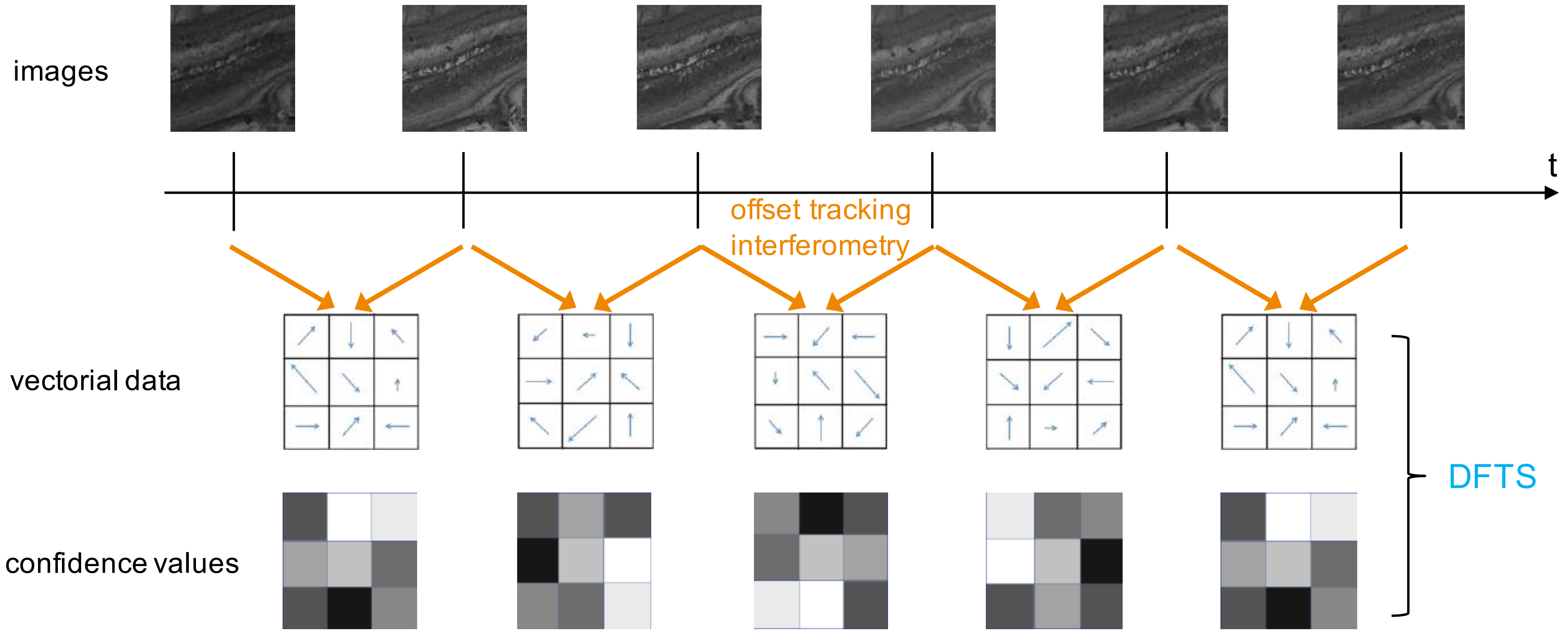
Mining displacement field time series with DFTS-P2miner

MDIS-2019, Tuesday 15 October 2019, Strasbourg

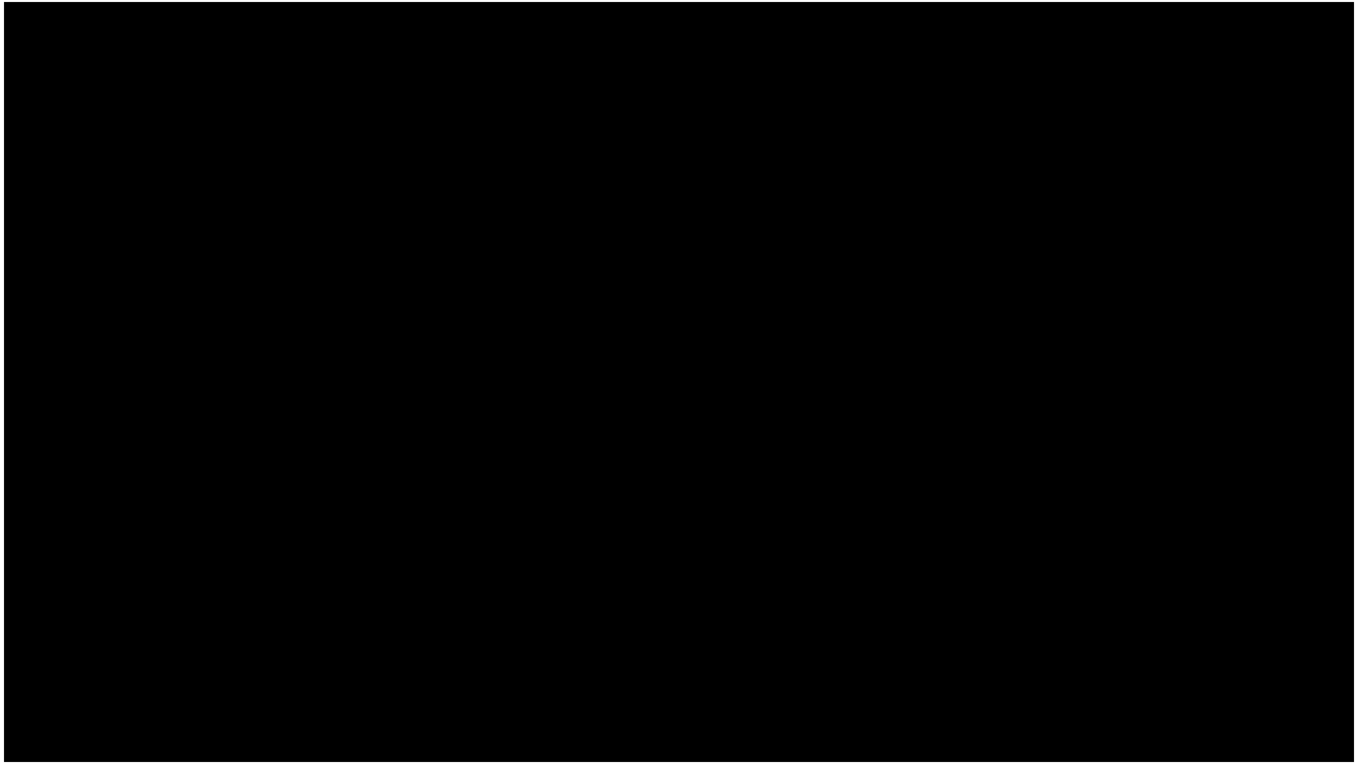
Nicolas Méger, Christophe Rigotti, Catherine Pothier,
Emmanuel Trouvé, Tuan Nguyen

Partial funding for this project was provided by a grant from la Région Auvergne-Rhône-Alpes (Tuan Nguyen's grant) and an ANR grant (PHOENIX ANR-15-CE23-0012).

