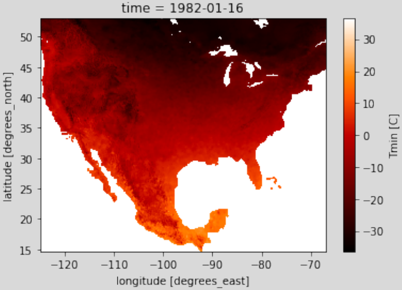
# Robust and interpretable unsupervised machine learning techniques for analyzing the climate system

Derek DeSantis†, Phil Wolfram, Boian Alexandrov, Raviteja Vangara, Erik Skau, Katrina Bennett

March 5, 2021


Figure: OpenAI CoastRunners misspecified reward function




- Gridded climate data set of North America.
- Grid cell is monthly data from 1950-2013, six kilometers across.
- Available variables used: precipitation, maximum temperature, minimum temperature.

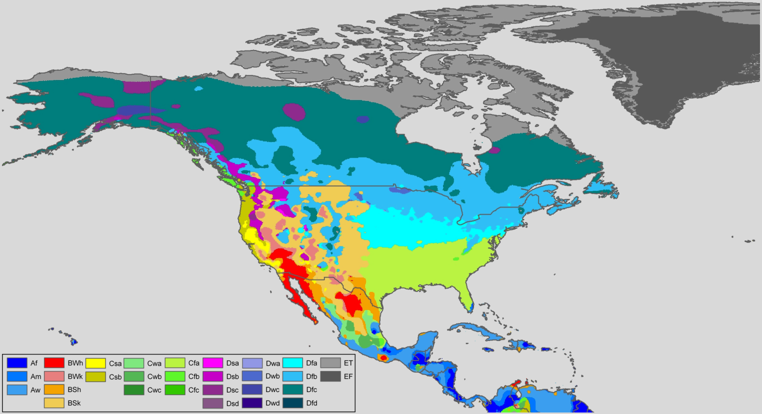



Figure: Köppen-Geiger map of North America (Peel et. al.)

## Problem

- Climate depends on more than temperature and precipitation.

## Problem

- Climate depends on more than temperature and precipitation.
- Can only resolve land.

## Problem

- Climate depends on more than temperature and precipitation.
- Can only resolve land.
- Does not adapt to changing climate.

## Problem

- Climate depends on more than temperature and precipitation.
- Can only resolve land.
- Does not adapt to changing climate.
- The cut-offs in model are, to some extent, arbitrary.

## Problem

- Climate depends on more than temperature and precipitation.
- Can only resolve land.
- Does not adapt to changing climate.
- The cut-offs in model are, to some extent, arbitrary.
- No universal agreement to how many classes there should be.

## Problem

- Dependence on algorithm of choice and hyperparameters.

## Problem

- Dependence on algorithm of choice and hyperparameters.

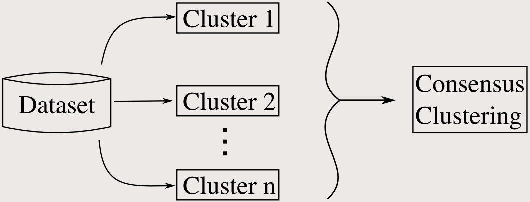



Figure: Many clusterings combined into a single **consensus clustering**.

## Problem

- Dependence on algorithm of choice and hyperparameters.

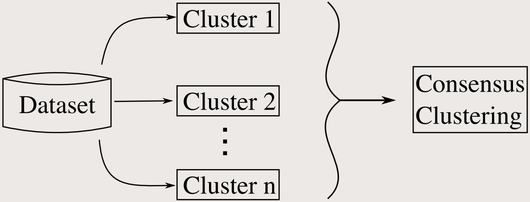



Figure: Many clusterings combined into a single **consensus clustering**.

- Clustering ill-posed - lack measurement of “trust”.

## Problem

- Dependence on algorithm of choice and hyperparameters.

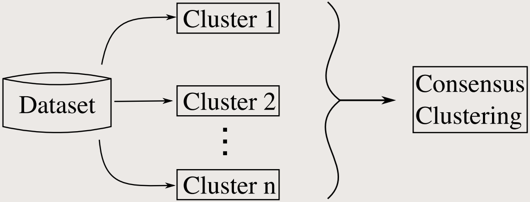



Figure: Many clusterings combined into a single **consensus clustering**.

- Clustering ill-posed - lack measurement of “trust”.
- Dependence on “hidden parameters” - **scale of data**.

