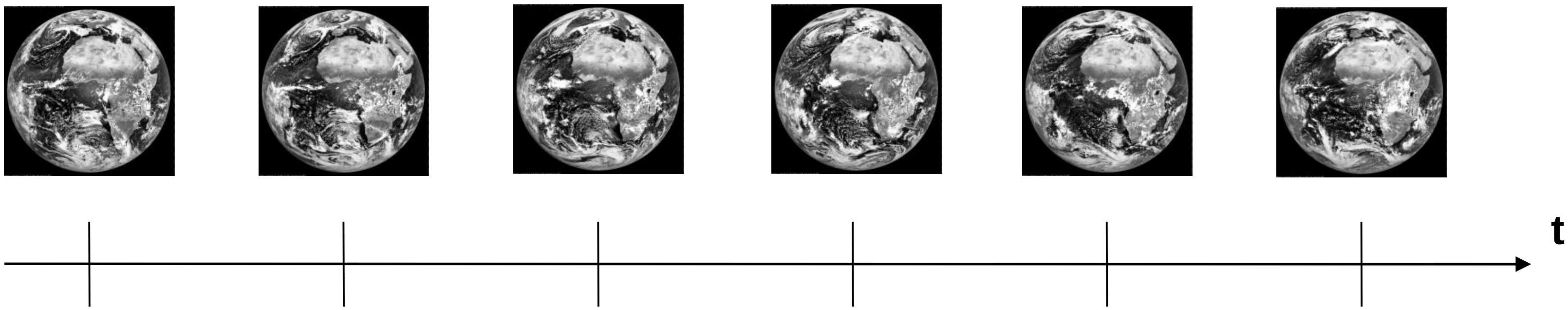
A swap randomization approach for mining motion field time series over the Argentière glacier

Youen Pericault, Catherine Pothier, Nicolas Méger, Christophe Rigotti, Flavien Vernier, Ha Thai Pham, Emmanuel Trouvé

MultiTemp 2015, Annecy.

Previous work

- Satellite Image Time Series (SITS) can be mined using Grouped Frequent Sequential patterns (GFS-patterns) [Julea11].
  - Unsupervised techniques implemented at the pixel level
  - GFS-patterns can be numerous
- End users can be guided towards the most promising GFS-patterns thanks to a swap randomization ranking method [Méger12].
  - The stability of the ranking method w.r.t the number of swaps has been evidenced for a single randomization seed.
  - The focus was on patterns with occurrences strongly affected by a randomization of the SITS.



Proposal

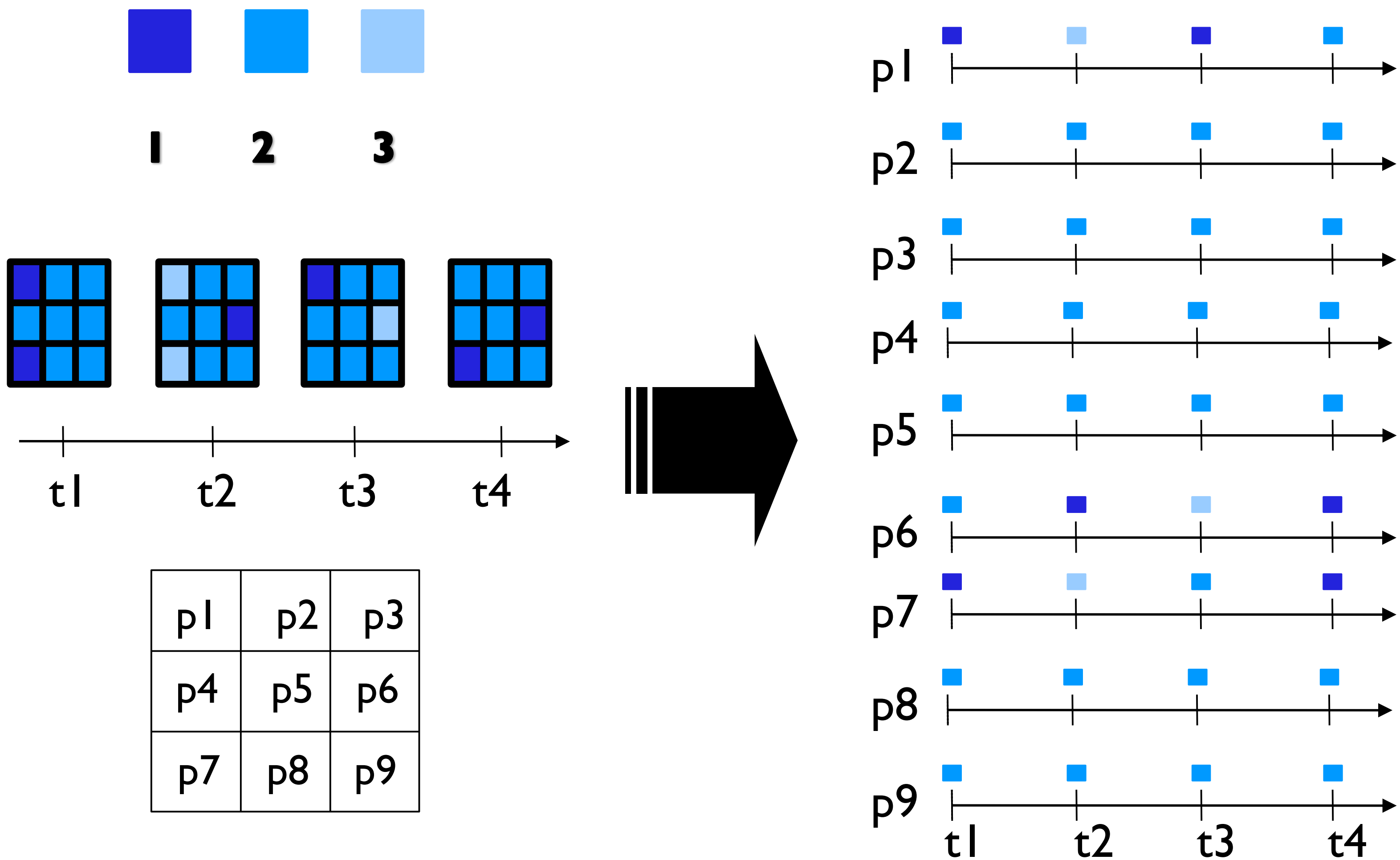
- Investigating the extraction and ranking of GFS-patterns on a Displacement Field Time Series (DFTS) obtained by terrestrial photogrammetry.

- Assessing the stability of the swap randomization ranking method for multiple randomizations of the DFTS.

- Extending the focus on GFS-patterns with occurrences destroyed OR maintained by randomization.

SITS and base of sequences

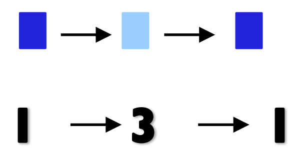
Pixel values are described using symbols (e.g., quantization, clustering).



Base of sequences [Agrawal95]

Frequent sequential patterns

- Model: sequential patterns

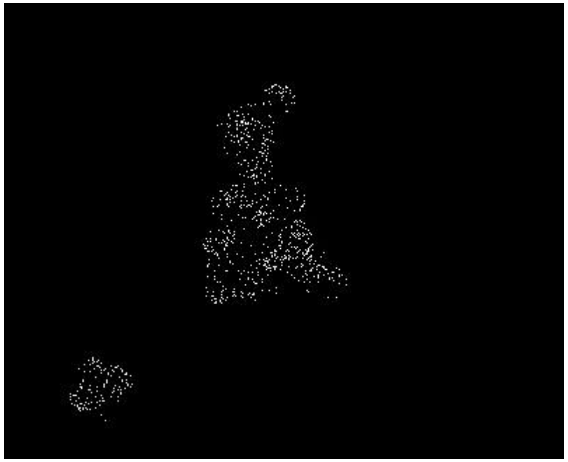


- Measure: support, i.e., the number of pixels affected by a sub-evolution.
- Sequential patterns covering at least σ pixels are frequent sequential patterns.
- σ, the minimum support, is used to prune the search space.

[Agrawal95]

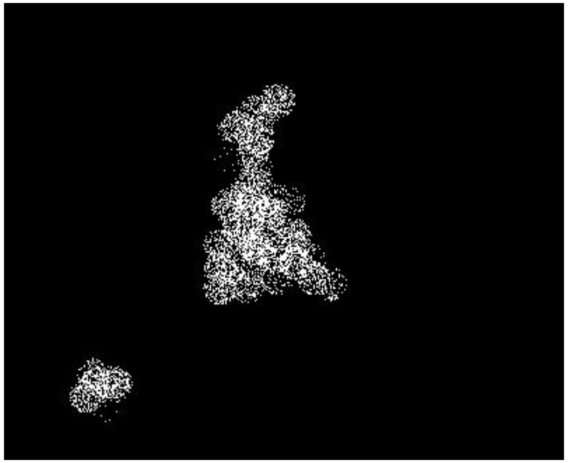
Towards spatiality

σ is a minimum surface



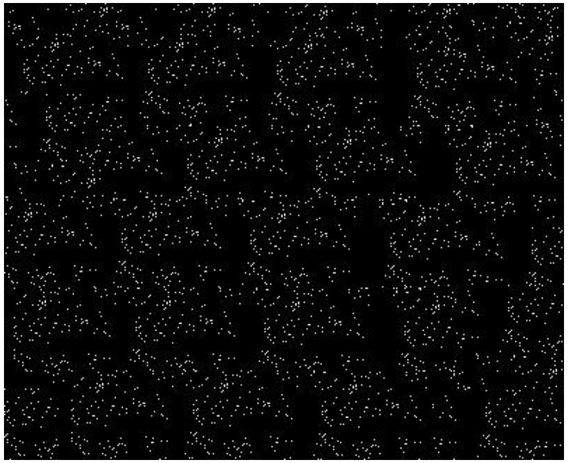
1 → 2 → 2 → 1

support < σ



1 → 3 → 1

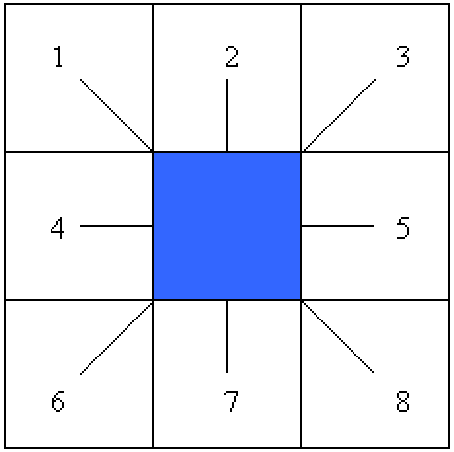
support ≥ σ



3 → 3 → 1

support ≥ σ

Incorporating spatiality: the GFS-patterns



- Measure: the average connectivity (AC) of evolution α, i.e. the average number of pixels covered by α in the 8-neighborhood

- Frequent sequential patterns covering pixels that are sufficiently connected (AC(α) ≥ κ) are Grouped Frequent Sequential patterns (GFS-patterns).

