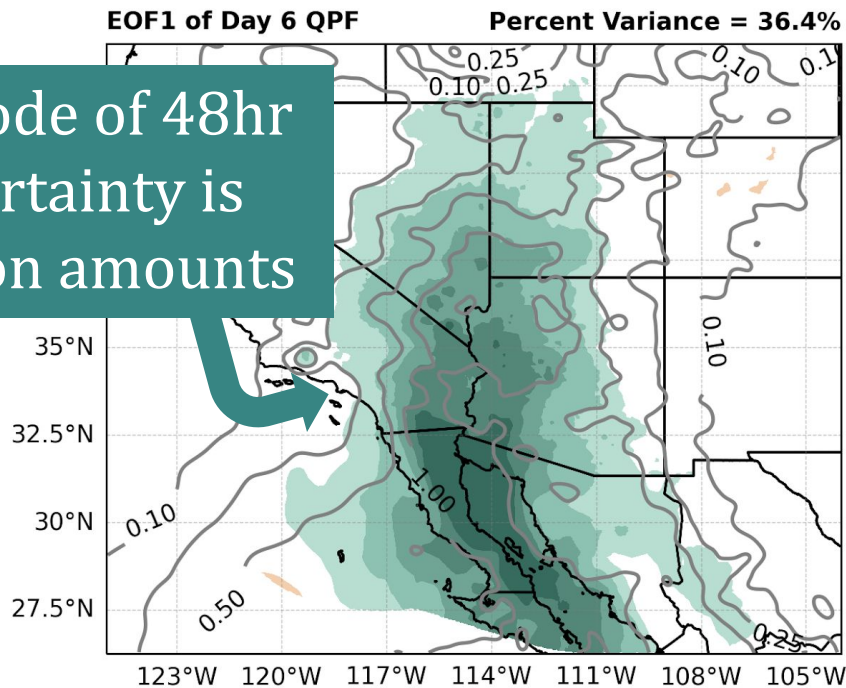


We also have a page that clusters on QPF uncertainty!

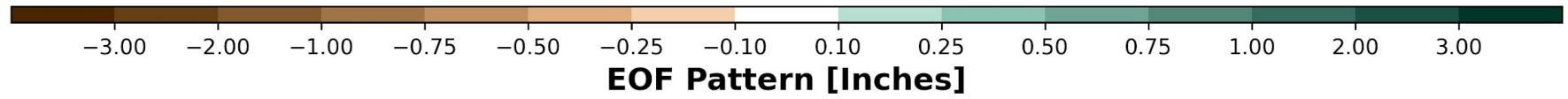
Based on NBM QMDs (100 members)

EOFs of 48-hour QPF [Inches]
Init: 0000 UTC Tue Aug 15 2023 --> Valid: 48-hours Ending 0000 UTC Tue Aug 22 2023