## Solution

1. Leverage discrete wavelet transform to classify across a multitude of scales.

## Solution

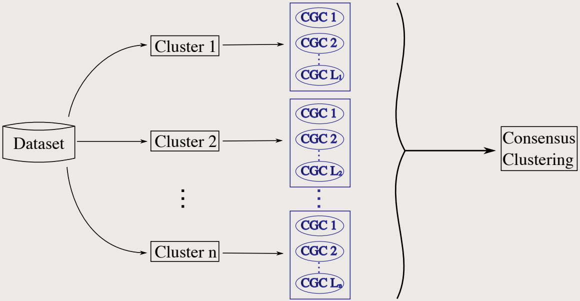
1. Leverage discrete wavelet transform to classify across a multitude of scales.
2. Use information theory to discover most important scales to classify on.

## Solution

1. Leverage discrete wavelet transform to classify across a multitude of scales.
2. Use information theory to discover most important scales to classify on.
3. Taking these scales, combine classifications to produce a **fuzzy** clustering that assess the trust at each point.

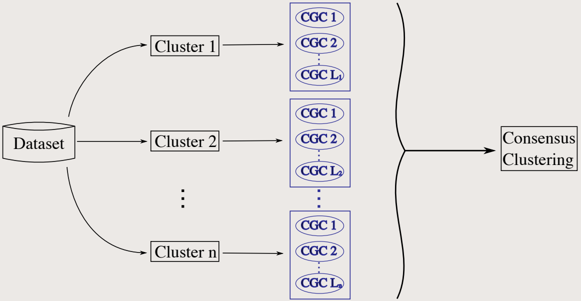
# Solution

1. Leverage discrete wavelet transform to classify across a multitude of scales.
2. Use information theory to discover most important scales to classify on.
3. Taking these scales, combine classifications to produce a **fuzzy** clustering that assess the trust at each point.



## Solution

1. **Leverage discrete wavelet transform to classify across a multitude of scales.**
2. Use information theory to discover most important scales to classify on.
3. Taking these scales, combine classifications to produce a fuzzy clustering that assess the trust at each point.