- κ is the minimum average connectivity and is also used to prune the search space.

[Julea11]

Maximal GFS-Patterns

- Definition : A maximal GFS pattern has the property of not being a sub-pattern of any other of the extracted GFS patterns.

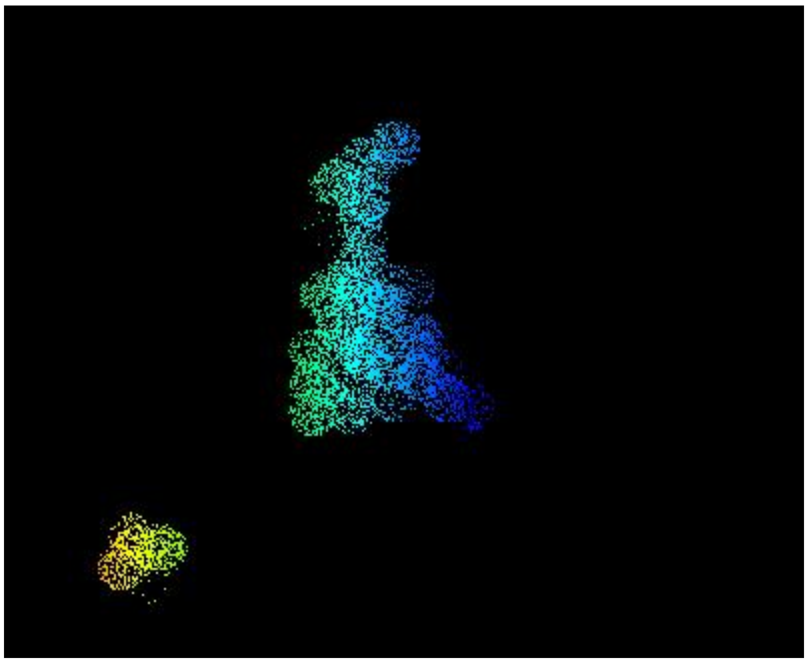
- (2,1) and (2,2) are sub-patterns of (2,1,2) therefore (2,1,2) is the only maximal GFS pattern of the set {(2,1),(2,2) ,(2,1,2)}.

- Selecting maximal patterns enables to focus on the most specific evolutions.

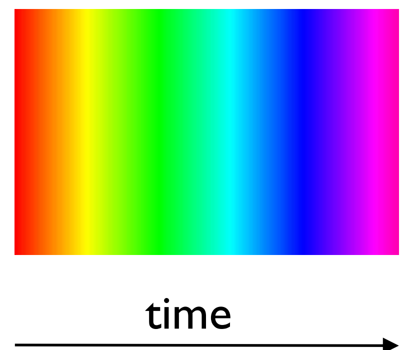
- Experiments on an optical Landsat SITS over New Caledonia : 15620 GFS-Patterns ; 295 maximal GFS-Patterns.

[Julea12]

SpatioTemporal Localization maps (STL-maps)

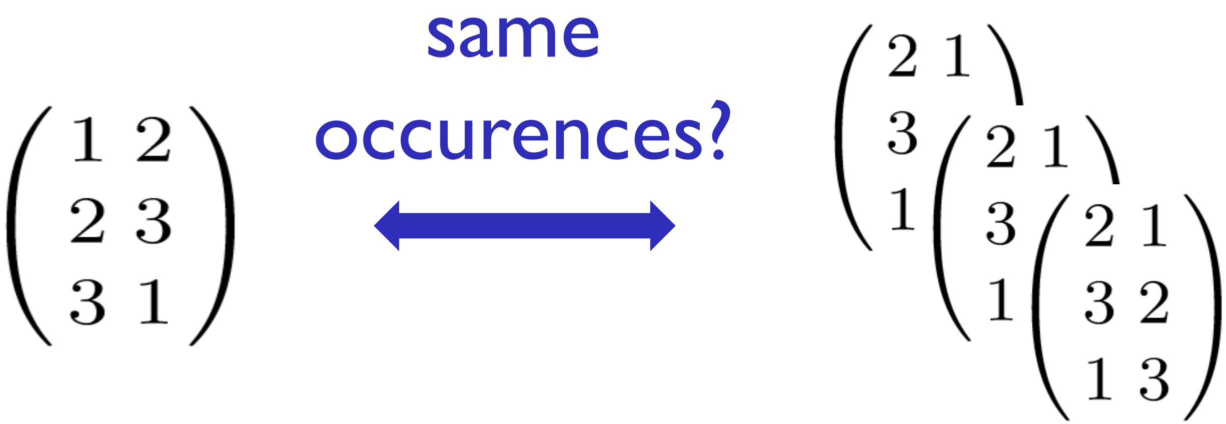


1 → 2 → 2



[Méger11]

Swap randomization for symbolic matrices



- A base of sequences expressing SITS can be expressed as a symbolic matrix: row ⇔pixel, column ⇔date

- To assess GFS-patterns obtained from symbolic matrices representing a SITS.

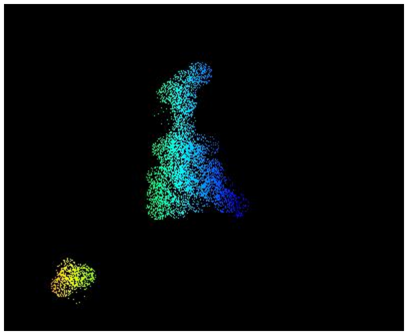
- The spatiotemporal nature of the observed phenomena must be preserved.

Swap randomization for symbolic matrices

$$C = \begin{pmatrix} \underline{3} & \underline{2} \\ 1 & 1 \\ \underline{2} & \underline{3} \end{pmatrix}, C' = \begin{pmatrix} \underline{2} & \underline{3} \\ 1 & 1 \\ \underline{3} & \underline{2} \end{pmatrix}$$

- Pairs of elements sharing the same symbol are chosen at random.
- If a pair B(i,j) = B(k,l) = α and if B(k,j) = B(i,l) = β (α ≠ β) then α's and β's are swapped.
- Column and row margins are maintained while GFS-patterns occurrences are affected.
- Failed swap attempts are counted as self-loops to explore equiprobable matrices.

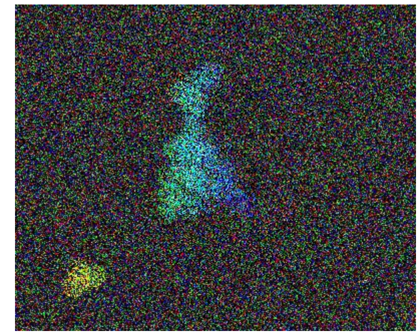
Normalized Mutual Information (NMI)



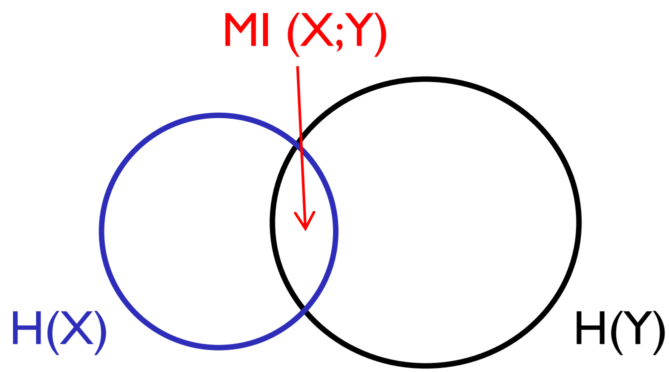
STL-map
original SITS

How similar?

X ⟷ Y



STL-map
swap-randomized SITS

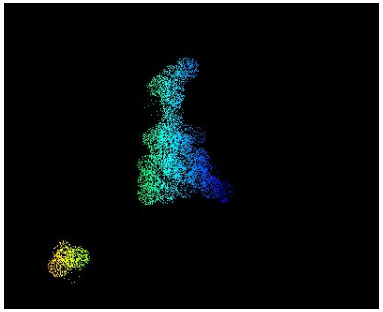


$$NMI(X;Y) = \frac{\sum_{x,y\in\Omega^2} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}}{min(H(X),H(Y))}$$

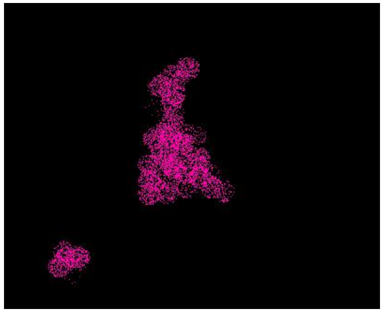
$$H(X) = -\sum_{x\in\Omega} P(x)\log P(x),$$

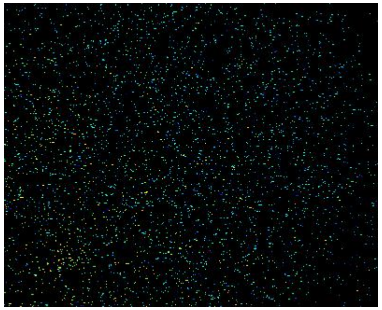
[Méger12]

