



# Outline

- 1 Introduction
- 2 Methodological approach
- 3 Experimental Design
- 4 Results
- 5 Conclusions

# Outline

- 1 Introduction
- 2 Methodological approach
- 3 Experimental Design
- 4 Results
- 5 Conclusions

# Satellite image time series

- Satellite images are a highly valuable and abundant resource for many real world applications.
- Earth monitoring using satellite images is essential for the identification, mapping, assessment and monitoring the land.
- Public access to satellite imagery has favored the interest of a growing number of researchers in the analysis of satellite image time series (SITS).
- The huge data volume and the complexity of SITS analysis have promoted the use of machine learning methods.

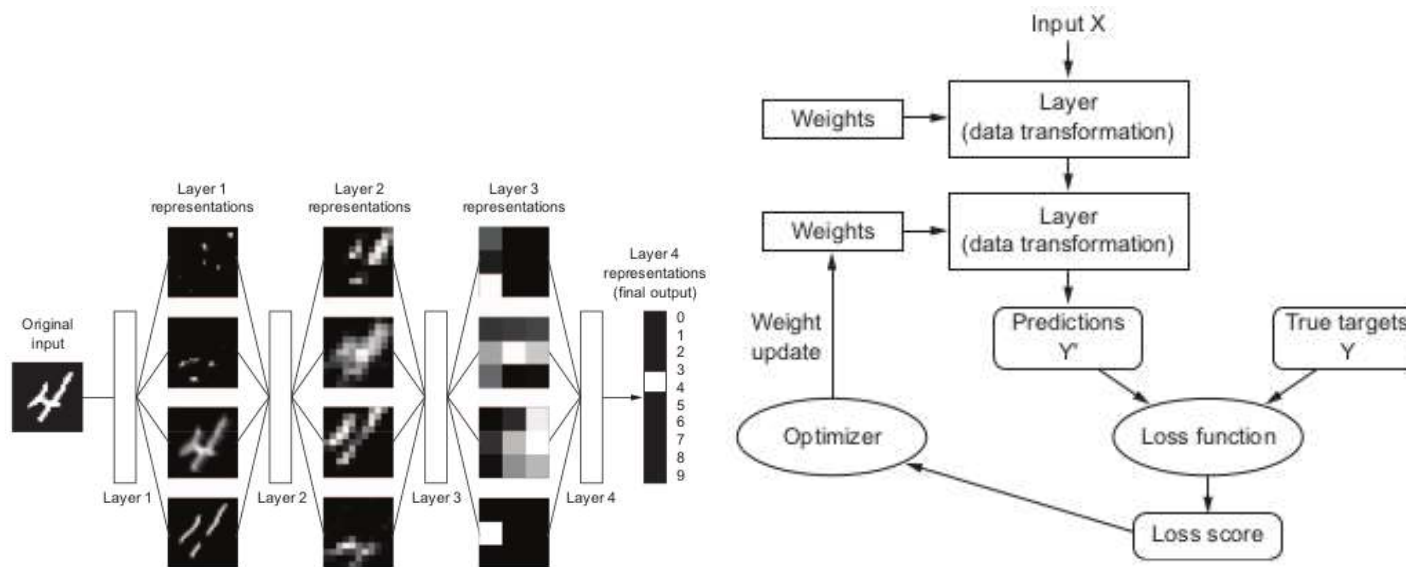


# Satellite image time series

- The labeled data needed to train most machine learning models are scarce and difficult to obtain.
- The class label of a region can change over time, especially when the time series is long.
- Semi-supervised and clustering methods are gaining increasing relevance in SITS analysis.
- Lack of methods to analyze a region from a global perspective.



# Deep learning



- Extract distinctive patterns behind an image through a complex architecture of hidden layers.
- Image representations that are increasingly different from the original and increasingly informative about the final result.
- The loss score is the feedback signal to adjust the values of the weights (backpropagation).

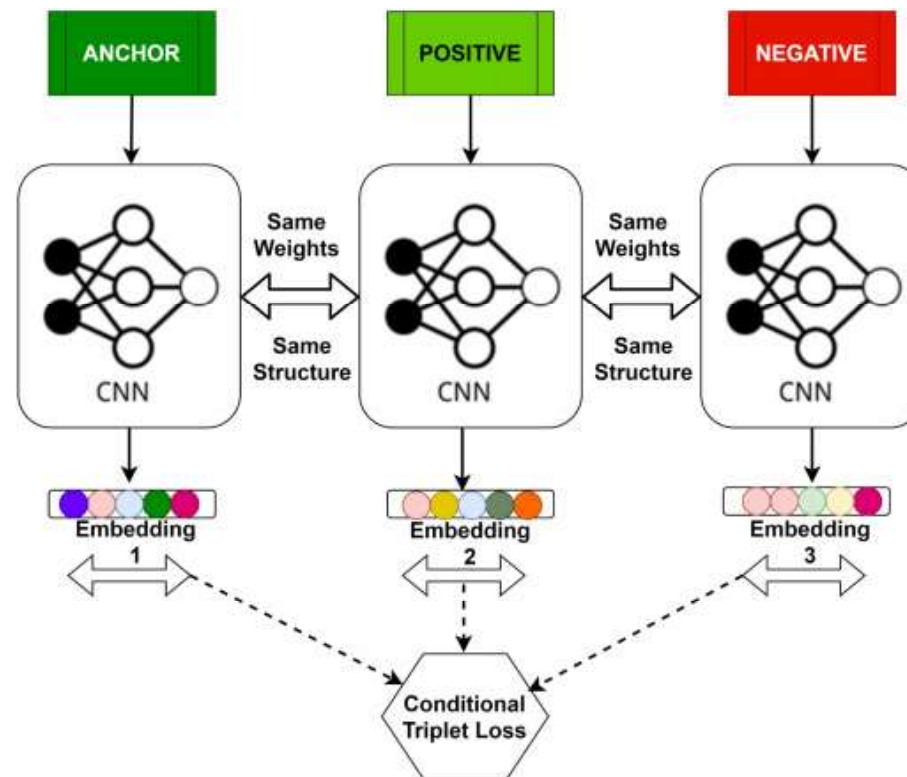
# Deep learning for satellite images

- Deep learning techniques, and more specifically convolutional neural networks (CNN), represent a revolution in the field of image analysis in general and in the analysis of SITS data in particular.
- These techniques are able to extract patterns from vast amounts of complex data. Therefore, they become a natural candidate to solve problems in the field of remote sensing.
- *Examples:* land use classification, wildfire forecasting, predicting sea ice motion, crop type mapping by combining data from satellites and farmer smartphones, segmentation, ...

# Semantic embeddings

- Complex multi-spectral satellite images can be encoded as vectors of bounded size by means of CNNs.
- A semantic embedding not only provides meaningful vectors, but also the distance between them represents the degree of semantic similarity (*vectors of similar images are closer than vectors of dissimilar images*).
- We use a method based on Tile2Vec, an algorithm to create semantically meaningful embeddings where simple arithmetic operations within the space conserve semantic properties.
- This semantic embedding is learned in an unsupervised way.

# Unsupervised learning based on triplets



$$L(x_a, x_p, x_n) = [\|f(x_a) - f(x_p)\|_2 - \|f(x_a) - f(x_n)\|_2 + \delta]_+, \quad (1)$$

# Main target

To propose a fully unsupervised methodology that, given a SITS of a region of interest, creates a partition of the ground where each cluster has both similar semantic properties and similar temporal evolution.

# Outline

- 1 Introduction
- 2 Methodological approach
- 3 Experimental Design
- 4 Results
- 5 Conclusions

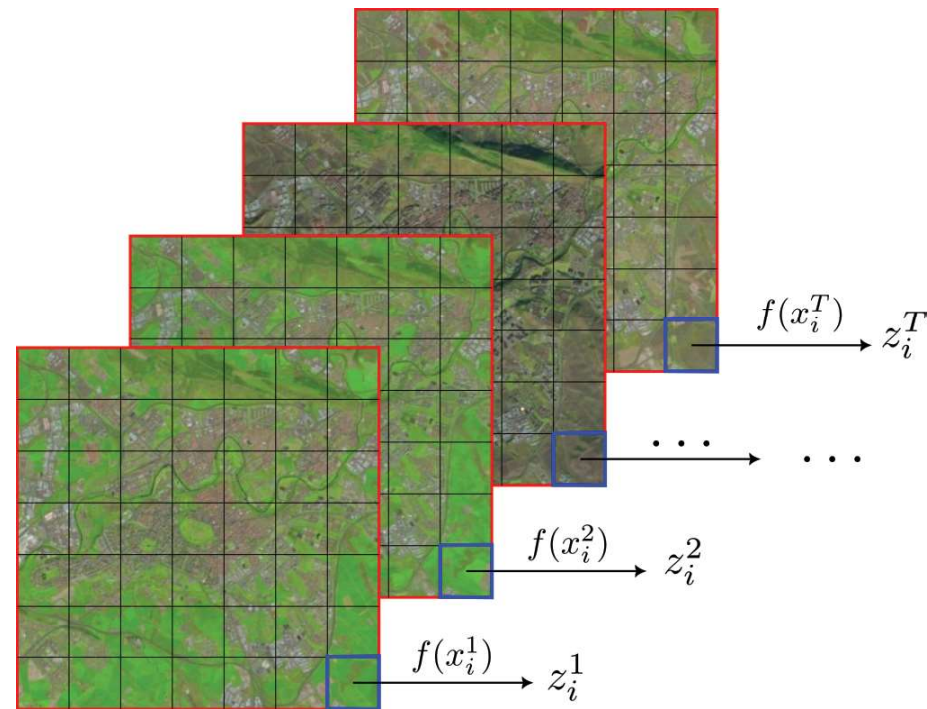


# Proposed methodology

- 1 Unsupervised learning of the embedding from SITS according to the geographic neighborhood (*distributional hypothesis*).
- 2 The sequence of images is decomposed into a collection of multivariate time series (MTS) by means of the embedding.
- 3 Identify areas with similar semantic, both in space and time, using the  $K$ -means clustering method for MTS.
- 4 Refine the embedding by considering the neighborhood given by the previous clustering partition of MTS.
- 5 Run a second clustering based on the refined embedding.
- 6 Provide a final semantic summary with a graph representation.

