# Using Clustering to Evaluate the CESM1 Large Ensemble's Ability to Simulate Regional Variations in Heat Waves

Savanna Wolvin[1], Andrea Jenney[2], David A. Randall[2]
St. Cloud State University, St. Cloud[1], MN and Colorado State University, Fort Collins, CO[2]

## Motivation

- Nighttime heat waves are known to kill people
- Can global climate models simulate regional variations in nighttime heat waves?
- And if not, what processes are being modeled incorrectly?

## Data & Methods

Data used:

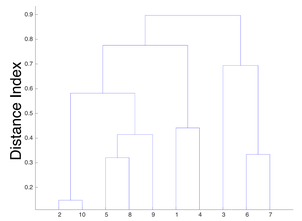
- Observations: Global Historical Climatology Network – Daily; 1948 - 2017
- Model: Community Earth System Model version 1 (CESM1), Large Ensemble (LE) Project[1], 1948 – 2005

Defining a Nighttime Heat Wave:

- Summer (JJA) Daily Minimum Temperature Anomalies
- 90th Percentile Exceedance

Analysis Methods:

- Hierarchical Clustering – Each location has a binary data vector indicating which days it experiences a nighttime heat wave. These data vectors are grouped to minimize the average Jaccard distance between all pairs within a group.




Example of hierarchical clustering, grouping values based on creating the lowest average Jaccard Distance.

**Distance Index**

1 — Completely Unalike
0.8 — More Dissimilar than Similar
0.6 — Neutral
0.4 — More Similar than Dissimilar
0.2 — Identical
0

- Distance Index between Hierarchical Clusterings[2] – Example:

| Location Pairs | $X_1$ | $X_2$ |
|---|---|---|
| A - B | 1 | 1 |
| A - C | 0 | 0 |
| A - D | 0 | 1 |
| B - C | 0 | 0 |
| B - D | 1 | 1 |
| C - D | 1 | 0 |
| Sum | 3 | 3 |

$$Z = \frac{\| X_1 - X_2 \|}{\| X_1 \| + \| X_2 \|}$$

$X_1$ – Data vector for first hierarchical clustering

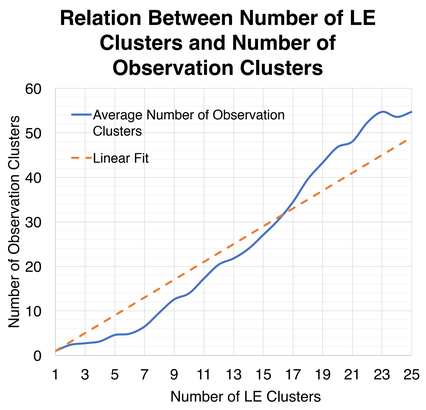
$X_2$ – Data vector for second hierarchical clustering

0 – pair in same cluster

1 – pair in different cluster

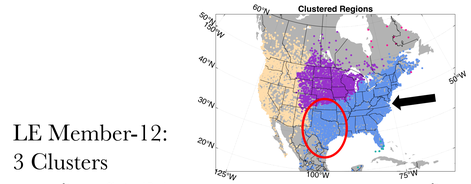
$$Z = \frac{2}{3+3} = 0.334$$

## LE & Observations Cluster at Different Rates



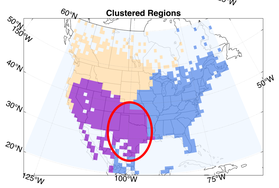

- Need a way to compare Observations to LE
- Figure shows number of observation groups for smallest distance index between Observations and LE
- K – Number of clusters
- Linear Fit: $K_{obs} = 2 \times K_{LE} - 1$
- $R^2 = 0.972$

## Comparing Clusters



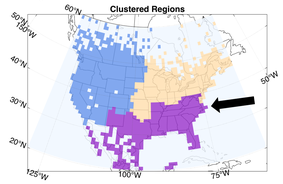

Observations: 5 Clusters

LE Member-12: 3 Clusters

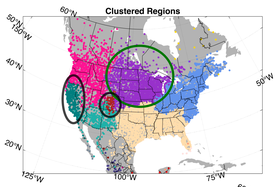



- Texas grouped with southwest US (red circle)

LE Member-6: 3 Clusters

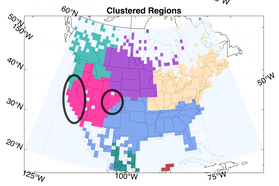



- Has NE-SE border (black arrow)



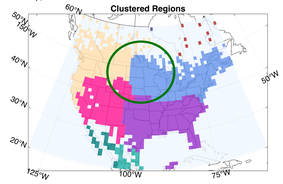

Observations: 15 Clusters

LE Member-9: 8 Clusters

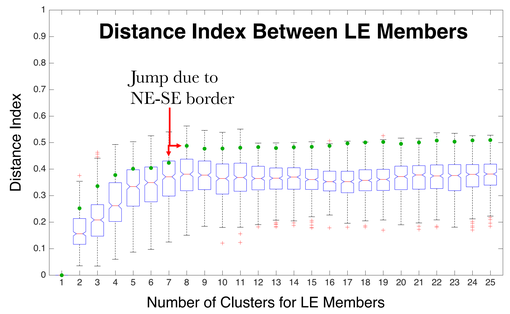



- Lacks Colorado and California cluster (black circles)