NMI-based ranking

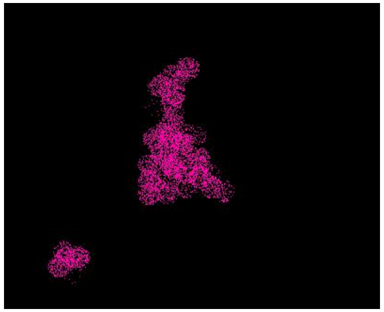


Original SITS





RandomizedSITS



Destroyed by randomization

Hardly altered by randomization

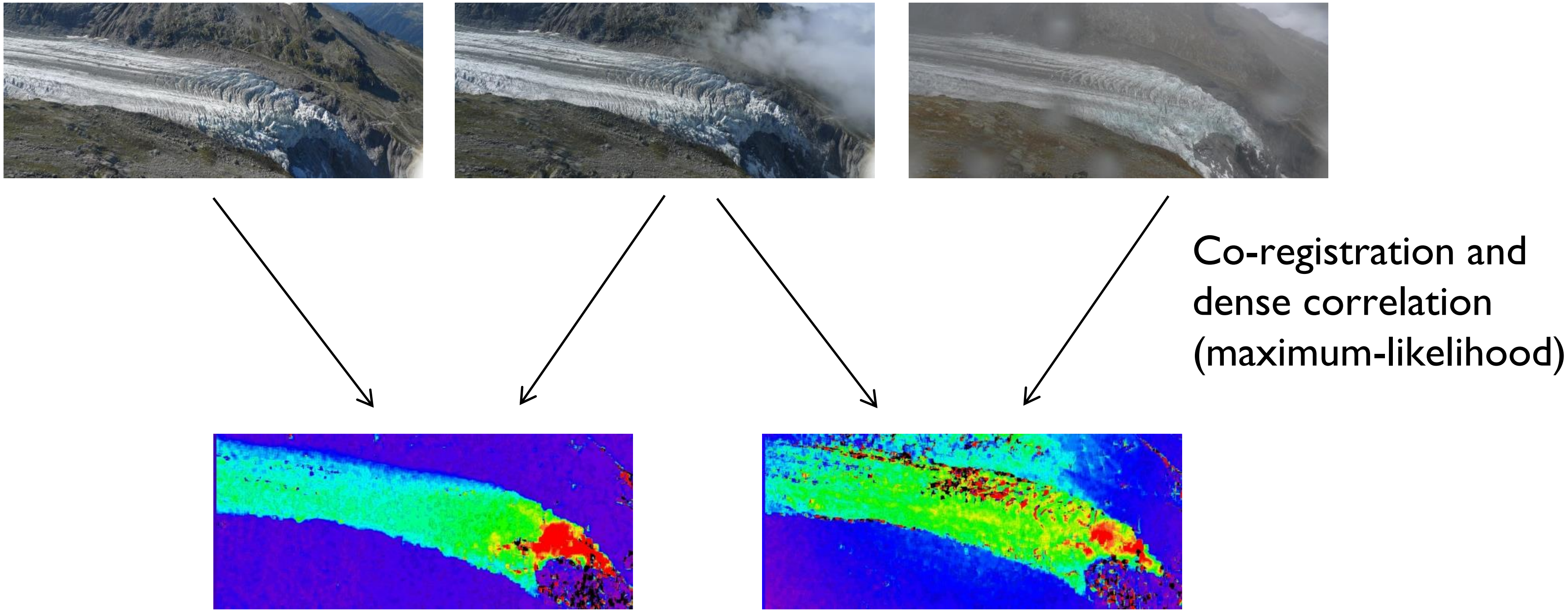
0 ⟵ **NMI** ⟶ 1

2 → 1 → 3

2 → 2 → 2 → 2 → 2 → 2

Monitoring of the Argentière glacier

- 21 images of the glacier automatically acquired every 2 days from September to November 2013.
- Optical camera Leica, DMC-LX, 10 mega-pixels.

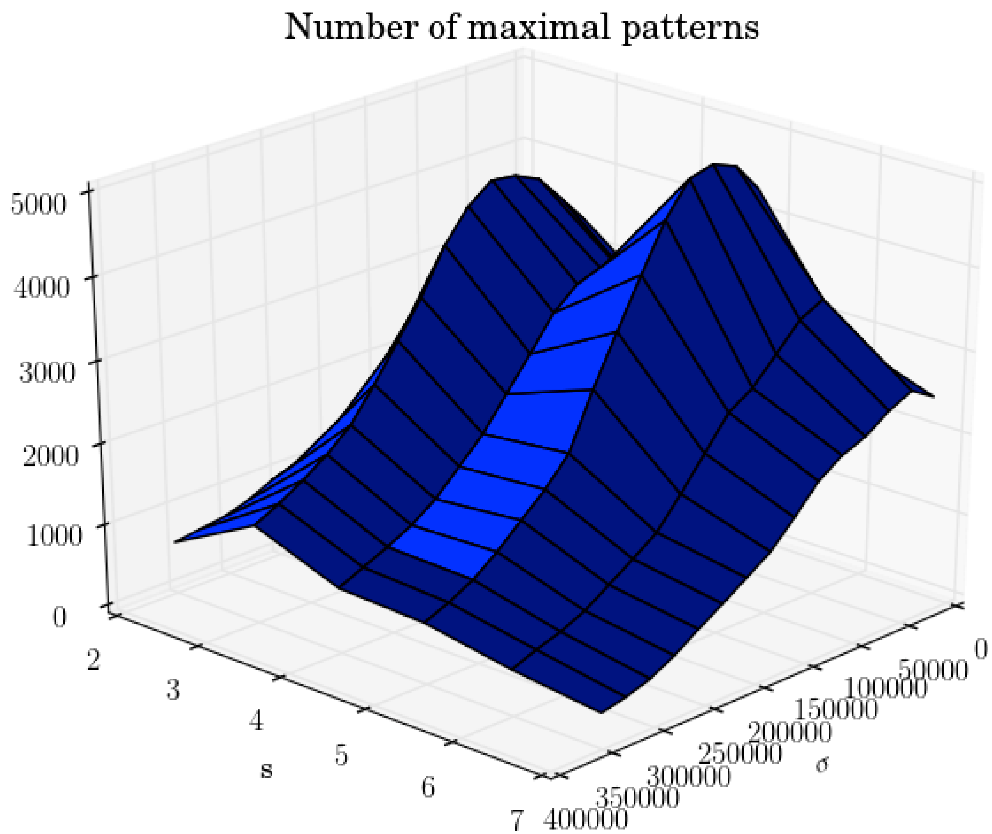


- 20 displacement fields in pixels (horizontal component). Size : 1675*700 pixels.

[Benoit15]

Setting the quantization and extraction parameters

- κ set to 5 (standard setting)
- σ ranging in [2%;20%]
- s ranging in [2;7]
- Multiple quantization and extraction in order to maximize the number of maximal GFS-Patterns.
- Maximum of 4900 maximal GFS-patterns for s=5 and σ=6% (70350 pixels)



GFS-pattern extraction

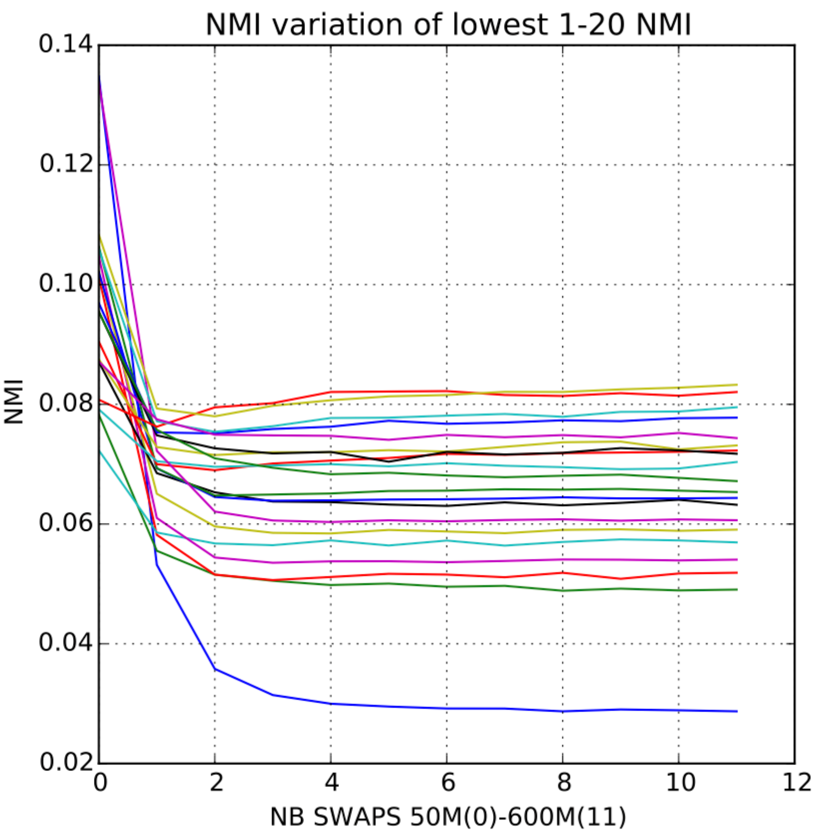
- prototype: C/Python
- platform : a single core on 2.9 GHz Intel Core i7

