Multiresolution Cluster Analysis—Addressing Trust in Climate Classifications

Derek DeSantis[†], Phil Wolfram, Boian Alexandrov

Los Alamos National Laboratory, Center for Nonlinear Studies[†]

AMS Annual Meeting, January 2020

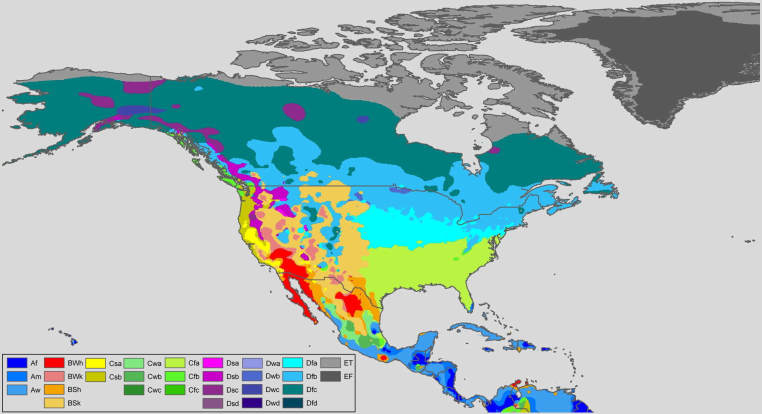


Figure: Köppen-Geiger map of North America (Peel et. al.)

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.

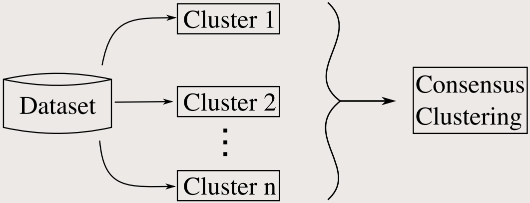


Figure: Many clusterings combined into a single **consensus clustering**.

Problem

- Dependence on algorithm of choice and hyperparameters.

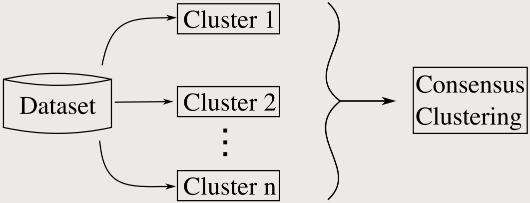


Figure: Many clusterings combined into a single **consensus clustering**.

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

Problem

- Dependence on algorithm of choice and hyperparameters.

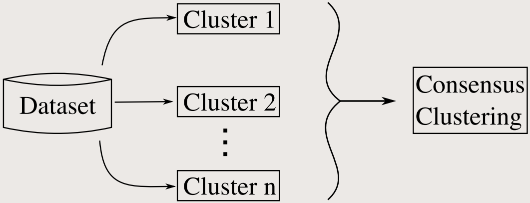


Figure: Many clusterings combined into a single **consensus clustering**.

- 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

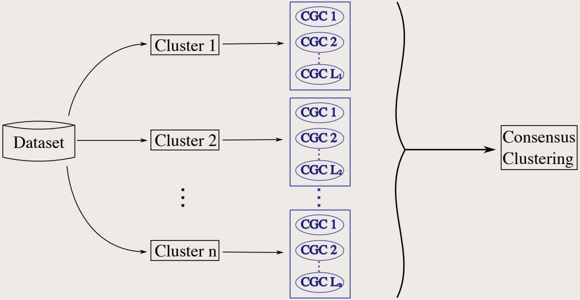
1. Leverage discrete wavelet transform to classify across a multitude of scales.
2. Use information theory to discover most important scales to classify on.

Solution

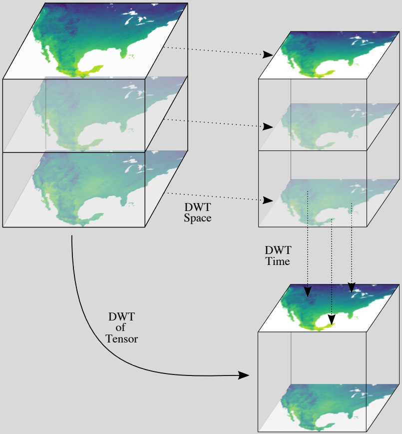
1. Leverage discrete wavelet transform to classify across a multitude of scales.
2. Use information theory to discover most important scales to classify on.
3. Taking these scales, combine classifications to produce a **fuzzy** clustering that assess the trust at each point.