Leading mode of 48hr QPF uncertainty is precipitation amounts



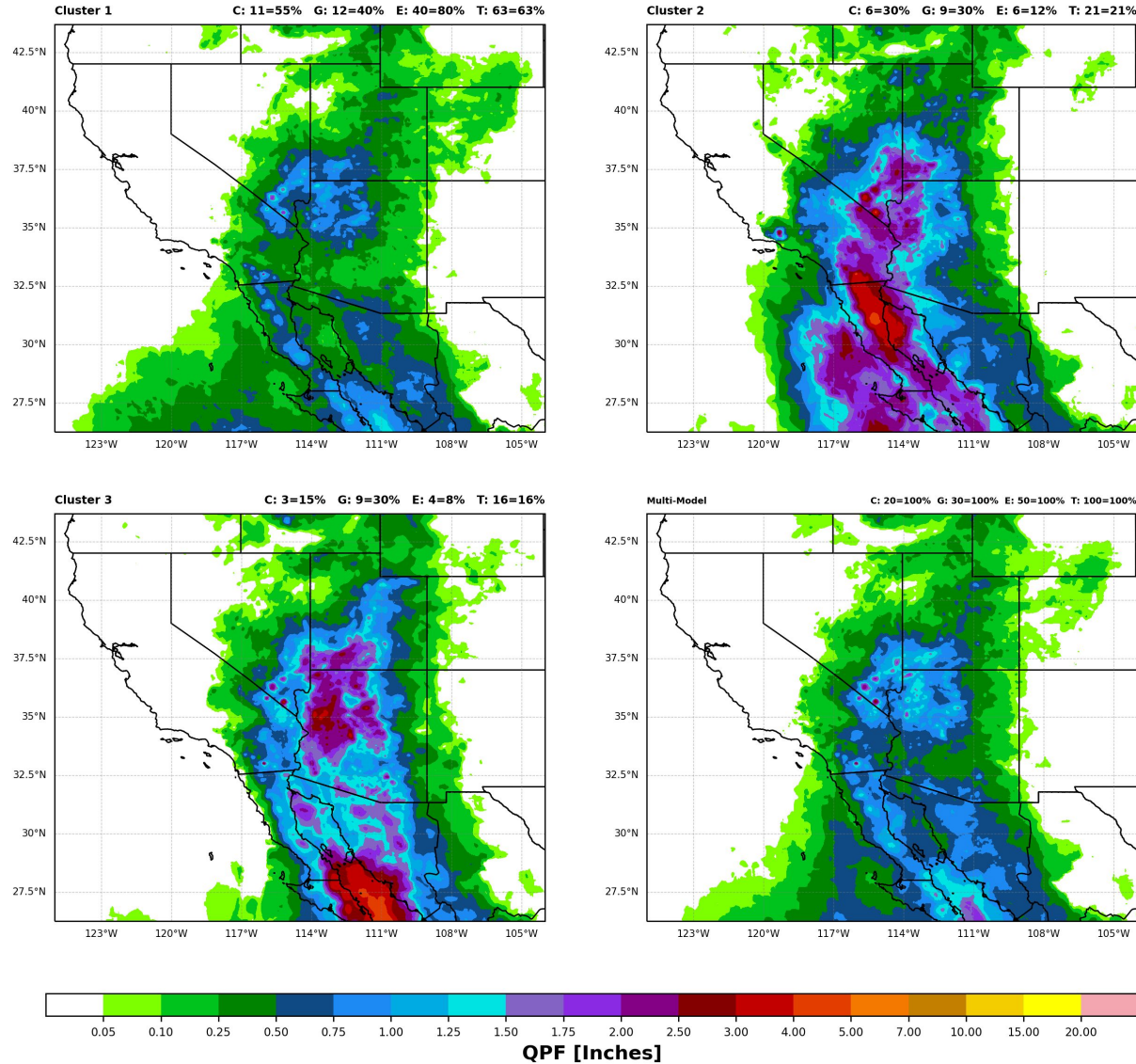
Secondary mode of 48hr QPF uncertainty is position of precipitation max



We also have a page that clusters on QPF uncertainty!

48-hr QPF 50th Percentile by Cluster

48-hour QPF 50th Percentile [Inches]
Init: 0000 UTC Tue Aug 15 2023 --> Valid: 48-hours Ending 0000 UTC Tue Aug 22 2023



Forecasters regularly use this guidance!

- Clustering technique has quickly gained popularity within the NWS (largely due to Western & Central Region championing it!)
 - In 2019, ensemble clustering tool was mentioned **38x in NWS AFDs**
 - By 2021, it was referenced over **3,600x in AFDs**
- Clustering also used as a centerpiece of the experimental ensemble visualization platform known as the Dynamic Ensemble-based Scenarios for IDSS (DESI)
 - As part of DESI, the clustering technique is used as both a scientific forecasting tool and a communication tool



