Making the Most out of Operational Ensembles with Clustering and Sensitivity Analysis

Austin Coleman[1], Jim Nelson[2], Brian Colle[3], Travis Wilson[4], and Samantha Lankowicz[3]

VLab Forum | Feb 28th 3 – 4pm ET

[1]CIRES/WPC, [2]NOAA WPC, [3]Stony Brook University, [4]NOAA GSL

Motivation

As technology improves and NWS responsibilities expand

- Forecasters have access to **more data** with simultaneously **less time** to interrogate those data

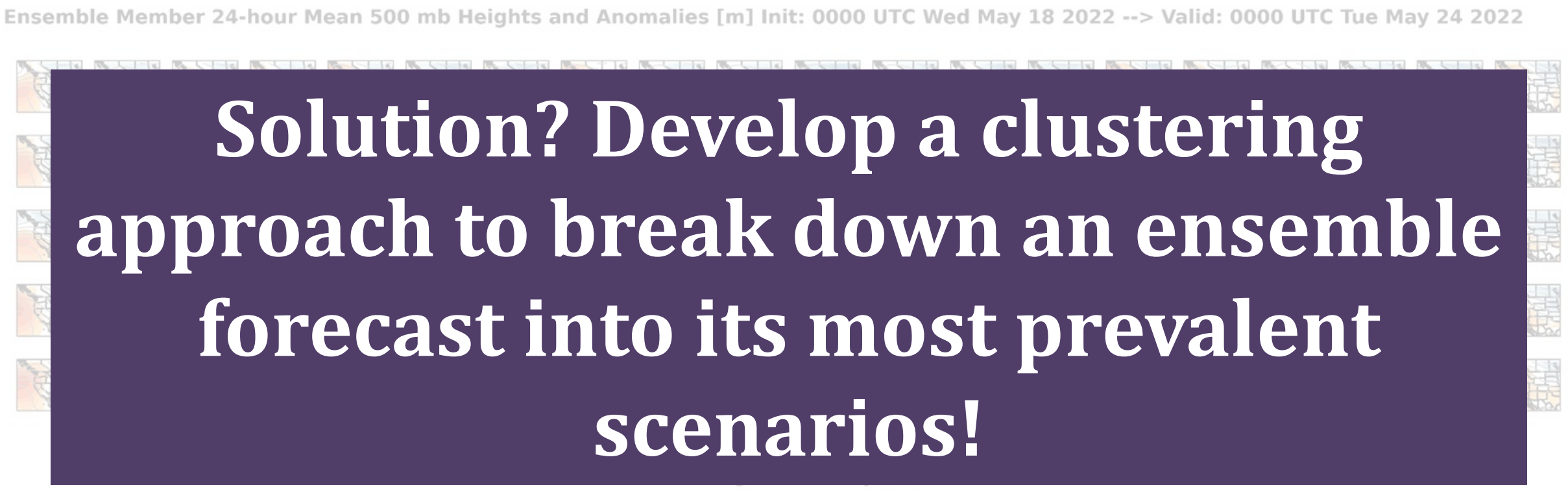
The National Blend of Models (NBM) is frequently used as a first-guess for said forecasts

- 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!

This isn't a viable answer either!

Motivation

WPC runs an ensemble clustering tool that breaks down an ensemble forecast (composed of 100 CMCE/GEFS/ECMWF members) to its prevalent scenarios

Cluster scenarios prove more skillful than the ensemble mean (Lamberson et al. 2023)

NWS fo SI – Dynam

In 2021, clusters were referenced over 3,000x in Area Forecast Discussions

Ensemble clusters tell you the **"what"**
Ensemble sensitivity tells you the **"why"**

Forecasters often request additional context about what leads to the different cluster outcomes → cue **ensemble sensitivity analysis**!

PART 1

Ensemble Clustering | “The What”

But first, how does clustering work?

Relies heavily on Empirical Orthogonal Function (EOF) Analysis, traditionally known in statistics as Principal Component Analysis (PCA)

Empirical Orthogonal Function (EOF) Analysis decomposes a series of spatial fields into mathematically-independent (orthogonal) modes

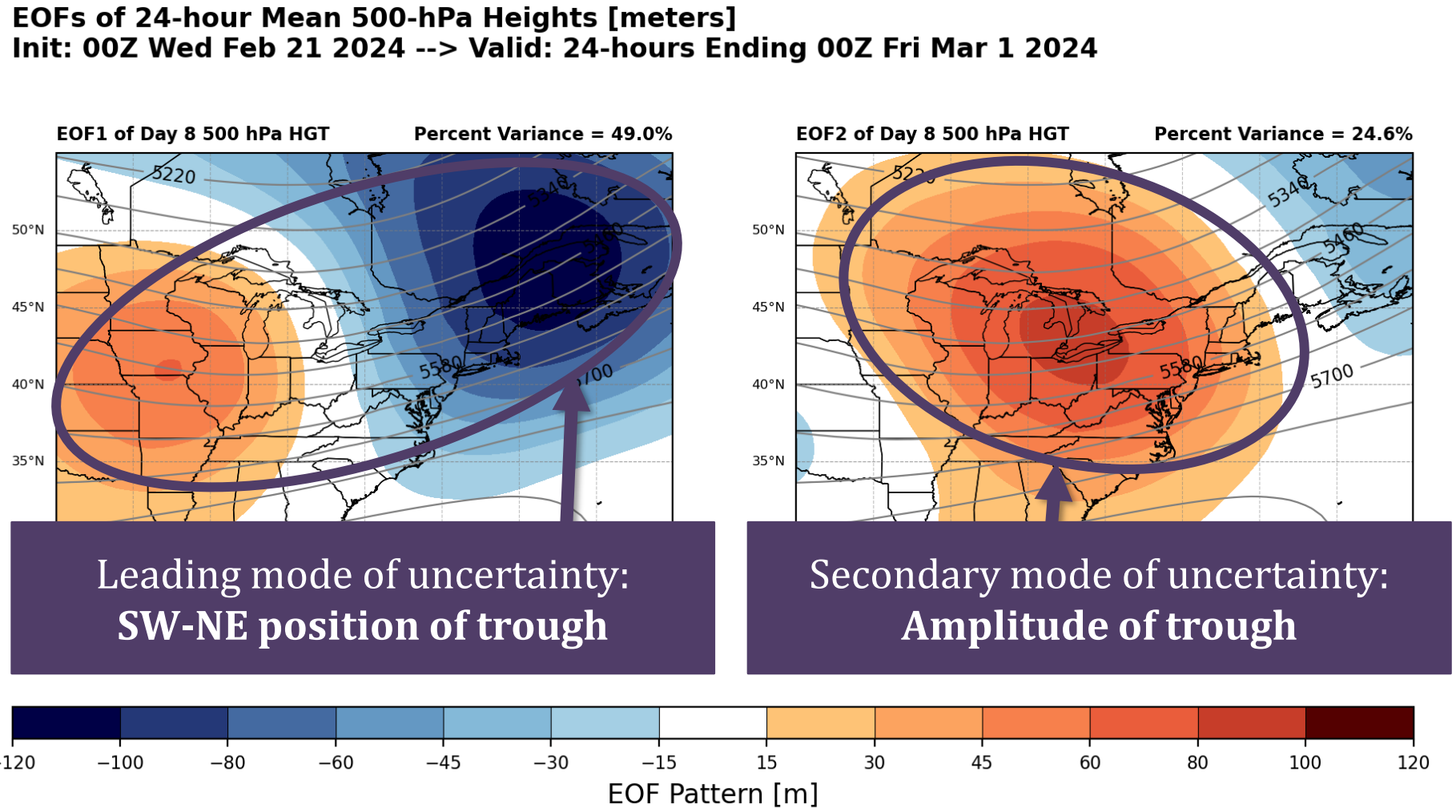
Method most often used by climatologists to understand leading spatial modes of variability in a time series

> For example, it could be used to answer the question: *What wind patterns are typically associated with various phases of the North Atlantic Oscillation?*

For ensemble clustering, we seek to identify the **leading spatial modes of variability amongst the ensemble membership**

("THE CLIMATE DATA GUIDE: EMPIRICAL ORTHOGONAL FUNCTION (EOF) ANALYSIS AND ROTATED EOF ANALYSIS", 2013)

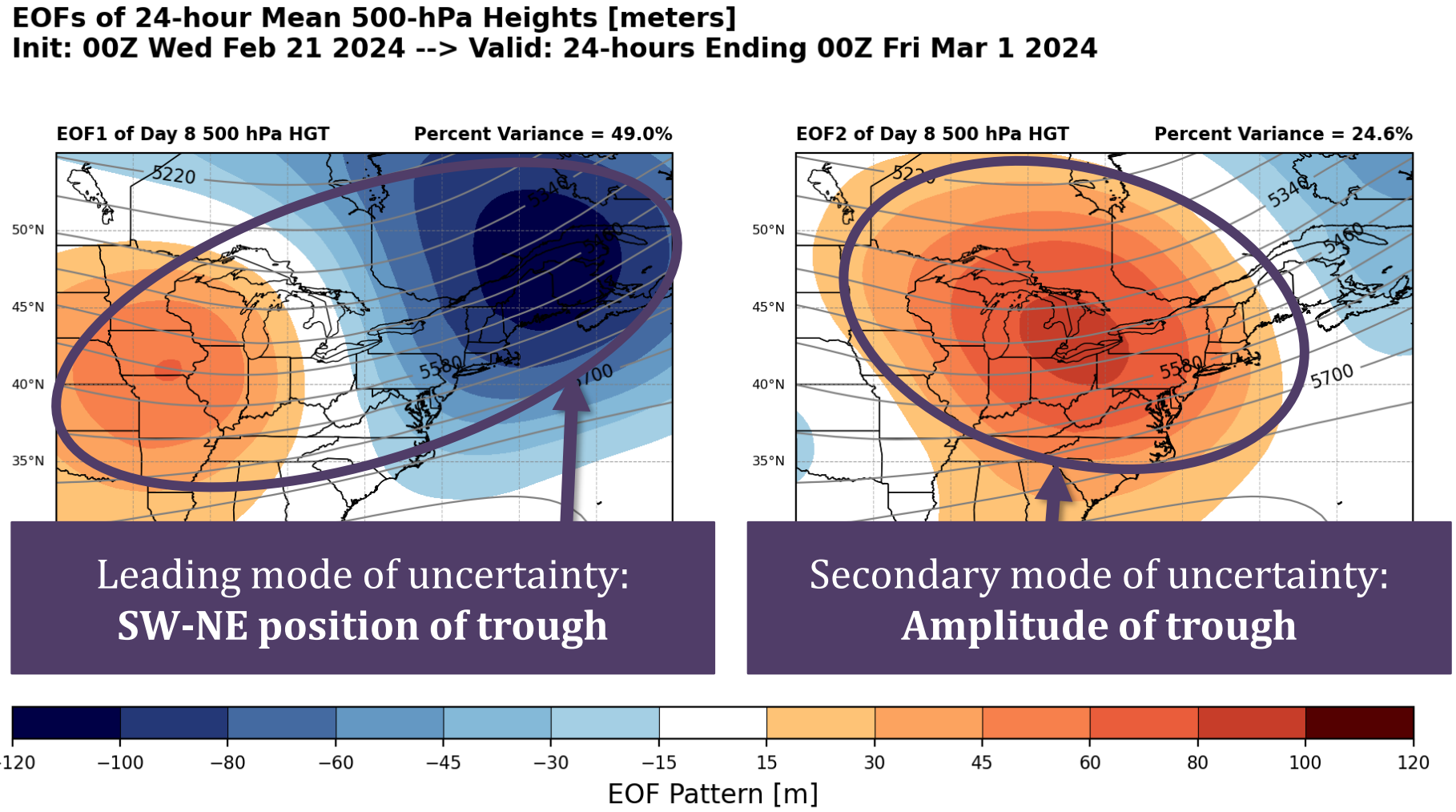
We break down the forecast (super-ensemble of CMCE, GEFS, & ECMWF) into its leading modes of variability via EOF Analysis



Common source of confusion:

What do the positives and negatives mean?