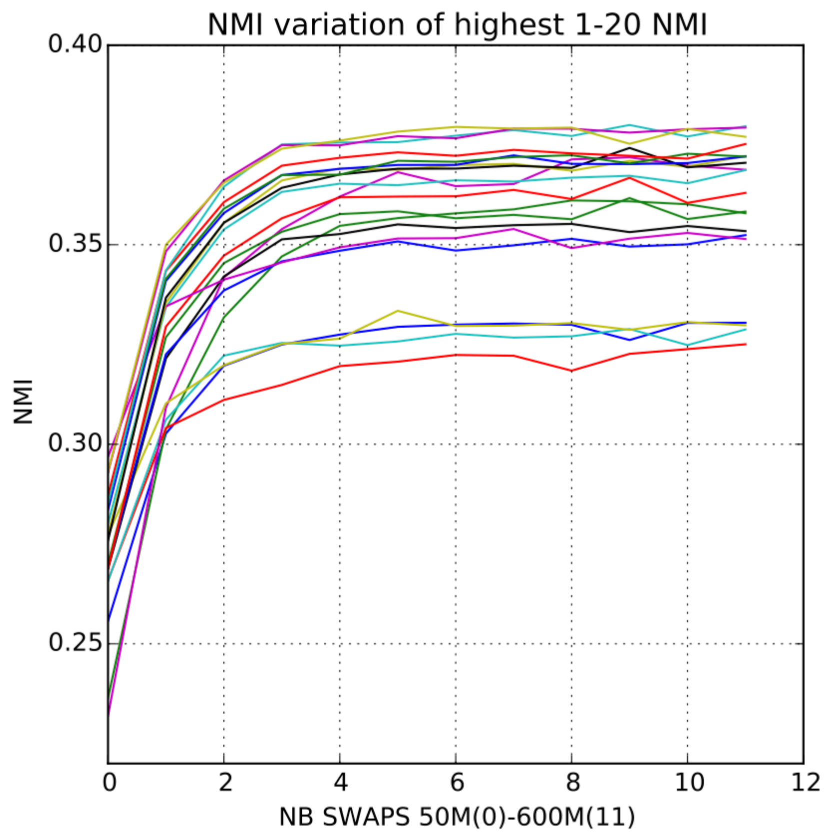
- Quantization with 5 symbols (1,2 : small negative displacements ; 3,4,5 small to large positive displacements)

- σ = 70350 (6%, set w.r.t. the maximum number of maximal patterns)
- k = 5
- Number of GFS-patterns: 19234
- Number of maximal GFS-patterns: 4900

- space/time requirements: 2,93 GB, 84 minutes.

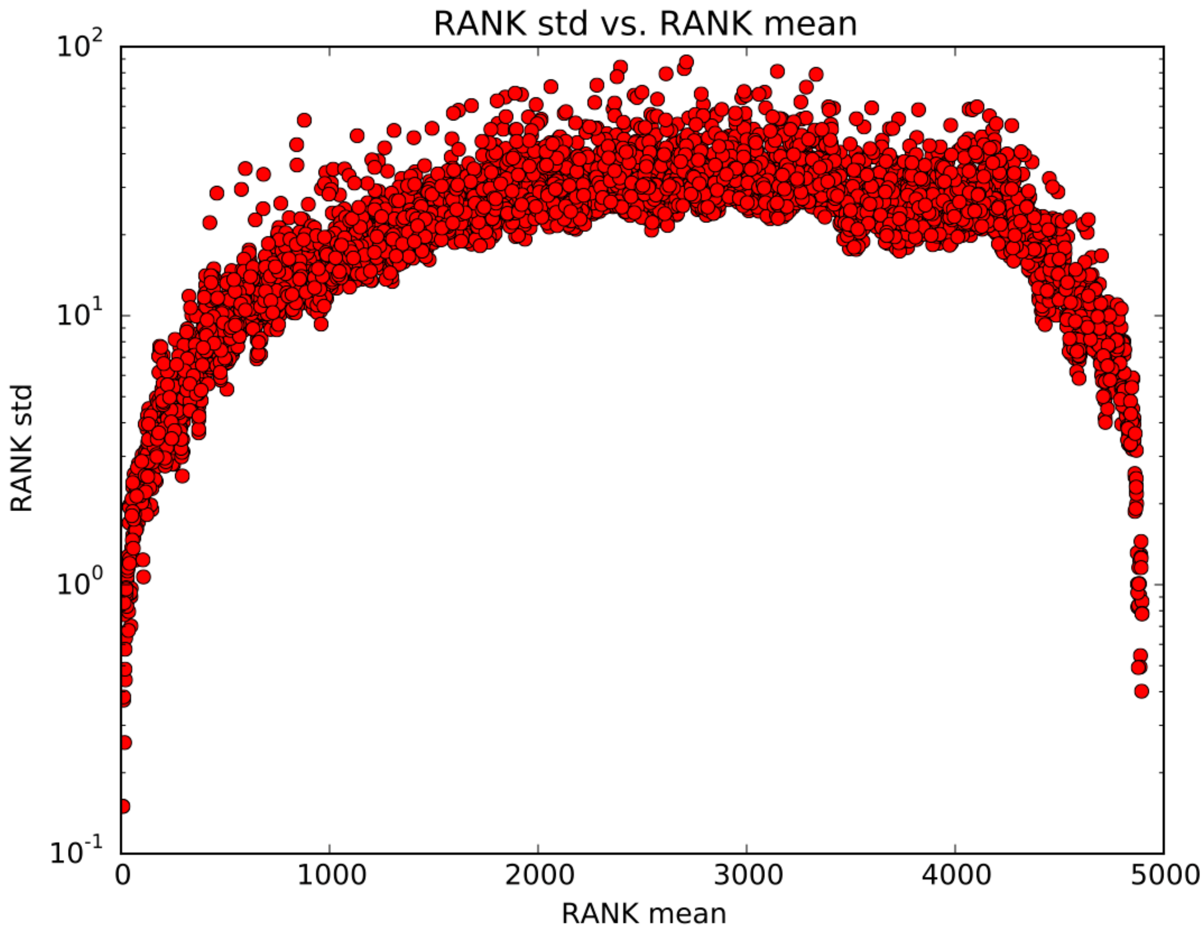
Setting the number of swap attempts





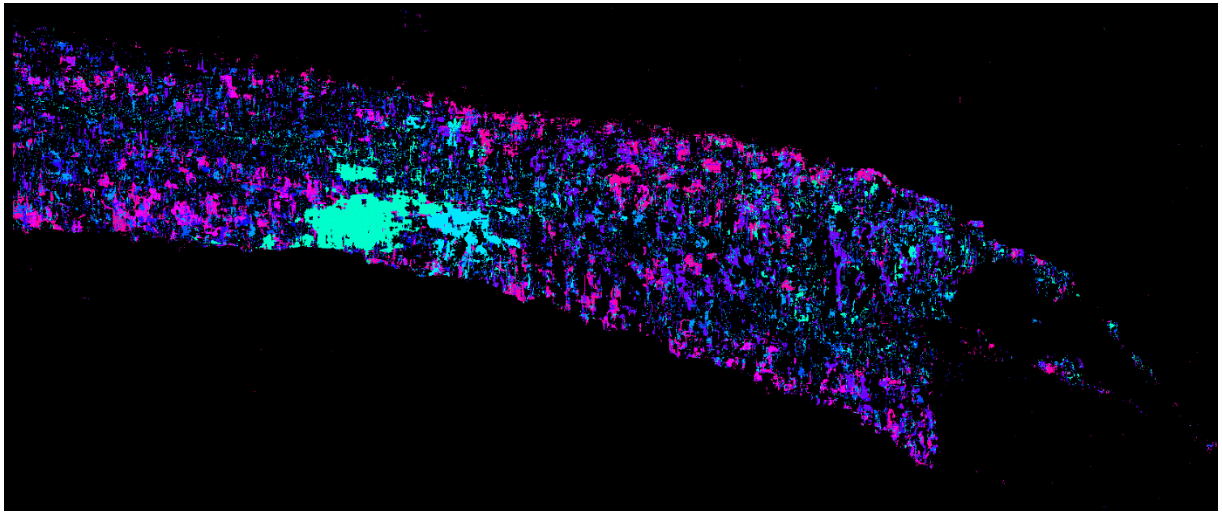
- Reference is 500M of swap attempts

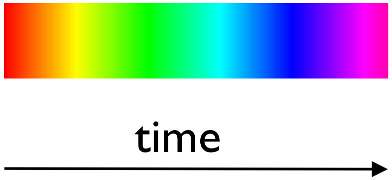
Ranking stability



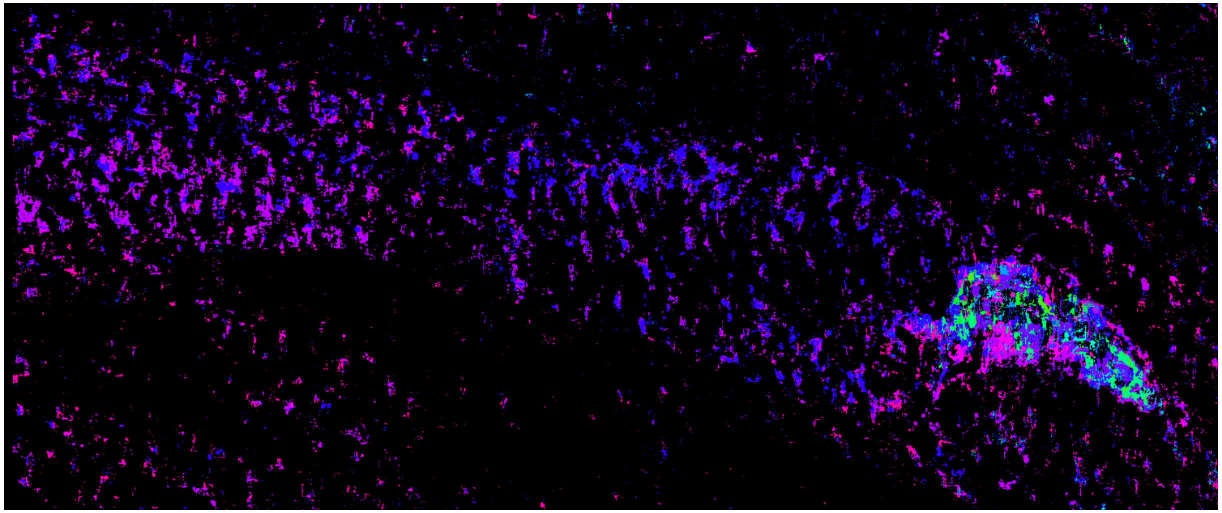
- The ranking experiment has been repeated 1000 times.