Displacement Field Time Series - DFTS



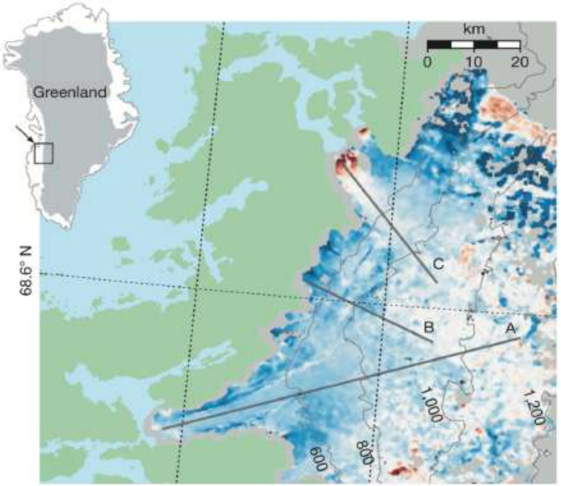
DFTS are complex datasets



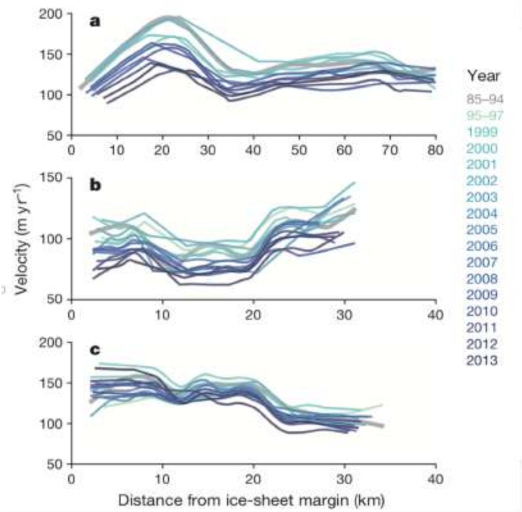
DFTS analysis: standard approach

Raucoules et al. 2013; Tedstone et al. 2015; Altena et al. 2018

- Low confidence data points are filtered out (if any)
- Spatiotemporal simplification by information selection & aggregation



velocity evolution profiles along transects
Tedstone et al. 2015



➔ Hypothesis testing, expert-oriented/biased, information loss.

DFTS analysis: what about knowledge discovery?

confirm infirm enrich

☺ ☹ ☺

users' knowledge

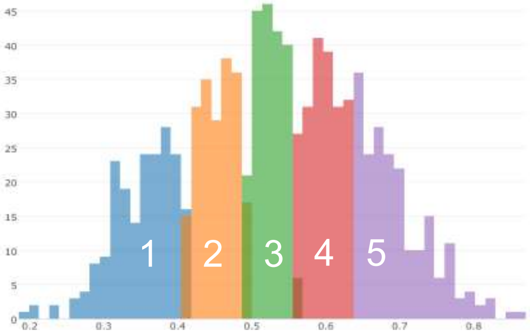
hypothesis testing ➔ hypothesis formation

DFTS analysis: data mining

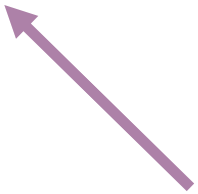
- Pattern discovery in large databases using artificial intelligence, computer science and statistics
- Mature field: itemsets, association rules – Agrawal et al. 1993, sequential patterns - Agrawal et al. 1995, episodes - Mannila et al. 1997
- Method: Reliable Grouped Frequent Sequential pattern (RGFS-pattern) extraction
- Guidelines:
  - basic preprocessing (direction and/or magnitude quantization, confidence values left unchanged)
  - unsupervised (no prior object/evolution identification)
  - easy-to-read models/patterns
  - noise-tolerant (atmospheric perturbations, sensor defects)
  - green IT (as much as possible ...)

