Constrained Clustering: Why and How?

Thomas Lampert – Pierre Gançarski – Baptiste Lafabregue

Université de Strasbourg

Project

- Collaborative project R&T A²CNES
  - ICube Université de Strasbourg: Thomas Lampert, Baptiste Lafabregue, Pierre Gançarski
  - LIFO Université d'Orléans: Thi-Bich-Hanh Dao, Nicolas Serrette, Christel Vrain
- Objective:
  - Evaluate the applicability of constrained clustering to time-series (remote sensing images)
- Financed by: CNES

Contents

- Problem background
- Types of constraints
- Overview of constrained clustering algorithms
- Application to time-series
- Experiments/Results
- Discussion of results
- Conclusions

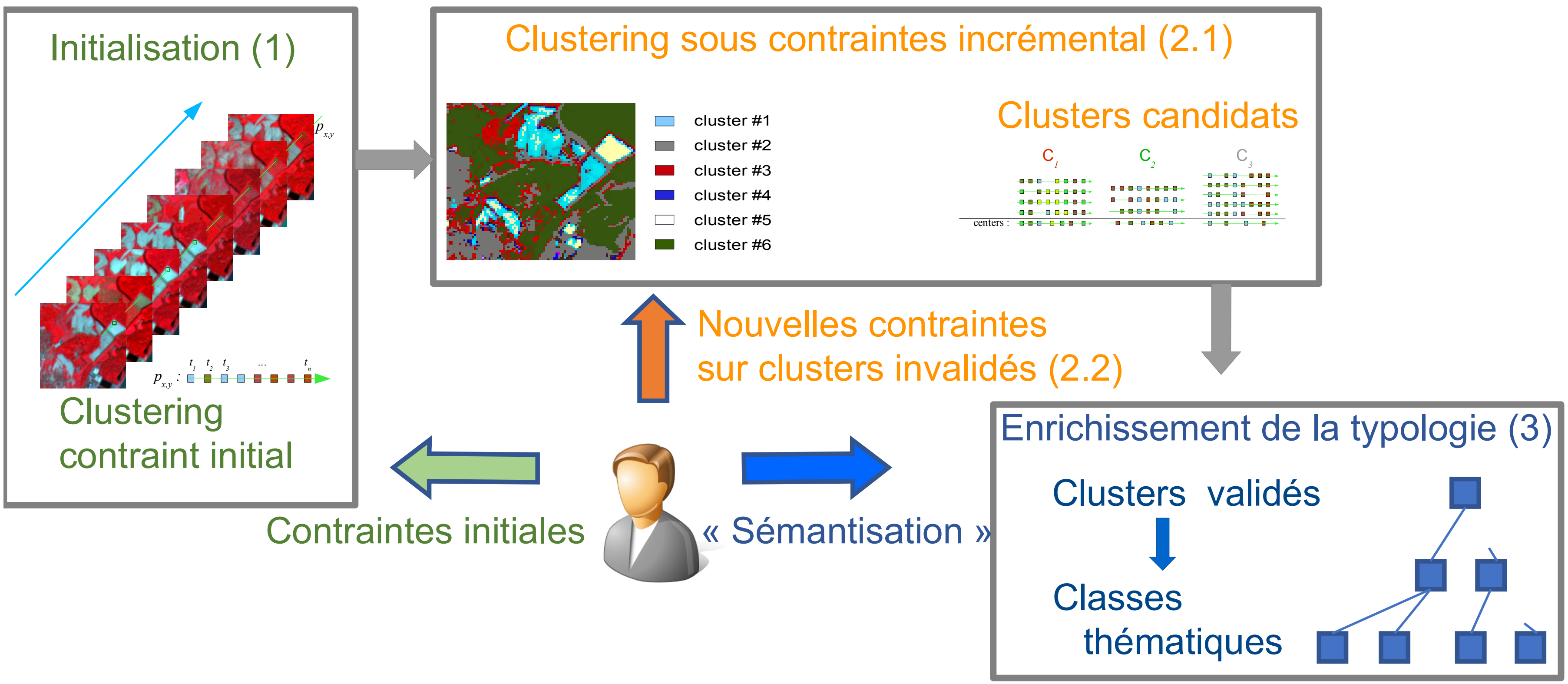
Problem Background

- Supervised learning (SVM, deep learning, etc)
  - Offer excellent classification performance in several areas
  - Require large volumes of training data
- Unsupervised learning (clustering, EM, etc)
  - Data exploration, offer insights into the data
  - Do not require labelled examples
  - Results may not be relevant to the user’s requirements

Problem Background

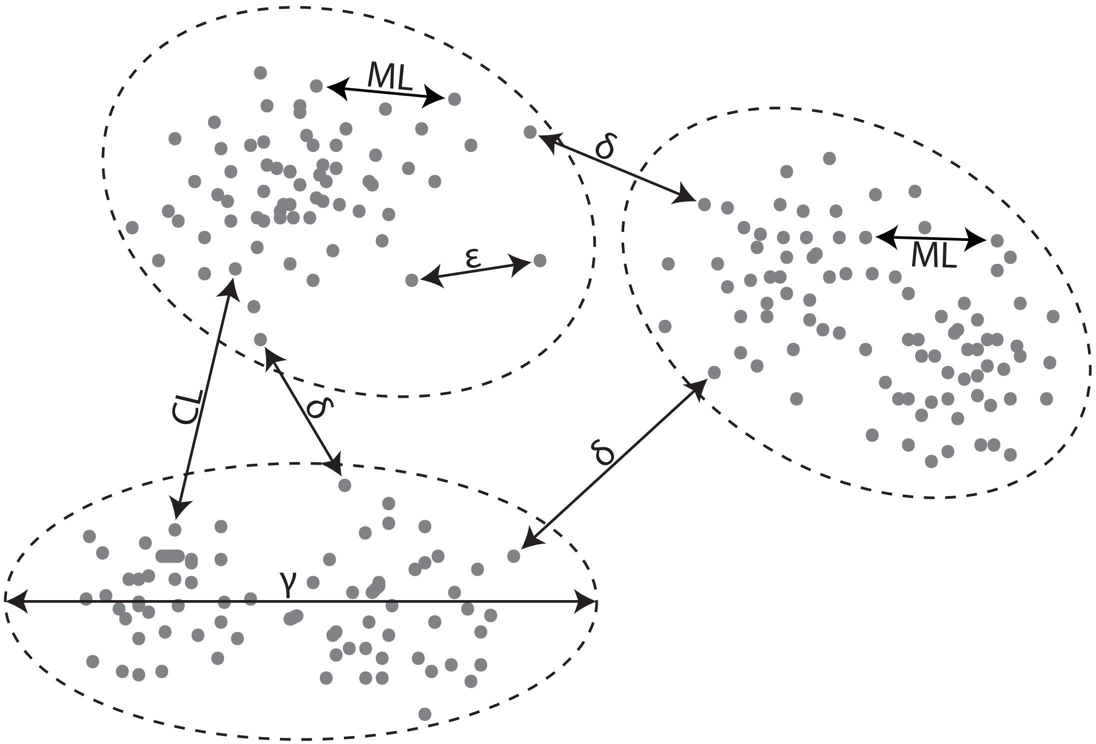
- Supervised learning (SVM, deep learning, etc)
  - Offer excellent classification performance in several areas
  - Require large volumes of training data
- Unsupervised learning (clustering, EM, etc)
  - Data exploration, offer insights into the data
  - Do not require labelled examples
  - Results may not be relevant to the user's requirements
- Semi-supervised learning (constrained clustering)
  - Fits between the two, large amount of unlabelled data with some information for a (generally) small subset

