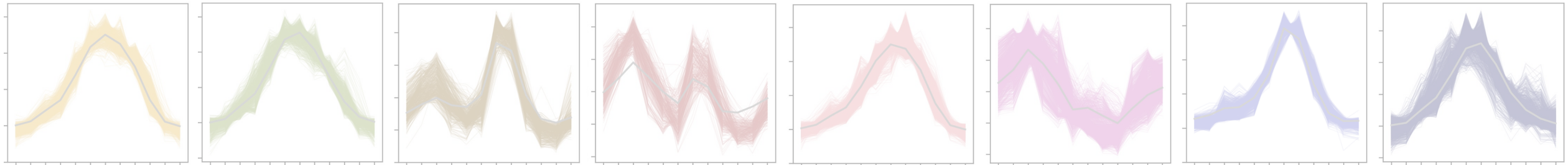
UNIVERSITY OF GOTHENBURG

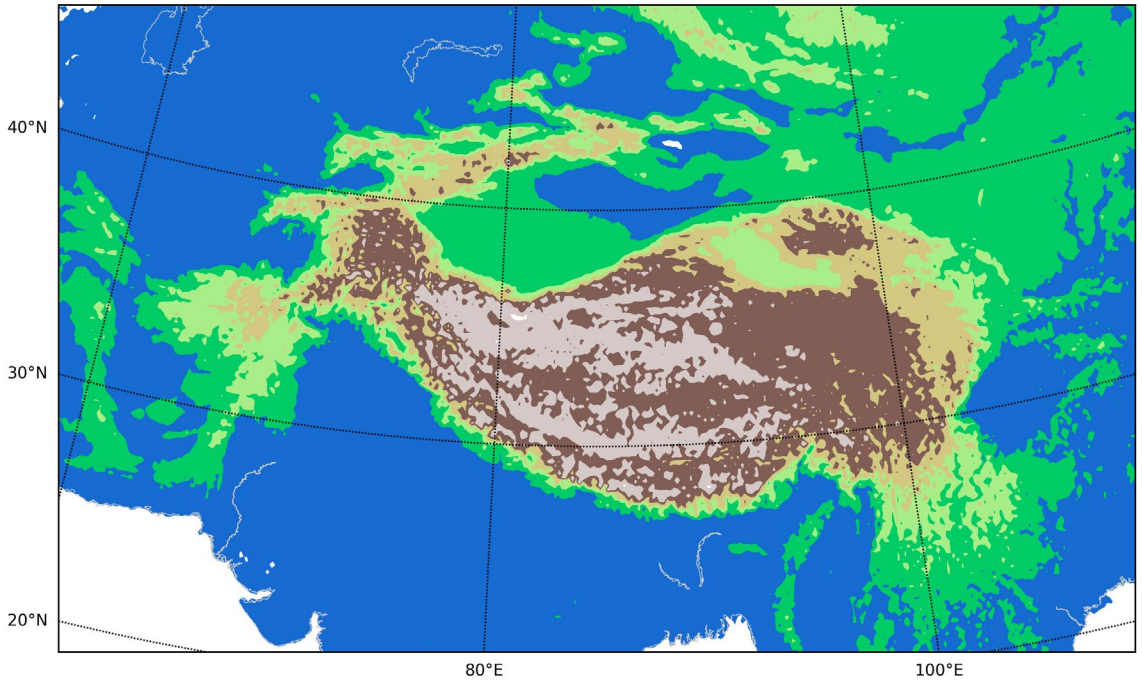
Using unsupervised learning algorithms to classify precipitation over the Tibetan Plateau for 1979 - 2018

Hui-Wen Lai
Supervisor: Deliang Chen

RCG meeting
22 November 2019



Dataset and study domain



Dataset: ERA-5

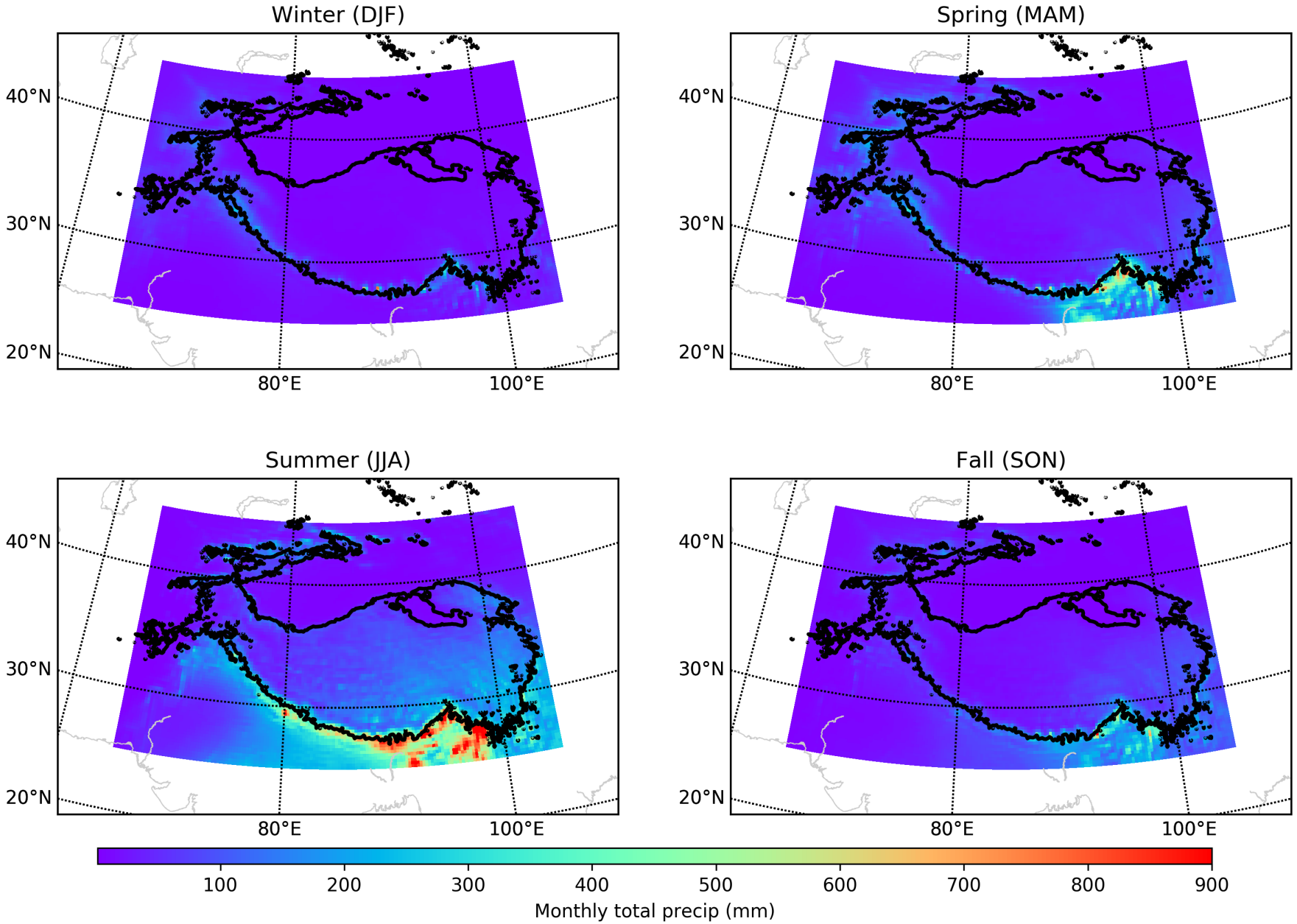
Domain: 25° N– 45°N, 65°E - 105°E

Period: 1979 – 2018

Horizontal resolution: ~31 km

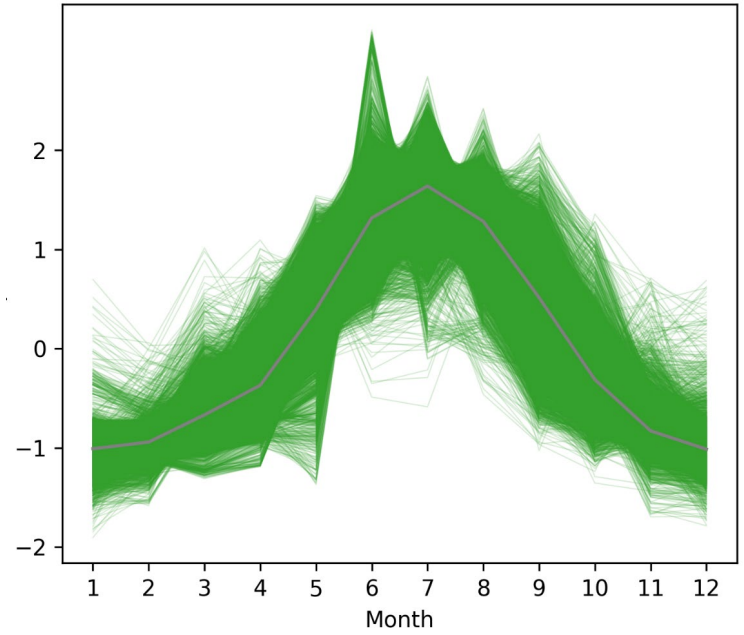
Variable: monthly mean total precipitation

Monthly total precipitation from 1979 - 2018



Data preprocessing

Training feature: season cycle



Data for training and predicting has been standardized.