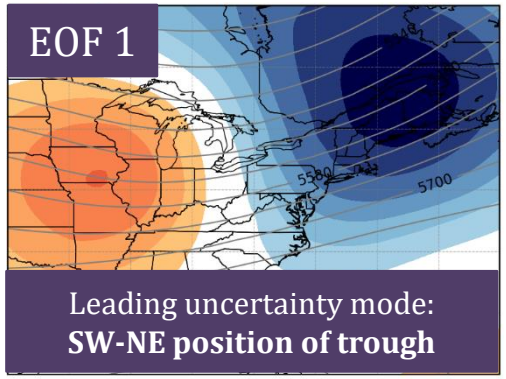
- Sign doesn't matter when looking at the EOFs themselves
- Only becomes important once we start clustering or looking at members in PC phase space



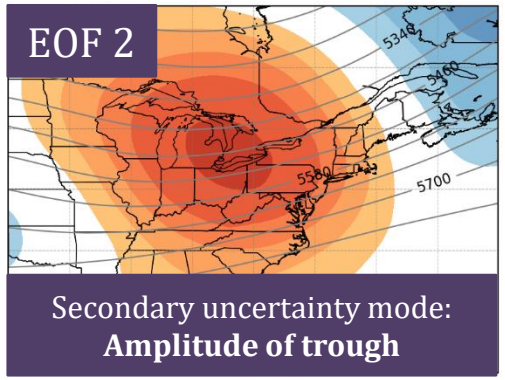
How does it work?

Next, we group similar ensemble solutions together with clustering

First two EOFs for reference



Leading uncertainty mode:
SW-NE position of trough



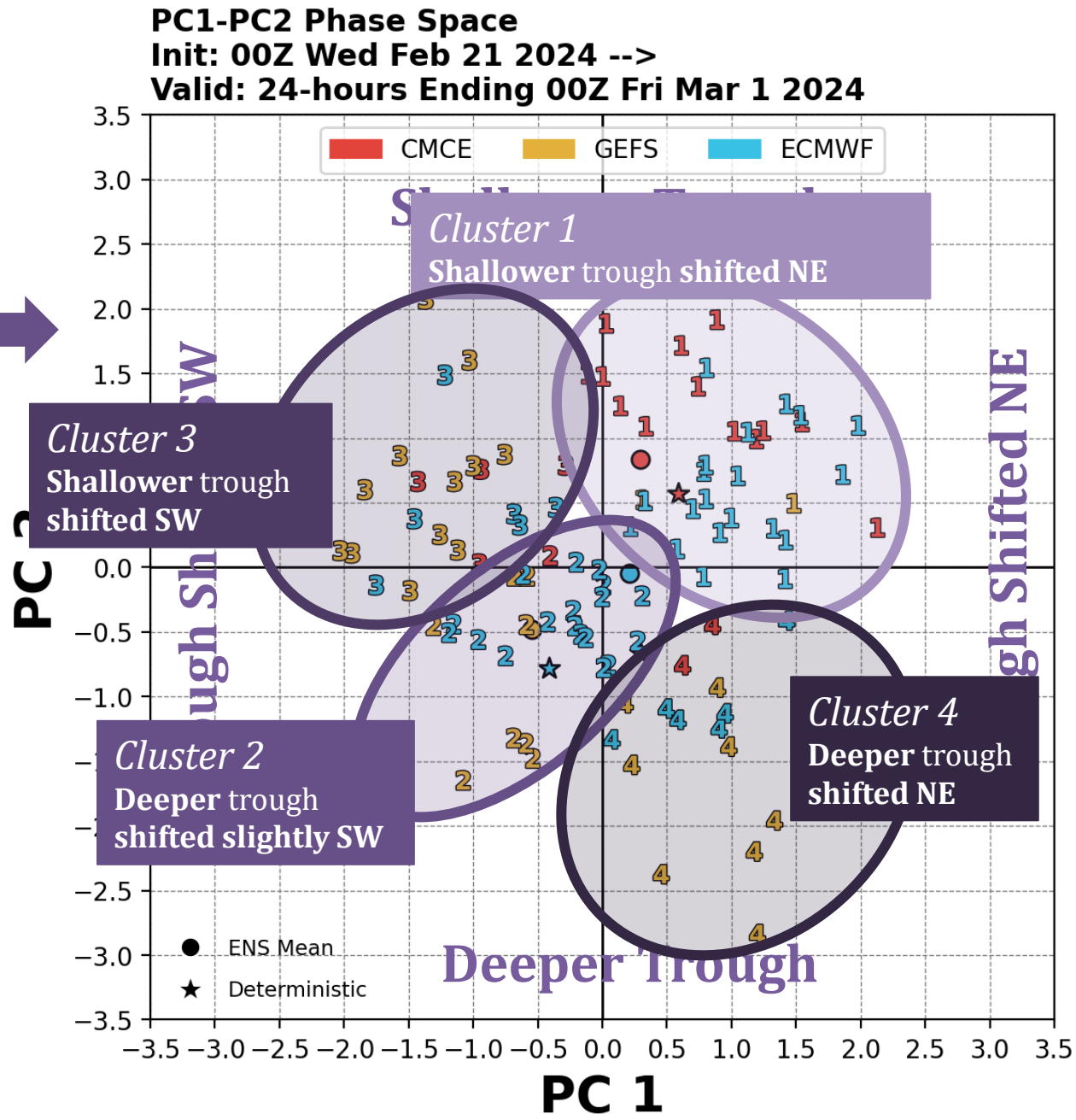
Secondary uncertainty mode:
Amplitude of trough

Projecting ensemble members into PC phase space shows us the forecast scenario for each member

Members with positive PC1 will look more like EOF1 *(trough shift NE)*

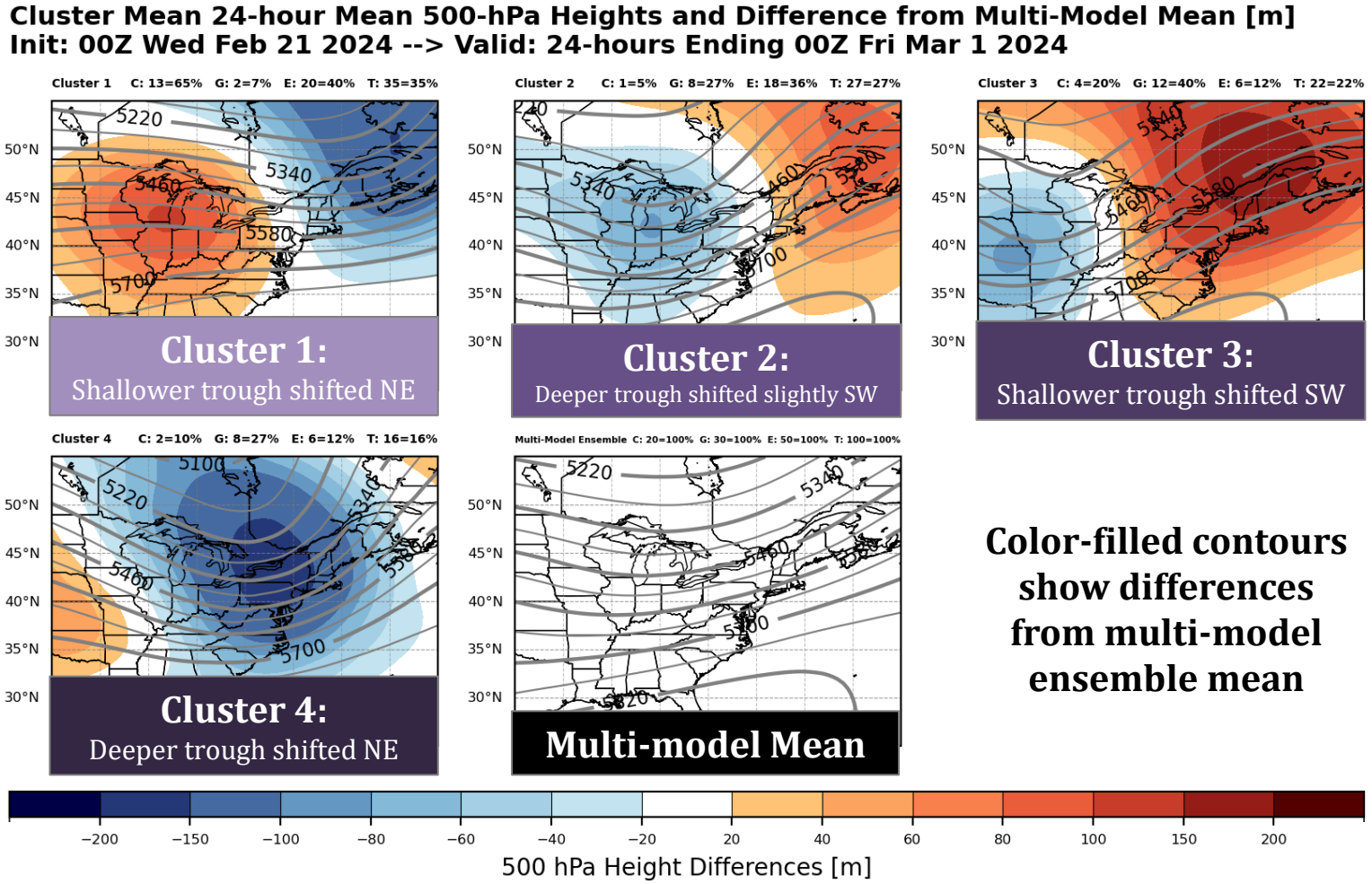
whereas members with negative PC1 will look opposite EOF1 *(trough shift SW)*