Goal



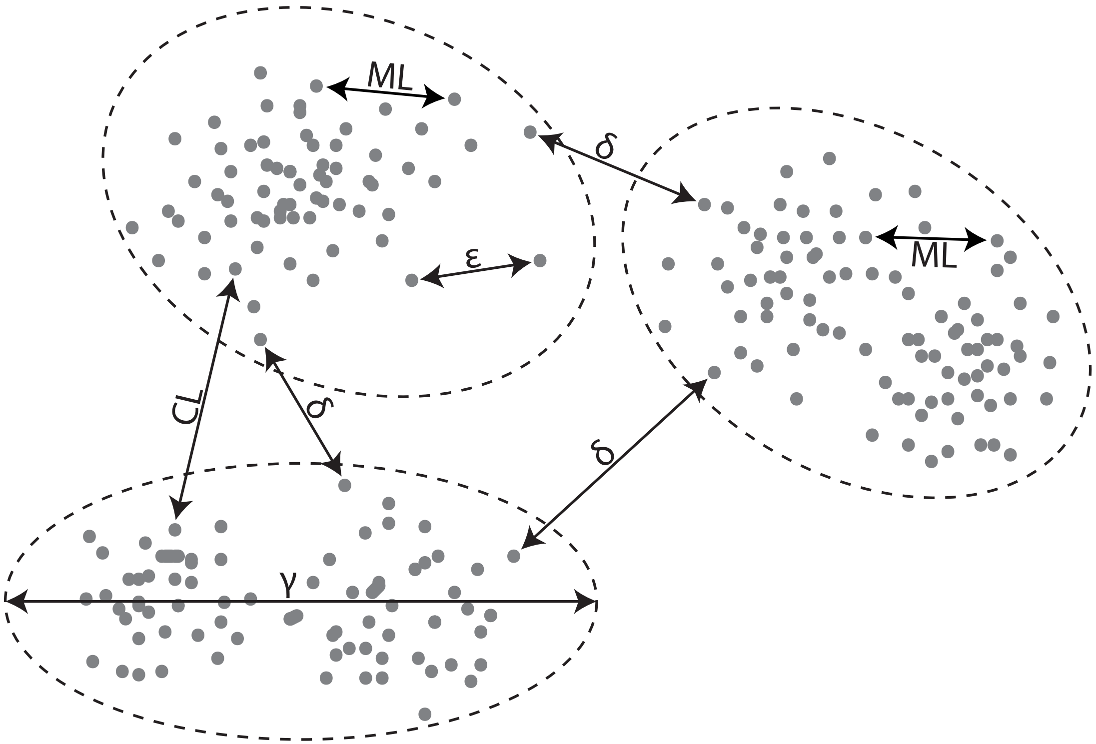
Initialisation (1)
Clustering
contraint initial
Clustering sous contraintes incrémental (2.1)
cluster #1
cluster #2
cluster #3
cluster #4
cluster #5
cluster #6
Clusters candidats
centers :
Nouvelles contraintes
sur clusters invalidés (2.2)
Contraintes initiales
« Sémantisation »
Enrichissement de la typologie (3)
Clusters validés
Classes
thématiques

Types of Constraints: Cluster-Level



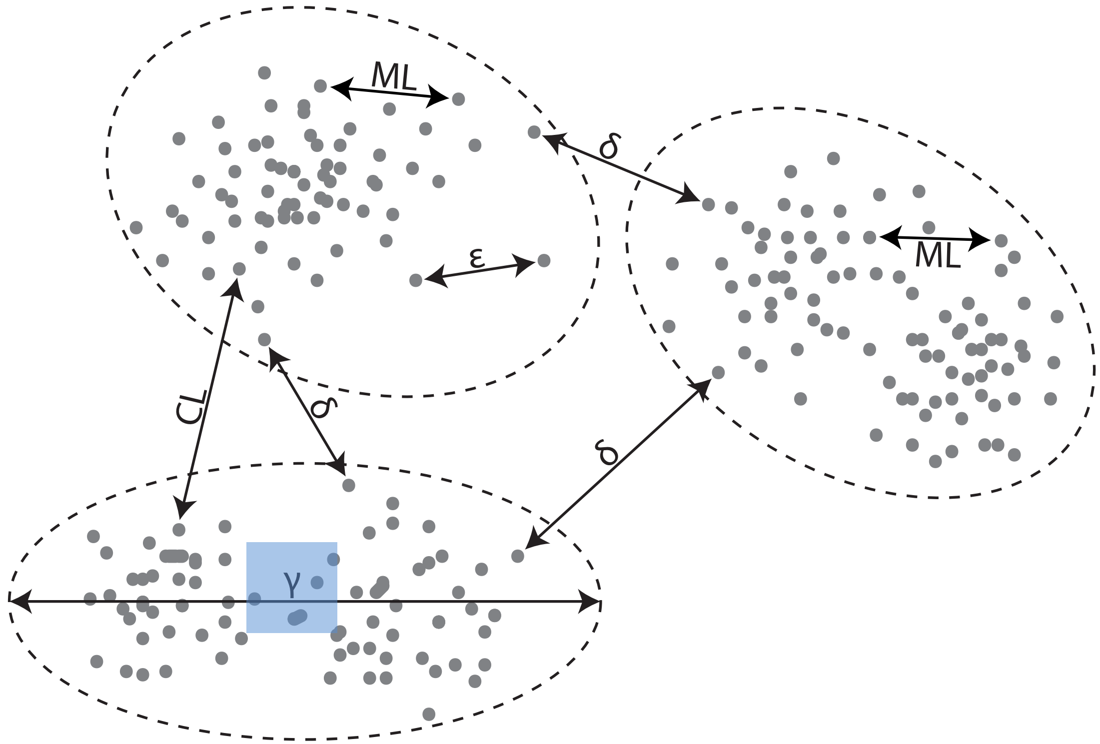
- Number of clusters
  - User: number of distinct objects

