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

or:

Predictability of weather on subseasonal to seasonal time scales

David M. Straus[1]

Greg Jennrich[2], Muthuvel Chelliah[3], and Cory Baggett[1]

[1] George Mason University / COLA / AOES Department

[2] ERT, Inc.

[3] NOAA Climate Prediction Center

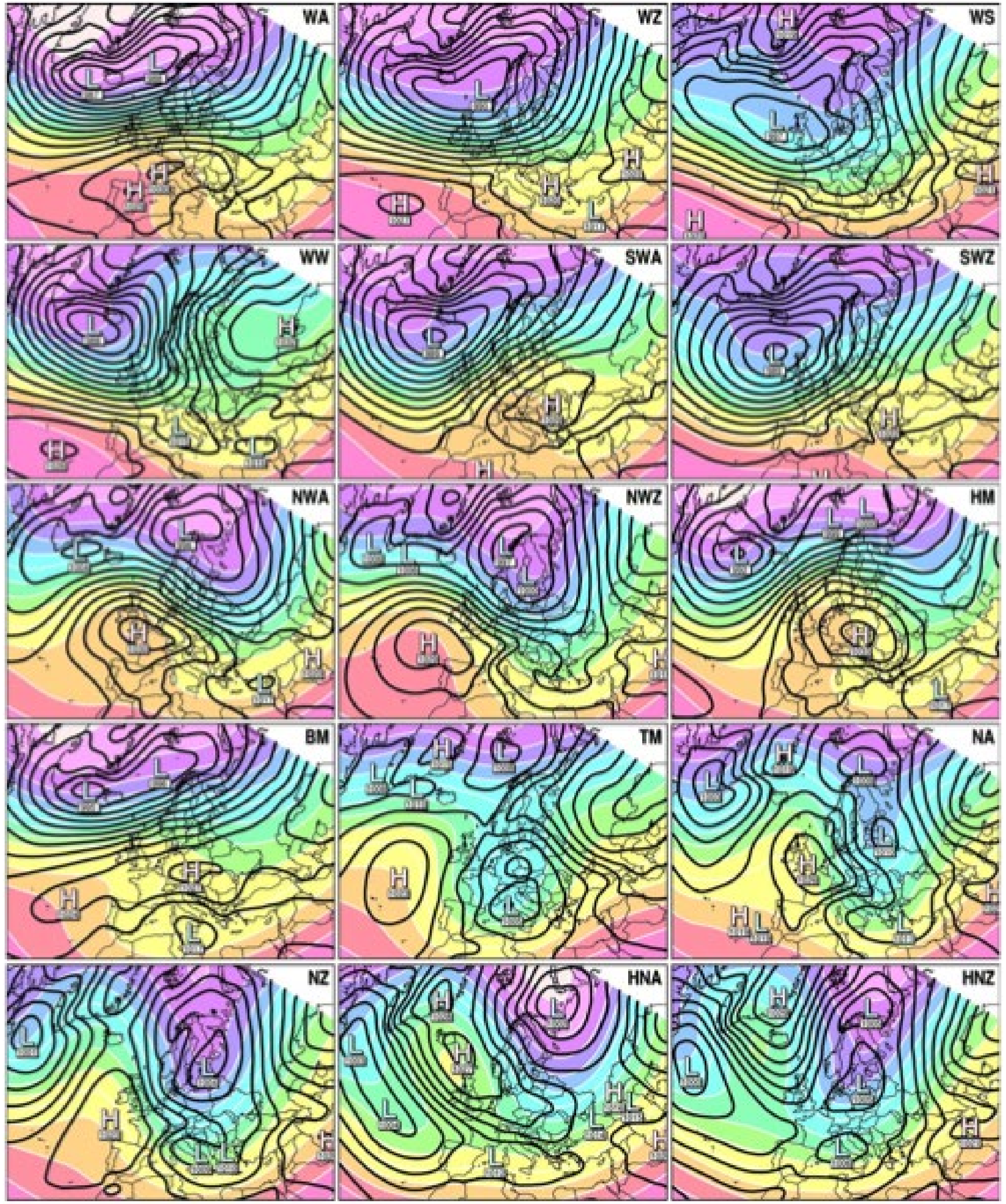
Predictability of (Extreme) weather on subseasonal to seasonal time scales

- *The Challenge*: Precipitation (and weather in general) is known to be poorly represented in models (reliance on uncertain parameterizations)

- The most predictable components of the atmosphere are the planetary wave components

- Observational results show how the planetary wave components are linked to the probability of extreme weather

- Can we use this control to better forecast the predictability of (extreme) weather on subseasonal to seasonal time scales?



Climate Dynamics (2006) 27: 215–231 DOI 10.1007/s00382-006-0133-9
P. M. James
“An assessment of European synoptic variability in Had
Centre Global Environmental models based on an
objective classification of weather regimes”

Categorizing weather regimes has been happening since the 1940s – the *Grosswetterlagen* (GWL) system of the German Weather Service, which has been updated (James, 2006).

Here we show 15 out of the 29 weather regimes of James for central Europe (SLP contour, Z500 color)

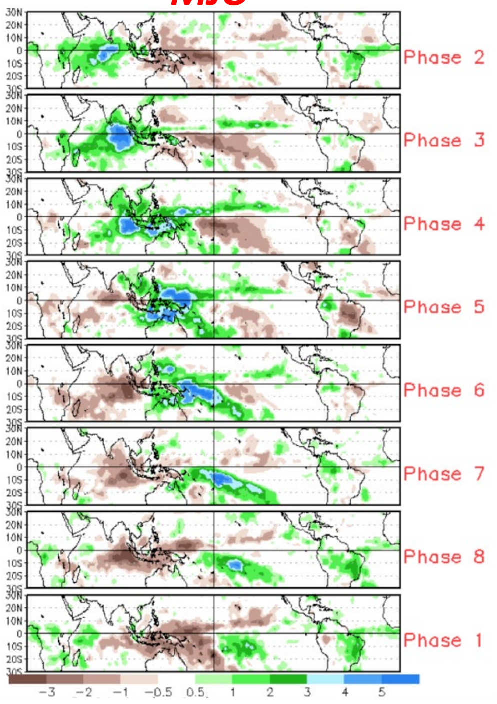
a Climatological composites of GWLs 1-15 for the winter half-year, showing Geopotential height at 500 hPa (colour-filled field, contour interval 6 dam) and MSLP (black contours, interval 3 hPa, with plotted central pressure values with High, Low centre symbols).

- Probably more ”Coarse-Graining” is needed. By widening the area and reducing the number of “regimes” we focus more on the planetary wave component environment.

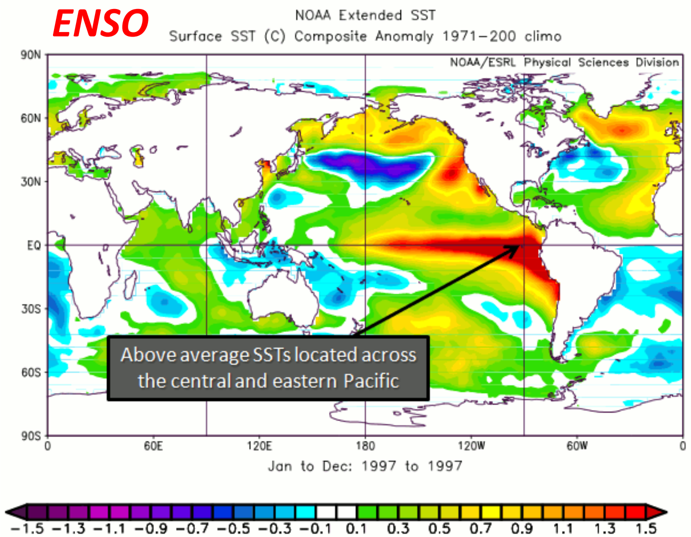
- The planetary wave component modulates the effect of tropical forcing on extreme weather.