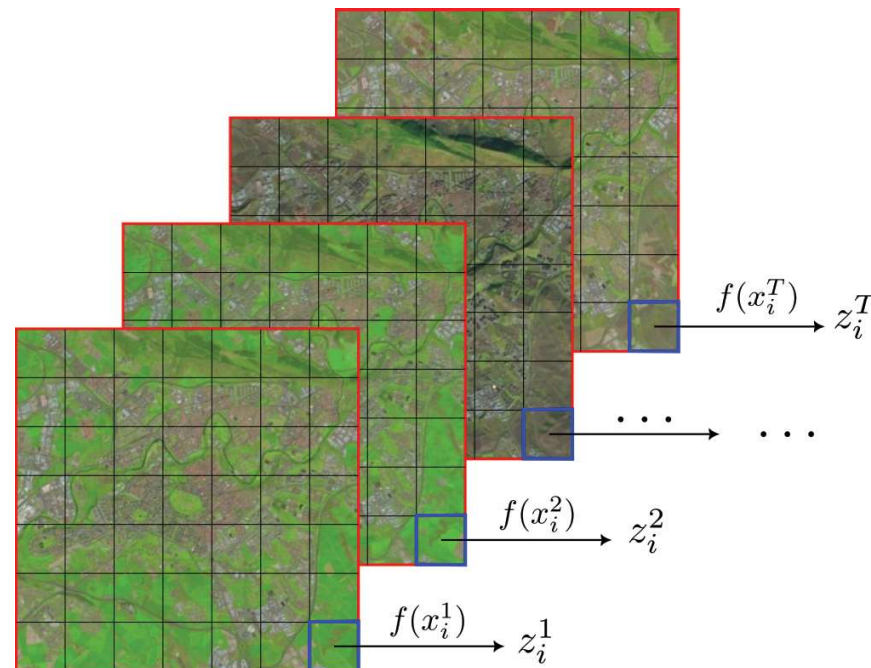
# Definitions

- Let  $X = \{x_1, \dots, x_m\}$  be an image of a region that is decomposed into a grid of tiles  $x_i$ .
- Let  $(X^1, \dots, X^T)$  be a temporal sequence of satellite images of the same region, where  $X^t$  is the image at time  $t$ .



# Embedded sequences of tiles

- From these images, we get sequences of tiles (STs)  
 $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ , where  $\mathbf{x}_i = (x_i^1, \dots, x_i^T)$  corresponds to the  $i$ -th ST.
- Based on the Tile2Vec approach, we represent a ST,  $\mathbf{x}_i$ , as a multivariate time series (MTS) of embedded vectors  $\mathbf{z}_i = (f(x_i^1), f(x_i^2), \dots, f(x_i^T))$ .



# Tile2Vec

Tobler's first law of geography: *everything is related to everything else, but near things are more related than distant things*

- Tile2Vec learns an embedding,  $f$ , from a training set,  $D$ , of triplets of tiles  $(x_a, x_b, x_c)$ , where  $x_a$  denotes the anchor tile,  $x_b$  the neighbor tile and  $x_c$  the distant tile.
- The embedding function,  $f$ , maps multi-spectral tiles  $x \in \mathcal{X}$  to a  $d$ -dimensional vector  $z$ ,  $f : \mathcal{X} \rightarrow \mathbb{R}^d$ .
- The embedding  $f$  is a CNN with a modified input, to be able to handle multi-spectral tiles, and without the final classification layer.

## Tile2Vec loss function

- The parameters of the embedding  $f$  are those that minimize the distance between geographically neighboring tiles,  $(x_a, x_b)$ , and maximize the distance between distant tiles  $(x_a, x_c)$ .
- $f$  is found by minimizing the following loss function for each triplet:

$$L(x_a, x_b, x_c) = [\|f(x_a) - f(x_b)\|_2 - \|f(x_a) - f(x_c)\|_2 + \delta]_+, \quad (2)$$

- The full objective function is the sum of loss for the whole training set of  $N$  triplets  $D = \{(x_a, x_b, x_c)\}$ :

$$\arg \min_{\theta} \sum_{i=1}^N \left[ L(x_a^{(i)}, x_b^{(i)}, x_c^{(i)}) + \lambda \left( \|f(x_a^{(i)})\|_2 + \|f(x_b^{(i)})\|_2 + \|f(x_c^{(i)})\|_2 \right) \right]$$

# Clustering of STs

- Given the embedding of a grid of sequences of tiles,  $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_m\}$ , we would like to identify  $K$  differentiated groups of sequence of tiles by using partitional clustering.
- $K$  determines the number of clusters of a partition of  $\mathbf{Z}$ ,  $\mathcal{P} = \{P_1, \dots, P_K\}$ .
- We solve the  $K$ -means problem for  $\mathbf{Z}$  by minimizing the error:

$$E(\mathcal{P}) = \sum_{k=1}^K \sum_{\mathbf{z} \in P_k} d(\mathbf{z}, \mathbf{c}_k)^2, \quad (3)$$

where  $d(\mathbf{z}, \mathbf{z}') = \sum_{t=1}^T \|\mathbf{z}_t - \mathbf{z}'_t\|_2$  is the Euclidean distance between the MTSs  $\mathbf{z}$  and  $\mathbf{z}'$ , and  $\mathbf{c}_k = \frac{1}{|P_k|} \sum_{\mathbf{z} \in P_k} \mathbf{z}$  is the centroid of the cluster  $P_k$  which corresponds to the average of the MTS within this cluster.

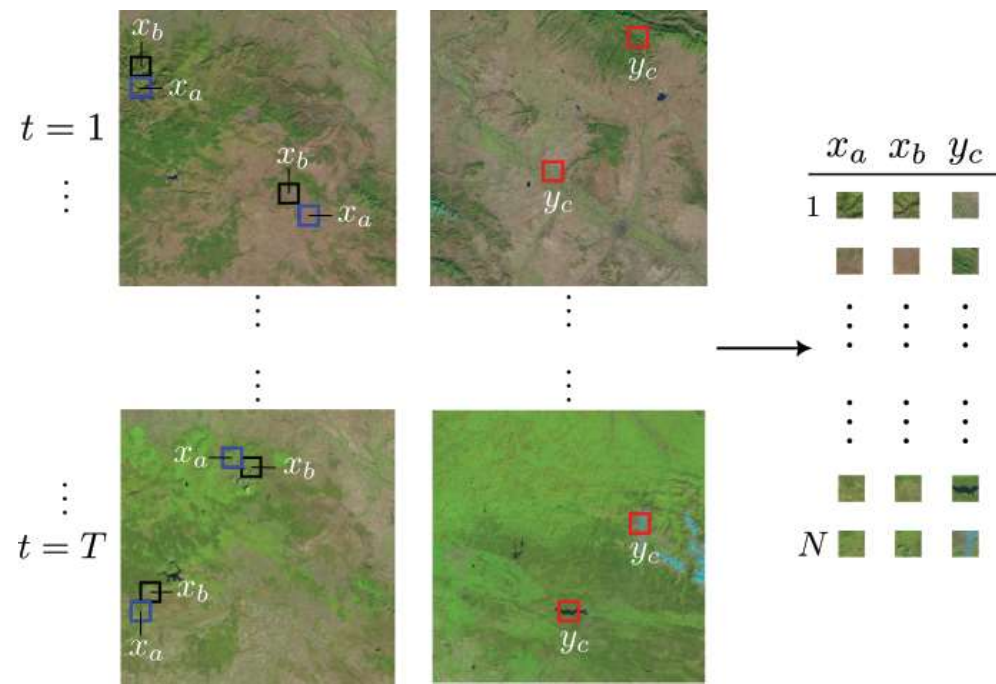
# Semantic clustering procedure

The semantic clustering over a sequence of images can be summarized as follows:

- Learn a geographic-based embedding of STs,  $f^g$ .
- Clustering of STs,  $\mathcal{P}^g$ , using the embedding  $f^g$ .
- Learn a clustering-based embedding of STs,  $f^c$ .
- Clustering of STs,  $\mathcal{P}^c$ , using the embedding  $f^c$ .

# Geographic-based embedding $f^g$

The training set,  $D^g$ , is generated by using two sequences of images  $(X^1, \dots, X^T)$  and  $(Y^1, \dots, Y^T)$  subject to the next constraint: the triplet must belong to images from the same time,  $(x_a^t, x_b^t, y_c^t)$  where  $x_a^t, x_b^t \in X^t$  and  $y_c^t \in Y^t$ .



$f^g$  is learned from  $D^g$  by minimizing Equation 2



## Clustering-based embedding $f^c$

We refine geographic-based embedding,  $f^g$ , by using information of the clustering  $\mathcal{P}^g$ . Each triplet of  $D^c$  is generated as follows:

- Select a cluster index  $k$  at random from  $\{1, \dots, K\}$  with a probability proportional to the size of the cluster  $|P_k^g|$ .
- Select an anchor ST  $\mathbf{x}_a$  uniformly at random from the cluster  $P_k^g$ .
- Select a neighbor ST  $\mathbf{x}_b \neq \mathbf{x}_a$  uniformly at random from the cluster  $P_k^g$ .
- Select a cluster index  $j$  at random from  $\{1, \dots, k-1, k+1, \dots, K\}$  with a probability proportional to the size of the cluster  $|P_j^g|$ .
- Select a distant ST  $\mathbf{x}_c^t$  uniformly at random from  $P_j^g$ .
- Construct the triplet  $(\mathbf{x}_a^t, \mathbf{x}_b^t, \mathbf{x}_c^t)$  by selecting  $t$  uniformly at random from  $\{1, \dots, T\}$ .