Types of Constraints: Cluster-Level



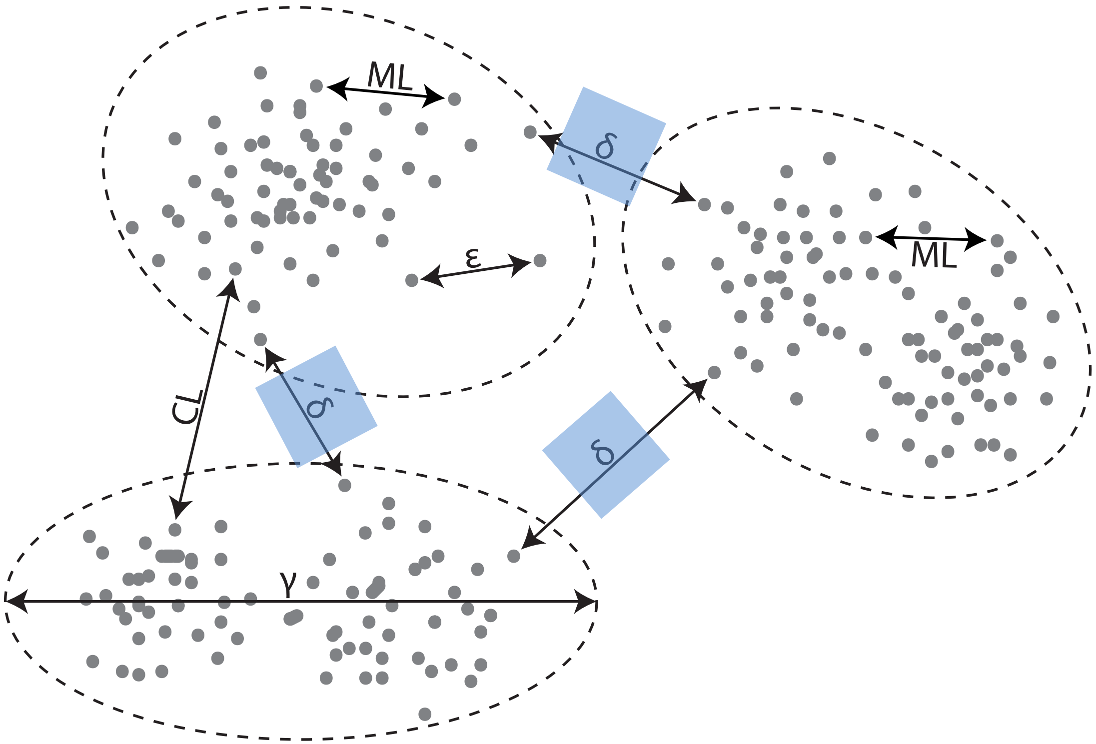
- Absolute (or relative) maximal (or minimal) size
  - User: number of samples for each object

Types of Constraints: Cluster-Level



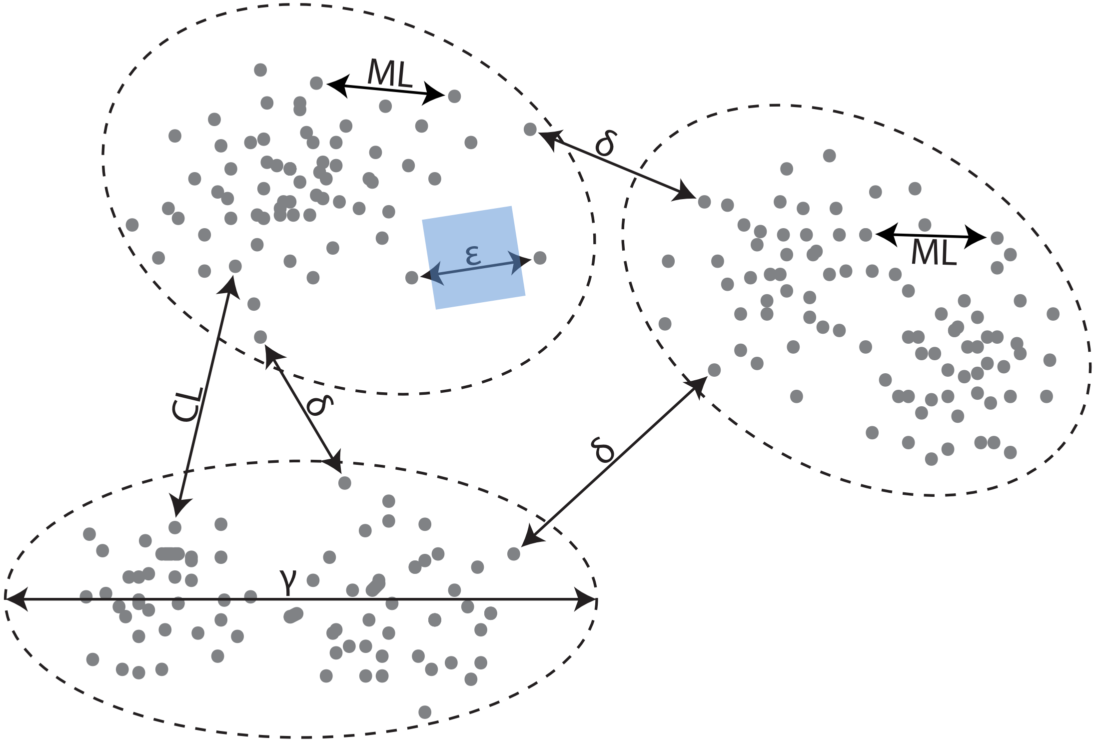
- Maximum diameter γ
  - User: inter class dissimilarity