Solution

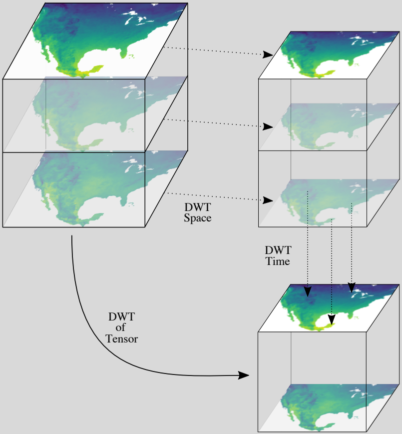
1. Leverage discrete wavelet transform to classify across a multitude of scales.
2. Use information theory to discover most important scales to classify on.
3. Taking these scales, combine classifications to produce a **fuzzy** clustering that assess the trust at each point.



- 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).

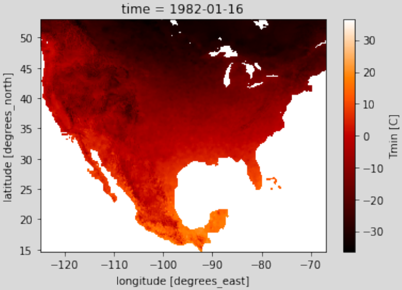


- The DWT splits a signal into high and low frequency
- Low temporal signal captures climatology (seasons, years, decades), while low spatial signal captures regional features(city, county, state).



Definition

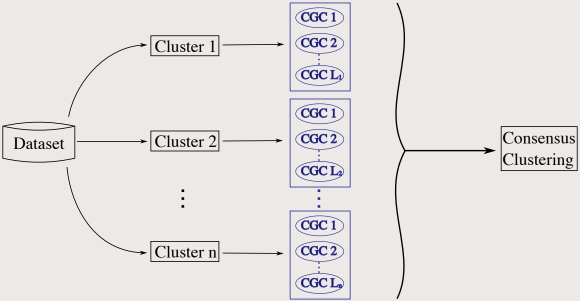
Given partitions of data $U = \{U_j\}_{j=1}^{k}, V = \{V_j\}_{j=1}^{l}$, the **Mutual Information** $\mathcal{NI}(U, V)$ measures how knowledge of one clustering reduces our uncertainty of the other.

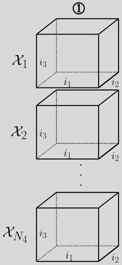


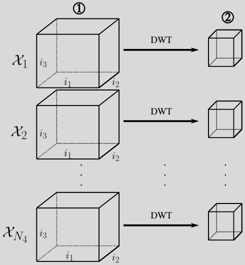
- Gridded climate data set of North America.
- Grid cell is monthly data from 1950-2013, six kilometers across.
- Available variables used: precipitation, maximum temperature, minimum temperature.

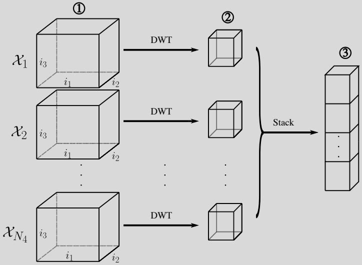
Solution

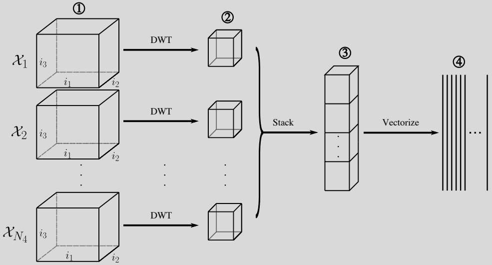
1. **Leverage discrete wavelet transform to classify across a multitude of scales.**
2. Use information theory to discover most important scales to classify on.
3. Taking these scales, combine classifications to produce a fuzzy clustering that assess the trust at each point.

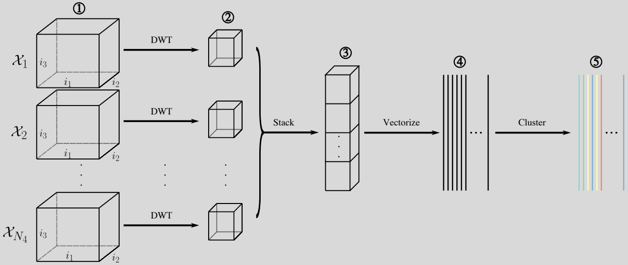


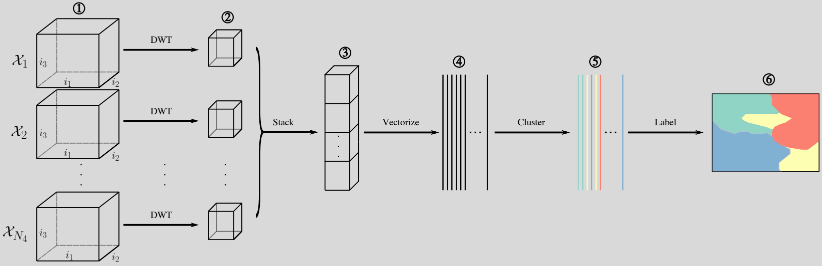

①
$\mathcal{X}_1$
i_3
i_1
i_2
$\mathcal{X}_2$
i_3
i_1
i_2
$\mathcal{X}_{N_4}$
i_3
i_1
i_2