# Clustering-based embedding $f^c$

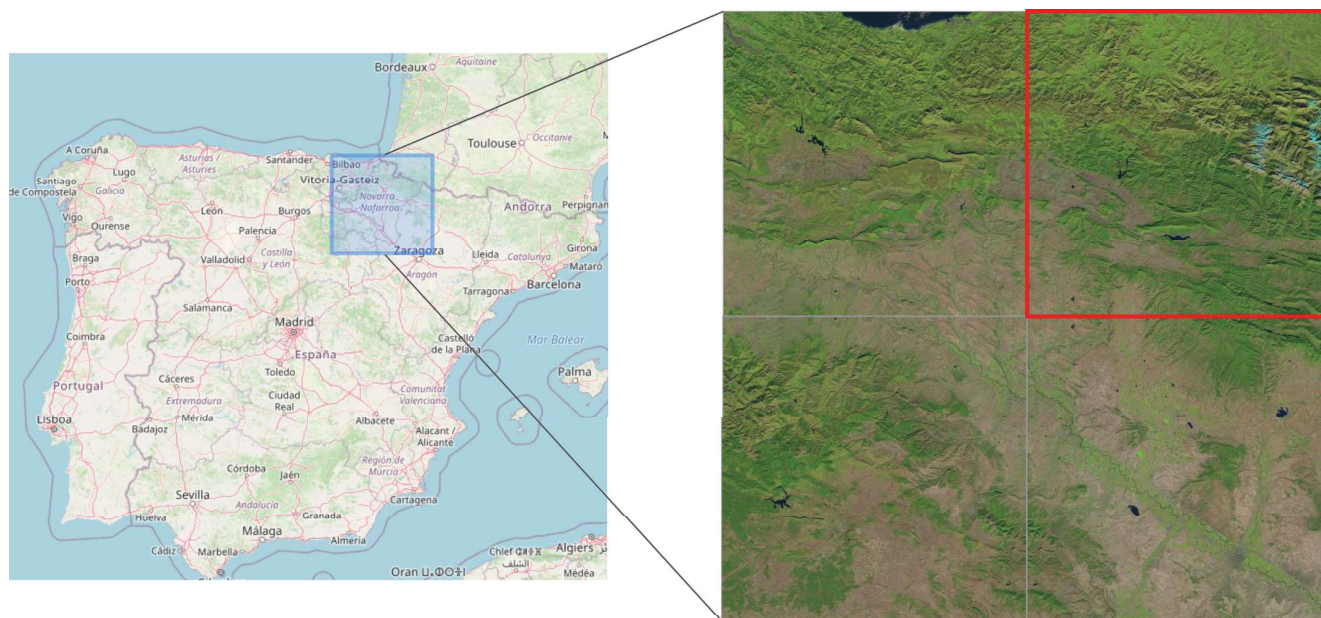
- Since the clustering  $\mathcal{P}^g$  captures spatio-temporal patterns, the triplets sampled from  $\mathcal{P}^g$  contain additional spatial and temporal information.
- The neighbor tiles not only tend to have similar semantic properties to the anchors, but also they belong to regions that change similarly over time. The opposite is true for the distant tiles.
- This step aims at reinforcing the target of the proposed method which is, in essence, to group areas with a similar evolving semantic.

# Outline

- 1 Introduction
- 2 Methodological approach
- 3 Experimental Design**
- 4 Results
- 5 Conclusions

# Image dataset

- Sequences of Sentinel-2 images (image size  $10980 \times 10980$ , 10 meters per pixel) from the region of Navarre.
- Time series with the four seasons of the year during the last five years (2017-2021),  $T = 20$ .



# Training parameters

- $D^g$  contains  $N = 100000$  triplets sampled from the 4 sequences of images.
- The size of the tiles is  $100 \times 100$  pixels (covering  $1 \text{ km}^2$ ).
- The geographical neighborhood is given by a ball of radius  $r = 1 \text{ km}$ .
- The training process is iterated 50 epochs, with a margin of  $\delta = 50$ .
- The last layer of the network has  $d = 512$  features.
- $D^c$  contains  $M = 20000$  triplets sampled from the sequence chosen for clustering. The neighborhood is given by a partitional clustering of size  $K = 5$  and the training process is iterated 25 epoch.

# Methods of analysis

## Geographic representations

- We assign the same color to the tiles belonging to the same cluster in the space of the original images.
- The colors are generated using principal component analysis (PCA).

## 2D projections

The 2D projection of the embedded space tries to represent the original distances between MTS. We use:

- t-Distributed stochastic neighbor embedding (t-SNE)
- Multidimensional scaling (MDS).

# Methods of analysis

## Representatives, interpolations and minimum spanning tree

- The spatio-temporal semantic patterns are exhibited by the representative ST of each cluster and the linear interpolations between them.
- The representative ST is the closest ST to the centroid (cluster medoid).
- The linear interpolations between two centroids  $\mathbf{c}_k$  and  $\mathbf{c}_{k'}$  are intermediate MTS  $\mathbf{z}_w = w \cdot \mathbf{c}_k + (1 - w) \cdot \mathbf{c}_{k'}$ , for  $w \in [0, 1]$ .
- We obtain the interpolated representative ST,  $\mathbf{x}_w$ , as the closest ST to  $\mathbf{z}_w$ .
- The clustering representatives and the interpolations with  $w = 0.5$  are arranged in a meaningful way by calculating the minimum spanning tree from the distance matrix among clusters.

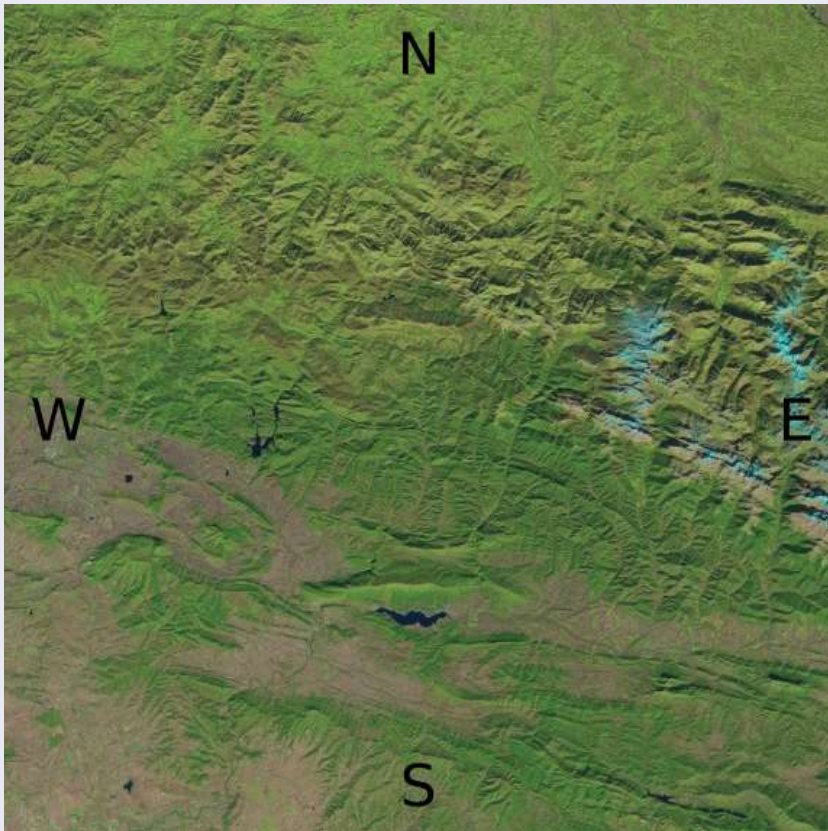
# Outline

- 1 Introduction
- 2 Methodological approach
- 3 Experimental Design
- 4 Results**
- 5 Conclusions