LE Member-22: 8 Clusters




- Missing Midwest grouping (green circle)

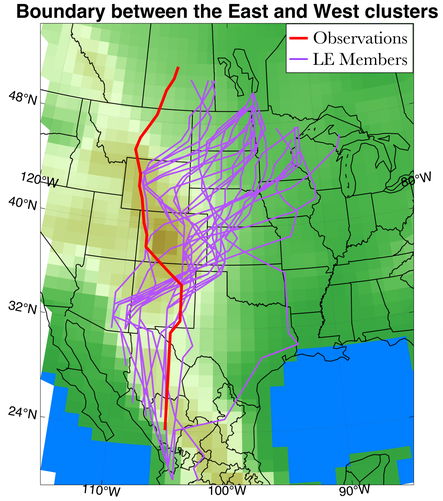



Green Plot: Mean distance index between Observations and Large Ensemble Members

- Range in Distance Indices between LE Members shows how clusters vary due to internal/random variability within the model
- Mean LE distance index and mean Observation distance index becomes relatively stable after 7 and 8 LE clusters respectively
- Jump in mean distance index for Observations from LE clusters 7-8 due to NE-SE border forming in Observations (red arrows)
- Observations are most similar to LE Members for smaller numbers of clusters
- Observations surpass 95th percentile of LE Members spread at 11 LE clusters

## Comparing Cluster Boundaries to Reanalysis

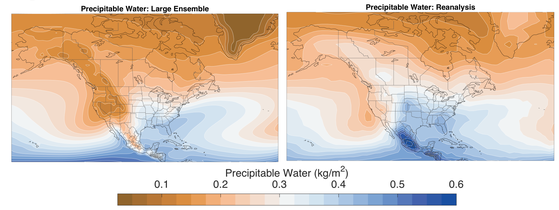
**Boundary between the East and West clusters**



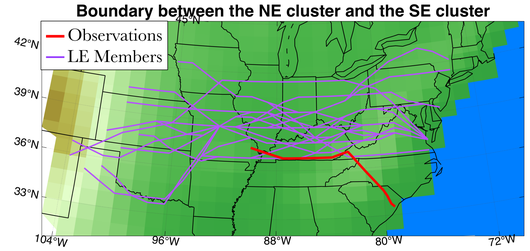

Left: East-West Group Border

- LE Member's E-W borders are east of observations
- Spread in LE Member borders in northern plains
- LE Member-12 border clusters Texas with western US
- Dry bias found over topography in LE (figure 1) could cause northern plains to be grouped with western US

**Figure 1: Precipitable water for LE and NCEP/NCAR Reanalysis data**



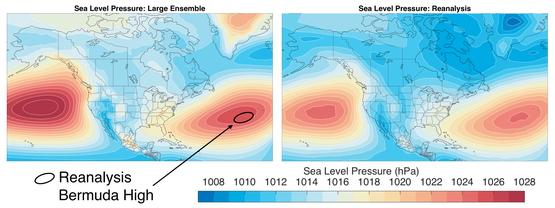

**Boundary between the NE cluster and the SE cluster**




Left: NE-SE Group Border

- LE Member's border clusters farther north than Observations
- LE borders reach farther west than Observations
- Bermuda high shifted farther north and is stronger in LE (figure 2), which could cause the northern shift in LE Member's NE-SE border

**Figure 2: Sea Level Pressure for LE and NCEP/NCAR Reanalysis data**




## Conclusions

- Observations most similar to the LE Members at lower numbers of clusters
- Dry bias over topography in LE, could cause the East-West border to be east of Observations
  - Could cause prediction of less deadly nighttime heat waves and less energy use
- Bermuda High is farther north and stronger in LE, could cause NE-SE border to shift north of Observations

## References & Acknowledgements

1. Kay, J. E., and Coauthors, 2015: The community earth system model (CESM) large ensemble project : A community resource for studying climate change in the presence of internal climate variability. Bull. Am. Meteorol. Soc., 96, 1333–1349.

2. Morlini, I., and S. Zani, 2012: An Overall Index for Comparing Hierarchical Clusterings. Challenges at the Interface of Data Analysis, Computer Science, and Optimization, W.A. Gaul, A. Geyer-Schulz, L. Schmidt-Thieme, and J. Kunze, Eds., Studies in Classification, Data Analysis, and Knowledge Organization, Springer-Verlag, Berlin Heidelberg, 29–36.

This work has been supported by the National Science Foundation Research Experiences for Undergraduates Site in Climate Science at Colorado State University under the cooperative agreement No. AGS-1461270.