Qualitative results



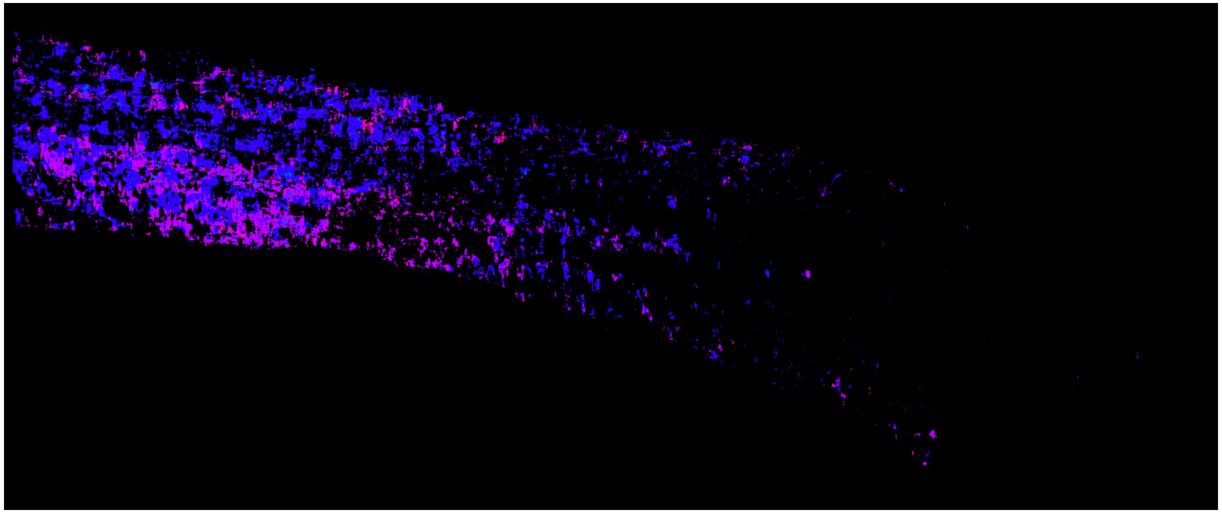


- 14th lowest NMI pattern
  5,4,4,4,4,4,4,2

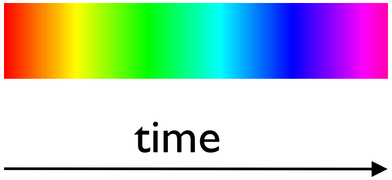


- 15th lowest NMI pattern
  5,5,5,5,5,5,4

Qualitative results



- 12th highest NMI pattern
  4,4,4,4,4,3,4,4,1,5,1,4



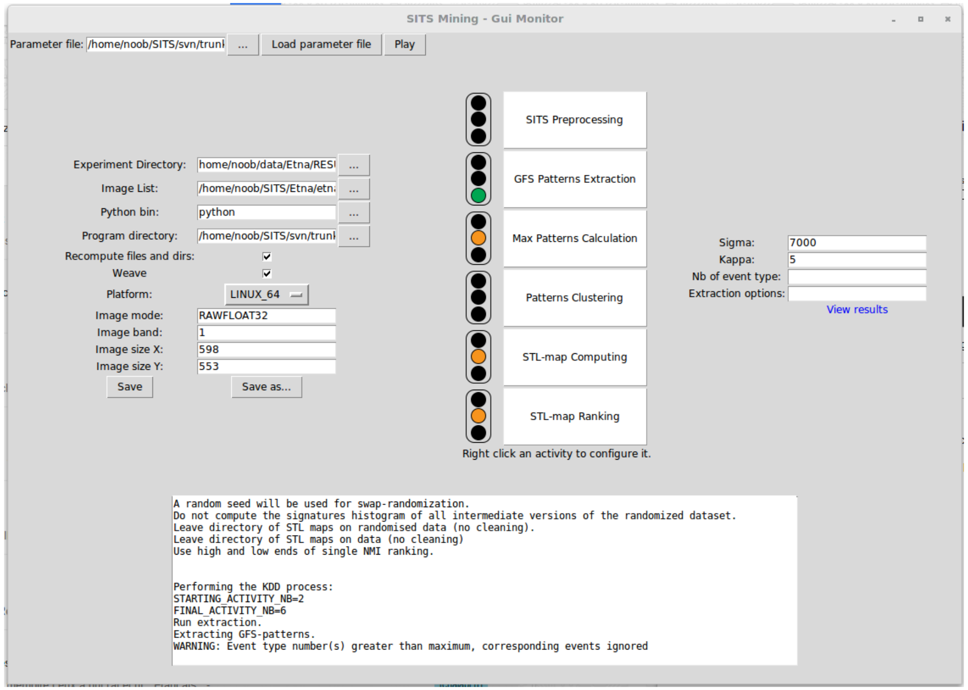
Conclusion & future work directions

- Encouraging qualitative results on a displacement field time series obtained by terrestrial photogrametry.
- Both ends of the NMI ranking can be of interest.
- Stability of both ends of the ranking for multiple randomizations.
- An operating point could be found to maximize the number of maximal GFS-patterns / get the richest description.

- Future work directions : raw optical images / considering the uncertainties during the extraction / longer time series.

SITS Miner

- Quantize SITS
- Extract GFS-patterns
- Select maximal ones
- Cluster patterns
- Compute STL-maps
- Rank STL-maps/patterns



- C/Python 2.7, free, open source, multiplatform (Linux, Mac OS, Windows)

Open source package soon available for download

References

[Agrawal95] Agrawal, R., Srikant, R.: Mining sequential patterns. In: Proceedings of the Eleventh International Conference on Data Engineering, Taipei, Taiwan. pp. 3– 14 (1995)

[Benoit15] Benoit L., Dehecq A., Pham H.-T., et al. Multi-method monitoring of glacier d'Argentière dynamics. In: Annals of Glaciology, vol. 56, no. 70, 2015.

[Gionis07] Gionis, A., Mannila, H., Mielikainen, T., Tsaparas, P.: Assessing data mining results via swap randomization. TKDD 1(3) (2007)

[Julea11] Julea A., Méger N., Bolon P., et al. Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns. In: IEEE Transactions on Geoscience and Remote Sensing, vol. 49, 4, 2011, pp. 1417–1430.

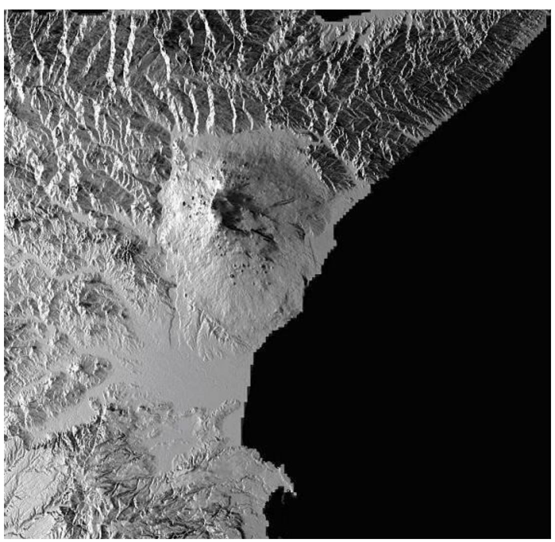
[Julea12] Julea A., Méger N., Rigotti C., et al. Efficient Spatiotemporal Mining of Satellite Image Time Series for Agricultural Monitoring. In: Transactions on Machine Learning and Data Mining, vol. 5, 1, 2012, pp. 23–45.

[Méger11] Méger N., Jolivet R., Lasserre C., et al. Spatio-temporal mining of ENVISAT SAR interferogram time series over the Haiyuan fault in China. In: Proc. of the 6th Int. Workshop on the Analysis of Multitemporal Remote Sensing Images (MULTI-TEMP'2011), July 2011, pp. 137–140.

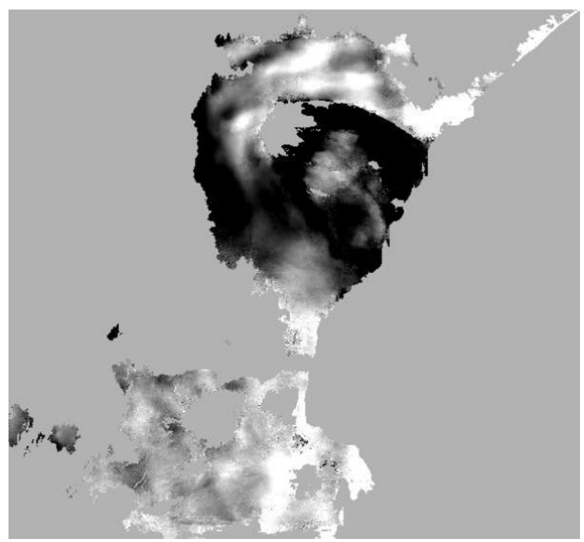
[Méger12] Méger N., Rigotti C., Gueguen L, et al. Normalized Mutual Information-Based Ranking of Spatio-Temporal Localization Maps. In: Proc. of 8th European Spatial Agency (ESA) - EUSC - JRC Conference on Image Information Mining, Oct. 2012, pp. 11–14.

Deformation monitoring of Mount Etna

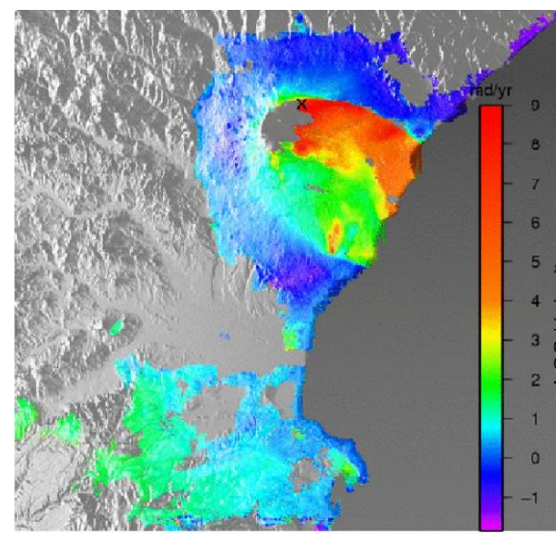
- 16 co-registered total phase delays images (553X598), 2003-2010, SAR geometry, ≈160 m, over the site of the Mount Etna volcano.
- Data produced by M-P. Doin, laboratory ISTerre.
- Phase delays were quantized into 3 levels (33rd et 66th percentiles).



