Tools to Utilize Ensembles During the Forecast Process: A Stony Brook CSTAR Perspective

BRIAN A. COLLE, EDMUND CHANG, TAYLOR MANDELBAUM, MINGHUA ZHENG, AND RUI ZHANG

Stony Brook University

School of Marine and Atmospheric Sciences, Stony Brook, NY

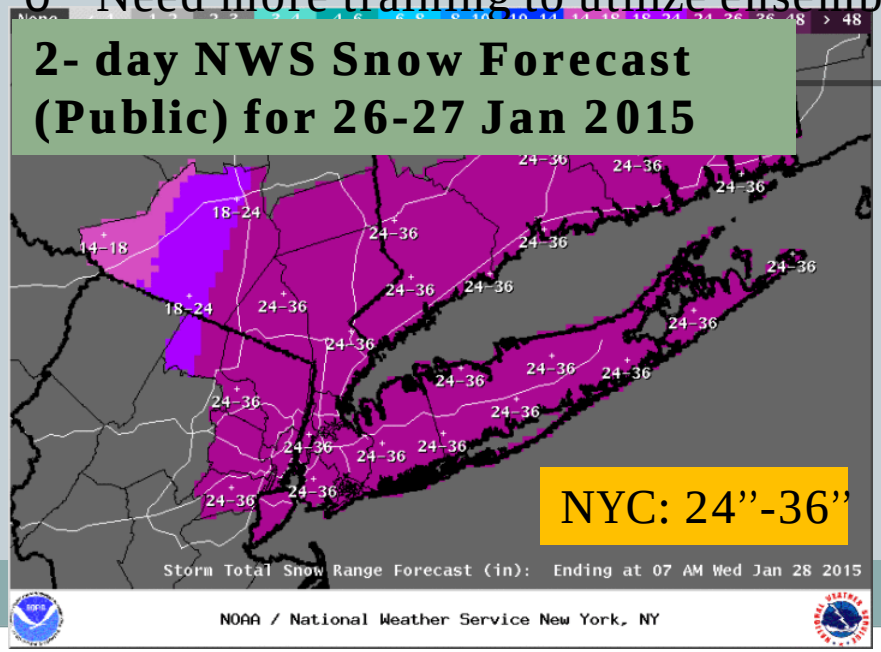
AND OUR CSTAR PARTNERS:

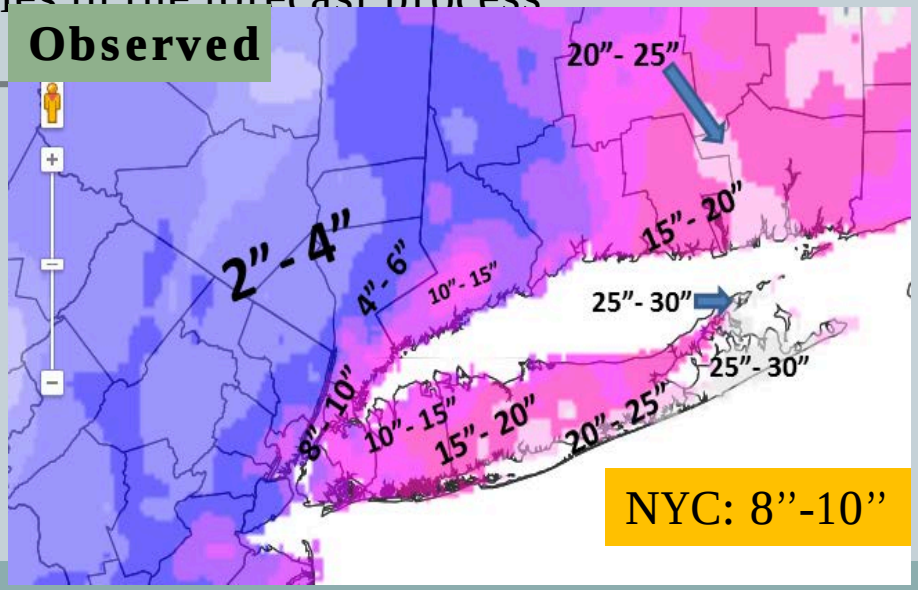
NYC, BOSTON, PHILLY, ALBANY NWS WFOS; NWS-EREGION, NCEP-EMC AND NCEP-WPC

* This work is supported by NOAA-CSTAR

Some Operational Challenges

- Underutilization of ensemble forecasts in operations (2014 NWS CSTAR survey – ~50 forecasters):
  - **Lack of graphics/tools to display and understand ensemble predictions (highest rank in the survey)**
  - Limited ensemble data in the office (bandwidth issues)
  - Limited time to synthesize ensemble data during an operation forecast process
  - Need more training to utilize ensembles in the forecast process





Select CSTAR Tools (2012-present)

- Ensemble Sensitivity: Determines upstream features leading to ensemble spread or dModel/dt
- Fuzzy Clustering: Scenario determination and maps for 4-5 different clusters (EC+GEFS+CMC).
- Ensemble Cyclone Tracks: GEFS+CMC+FNOC+SREF tracks, track probabilities, and GEFS bias correction using cyclone verification.
- Ensemble Rossby Wave Packets: GEFS wave packet amplitude probabilities and spread .
- Spread-Anomaly Tool: GEFS spread anomalies based on reforecast dataset
- http://breezy.somas.stonybrook.edu/CSTAR/

Motivation for Fuzzy Clustering

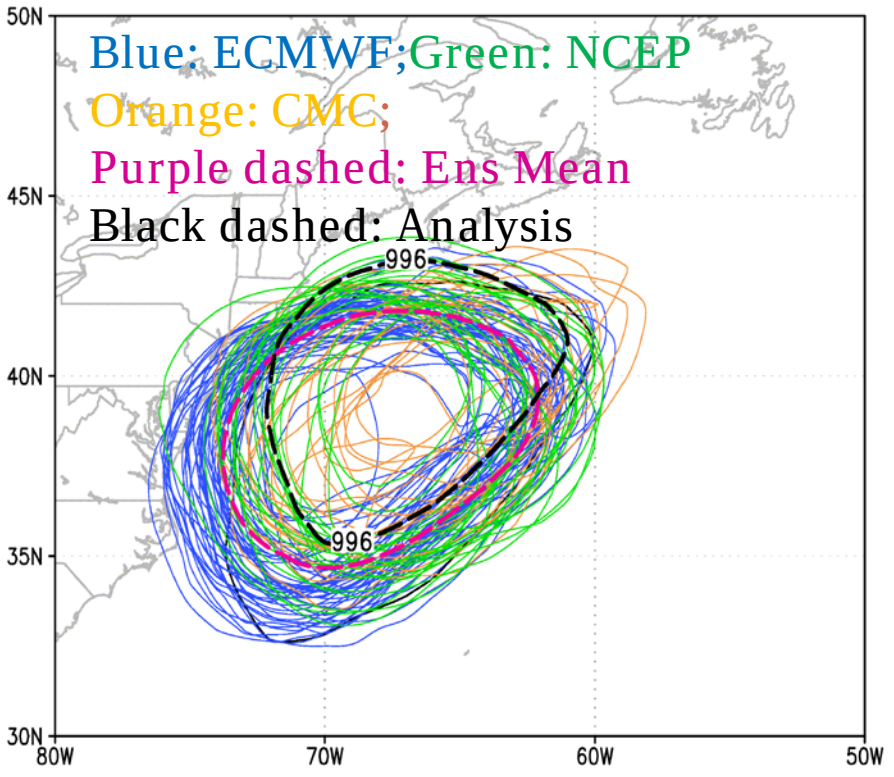
- **Operational**
  - ✓ To quickly separate forecast scenarios among a large ensemble set in a forecast.
  - ✓ Provide scenarios based on a mix of ensembles, rather ensemble A versus ensemble B (e.g., EC vs GEFS).
- **Some research questions**
  - ✓ Can fuzzy clustering efficiently separate forecast scenarios in multi-model ensemble?
  - ✓ Which ensemble system is more reliable in terms of capturing scenarios associated with cyclone intensity and track for East Coast storms?

Fuzzy Clustering Data and Methods

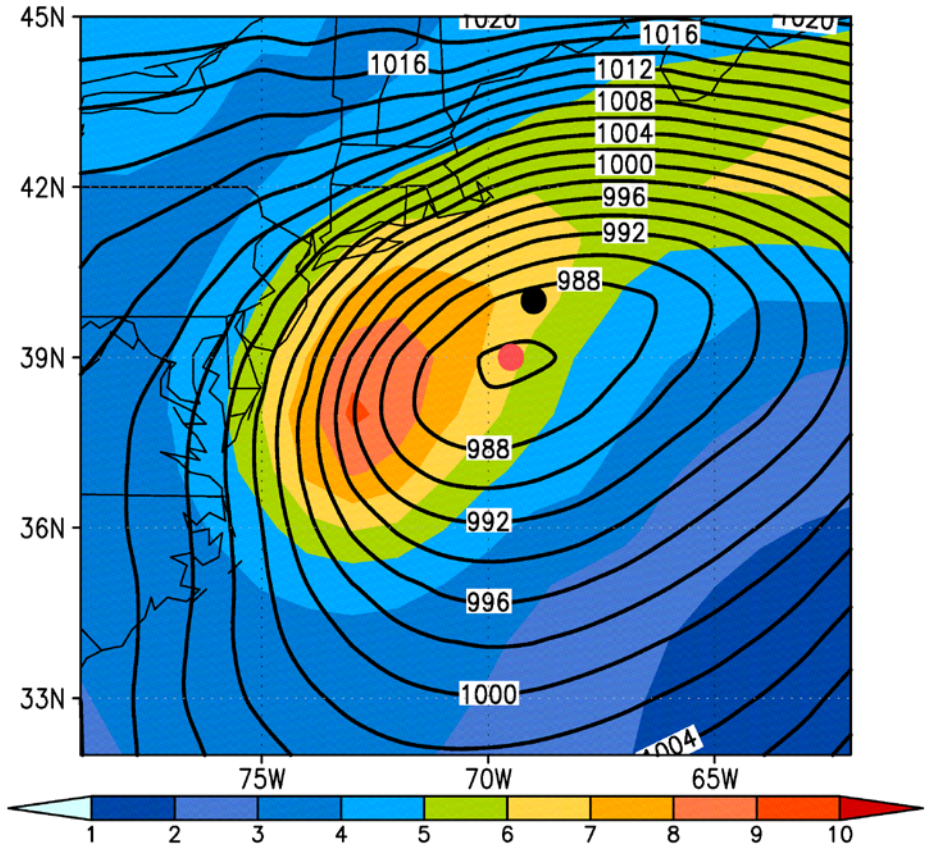
- Data:
  - TIGGE Ensemble forecast archive: NCEP (20 mem) + CMC (20 mem) + ECMWF (50 mem)
  - For Real-time (0000 and 1200 UTC) – Scripts run at EMC – Thank you Yan Luo and Yuejan Zhu; WPC also runs a version
  - Cluster Validation: NCEP operational analysis
  - Variables: MSLP, Z500, precipitation, and 925 hPa temp
  - Historical cases selections: 124 (114 for US East coast region) cyclone cases (minimum pressure <996 hPa) from 2007 to 2014 cool seasons (NDJFM) using Hodges cyclone tracker.
- Approach (see Zheng et al. 2017; 2019)
  - Empirical Orthogonal Function (EOF) analysis on ensemble spread
    - To quantify dominate ensemble SLP spread patterns
  - Fuzzy clustering analysis
    - To group ensemble members based on EOF PCs.

Example 1 (NYC "Blizzard": 3-Day forecast, IT: 12 UTC Jan 24 2015; VT: 12 UTC Jan 27 2015

Spaghetti plot of 996 hPa MSLP ([hPa]) and analysis at Jan 27th 2015 12Z

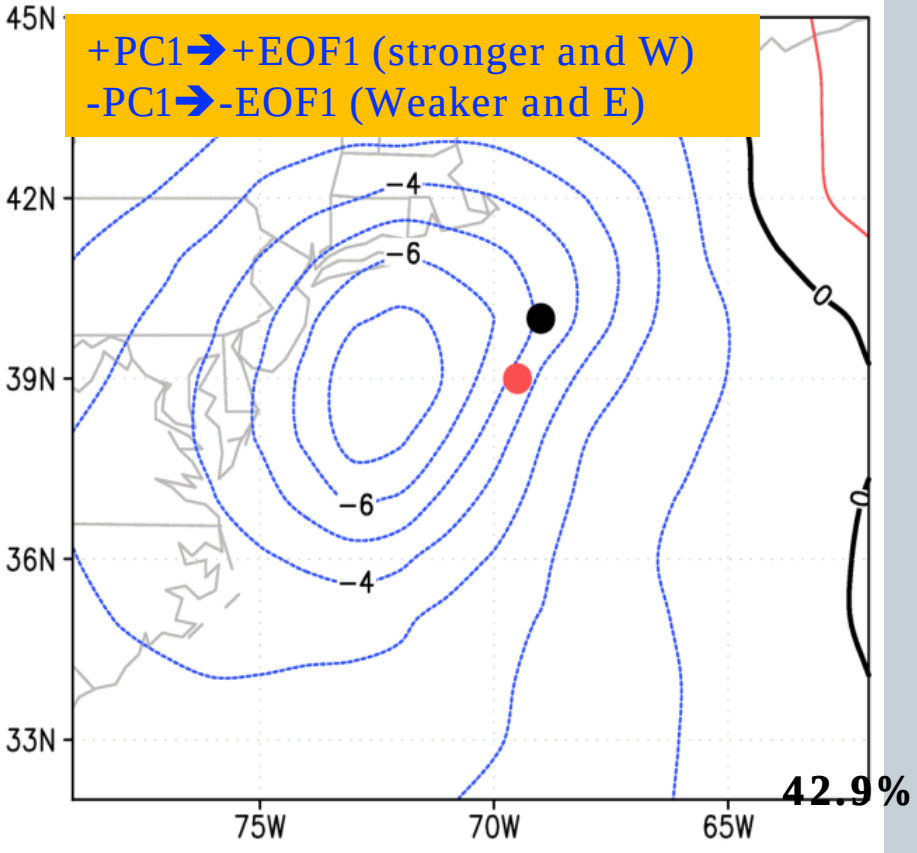


Ensemble mean (contours) and spread (shades) of MSLP, [hPa]

