Ensemble Predictability of Week 3 to 4 Precipitation and Temperature over the United States via Cluster Analysis of the Large-Scale Circulation.

Ensemble Predictability of Week 3/4 Precipitation and Temperature over the United States via Cluster Analysis of the Large-Scale Circulation

Gregory Jennrich,[a,c] David Straus,[b] Muthuvel Chelliah,[c] Cory Baggett,[c]

[a] *Earth Resources Technology Inc.* [b] *George Mason University, AOES Department* [c] *NOAA Climate Prediction Center*

Symposium on the Weather, Water, and Climate Enterprise
AMS 2024

Background

It has been recognized for some time that forecast models are better at predicting the large-scale flow than at predicting surface variables such as temperature and precipitation.

There are good reasons for this:
- The large-scale flow is explicitly resolved
- The large-scale flow responds to long-lived tropical forcing (e.g. ENSO)
- Precipitation in particular is generated by suspect parameterizations

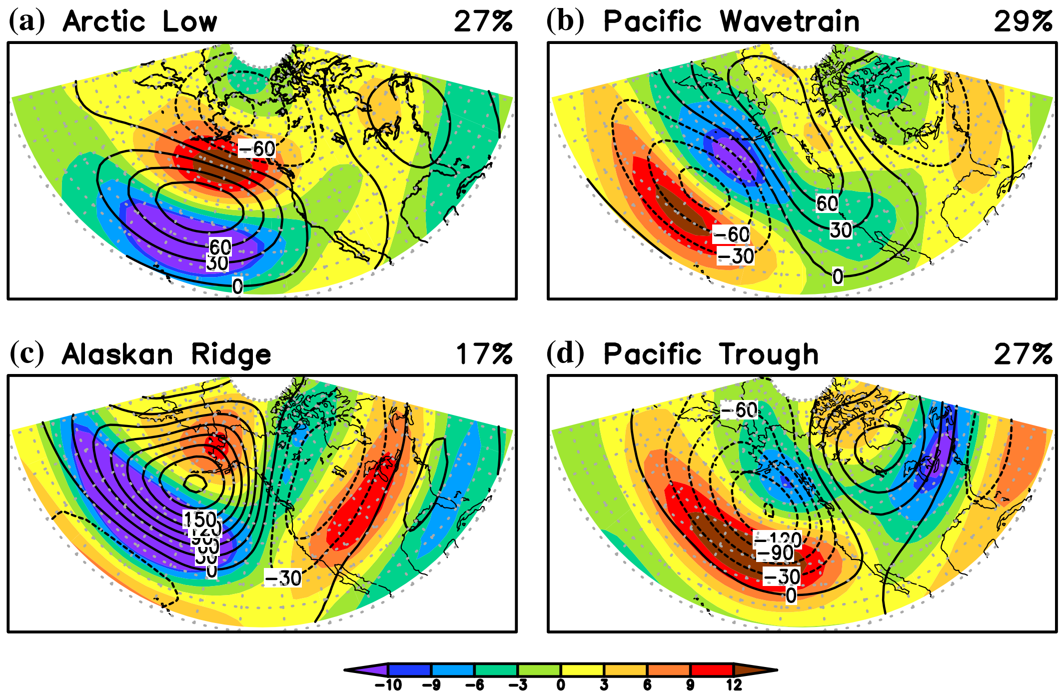
Can we leverage the prediction of the large-scale flow to improve predictions of surface variables?

The idea is to assign the Z500 forecast to one of a set of observed circulation regimes (or characteristic patterns), and then predict the surface variables using the observed association with that regime.