DEM of the Mount Etna area



Phases delays 2003/01/22



Average velocity in rad/yr [Doin11]
2π = 2.8 cm

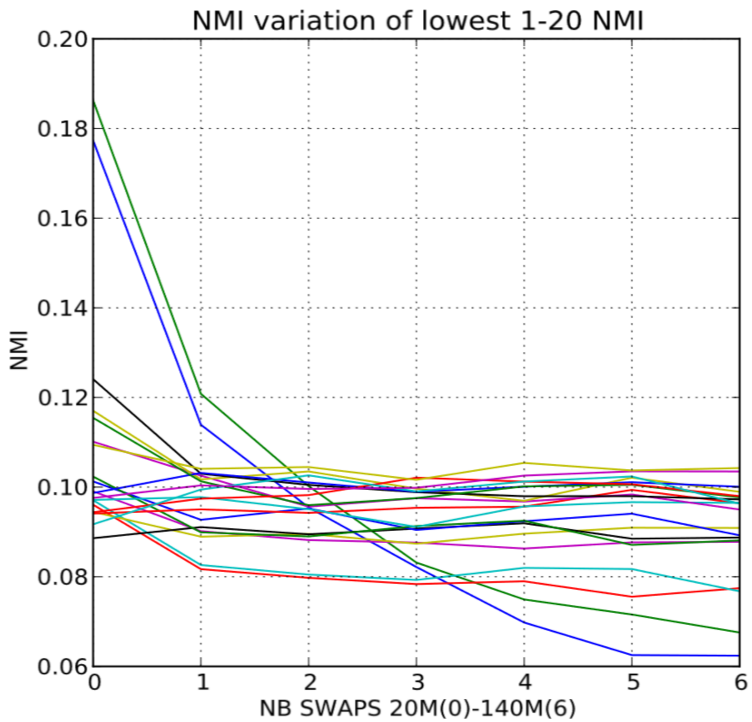
GFS-pattern extraction and ranking

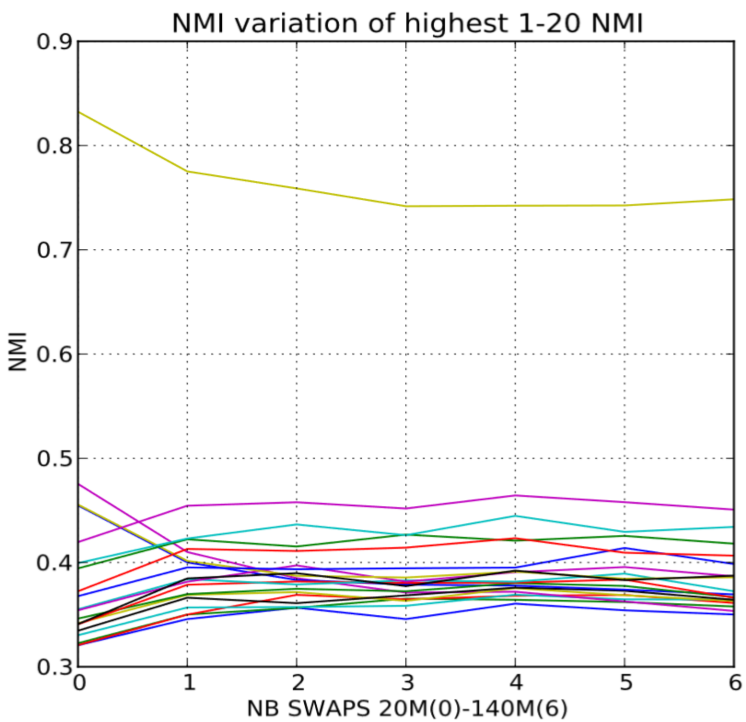
- prototype: C/Python
- platform : a single core on 2.7 GHz Intel Core i7

- σ = 7000 (2.1%, set w.r.t. the maximum number of maximal patterns)
- k = 5
- Number of GFS-patterns: 2658
- Number of maximal GFS-patterns: 508
- #swaps: 100M

- space/time requirements: 1.66 GB, 700 s.

#swaps: 100 M


NMI variation of lowest 1-20 NMI
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
NMI
0
1
2
3
4
5
6
NB SWAPS 20M(0)-140M(6)


NMI variation of highest 1-20 NMI
0.9
0.8
0.7
0.6
0.5
0.4
0.3
NMI
0
1
2
3
4
5
6
NB SWAPS 20M(0)-140M(6)

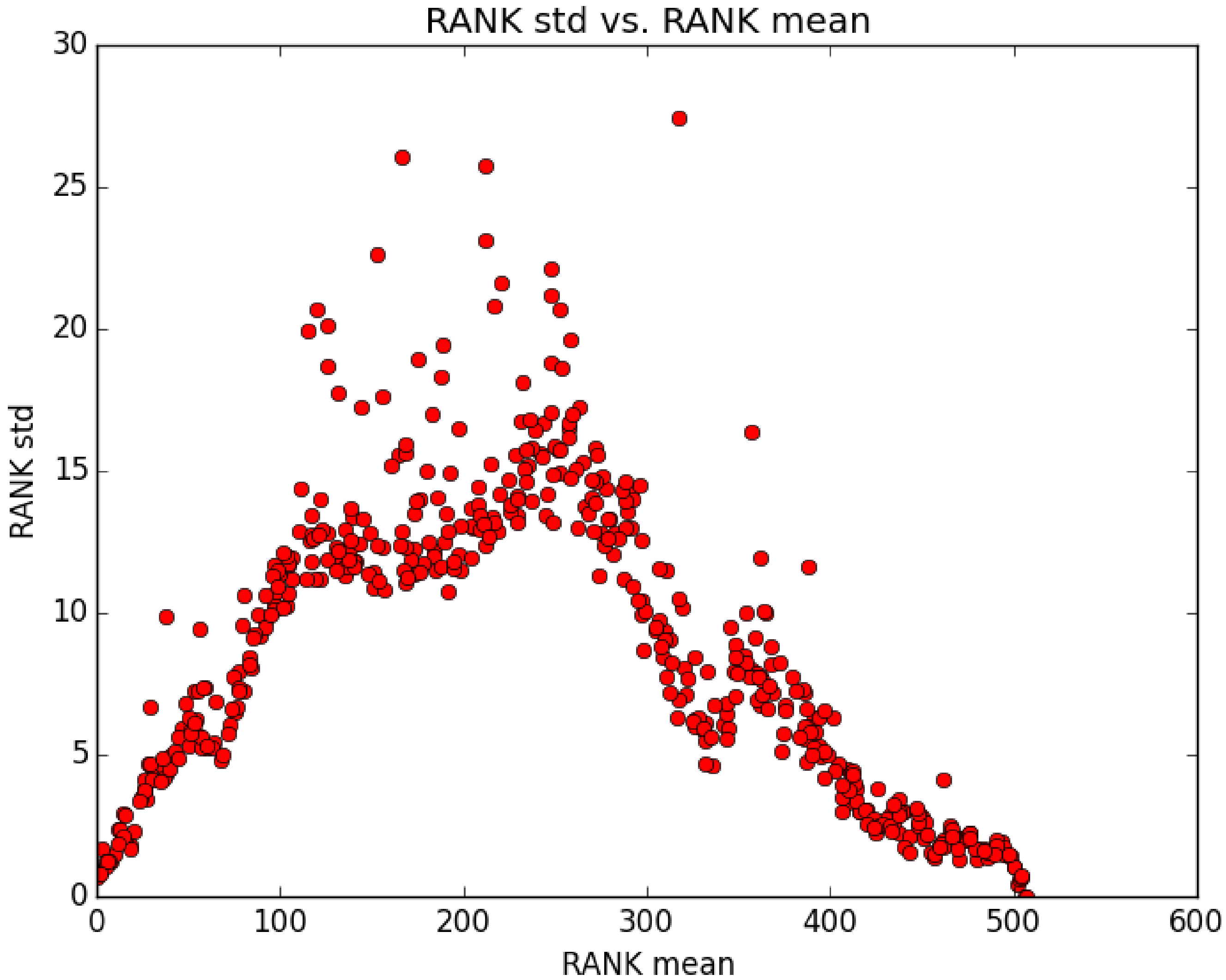
Randomization procedure behavior

73,9% of the dataset can not be swapped

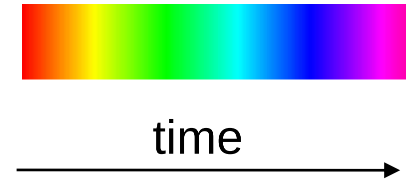
- 1000 swap randomizations
- Average swapped events rate: 6,5% (standard deviation ≈ 0, stable)

- 1 swap randomization
  - 1,070,219 matrices are explored.
  - All matrices are reached only once (except one, 8 times)

Ranking stability (over 1000 matrices)