K-means Clustering groups members with similar solutions based on our leading modes of uncertainty



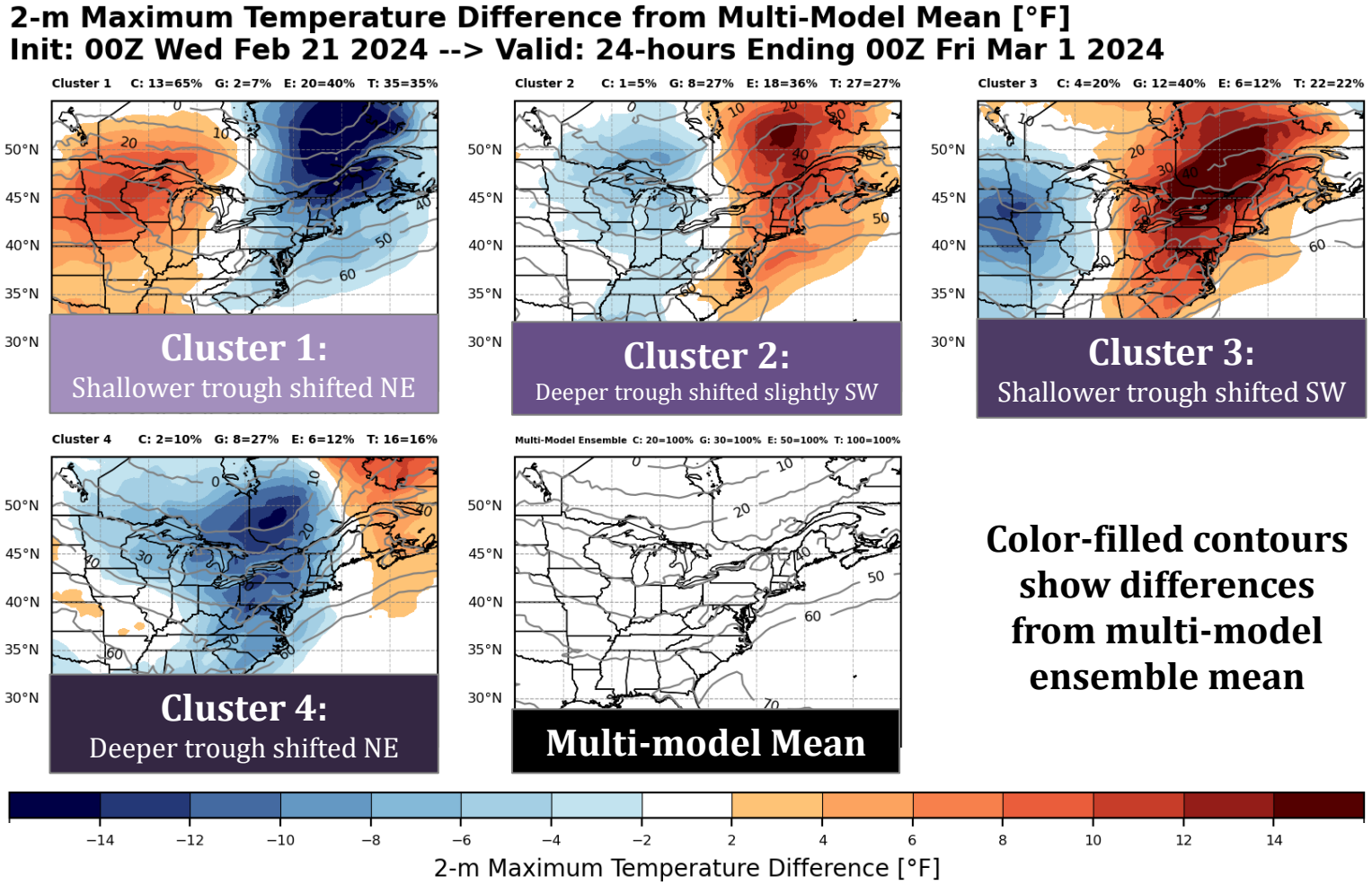
Don’t even need to look at EOFs or PCs to use!