Example of extreme weather associated with circulation regimes in the Pacific North-America region

Amini and Straus (2019)* one of many, many examples in the literature

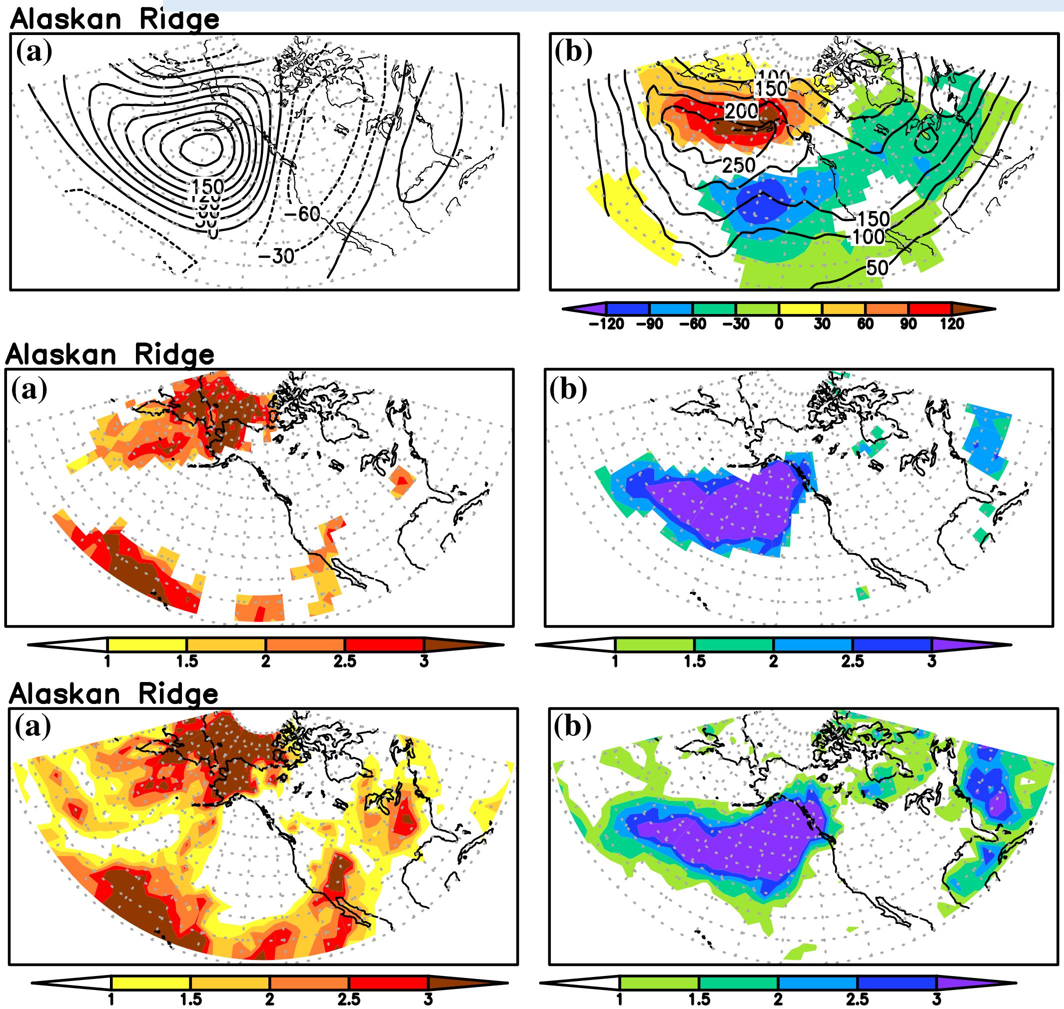
4 circulation regimes of Z500/U250 using the k-means clustering algorithm



Composites of Z500 (contours) and U250 (shading) for four clusters from k-means

*Amini, S., Straus, D.M. Control of Storminess over the Pacific and North America by Circulation Regimes. *Clim Dyn* **52**, 4749–4770 (2019). https://doi.org/10.1007/s00382-018-4409-7

Example of Extreme Weather Associated with one circulation regime



(a) Z500 for Alaskan Ridge Regime

(b) Composite of upper–level storm track indicator (variance of $v_{highpass}$)
Colors show anomalous storm tracks

(a) Ratio of number of Alaskan Ridge days for which precip is in top 5th percentile to the number expected based only on the number of days in this regime
(b) Same ratio for bottom 5th percentile.