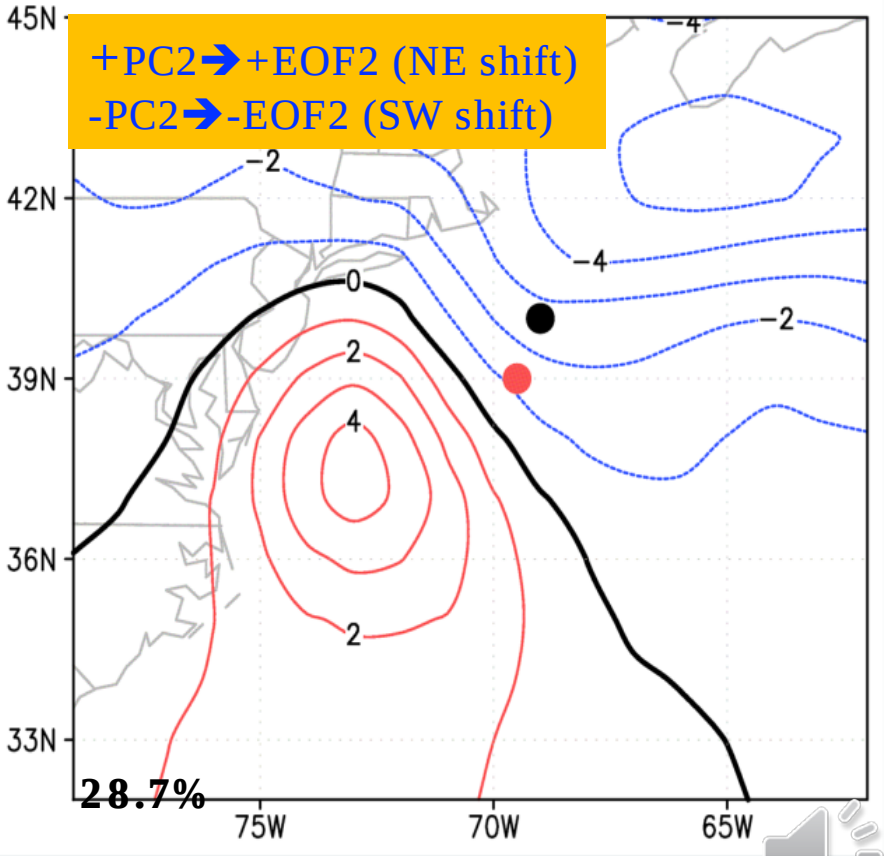


STEP1: EOF analysis of MSLP on *90* members of forecasts at VT

EOF1 (43% variance)
MSLP anomaly pattern, [hPa]



EOF2 (23% variance)
MSLP anomaly pattern, [hPa]



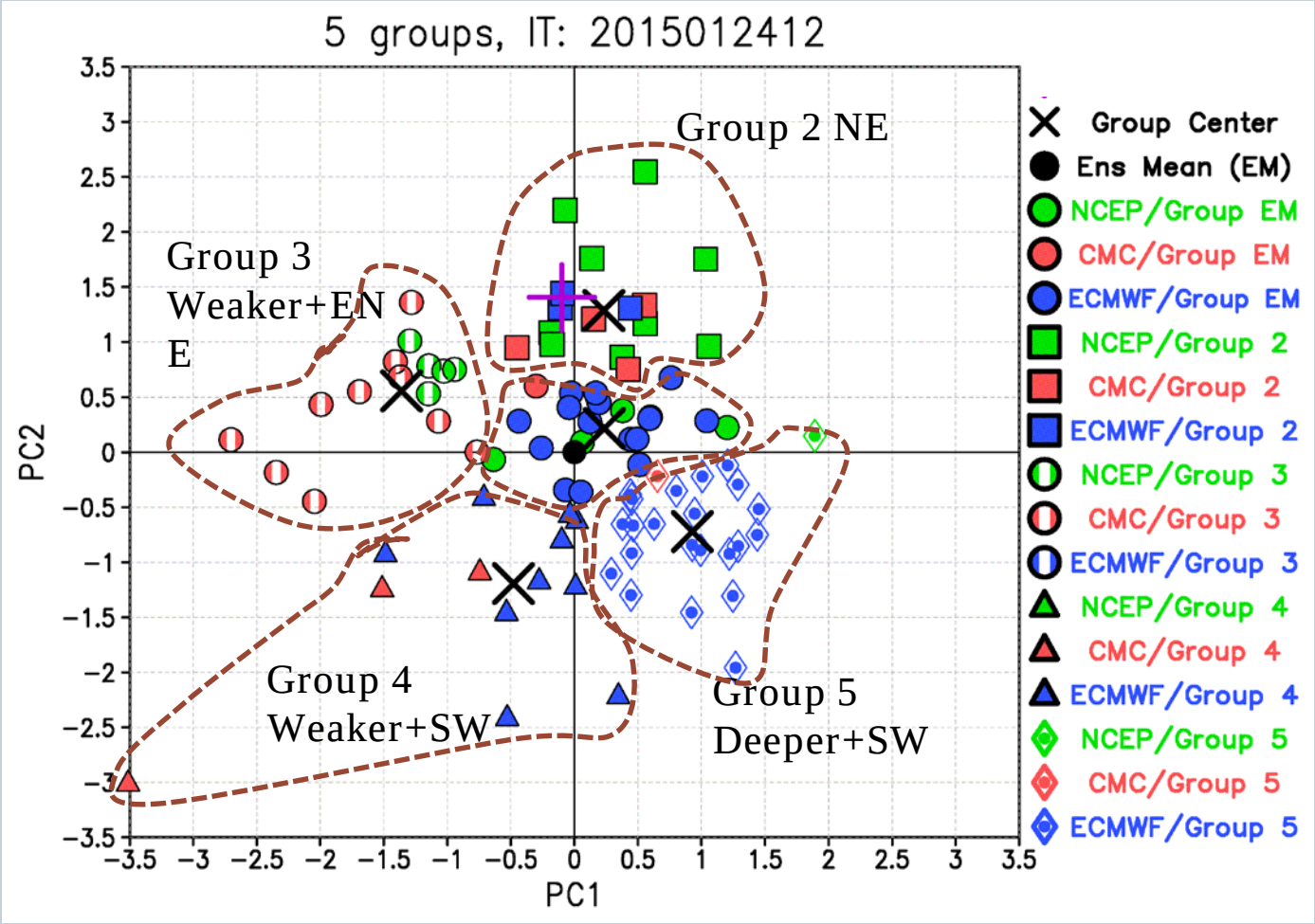
Each member has a PC value to represent its projection on each pattern

STEP2: group ensemble members into 5 clusters based on PCs using Fuzzy clustering scatter plots

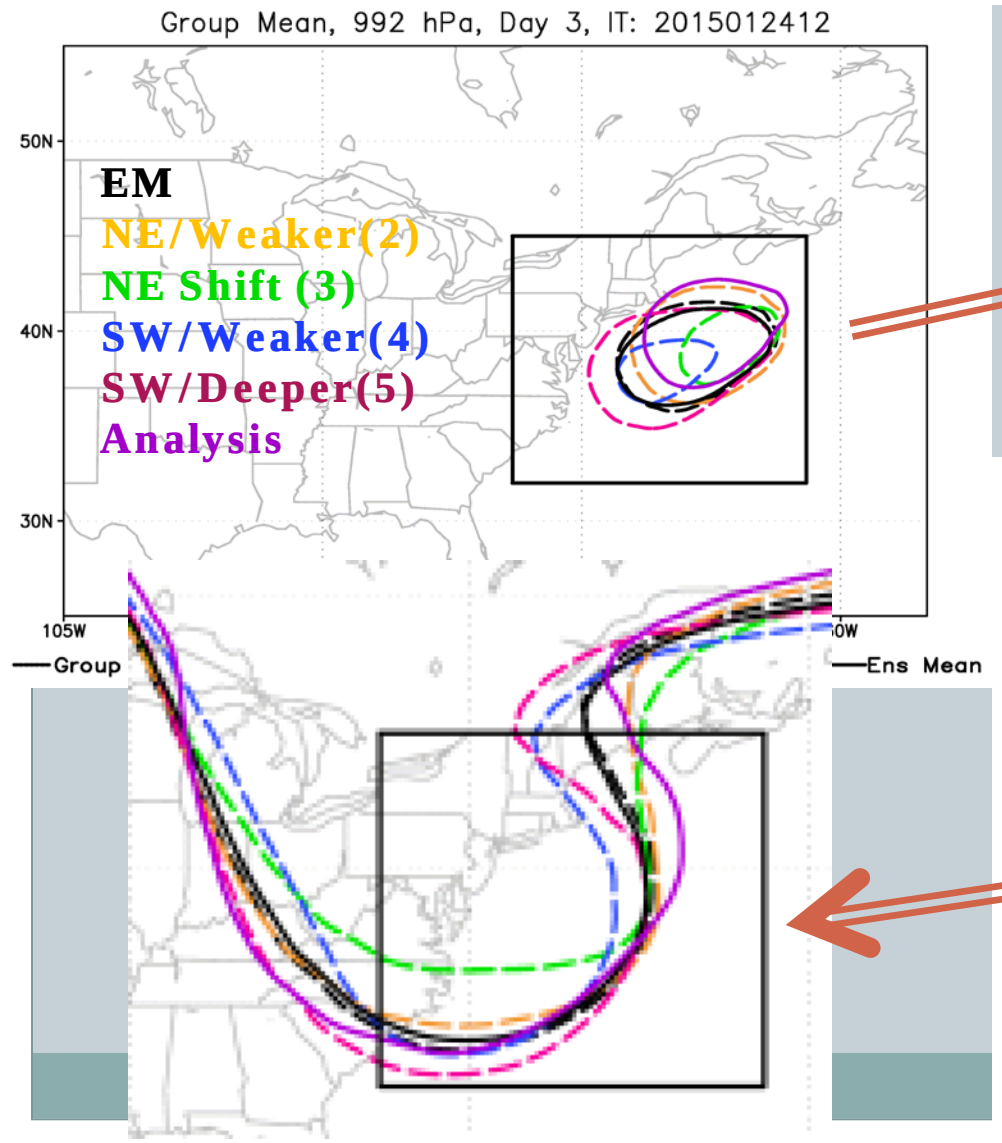
+PC1: Deeper +W
-PC1: Weaker + E
+PC2: NE
-PC2: SW

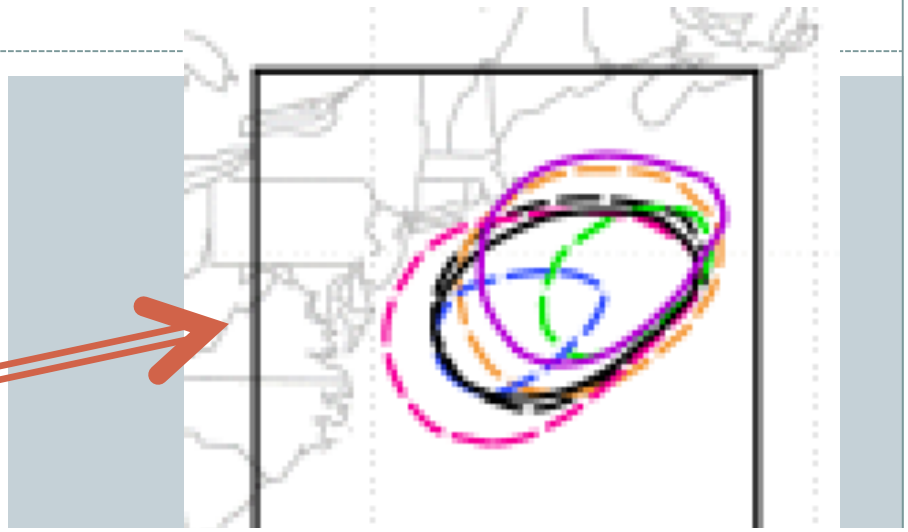
Fuzzy cluster:
Harr et al. (2008)

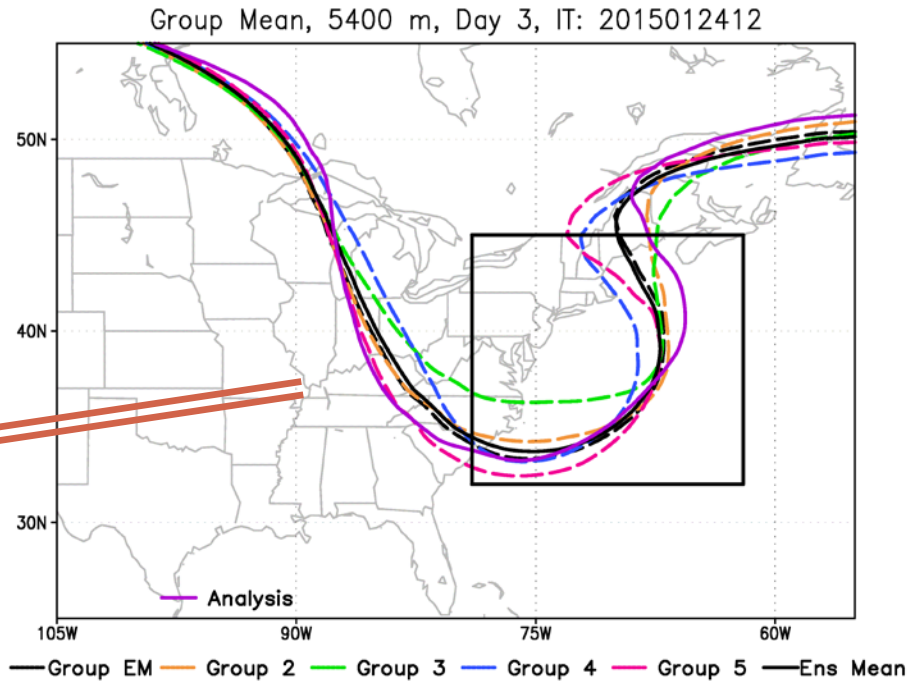
Each member is assigned a weight that identifies its relative strength of membership to each of the five clusters depending on its distance from the cluster mean in the PC phase space



STEP3: Pick up a contour line and plot group mean summary based on the partitions of clusters