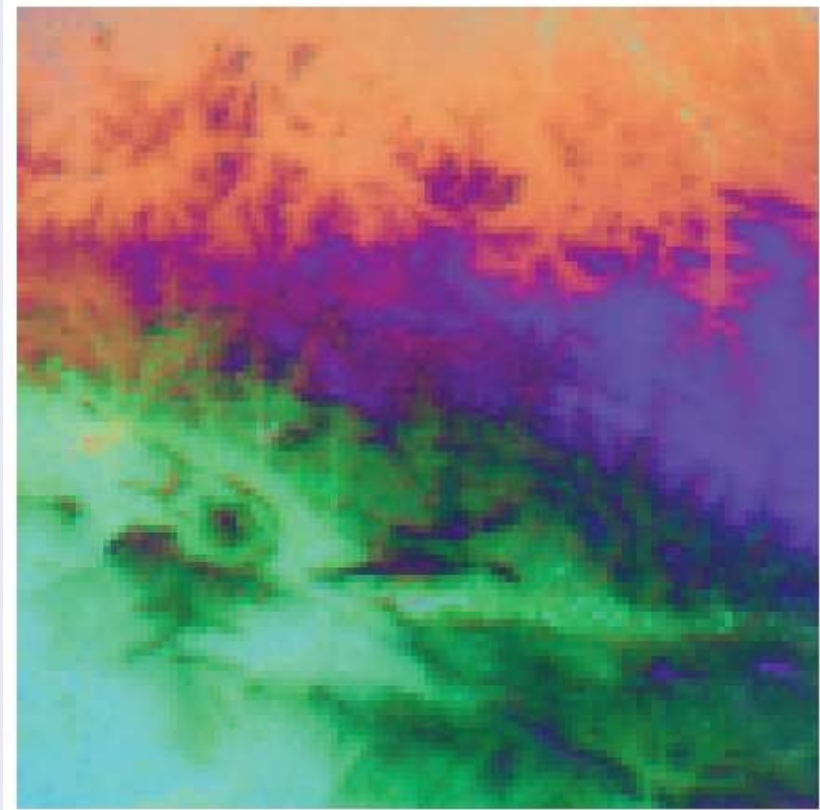


# Preliminary analysis

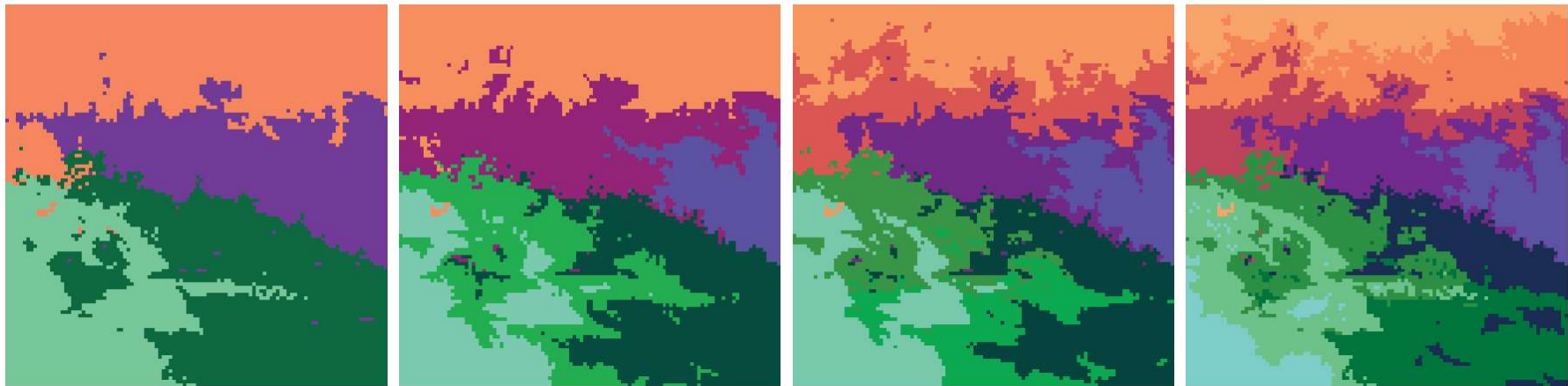
Region of interest



PCA colors for each ST



# Representations of $\mathcal{P}^g$ with different $K$

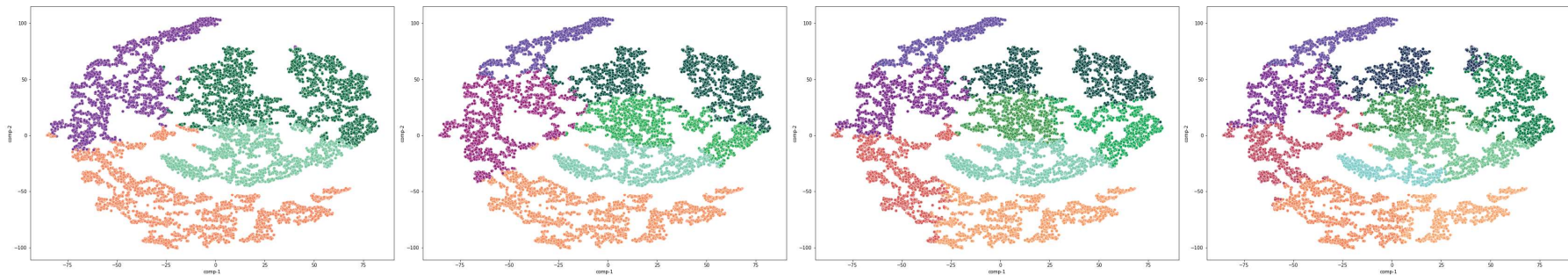


(e)  $K = 4$

(f)  $K = 6$

(g)  $K = 8$

(h)  $K = 10$



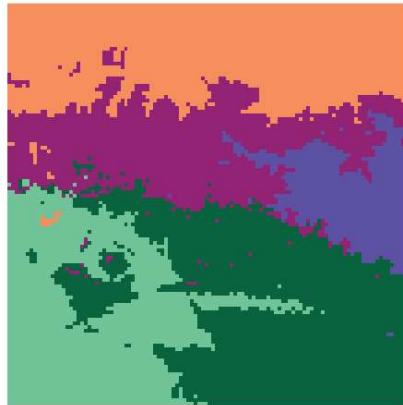
(i)  $K = 4$

(j)  $K = 6$

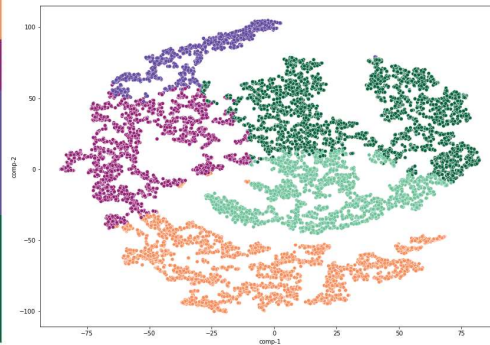
(k)  $K = 8$

(l)  $K = 10$

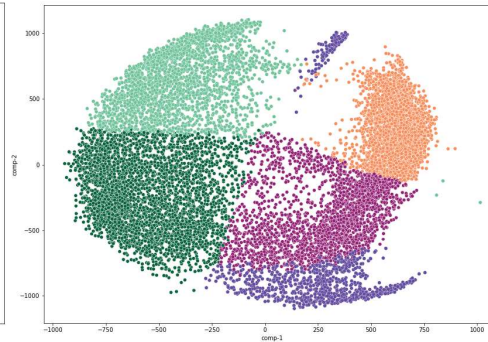
# Comparison between the clusterings $\mathcal{P}^g$ and $\mathcal{P}^c$



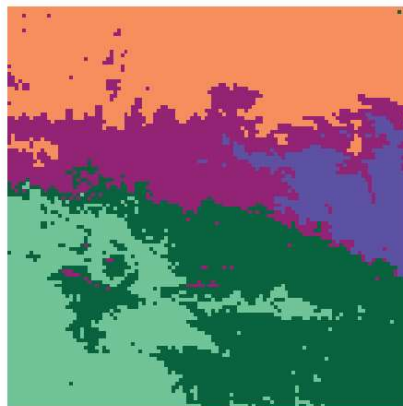
Land according to  $\mathcal{P}^g$



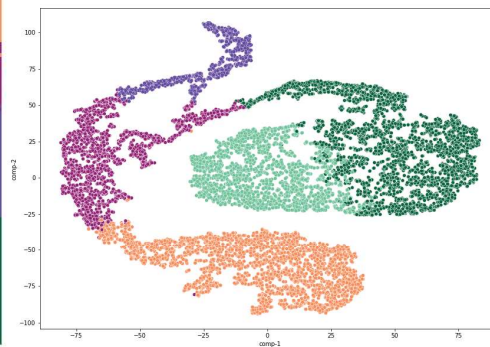
t-SNE of  $\mathcal{P}^g$



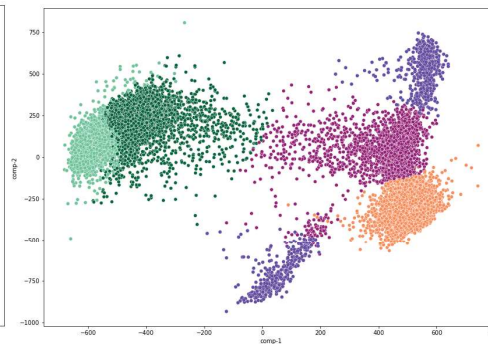
MDS of  $\mathcal{P}^g$



Land according to  $\mathcal{P}^c$



t-SNE with  $\mathcal{P}^c$

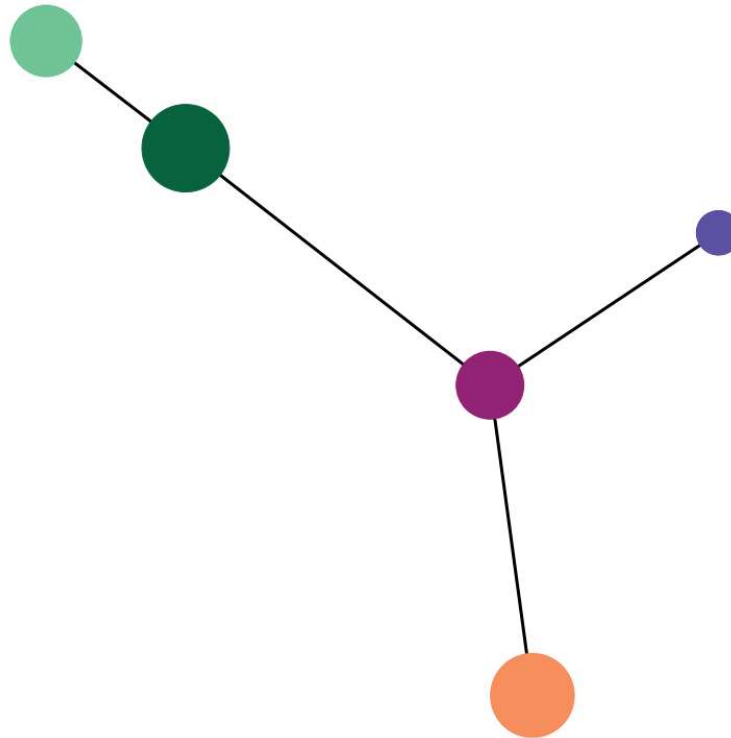
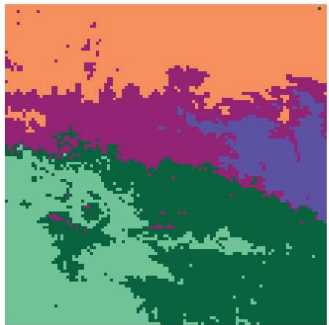