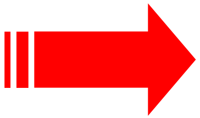
RGFS-patterns: preprocessing example

direction: equal interval bucketting



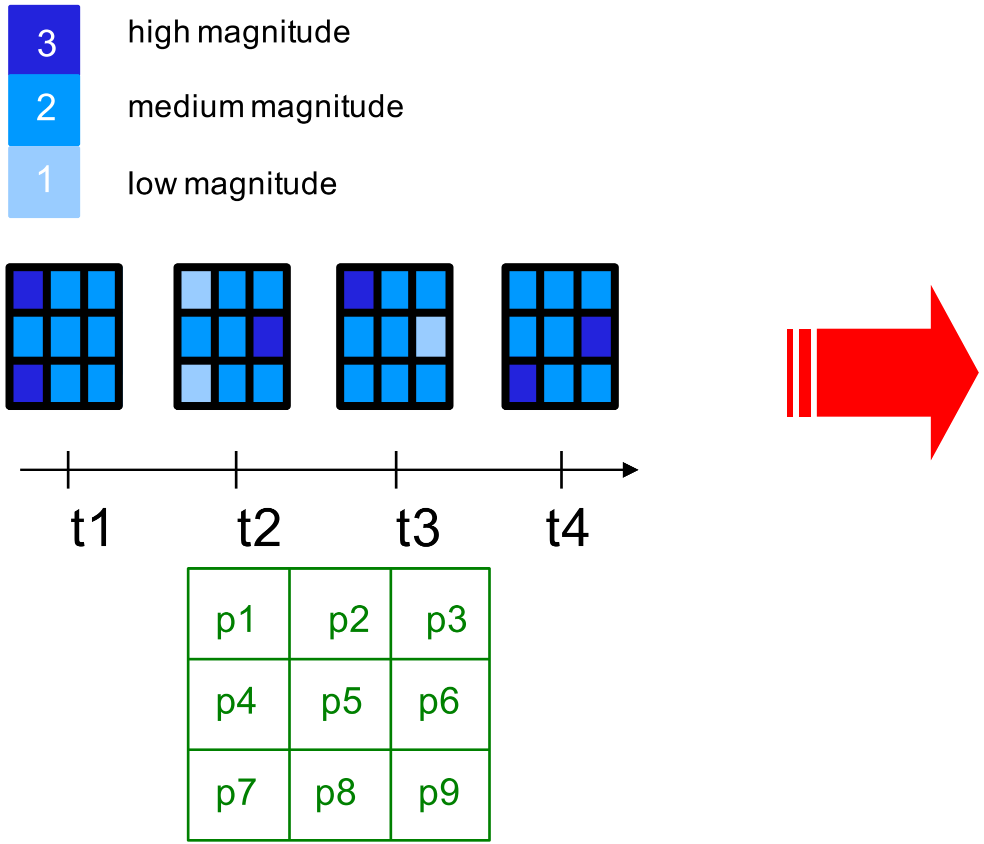
magnitude: equal frequency bucketting

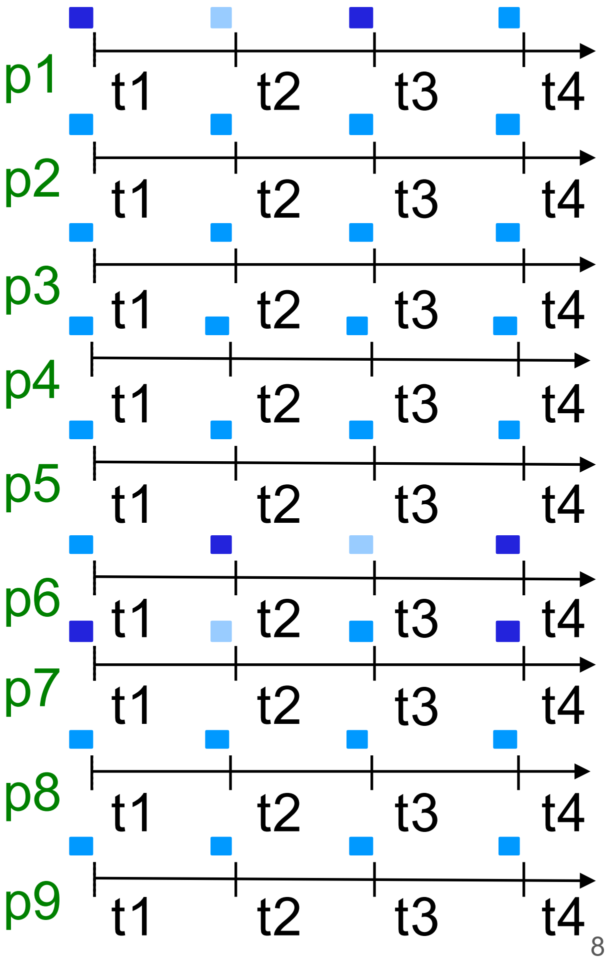




direction ➔ 11
magnitude ➔ 5

RGFS-patterns: base of sequences





RGFS-patterns: sequential patterns

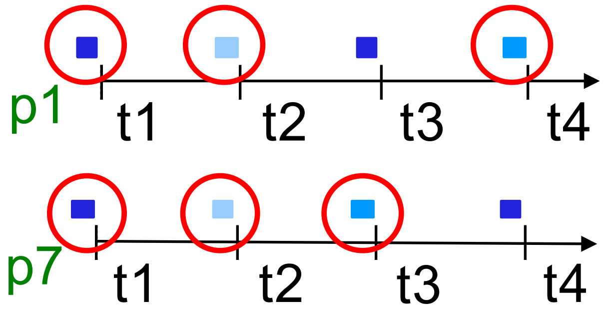
3 → 1 → 2

- easy-to-interpret
- all patterns and occurrences
- time shifts and gaps allowed (not substrings)

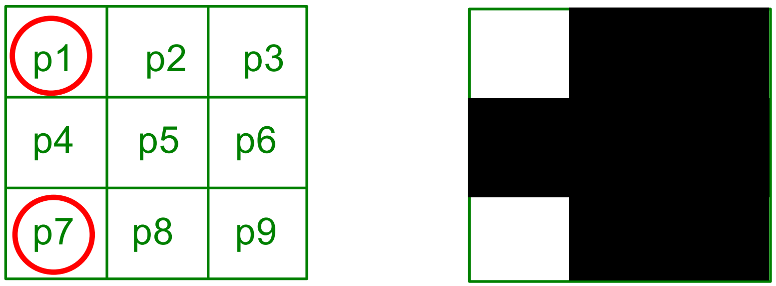
➔ noise-tolerant and no synchronization

RGFS-patterns: frequent sequential patterns

- Pattern support: |sequences in which it occurs| = |pixels covered by the pattern|
- A pattern is frequent if its support ≥ σ, the minimum support (or surface)
- Ex.: if σ=2, 3 → 1 → 2 is frequent