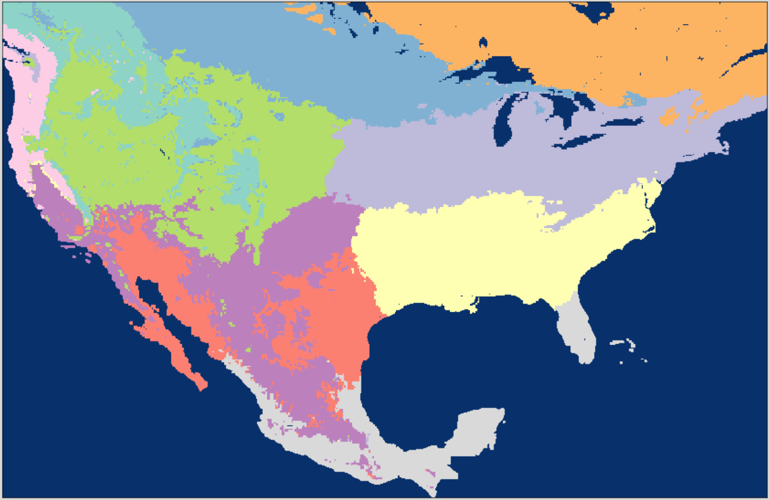


Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (1, 1)$

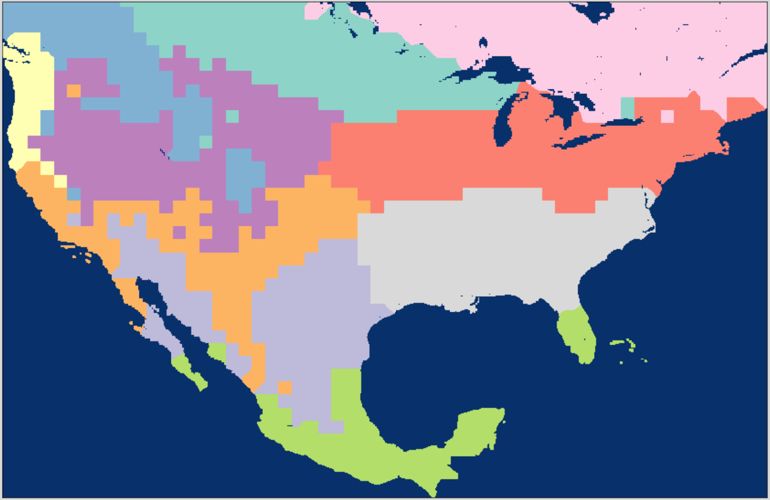


Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (4, 1)$

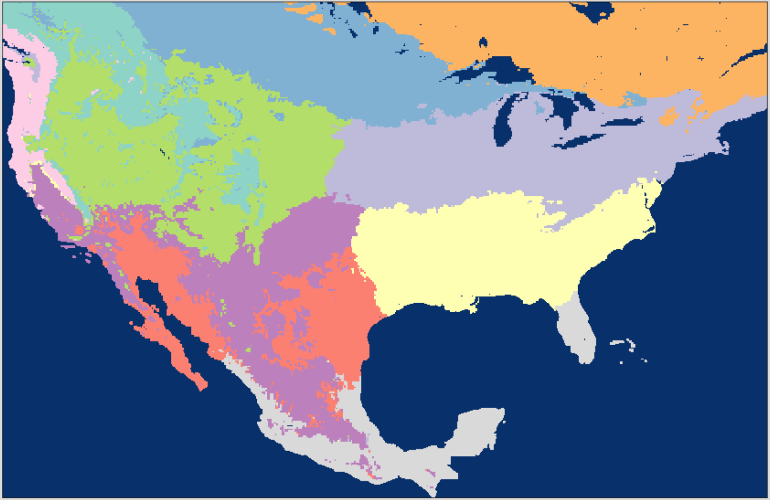


Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (1, 1)$

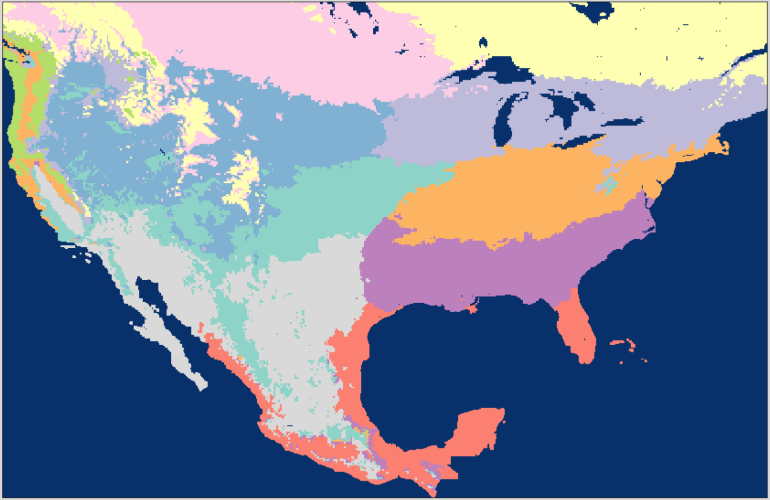


Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (1, 6)$

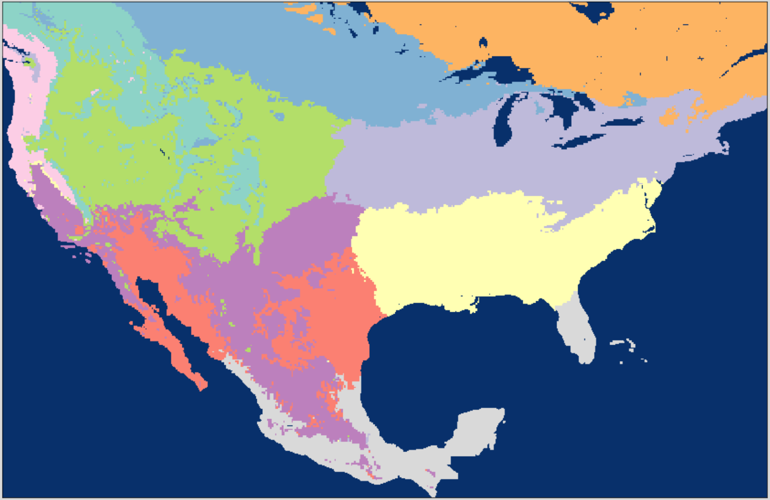


Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (1, 1)$

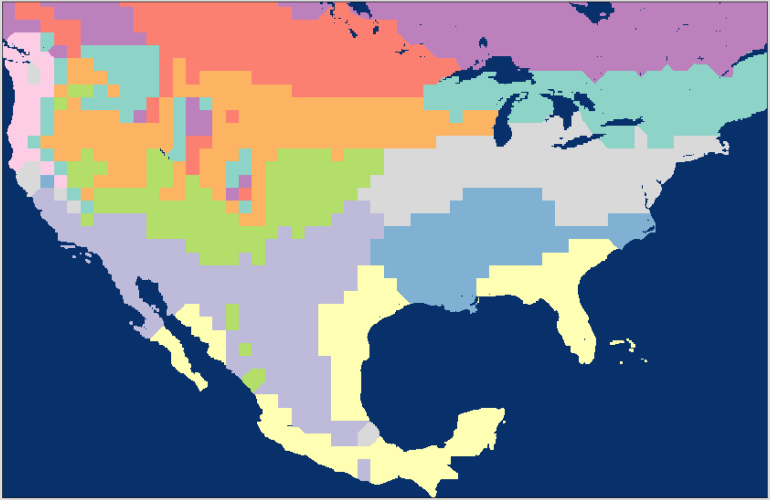
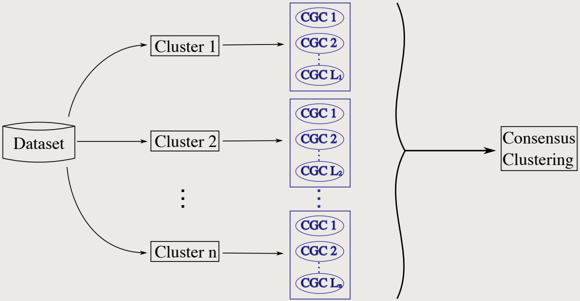
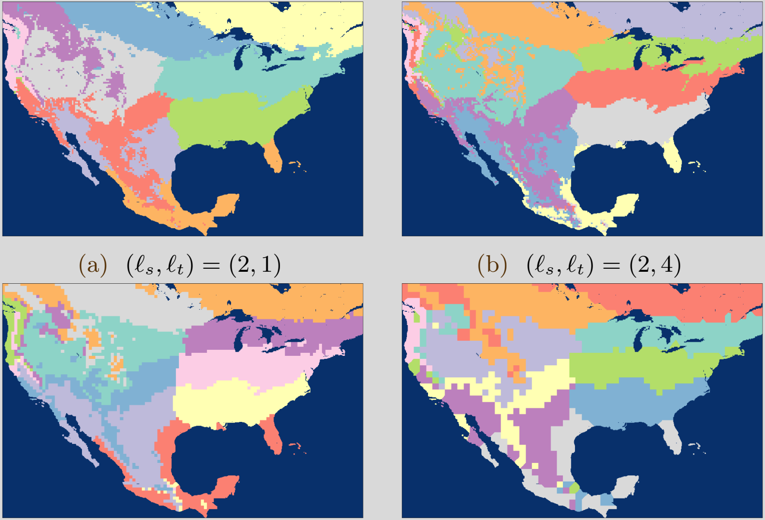


Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (4, 6)$