Tropical Forcing

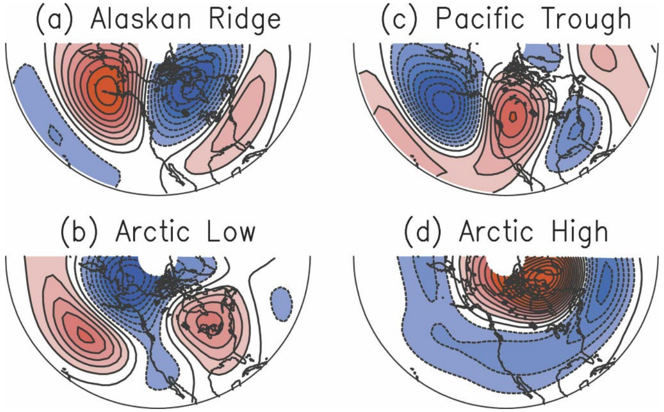
MJO



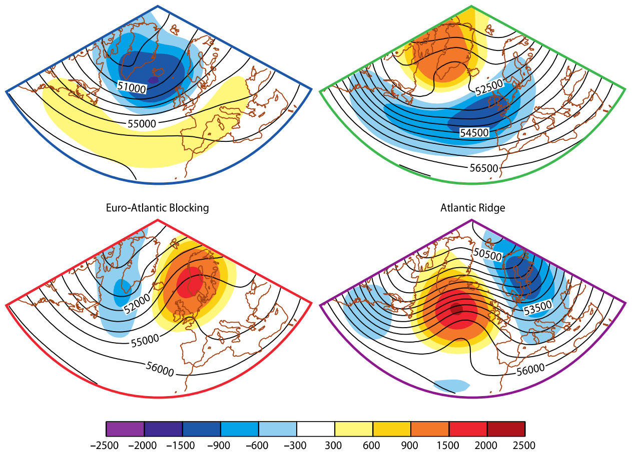
ENSO



Planetary Scale Preferred Regimes



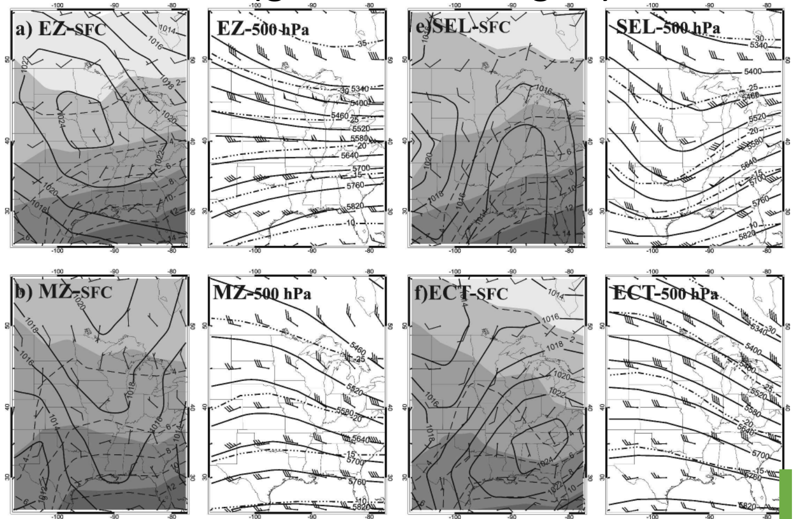
Regimes over **PNA region** (Riddle et al. 2013; Straus et al., 2007)



Regimes over **Euro-Atlantic region** (Michelangeli et al., 1995; Cassou, 2008; Ferranti and Corti, 2011)

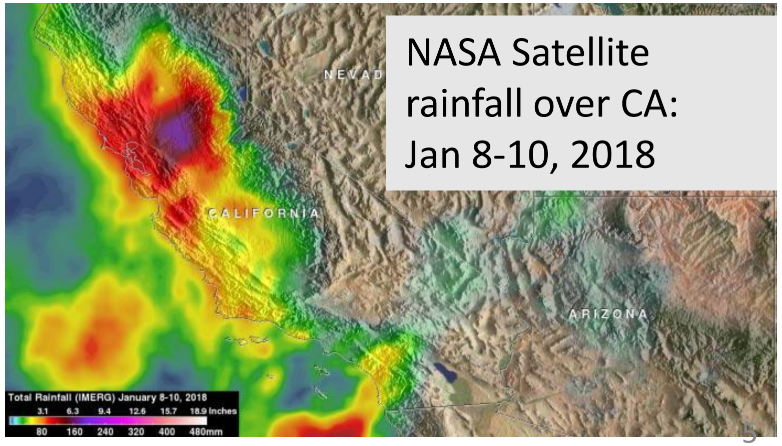
Regional Weather Types

(grosswetterlagen)



Four of the US weather types defined over the region 25°- 55° N, 7.5° - 107.5°W by Coleman and Rogers (2007)

Severe Weather



NASA Satellite rainfall over CA: Jan 8-10, 2018

Obtaining Large Scale Circulation Regimes via Cluster Analysis

Partitioning Method (k-means)

- The number of clusters (which is denoted by k) is determined a priori, after which all states are assigned uniquely to one of the k clusters, based on a similarity criterion between states.

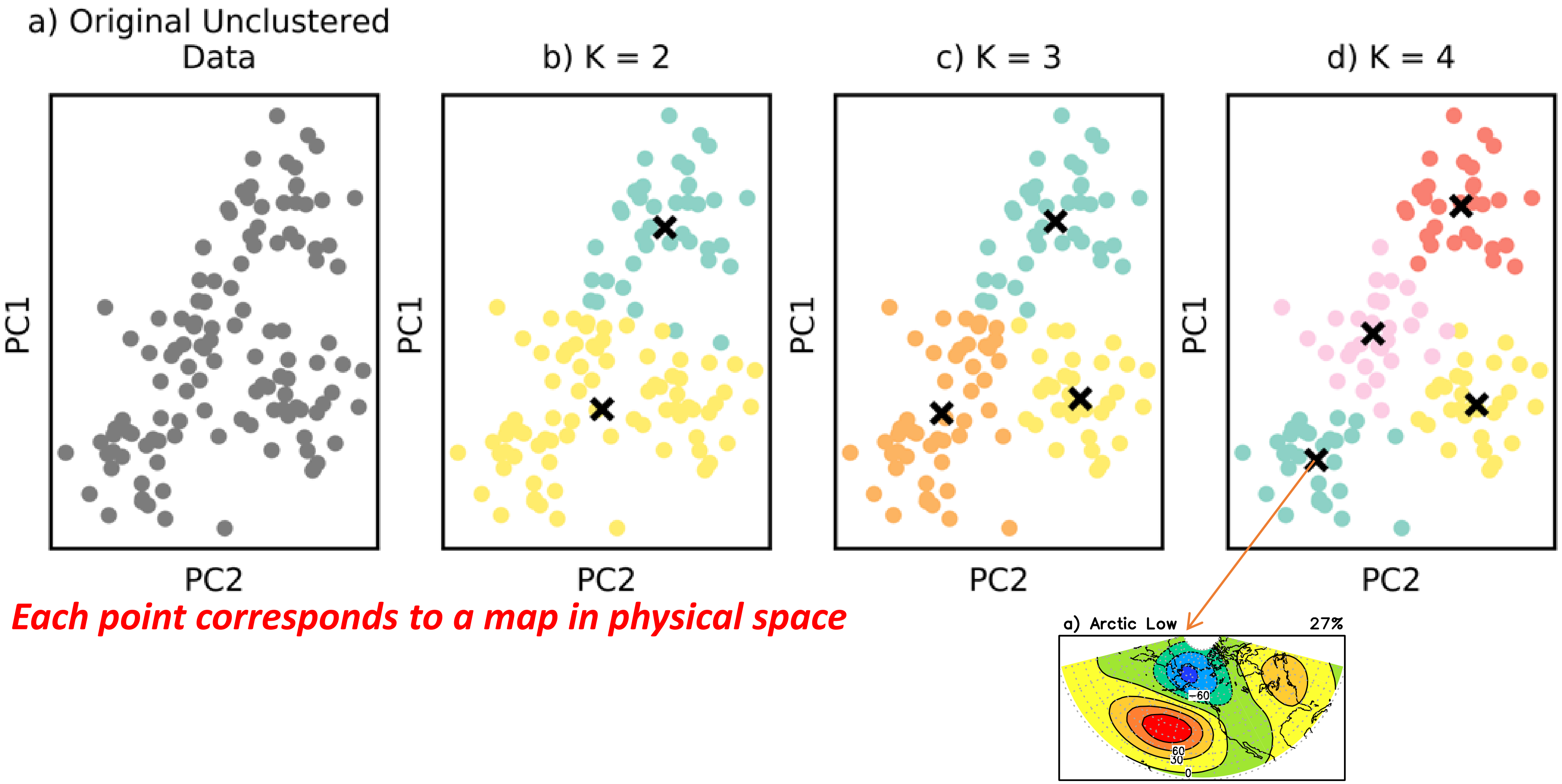
- The partitioning is almost carried out in a reduced dimensional space, via the use of Principal Component Analysis. The data vector (such as Z500 field) for each time is then represented as a point in the state space of the PCs, which provide the coordinates.

- The *cluster centroid* is then defined as the point whose coordinates are the average of the coordinates of all the points assigned to the cluster

- The strength of the clustering ("bunching") is measured by the variance ratio R: Variance of cluster centroids (weighted by number of states in a cluster) divided by the average within-cluster variance.