MDS of  $\mathcal{P}^c$

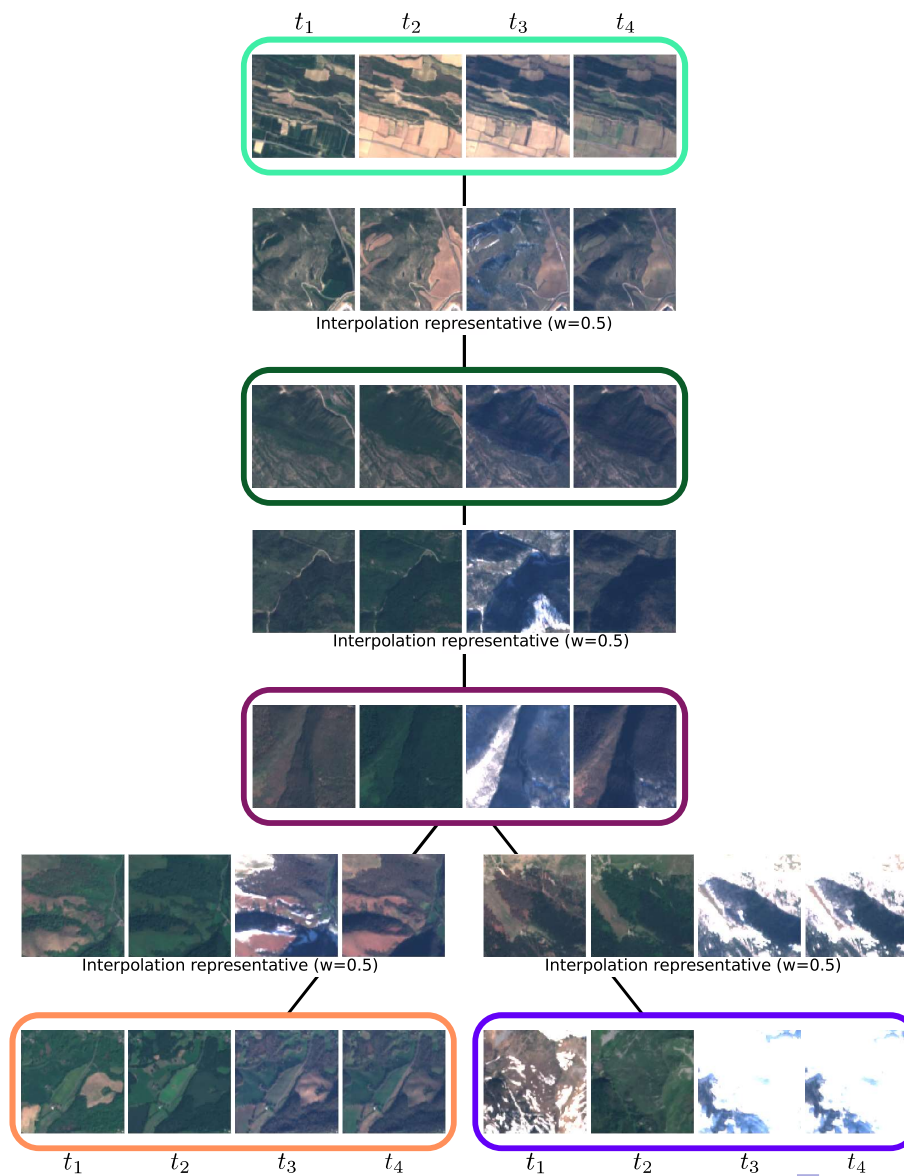
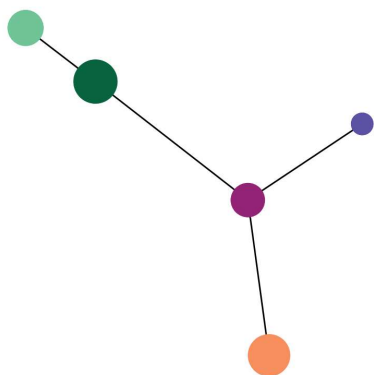
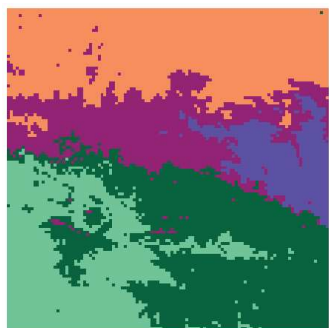
$K$ -means error  $\mathcal{P}^g$ : 13293;  $K$ -means error  $\mathcal{P}^c$ : 2524



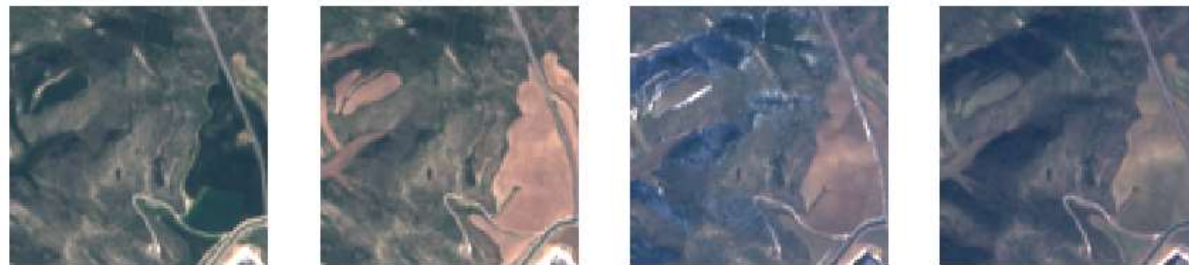
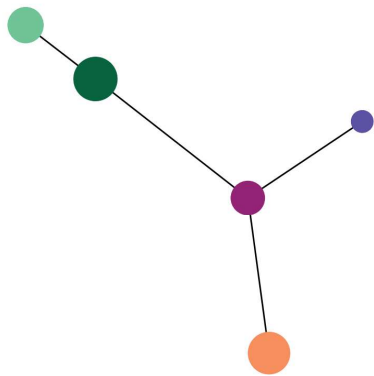
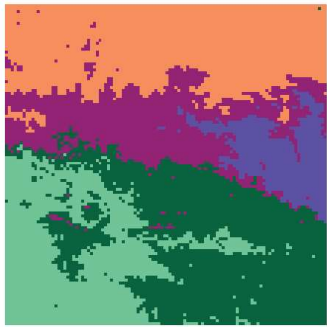
# Minimum spanning tree



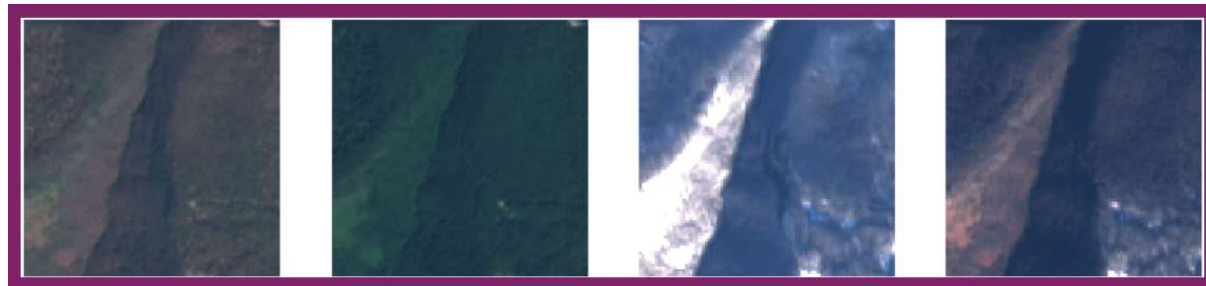
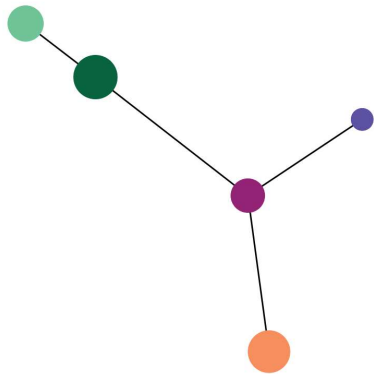
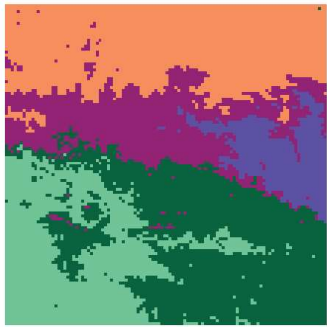
# Graph with representatives and interpolations $w = 0.5$