Same as above, but for more persistent episodes of Alaskan Ridge (> 14 days of duration)

Application
to CPC
week 3 – 4
Forecasts

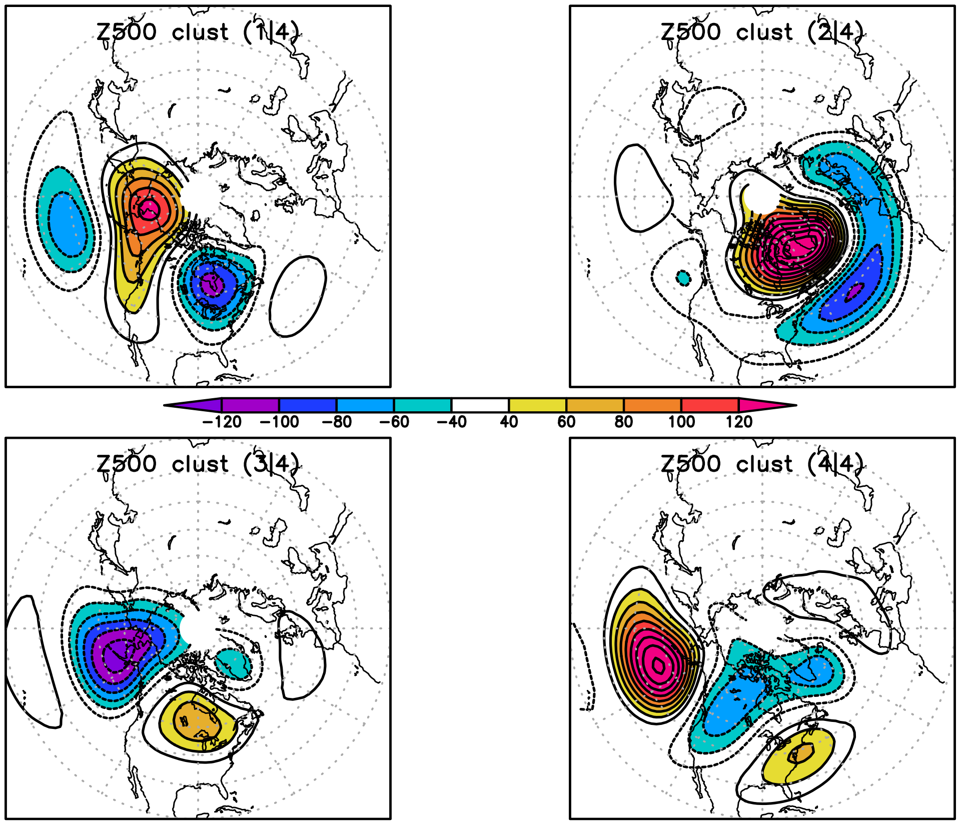
Ensemble Predictability of Week 3/4 Precipitation and Temperature over the United States via Cluster Analysis of the Large-Scale Circulation

Application to "week 3 to 4" (two-week) boreal winter operational forecasts at CPC

Circulation Regimes based on running 14-day means in boreal winter
Applicable to "week 3 to 4" (two-week) operational forecasts at CPC

4 circulation regimes of Z500 using the k-means clustering algorithm

Note the wider domain



How many clusters (regimes) should we use?

No. of Clusters	Confidence Level	S. score
2	82.0%	0.1277
3	88.0%	0.1084
4	100.0%	0.1102
5	100.0%	0.1069
6	100.0%	0.0986
7	100.0%	0.0950
8	100.0%	0.0891

Confidence level is based on testing the significance of the clustering with respect to synthetic data sets. (Higher is better).

Silhouette score is another measure of the tightness of the clusters (higher is better)

How many clusters (regimes) should we use?