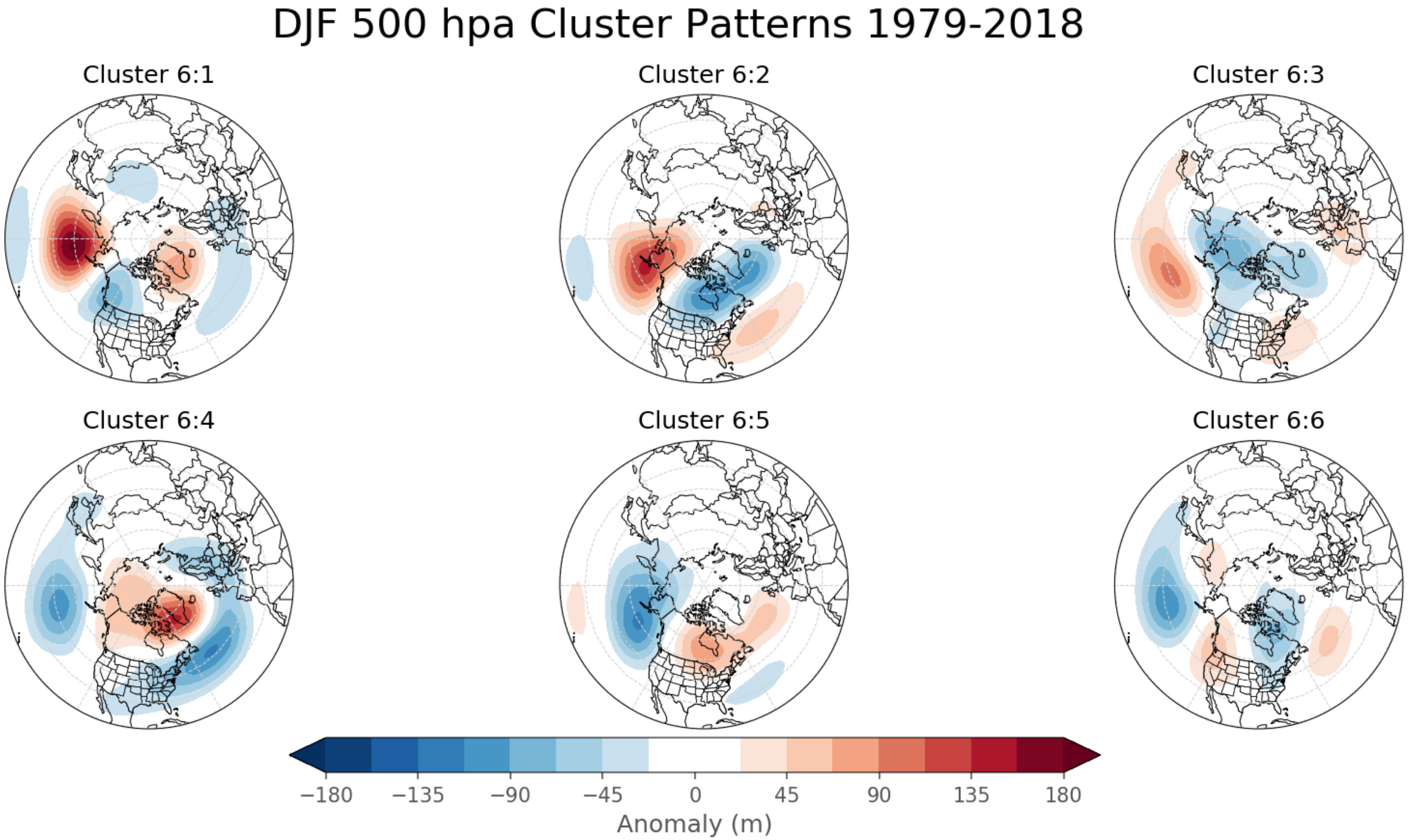
Cluster analysis finds groups of states that are clumped together.

Illustration: Applying k-means algorithm in a 2D phase space
(Each dot represents the Z500 field expressed in terms of the
values of the two leading principal compos



Each point corresponds to a map in physical space

Categorizing the circulation pattern via k-means clustering
Apply to running 14-day means for DJF (from ERA5)
(patterns shown using 10 PCs)



*No significance in the order of clusters!

Some check on robustness of clusters: Use Half-Length datasets

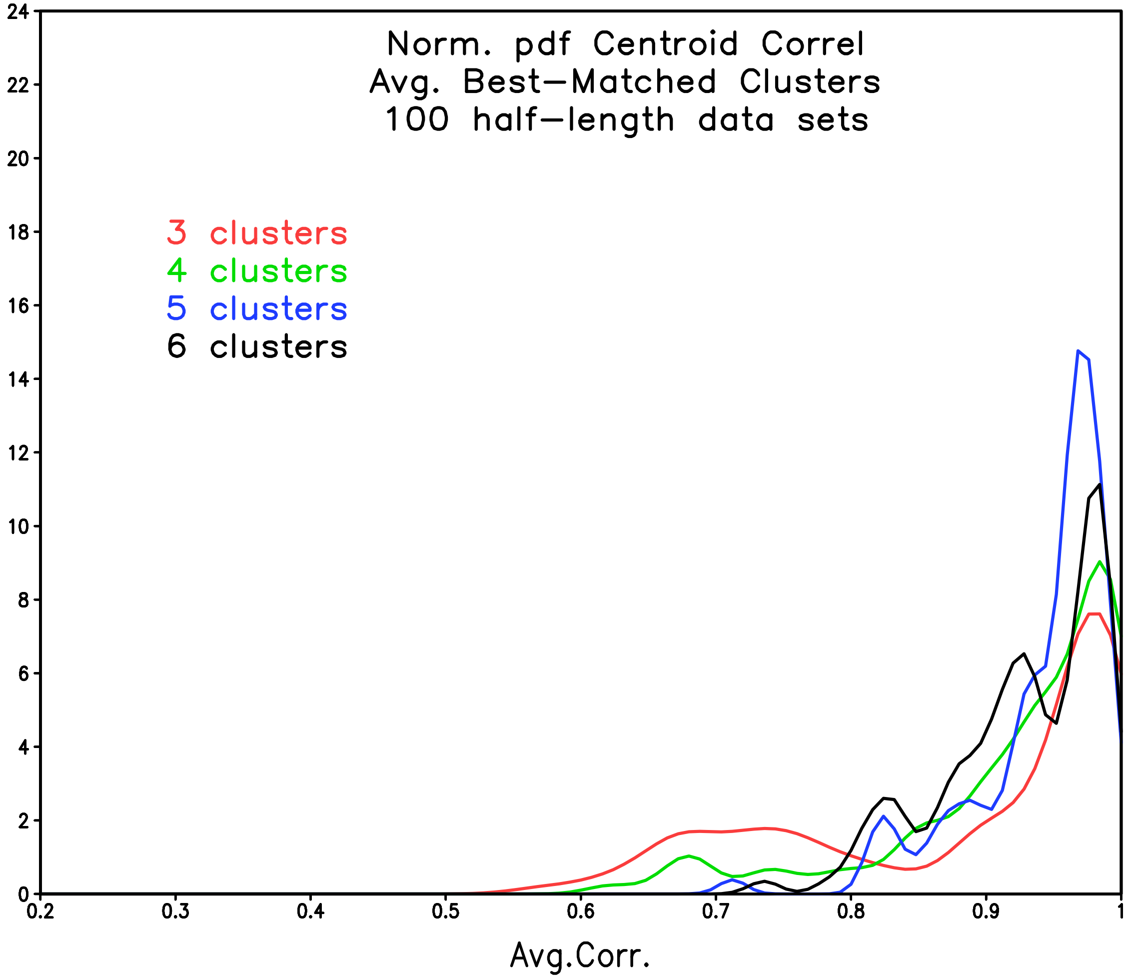
Randomly pick half the days within each season and recompute the cluster centroids. This is repeated for many trials.

For each trial, match the centroids to the reference (full data set centroids) as follows: Each centroid is assigned to a corresponding reference centroid on the basis of pattern correlation, so the matching may not necessarily lead to a one-to-one assignment. For each trial compute an average pattern correlation across the k centroids

Measures of reproducibility:

(i) What is the pdf of average pattern correlation estimated from the 100 trials?

(ii) For what percentage of trials is the matching one-to-one?