Types of Constraints: Cluster-Level



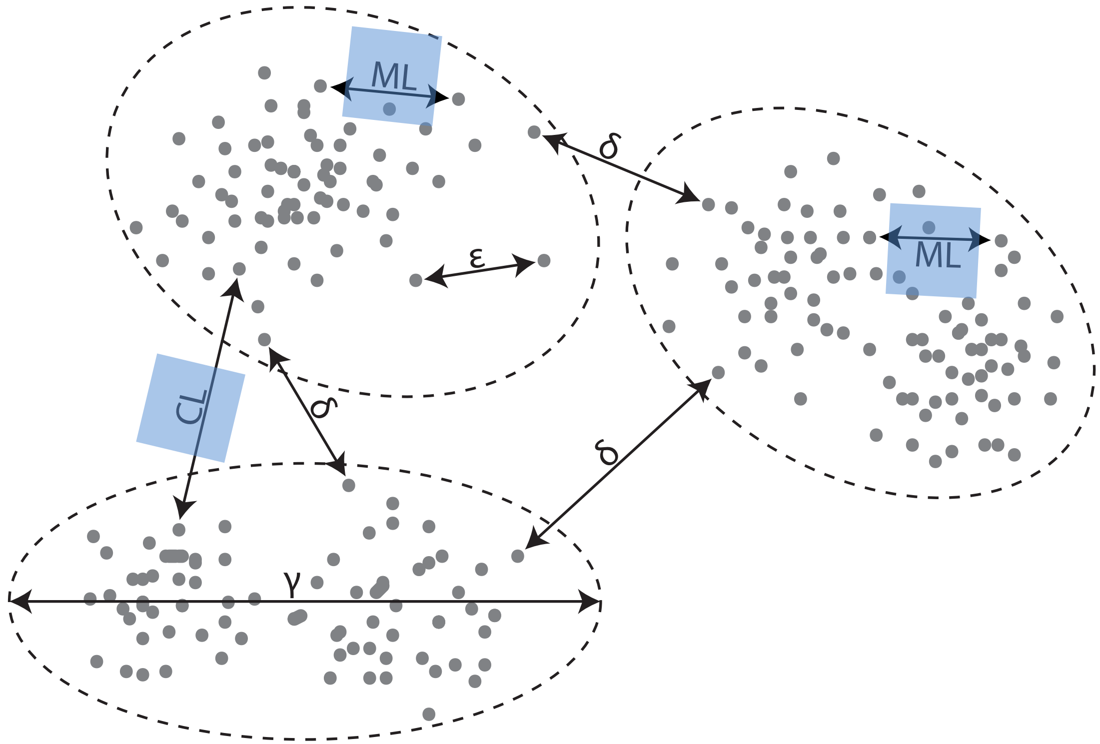
- Split: clusters must be separated by at least δ [Davidson and Ravi, 2005]
  - User: intra class similarity

Types of Constraints: Cluster-Level



- ε-constraint: each point has within a radius ε at least one other point in the same cluster [Davidson and Ravi, 2005]
  - User: difficult to interpret

Types of Constraints: Instance-Level



- Must/Cannot Link
  - User: objects of the same/different type(s)

iCUBE

Constrained Clustering Algorithms

- 6 general approaches:
  - k-Means
  - Metric Learning
  - Spectral Graph Theory
  - Ensemble Clustering
  - Collaborative Clustering
  - Declarative

Constrained Clustering Algorithms

- 71 papers proposing algorithms

Category	Method
k-Means	COP-COBWEB (Wagstaff and Cardie, 2000) COP-KMeans (Wagstaff et al, 2001) Seed-KMeans (Basu et al, 2002) Constrained-KMeans (Basu et al, 2002) ICOP-KMeans (Tan et al, 2010) Sequenced Assignment COP-KMeans (Rutayisire et al, 2011) MLC-KMeans (Huang et al, 2008) SCREEN (Tang et al, 2007) GA Dispersion & Impurity (Demiriz et al, 1999) CVQE (Davidson and Ravi, 2005) LCVQE (Pelleg and Baras, 2007) PCK-Means (Basu et al, 2004b) Lagrangian Relaxation (Ganji et al, 2016) Tabu Search (Hiep et al, 2016) Fuzzy CMeans (Grira et al, 2006) Non-Negative Matrix Factorisation (Li et al, 2007) Mathematical Program (Ng, 2000) Minimal Capacity Constraints (Bradley et al, 2000) Balanced Clustering (Banerjee and Ghosh, 2006) Minimal Size (Demiriz et al, 2008) Minimal Size & Balanced Clustering (Ge et al, 2007)
Metric Learning	Euclidean (Klein et al, 2002) Mahanalobis (Bar-Hillel et al, 2003, 2005; Xing et al, 2002) Kullback-Leibler Divergence (Cohn et al, 2003) String-Edit Distance (Bilenko and Mooney, 2003) LRML (Hoi et al, 2008, 2010) Partially Observed Constraints (Yi et al, 2012)
k-Means & Metric Learning	MPCK-Means (Bilenko et al, 2004) HMRF-KMeans (Basu et al, 2004b) Semi-Supervised Kernel k-Means (Kulis et al, 2005, 2009) CLWC (Cheng et al, 2008)
Spectral Graph Theory	Adjacency Matrix Modification (Kamvar et al, 2003) Out-Of-Sample Adjacency Matrix Modification (Alzate and Suykens, 2009) CSP (Wang and Davidson, 2010a; Wang et al, 2014) Constraint Satisfaction Lower Bound (Wang et al, 2010) Inconsistent Constraints (Rangapuram and Hein, 2012) Logical Constraint Combinations (Zhi et al, 2013) Distance Modification (Anand and Reddy, 2011) Constraint Propagation Binary Class (Lu and Carreira-Perpiñán, 2008) Constraint Propagation Multi-Class (Lu and Ip, 2010; Chen and Feng, 2012; Ding et al, 2013) Kernel Matrix Learning (Zhang and Ando, 2006; Hoi et al, 2007; Li and Ding, 2008; Li and Liu, 2009) Guaranteed Quality Clustering (Cucuringu et al, 2016)
Ensemble Clustering	SCEV (Iqbal et al, 2012) Consensus Function (Al-Razgan and Domeniconi, 2009; Xiao et al, 2016; Dimitriadou et al, 2002)
Collaborative Clustering	Samarah (Forestier et al, 2010a) Penta-Training (Domeniconi and Al-Razgan, 2008)
Declarative Approaches	SAT (Davidson et al, 2010) CP (Dao et al, 2013, 2016, 2017; Guns et al, 2016) ILP Column Generation (Merle et al, 1999; Aloise et al, 2012; Babaki et al, 2014) Restricted Cluster Candidates (Mueller and Kramer, 2010; Ouali et al, 2016)
Miscellaneous	Constrained EM (Shental et al, 2013) Evolutionary Algorithm (Handl and Knowles, 2006) Random Forest (Zhu et al, 2016)

Algorithms: k-Means

- Extends the classic k-Means algorithm
  - Check whether proposed clustering modifications violate constraints
  - Or add penalty for violating constraints
- Some guarantee constraint satisfaction
  - Problem with noisy constraints
- Iterative
  - May only find local optimum