Width: 158 Height: 79

Clustering Area: Northeast

Clustering Variable: 500 mb Height

Clustering Period: Day 8: 00Z Aug 22 - 23

Dataset: LREF-NH

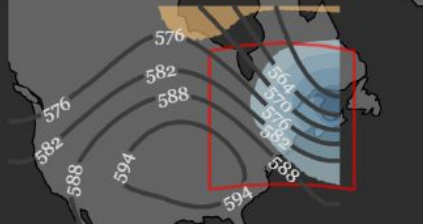
Forecast Runtime: 12Z August 14, 2023

Group By: Cluster

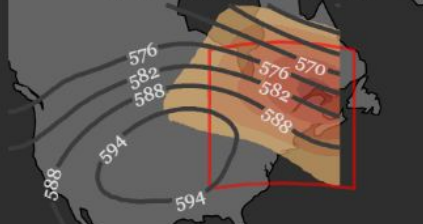
Select Member: Grand Ensemble

- Field Options
- Min
 - 5th
 - 10th
 - 25th
 - 50th
 - Mean

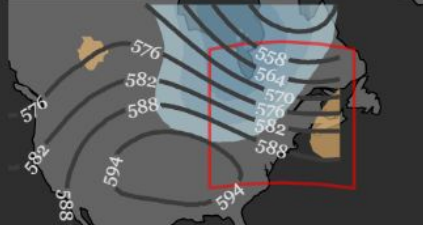
Cluster-1
GEFS: 53% EPS: 40% CMC: 35% T: 43%



Cluster-2
GEFS: 20% EPS: 40% CMC: 35% T: 33%



Cluster-3
GEFS: 27% EPS: 20% CMC: 30% T: 24%



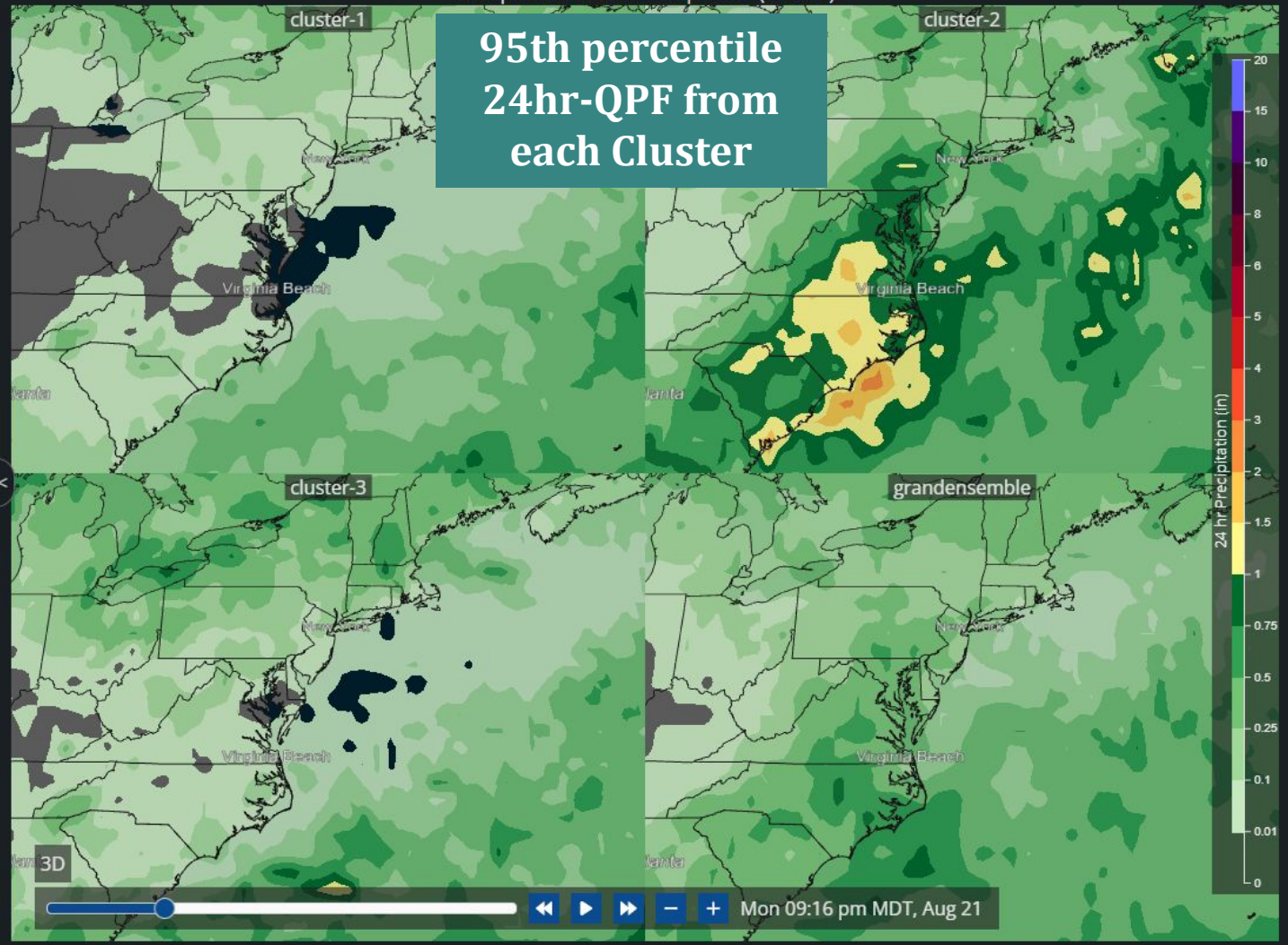
Grand Ensemble
GEFS: 100% EPS: 100% CMC: 100% T: 100%