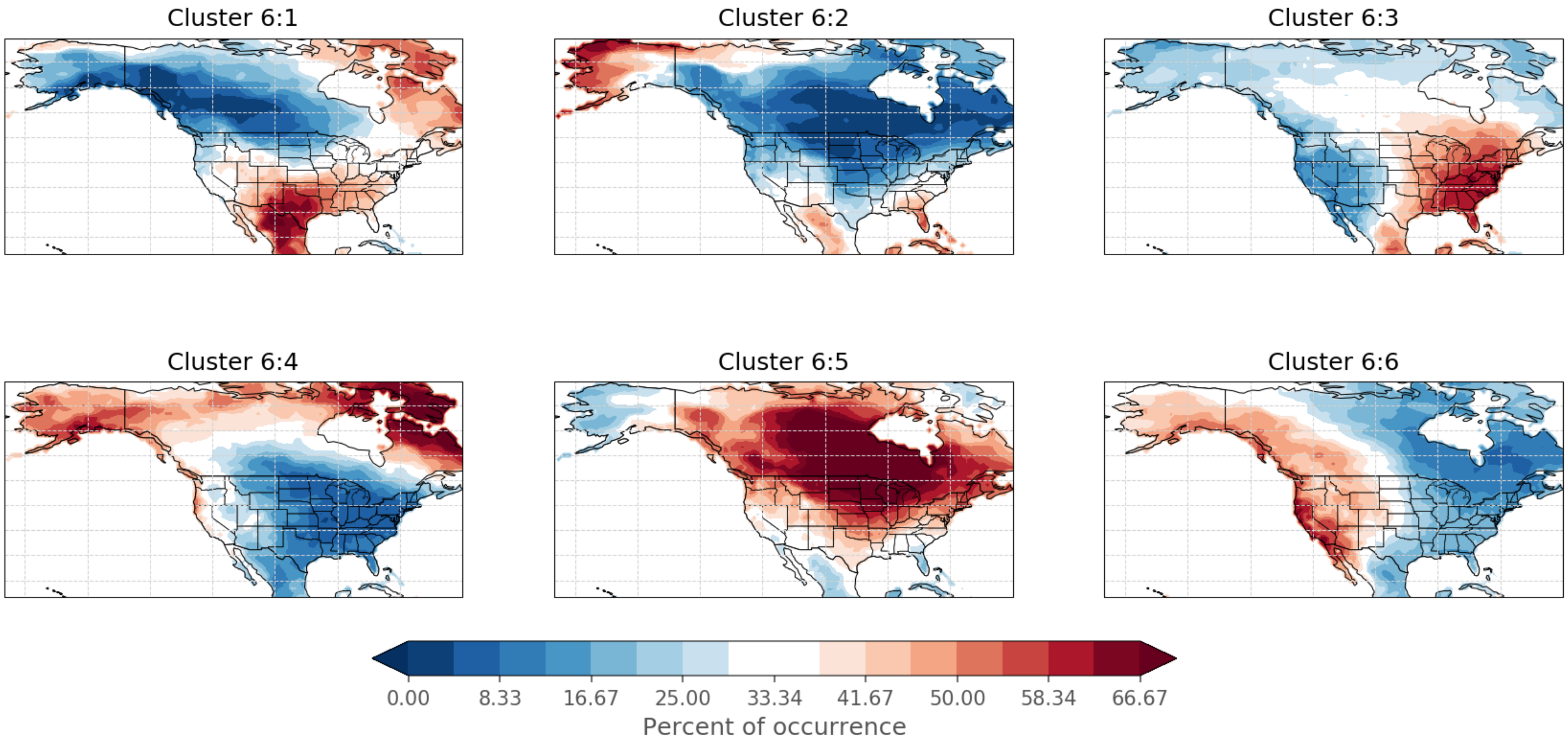
Percentage of trials for which the matching is one-to-one

k=2	100%
k=3	100%
k=4	90%
k=5	89%
k=6	64%

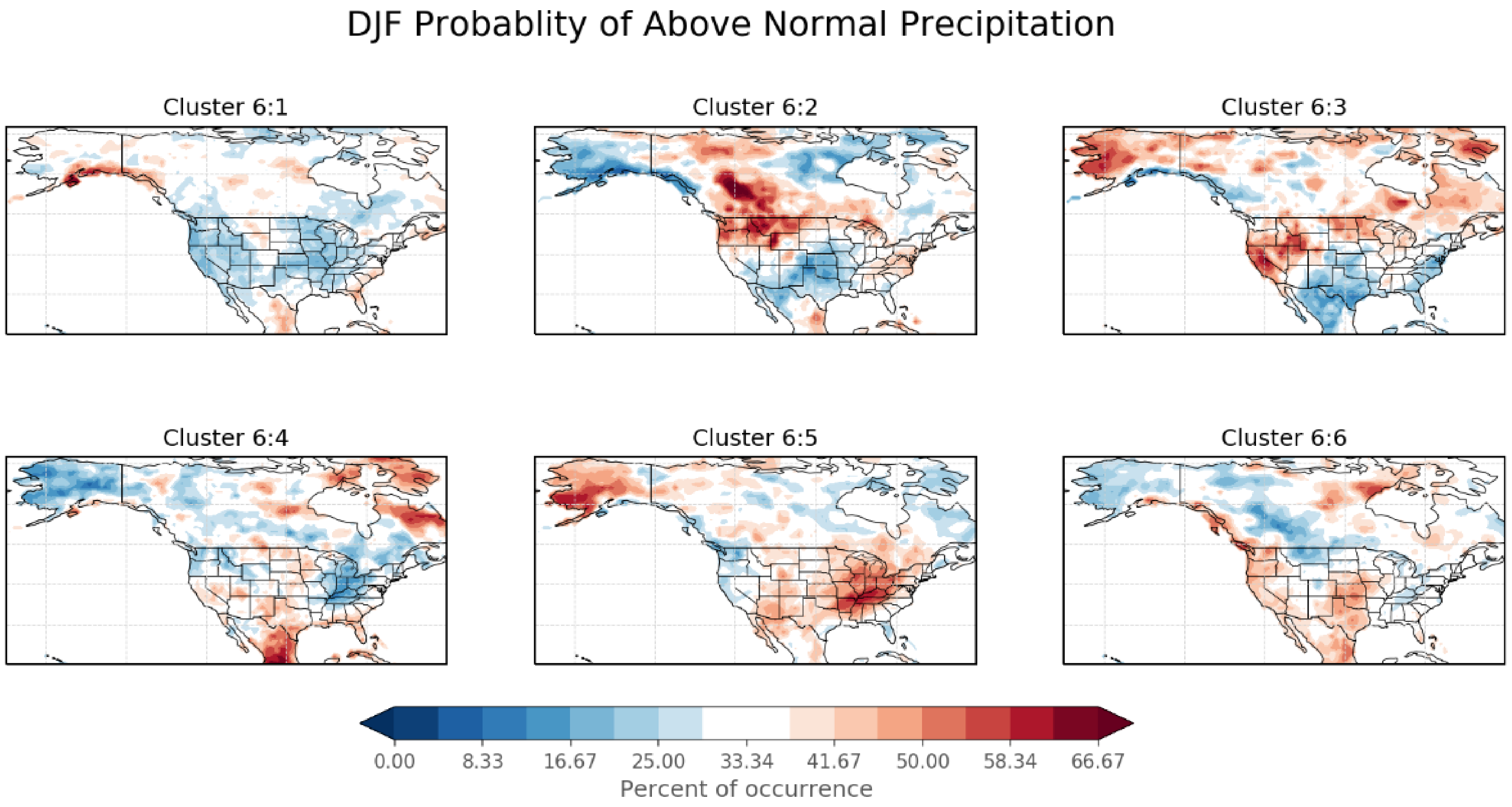
ERA Interim 1997-2015 Regimes
DAILY DATA for Oct-Mar (ONDJFM)
Z500
5 PCs retained

Associated Cluster Composites of Weather (2m T)

Probability of Above Normal **Temperature**



Associated Cluster Composites of Weather (precip)

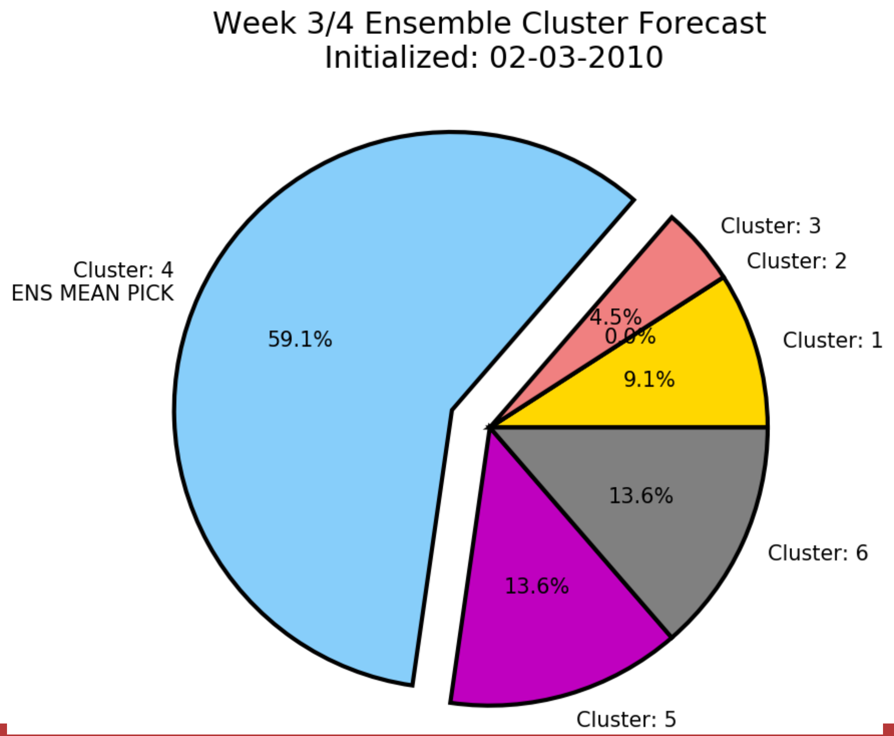


Associated Cluster Composites

- Now that we have a cluster assignment for each 14-day period, we can find the associated anomaly composites for any variable. We first investigate temperature and precipitation terciles.
- Data:
  - Temperature: CPC 2-meter Daily Reanalysis
  - Precipitation: CPC Global Unified Gauge-based Daily Precipitation 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