Algorithms: Metric Learning

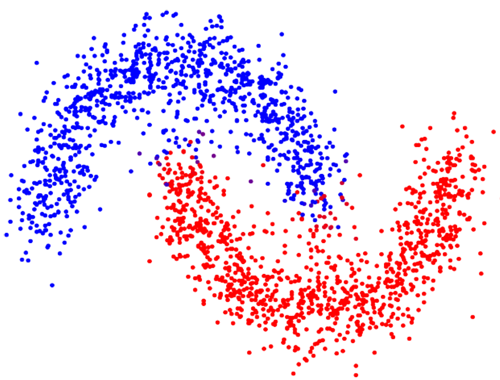
- Uses constraints to build
  - Set of similar points from must-link
  - Set of dissimilar points from cannot-link
- These are used to learn the metric
- Each algorithm is dependent upon the metric upon which it is based (i.e. Euclidean, Mahalanobis, etc)

Algorithms: Constrained Spectral Clustering

- Based on the graph cut of a graph $\mathrm{G} = (X, A)$
  - Edges described by a similarity matrix A:
    - $a_{ij} = \exp\left(\frac{-d_{ij}}{2\sigma^2}\right)$,
    - d_{ij} — distance (dissimilarity) between x_i and x_j
    - σ — user defined parameter
- Method:
  - Construct normalised Laplacian matrix: $L = D^{1/2}(D - A)D^{1/2}$,
    - D — degree matrix
  - Determine the first k eigenvectors
  - Embed graph into low dimensional space (another parameter)
    - Perform k-Means in this space

Algorithms: Constrained Spectral Clustering

- Integrate constraints directly into the similarity matrix
  - Force values of A to be 0 (resp. 1) for points under CL (resp. ML) constraints
  - Use constraints as regularisation factors
- Advantages
  - Clusters are not necessarily connected
  - Polynomial time calculation
  - Accepts over constrained problems
- Limits
  - No guarantee of optimality
  - Not incremental
  - No cluster level or label constraints
  - Eigenvalues for large graphs may be ill conditioned



Algorithms: Ensemble

- Apply many diverse algorithms (or same with different parameters) and take consensus
- Constraints integrated:
  - each learning agent integrates them in its own fashion
    - Problem: constraints restrict diversity
  - or apply them in the consensus function
    - Builds a graph clustering approach on top, constraints guide this

Algorithms: Collaborative

- Collaborative approach
  - Execute several clustering algorithms in parallel
  - Process guided by a coefficient taking into account:
    - Similarity of the results
    - Quality of the results
    - Level of constraint satisfaction

$$
\begin{aligned}
\gamma^{(i,j)} = & p_s . \left(\frac{1}{n_i} \sum_{k=1}^{n_i} \omega_k^{(i,j)} + \frac{1}{n_j} \sum_{k=1}^{n_j} \omega_k^{(j,i)} \right) \\
& + p_q . \left(\delta^{(i)} + \delta^{(j)} \right) \\
& + p_c . \left(\theta^{(i)} + \theta^{(j)} \right)
\end{aligned}
$$

Algorithms: Collaborative

- Advantages
  - Clusters not necessarily convex
  - Accepts over constrained problems
  - Incremental
- Limits
  - Long processing time
  - No guarantee of optimality
  - No cluster level or label constraints

Algorithms: Declarative

- Optimisation problem under constraints $S = (X, C, F)$
  - S — clustering to be determined
  - $X = \{x_1, \dots, x_n\}$ — data
  - $C = \{c_1, \dots, c_p\}$ — constraints
  - $F(X, C, F)$ — objective function (e.g. intra-class inertia)
- Search for a/the best solution — PPC
  - Branch-and-bound optimisation: guarantees a global optimum (if it exists) satisfying all constraints
- Large choice of constraints

Algorithms: Declarative

- CPClustering [Dao et al. 2017]
  - Number of clusters: limited, i.e. $K \in [K_{min}, K_{max}]$
  - Optimisation criteria: cluster diameter; between cluster distance; intra/inter-class inertia, …
  - Constraints: must-link/cannot-link, size, diameter or density of clusters, …
- Limits:
  - Processing time
  - Amount of data
  - No solution if over-constrained
  - Not incremental