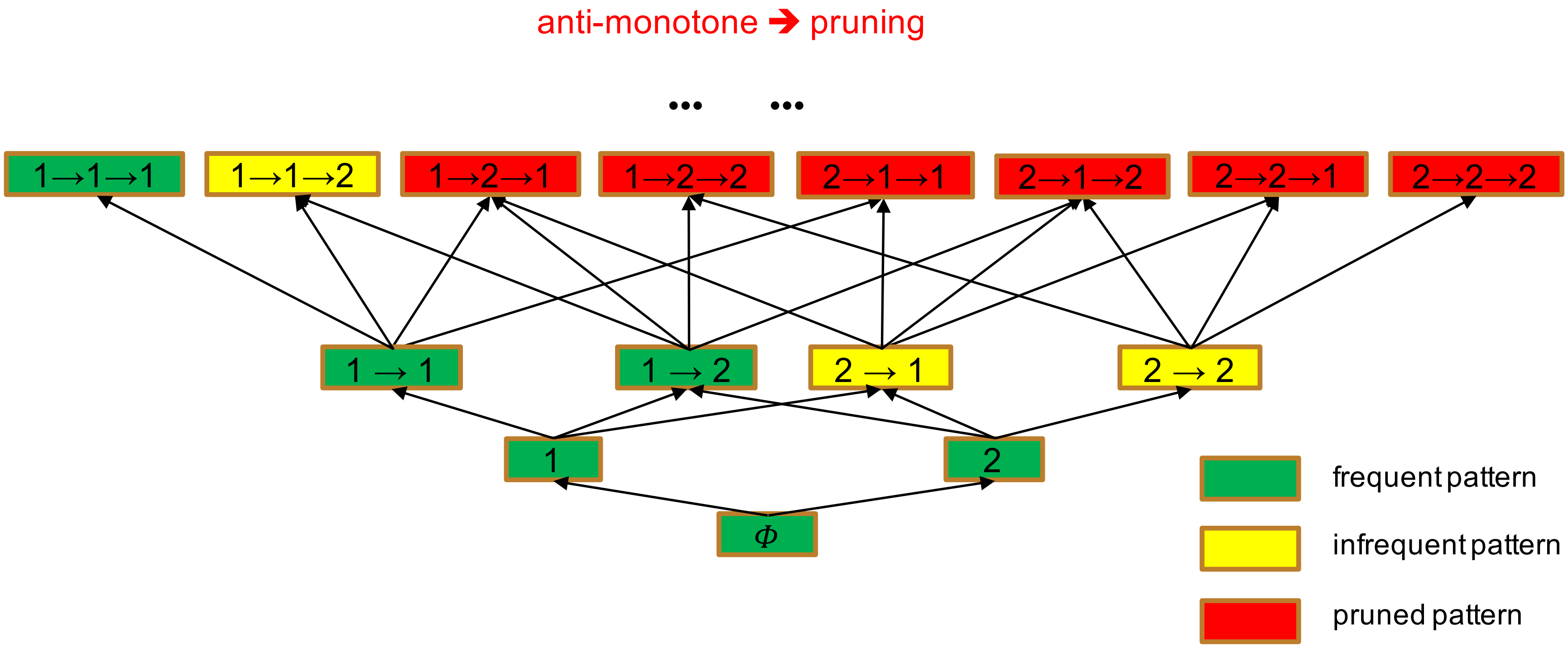


occurrence temporal localization

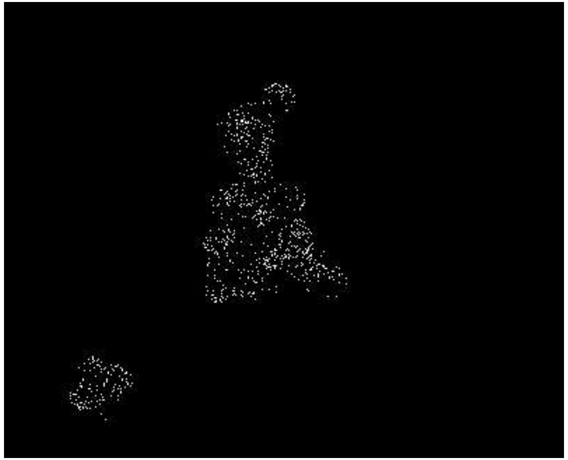


occurrence spatial localization

RGFS-patterns: frequency (or surface) constraint

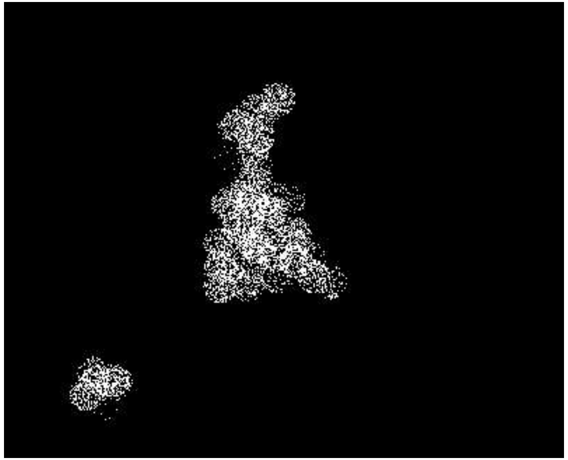


RGFS-patterns: towards spatiality



1 → 3 → 2
support < σ

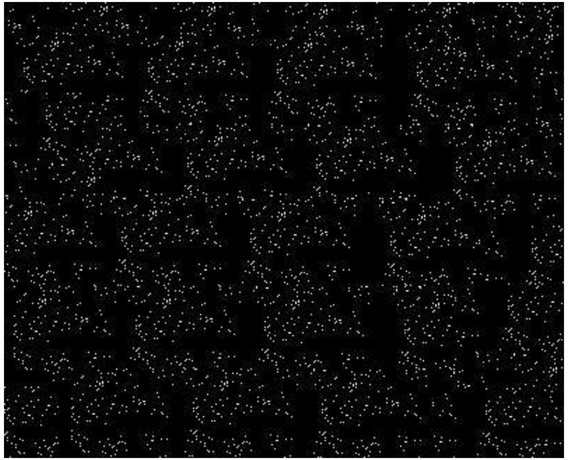
1 → 3
support ≥ σ



3 → 1 → 2
support ≥ σ

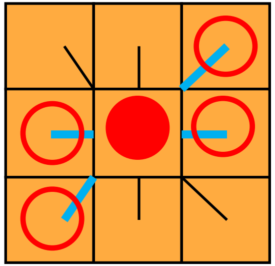
only noise

…



RGFS-patterns: Grouped Frequent Sequential patterns – GFS-patterns

- Pattern Average Connectivity (AC): average number of the pixels covered by a given pattern in the 8-neighborhood of its occurrences.



For a pattern α:

Links(α) = sum for all pixels covered by α of the number of their neighbors that are covered by α

AC(α) = links(α)/support(α)

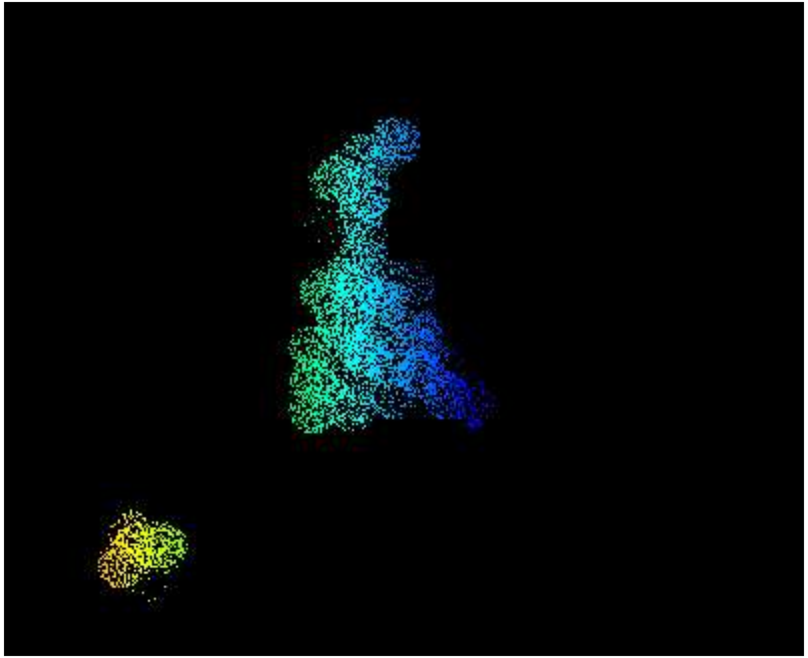
- A frequent sequential pattern is grouped if its AC ≥ κ, the minimum AC.
- This grouping constraint is not anti-monotone but …

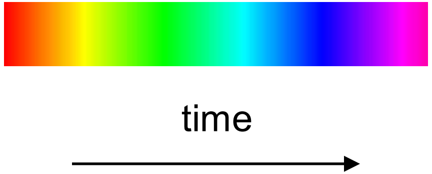
RGFS-patterns: partial pushing of the grouping constraint

- $AC(\alpha) \leq links(\alpha)/\sigma$ (upper bound)
- $links(\alpha)/\sigma \geq \kappa$ is anti-monotone
- Partial pushing
  - pruning using $links(\alpha)/\sigma \geq \kappa$
  - selection of the pattern such that $AC(\alpha) \geq \kappa$
- Complementary to support pruning (up to 2x faster)