# Graph with representatives and interpolations $w = 0.5$

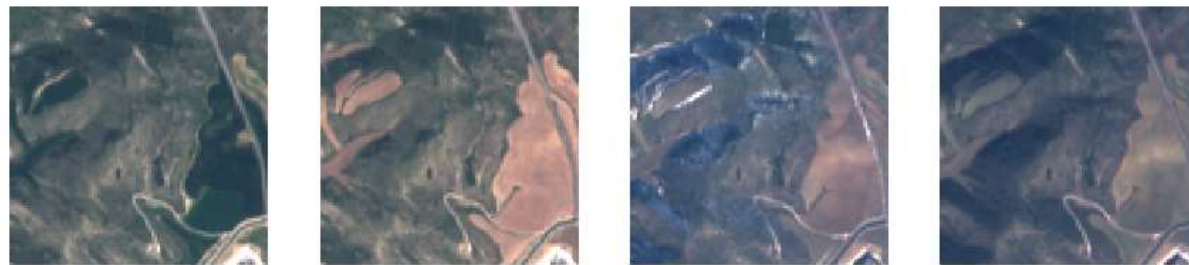
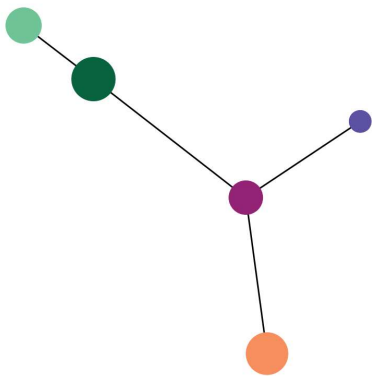
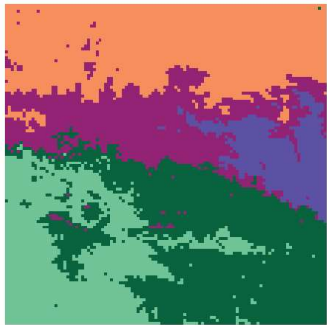


# Centroids of $\mathcal{P}^c$ and interpolation with $w = 0.5$



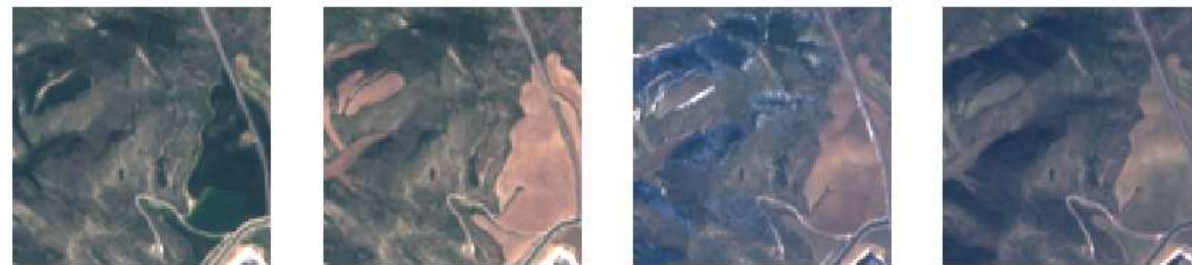
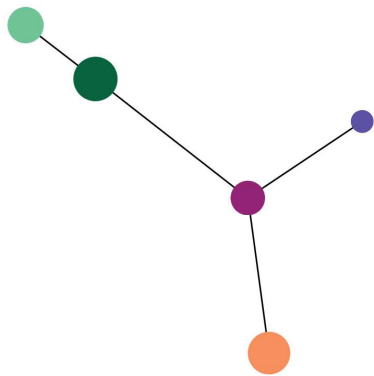
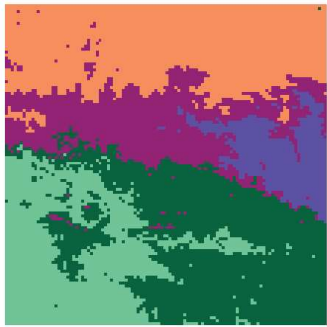


# Graph with representatives and interpolations $w = 0.5$

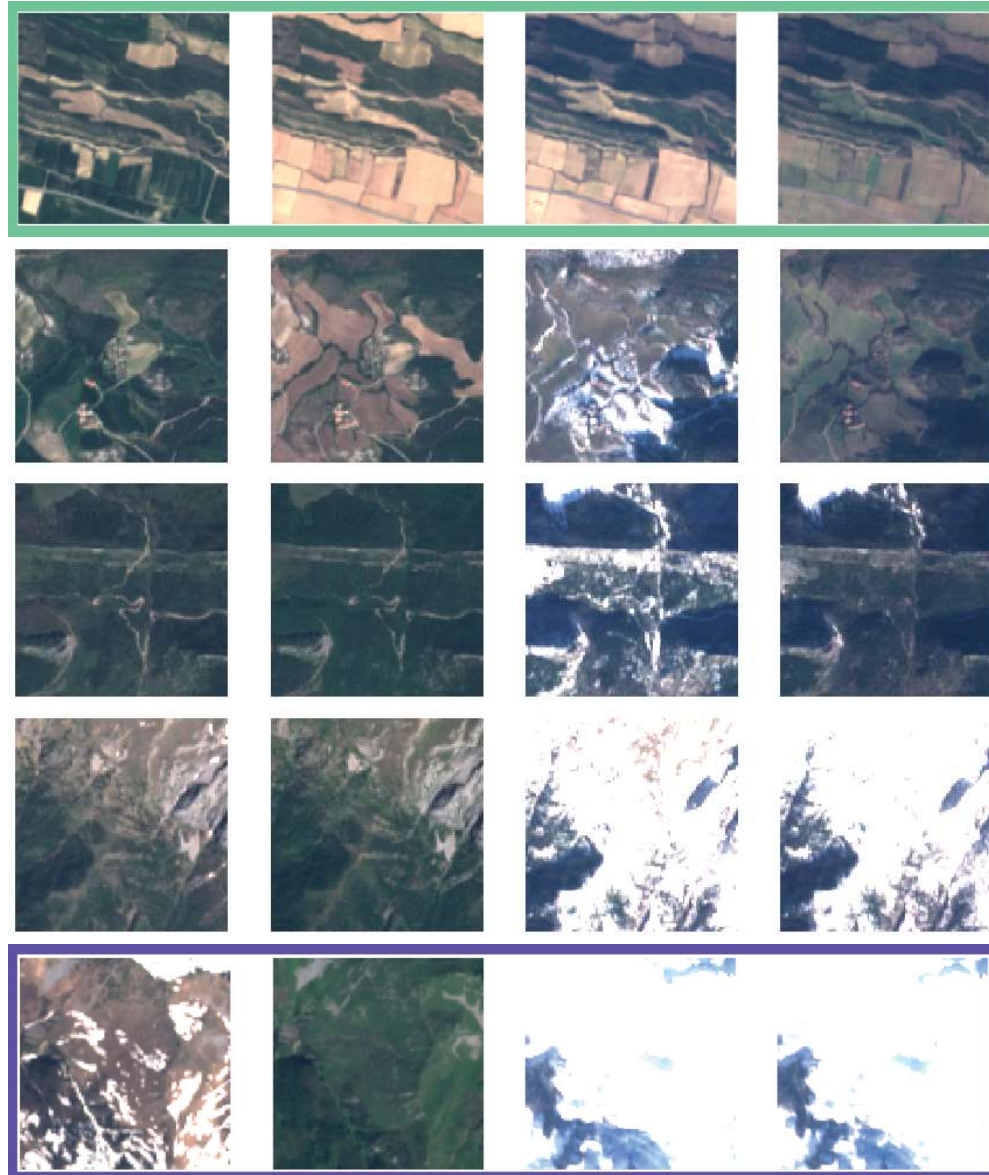
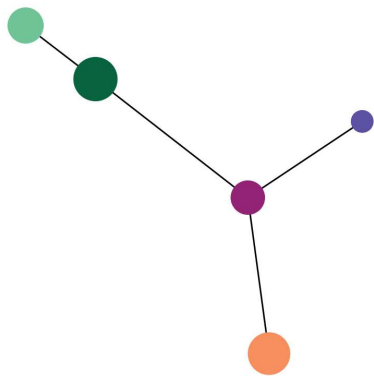
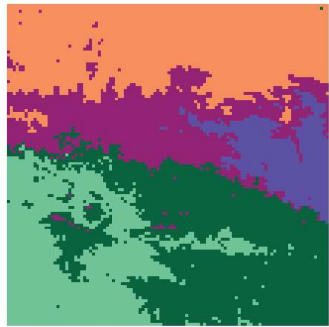




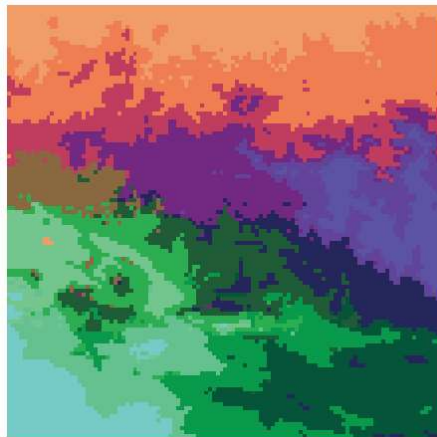
# Graph with representatives and interpolations $w = 0.5$