Constrained Clustering Implementations

- 12 implementations available

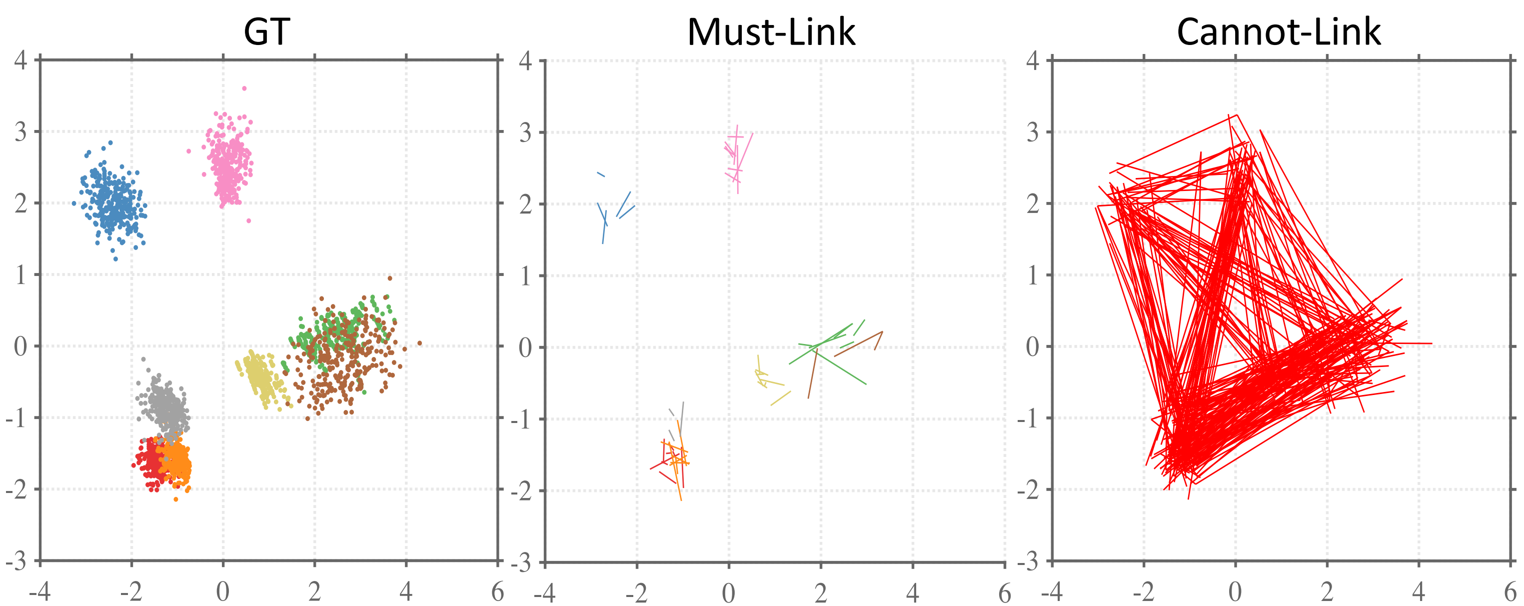
Category	Method
k-Means	COBWEB (Fisher, 1987) COP-KMeans (Wagstaff et al, 2001) Seed-KMeans (Basu et al, 2002) Constrained-KMeans (Basu et al, 2002) ICOP-KMeans (Tan et al, 2010) Sequenced Assignment COP-KMeans (Rutayisire et al, 2011) MLC-KMeans (Huang et al, 2008) SCREEN (Tang et al, 2007) GA Dispersion & Impurity (Demiriz et al, 1999) CVQE (Davidson and Ravi, 2005) LCVQE (Pelleg and Baras, 2007) PCK-Means (Basu et al, 2004b) Lagrangian Relaxation (Ganji et al, 2016) Tabu Search (Hiep et al, 2016) Fuzzy CMeans (Grira et al, 2006) Non-Negative Matrix Factorisation (Li et al, 2007) Mathematical Program (Ng, 2000) Minimal Capacity Constraints (Bradley et al, 2000) Balanced Clustering (Banerjee and Ghosh, 2006) Minimal Size (Demiriz et al, 2008) Minimal Size & Balanced Clustering (Ge et al, 2007)
Metric Learning	Euclidean (Klein et al, 2002) Mahanalobis (Bar-Hillel et al, 2003, 2005; Xing et al, 2002) Kullback-Leibler Divergence (Cohn et al, 2003) String-Edit Distance (Bilenko and Mooney, 2003) LRML (Hoi et al, 2008, 2010) Partially Observed Constraints (Yi et al, 2012)
k-Means & Metric Learning	MPCK-Means (Bilenko et al, 2004) HMRF-KMeans (Basu et al, 2004b) Semi-Supervised Kernel *k*-Means (Kulis et al, 2005, 2009) CLWC (Cheng et al, 2008)
Spectral Graph Theory	Adjacency Matrix Modification (Kamvar et al, 2003) Out-Of-Sample Adjacency Matrix Modification (Alzate and Suykens, 2009) CSP (Wang and Davidson, 2010a; Wang et al, 2014) Constraint Satisfaction Lower Bound (Wang et al, 2010) Inconsistent Constraints (Rangapuram and Hein, 2012) Logical Constraint Combinations (Zhi et al, 2013) Distance Modification (Anand and Reddy, 2011) Constraint Propagation Binary Class (Lu and Carreira-Perpiñán, 2008) Constraint Propagation Multi-Class (Lu and Ip, 2010; Chen and Feng, 2012; Ding et al, 2013) Kernel Matrix Learning (Zhang and Ando, 2006; Hoi et al, 2007; Li and Ding, 2008; Li and Liu, 2009) Guaranteed Quality Clustering (Cucuringu et al, 2016)
Ensemble Clustering	SCEV (Iqbal et al, 2012) Consensus Function (Al-Razgan and Domeniconi, 2009; Xiao et al, 2016; Dimitriadou et al, 2002)
Collaborative Clustering	SAMARAH (Forestier et al, 2010a) Penta-Training (Domeniconi and Al-Razgan, 2008)
Declarative Approaches	SAT (Davidson et al, 2010) CP (Dao et al, 2013, 2016, 2017; Guns et al, 2016) ILP Column Generation (Merle et al, 1999; Aloise et al, 2012; Babaki et al, 2014) Restricted Cluster Candidates (Mueller and Kramer, 2010; Ouali et al, 2016)
Miscellaneous	Constrained EM (Shental et al, 2013) Evolutionary Algorithm (Handl and Knowles, 2006) Random Forest (Zhu et al, 2016)

Experiments

- 5 implementations evaluated
  - COP-KMeans [Wagstaff et al, 2001]
  - Spectral clustering (adjacency modification) [Kamvar et al, 2003]
  - Spectral clustering (with regularisation) [Li and Liu, 2009]
  - SAMARAH (collaborative) [Forestier et al, 2010]
  - CPClustering (declarative) [Dao et al, 2017]
- 9 UCR datasets (time-series)

Experiments

- Create N_c constraints using the reference data
  - $N_c \in \{5\%, 10\%, 15\%, 50\%\}$ of the number of objects
  - Total number of possible constraints: $\frac{N(N-1)}{2}$
  - i.e. a very small ratio (50%): $\frac{1}{N-1}$



Experiments

- Create N_c constraints using the reference data
    - $N_c \in \{5\%, 10\%, 15\%, 50\%\}$ of the number of objects
    - Total number of possible constraints: $\frac{N(N-1)}{2}$
    - i.e. a very small ratio (50%): $\frac{1}{N-1}$
- Validation:
    - ARI: Adjusted Rand Index
    - CSR: Constraints Satisfaction Ratio

Results

- ARI difference: constrained ARI – unconstrained ARI

Method	5%	10%	15%	50%
COP-KMeans	0.004 (0.053)	0.011 (0.061)	−0.001 (0.052)	0.011 (0.067)
Spec	−0.034 (0.068)	−0.030 (0.062)	−0.043 (0.078)	−0.049 (0.101)
SpecReg	0.262 (0.298)	0.263 (0.297)	0.263 (0.296)	0.261 (0.296)
CPClustering	−0.029 (0.089)	−0.023 (0.067)	−0.021 (0.076)	−0.036 (0.085)
SAMARAH	0.066 (0.091)	0.070 (0.084)	0.064 (0.084)	0.076 (0.094)

- CSR difference: constrained CSR – unconstrained CSR

Method	5%	10%	15%	50%
COP-KMeans	0.133 (0.078)	0.133 (0.078)	0.134 (0.077)	0.136 (0.075)
Spec	−0.006 (0.063)	0.001 (0.037)	−0.011 (0.052)	−0.012 (0.050)
SpecReg	0.099 (0.129)	0.100 (0.125)	0.104 (0.123)	0.099 (0.124)
CPClustering	0.215 (0.063)	0.215 (0.063)	0.215 (0.063)	0.215 (0.063)
SAMARAH	0.057 (0.043)	0.050 (0.038)	0.044 (0.034)	0.040 (0.034)