RGFS-patterns: SpatioTemporal Localization maps - STL-maps

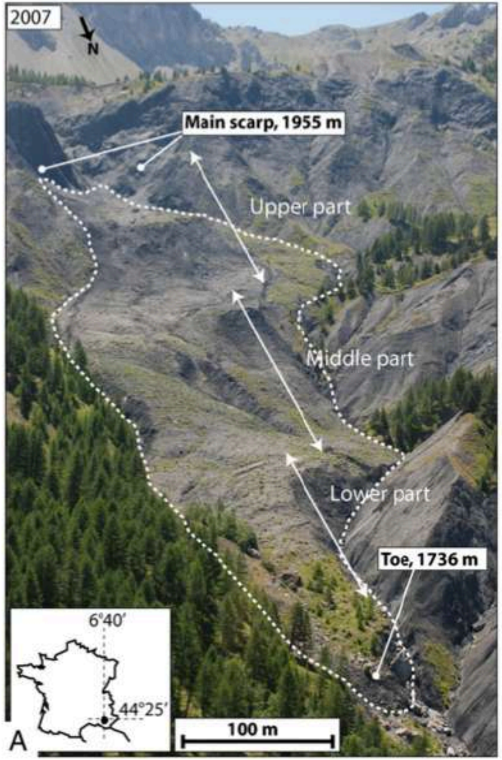
1 → 3
support ≥ σ
AC ≥ κ





A first example: the Super-Sauze landslide

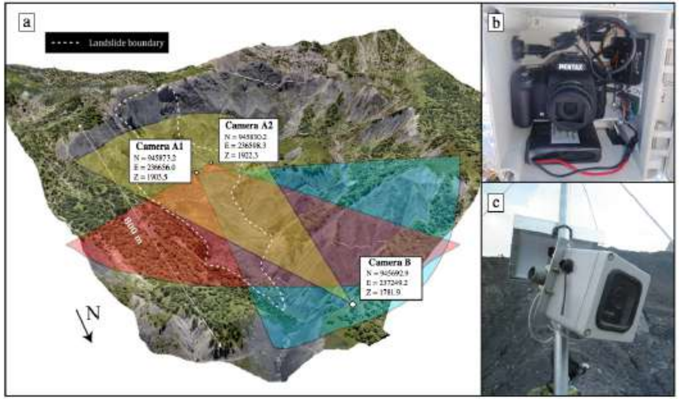
- Triggered during the 60's
- Filling the talweg of the Sauze torrent progressively
- 20 cm ≥ velocities ≥ 5cm a day
- Some surges measured at several meters a day
- About 560.000 m^3 of moving materials



Travelletti et Malet 2012

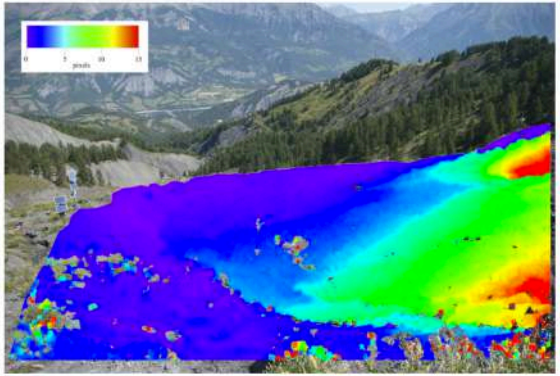
A first example: the Super-Sauze time-lapse

- Collab. IPGS (J.-P. Malet)
- Camera: Pentax K200D
- Resolution: 3872 x 2592
- Sensor size: 23.5 x 15.7 mm
- Focal distance: 25.68 m
- Period: 07/09/2011 – 08/23/2011
- Frequency: 1 image/day
- Number of images: 40

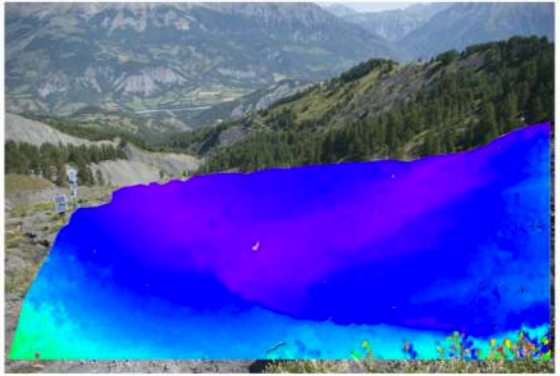


Mining the magnitudes

input DFTS: 37 fields of size 1936 x 880 obtained by offset tracking (EFIDIR Tools)

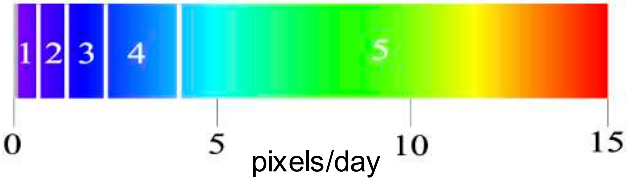


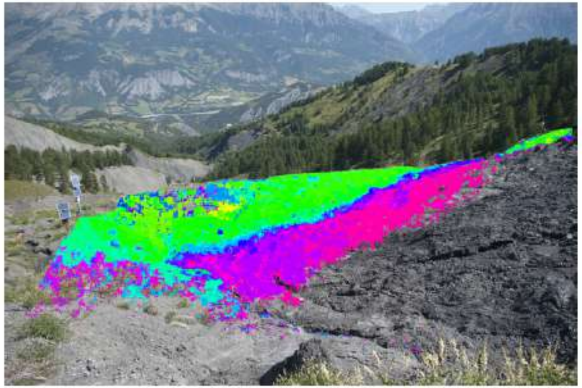
(a) magnitudes, 19-20 July 2011



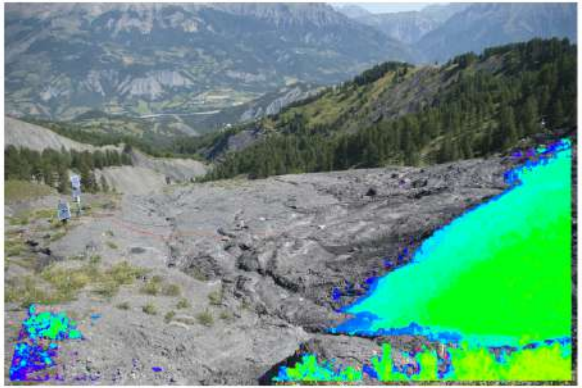
(b) magnitudes, 2-3 August 2011

- Parameters set to get as many patterns as possible
- nb symbols = 5 symbols (equal frequency bucketting)
- σ = 170367 pixels (10%)
- K = 7
- maximum time span = 10 days