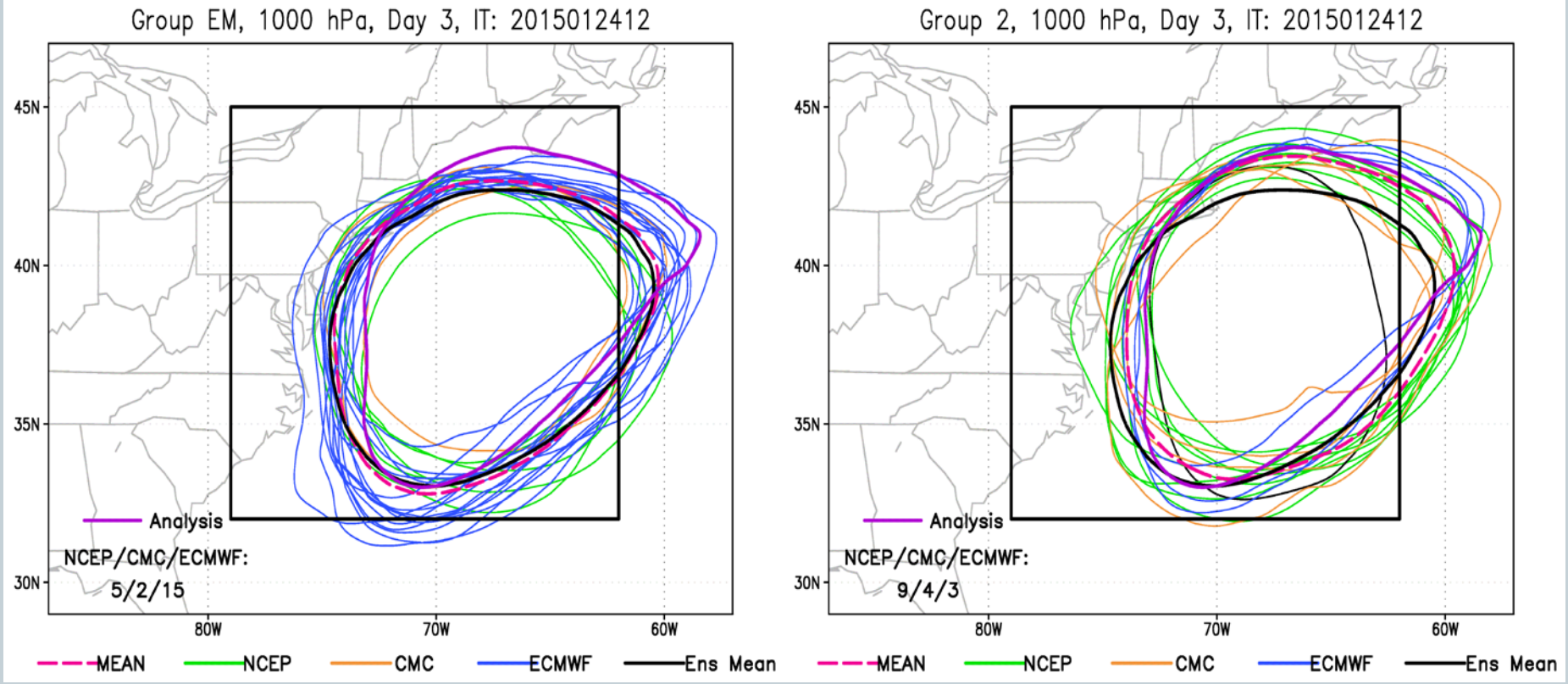




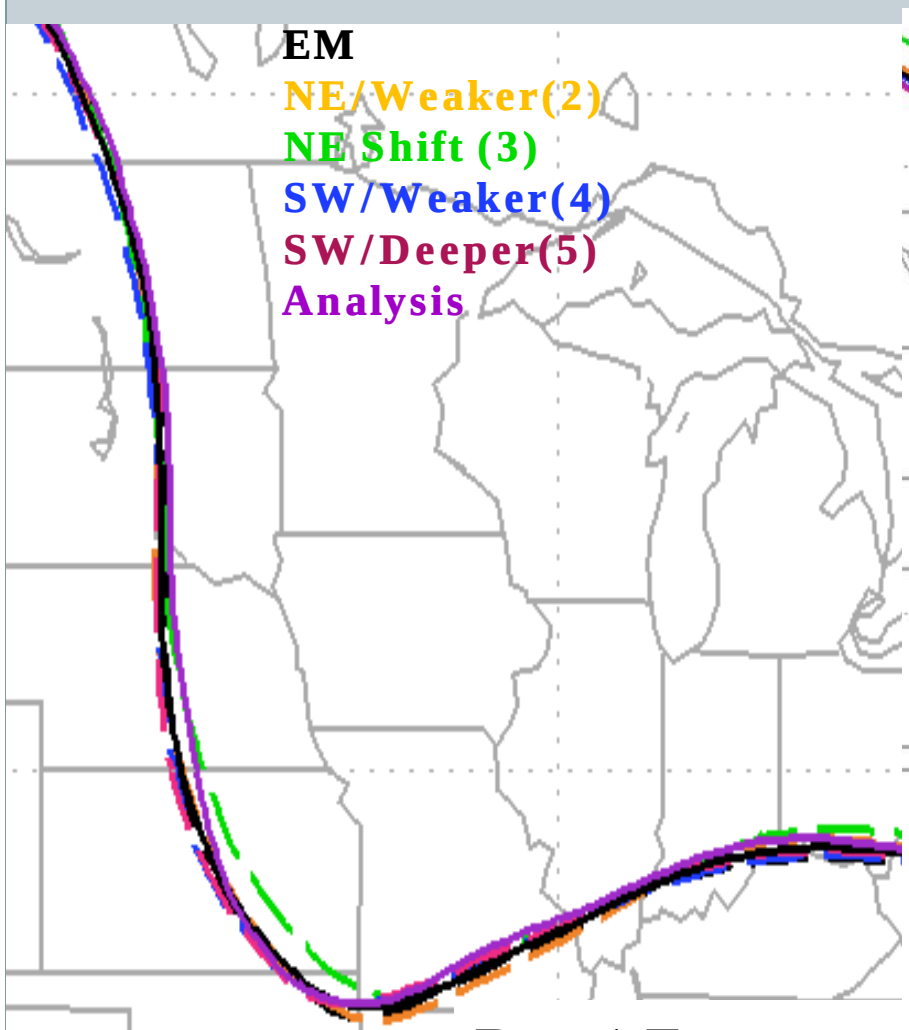
STEP4: Can plot spaghetti plots for each group, e.g., Group EM and Group 2

Purple solid: analysis; Magenta dashed: cluster mean

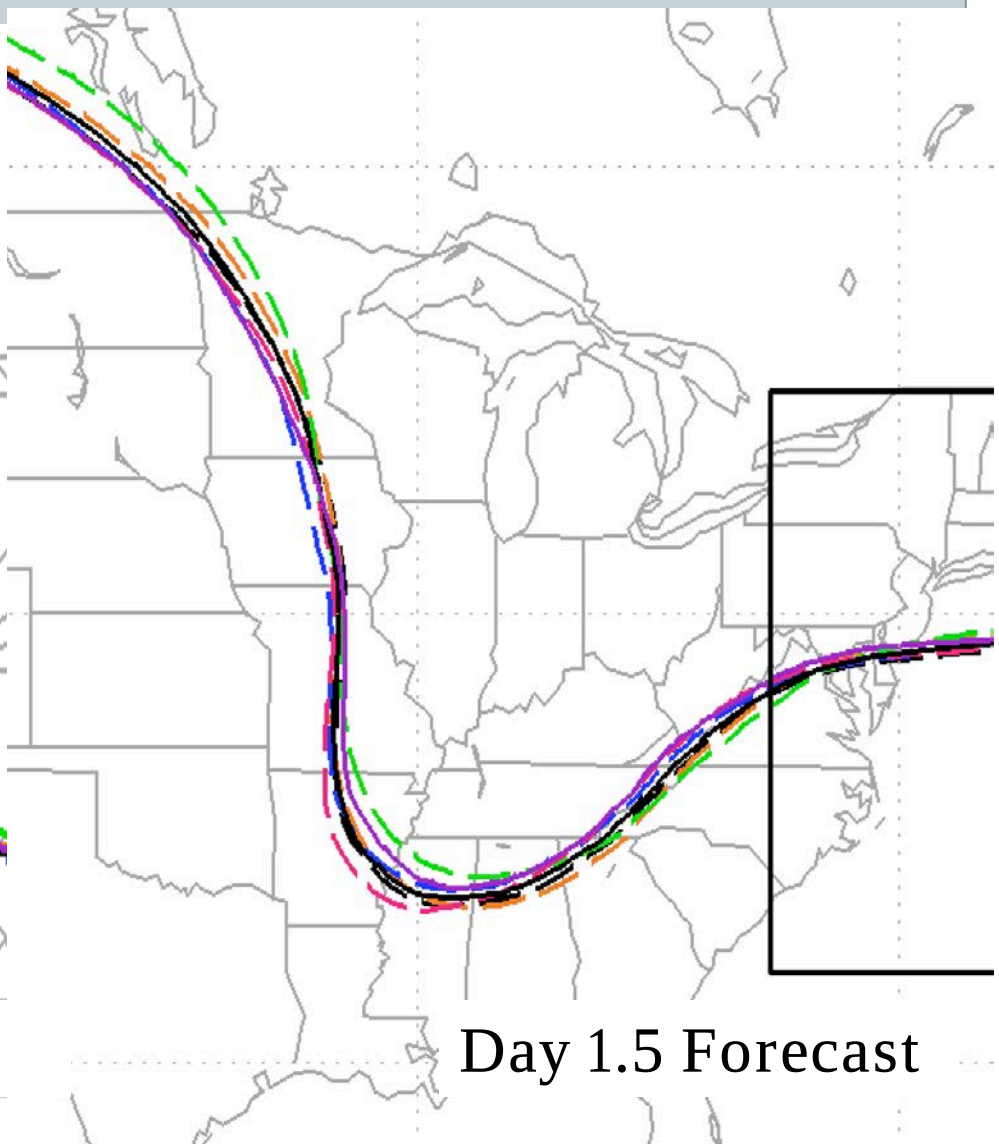
Black: ensemble mean



STEP5: Look at the evolution of the clusters upstream and how they compare to analysis as they become available...