No. of Clusters	Confidence Level	S. score
2	82.0%	0.1277
3	88.0%	0.1084
4	100.0%	0.1102
5	100.0%	0.1069
6	100.0%	0.0986
7	100.0%	0.0950
8	100.0%	0.0891

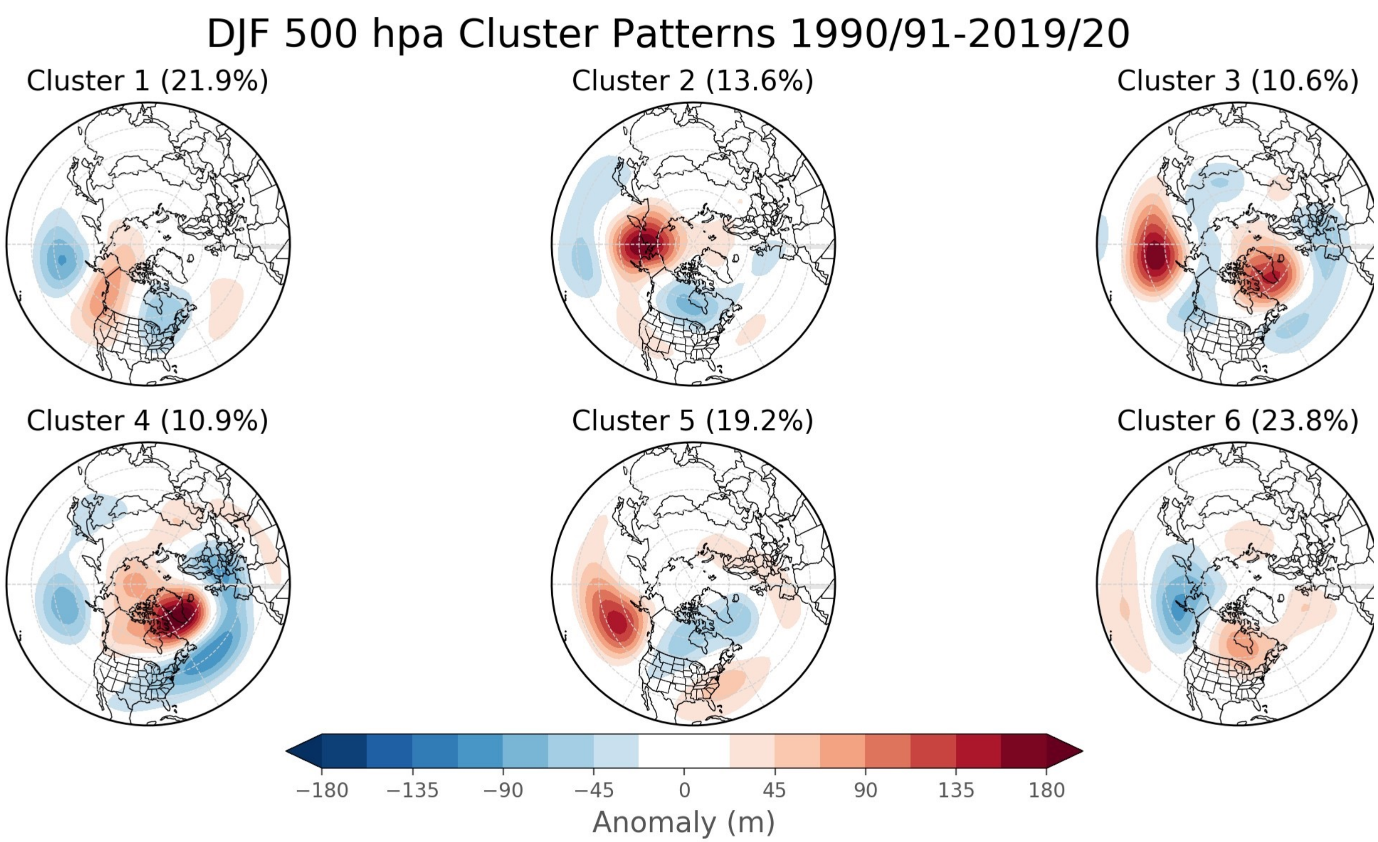
Confidence level indicates more than 3 clusters should be used.