Cluster-Based Forecasts

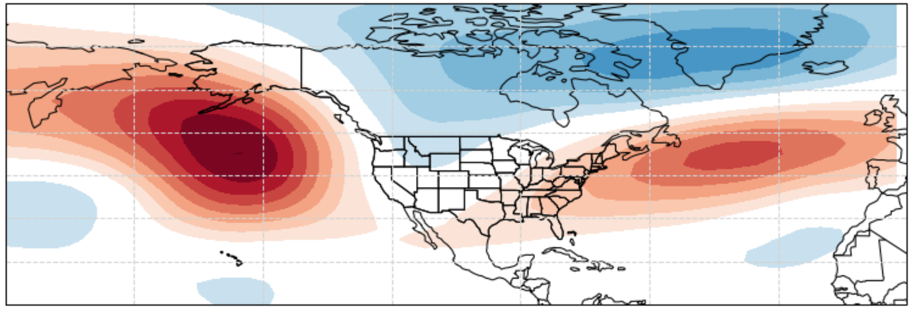
- Method:
  - We can make forecasts using this analysis by assigning the average Week 3-4 z500 forecast to a single observed cluster (via minimum RMSE)
  - Composite maps are weighted by the number of ensemble assignments. Weighted averages of the corresponding *observed* T2m and precip give a final probabilistic forecast, comprising chances of above, near and below average



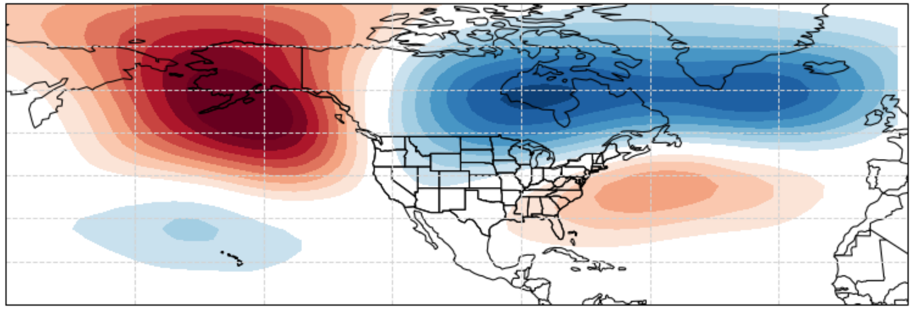
- Data:
  - GEFSv12 and ECMWF hindcast z500 forecasts
  - 22 total ensemble members
  - Weekly Initialized between 11/15-2/15
  - For years 2000-2019
  - 252 samples

Cluster z500 Week 3/4 Forecast: 02-02-2022 to 02-15-2022

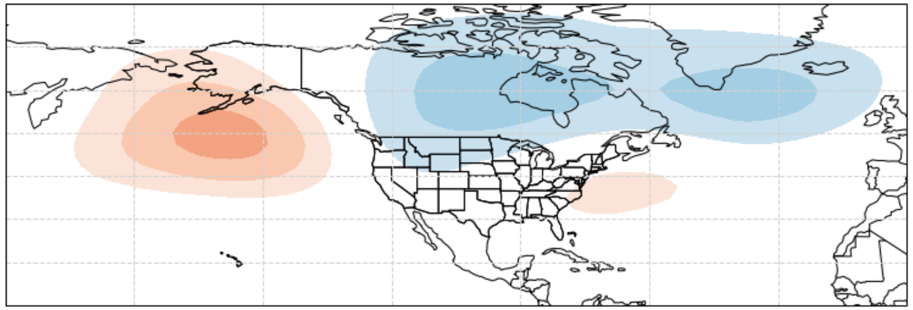
GEFSv12(31)/ECMWF(51) ENS Member Mean



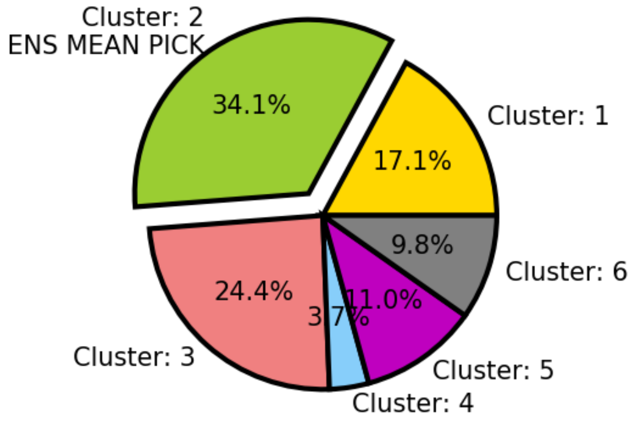
Ensemble Mean Cluster Assigment: 2

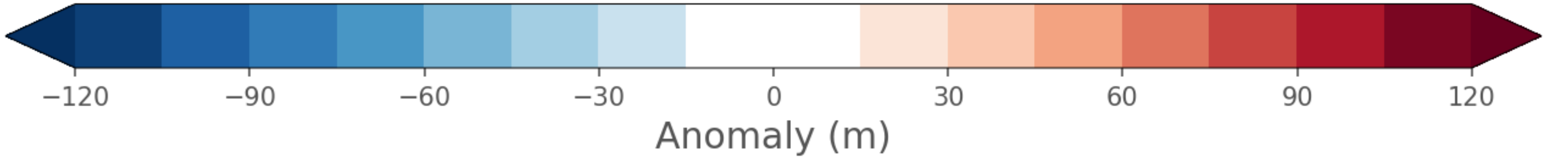


ENS Member Weighted Cluster Composite

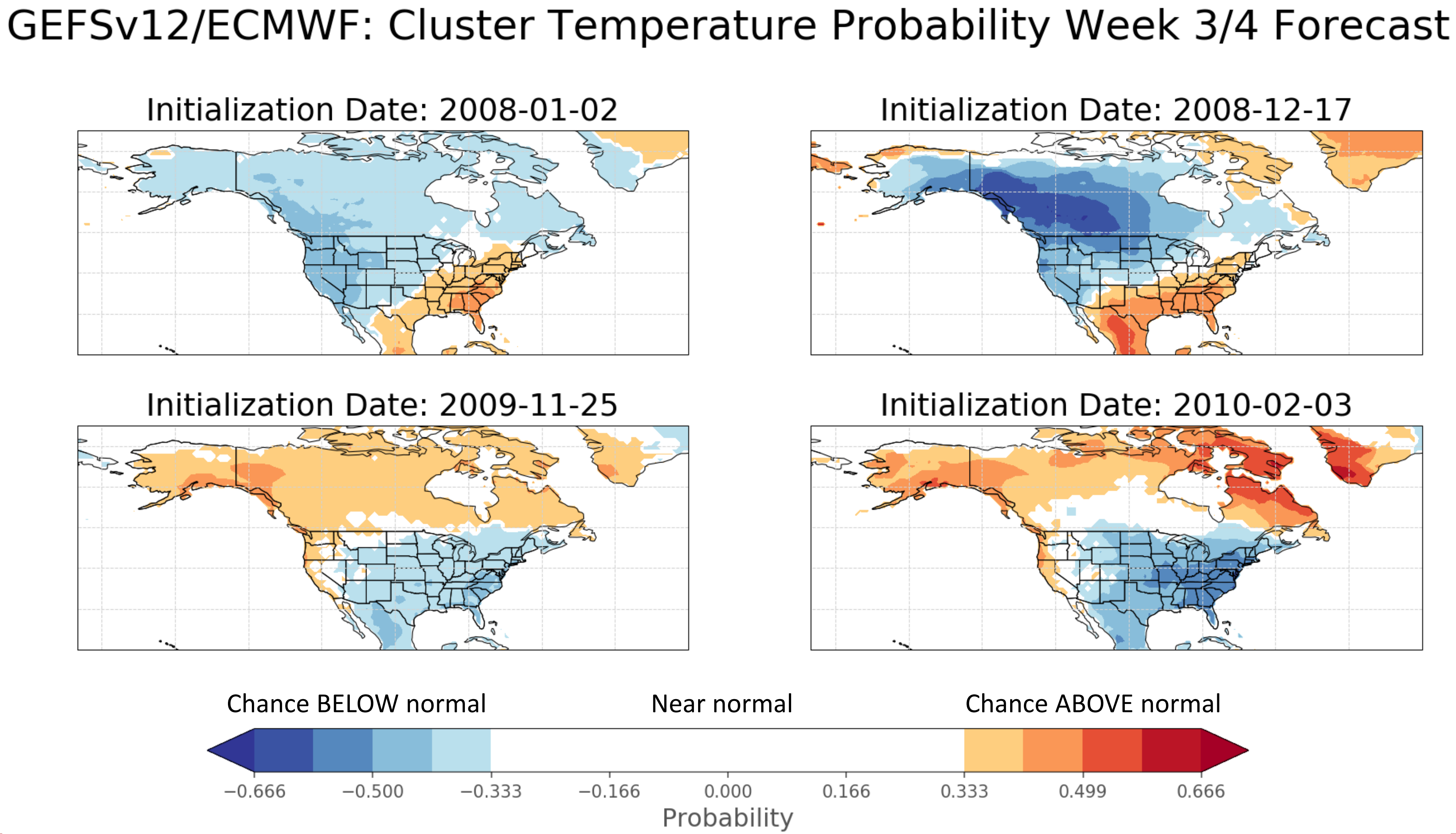


Ensemble Cluster Forecast: N=82





Cluster-Based Forecasts



Cluster-Based Forecasts

- Initial skill scores:
  - Scoring all 252 hindcast samples with the same observational datasets (cluster forecasts, original forecasts)
  - Score using the Heidke Skill Score (HSS)
  - Scoring only the CONUS/AK region