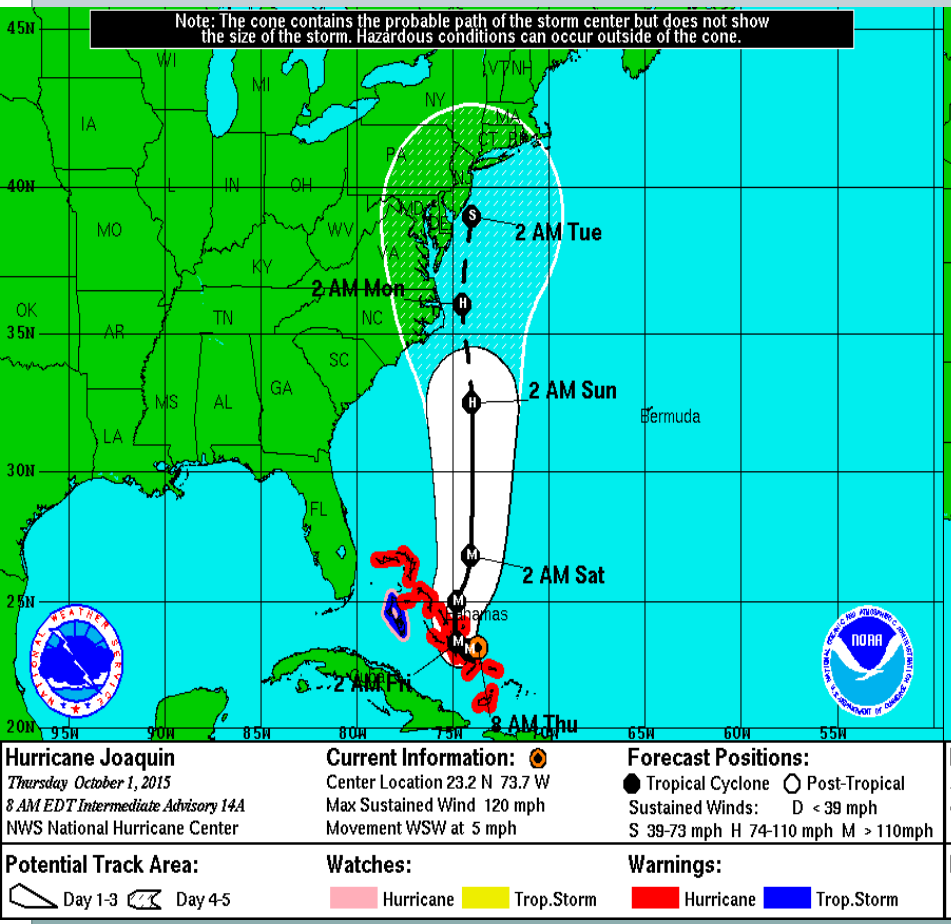
Day 1 Forecast



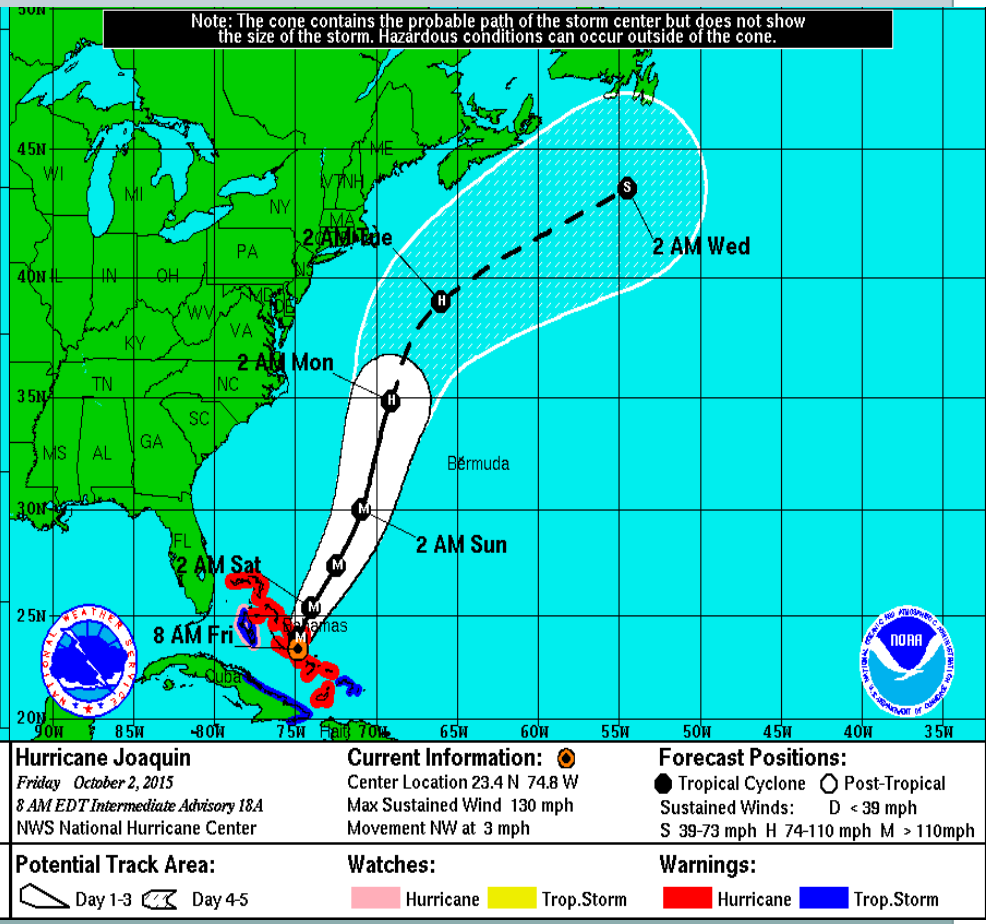
Day 1.5 Forecast

Hurricane Joaquin Track Uncertainty

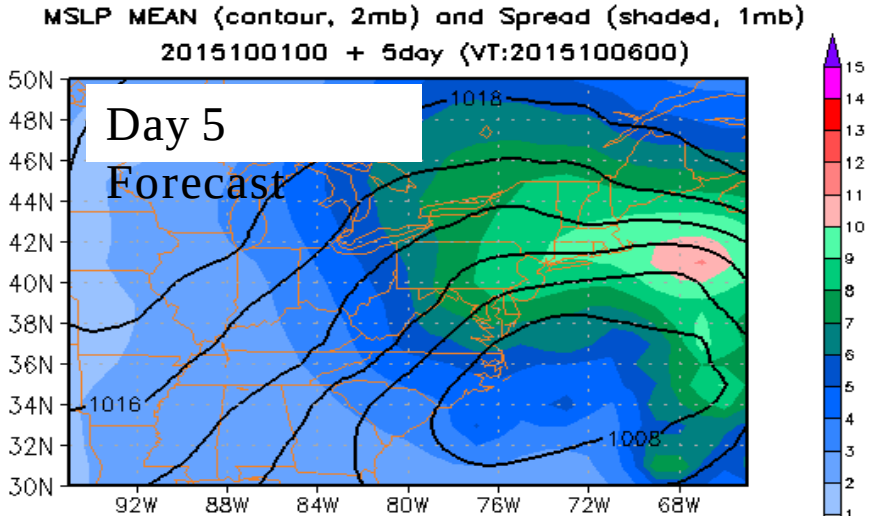
1200 UTC 1 Oct 2015

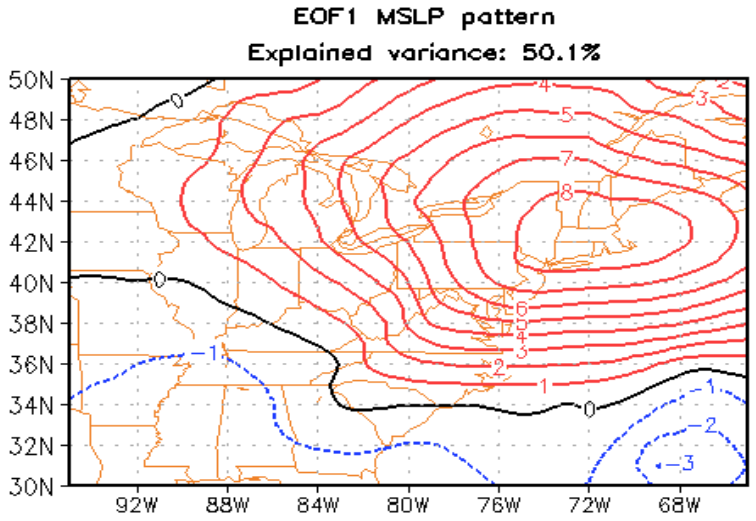


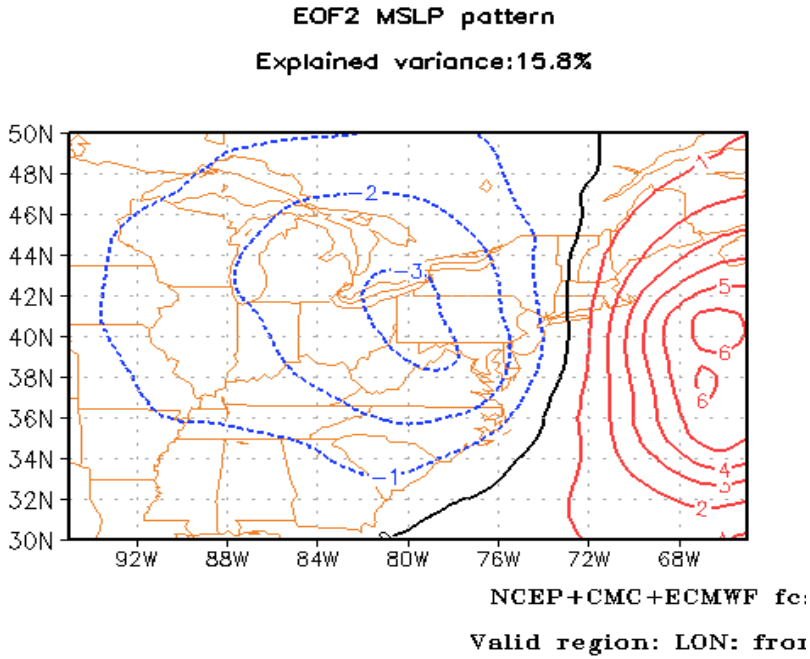
1200 UTC 2 Oct 2015

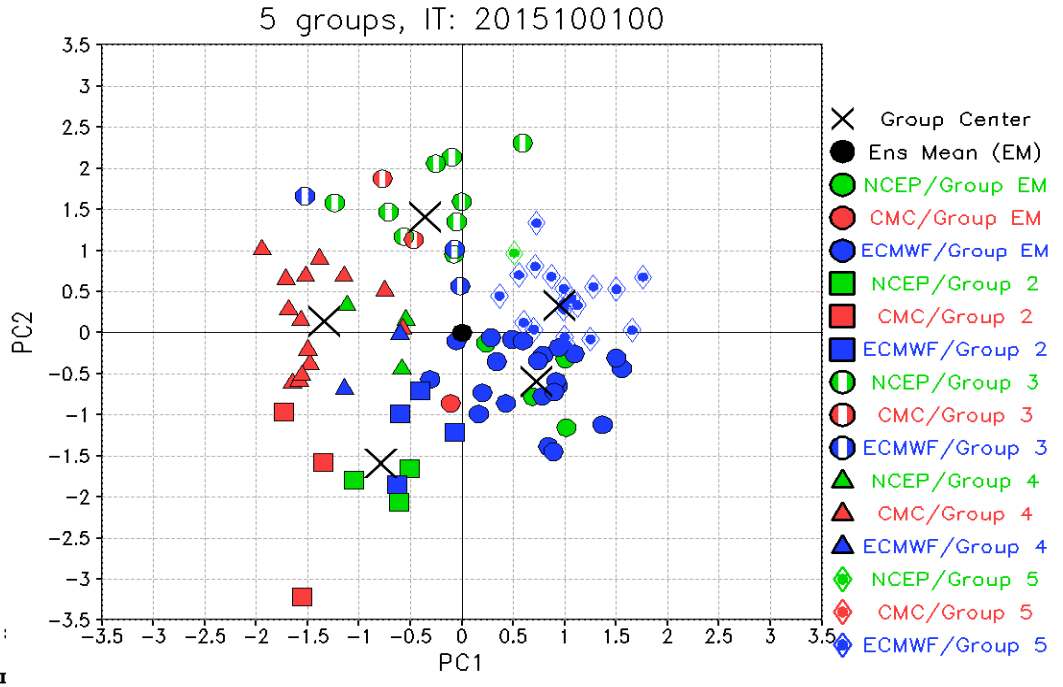


Spread and EOFs for 0000 UTC 1 October 2015 Run (GEFS+EC+CMC)

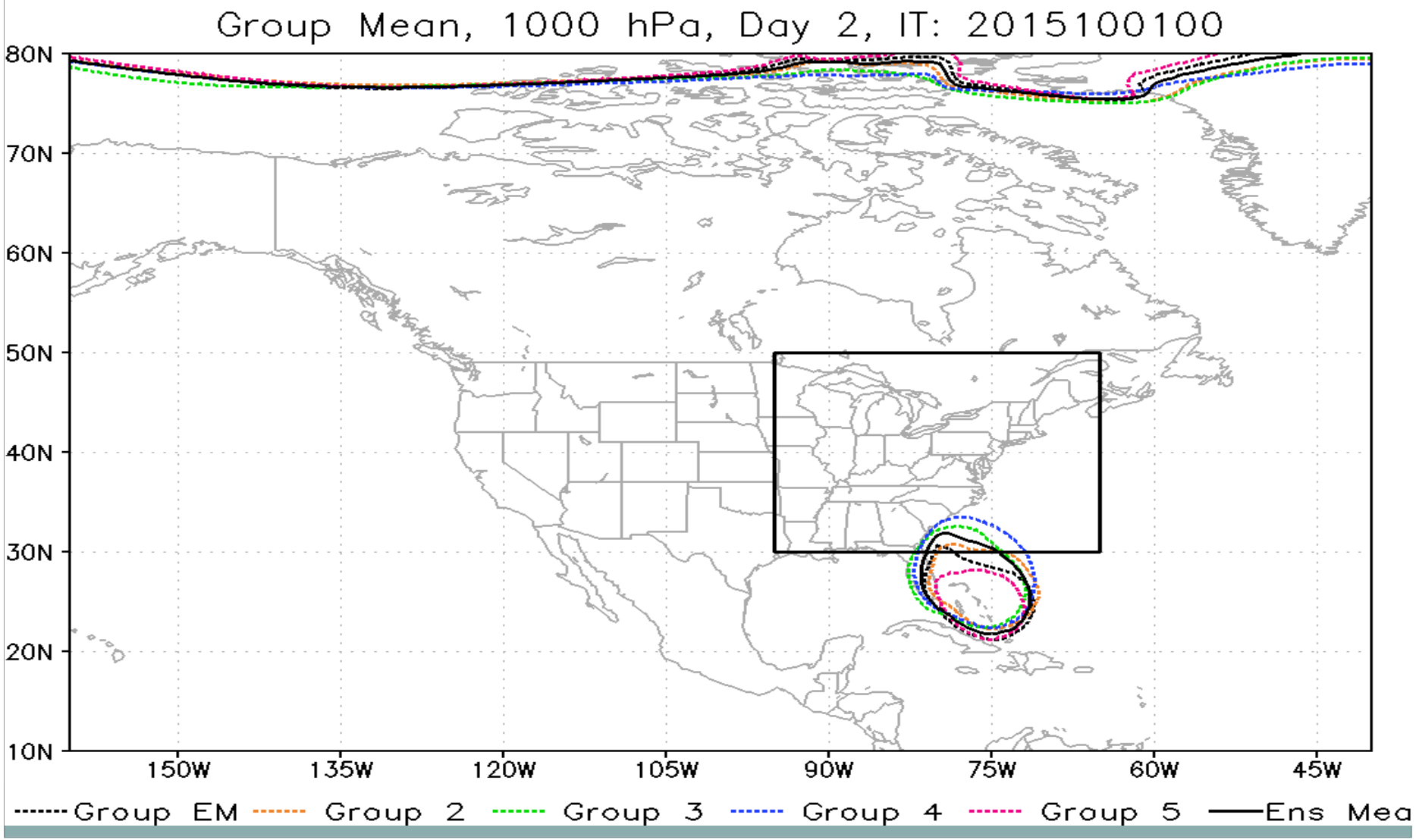




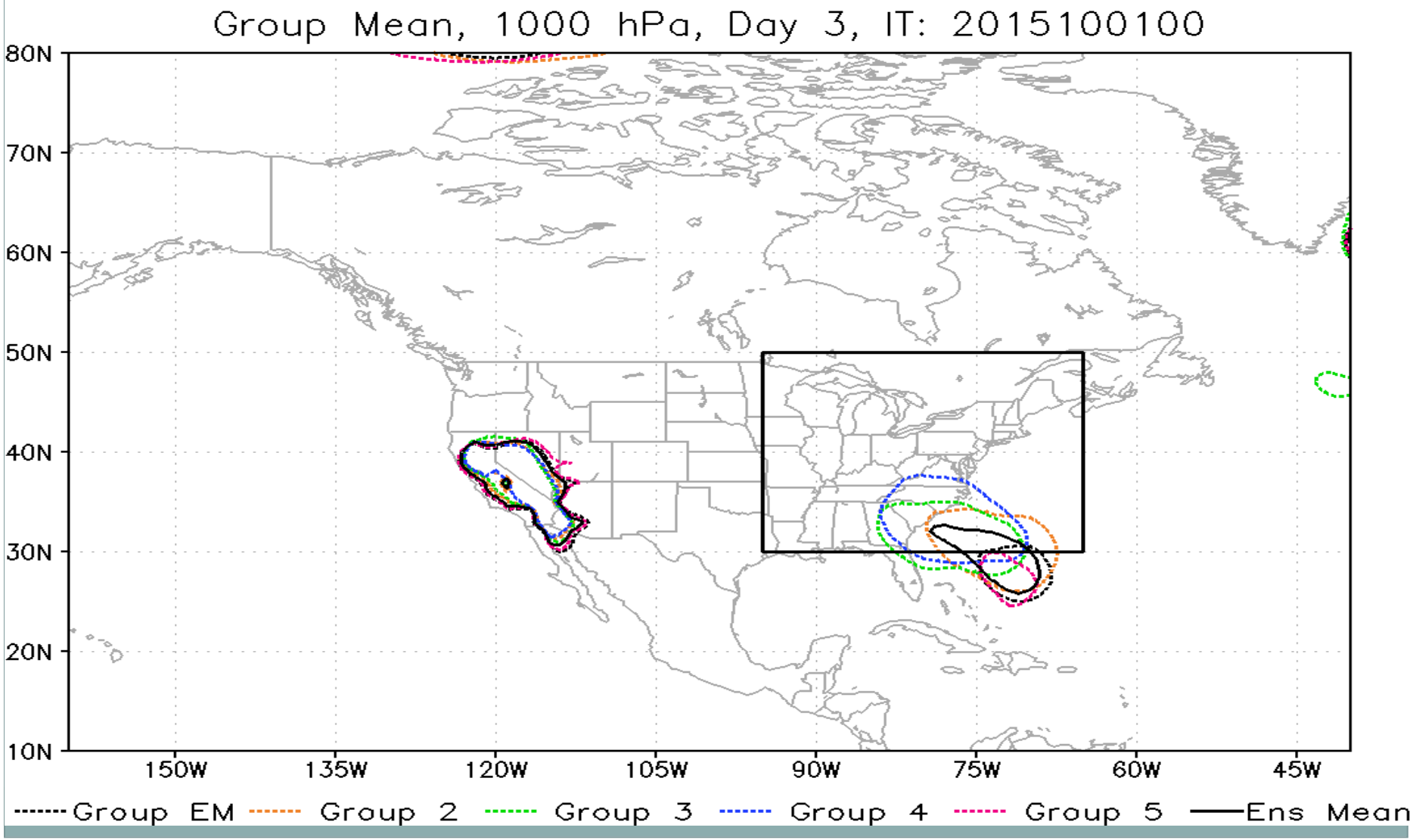




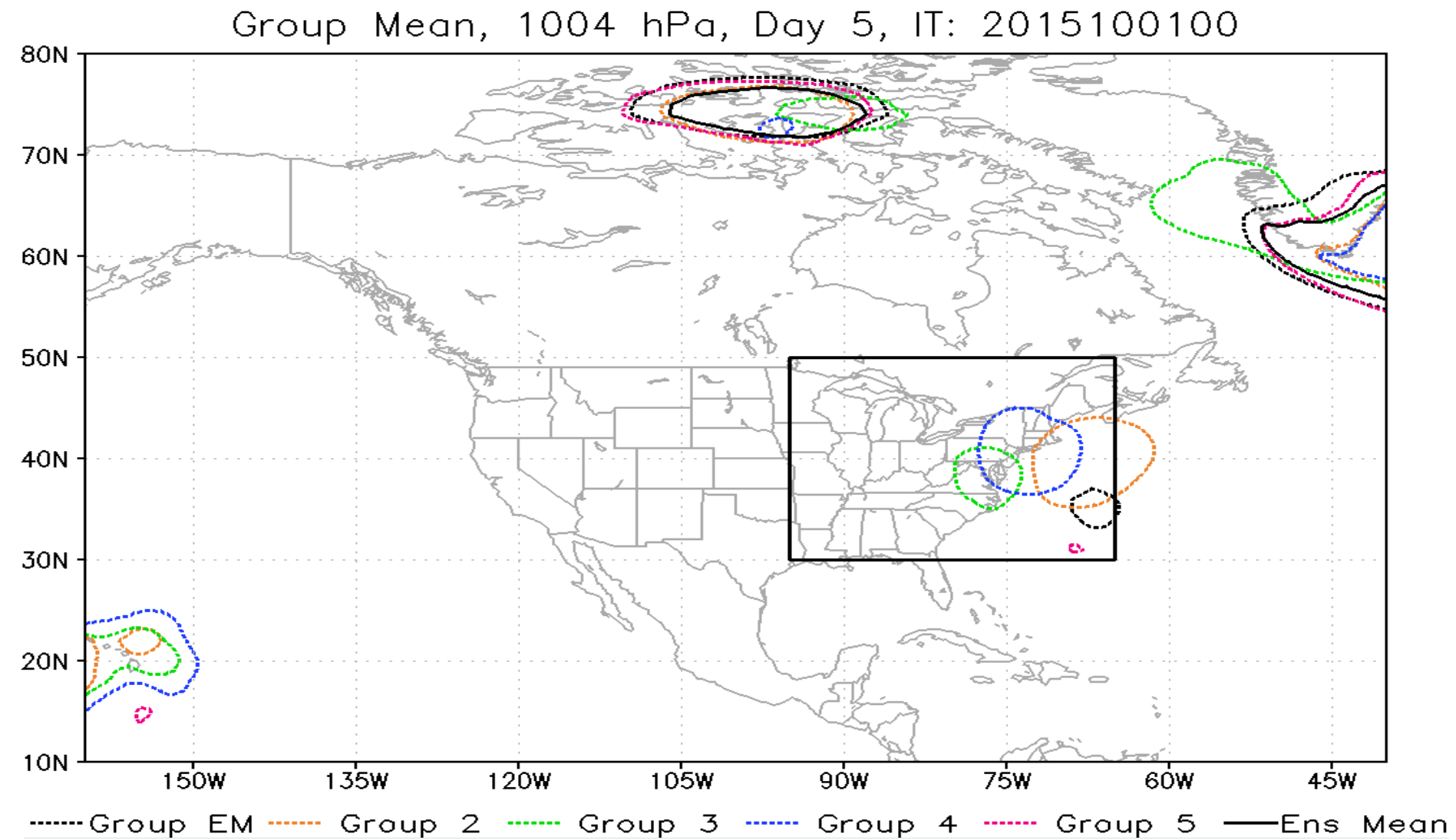
Cluster Groups and Ens Mean
For 0000 UTC 1 Oct Cycle (1000 hPa SLP)