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



Can use 500-hPa height clusters to predict other fields

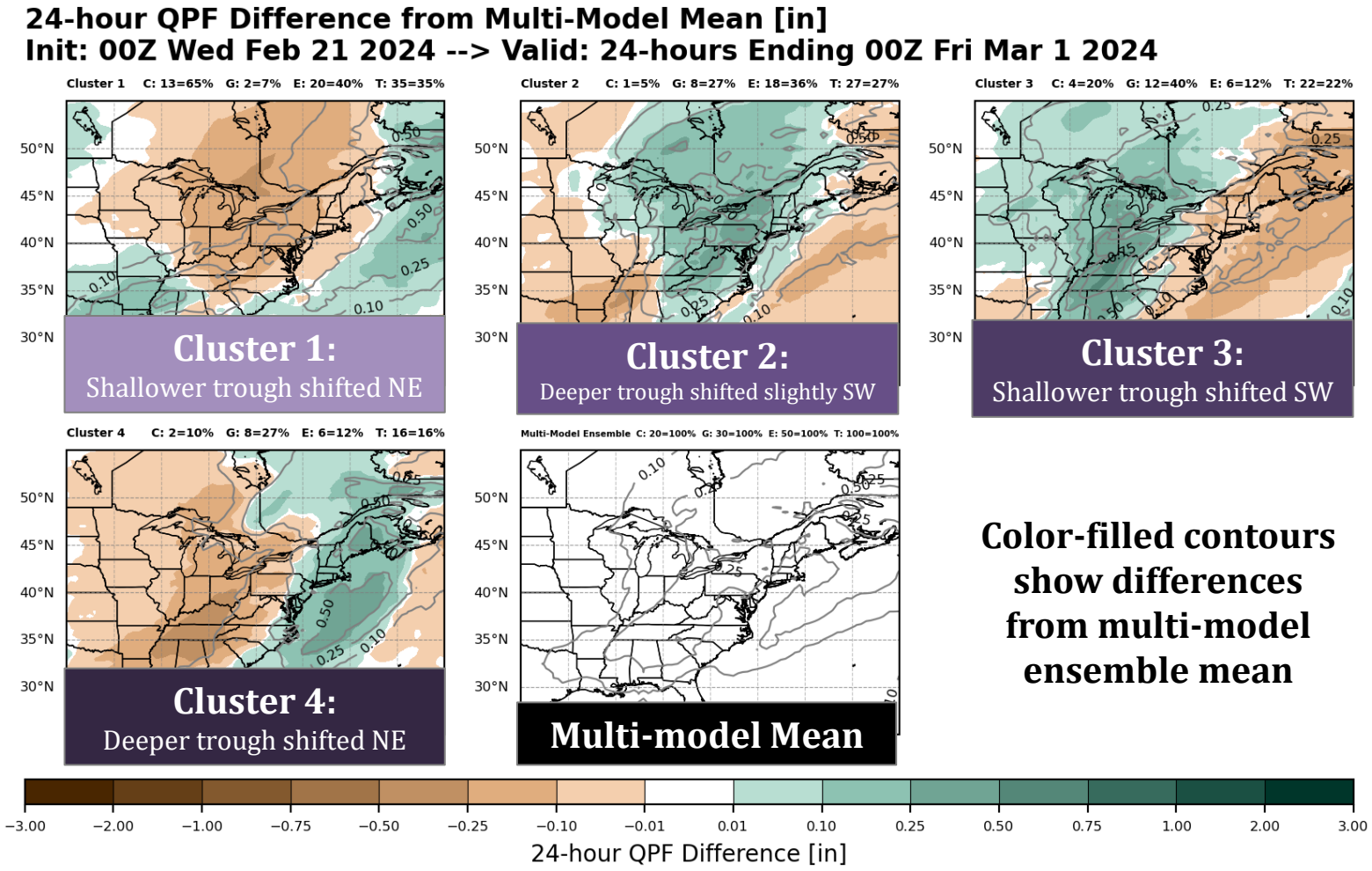
Maximum Temperatures



Color-filled contours show differences from multi-model ensemble mean

Can use 500-hPa height clusters to predict other fields

24-hr QPF

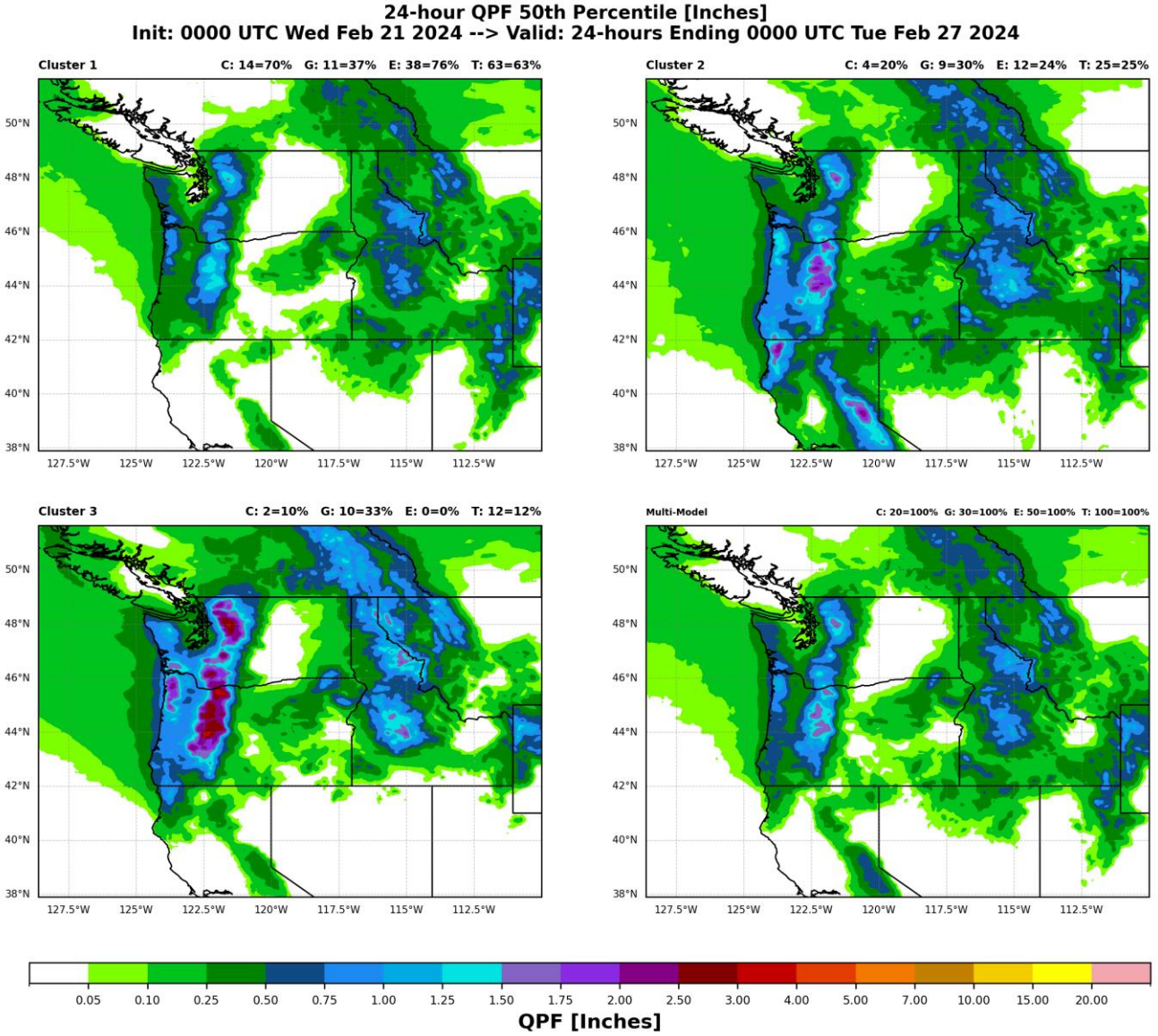


Color-filled contours show differences from multi-model ensemble mean

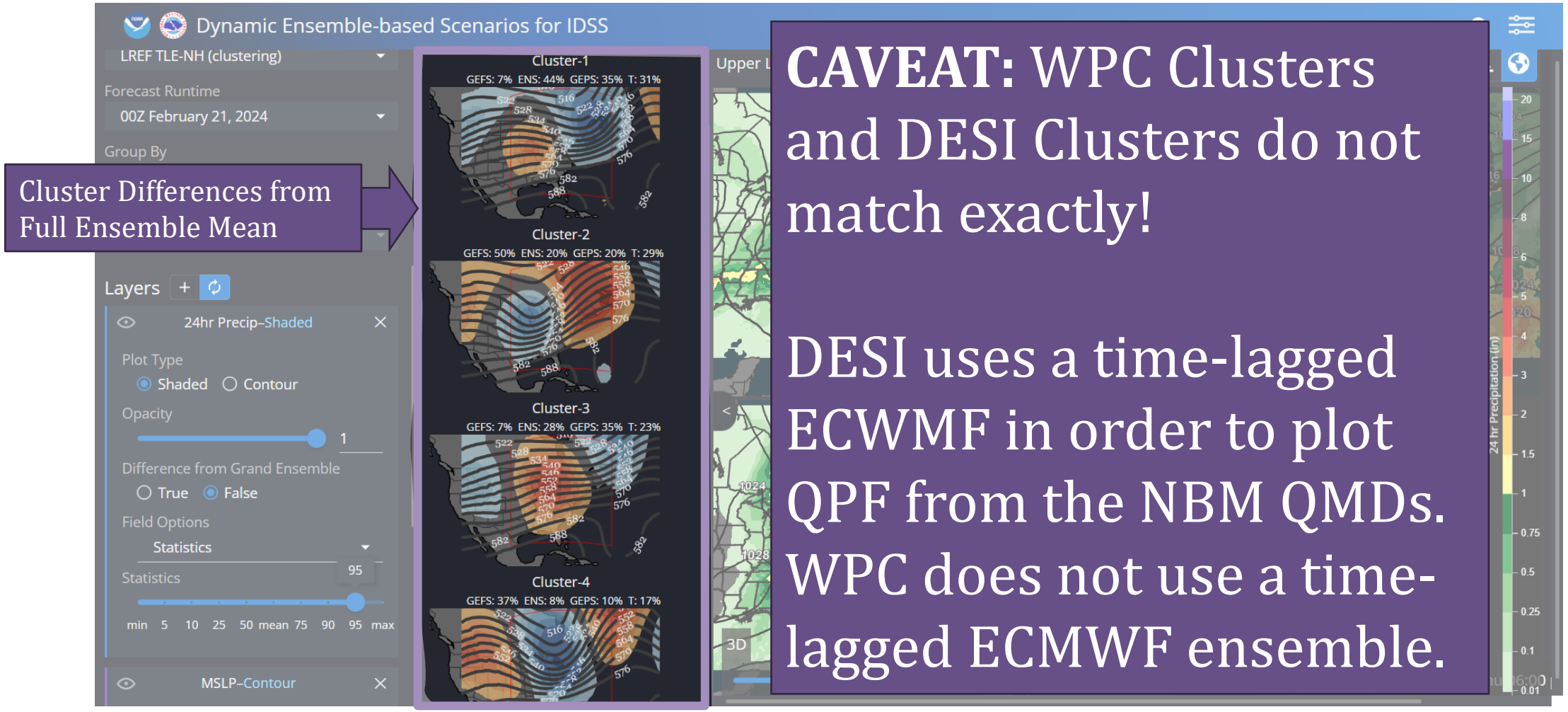
We additionally have a WPC page that clusters directly on QPF!

Algorithm

- ✓ Identifies QPF objects (0.50” Day 5 QPF object shown here)
- ✓ Picks a varying number of clusters based on the silhouette score (3 clusters picked here)
- ✓ **Provides more skillful QPF scenarios than using QPF derived from 500-hPa height clusters!** (Colle, personal communication)



You can interrogate the clusters even more effectively with the Dynamic Ensemble-based Scenarios for IDSS (DESI)!

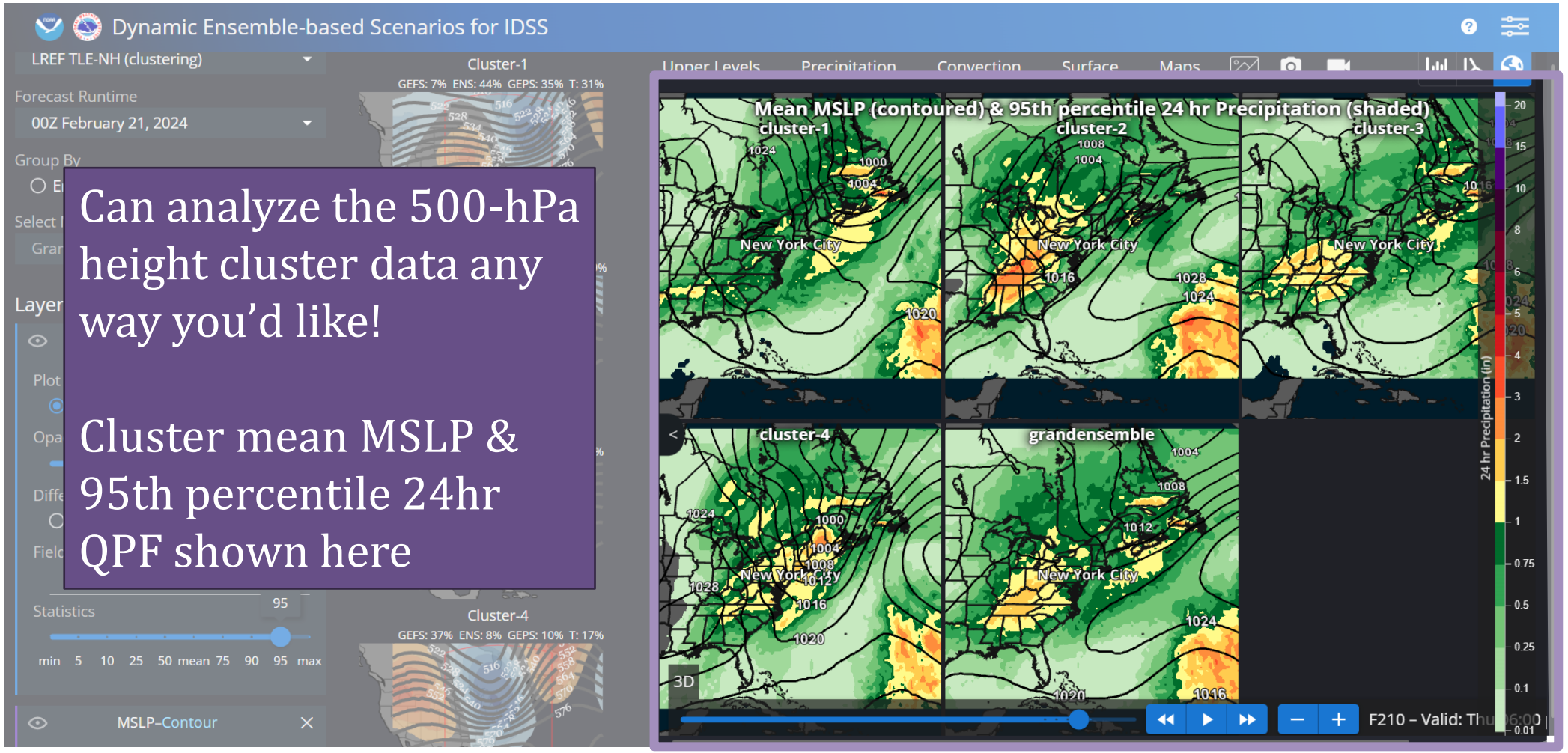


Cluster Differences from Full Ensemble Mean

CAVEAT: WPC Clusters and DESI Clusters do not match exactly!

DESI uses a time-lagged ECWMF in order to plot QPF from the NBM QMDs. WPC does not use a time-lagged ECMWF ensemble.

You can interrogate the clusters even more effectively with the Dynamic Ensemble-based Scenarios for IDSS (DESI)!



Can analyze the 500-hPa height cluster data any way you'd like!

Cluster mean MSLP & 95th percentile 24hr QPF shown here

You can interrogate the clusters even more effectively with the Dynamic Ensemble-based Scenarios for IDSS (DESI)!

PART 2

Ensemble Sensitivity Analysis | “The Why”

What is Ensemble Sensitivity Analysis?