Forecast Type	Forecast Category	Forecast Count	Temperature	Precipitation
Cluster: Maximum Category	Above/Near/Below, Near=Ignore	252	18.9	7.9
GEFS/ECMWF Raw Anomaly Forecast	Above/Below	252	22.9	10.3
Cluster: Top 2 Cluster Assignments sum to > 75%	Above/Near/Below, Near=Ignore	50	27.0	**17.6**
GEFS/ECWMF Anomaly Forecast: For same forecasts as Row 3	Above/Below	50	28.2	15.6

T scores for cluster forecasts are competitive with (although slightly less skillful) than original forecasts

Comparison of cluster method for DJF with the other 3 seasons (MAM, JJA, and SON)

Cluster Period	# Hindcasts	Best K #	Temp HSS	Precip HSS	Z500 Pattern Corr.
DJF Clusters	264	6	13.1	7.8	.234
NDJFM Clusters (DJF hindcasts)	264	6	14.9	8.0	.242
MAM Clusters	266	7	9.2	3.7	.145
JJA Clusters	277	7	4.4	4.6	.241
SON Clusters	269	8	1.9	1.1	.080

Skill scores for weighted clusters forecasts based on GEFSv12 hindcasts (2000-2020). Heidke skill scores are used for temperature and precipitation forecasts and pattern correlation for z500. The best performing number of clusters (k) is displayed for each but k=4 through k=8 were considered. Using the NDJFM clusters gave a slight boost in skill over the DJF period, but not over the NDJFM period (not shown).

The cluster method has the most utility during the DJF period.

Tentative Conclusions

- The dominant circulation regimes over North America are combinations of teleconnections.
- These clusters have relationships to temperature and precipitation anomalies, and likely other fields.
- Based on hindcast testing, the cluster framework provides forecasts competitive with the model probabilistic forecasts, particularly when the z500 forecast is skillful and can identify possible forecasts of opportunity.
- The cluster-based forecasts can provide additional information: probability of extremes, changes in storm tracks, and more.
- The method needs more development when applied to the transition and summer seasons.

Ongoing/Future Work

- Use of exremes information and storm tracks in the web-based tool

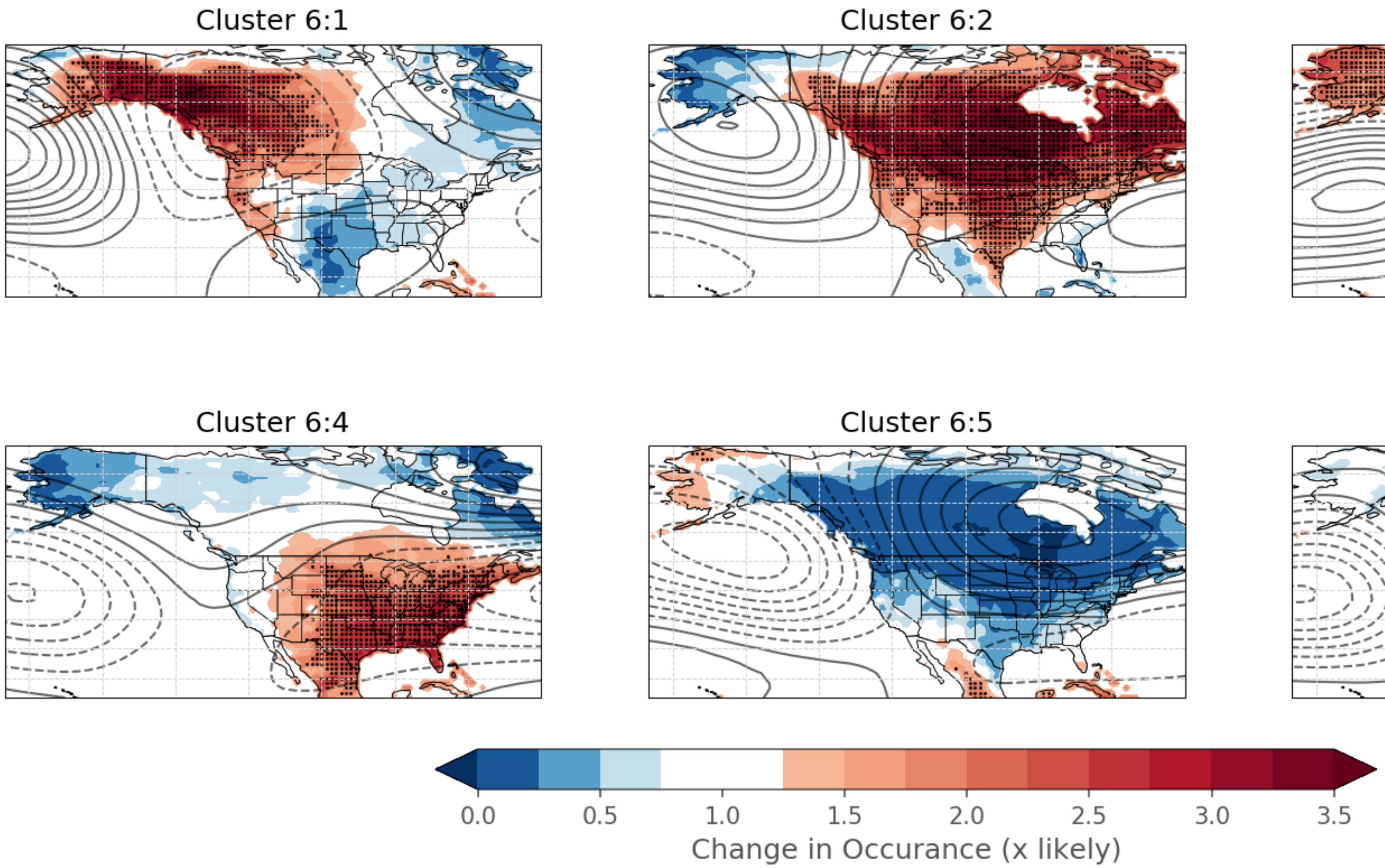
- A further investigation into forecast method/forecasts of opportunity

- Further documentation of difficulties forecast models have in reproducing the circulation regimes.

- Adaptation of the method for seasons other than boreal winter.

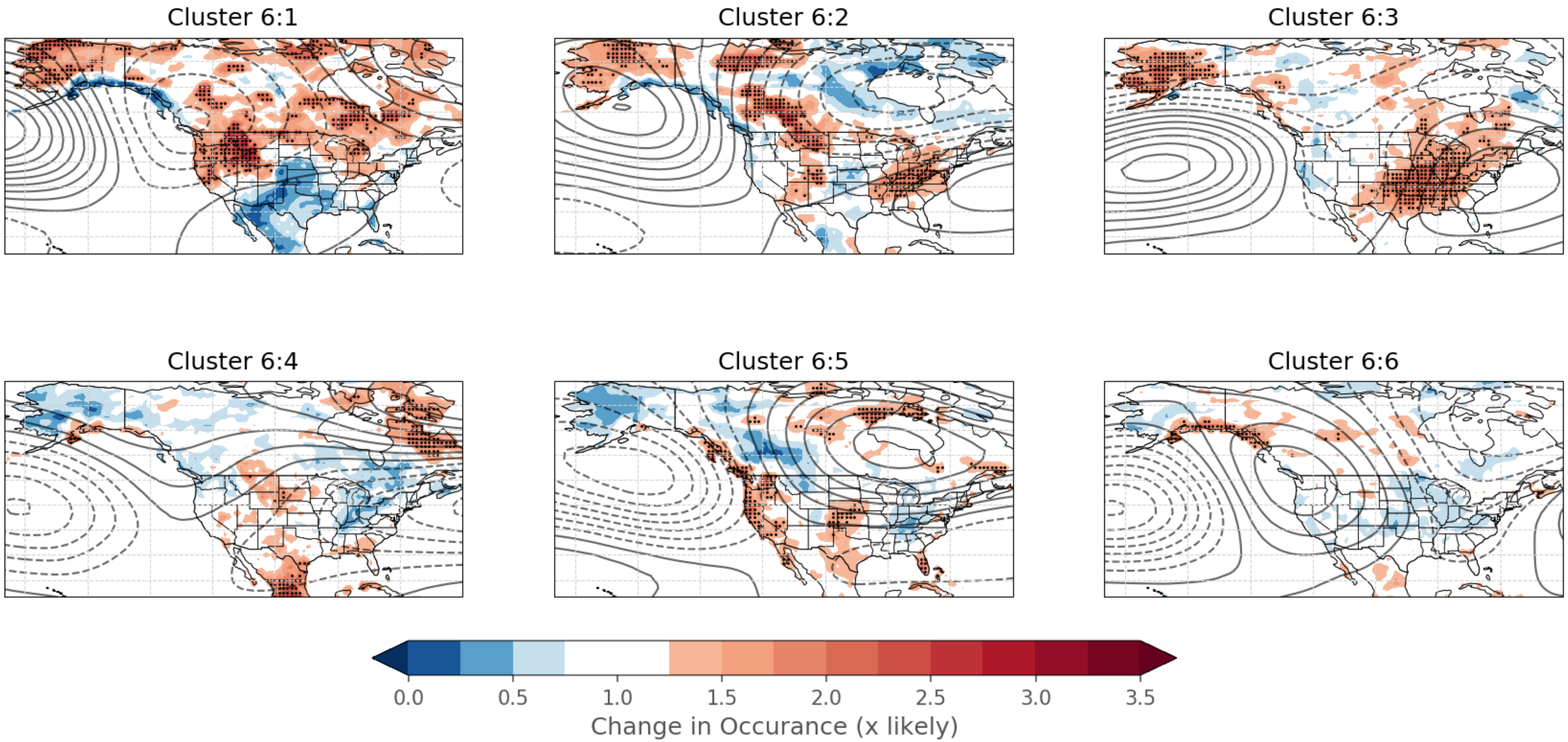
Associated Cluster Composites

Probability of **Extreme Cold** (15th Percentile Temperature)