①
②
$\mathcal{X}_1$
$\mathcal{X}_2$
$\mathcal{X}_{N_4}$
i_1
i_2
i_3
DWT


①
②
③
$\mathcal{X}_1$
$\mathcal{X}_2$
$\mathcal{X}_{N_4}$
i_1
i_2
i_3
DWT
Stack


①
②
③
④
$\mathcal{X}_1$
$\mathcal{X}_2$
$\mathcal{X}_{N_4}$
i_1
i_2
i_3
DWT
Stack
Vectorize


①
②
③
④
⑤
$\mathcal{X}_1$
$\mathcal{X}_2$
$\mathcal{X}_{N_4}$
i_1
i_2
i_3
DWT
Stack
Vectorize
Cluster


①
②
③
④
⑤
⑥
$\mathcal{X}_1$
$\mathcal{X}_2$
$\mathcal{X}_{N_4}$
i_1
i_2
i_3
DWT
Stack
Vectorize
Cluster
Label

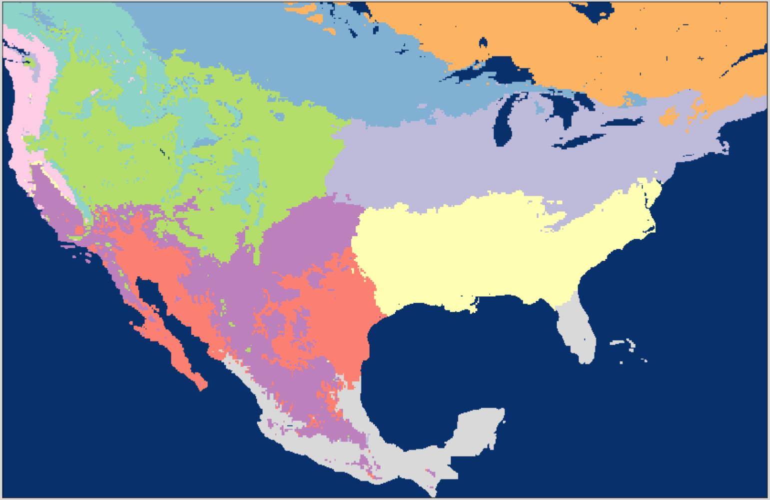


Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (1, 1)$