Reveals how meteorological conditions early in the forecast (**sensitivity variable**) are linked to the evolution of a chosen high-impact forecast feature (**response function**)

(Hakim and Torn 2008, Ancell and Hakim 2007, Torn and Hakim 2008)

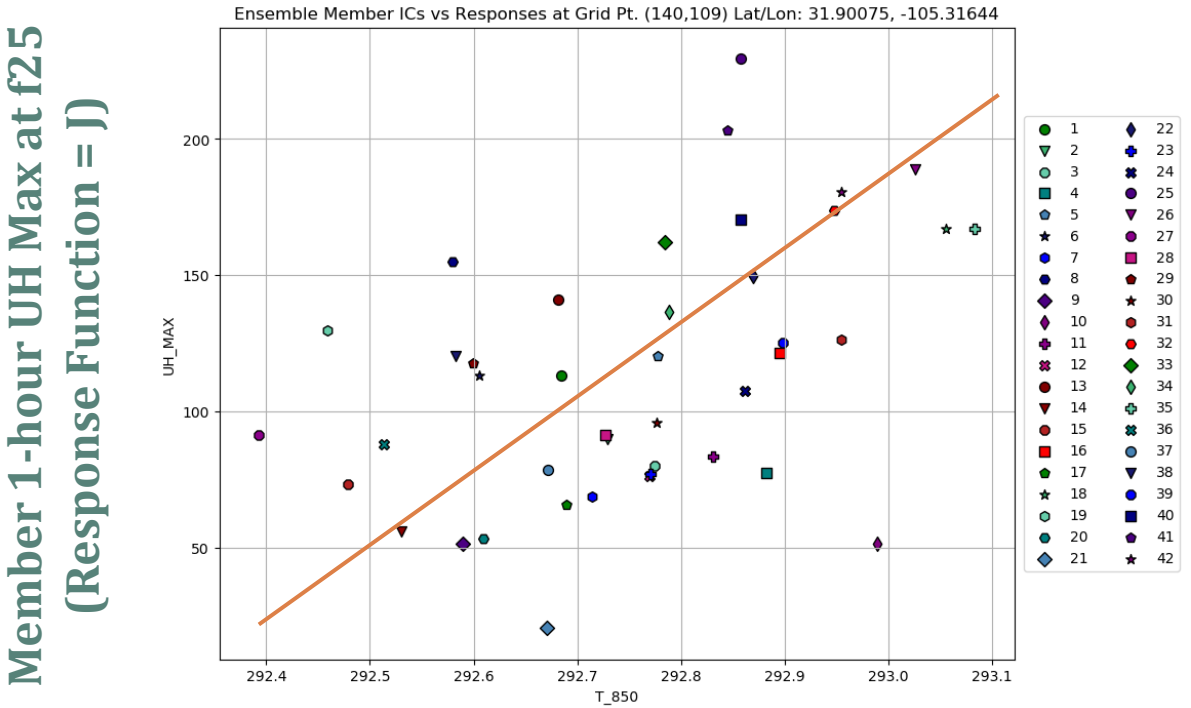
Simply the slope of a linear regression line:

$$\frac{\partial J}{\partial x_0}$$

$J \equiv response\ function$

$x_0 \equiv sensitivity\ variable$

Ensemble Sensitivity ≡ Slope of the Linear Regression



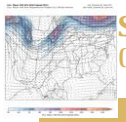
Member 1-hour UH Max at f25 (Response Function = J)

Member 850 hPa Temperature at f6 (Sensitivity Variable = x_0)

Ensemble Sensitivity Fields

Powerful tool: Sensitivity fields show us which early forecast features the ensemble "cares" most about in predicting high-impact weather!

More Examples of Sensitivity Variables and Response Functions:



Sensitivity Variables
(Predictors)

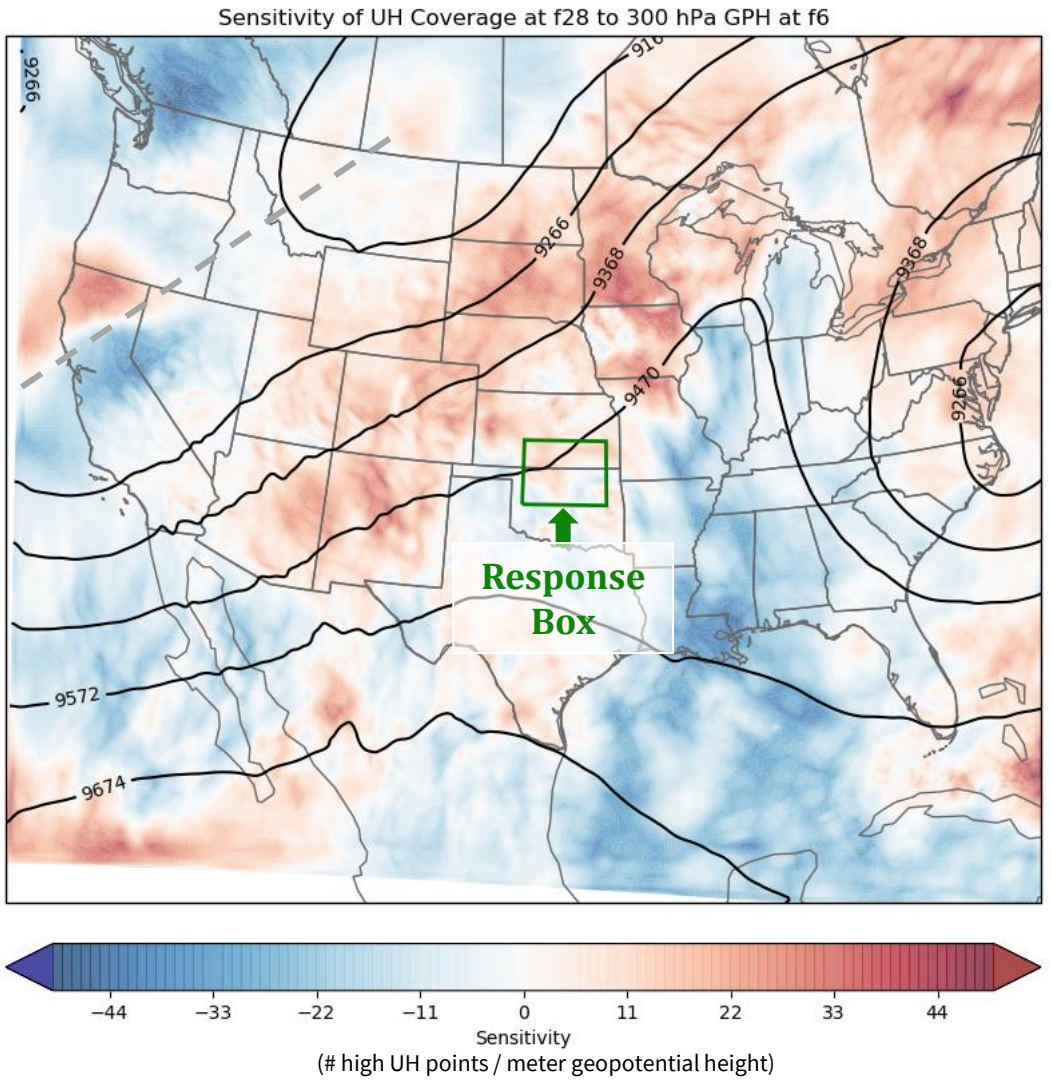
- 2m Temperature
- SLP
- 850 hPa Temperature
- 850 hPa Moisture
- 700 hPa Temperature
- 500 hPa GPH
- 300 hPa U, V

Response Functions
(High-Impact Forecast Features)

Magnitude Responses
- Max Simulated Reflectivity
- Max Precipitation Accumulation
- Max 10m Wind Speed

Coverage Responses
- High Reflectivity (>40 dBZ) Coverage
- High 24hr QPF (>2") Coverage
- High 10m Wind Speed (>40 mph) Coverage

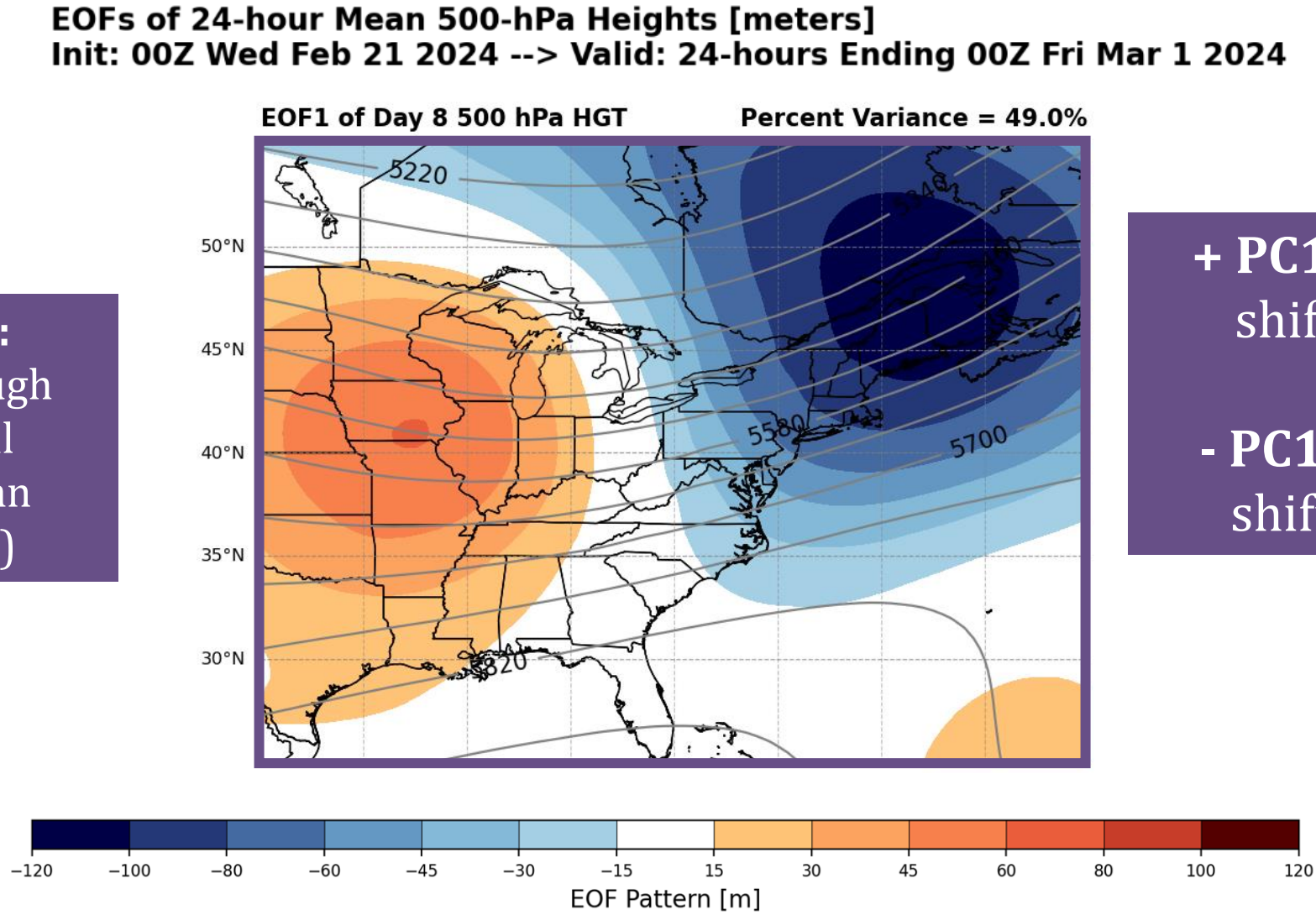
Sensitivity time (early) → Response time (later)



In our use case, ESA tells us how the atmosphere needs to evolve early on in order to look like a given EOF!