$P_{y,m,g}$ = Precipitation data from ERA5

y = Year (1979-2018)
m = Month (1-12)
g = Gridpoints (#)

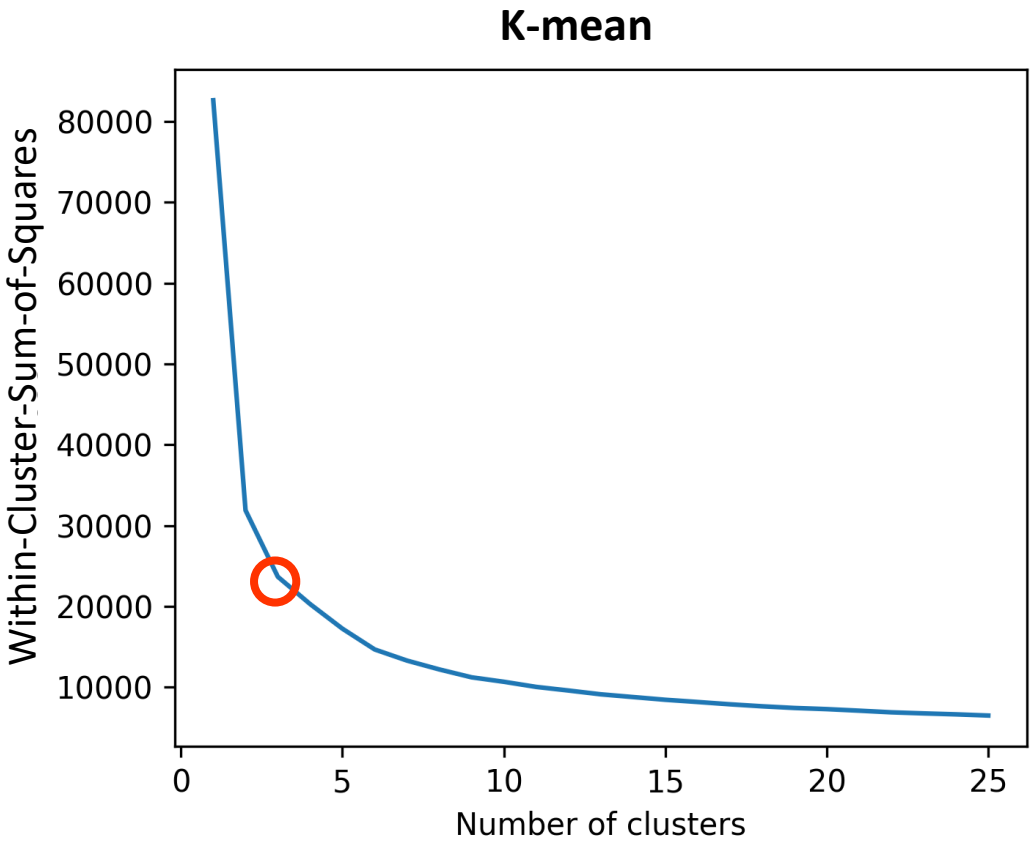
$$\bar{P}_{m,g} = \frac{\sum_{y=1}^{40} P_{y,m,g}}{I}$$

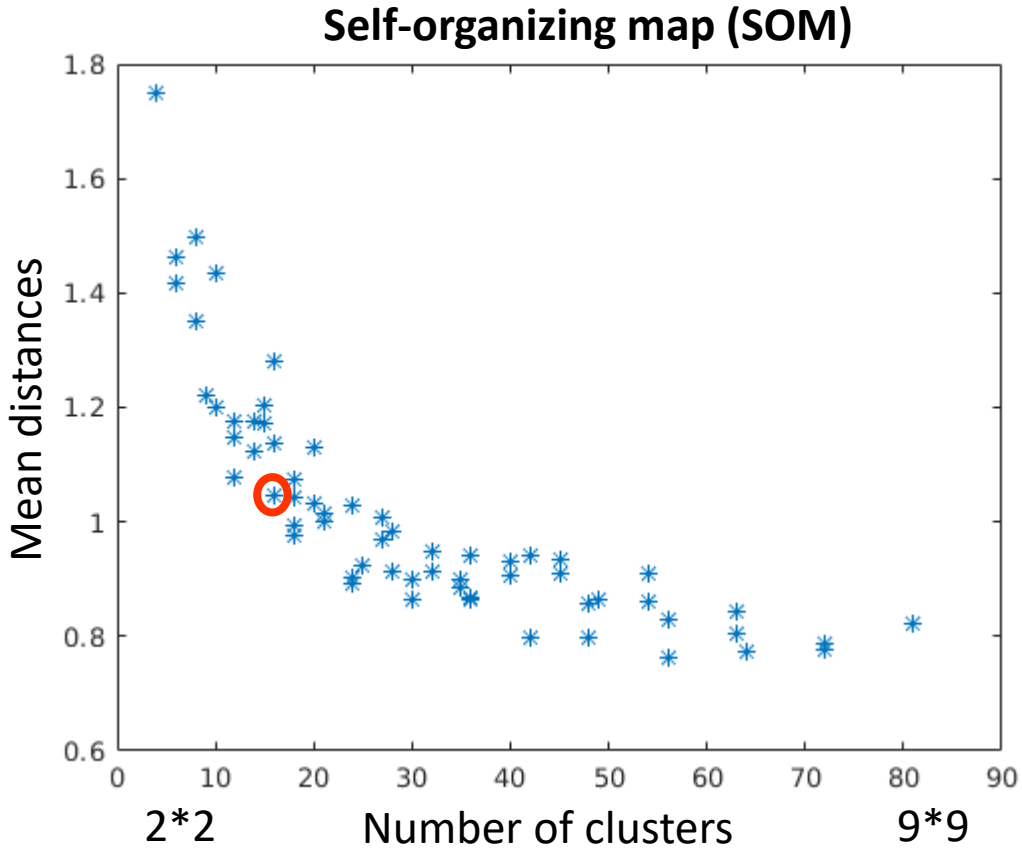
$$Pstd_{m,g} = \frac{\bar{P}_{m,g} - \bar{\bar{P}}_g}{\sigma}$$

How many clusters should we use?