## Solution

1. Leverage discrete wavelet transform to classify across a multitude of scales.
2. **Use information theory to discover most important scales to classify on.**
3. Taking these scales, combine classifications to produce a fuzzy clustering that assess the trust at each point.





(a) $(\ell_s, \ell_t) = (2, 1)$

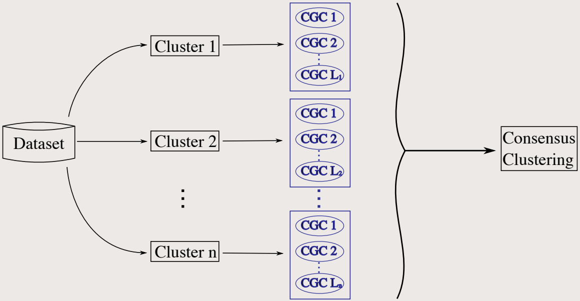
(b) $(\ell_s, \ell_t) = (2, 4)$

(c) $(\ell_s, \ell_t) = (3, 5)$

(d) $(\ell_s, \ell_t) = (4, 4)$

## Solution

1. Leverage discrete wavelet transform to classify across a multitude of scales.
2. Use information theory to discover most important scales to classify on.
3. **Taking these scales, combine classifications to produce a fuzzy clustering that assess the trust at each point.**