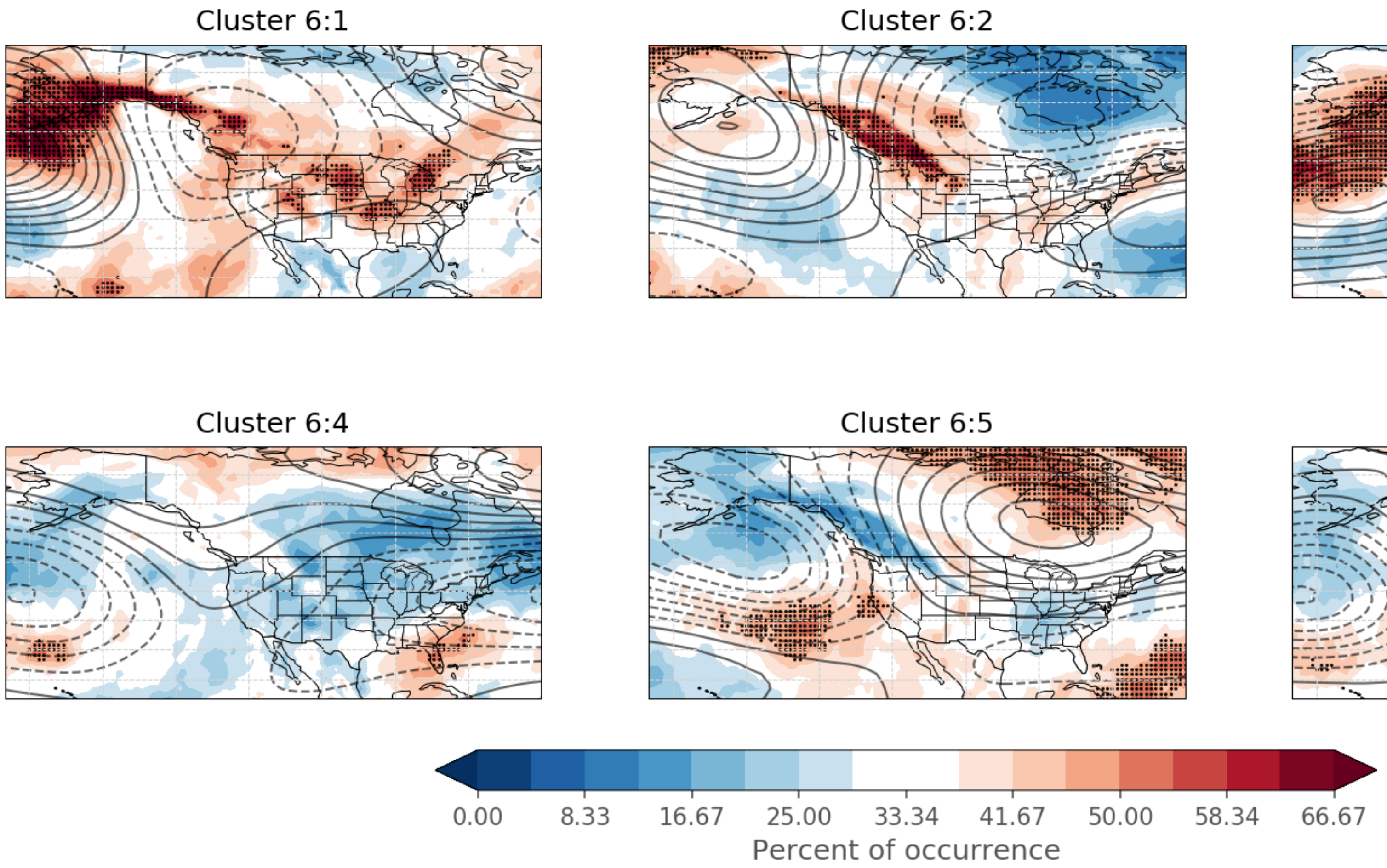
Associated Cluster Composites

Probability of **Extreme Precipitation** (85th Percentile)



Associated Cluster Composites

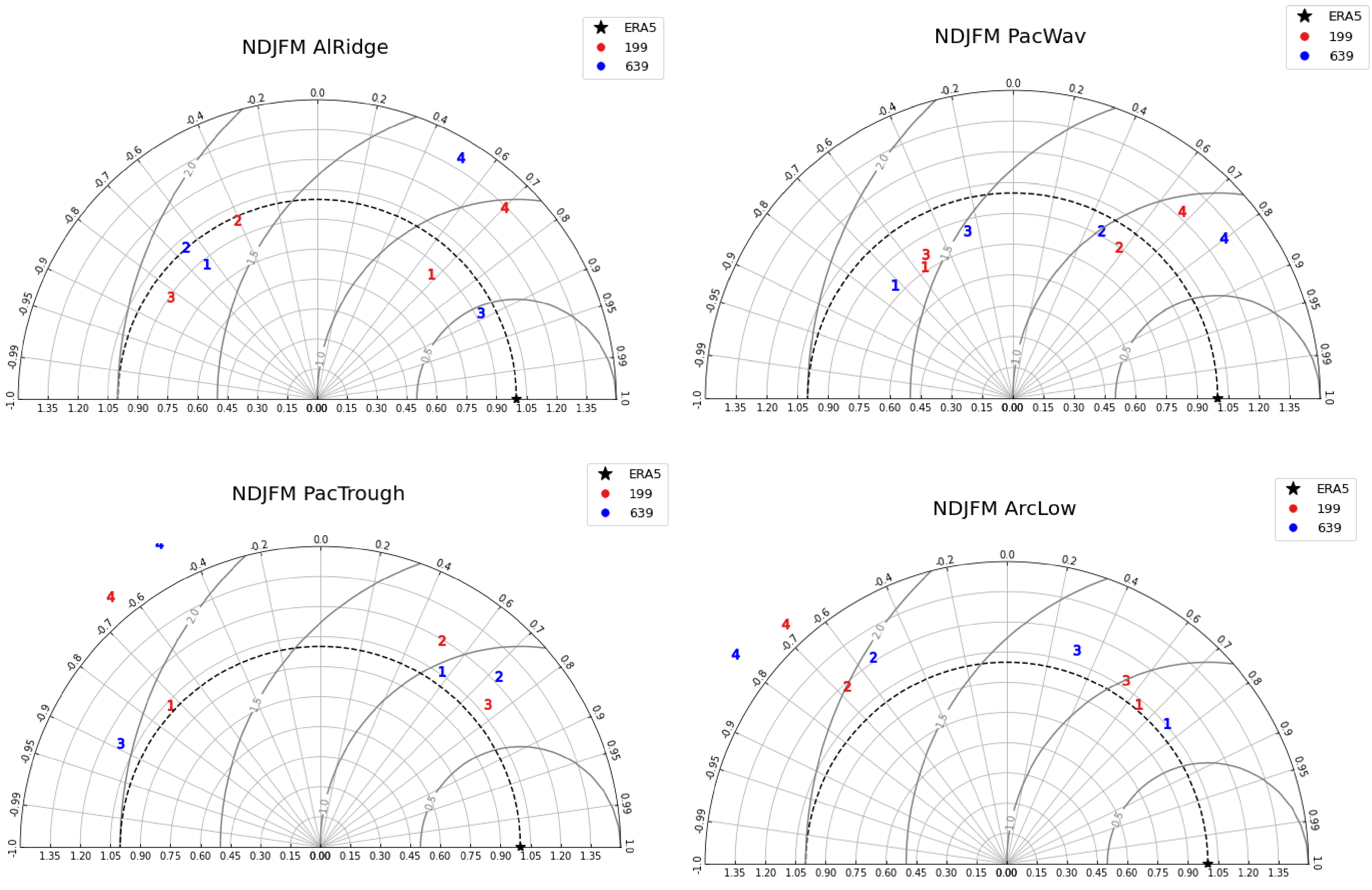
Probability of Above Normal **Storm Tracks** (v't'850)



Taylor diagrams:
Star represents ERA5.
Correlation on azimuthal angle, (grey lines radiating from origin), RMSE is distance from ERA5 (solid grey arcs centered on star), standard deviation is distance from origin (dashed black is ERA5 standard deviation)

Numbers indicate values for a particular cluster. Color indicates resolution.