Elbow method

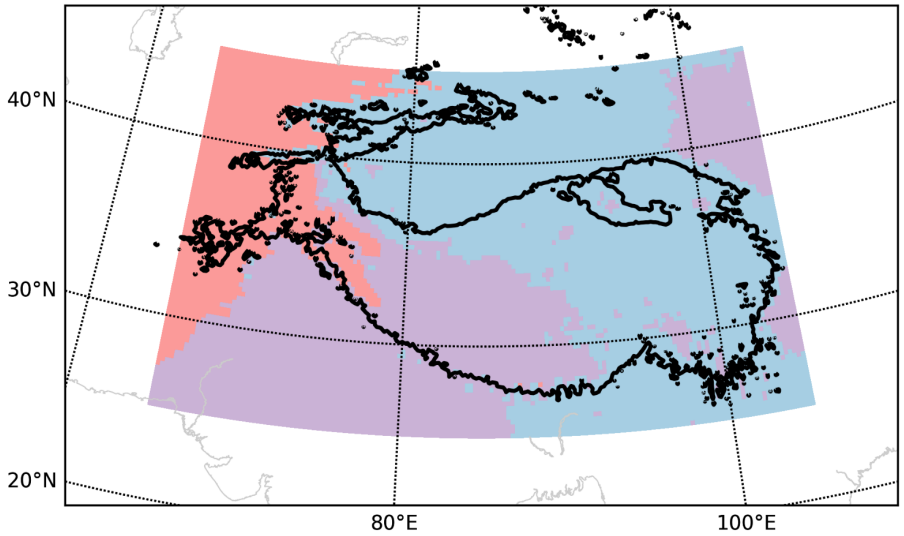
I decide 3 for K-means and 4x4 for SOM

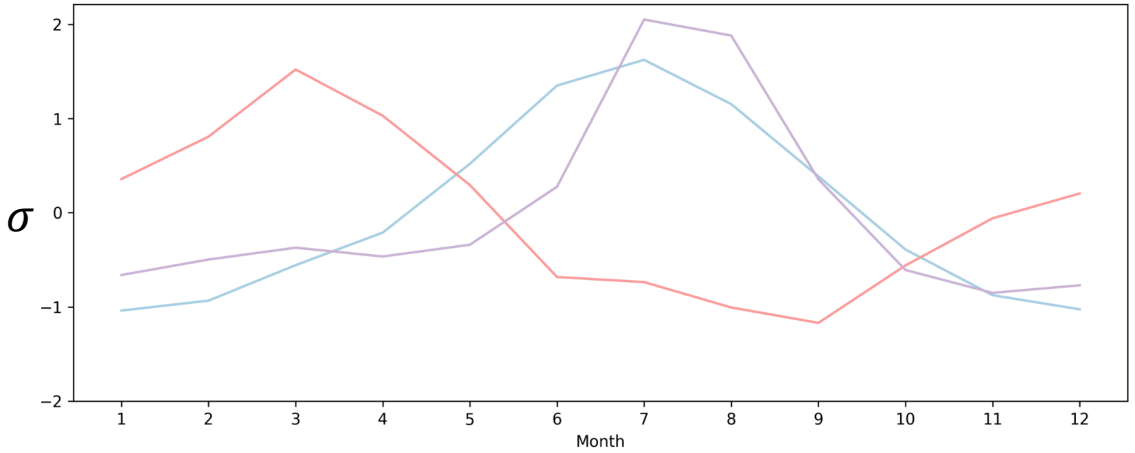




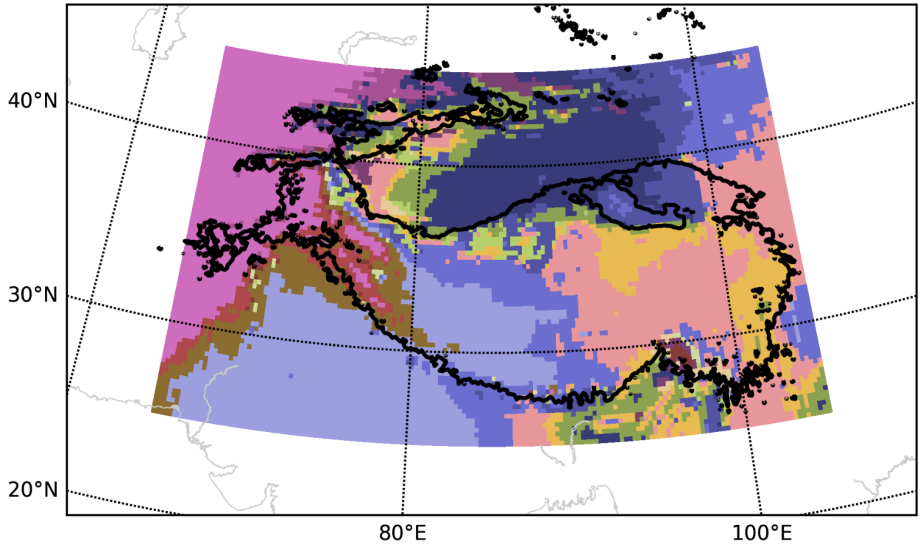
Maps of clustering results from K-mean and SOM

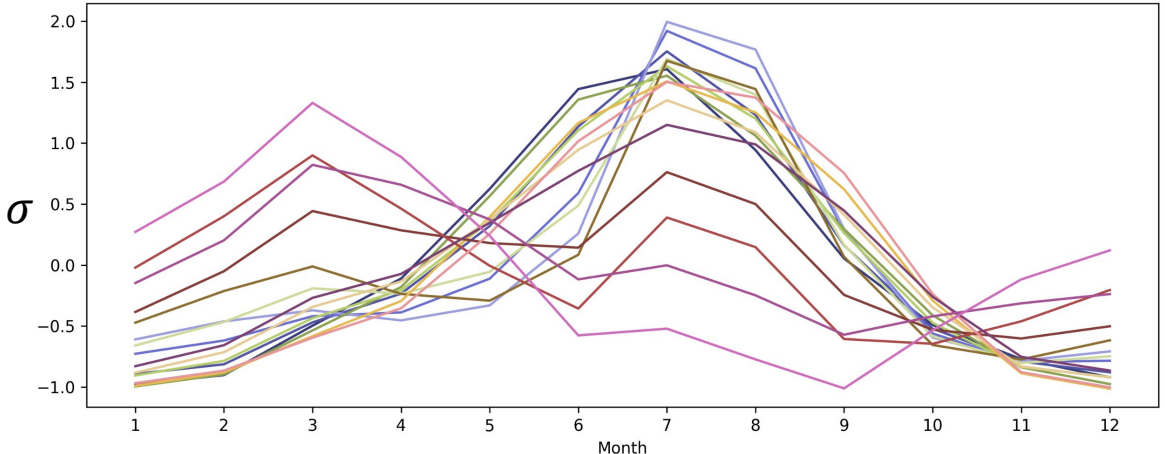
K-means (3 centroids)





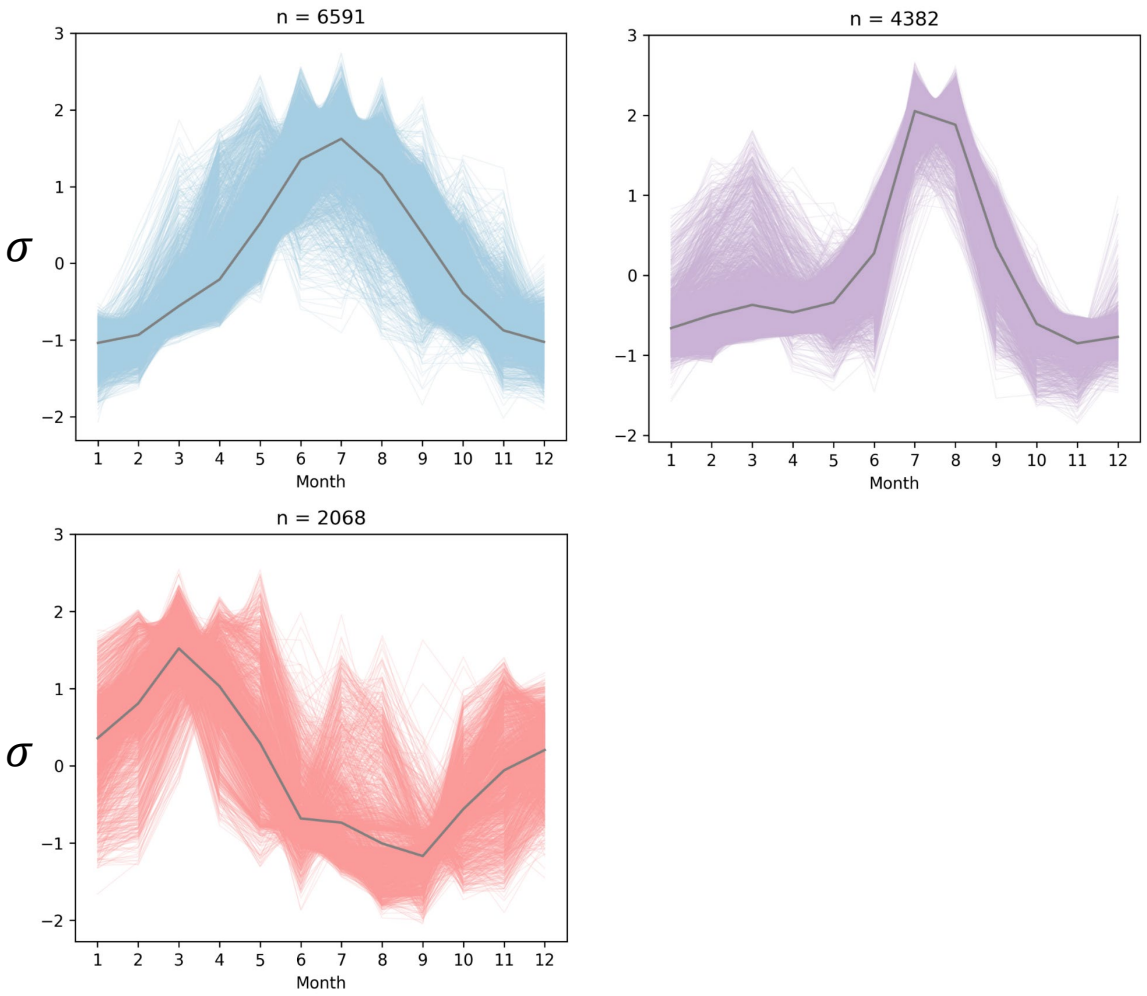
SOM (4 x 4)