# Centroids and interpolation with $w \in \{0.25, 0.5, 0.75\}$




# Agriculture-related clusters



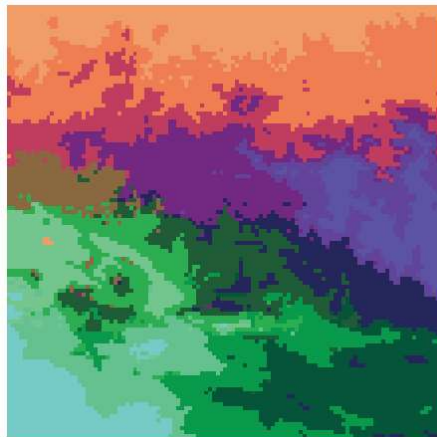
$K = 15$



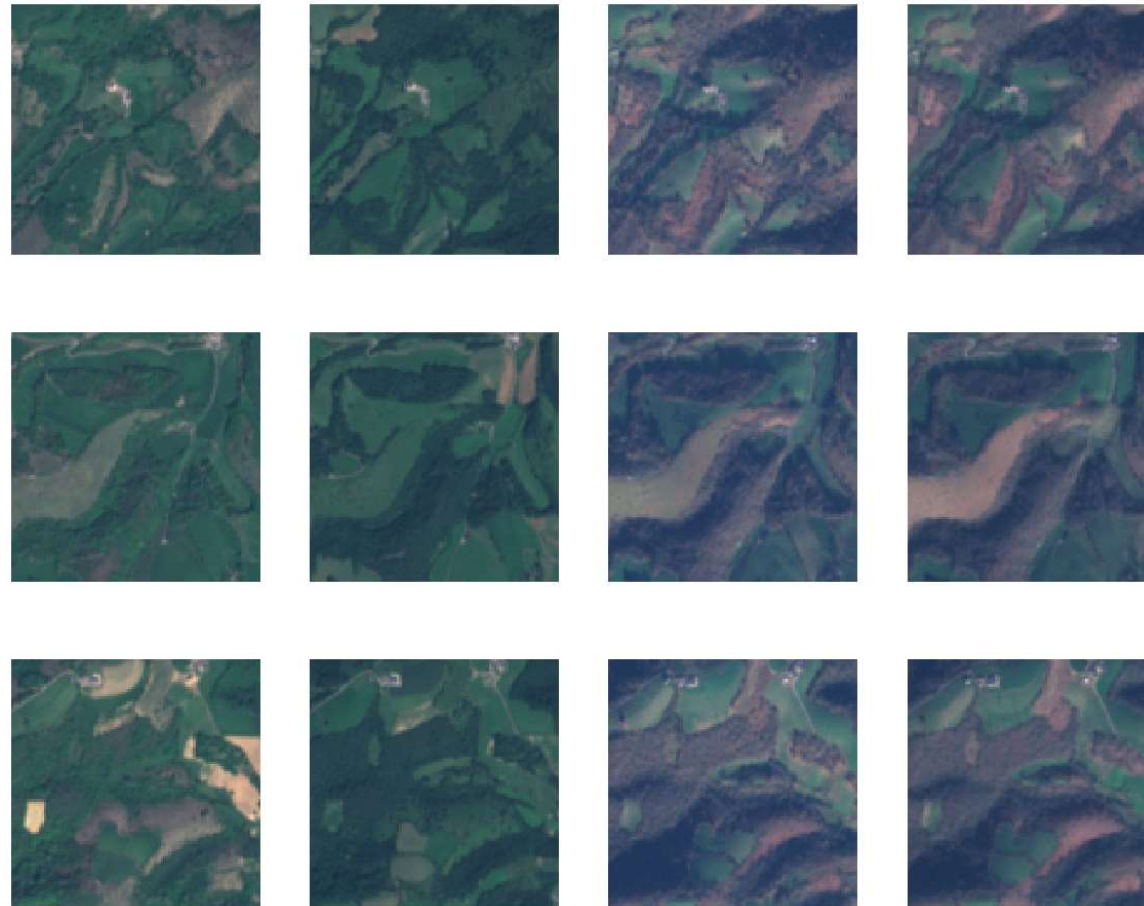
The 3 closest STs to the  cluster centroid