Upper Levels Precipitation Convection Surface Maps = A - B

95th percentile 24 hr Precipitation (shaded)

95th percentile 24hr-QPF from each Cluster



Forecaster Feedback on Clustering in DESI as a Communication Tool

“Helps the forecaster determine why uncertainty exists which can be passed along to the core partner which helps build trust.”

“Aids in communicating with our partners and the public. Cluster information gives me a better idea of where there is lots of uncertainty, and in those cases I avoid speaking in terms of absolutes.”



But they want more context about how the different clusters came to be...

...Cue ensemble sensitivity analysis (ESA)!

- Clusters tell you the different prevalent forecast outcomes, but don't provide any context on what leads to those outcomes
 - Forecasters can look at the forecast evolution of each cluster and infer which early forecast differences lead to different outcomes, but takes a lot of time/energy and can be tricky
 - Ensemble sensitivity analysis relates the possible forecast outcomes back to the early forecast state – quantifies which early features are most relevant to scenario
 - **Clustering describes the “what”, and ensemble sensitivity explains the “why”**
- Forecasters often ask for the “why” in DESI feedback surveys
- Actively developing this product as a web-based tool (should be available on the main WPC cluster page in the next few weeks)

Also in collaboration with Brian Colle at Stony Brook

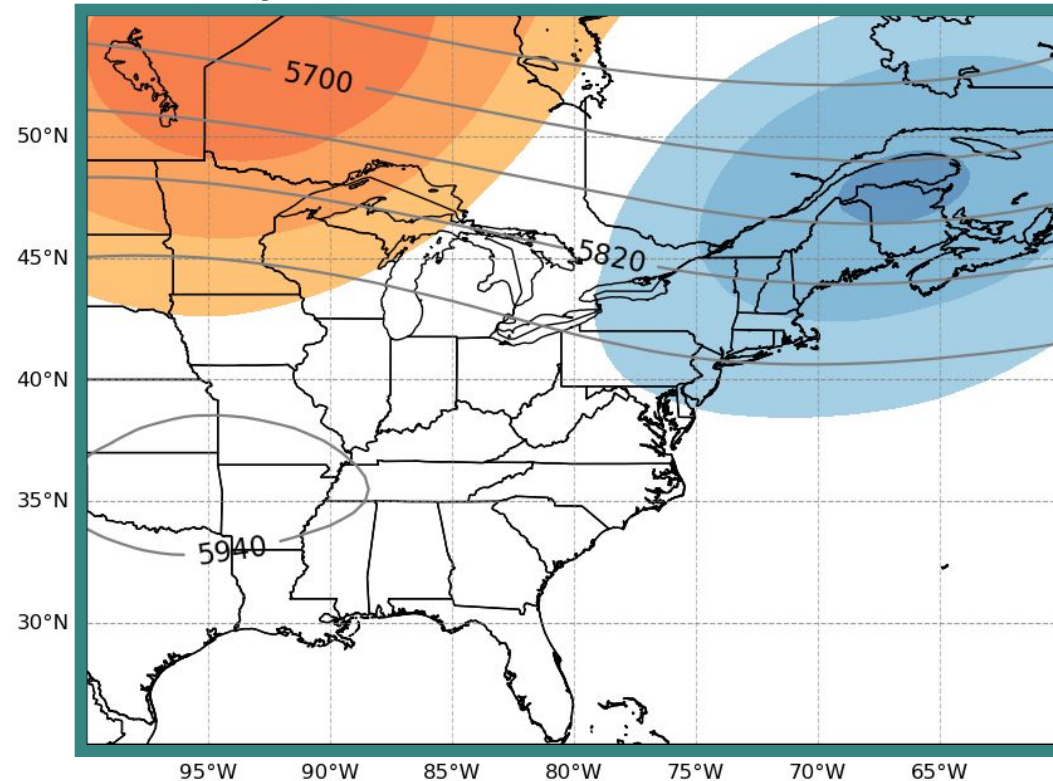
ESA tells us how the atmosphere needs to evolve early on in order to look like a given EOF!