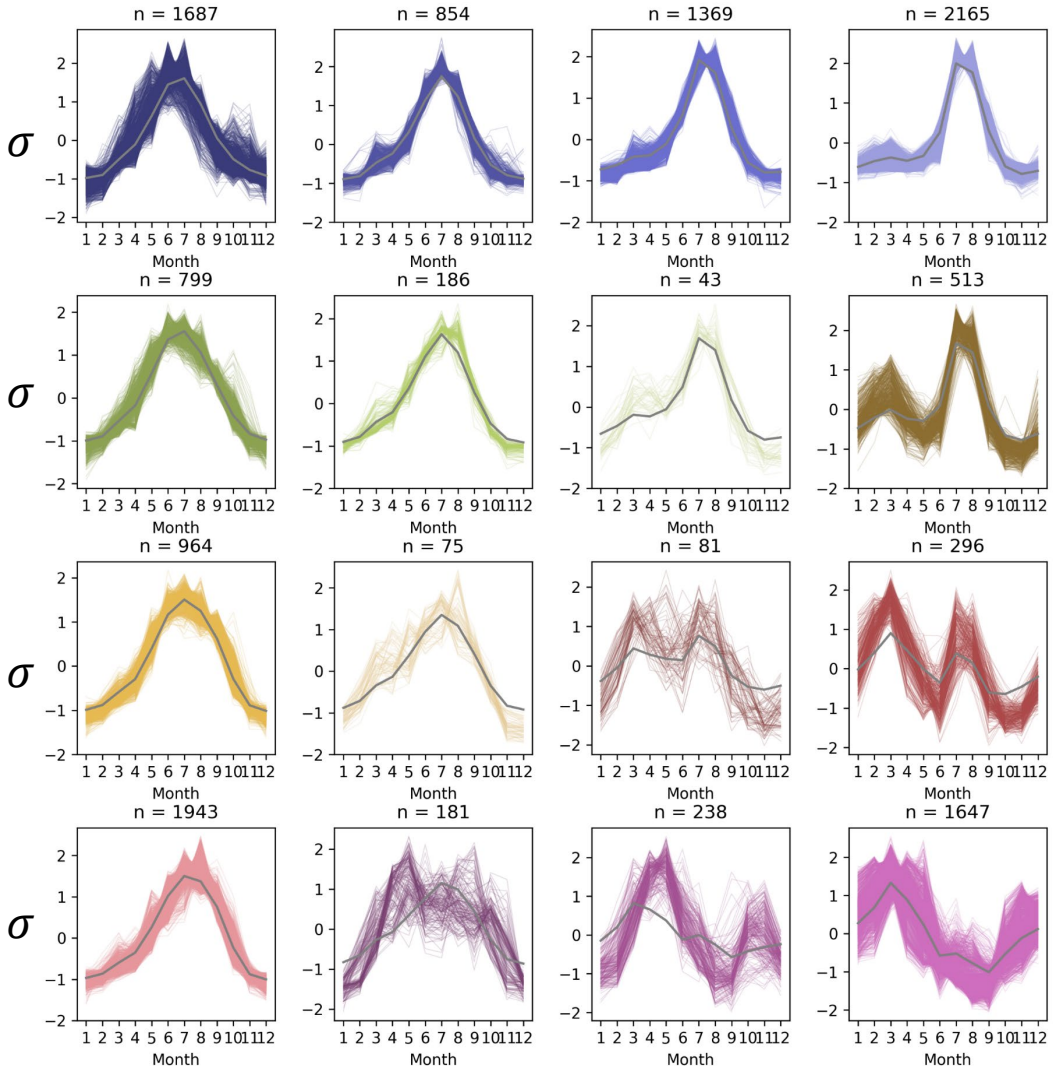


Seasonal cycles of nodes from K-mean and SOM

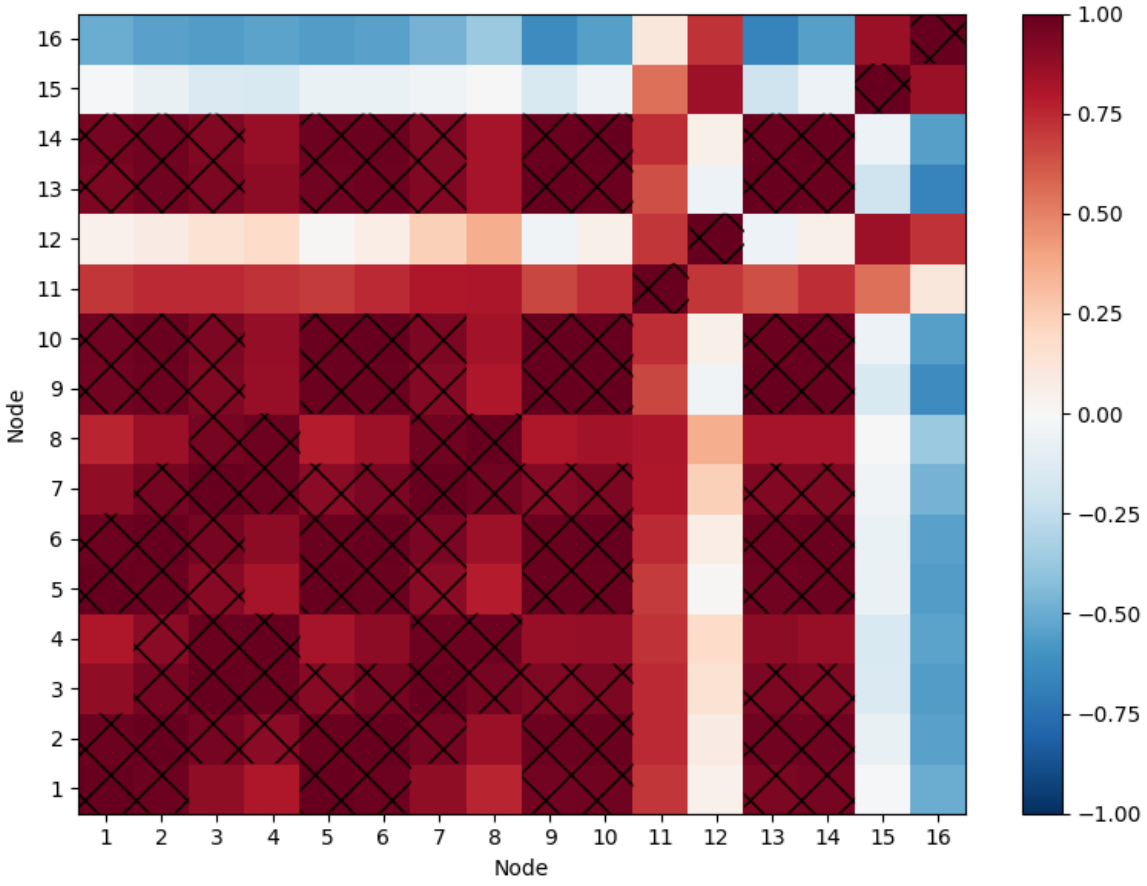
K-means (3 centroids)