The closer the number to the star, the better the representation.



Alaskan Ridge Composites (II)

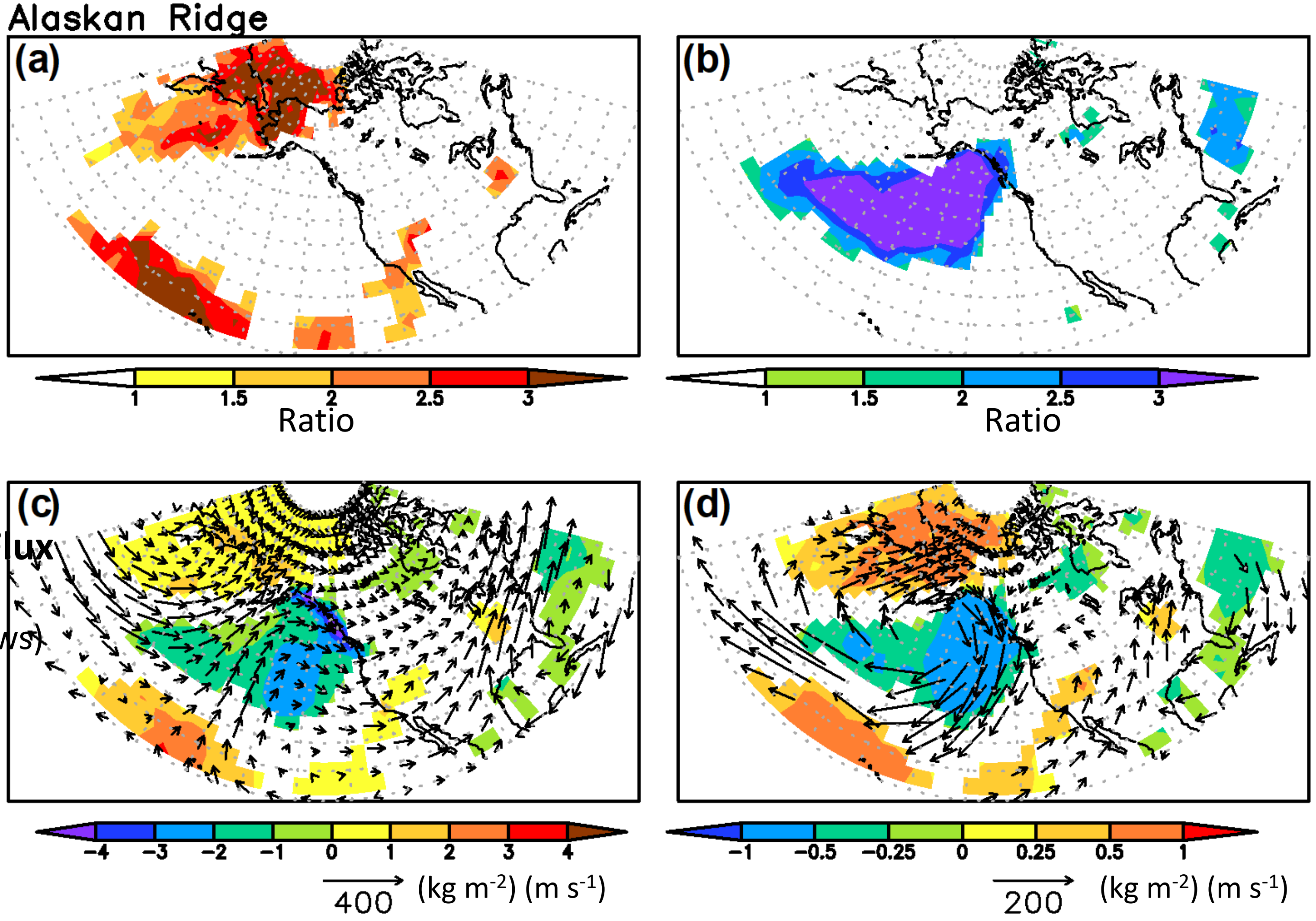
Ratio R

95th percentile precipitation

Ratio: number of regime states with 95% prec. / < number >$_{CLIM}$

5th percentile precipitation

Ratio: number of regime states with 5% prec. / < number >$_{CLIM}$



Precip. & Moisture Flux

Precip. in shading

V.I. moisture flux (arrows)

Contour interval: 10^{-3} m/day

Normalized Precip. & Moisture Flux (Anomaly)

Precip. in shading

V.I. moisture flux anom: arrows

Contour interval: stand. Dev.

Long-Lasting Alaskan Ridge Composites (>14 days)

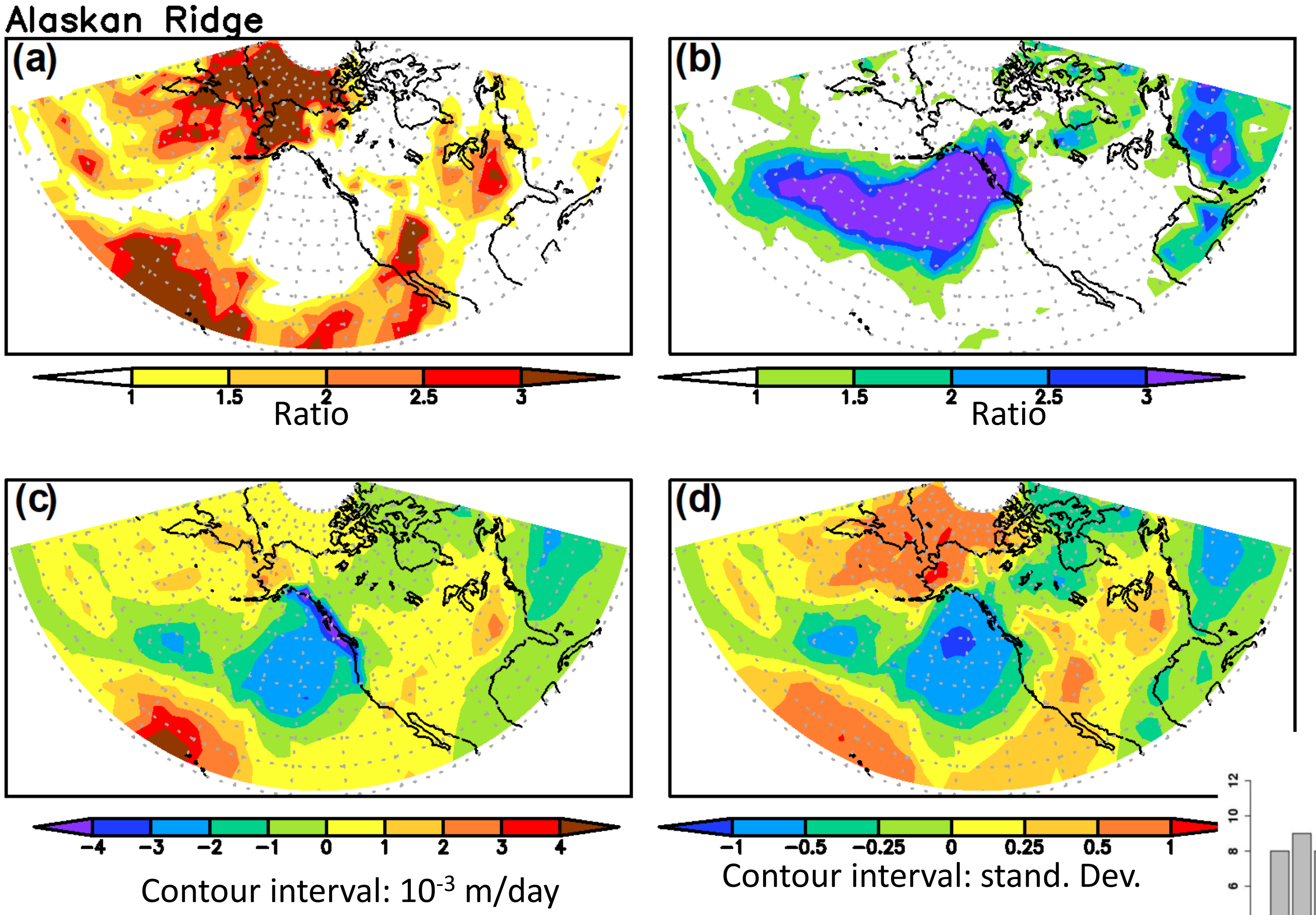
Ratio R

95th percentile precipitation

Ratio: number of regime states with 95% prec. / < number >$_{CLIM}$

5th percentile precipitation

Ratio: number of regime states with 5% prec. / < number >$_{CLIM}$



Precip.

Precip. in shading

Normalized Precip.

Precip. in shading