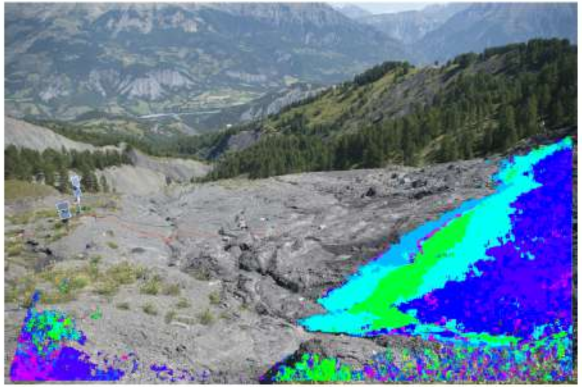




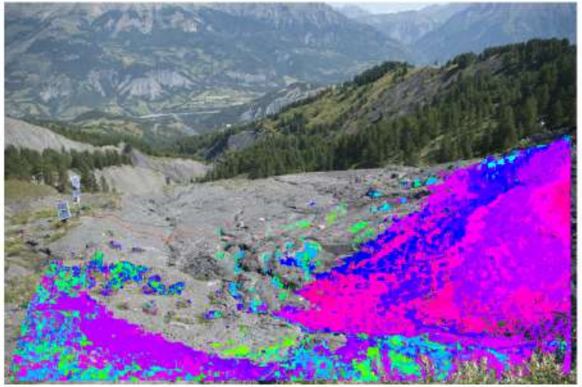
(a) 1,1,1,1,2,1,1,1



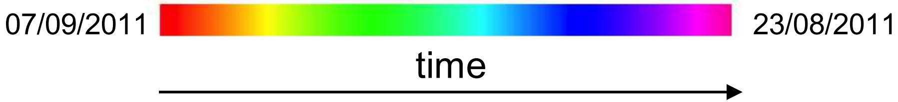
(b) 5,5,5,5,5,5,5,5,5,5



(c) 5,4,5,5,5,5,5,3



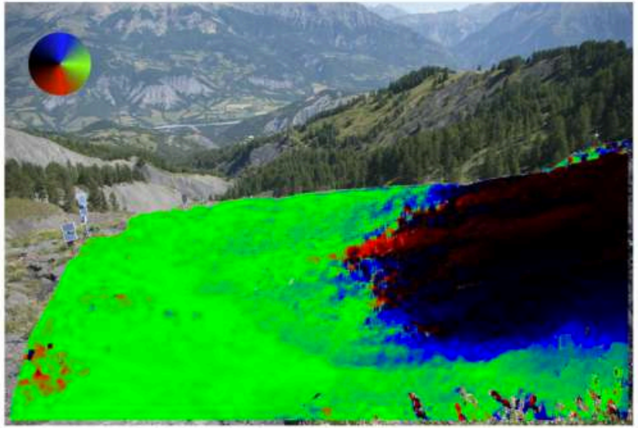
(d) 5,5,4,3,3,2



Mining the directions

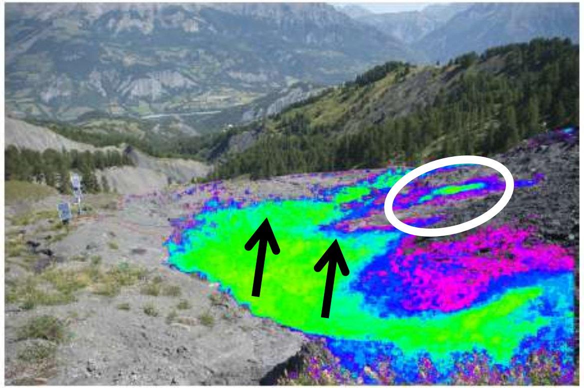
input DFTS: 37 fields of size 1936 x 880 obtained by offset tracking (EFIDIR Tools)

(a) directions, 2-3 August 2011

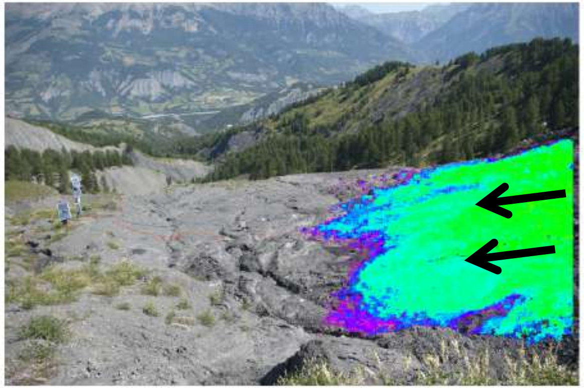


(b) directions, 17-18 August 2011

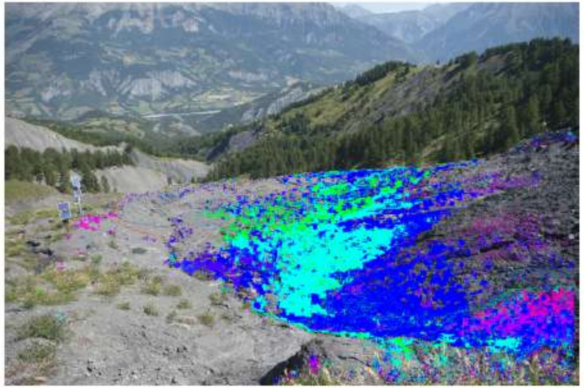
- Parameters set to get as many patterns as possible
- nb symbols = 5 symbols (equal frequency bucketting)
- σ = 200000 pixels (11.7%)
- K = 7
- maximum time span = 10 days