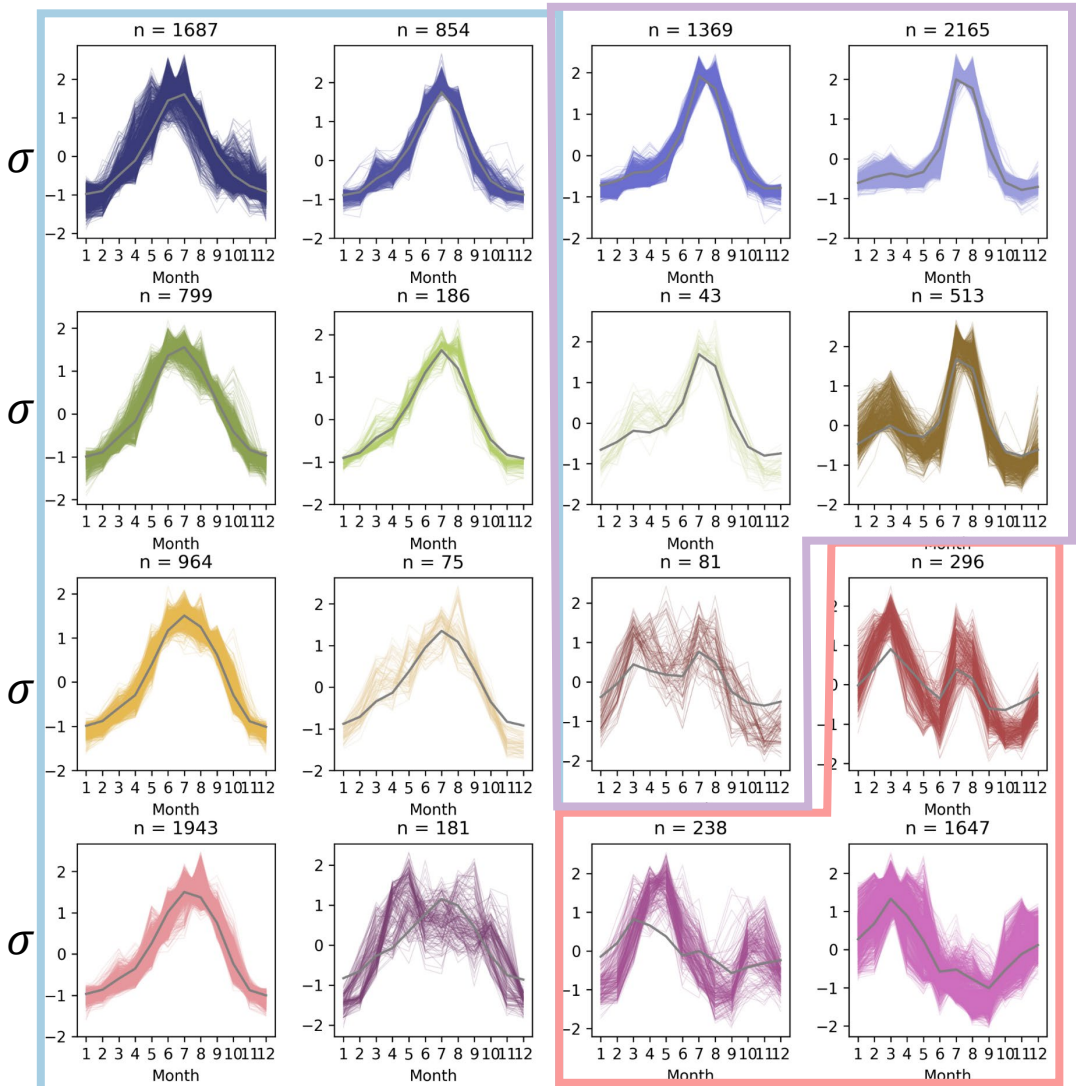


SOM (4 x 4)