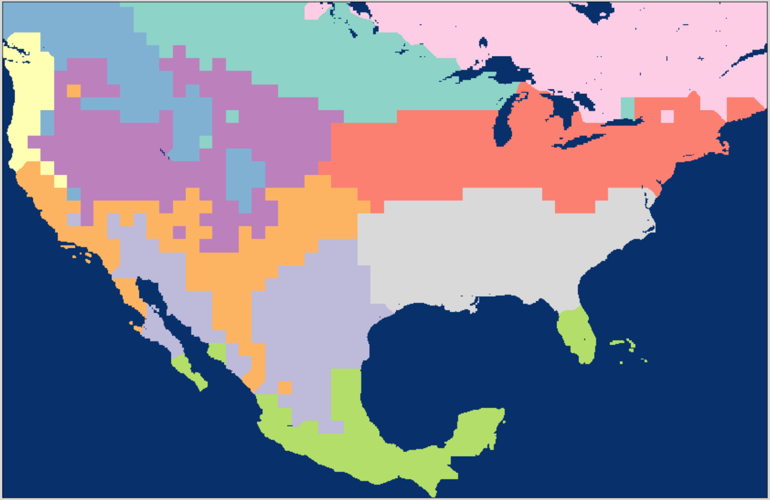


Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (4, 1)$

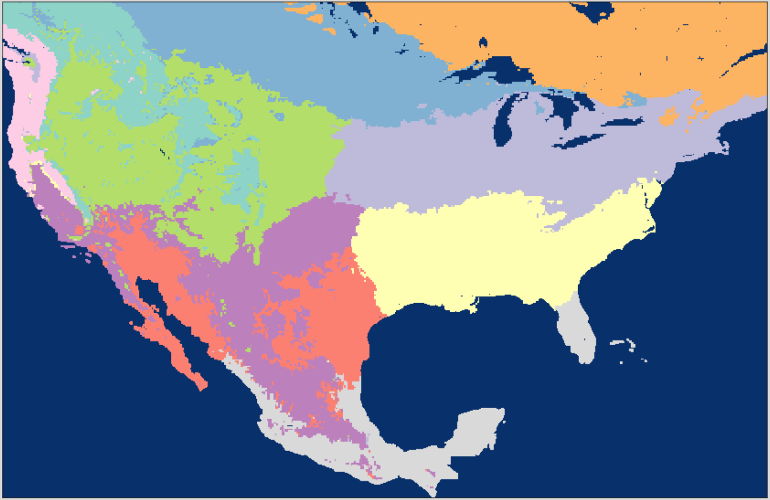


Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (1, 1)$

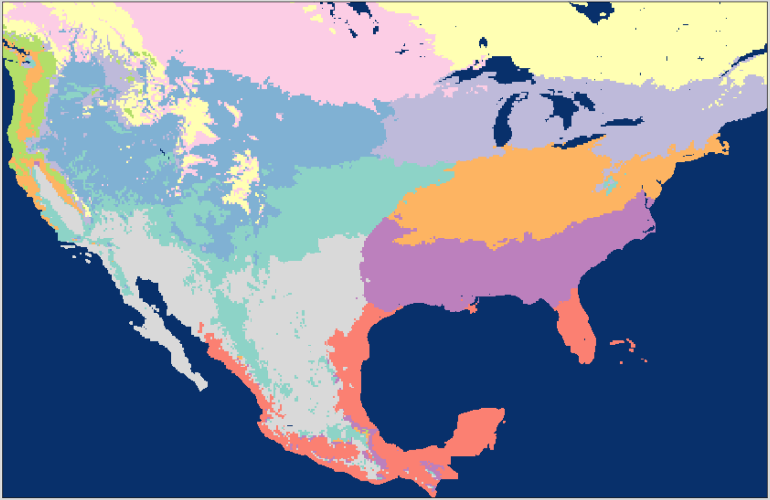


Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (1, 6)$

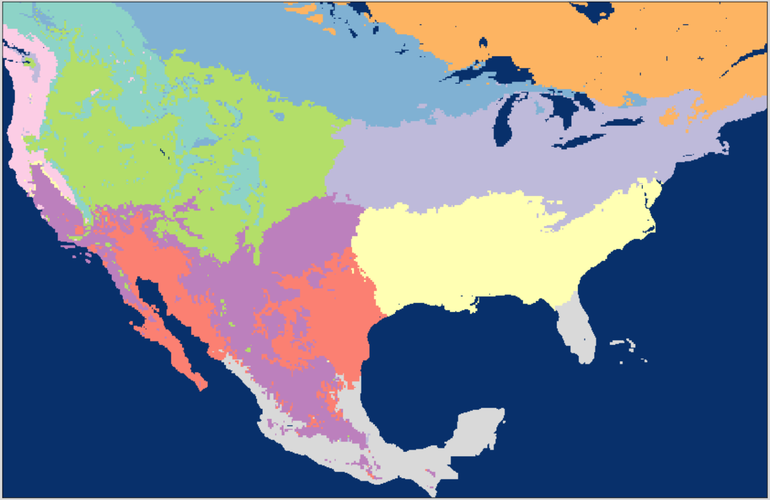


Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (1, 1)$

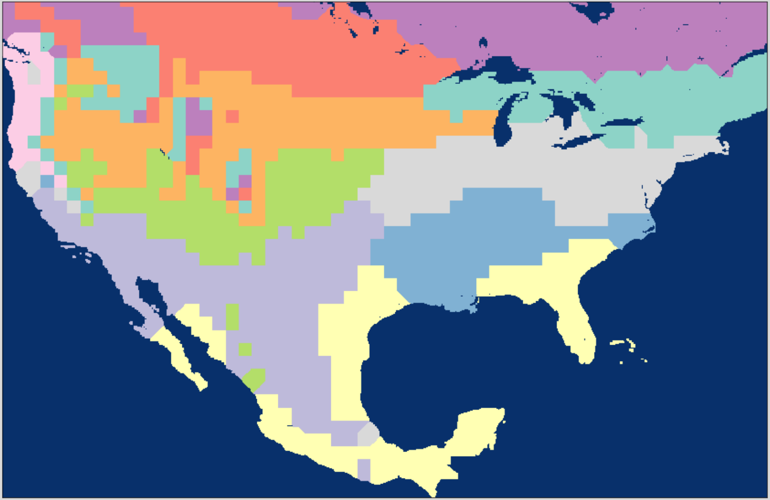
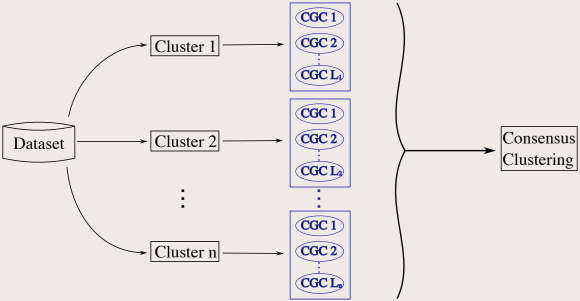
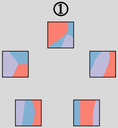