S Score is maximum for 4 clusters, but drops off only a little for $k > 4$

Higher k (more circulation regimes) means better resolution of forecasts in terms of clusters.

Don't know of a statistical measure of this!

ERA5 Patterns for 6 clusters (6 regimes)

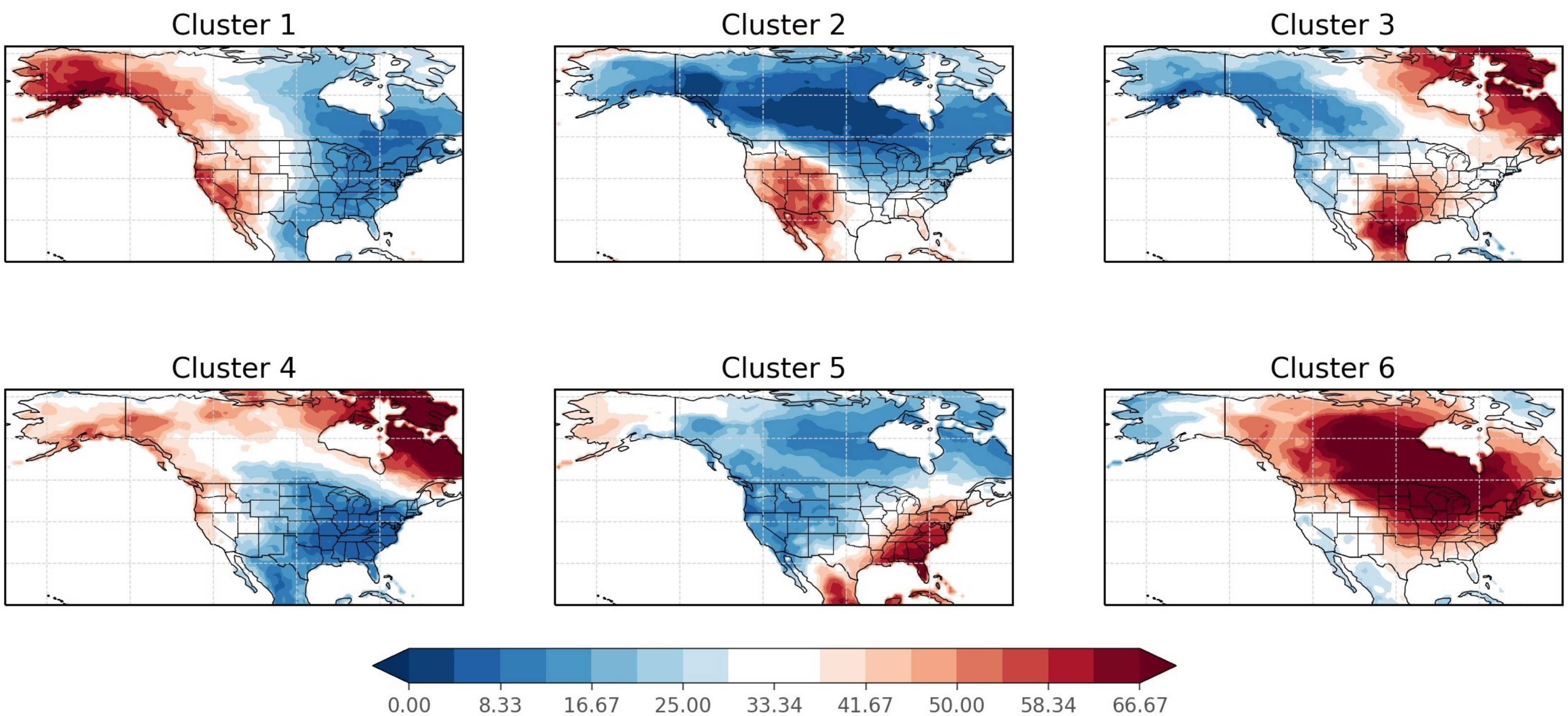


Extreme Weather Cluster Composites

- For each cluster we can find the associated anomaly composites for any variable. We first investigate temperature and precipitation terciles.
- Data:
  - Temperature: ERA5 Daily Reanalysis
  - Precipitation: ERA5 Daily Reanalysis