Let's regress the phase speed uncertainty of the pattern back onto the early 500-hPa height field

Uncertainty:
Position of trough relative to full ensemble mean (phase speed)

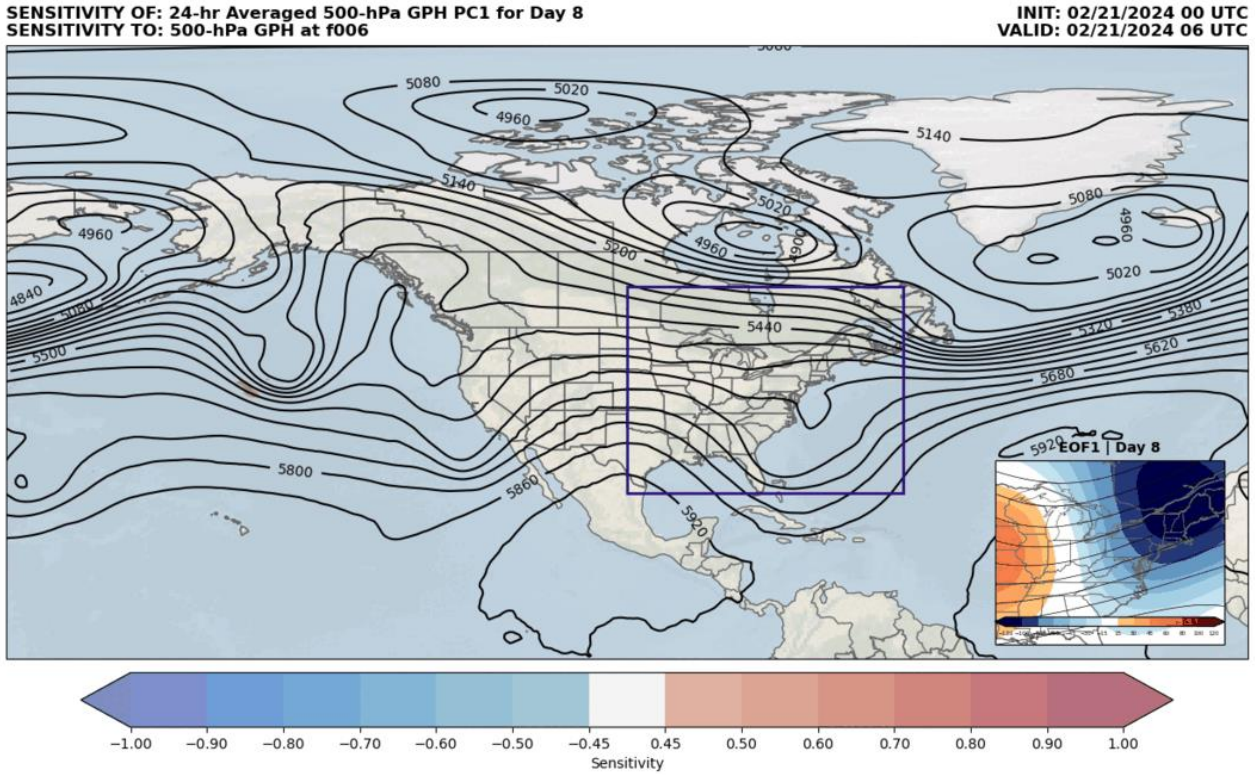


+ PC1 means trough shifted to the NE

- PC1 means trough shifted to the SW

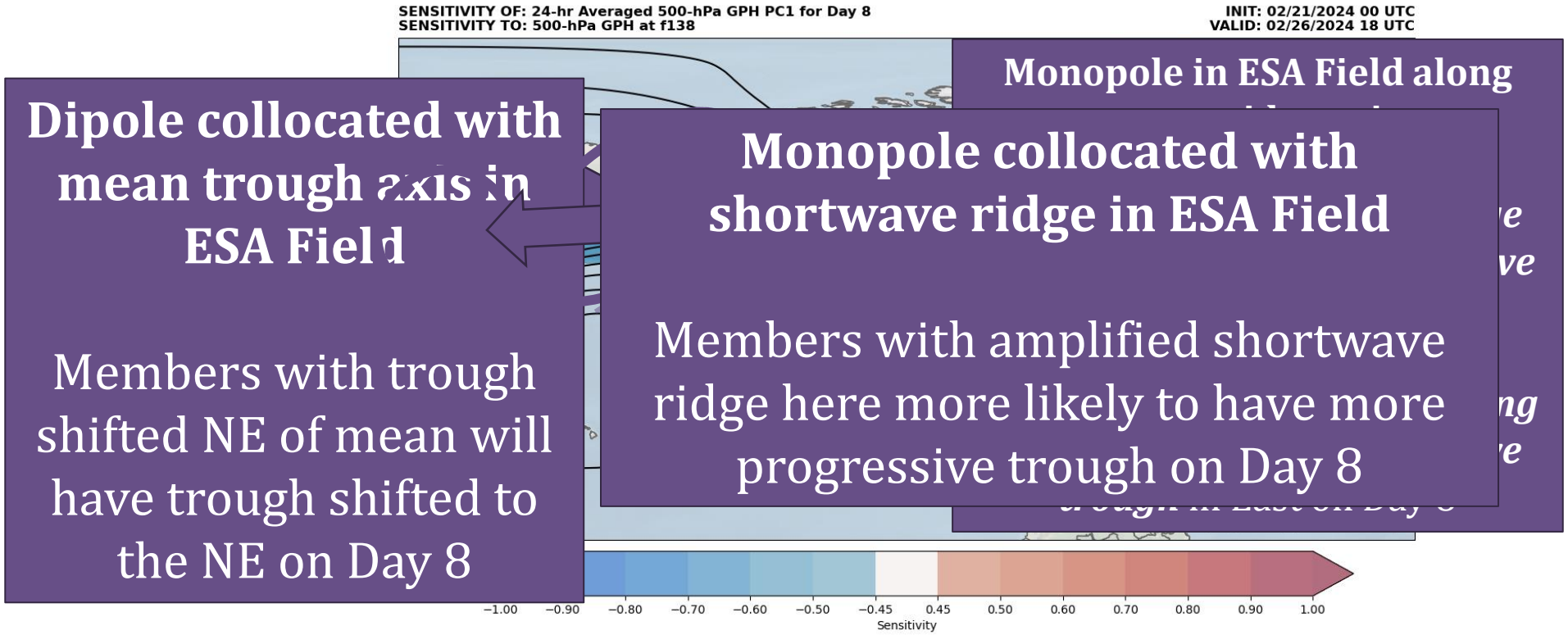
ESA shows us what the ensemble "cares about" most when predicting the position of the trough at Day 8

When we calculate the standardized sensitivity of PC values to the early forecast state, *the slope of the linear regression line becomes a* ***correlation***!



ESA shows us what the ensemble "cares about" most when predicting the position of the trough at Day 8

When we calculate the standardized sensitivity of PC values to the early forecast state, *the slope of the linear regression line becomes a* ***correlation***!

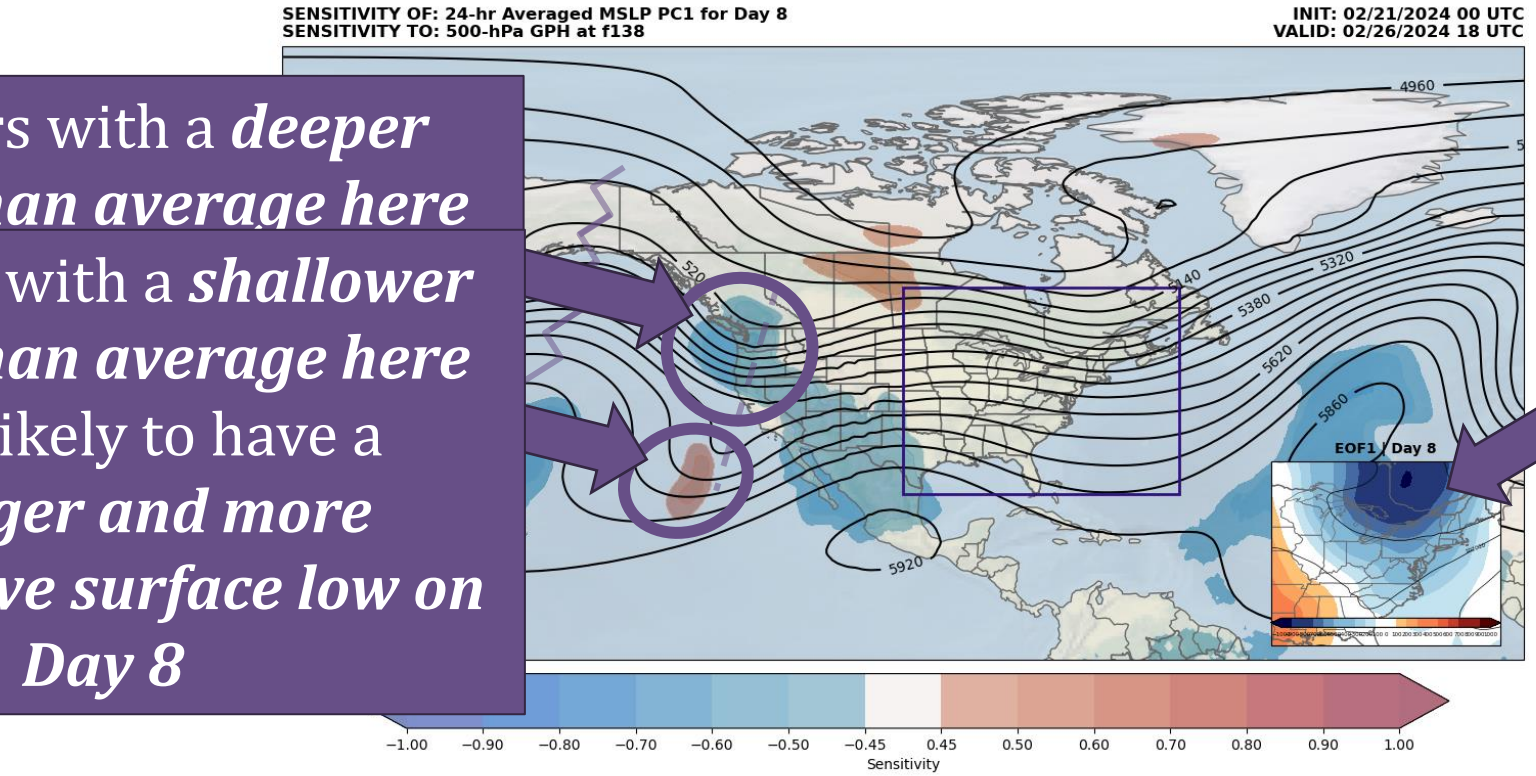


We can also conduct sensitivities of ***MSLP PCs*** back to the early forecast 500-hPa height field

This product could be particularly useful during the hurricane season for tropical cyclone sensitivity fields

Members with a ***deeper trough than average here***

Members with a ***shallower trough than average here*** more likely to have a ***stronger and more progressive surface low on Day 8***



Main Uncertainty: Depth & Position of Surface Low

Ongoing and Future Work | *Clustering*

- Recent Updates:

  Hawaii QPF Cluster Page (based on NBM 4.2 QMDs)

  Added Pacific 500Z Cluster Domain to WPC page for use in AR Forecasting and by Alaska region

  Worked with the wonderful team at COMET to publish a cluster training module! ☺

  Added ability for DESI clusters to handle missing members (requires >80% membership)

- Developing a medium-range WPC MSLP Cluster page for tropical cyclone and winter weather forecast applications *(expected mid-March)*

- Adopting a cluster consistency approach in DESI similar to that used by the Japan Meteorological Agency *(expected with fall DESI release)*

- Creating a verification dashboard with bulk long-term cluster statistics *(expected late 2024)*

Ongoing and Future Work | *Ensemble Sensitivity*

- Will replace WPC 500Z Cluster page with combined 500Z Cluster & ESA page once ECMWF section of ESA page is password-protected *(expected soon)*

  For now, it's linked here!