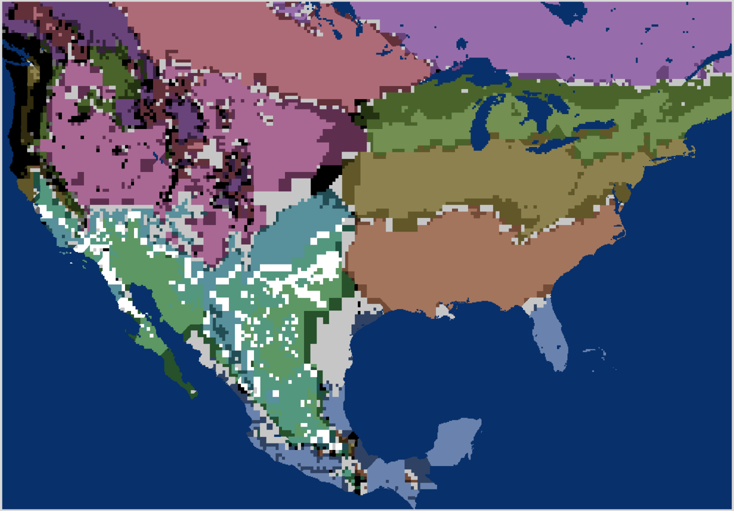
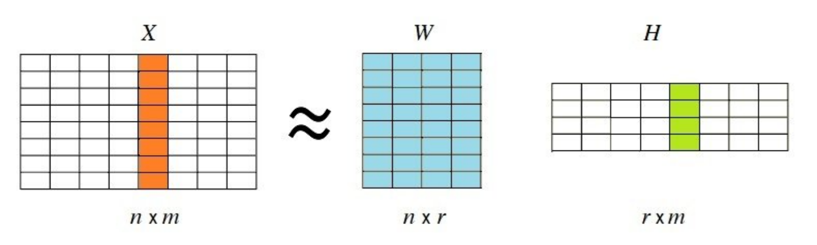


Figure: Consensus clustering from reduced ensemble of clusters for k=10, along with the trust. Grey = multi-class. Darker hue = lower trust.


X
W
H
≈
n x m
n x r
r x m

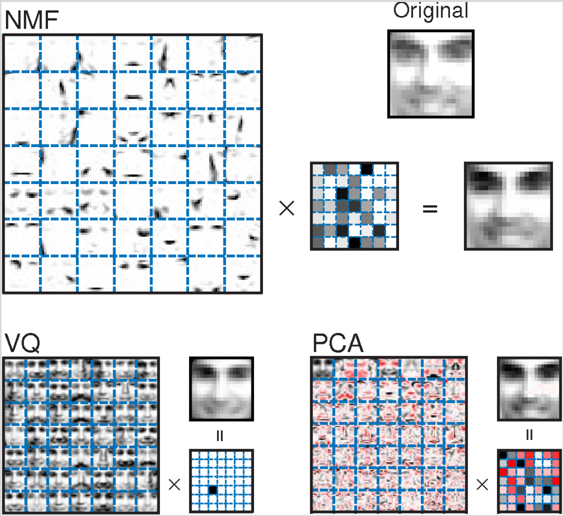
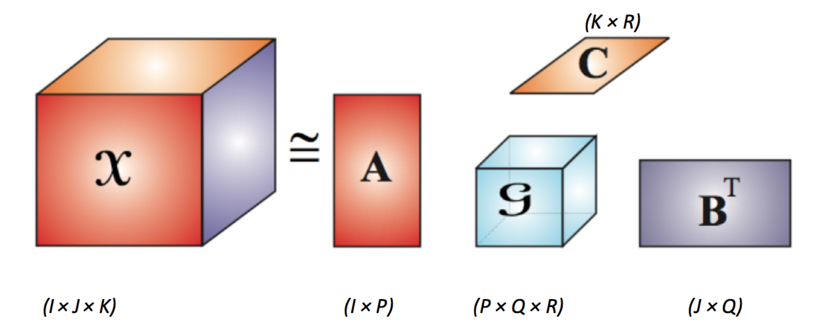
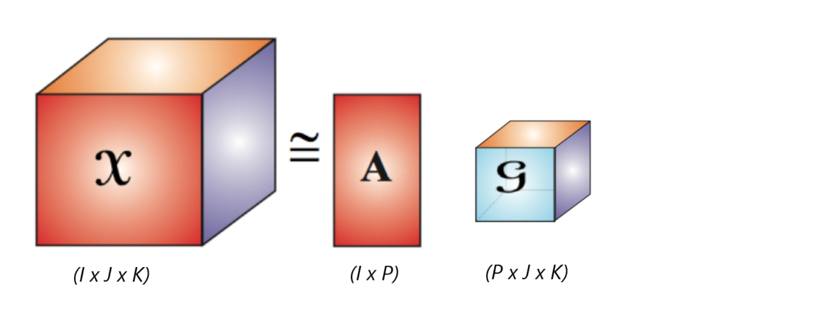



Figure: NMF versus other matrix decompositions (Lee, Seung)


$\mathcal{X}$
$\cong$
A
C
$\mathcal{G}$
$B^T$
(K × R)
(I × J × K)
(I × P)
(P × Q × R)
(J × Q)