  - 14-day running mean anomalies (temperature) or sum anomalies (precipitation) to match cluster periods
- Method:
  - Calculate terciles (33^{rd} and 67^{th}) for each running 14-day period
  - Smooth terciles (3^{rd} Harmonic smoothing)
  - Each period now can be classified as above, near, or below normal
  - For each of the 6 clusters, calculate the occurrence of each tercile
  - For example, a given point for Cluster 1 may have 70% occurrence of above, 20% near, and 10% below normal temperatures

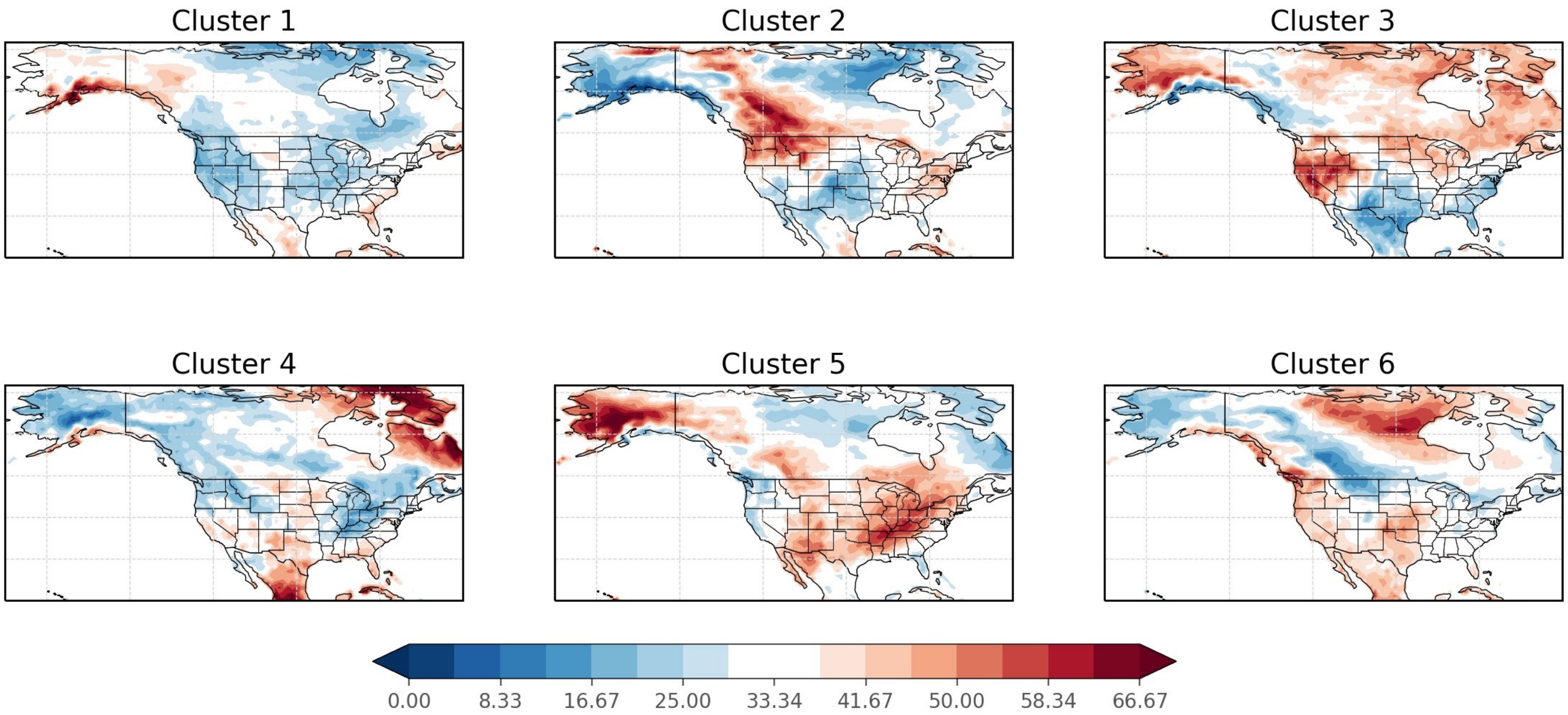
ERA5 Patterns for 6 clusters (6 regimes)