- Start soliciting feedback from forecasters and testing in the HMT testbeds *(over the next few years)*

- Work on a sensitivity-based ensemble subsetting application that objectively identifies most likely scenario to verify as event unfolds *(long-term goal)*

- Explore ESA & Clustering applications with RRFS-based convection-allowing ensemble systems *(long-term goal)*

Take-Home Points

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

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

- Allows you to hedge your bets on a particular scenario as the event unfolds

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/R2O in these spaces

BONUS SLIDES

Introduction

Forecasters often request additional context about what leads to the different cluster outcomes → cue **ensemble sensitivity analysis**!

Ensemble sensitivity analysis (ESA) offers a quick, efficient way to diagnose sources of high-impact forecast uncertainty

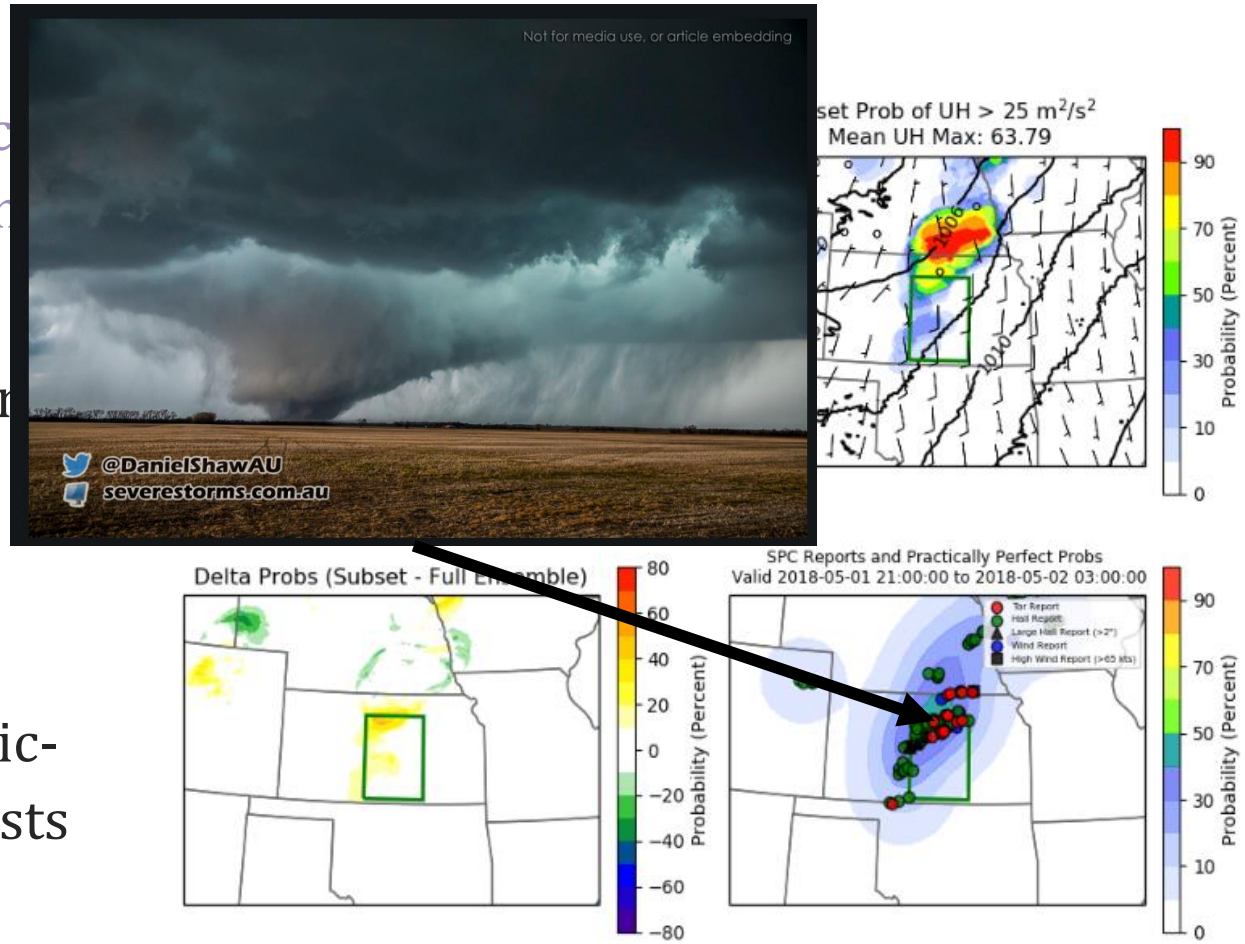
Sensitivity-based Ensemble Subsetting picks forecast solutions with best handle on early sensitive regions

Goal: Pick the ensemble members with the best forec of a chosen high-impact even

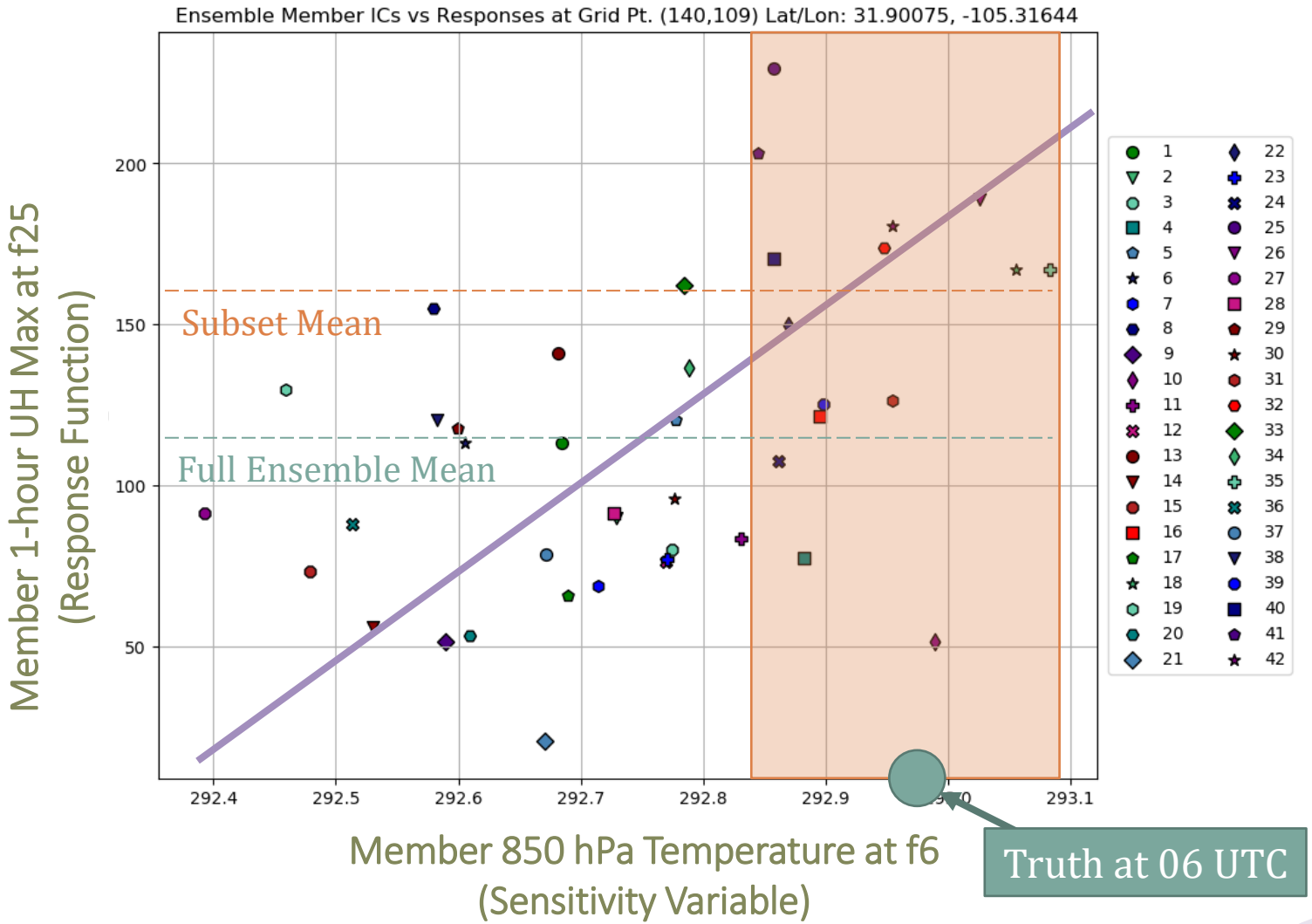
Combine ESA and analyses ("truth") at early forecast ti to pick members with best representation of sensitive regions!