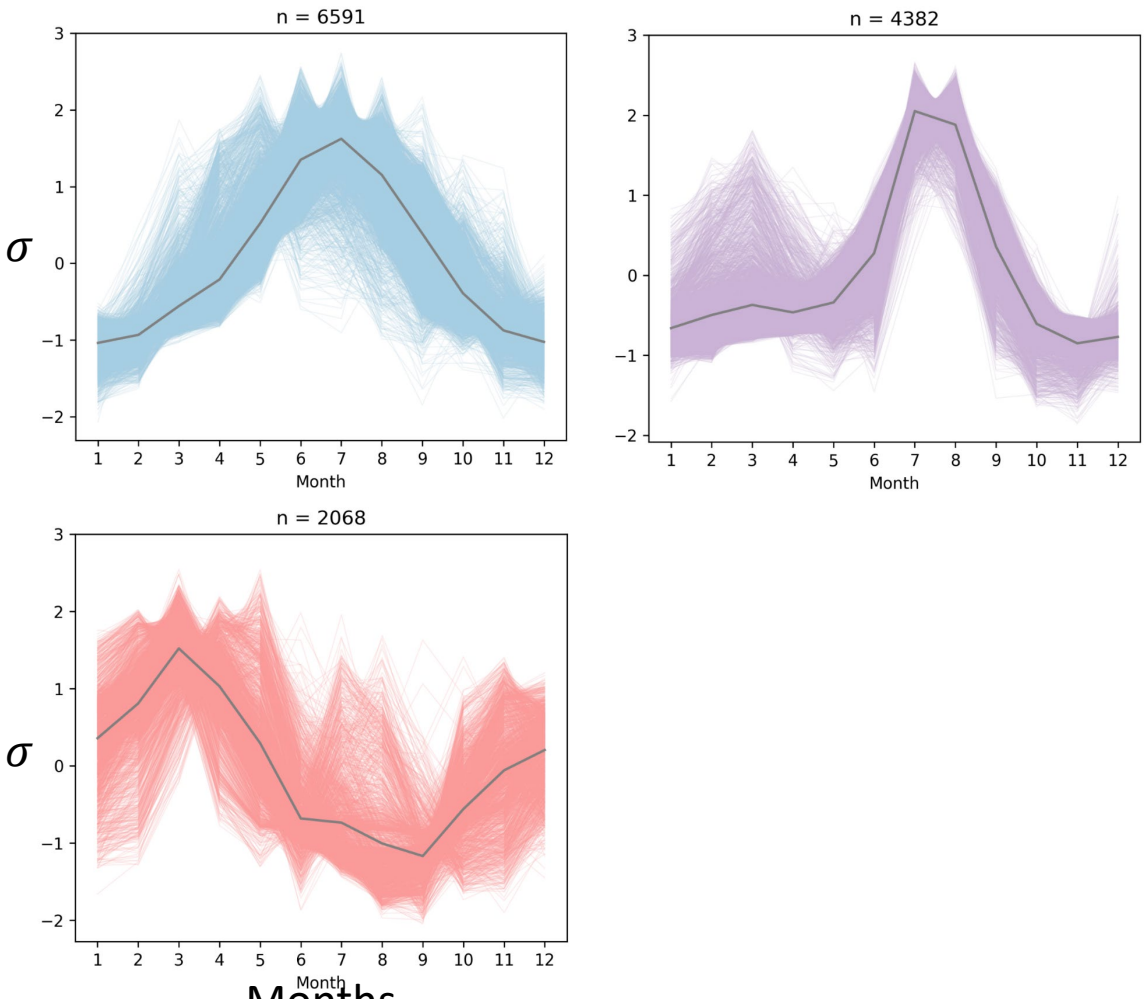
Reducing clusters for SOM



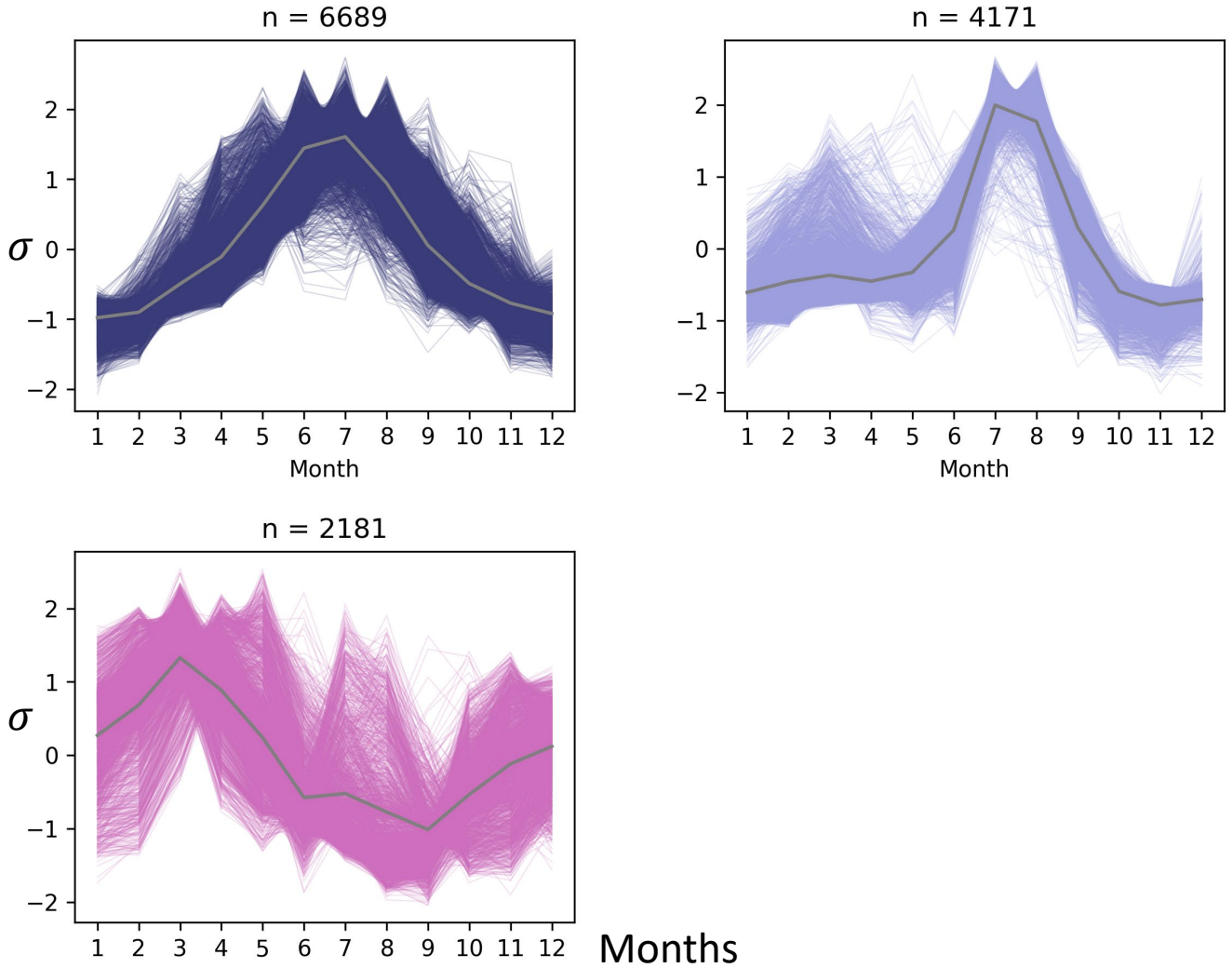


Seasonal cycles of nodes from K-mean and SOM

K-means (3 centroids)

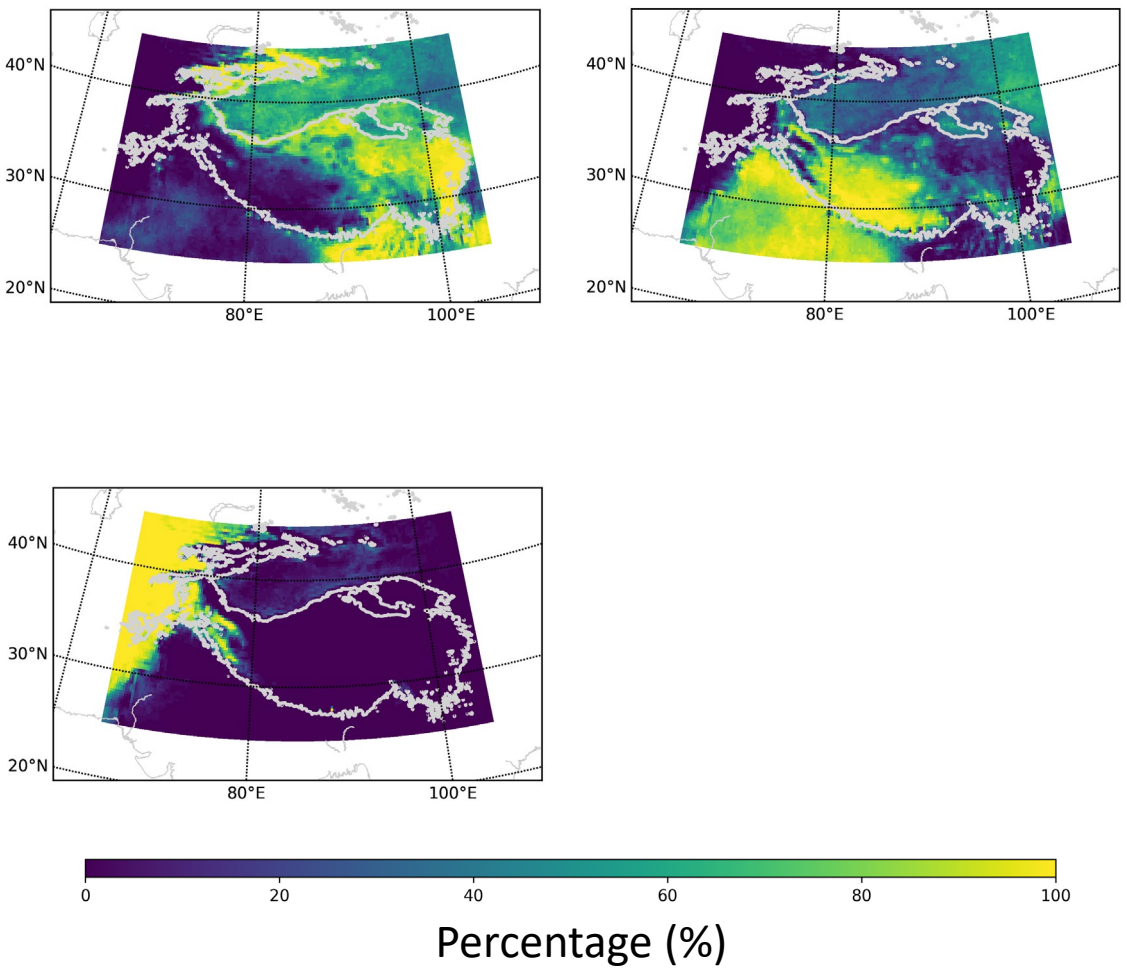


SOM (4 x 4)

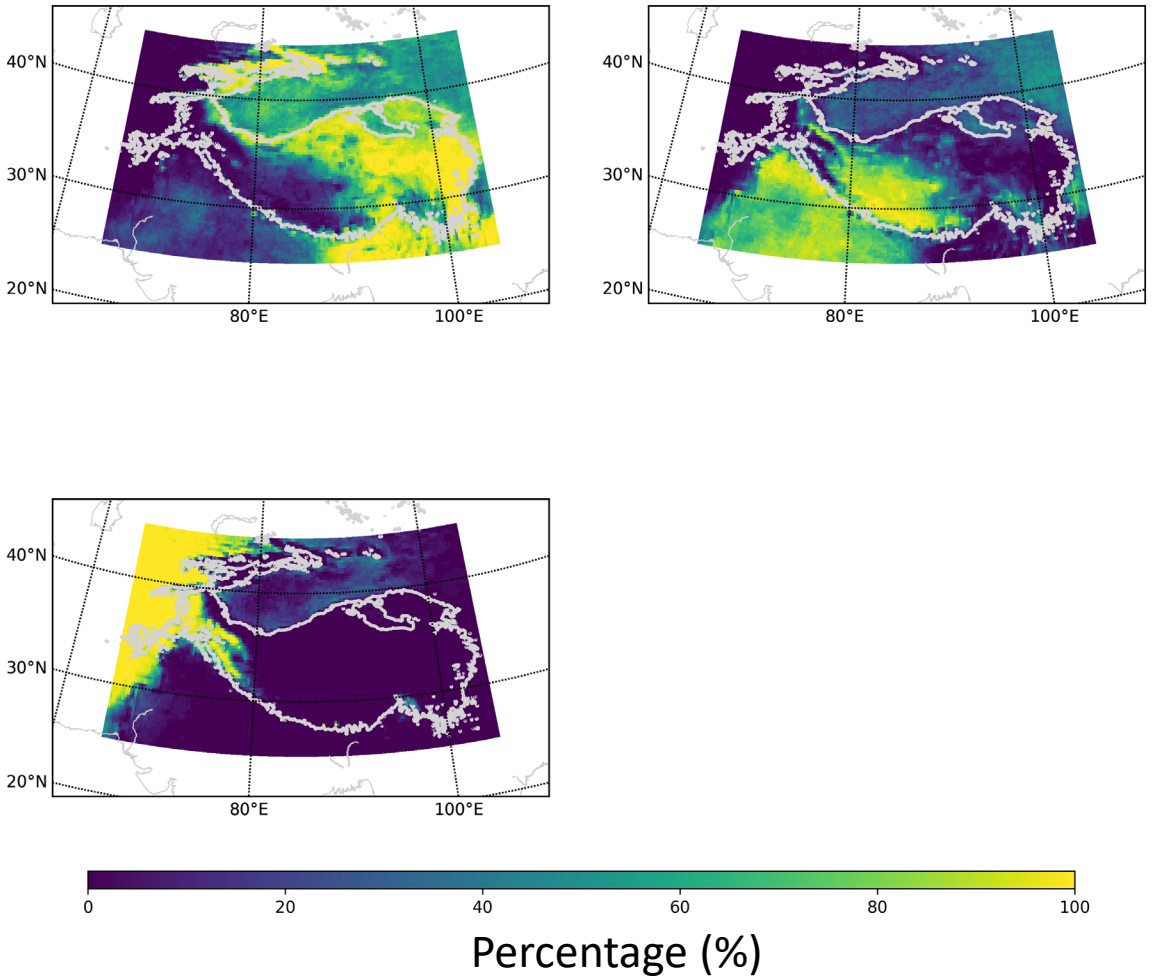


Heatmaps for 40-years prediction

K-means (3 centroids)