Cluster Groups and Ens Mean
For 0000 UTC 1 Oct Cycle (1000 hPa SLP)



Cluster Groups and Ens Mean
For 0000 UTC 1 Oct Cycle (1004 hPa SLP)

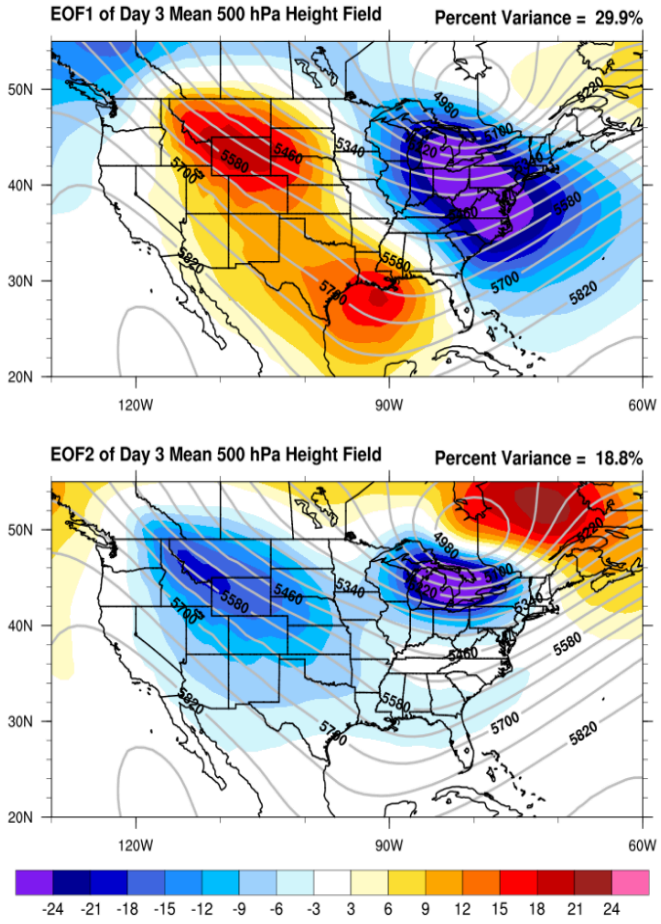


Clustering Implemented at WPC (Courtesy B. Lamberson)

WWE 2019 Cluster Prototype Page

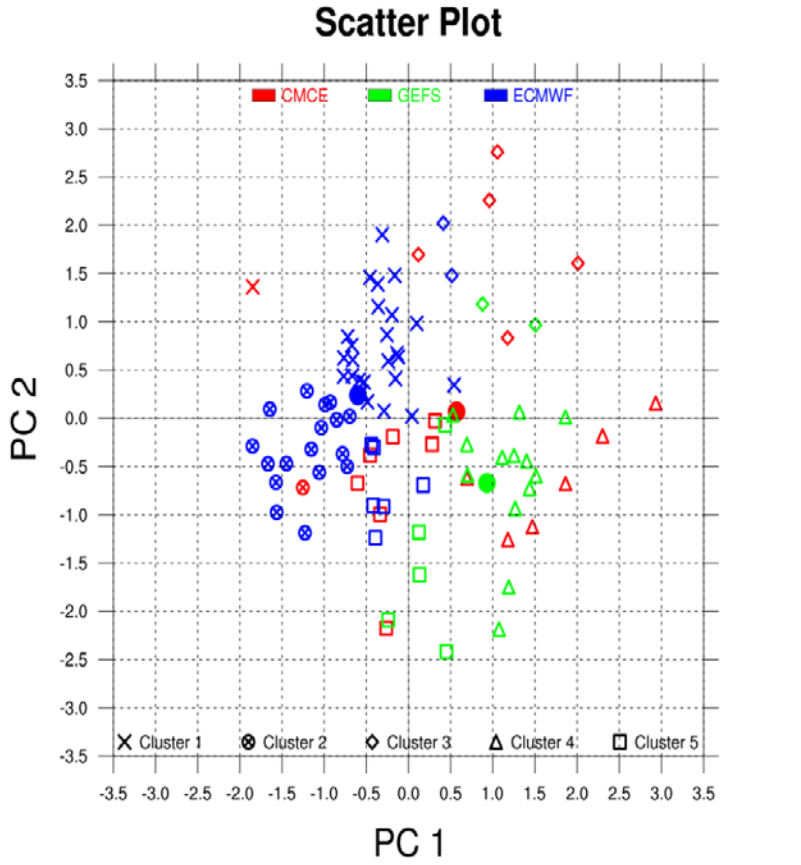
Day 3 | 500 mb EOF Patterns | F084 | CONUS

Init: 0000 UTC Jan 25 2019

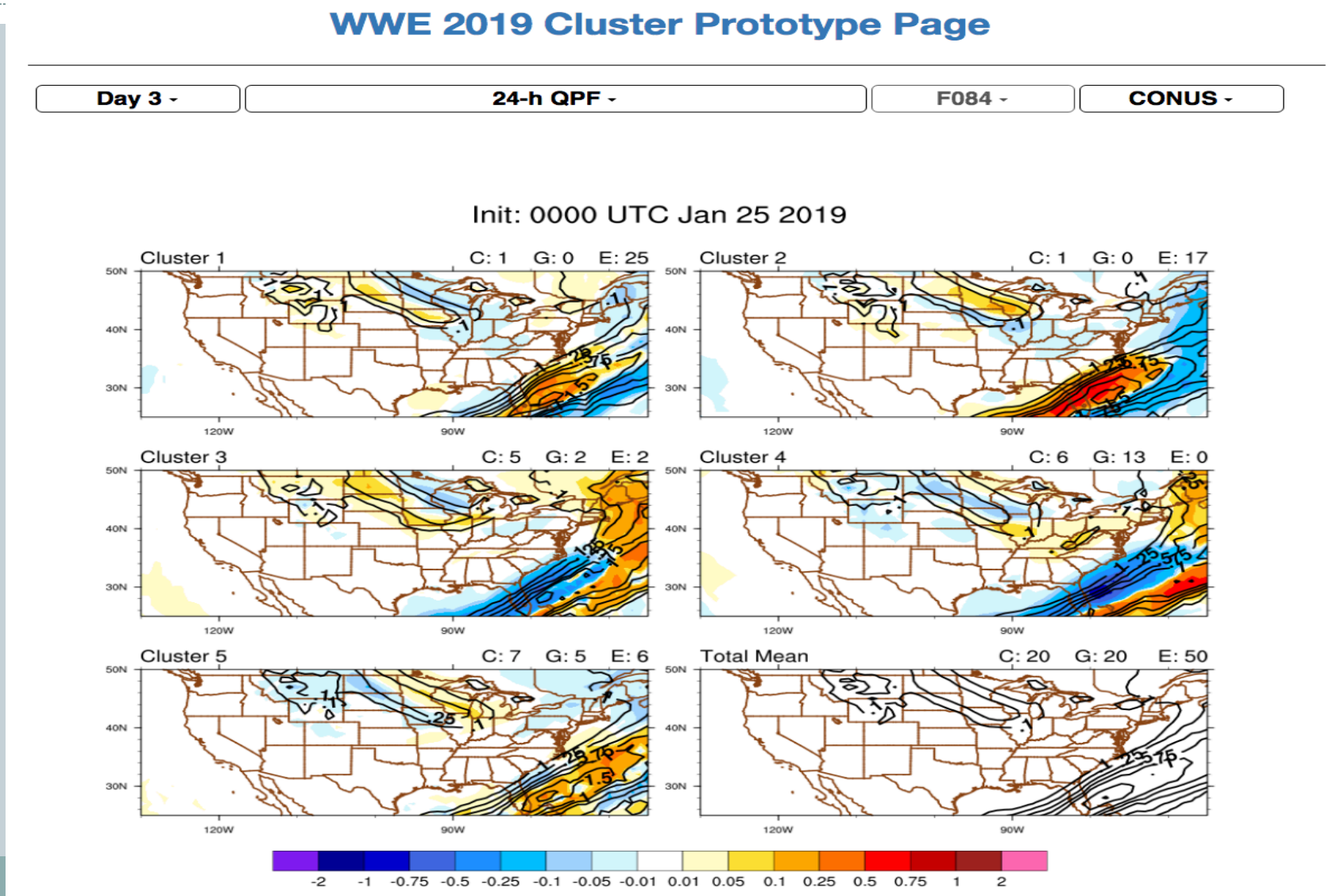


WWE 2019 Cluster Prototype Page

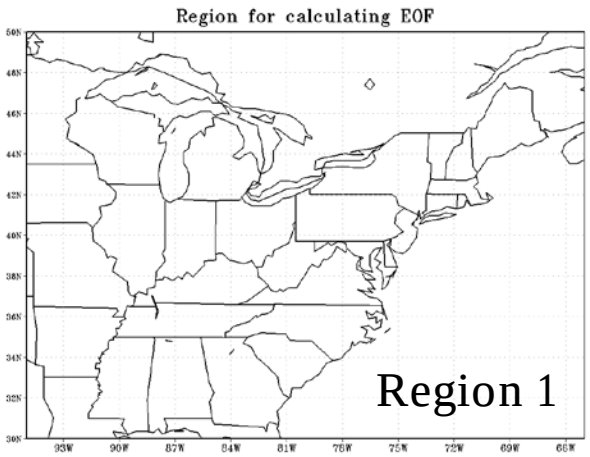
Day 3 | Cluster Phase Space | F084 | CONUS

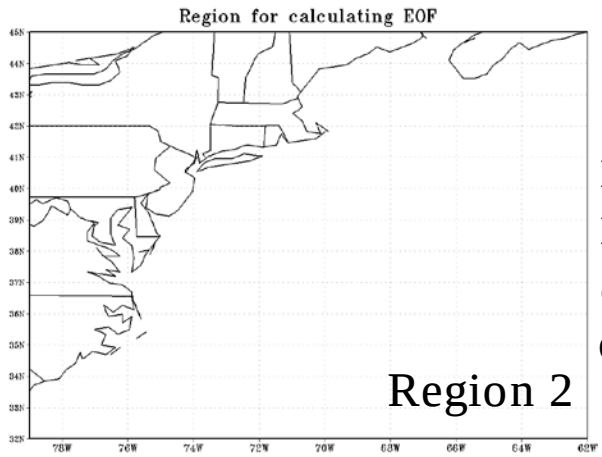


24-h Precip for the 500Z clusters and difference wrt to total mean (shaded)

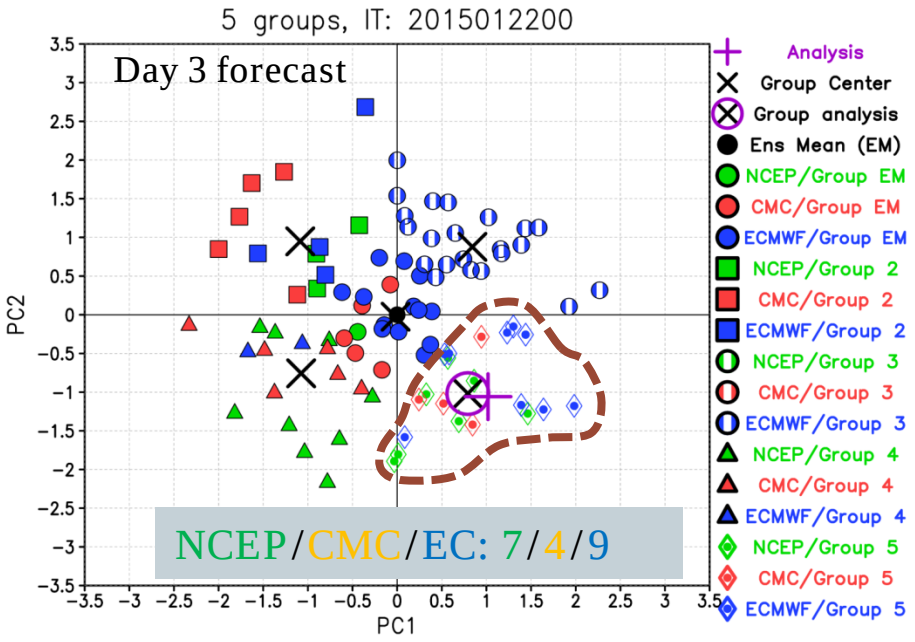


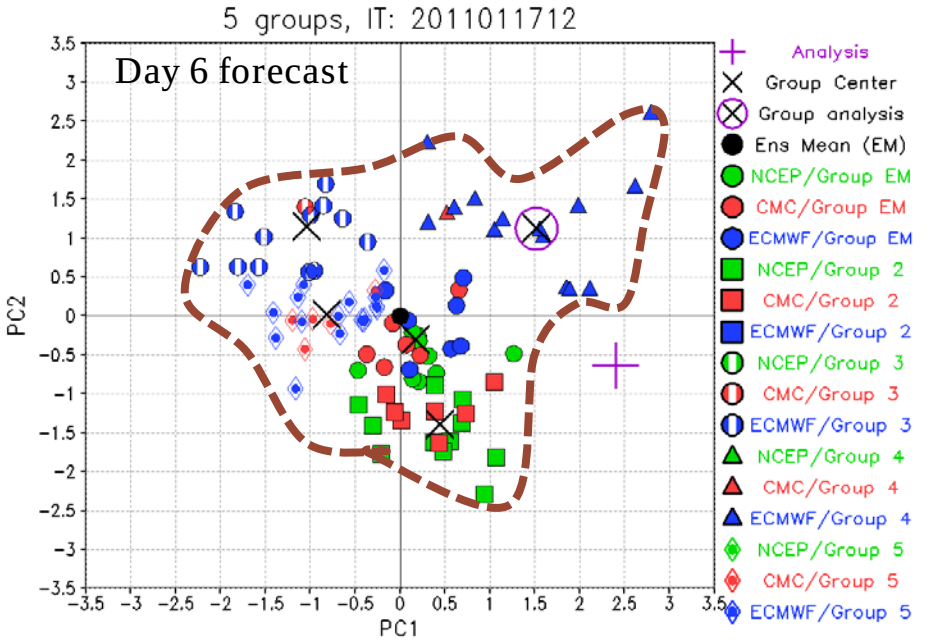
Historical evaluations using 124 (114) extratropical cyclones Nov-March 2007-2014)