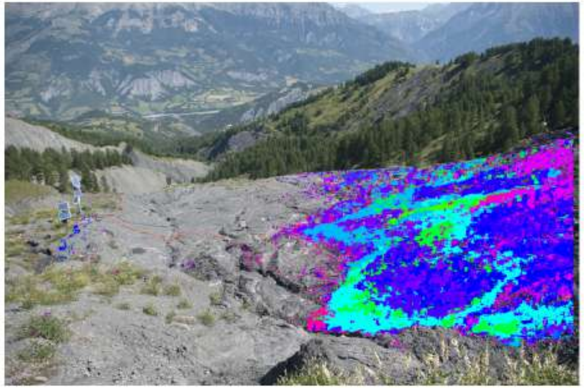
(a) 1,1,1,1,1,1,1



(b) 4,5,5,5,5,5,5,5,5,5

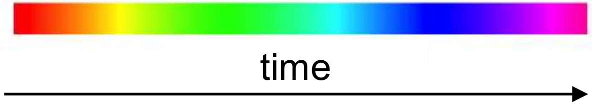


(c) 1,5,1,1,1,3



(d) 5,5,5,4,1,5,4

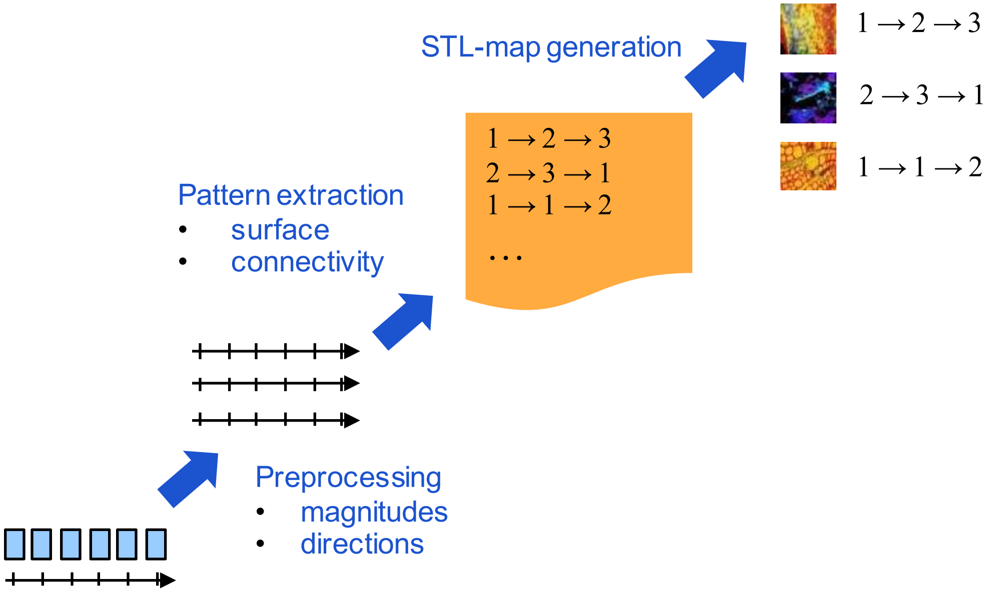
07/09/2011



23/08/2011

Workflow

Patterns can be numerous. What are the most promising ones?



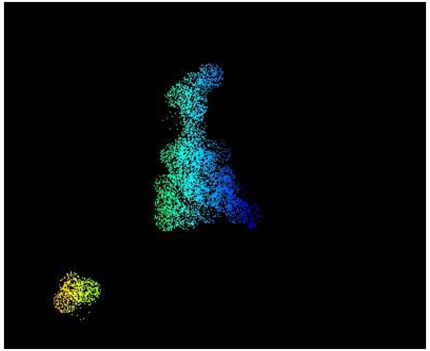
RGFS-patterns: pattern ranking

- Patterns can be numerous. What are the most promising ones?
  ➔ the most promising patterns have their occurrences destroyed OR maintained by randomization

- GFS-patterns occurrences contain spatiotemporal information: support or AC (via p-values or support ratios) are insufficient
  ➔ STL-maps

- Standard tests (e.g. p-value) require lots of randomized datasets
  ➔ a single randomized dataset

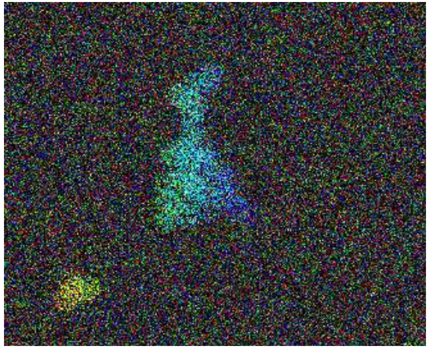
RGFS-patterns: Normalized Mutual Information - NMI



X

how similar?

Y



STL-map of 1 → 3
original DFTS

$I(X;Y) = H(X) - H(X|Y)$

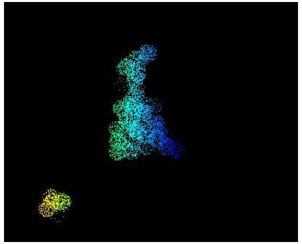
STL-map of 1 → 3
randomized DFTS

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

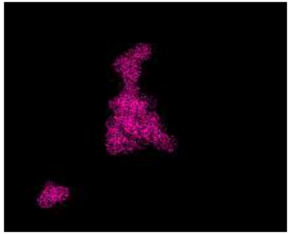
$$H(X) = -\sum_{x\in\Omega} P(x)\log P(x)$$

RGFS-patterns: NMI-based ranking

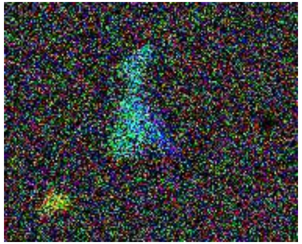
1 → 3

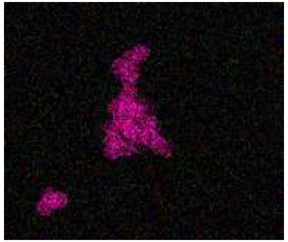


2 → 2 → 2 → 2 → 2 → 2



Original DFTS





Randomized DFTS

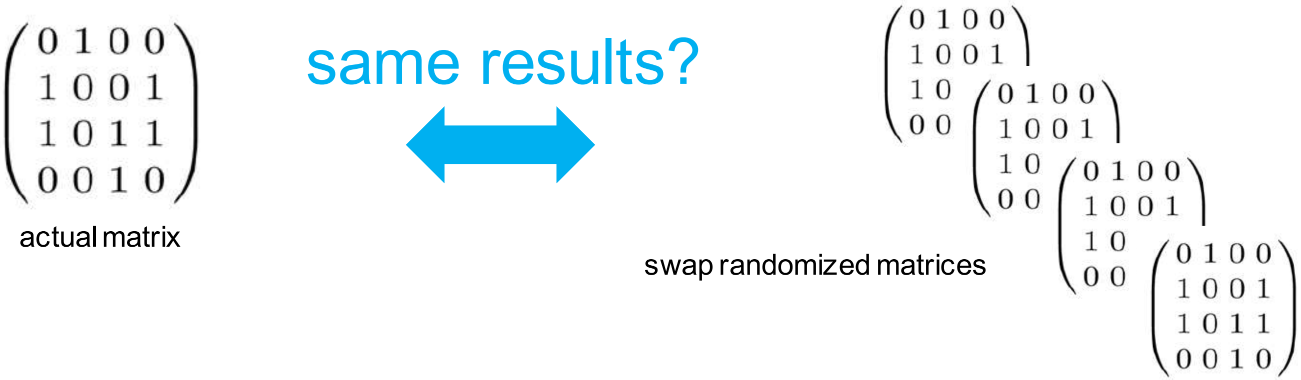
Destroyed by randomization

Hardly altered by randomization

NMI ranking

0 ↔ **1**

RGFS-patterns: swap randomization – Gionis et al. 2007



- Objective: to assess results (clusters, set of itemsets, itemsets, correlations, eigenvalues) obtained from Boolean matrices

- Null hypothesis: results are likely to be obtained on random matrices having the same column and row margins