# Agriculture-related clusters

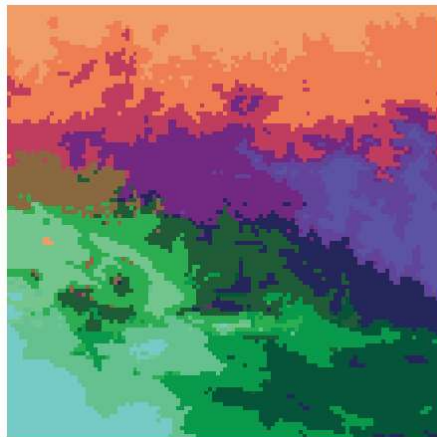


$K = 15$

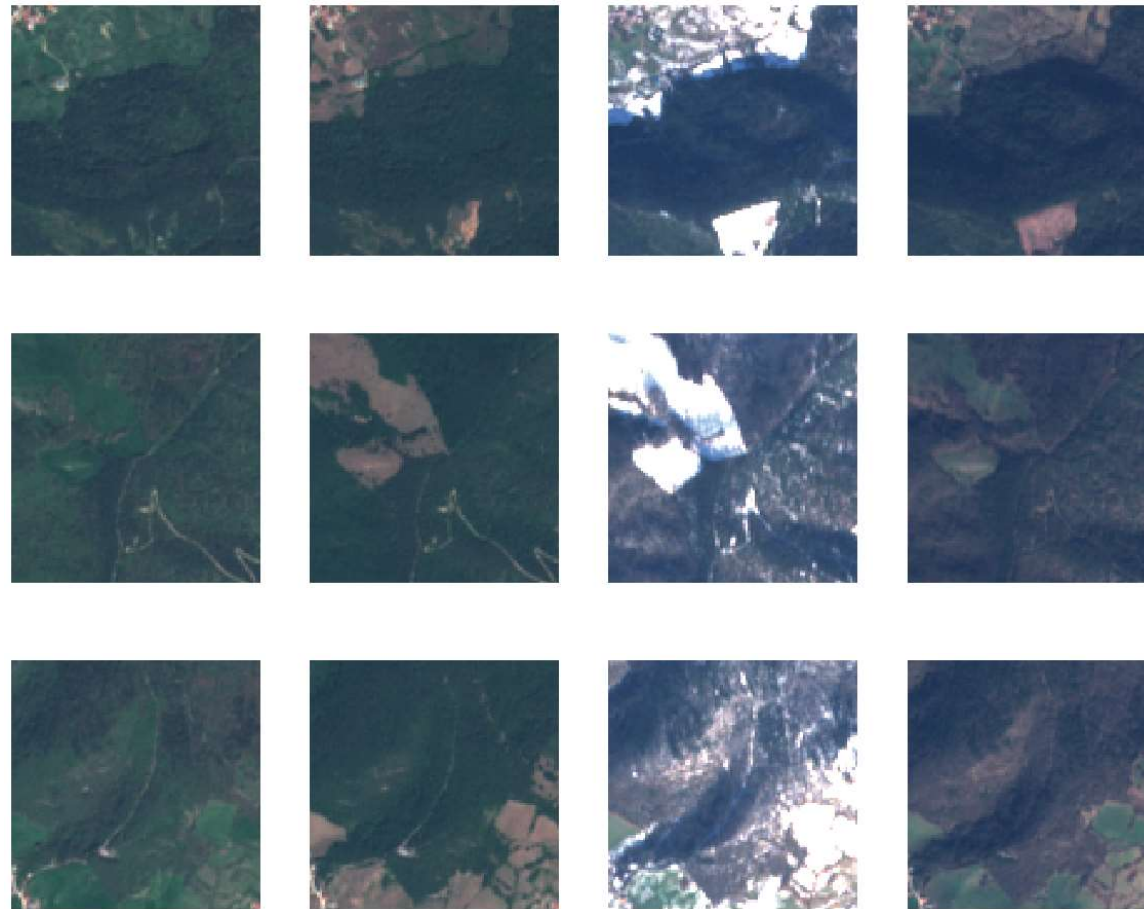



The 3 closest STs to the  cluster centroid

# Agriculture-related clusters

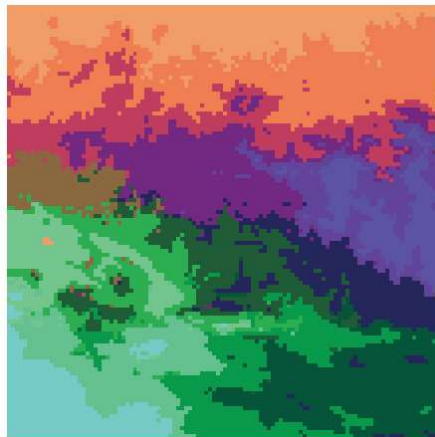


$K = 15$

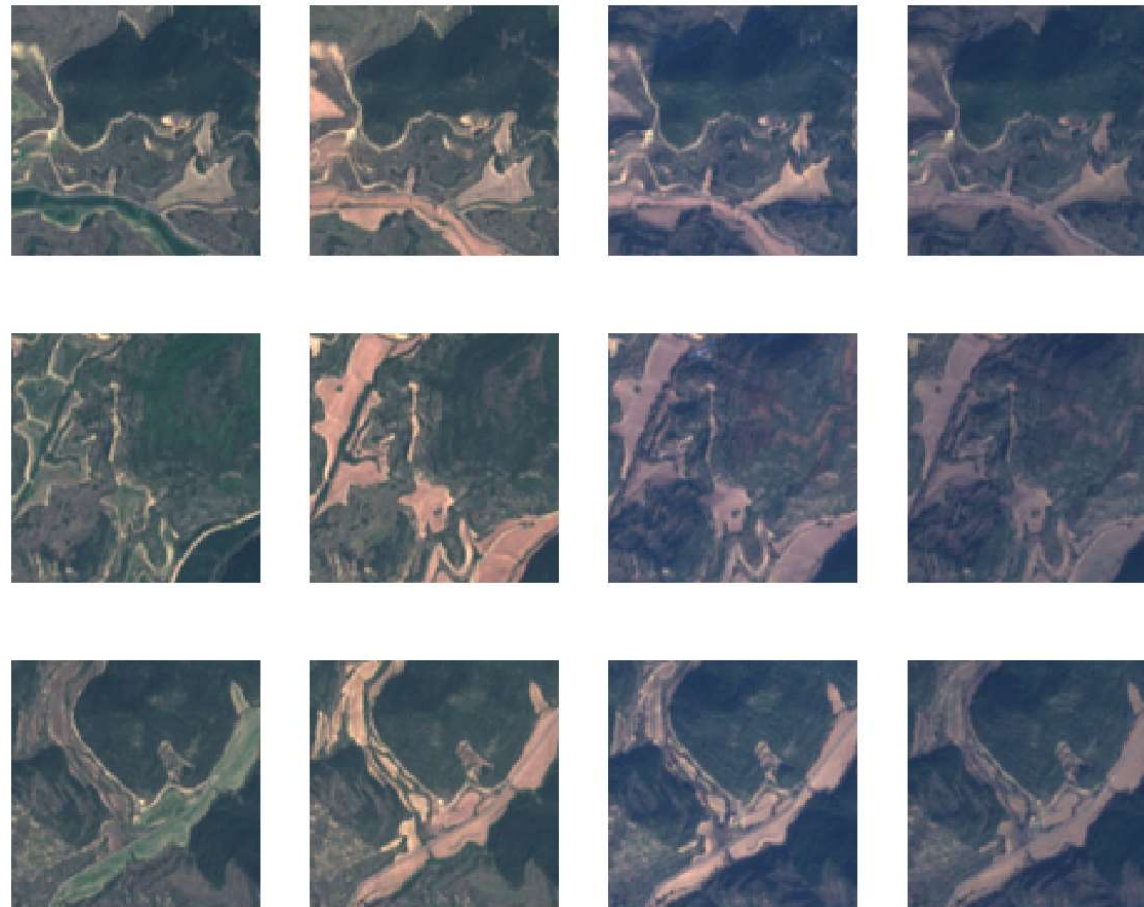



The 3 closest STs to the  cluster centroid

# Agriculture-related clusters



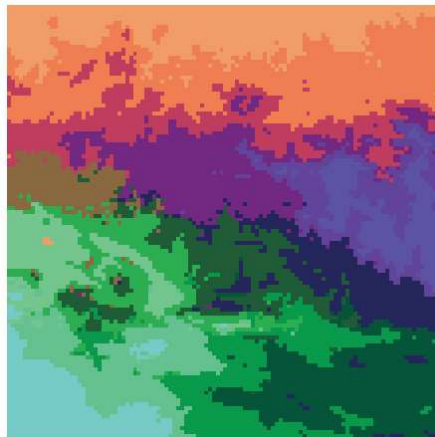
$K = 15$



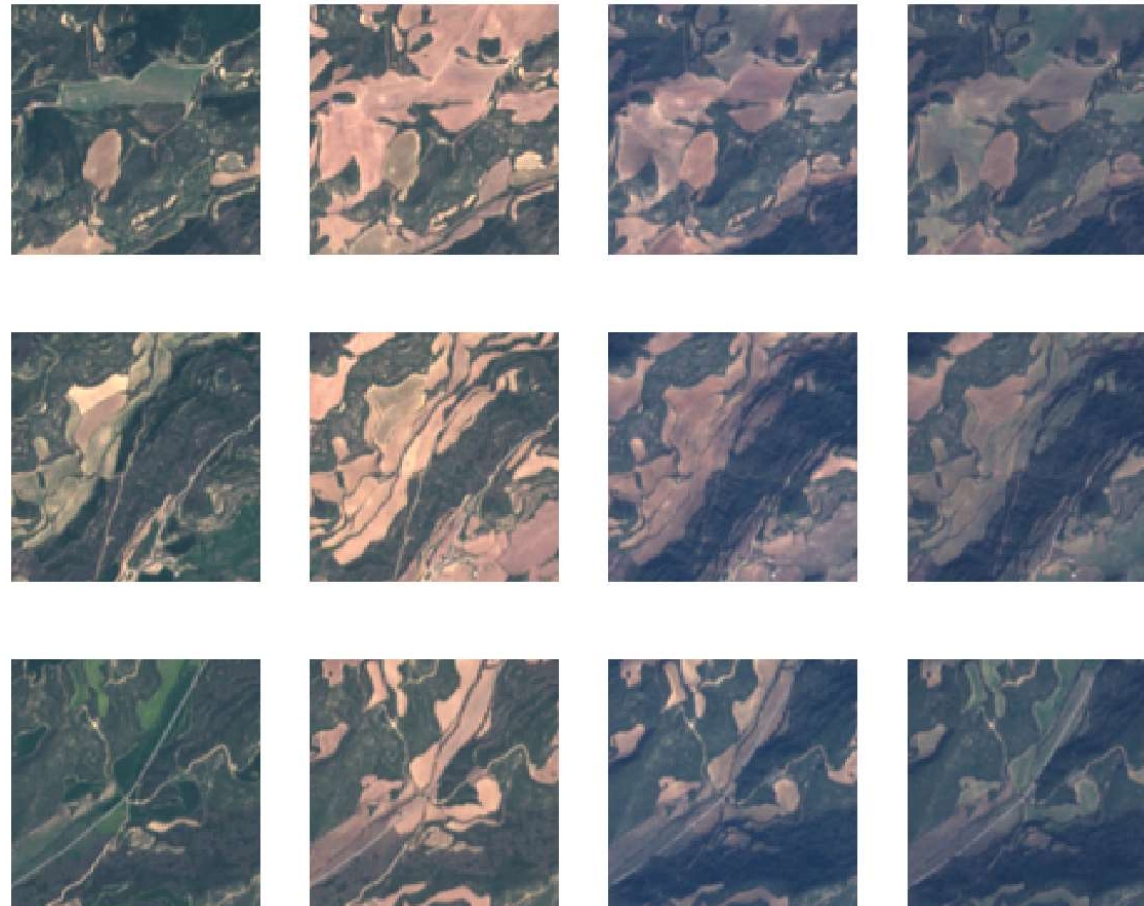
The 3 closest STs to the  cluster centroid




# Agriculture-related clusters

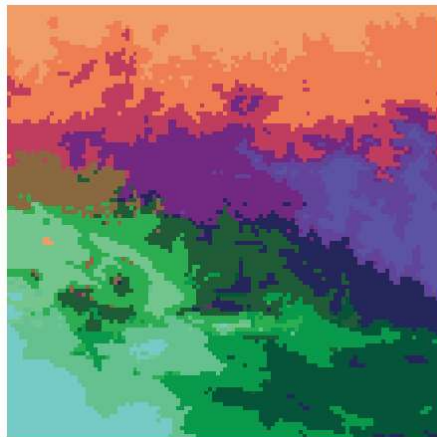


$K = 15$



The 3 closest STs to the  cluster centroid

# Agriculture-related clusters



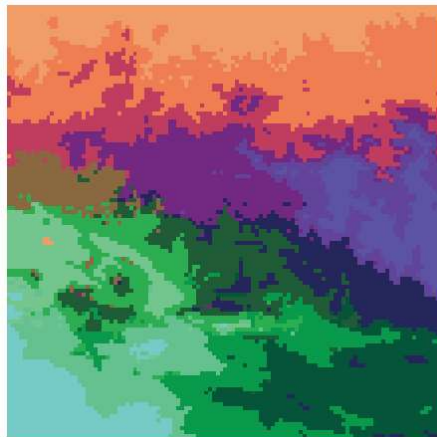
$K = 15$



The 3 closest STs to the █ cluster centroid



# Agriculture-related clusters

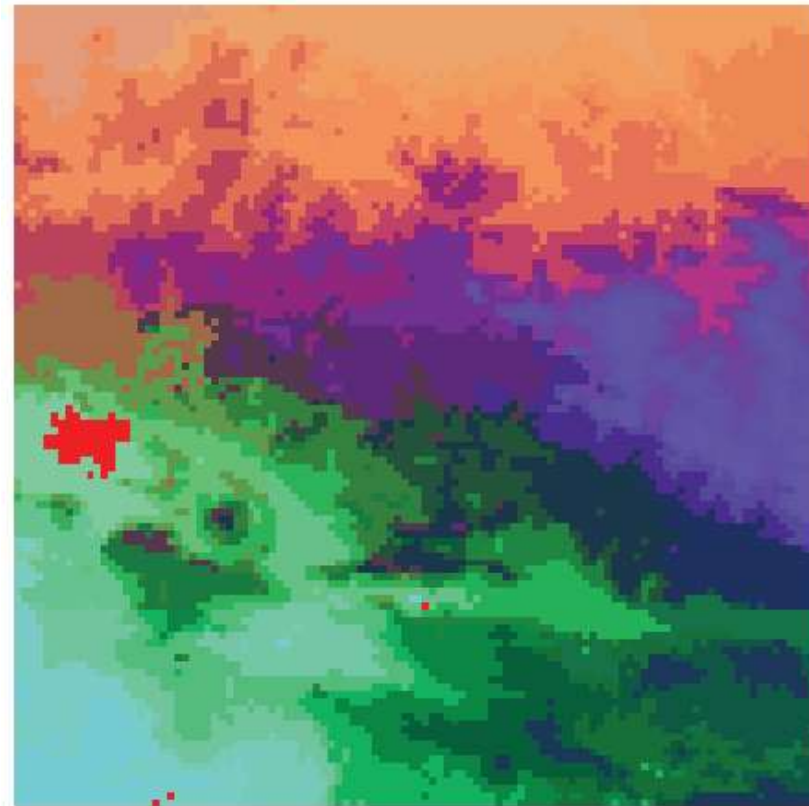


$K = 15$



The 3 closest STs to the  cluster centroid

# Urban cluster: Pamplona



# Outline

- 1 Introduction
- 2 Methodological approach
- 3 Experimental Design
- 4 Results
- 5 Conclusions

# Conclusions

- The clustering of MTS based on embedded vectors exhibits robust patterns of behavior in all the experiments.
- The semantic clustering can contribute a wealth of knowledge about the region of interest, from coarse-grained semantics to more detailed information, as the number of clusters increases.
- There exist a very close connection between the geographic and the embedded representation of the clustering.
- The clustering can be refined and enhanced by means of a second phase of training based on the clustering neighborhood.
- The clustering of MTS automatically captures precise spatio-temporal semantic information.