$\mathcal{X}$
$\cong$
A
$\mathcal{G}$
(I x J x K)
(I x P)
(P x J x K)

## Problem

- Increasing the number of hidden variables reduces reconstruction error

## Problem

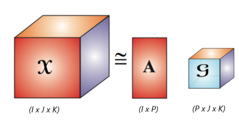
- Increasing the number of hidden variables reduces reconstruction error
- More hidden variables is harder to interpret

## Problem

- Increasing the number of hidden variables reduces reconstruction error
- More hidden variables is harder to interpret
- At a certain point, one is fitting noise and not signal

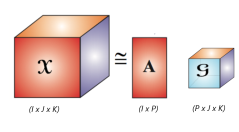
## Summary

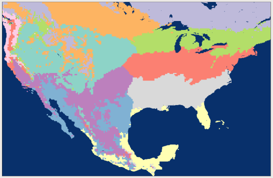
- NTF is finding interpretable climate signals



## Summary

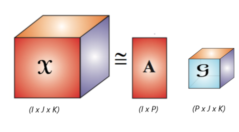
- NTF is finding interpretable climate signals
- As seen with clustering, scale is playing a role that we need to analyze



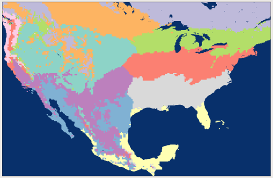



## Summary

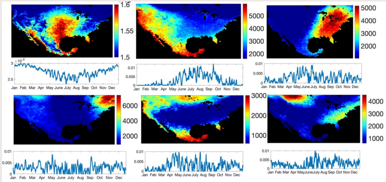
- NTF is finding interpretable climate signals




- As seen with clustering, scale is playing a role that we need to analyze



- Can we discover latent signatures of El Nino/La Nina?