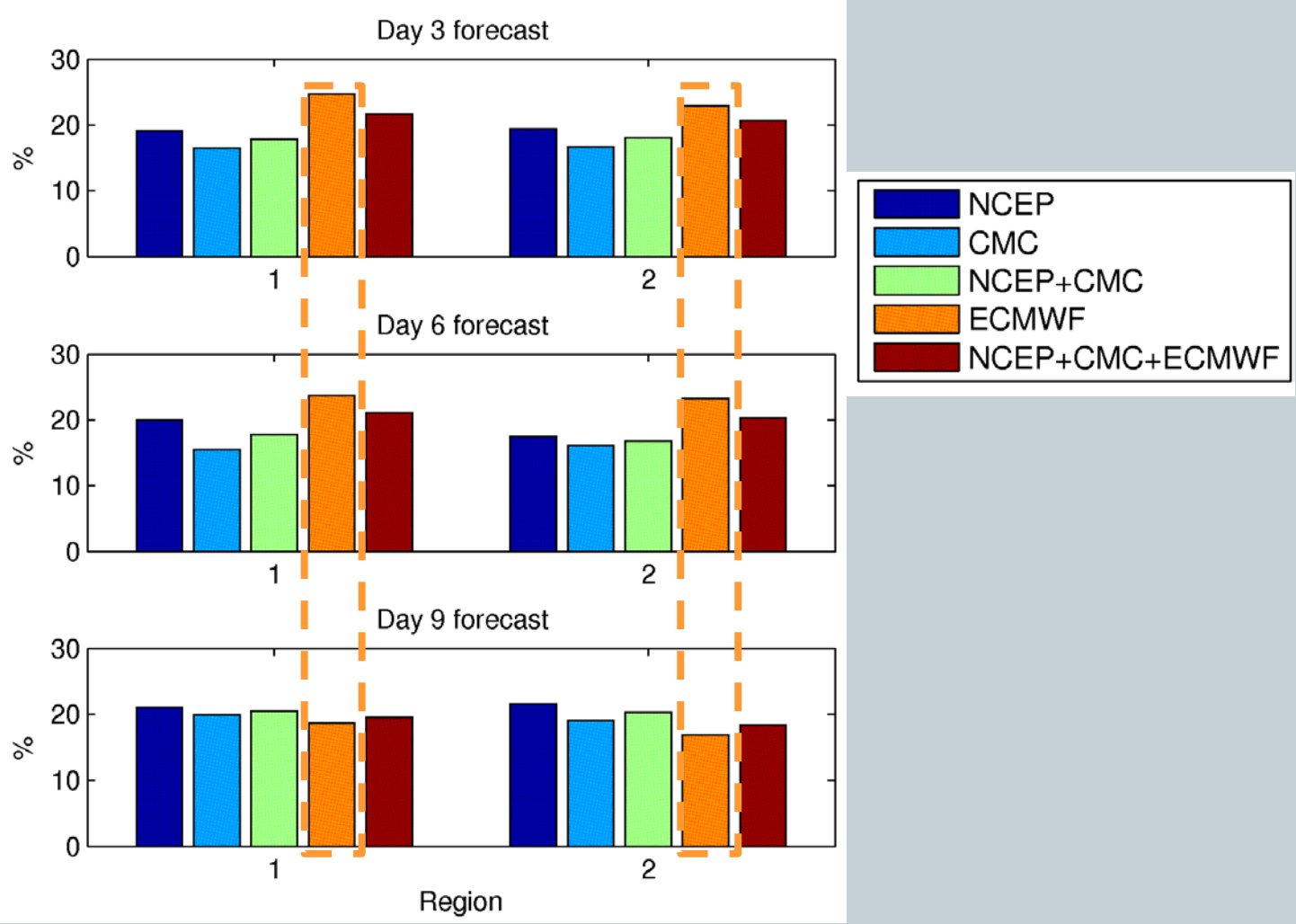
124 cyclone cases for region 1
114 cyclone cases for region 2
8 out-of-envelope or outlier cases are not included.



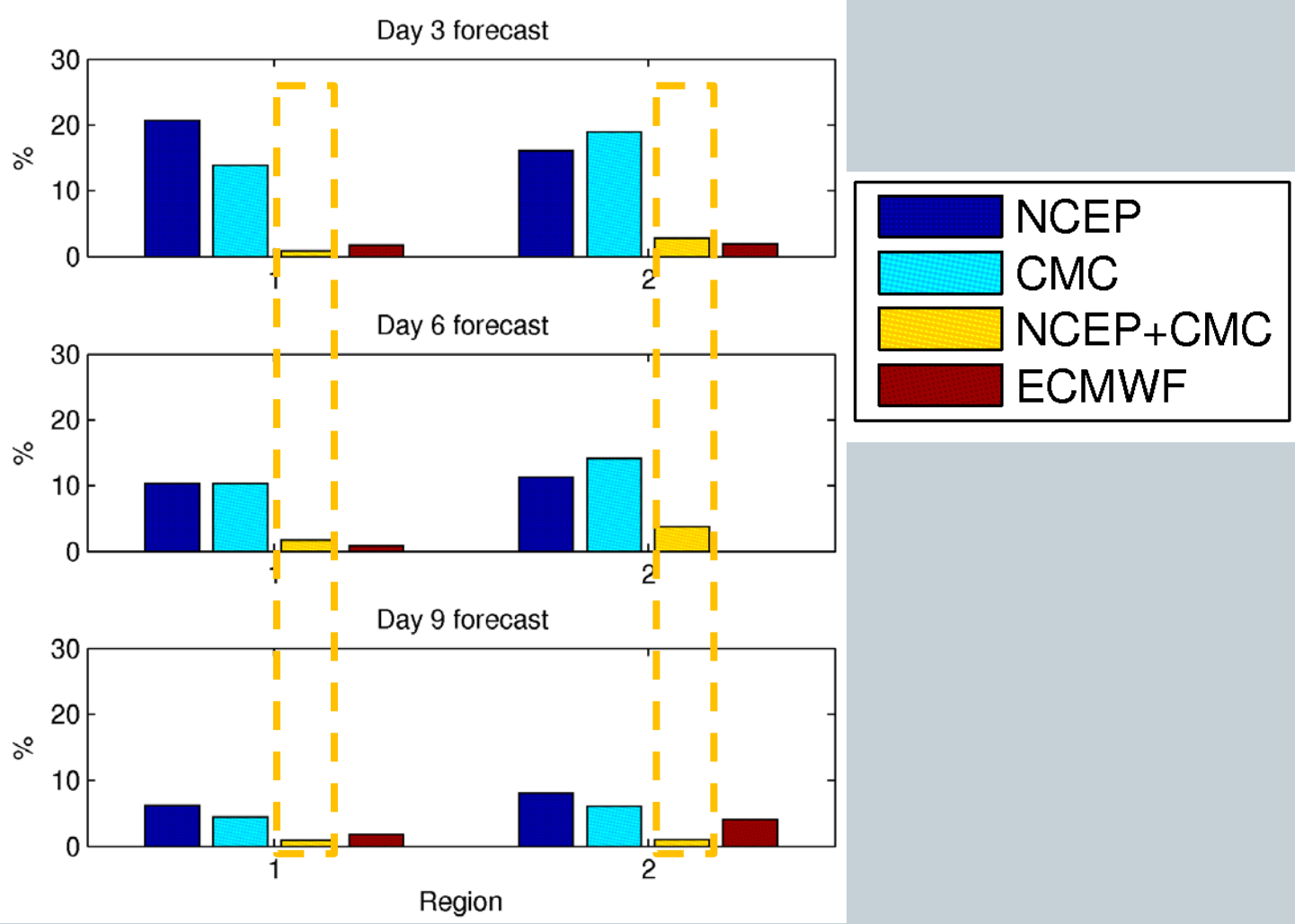


Historical evaluations using 124 (114) cyclone cases

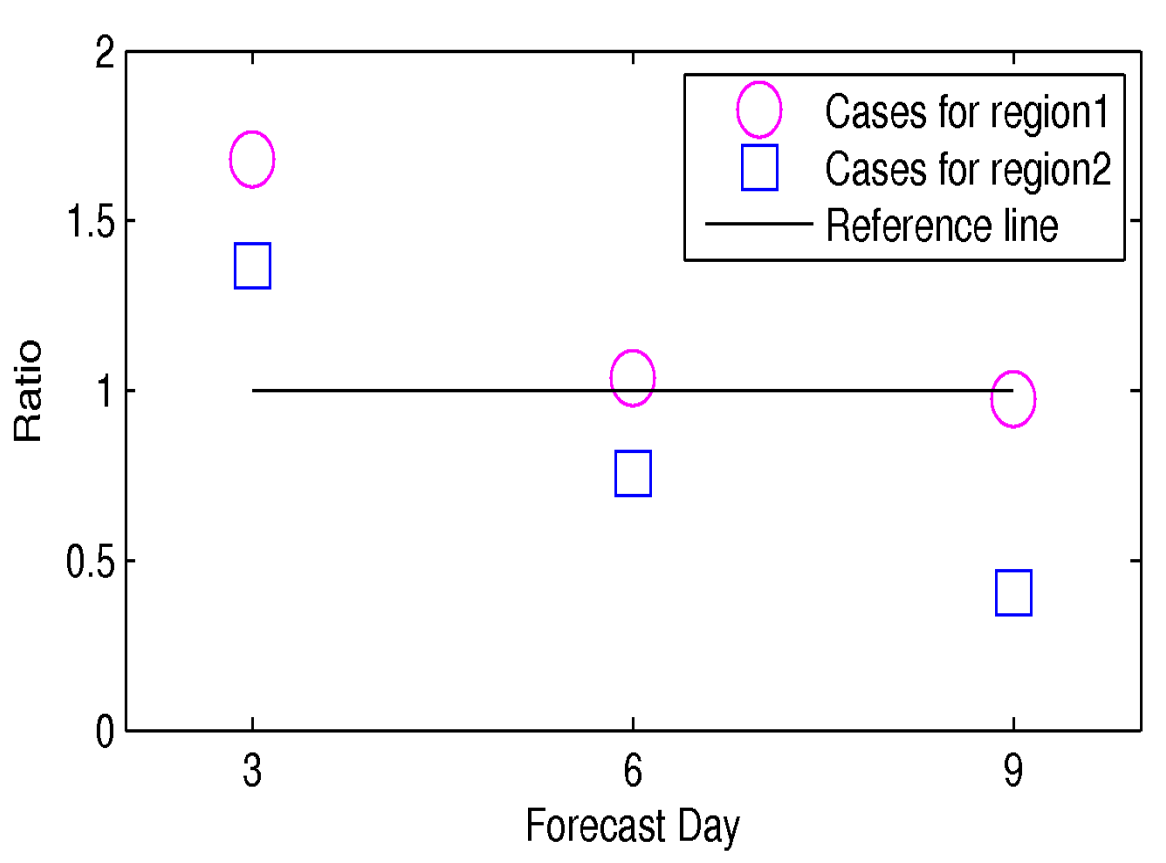
Percentage of each ensemble's members in Group ANA



Percentage of cases each model misses Group ANA



The ratio of the Analysis (ANA) in the Group EM group as compared to other cluster groups (for SLP cyclones) – no EM benefit after day 3



ANA more likely than average to be in Group EM

ANA less likely than average to be in Group EM

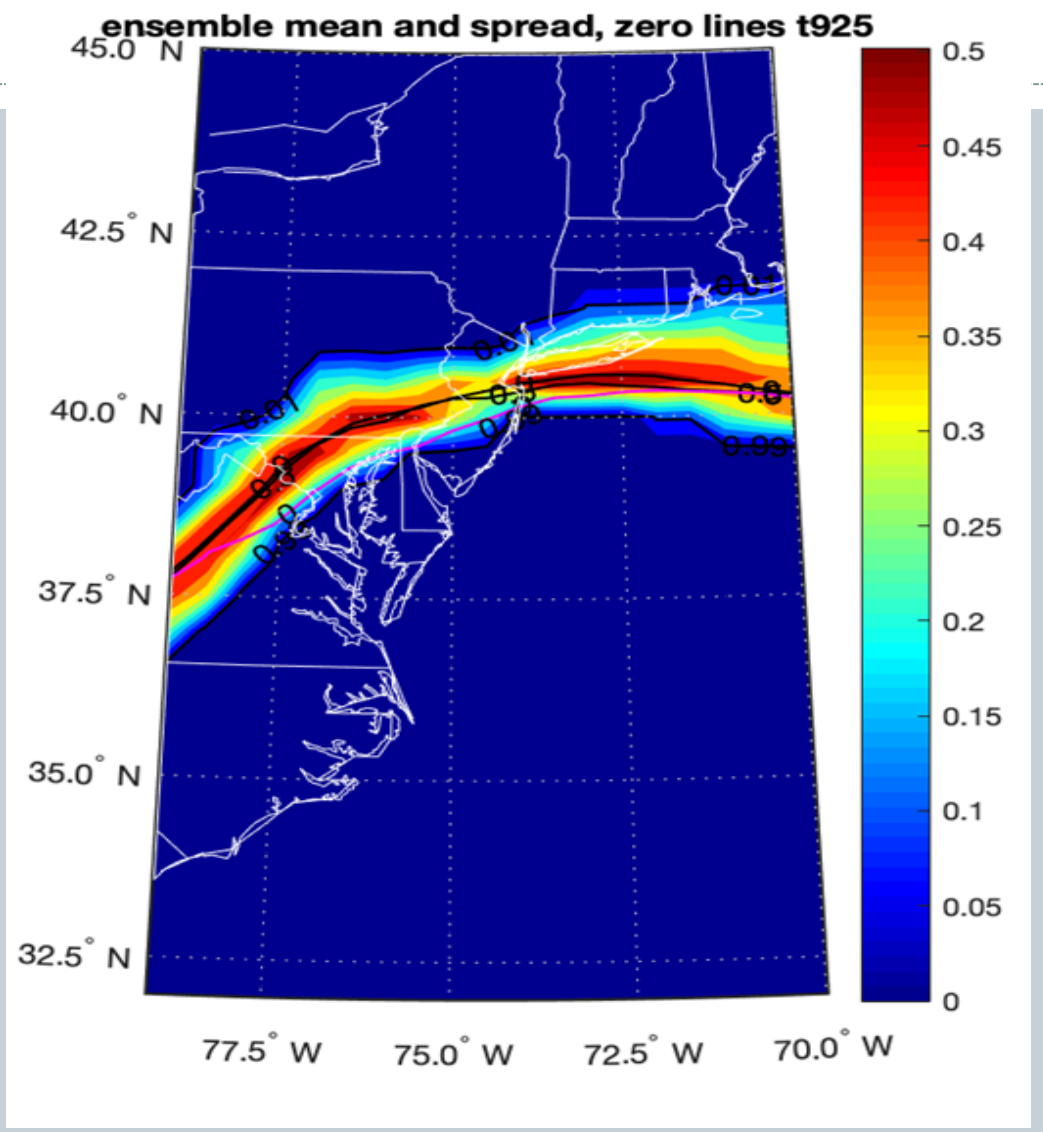
15 November 2018 NYC Snow "Surprise" (Evening Commute)