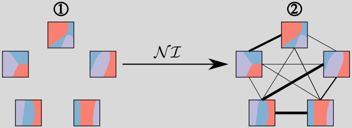
Figure: CGC: K-means $k = 10$, $(\ell_s, \ell_t) = (4, 6)$

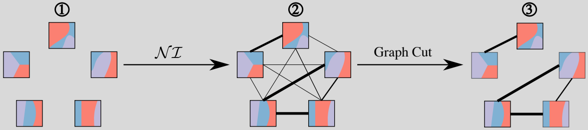
Solution

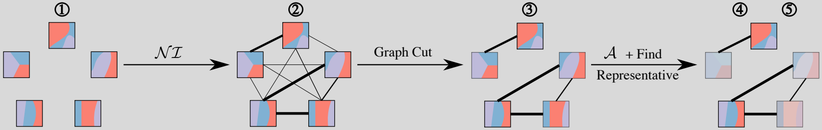
1. Leverage discrete wavelet transform to classify across a multitude of scales.
2. **Use information theory to discover most important scales to classify on.**
3. Taking these scales, combine classifications to produce a fuzzy clustering that assess the trust at each point.








①
$\mathcal{NI}$
②
Graph Cut
③


①
②
③
④
⑤
$\mathcal{NI}$
Graph Cut
$\mathcal{A}$ + Find
Representative

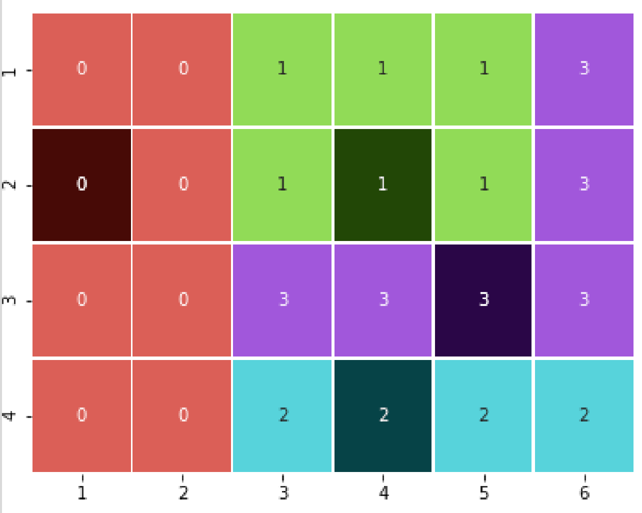
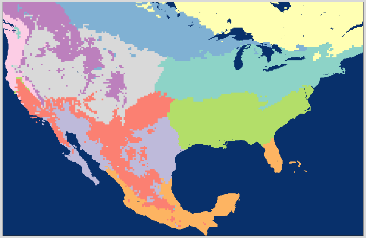
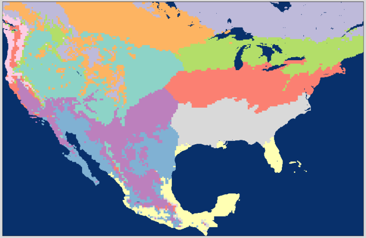


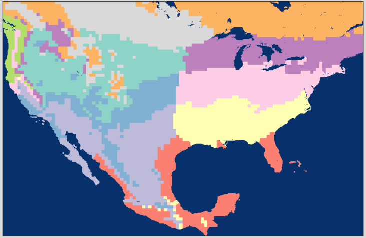
Figure: Results from graph cut algorithm. The highlighted resolutions are the final ensemble. Vertical number = l_s, horzontal bar = l_t.