Statistical Analysis

- Four potential ‘explanations’ for the loss/stagnation of quality:
  - Consistency: does the unconstrained clustering already fulfil the constraints [Wagstaff et al. 2006, Davidson et al. 2006]
  - Incoherence: amount of internal conflict between the constraints given a distance metric [Wagstaff et al. 2006, Davidson et al. 2006]
  - Cluster overlap: initial classes are not separable, i.e. overlapping
  - Algorithm: sensitivity of the algorithm to constraints

Statistical Analysis

- Four potential ‘explanations’ for the loss/stagnation of quality:
  - Consistency: does the unconstrained clustering already fulfil the constraints [Wagstaff et al. 2006, Davidson et al. 2006]
  - Incoherence: amount of internal conflict between the constraints given a distance metric [Wagstaff et al. 2006, Davidson et al. 2006]
  - Cluster overlap: initial classes are not separable, i.e. overlapping
  - Algorithm: sensitivity of the algorithm to constraints

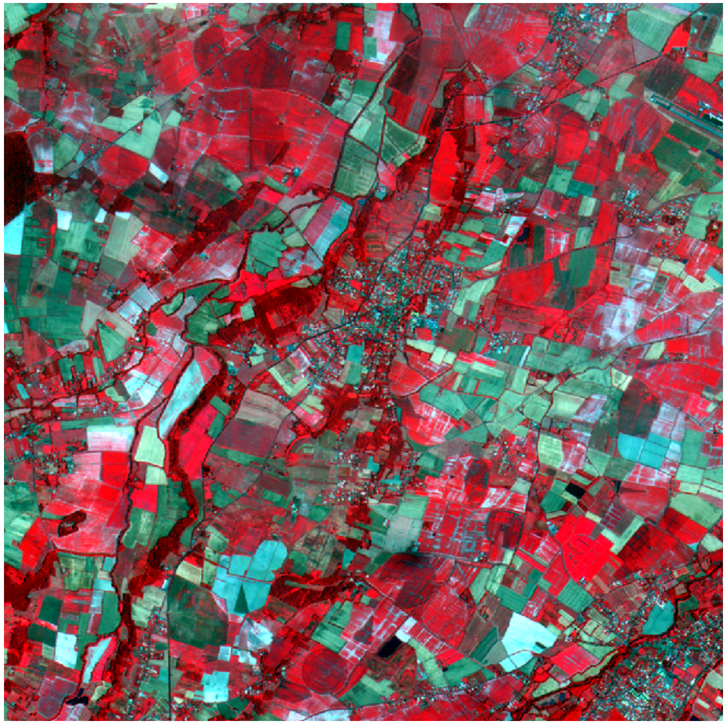
Predictor	Estimate	p-value
Consistency	-0.745	$4.923e{-87}$
Silhouette Score	0.419	$2.636e{-66}$
COP-KMeans	0.037	$3.572e{-4}$
Spec	-0.056	$3.553e{-15}$
SpecReg	0.220	$1.819e{-88}$
CPClustering	-0.059	$9.943e{-9}$
SAMARAH	0.088	$7.109e{-16}$

1635 samples (not all experiments finished on time)