Response Function: PC Values (in this case, positioning of the pattern)

EOFs of 24-hour Mean 500-hPa Heights [meters]

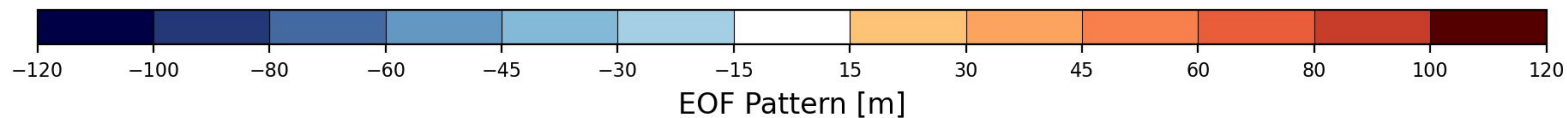
Init: 00Z Fri Aug 11 2023 --> Valid: 24-hours Ending 00Z Mon Aug 21 2023

EOF2 of Day 9 500 hPa HGT Percent Variance = 28.2%



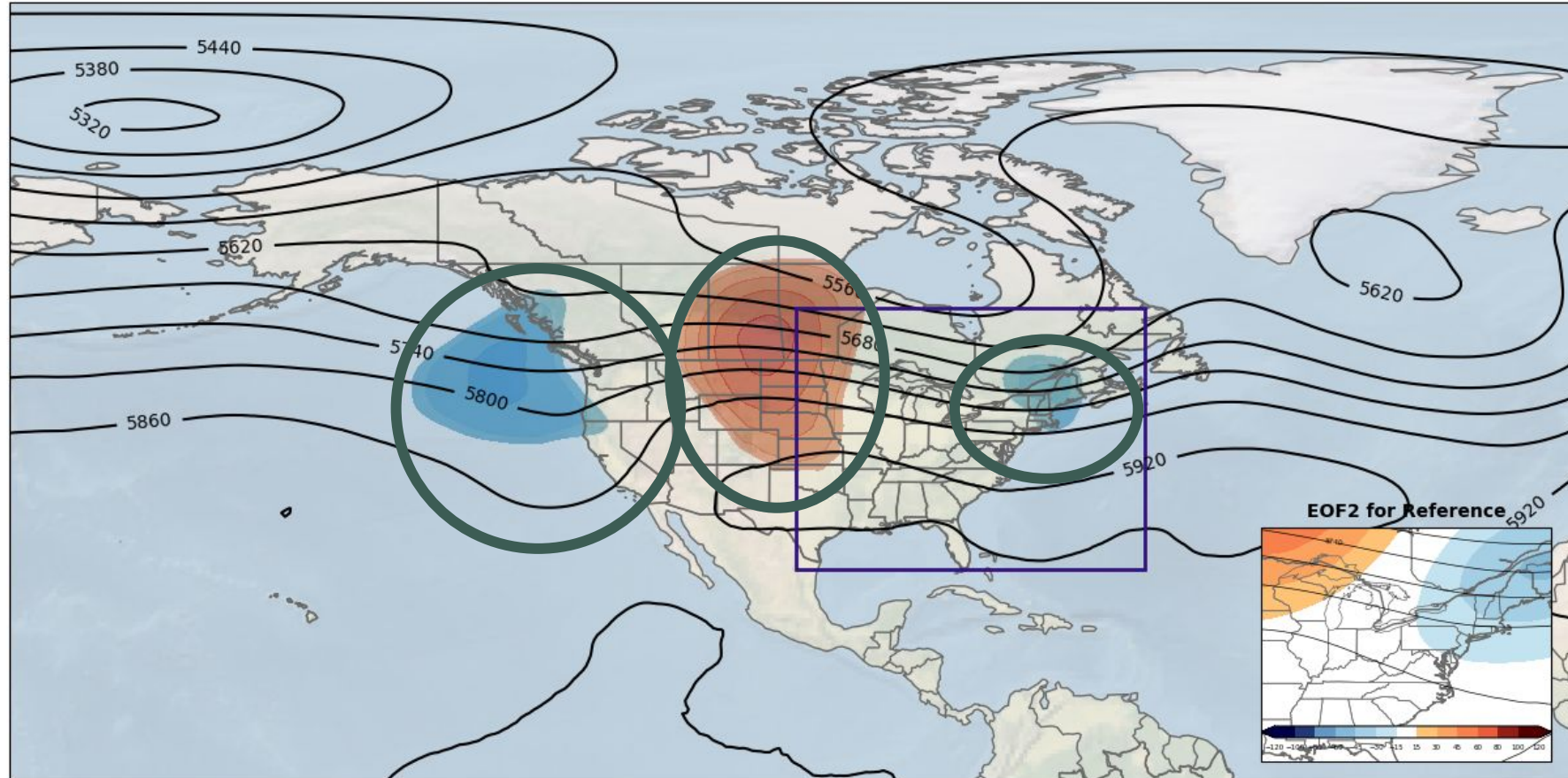
Uncertainty:
Position of Pattern in
Upper Midwest/NE