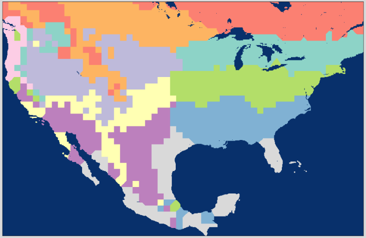
(a) $(\ell_s, \ell_t) = (2, 1)$



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



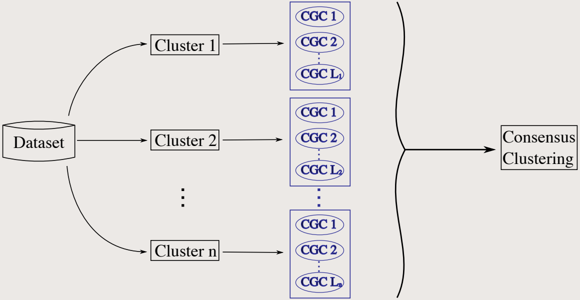
(c) $(\ell_s, \ell_t) = (3, 5)$



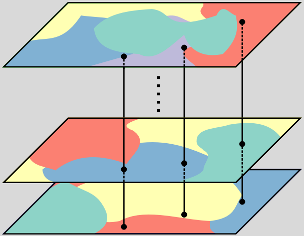
(d) $(\ell_s, \ell_t) = (4, 4)$

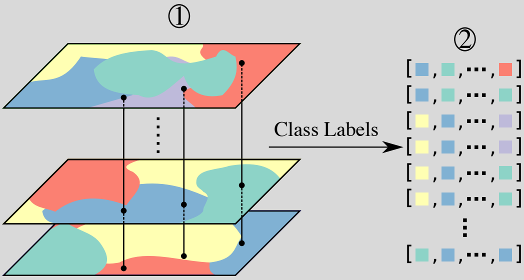
Solution

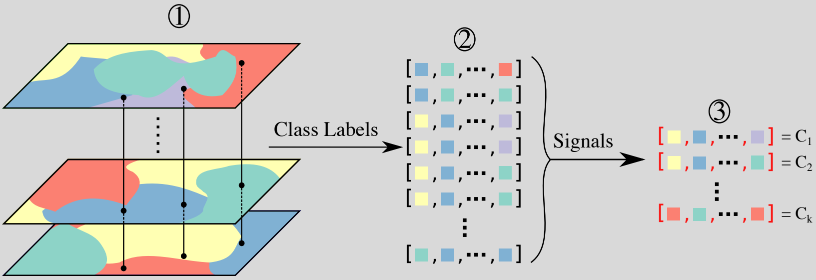
1. Leverage discrete wavelet transform to classify across a multitude of scales.
2. Use information theory to discover most important scales to classify on.
3. **Taking these scales, combine classifications to produce a fuzzy clustering that assess the trust at each point.**