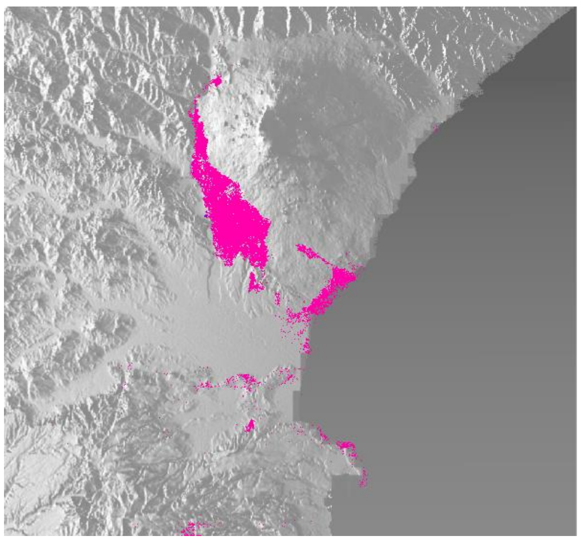


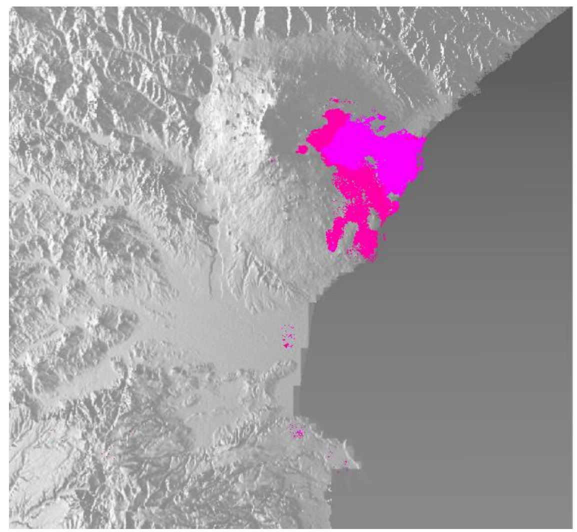
Qualitative results

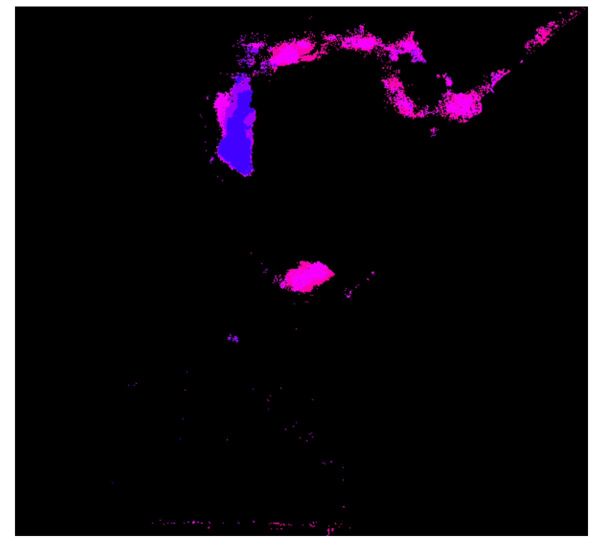


1→1 → 2 → 1 → 1 → 1 → 1 → 3

1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 1







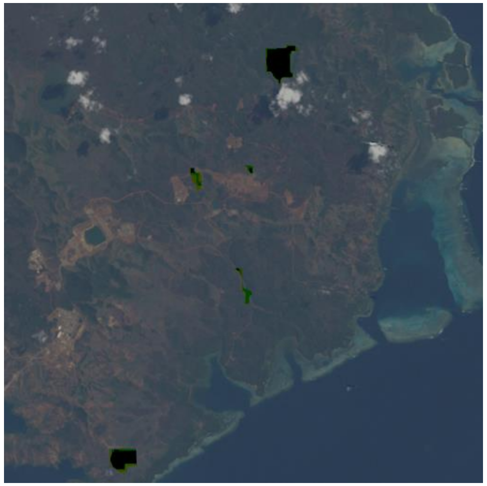
1 → 2 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3

Noisy/cloudy optical data can also be analysed

Monitoring the Yaté area

- 16 co-registered Landsat 7 images (513X513), 2000-2011, 30 m, over New Caledonia (Yaté rural district, nickel open cast mining activities, scrub fires, erosion, landsildes, UNESCO protected coral reefs and lagoons).
- Bands: blue (450-520nm), green (520-600nm), red (630-690nm), near infra-red (750-900nm).
- Synthetic band NDVI and ground truth provided by Bluecham SAS. NDVI values were quanitized into 3 levels (33rd et 66th percentiles).

Clouds ...

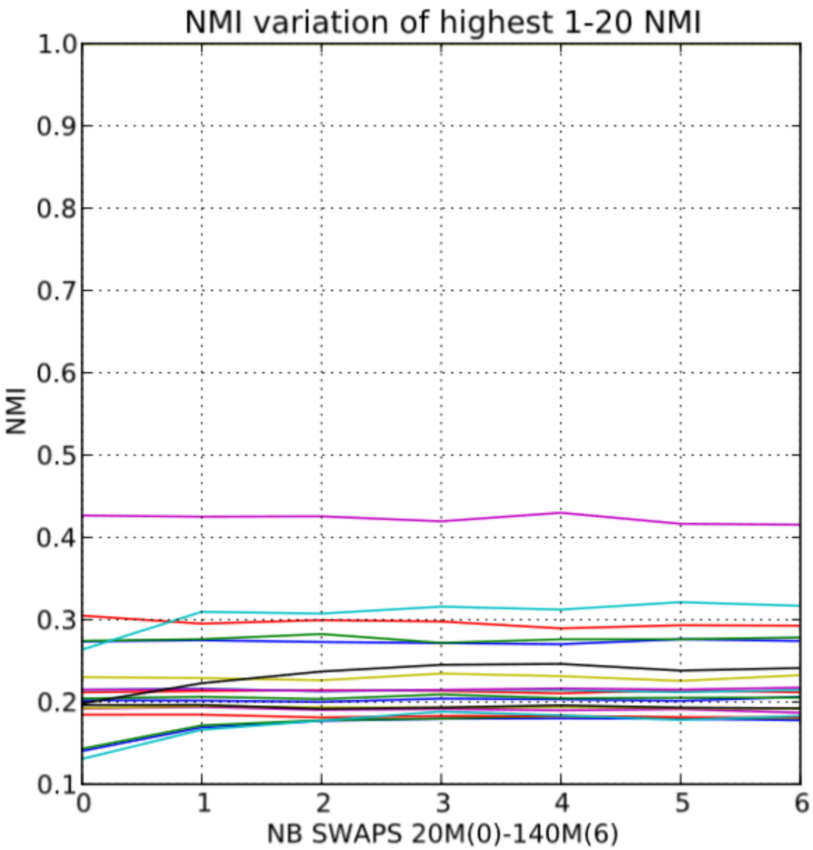


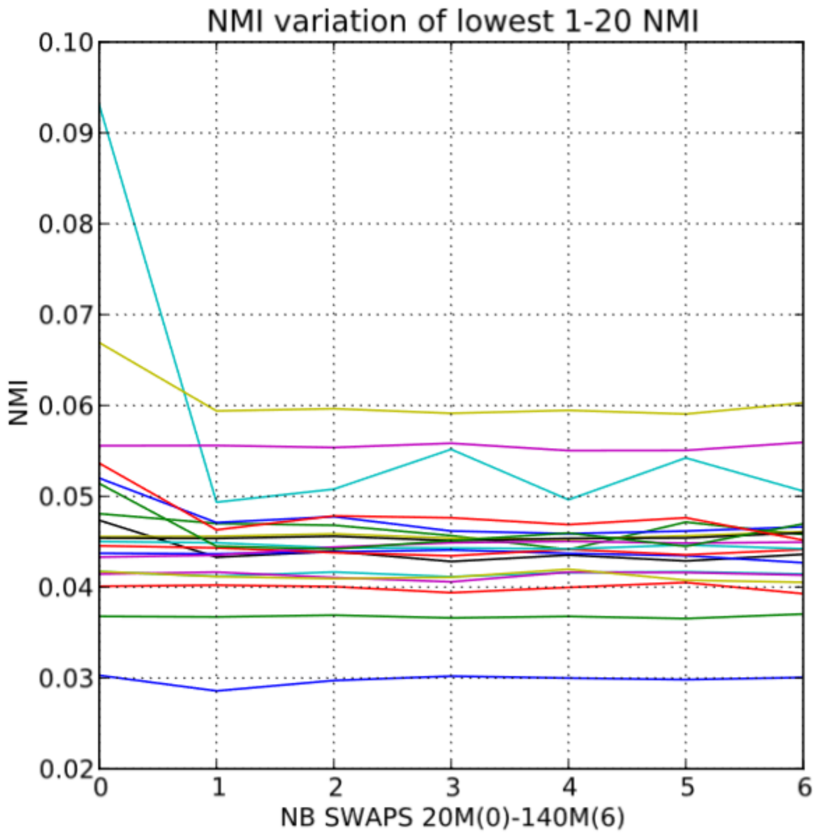
… artefacts

GFS-pattern extraction and ranking

- σ = 7000 (2.6%, set w.r.t. the maximum number of maximal patterns)
- k = 5
- Number of GFS-patterns: 15620
- Number of maximal GFS-patterns: 295
- NMI rankings at 100M swaps

#swaps: 100 M


NMI variation of highest 1-20 NMI
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
NMI
0
1
2
3
4
5
6
NB SWAPS 20M(0)-140M(6)


NMI variation of lowest 1-20 NMI
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
NMI
0
1
2
3
4
5
6
NB SWAPS 20M(0)-140M(6)