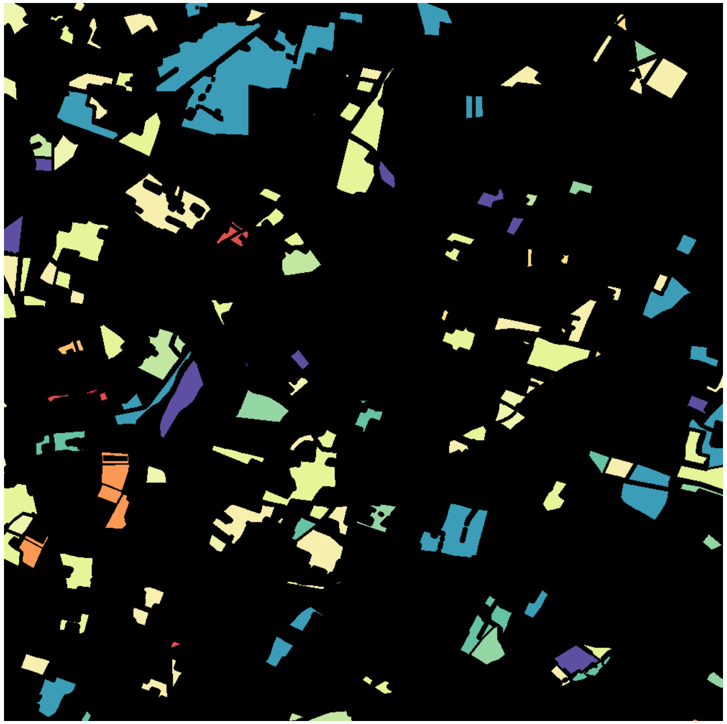
Satellite Image Time-Series

- Zone Sud-Ouest : 11 images, 2007
- Objective: agricultural classification

Image

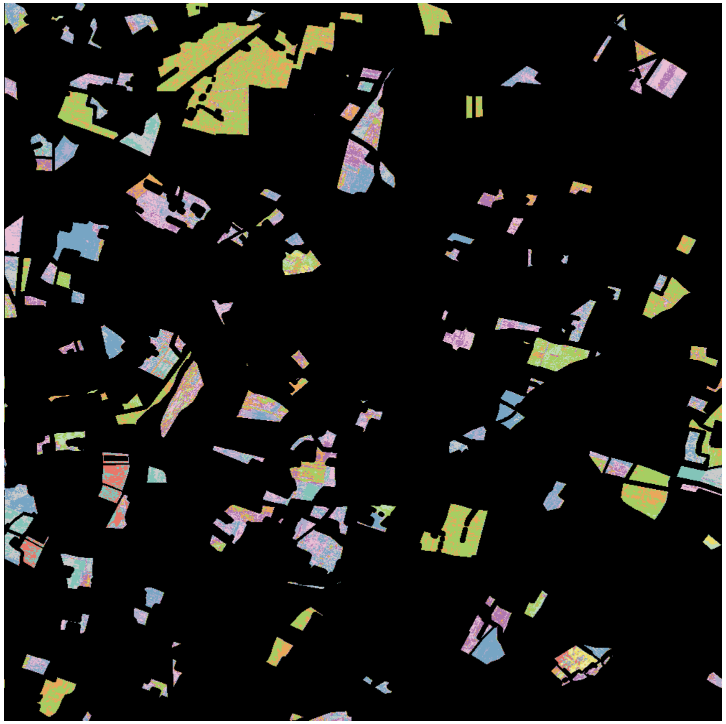


Ground Truth

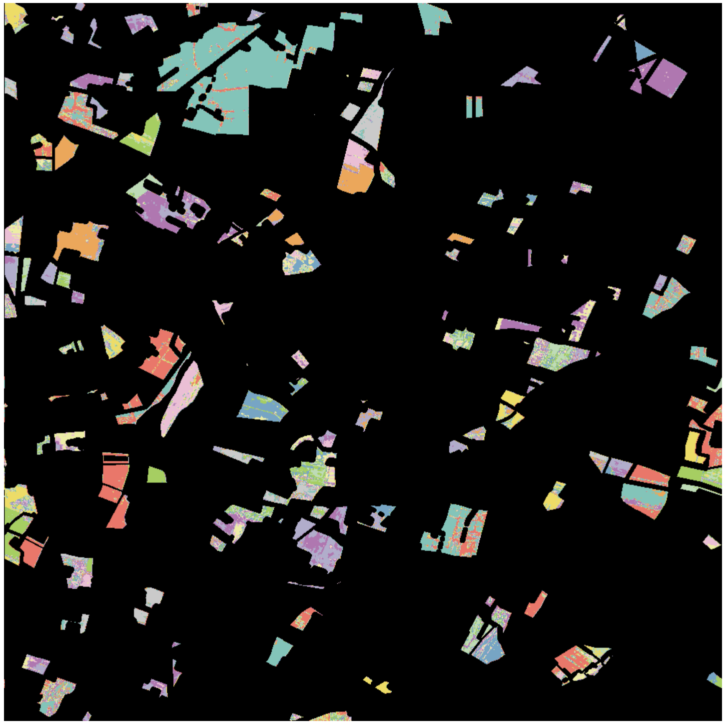


Satellite Image Time-Series

CPClustering

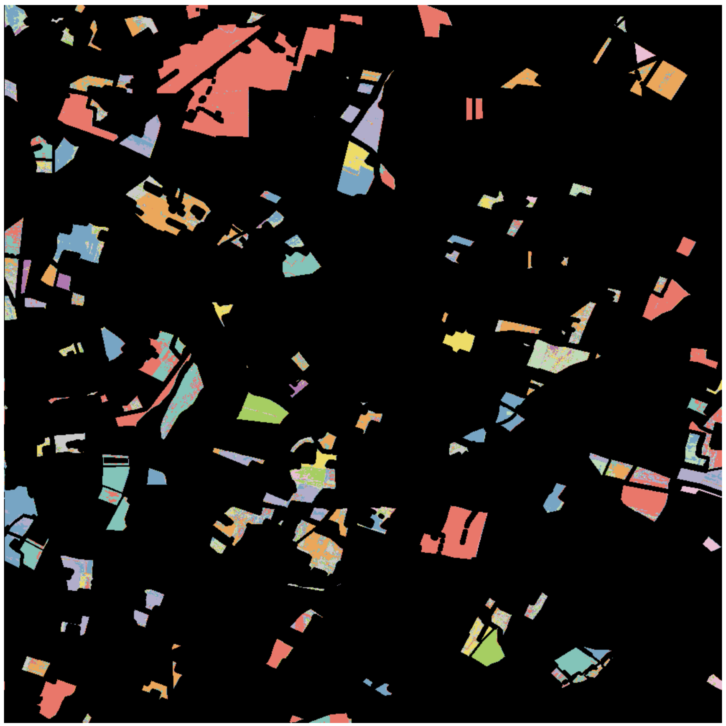


SpecReg



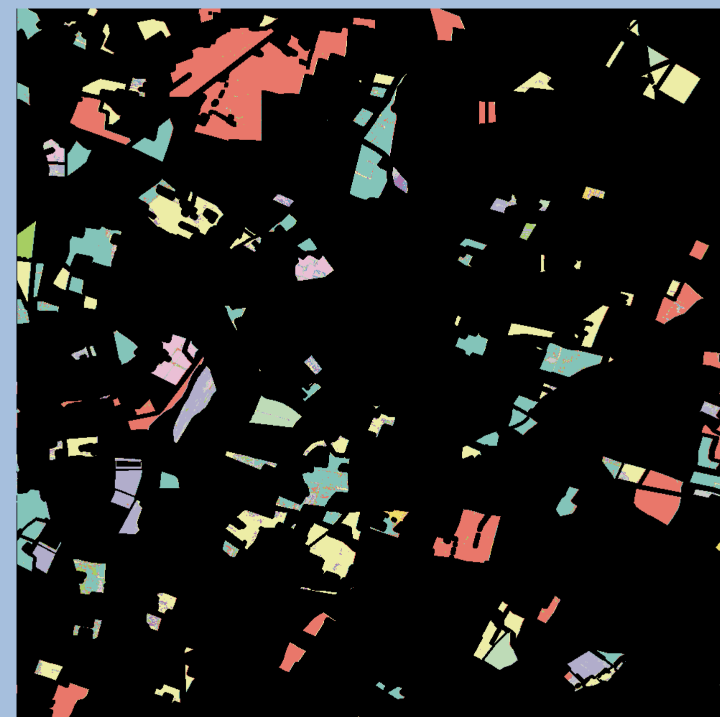
Satellite Image Time-Series

Spec



SAMARAH

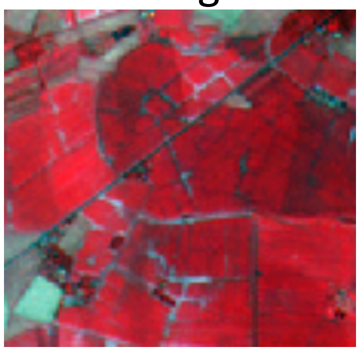
BEST



Satellite Image Time-Series

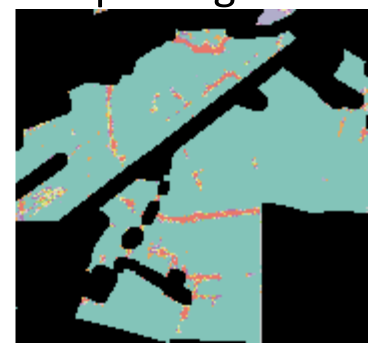
- Zoom …

Image



Spec

SpecReg



Ground Truth

CPClustering

SAMARAH

Conclusions

- Promising results on time-series
- Coherence and cluster overlap explain some of the variance observed

- Future
  - Incremental constraints
  - Sampling constraints
    - Coherence measure for constraints using DTW?
  - When do algorithms rely more on data vs constraints?