Proven beneficial to synoptic- and convective-scale forecasts in an idealized framework

(Ancell 2016; Coleman and Ancell 2020)

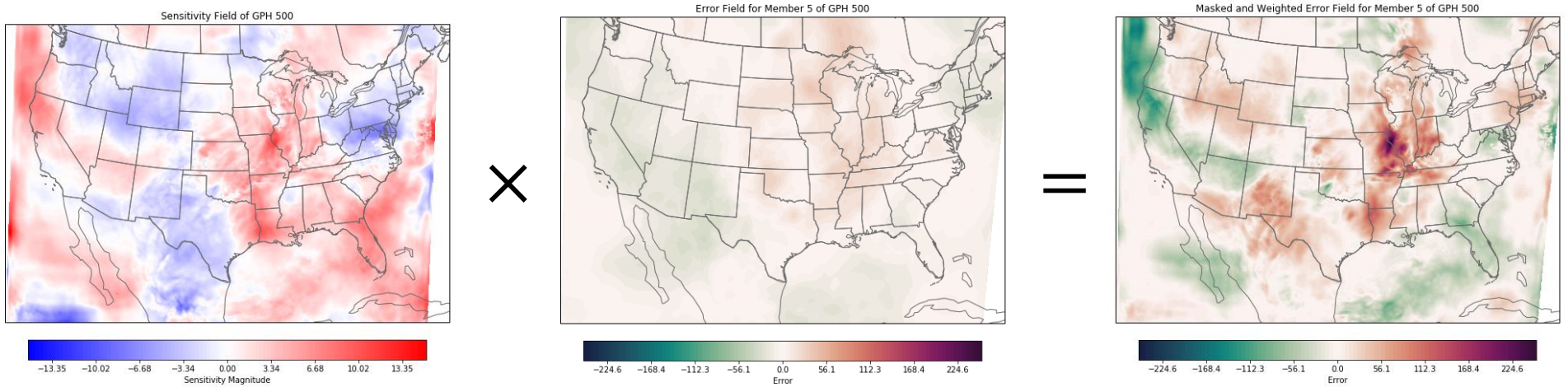


Subsetting Example with a Single Point

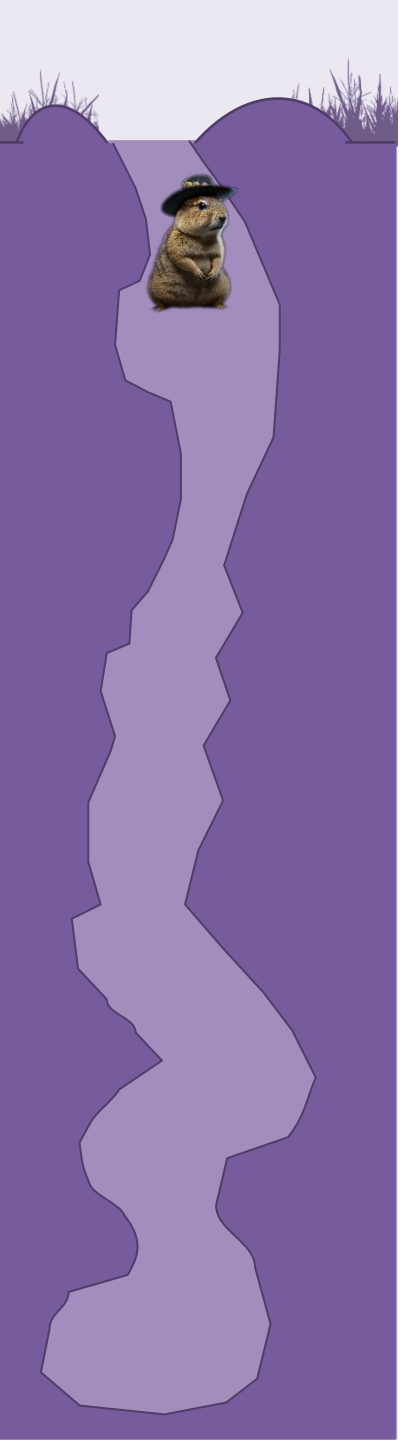


In practice, we use the Projection Technique

Sensitivity field is projected onto error field for each ensemble member to get an error field weighted by ensemble sensitivity.



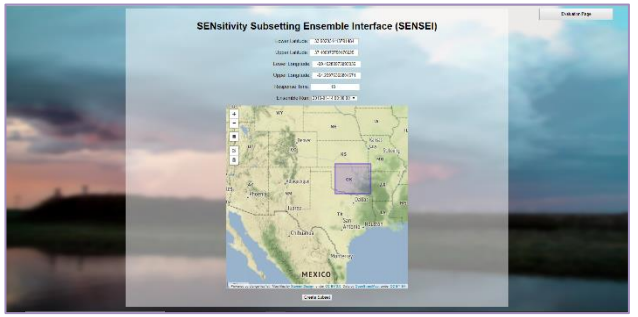
Repeat for each sensitivity variable of interest and sum over all sensitivity variables and grid points! Then repeat for each ensemble member and rank by summed sensitivity-weighted errors.

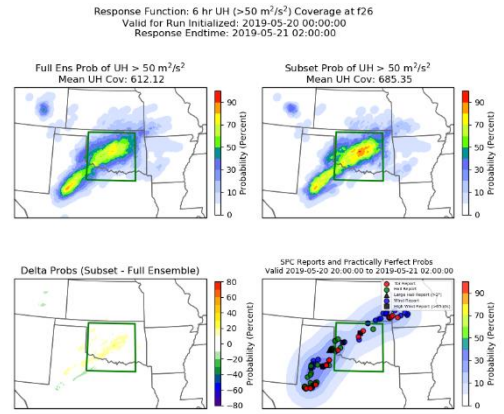


Hazardous Weather Testbed Spring Experiments

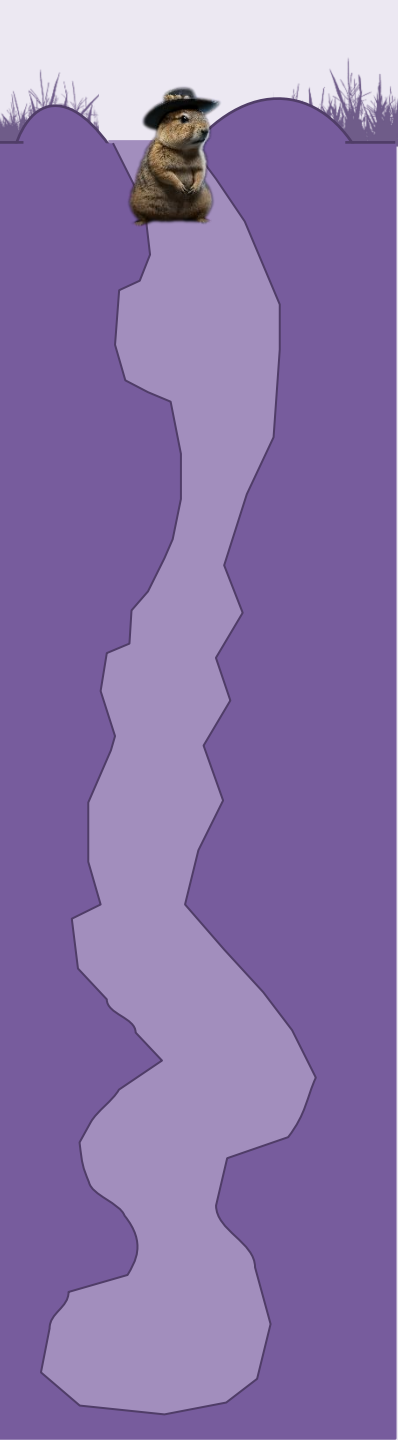
Variations of subsetting tool tested throughout its evolution at the 2018, 2019, 2020, and 2022 testbeds

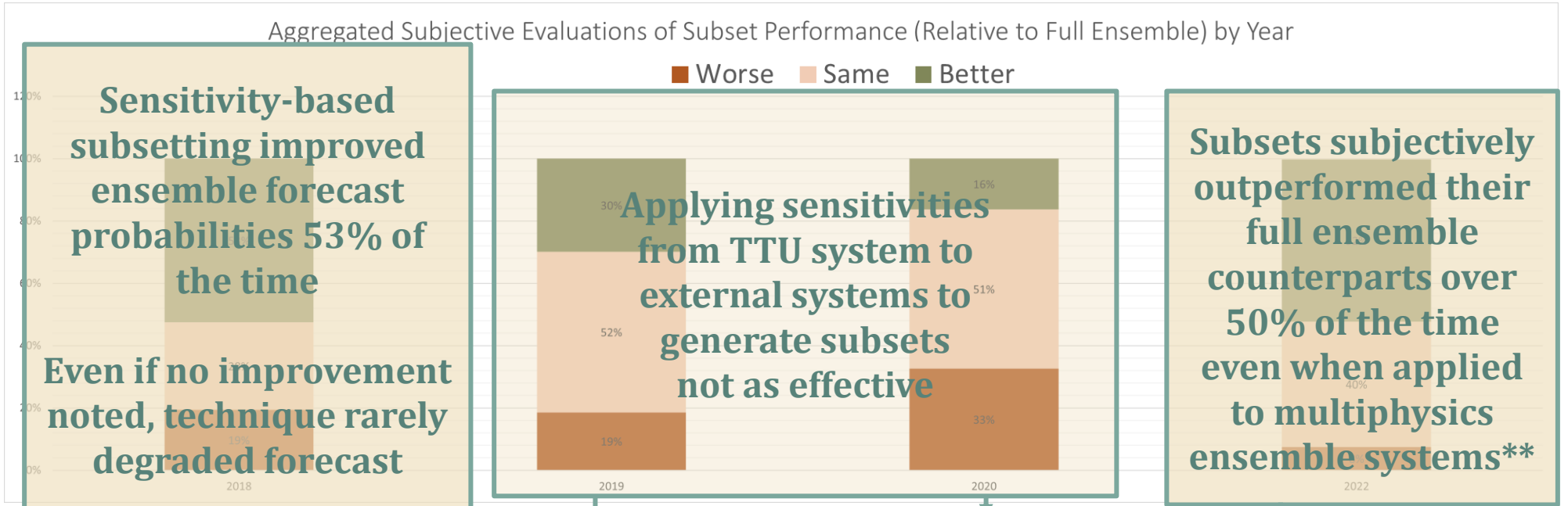
- Response box chosen by participants of HWT every day with a web GUI (SENSEI)
- Next day, participants compare subsets to full ensemble to reports and evaluate
- Testament to the importance of O2R in developing these tools! ☺





HWT SFE Overview of Results





Sensitivity-based subsetting improved ensemble forecast probabilities 53% of the time

Even if no improvement noted, technique rarely degraded forecast

Applying sensitivities from TTU system to external systems to generate subsets not as effective

Subsets subjectively outperformed their full ensemble counterparts over 50% of the time even when applied to multiphysics ensemble systems**

Sensitivities and subsets calculated within the TTU ensemble system

Sensitivities calculated within TTU system and applied to multiple IC ensemble systems from Community Leveraged Unified Ensemble (CLUE) to subset

Sensitivities calculated within TTU system and used to subset time-lagged HRRRE

Sensitivities and subsets calculated entirely within suite of multimodel/multiphysics ensemble systems (all FV3-based)

2018 2019 2020 2022

**Statistics only aggregated over 9 cases due to data flow issues throughout the 2022 HWT SFE