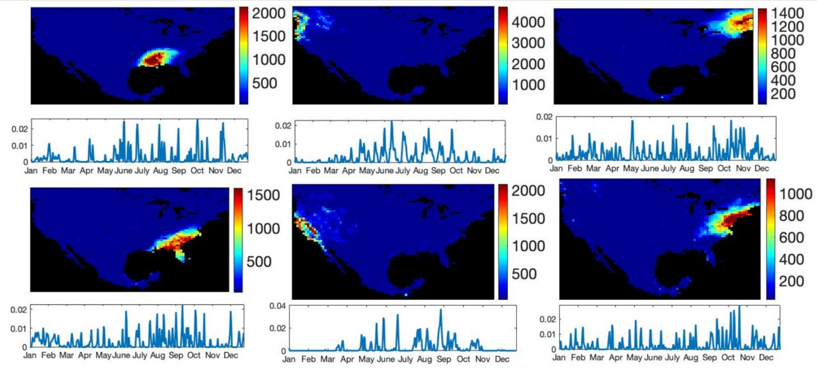



Figure: 1982 Precipitation Modes

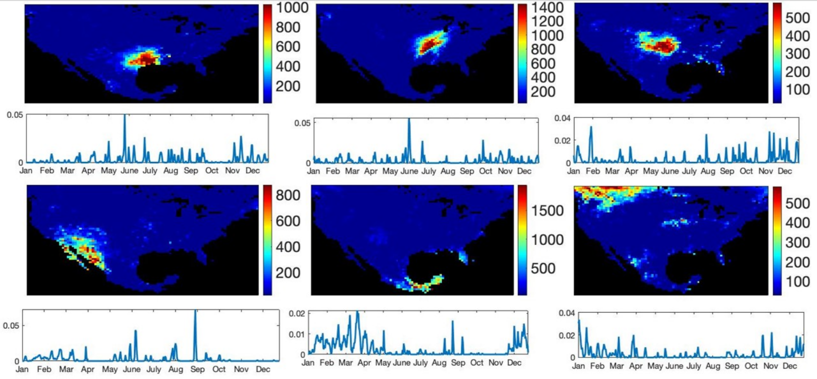



Figure: 1982 Precipitation Modes

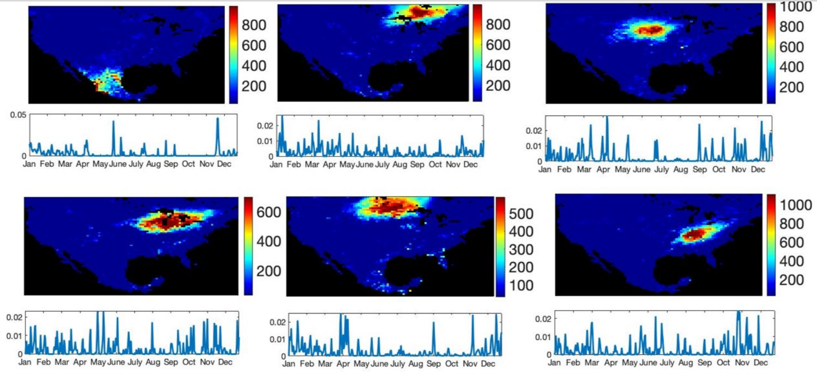



Figure: 1982 Precipitation Modes

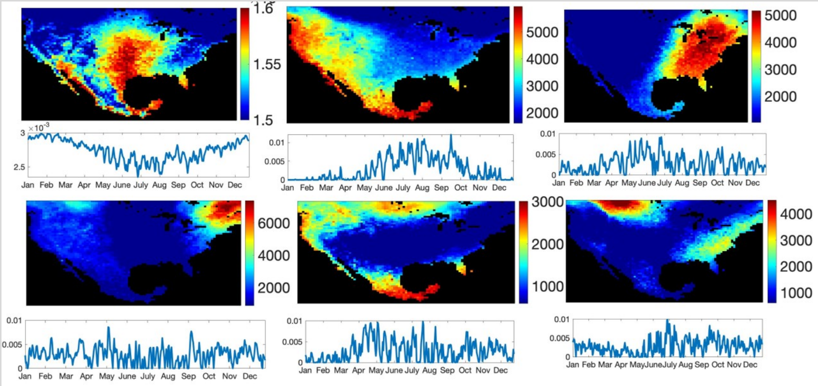



Figure: 1982 Temperature Modes

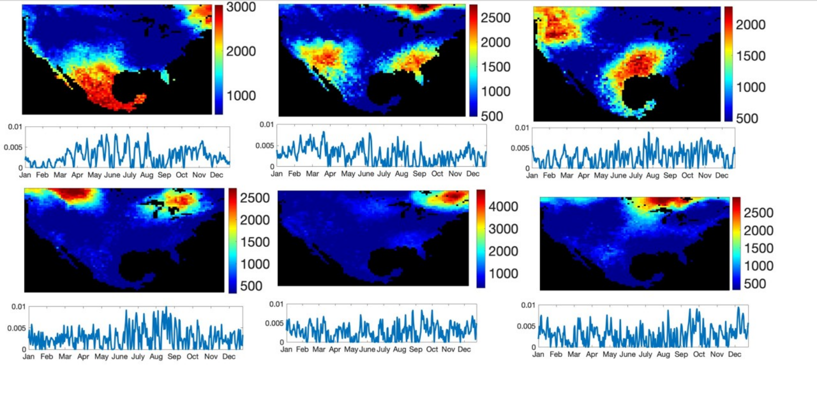
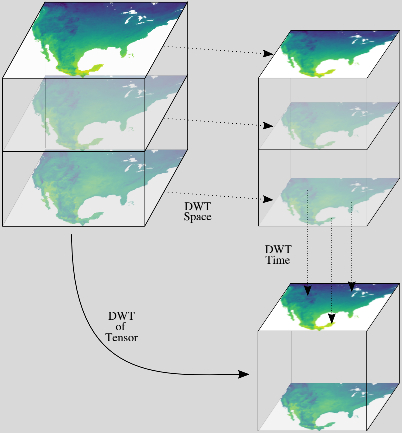



Figure: 1982 Temperature Modes

- The DWT splits a signal into high and low frequency
- Low temporal signal captures climatology (seasons, years, decades), while low spatial signal captures regional features(city, county, state).