Probability of Above Normal **Temperature**



ERA5 Patterns for 6 clusters (6 regimes)

Probability of Above Normal **Precipitation**



Cluster-Based Forecasts:

Determine which cluster(s) the forecast is like.

Then use observed temperature, precip. for that cluster (those clusters)

- **Ensemble Mean Approach:** Assign a cluster to the forecast ensemble mean based on minimum RMS error.
- Then use the observed weather composites to assign a tercile for Temperature and Precipitation

- **Member Weighted Approach:** The entire ensemble is given a weighted probability for each cluster, with the weights given by the inverse RMS error.
- The weather composites are made accordingly .

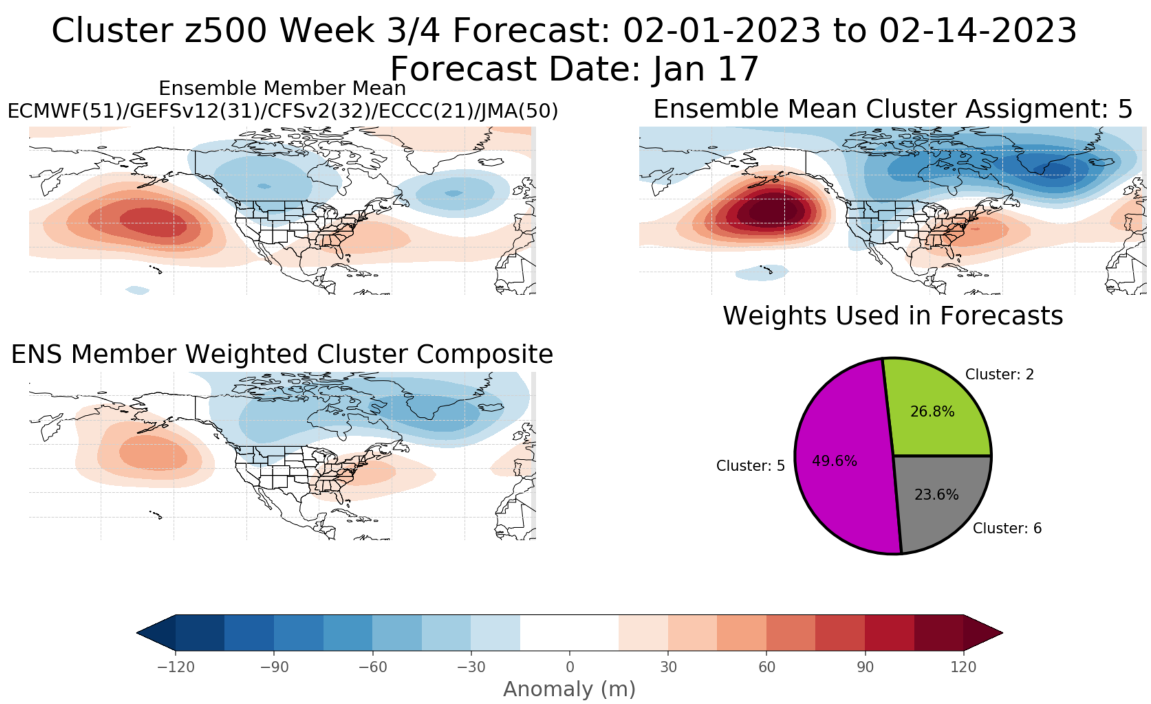
- Data:
  - GEFSv12 hindcast z500 forecasts
  - 11 total ensemble members
  - Weekly Initialized between 11/15-2/15
  - For years 2000-2019
  - 252 samples