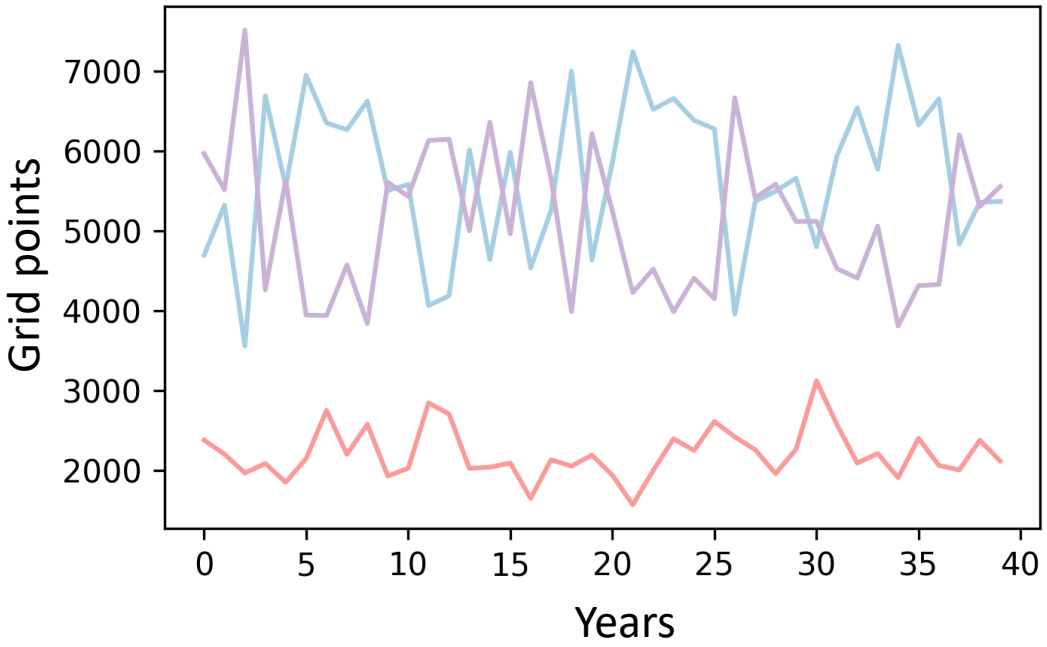


SOM (4 x 4)

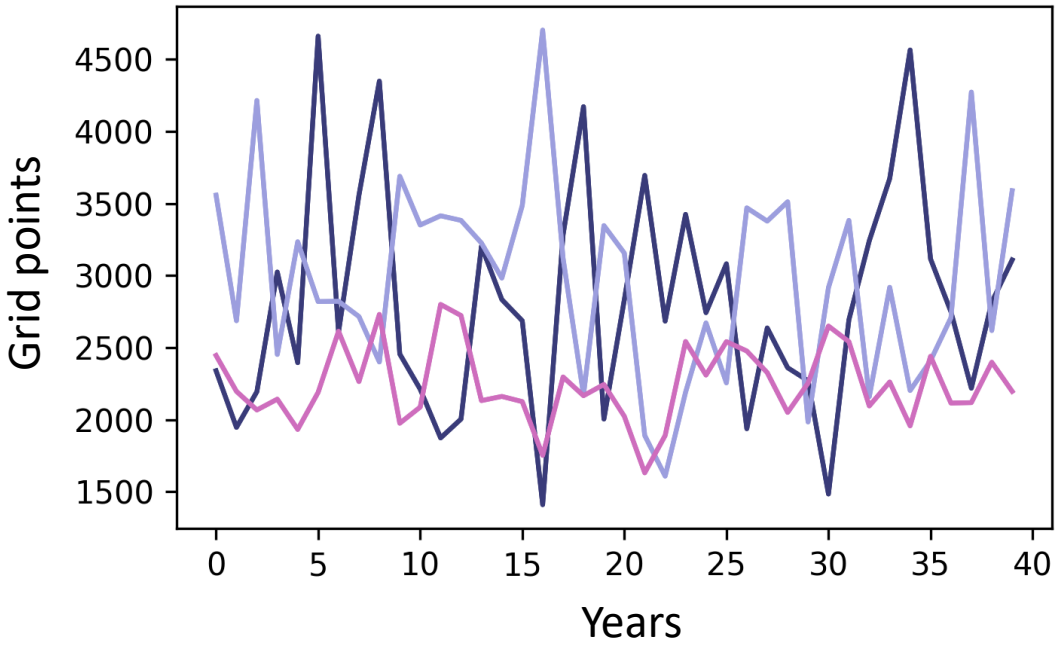


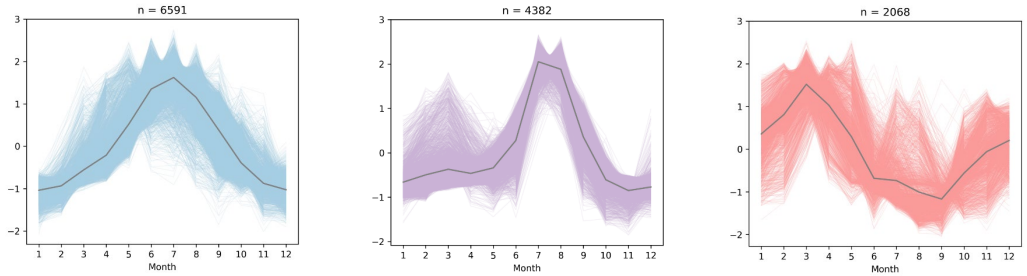
How have the sizes of these regions changed?

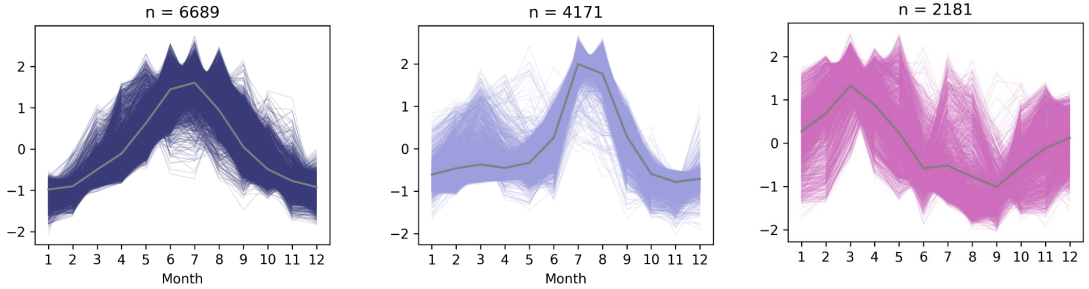
K-means (3 centroids) No significant trend