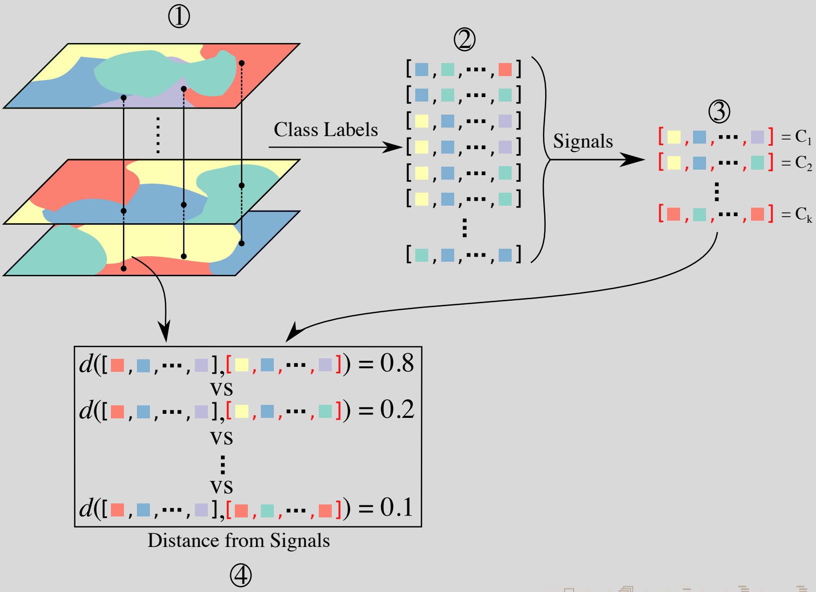


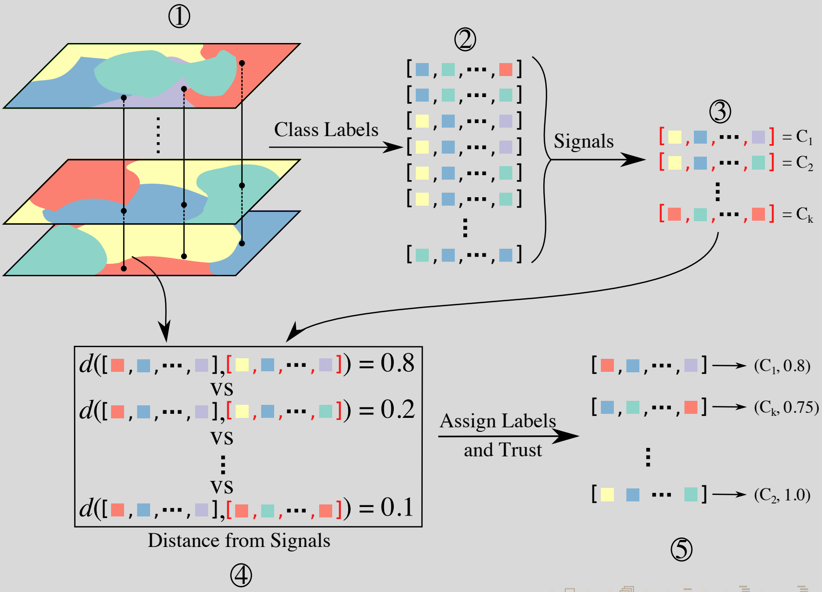
①




①
②
Class Labels


①
②
③
Class Labels
Signals
= C1
= C2
= Ck


①
Class Labels
②
Signals
③
= C1
= C2
= Ck
d([, , ··· ,],[, , ··· ,]) = 0.8
vs
d([, , ··· ,],[, , ··· ,]) = 0.2
vs
vs
d([, , ··· ,],[, , ··· ,]) = 0.1
Distance from Signals
④


①
②
Class Labels
Signals
③
$= C_1$
$= C_2$
$= C_k$
$d([\ldots],[\ldots]) = 0.8$
vs
$d([\ldots],[\ldots]) = 0.2$
vs
vs
$d([\ldots],[\ldots]) = 0.1$
Distance from Signals
④
Assign Labels and Trust
$(C_1, 0.8)$
$(C_k, 0.75)$
$(C_2, 1.0)$
⑤

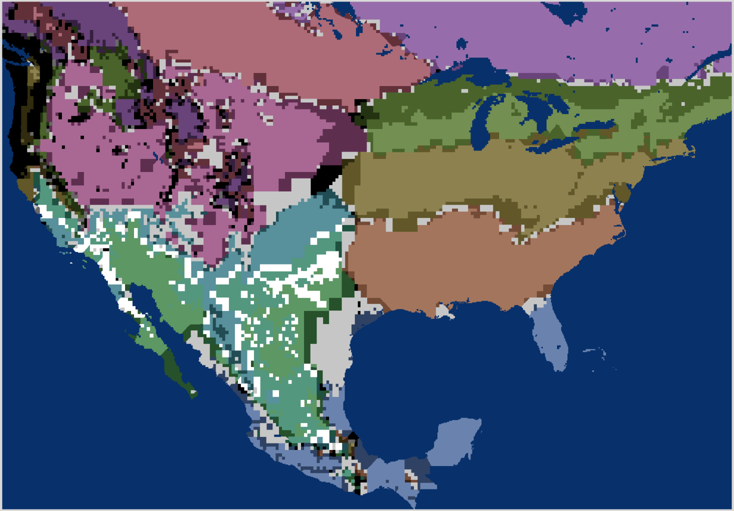
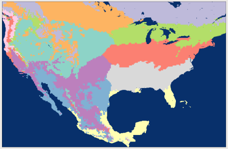


Figure: Consensus clustering from reduced ensemble of clusters for k=10, along with the trust. Grey = multi-class. Darker hue = lower trust.

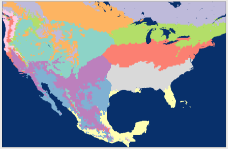
Summary

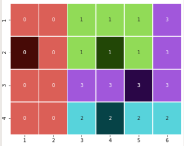
- The DWT brings forth structure hidden at different scales within the data.



Summary

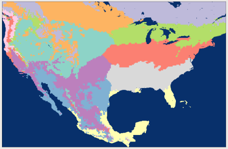
- The DWT brings forth structure hidden at different scales within the data.
- Mutual information allows us to effectively represent the diversity across all scales.

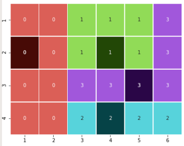


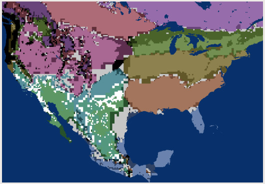


Summary

- The DWT brings forth structure hidden at different scales within the data.
- Mutual information allows us to effectively represent the diversity across all scales.
- Using this reduced ensemble, we produce a fuzzy clustering that has an interpretable trust metric at each point in space.