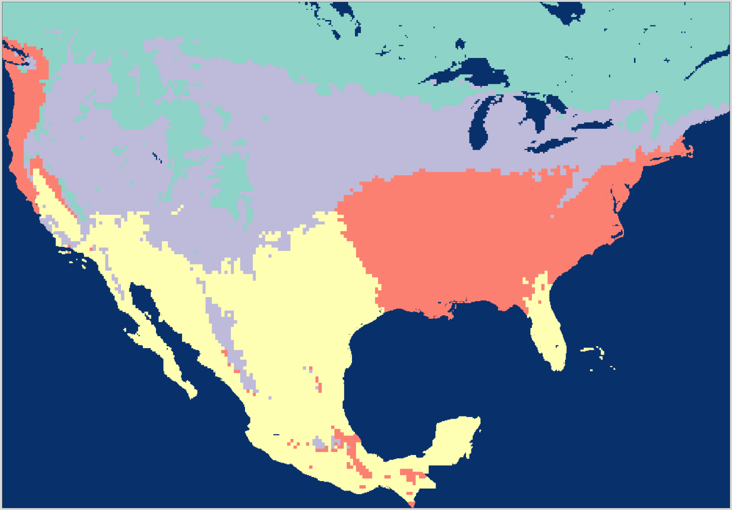


Figure: CGC: K-means $k = 4$, $(\ell_s, \ell_t) = (2, 3)$

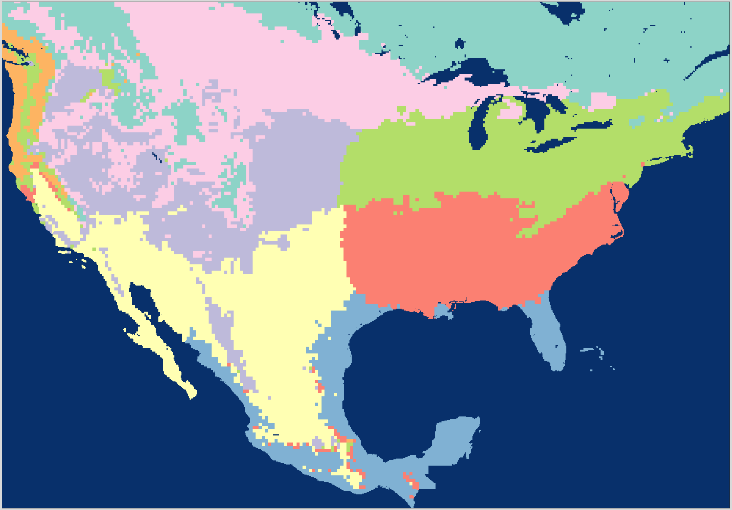


Figure: CGC: K-means $k = 8$, $(\ell_s, \ell_t) = (2, 3)$

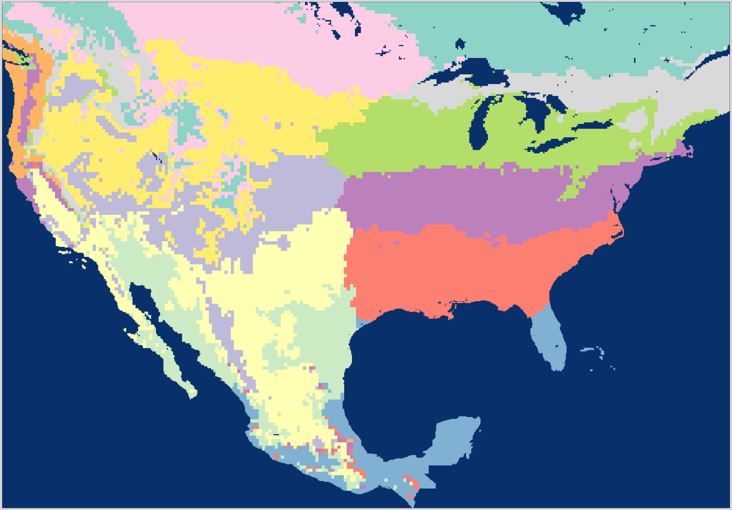


Figure: CGC: K-means $k = 12$, $(\ell_s, \ell_t) = (2, 3)$

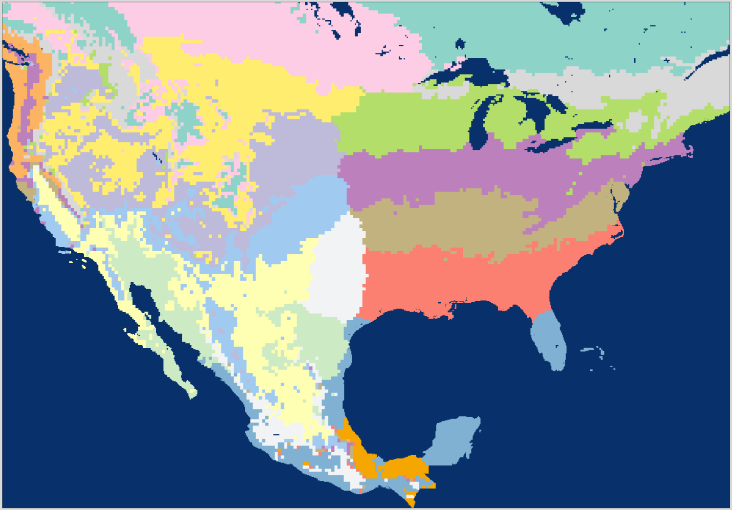


Figure: CGC: K-means $k = 16$, $(\ell_s, \ell_t) = (2, 3)$