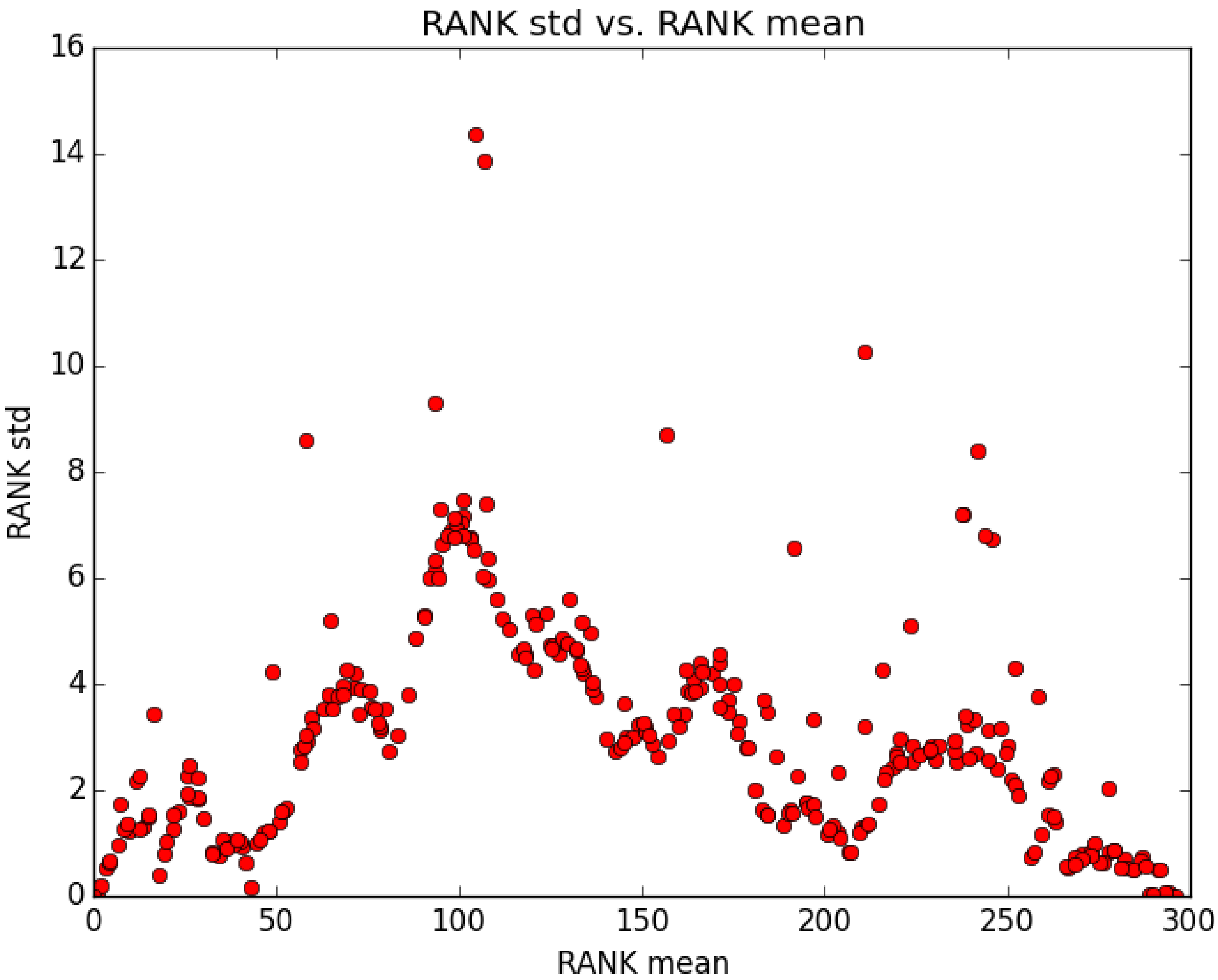
Randomization procedure behavior

16,2% of the dataset can not be swapped

- 1000 swap randomizations
- Average swapped events rate: 32,9% (standard deviation ≈ 0, stable)

- 1 swap randomization
  - 8,911,591 matrices are explored.
  - All matrices are reached only once (except one, 4189 times and another one, 44 times)

Ranking stability (over 1000 matrices)



Yaté qualitative results

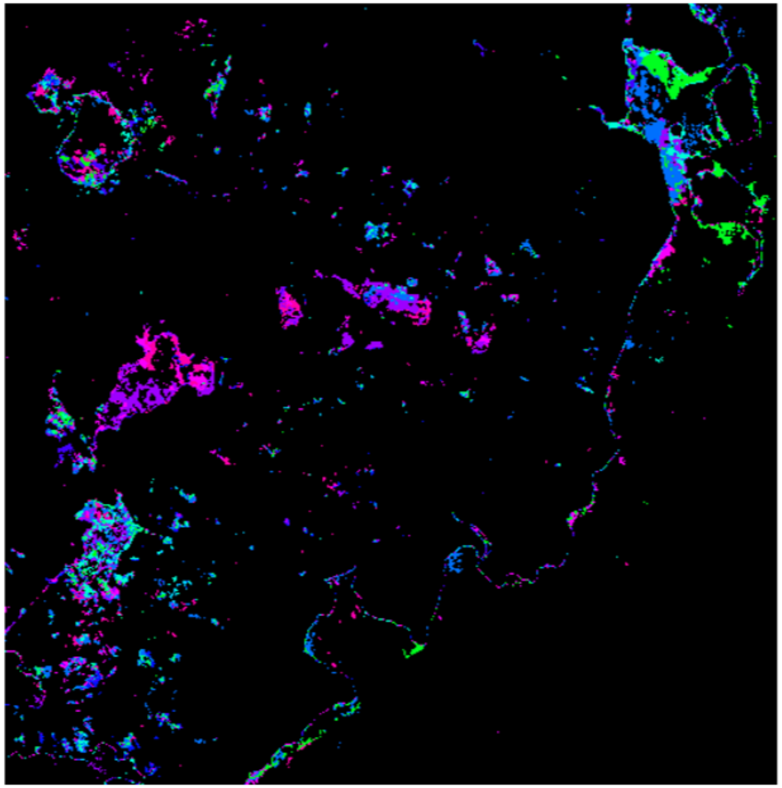


Fig. 13. STL-map: 6^{th} lowest NMI pattern $\langle 2,2,1,1,1,2 \rangle$, NC.

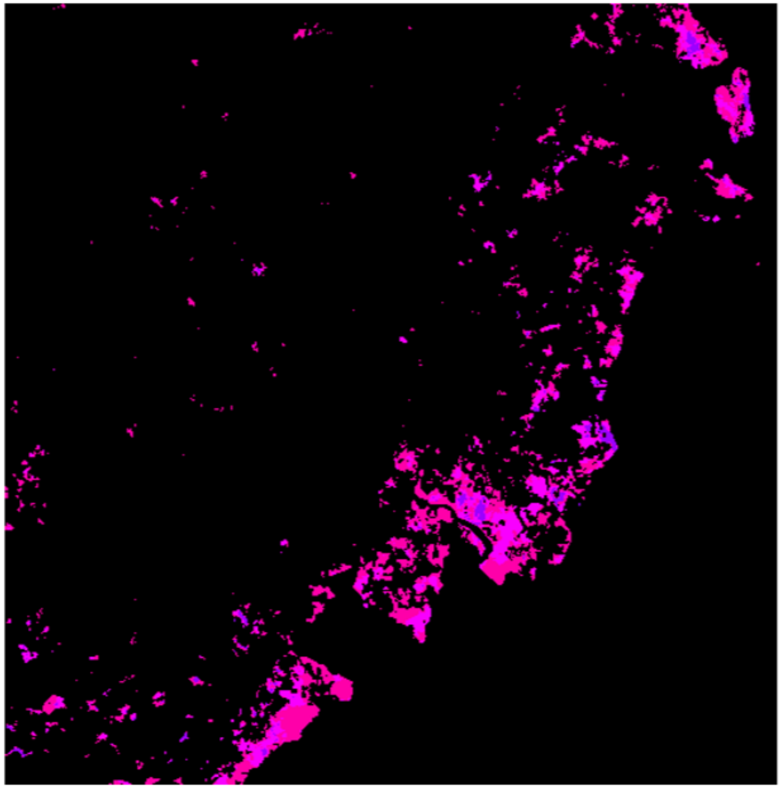
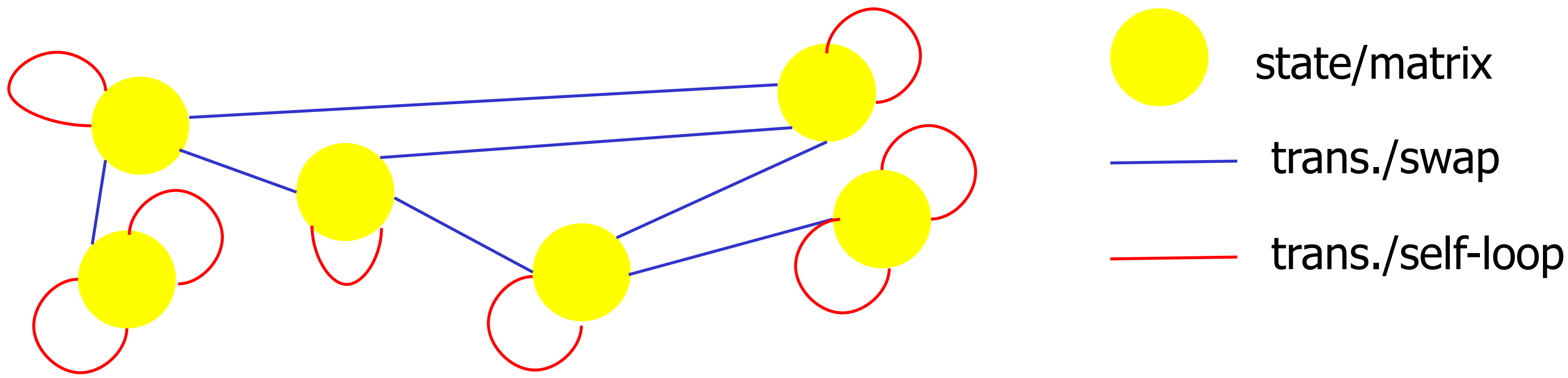


Fig. 14. STL-map: 2^{nd} highest NMI pattern $\langle 3,3,3,3,3,3,3,3,3,3,3,3,3,3 \rangle$, NC.

Equiprobable matrices and self-loops



- All matrices having the same structure are equiprobable if failed swap attempts are counted as self-loops.

  [Gionis07]

- A swap attempt = a step in Markov chain M(S,T)
  S – set of states/matrices, T – set of transitions/swap attempts

- Each state degree = P = |T|

- All states have the same degrees ➔ uniform distribution

Why maintaining column and row margins?

Maintaining margins ≈ preserving histograms

Colum margins: within a single image, the nature of the observed scene must not be modified.

Glaciers and forests should not be transformed into bare soils.

Row margins: the nature of a pixel evolution should be preserved.

Variations between snow and rocks should not be transformed into permanent vegetation.