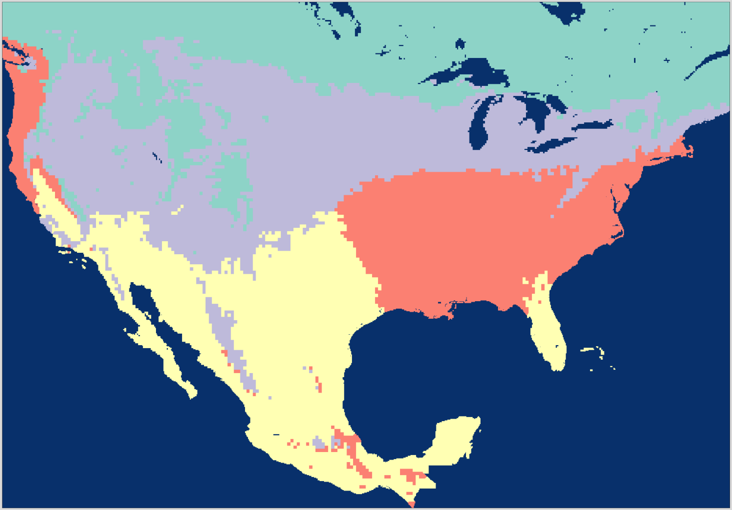


Figure: CGC: K-means $k = 4$, $(\ell_s, \ell_t) = (2, 3)$

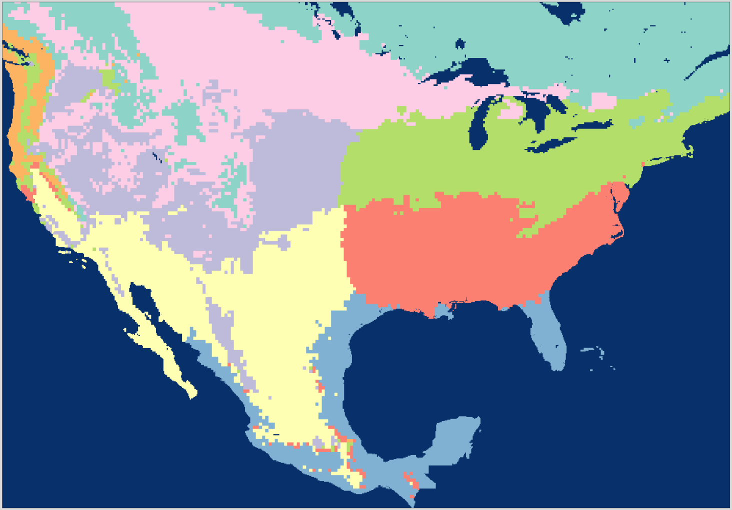


Figure: CGC: K-means $k = 8$, $(\ell_s, \ell_t) = (2, 3)$

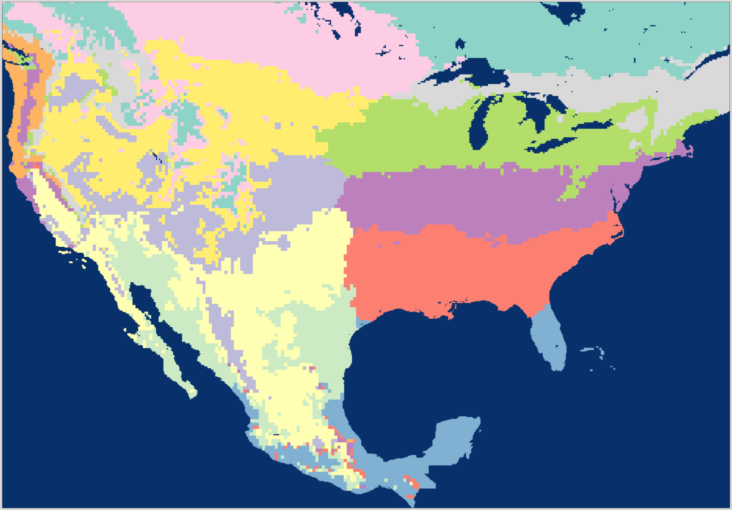


Figure: CGC: K-means $k = 12$, $(\ell_s, \ell_t) = (2, 3)$

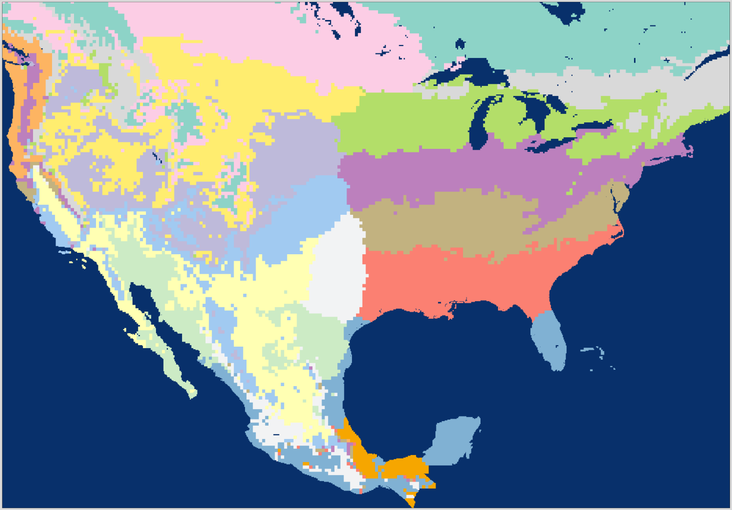


Figure: CGC: K-means $k = 16$, $(\ell_s, \ell_t) = (2, 3)$