SOM (4 x 4) No significant trend



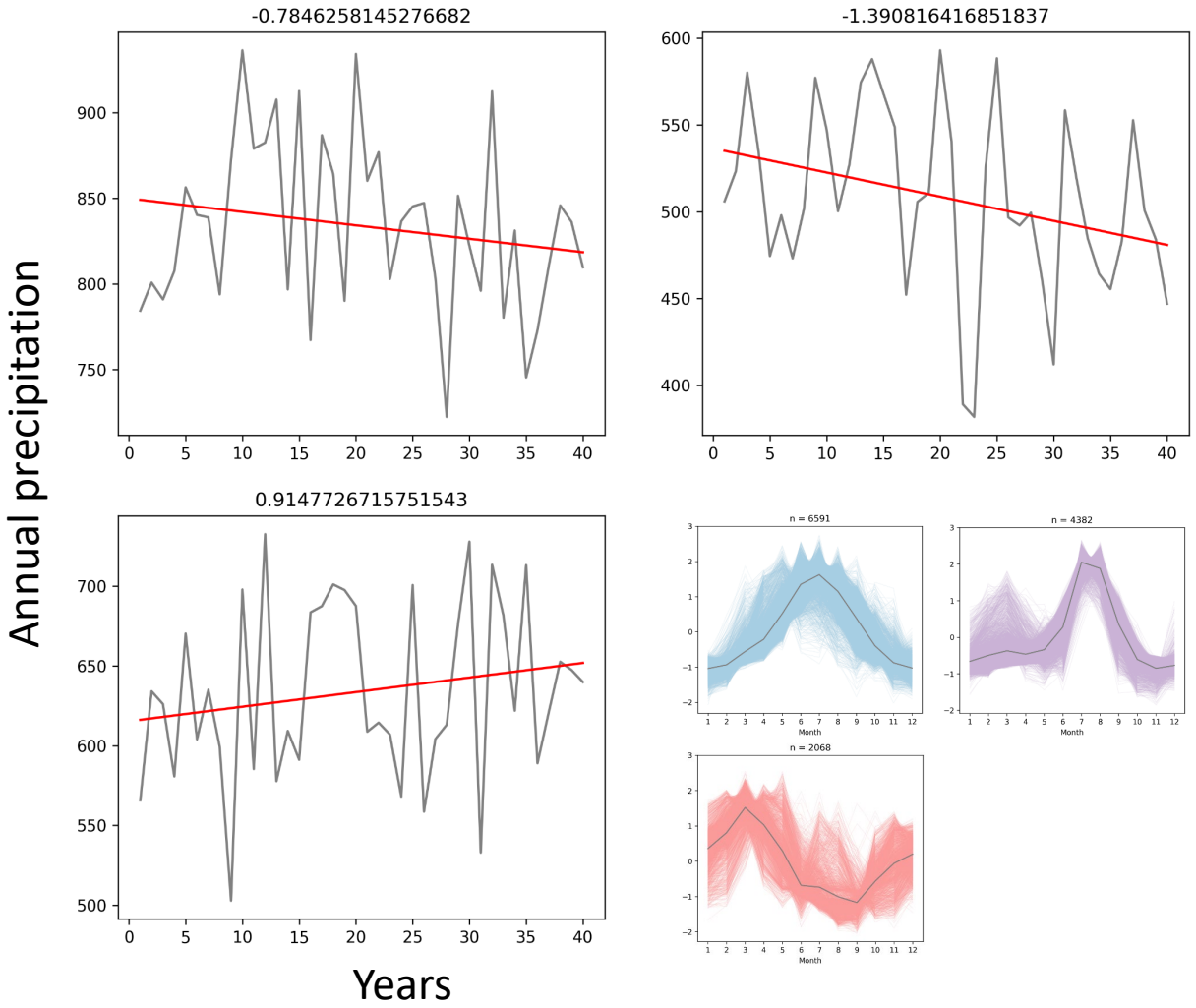


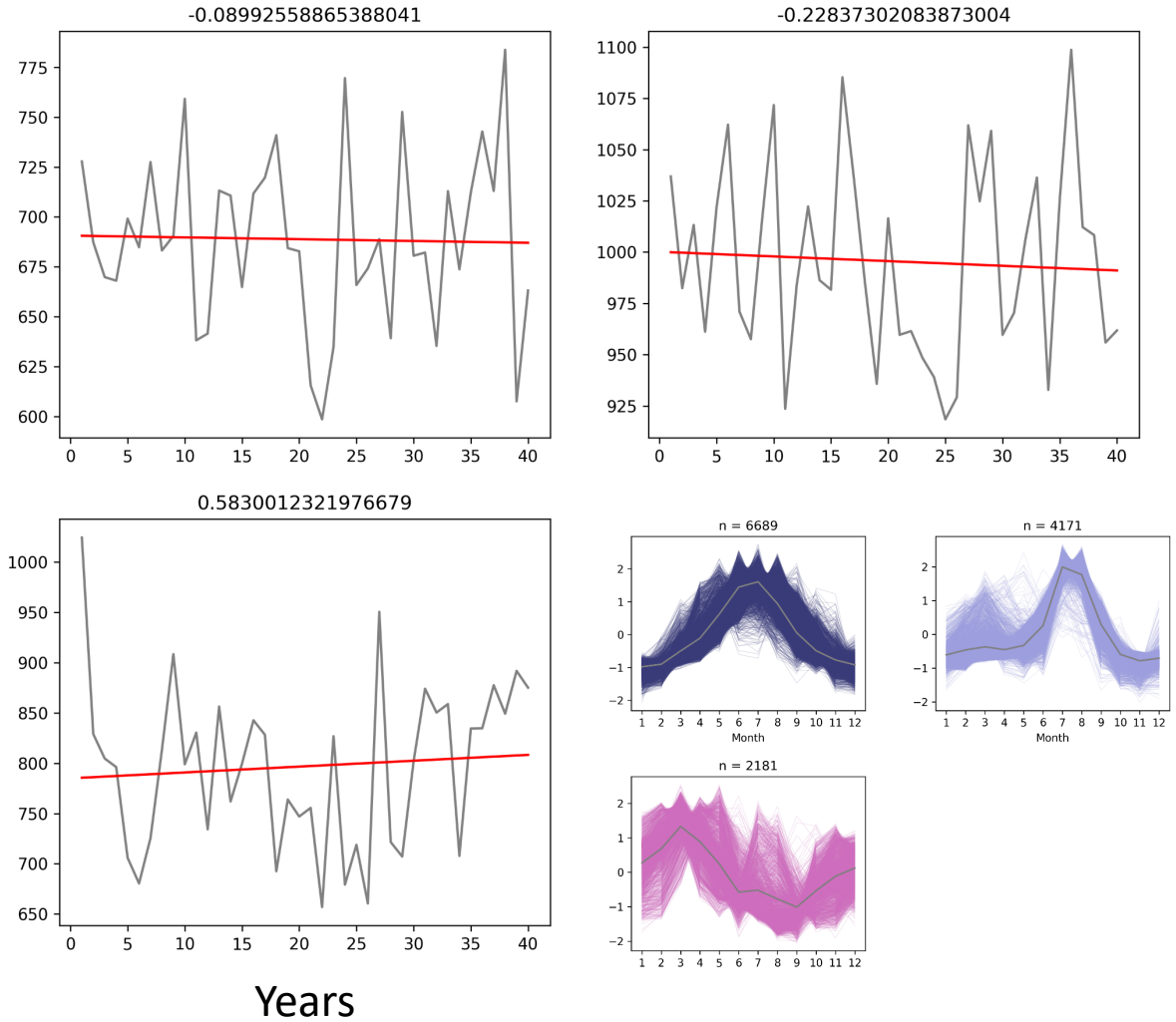


Has the precipitation increased/decreased in these regions?

K-means (3 centroids) No significant trend

SOM (4 x 4) No significant trend





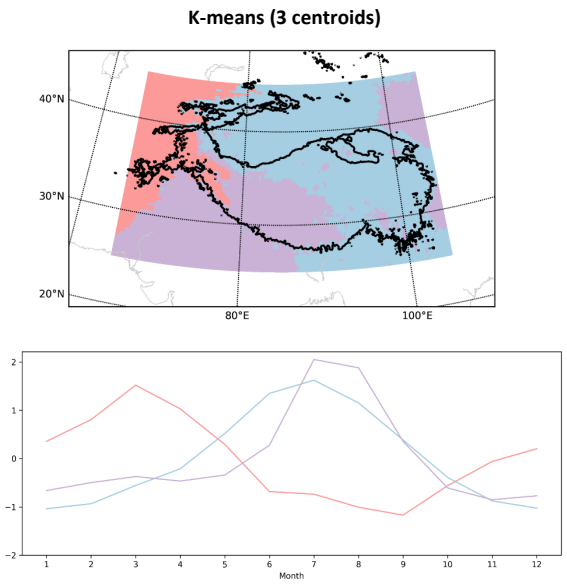
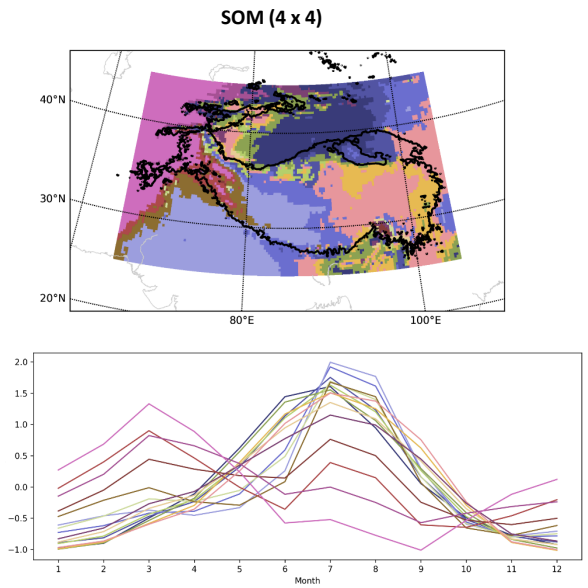
UNIVERSITY OF GOTHENBURG

Summary

- Classifying the regions according to seasonal cycles of precipitation (monsoon and westerly?)
- Quantifying the robustness of each region on interannual timescale
- Presenting the changes in amounts of precipitation and extent for each region over 40 years (no clear conclusion yet)

Issues

- Choice of number of nodes
- Methods of evaluations

Future work

- Compared to satellite observations
- Links to large-scale dynamics