**Positive PC2 means ridge
pattern shifted West and
enhanced troughing in the NE**

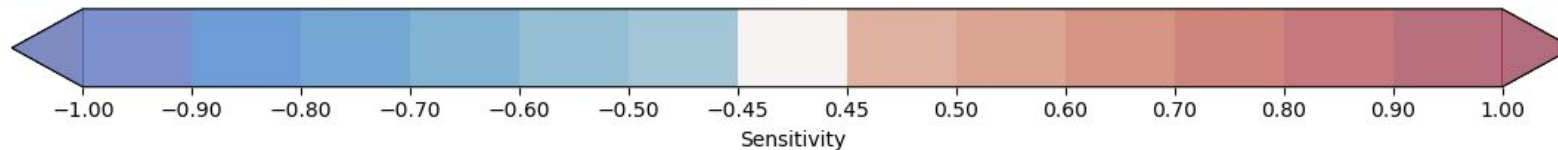


ESA tells us how the atmosphere needs to evolve early on in order to look like a given EOF!

Sensitivity of 24-hr Averaged 500-hPa GPH PC2 to 500-hPa GPH
Response Function (500-hPa GPH PC2) Valid: 2023-08-21 00:00:00 (Day 9 | f222-f240)
Sensitivity Variable (500-hPa GPH) Valid: 2023-08-19 12:00:00 (f204)



Positive PC2 means ridge pattern shifted West and enhanced troughing in the NE



Take-Home Points

Ensemble clustering is a quick way to distill an ensemble forecast down to its prevalent scenarios

- Lots of potential as a scientific tool (best cluster verifies better than best deterministic or best ensemble mean forecast)
- Feedback suggests utility as a communication tool as well

Ensemble sensitivity analysis (ESA) provides context on how the atmosphere must evolve to lead to different cluster scenarios

Testament to the potential of data mining ensemble systems

- As we continue to build techniques that extract information from these datasets, need to keep forecaster needs at the forefront (lots of room for O2R/R20 in these spaces)!

Useful Links!

WPC Days 3-7 500Z Clusters:

https://www.wpc.ncep.noaa.gov/wpc_ensemble_clusters/day_3_7/view.php

WPC Days 1-6 QPF Clusters:

https://www.wpc.ncep.noaa.gov/wpc_ensemble_clusters/qpf_clusters/view.php

WPC Days 8-10 500Z Clusters:

https://www.wpc.ncep.noaa.gov/wpc_ensemble_clusters/day_8_10/view.php