* This work is supported by NWS-CSTAR

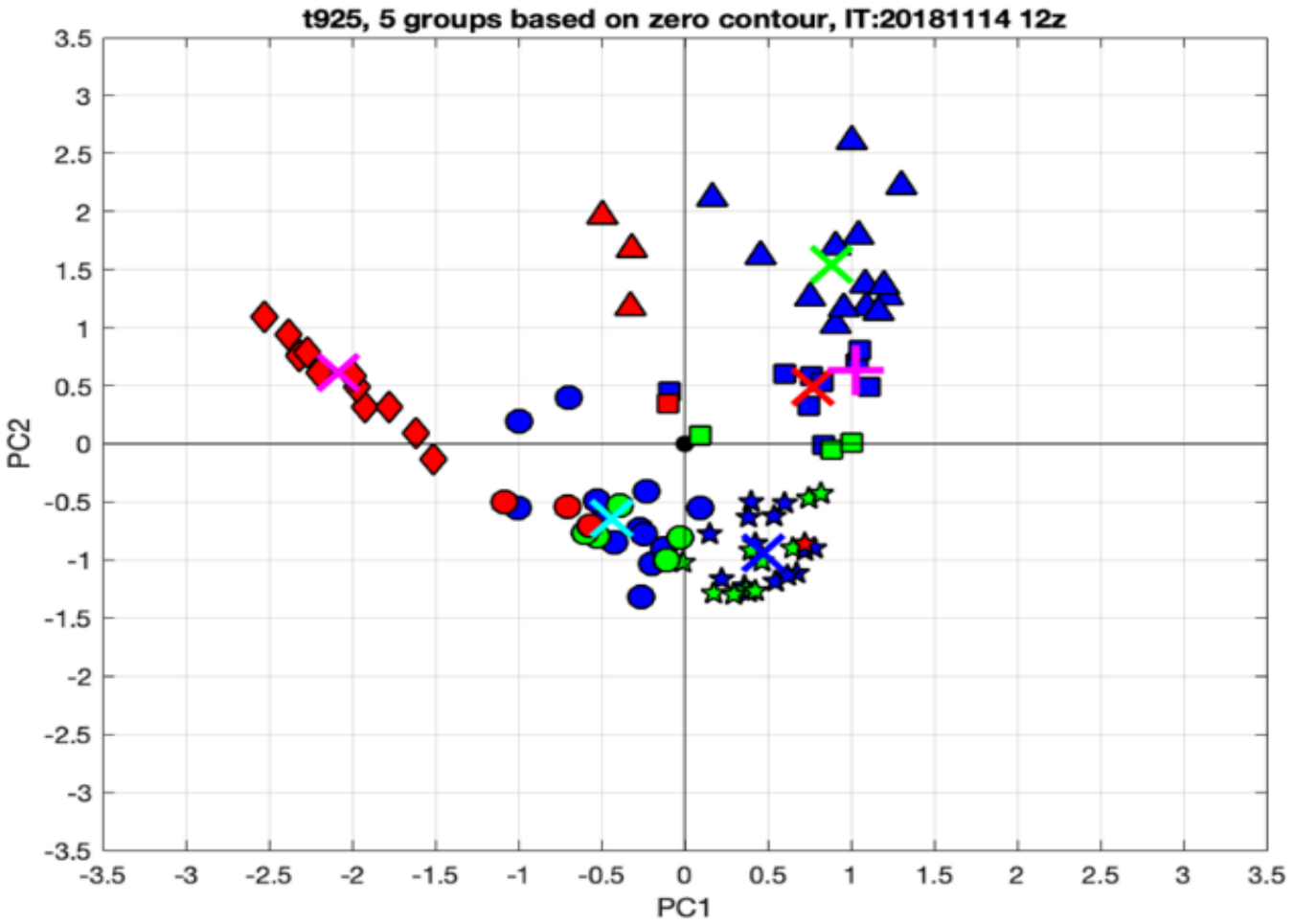
925-hPa Freezing Line Cluster

1.5 day Forecast

- temperatures larger than 0C are set to 1, while temperatures < 0C are set to zero.
- Ensemble mean and spread are calculated.
- First two EOFs are calculated.
- Clusters are determined using PC1 and PC2 phase space.



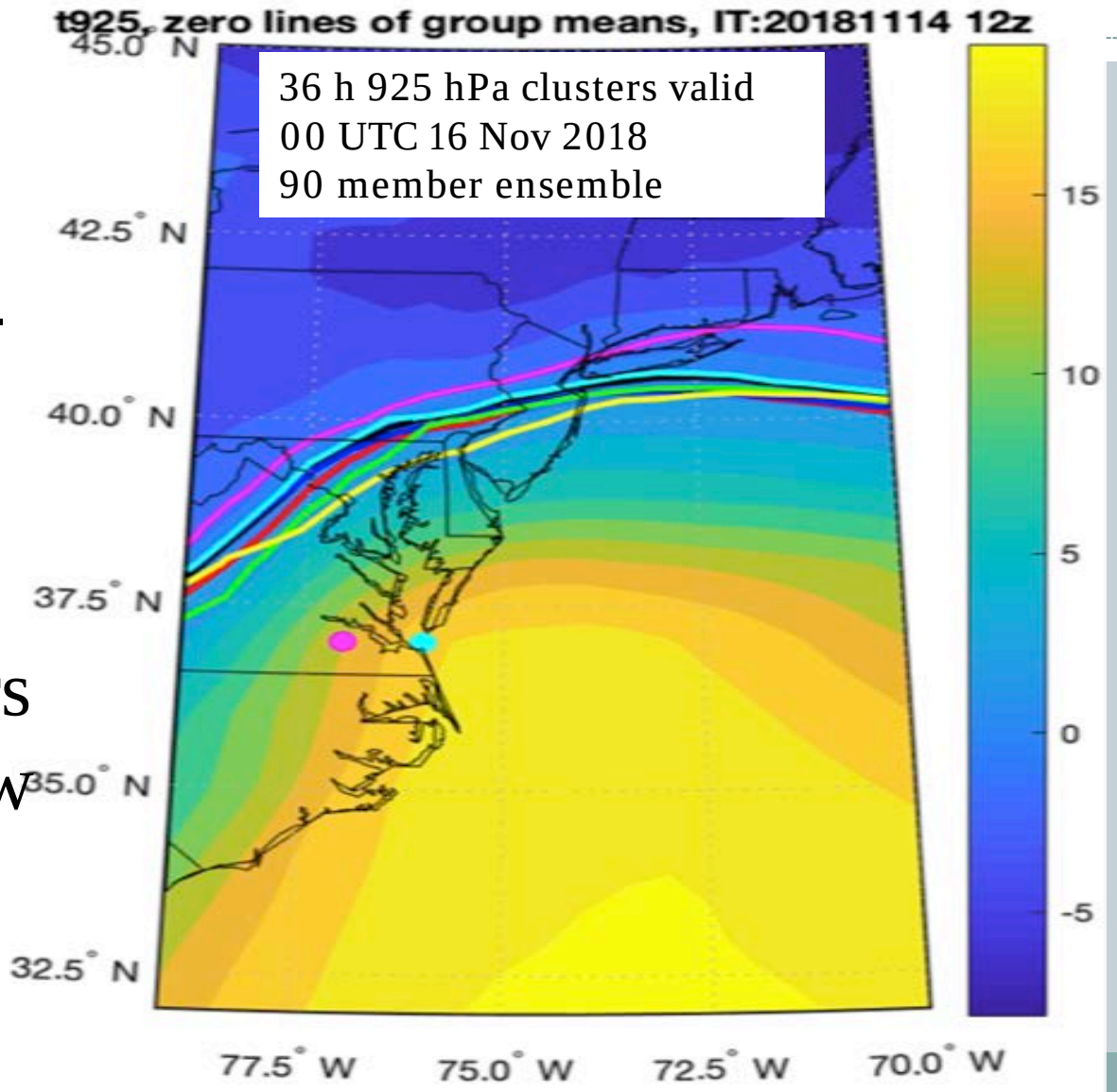
5 Clusters (X) on PC1-2 Phase Space



blue: ECMWF; green: NCEP; red: CMC

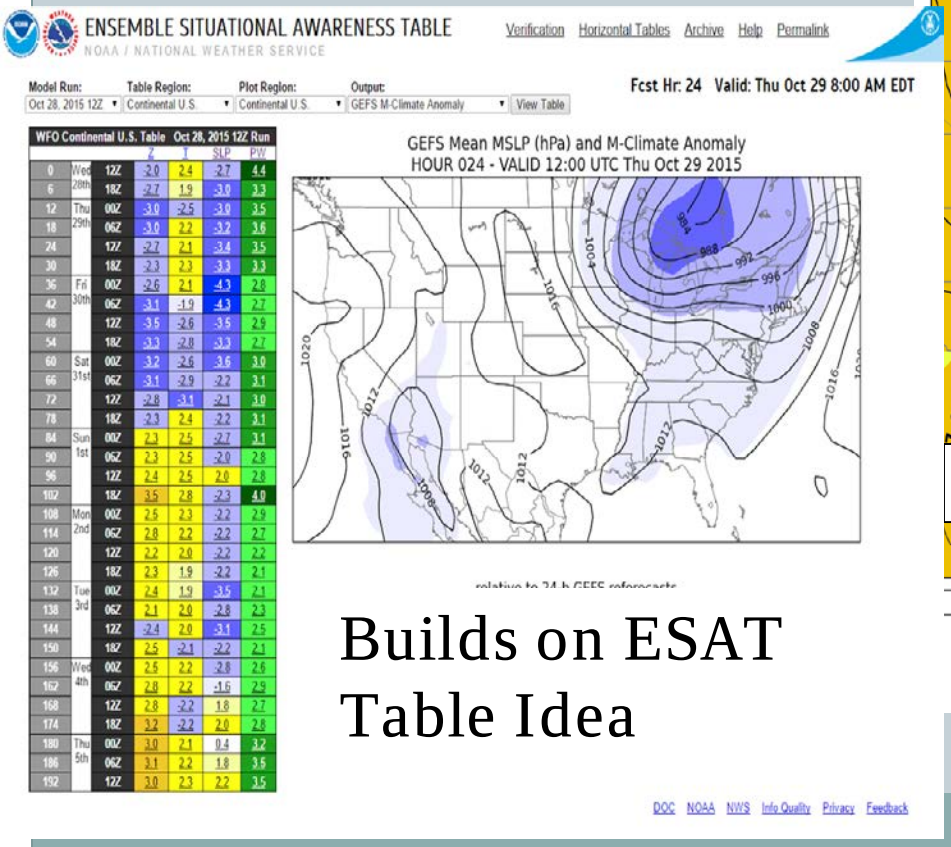
Each Freezing Line Cluster is a Scenario

- There was one warm cluster suggesting change-over to rain by rush-hour
- A few other clusters suggested still snow around rush-hour

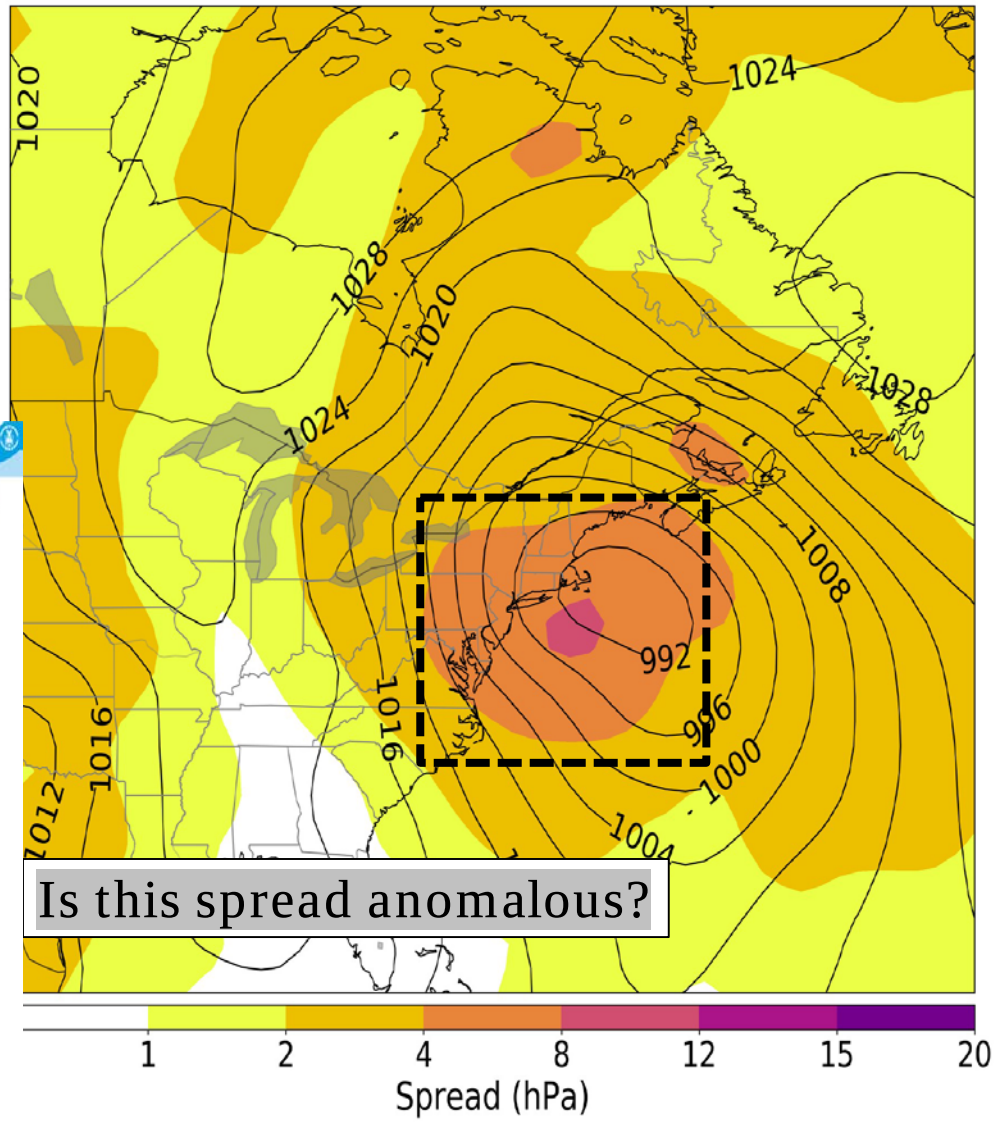


Motivation for Spread Anomaly Tool

- Allow forecasters to relate ensemble spread in context of previous forecasts of similar slp, 500Z,etc.. anomalies



Builds on ESAT Table Idea



Is this spread anomalous?