$$Inverse\ RMSE_i = \frac{1}{\sqrt{\sum_{n=1}^{n} \frac{\left(y_{pred} - y_{obs}\right)^2}{2}}}$$

$$W_i = \frac{Inverse\ RMSE_i}{\sum_i^{11} Inverse\ RMSE}$$

Ensemble Mean two-week forecast Z500

Forecast Z500 based on ensemble weighting approach



Cluster composite assigned to ensemble mean

Weighting of the clusters

Sample forecast for week 3 to 4 period (February 1-14, 2023) initialized (January 17, 2023)

Z500 and US Temperature Skill Scores for Hindcasts

Hindcast Skill Scores

Forecast Method	**K**	**z500**	**Temperature**	
			All Samples	**Cluster-Like**
Member Weighted	k=4	.223	14.0	**23.0 (70)***
	k=5	.228	13.7	**24.7 (77)***
	k=6	.223	12.7	**22.2 (84)**
	k=7	.244	15.9	**24.4 (94)***
	k=8	.248	14.5	20.6 (101)
Ensemble Mean	k=4	.205	12.6	**24.2 (70)***
	k=5	.200	13.1	**26.1 (77)***
	k=6	.210	11.3	**20.7 (84)**
	k=7	.207	12.0	**22.9 (94)***
	k=8	.209	11.6	18.3 (101)
GEFSv12	-	.291	17.6	19.1 (84)

Bold scores are significantly greater than the all-sample score.

Starred(*) scores denote a significant skill score increase for cluster-like initialization dates

(1) The z500 skill is scored via pattern correlation with ERA5 observations.
(2) Heidke skill scores (HSS) are used for temperature and precipitation scoring.
(3) **Cluster-like samples are based on the criteria that the ensemble mean GEFSv12 z500 forecast pattern has a correlation with one of the cluster composite z500 fields exceeding 0.7**

Z500 and US Temperature Skill Scores for Hindcasts

Hindcast Skill Scores

Forecast Method	K	z500	Temperature	
			All Samples	Cluster-Like
Member Weighted	k=4	.223	14.0	**23.0 (70)***
	k=5	.228	13.7	**24.7 (77)***
	k=6	.223	12.7	**22.2 (84)**
	k=7	.244	15.9	**24.4 (94)***
	k=8	.248	14.5	20.6 (101)
Ensemble Mean	k=4	.205	12.6	**24.2 (70)***
	k=5	.200	13.1	**26.1 (77)***
	k=6	.210	11.3	**20.7 (84)**
	k=7	.207	12.0	**22.9 (94)***
	k=8	.209	11.6	18.3 (101)
GEFSv12	-	.291	17.6	19.1 (84)

Bold scores are significantly greater than the all-sample score.

Starred(*) scores denote a significant skill score increase for cluster-like initialization dates

Z500 and US temperature scores for hindcasts are higher than that for cluster-based forecasts

Cluster-based forecasts have significantly higher Temperature scores for cluster-like forecasts than the model !!

Model scores for temperature increase significantly when the initialization is cluster-like !!

Some Encouraging Conclusions

Cluster-based forecasts have significantly higher Temperature scores for cluster-like forecasts than the model !!

Model scores for temperature increase significantly when the initialization is cluster-like !!

These results suggest it may be possible to assess the accuracy of the prediction at the time of the forecast!

Some Less Encouraging Conclusions

Improvement for cluster-like forecasts doesn't hold for precipitation

Ensemble mean Z500 forecasts are better than cluster-based forecasts.