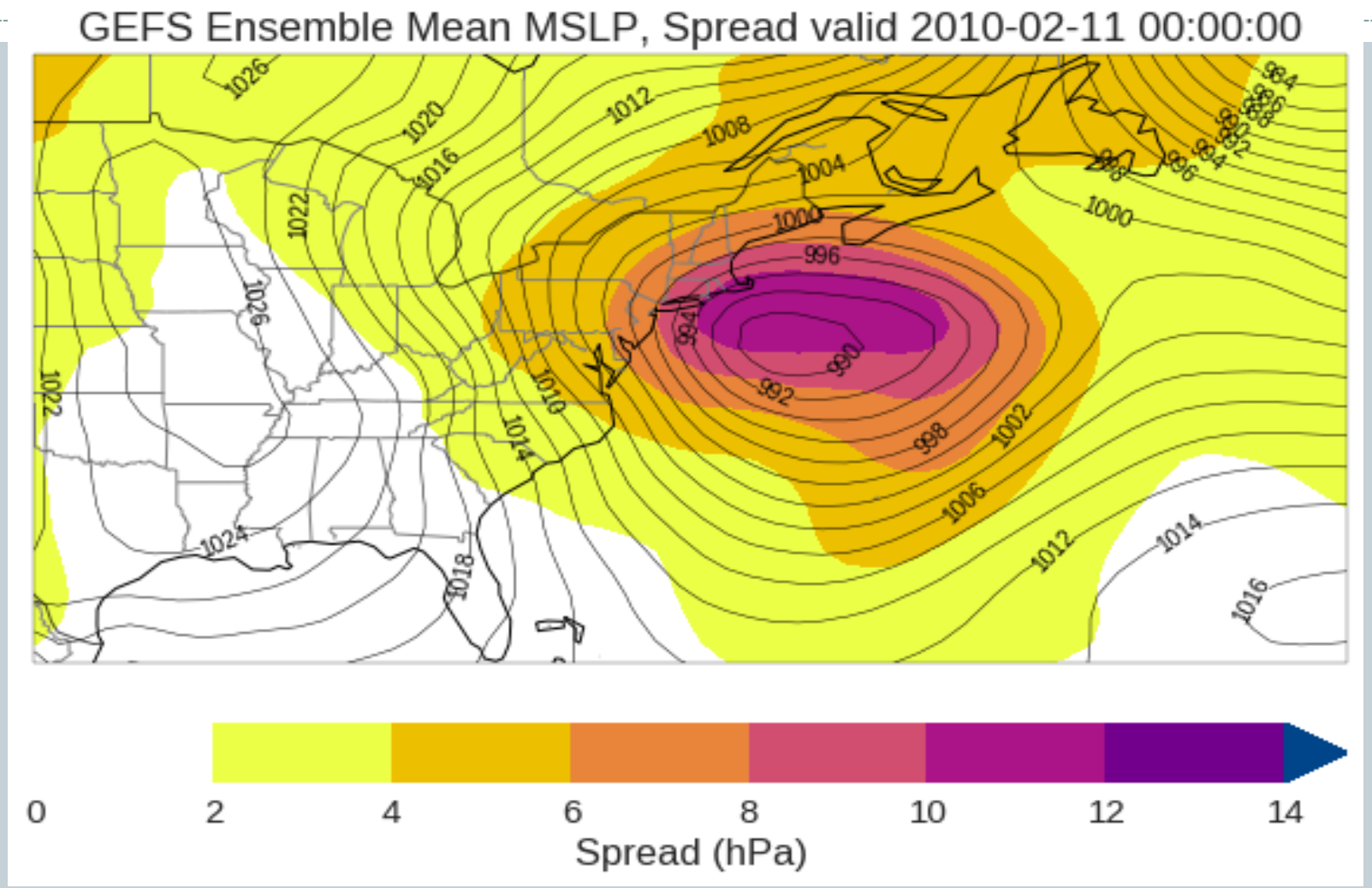
GEFS ensemble mean and spread forecast at hour 72, valid 26 Feb 2010 00z.

Brief Methodology

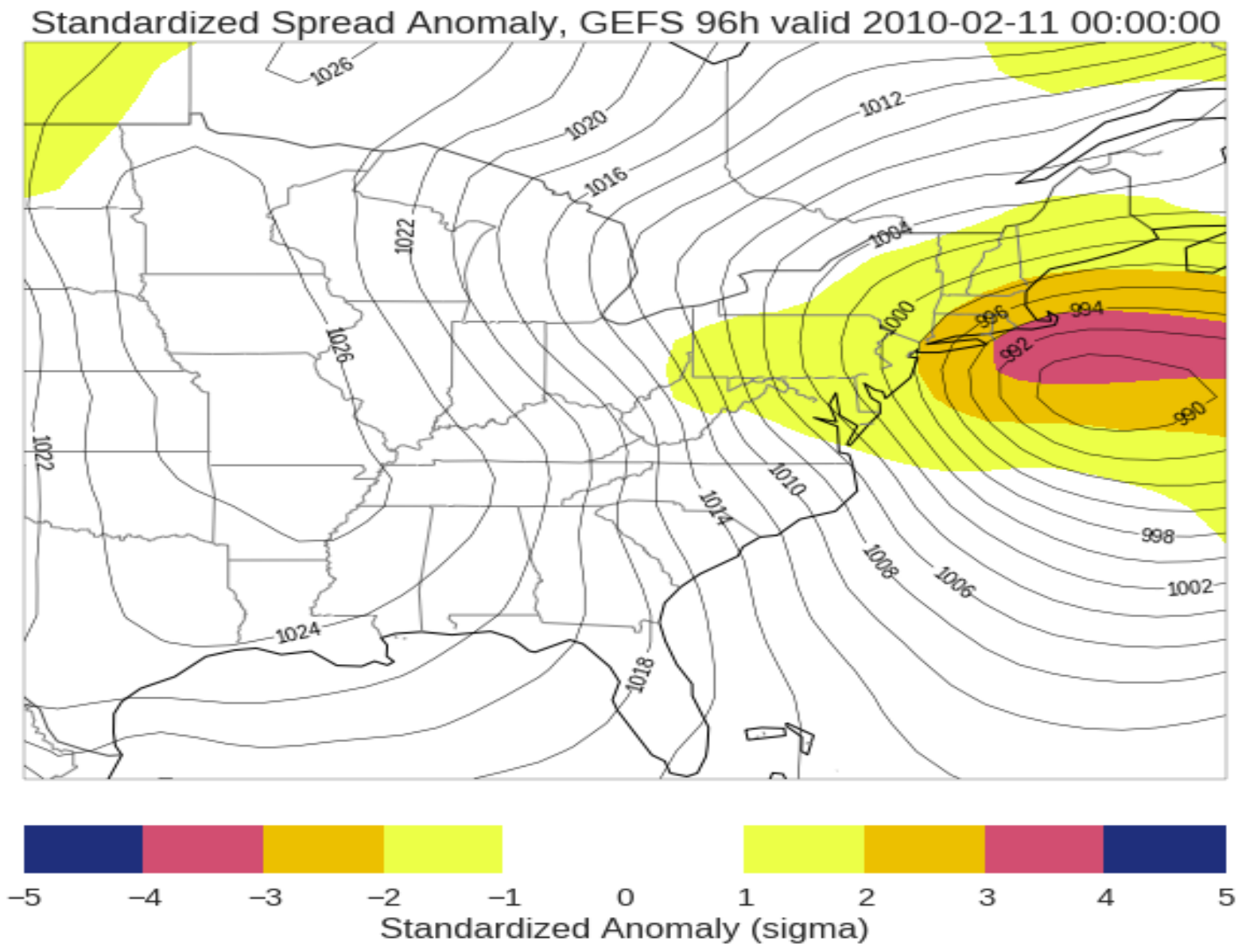
- The M-Climate data is generated for both ensemble mean and spread using GEFS Reforecast Ensemble.
- Data (for particular forecast hour) are taken from 21 days centered on the forecast day.
- The new distribution of SLP mean and spread is generated based on the +- 5% of the PDF centered on the SLP value of the grid point.
- The grid point's value is standardized based on the new subset distribution. It is transformed to a Gaussian distribution to generate standardized anomaly.

GEFS Mean/Spread (day 4)

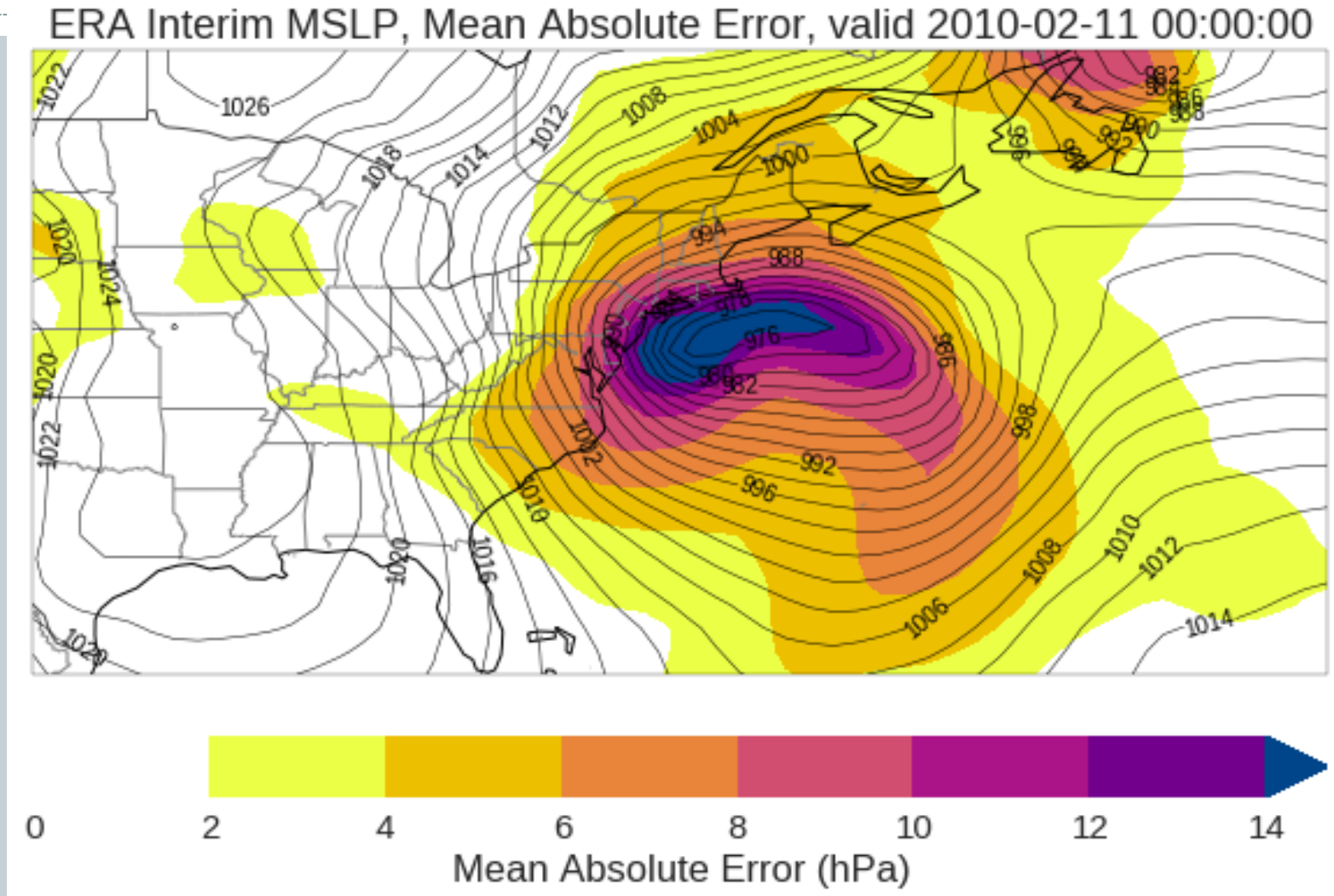
Valid: 0000 UTC 11 February 2010



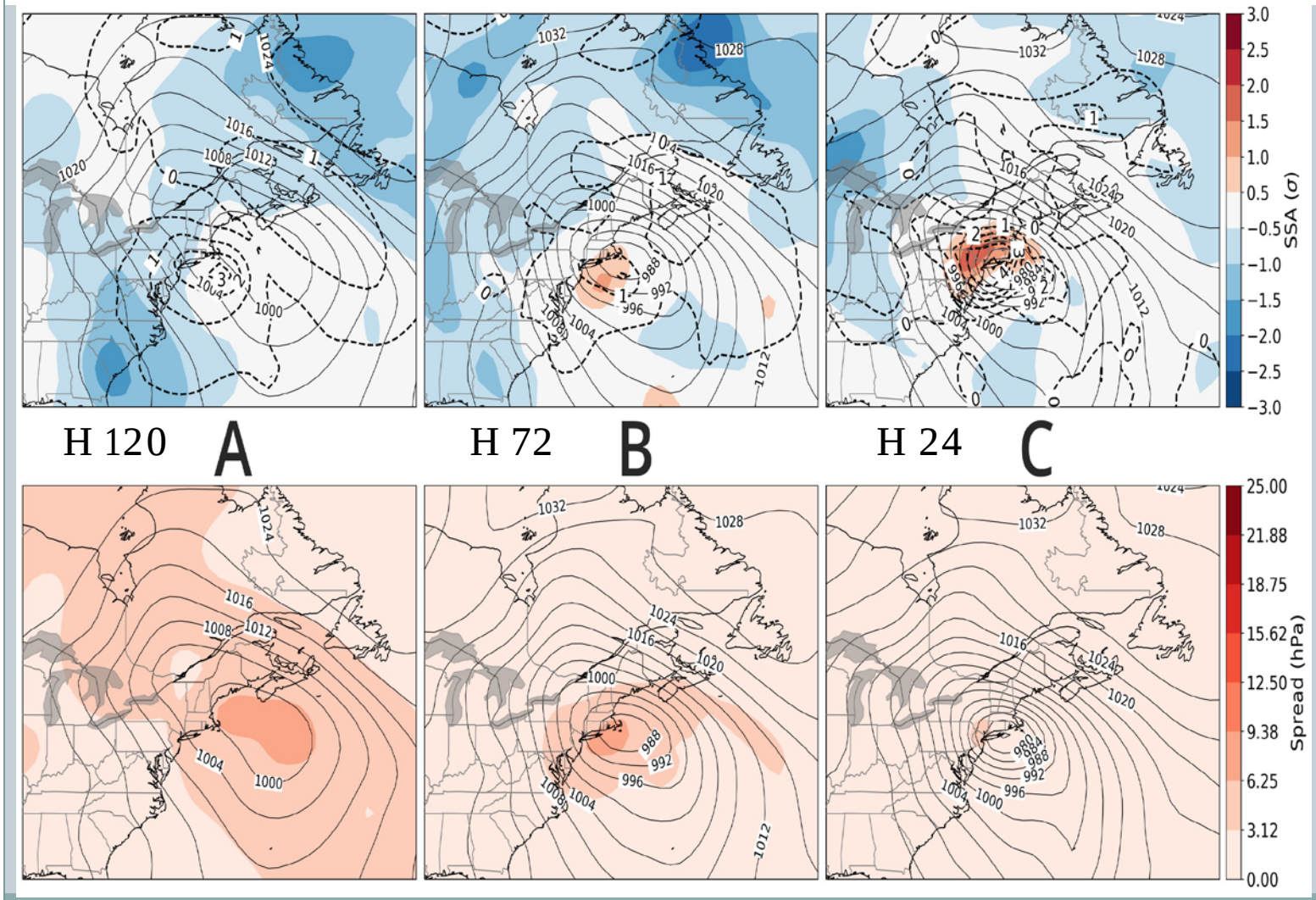
Standardized Spread Anomaly



Mean Absolute Error (hPa)



Sample Case (Valid 00 UTC 26 Feb 2010)



- Hours 120 (A), 72 (B), and 24 (C) valid 26 Feb 2010 00z.
- Shaded top is SSA, bottom is GEFS ensemble spread. Solid contours are MSLP, dashed contours on top are standardized MAE.
- Easier to denote the spread anomaly with SSA other a wide area which corresponds well with the MAE.

Conclusions

Stony Brook University
School *of* Marine and Atmospheric Sciences

- Tools have been developed to help distill ensemble information. This is only our effort. There are many others from Spring Experiment, WPC Winter Weather Expt, etc...).
- Fuzzy clustering can help generate scenarios to help with forecaster understanding and communication.
- Fuzzy clustering can also be used to validate ensembles (e.g., EC ensemble best in medium range, but it advantage for days 7-10 is less clear for E Coast storms.
- A spread-anomaly tool can help understand the uncertainty relative to other similar days.

- Some Challenges:
  - Ensembles are still underdispersed on extreme weather days.
  - How to use these tools in the forecast and communication process?

* Our Attempt to Address: Communication Uncertainty Workshops (March 2019, Nov 2019)

❖ *This research is supported by* ***NOAA CSTAR program***.

CSTAR Workshop with the Alan Alda Center for Communicating Science at Stony Brook – Mar 2019

Better Use of Ensembles in the Forecast Process: Scenario-Based Tools for Predictability Studies and Hazardous Weather Communication

Whats New? Click Here