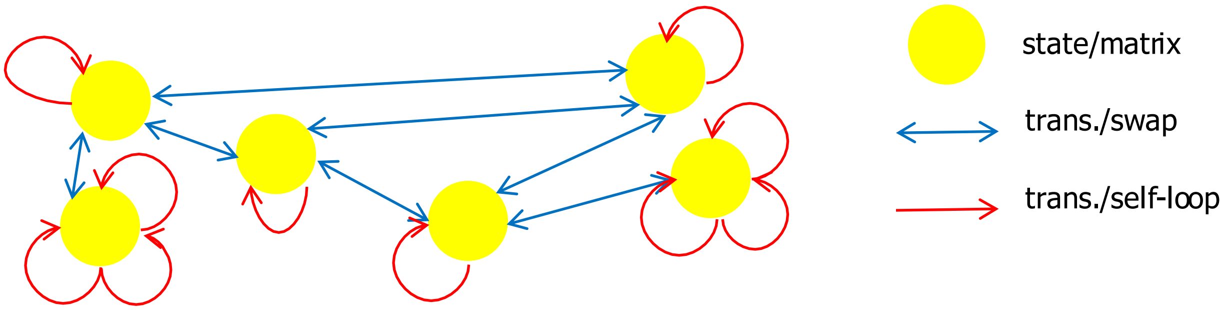
- Tests for frequent itemsets: p-values, support ratios.

Swap randomization: procedure

$$B = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \xrightarrow{\text{swap}} B' = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

- Randomized matrices are obtained by applying a series of swaps
- Pairs of 1's are chosen at random. Their number, P, is fixed
- If a pair B(i,j) = B(k,l) = 1 and if B(k,j) = B(i,l) = 0 then 1's and 0's are swapped
- Column and row margins are maintained
- All matrices having the same structure can be reached (Ryser 1957)

Swap randomization: equiprobable matrices and self-loops



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

- Failed swap attempts are counted as self-loops, each state degree = P ➔ uniform distribution

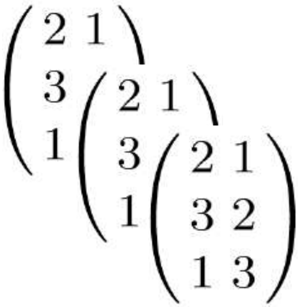
- All matrices having the same structure are equiprobable

- Mixing time is still an open research question

Swap randomization for symbolic matrices

$$\begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 1 \end{pmatrix}$$

same results?



- A base of sequences can be expressed as a symbolic matrix: row ⇔ pixel, column ⇔ date
- Objective: to assess patterns obtained from symbolic matrices representing a DFTS
- The spatiotemporal nature of the observed phenomena must be preserved
- Do we find the same pattern occurrences in random matrices having the same symbol distributions over rows and columns?

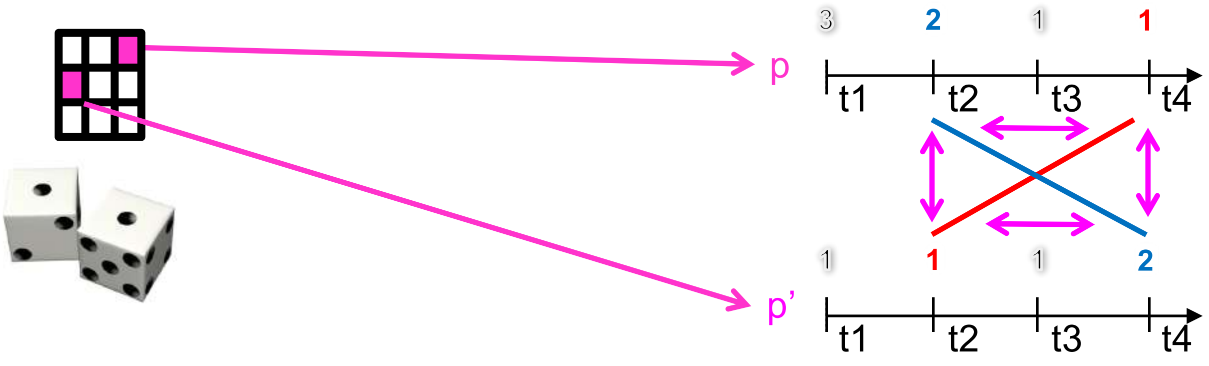
Swap randomization for symbolic matrices: procedure

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

- Pairs of elements sharing the same symbol are chosen at random.
- If a pair $B(i,j) = B(k,l) = \alpha$ and if $B(k,j) = B(i,l) = \beta$ ($\alpha \neq \beta$) then α's and β's are swapped
- Symbol distributions are maintained for each column and row while GFS-pattern occurrences are affected
- Not all matrices having the same structure can be reached
- Self-loops are also considered to explore equiprobable matrices

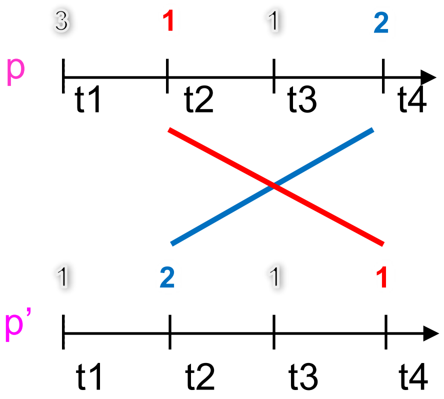
Swap randomization for symbolic matrices in a nutshell

A pair of pixel states sharing the same symbol is chosen randomly



symbols to be swapped

Swap randomization is done spatially and temporally

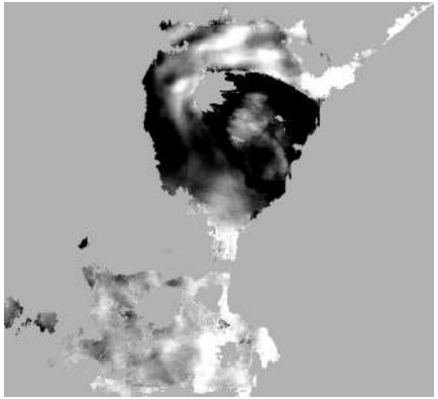


symbols once swapped

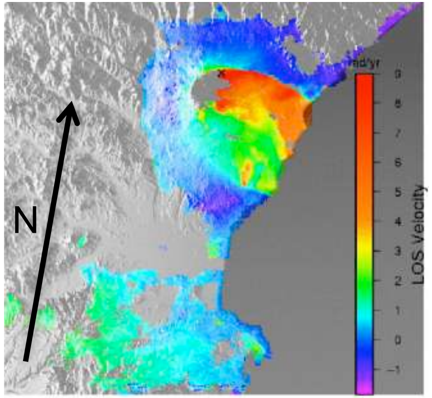
Mount Etna: deformation monitoring

- Acquisitions: Envisat ascending tracks (looking eastward)
- 16 co-registered total phase delay images (553X598), 2003-2010, SAR geometry, ≈160 m.
- Displacement magnitudes in the Line Of Sight (LOS).
- Data produced by M-P. Doin's team, NSBAS chain, ISTerre lab.

DEM of the Mount Etna area

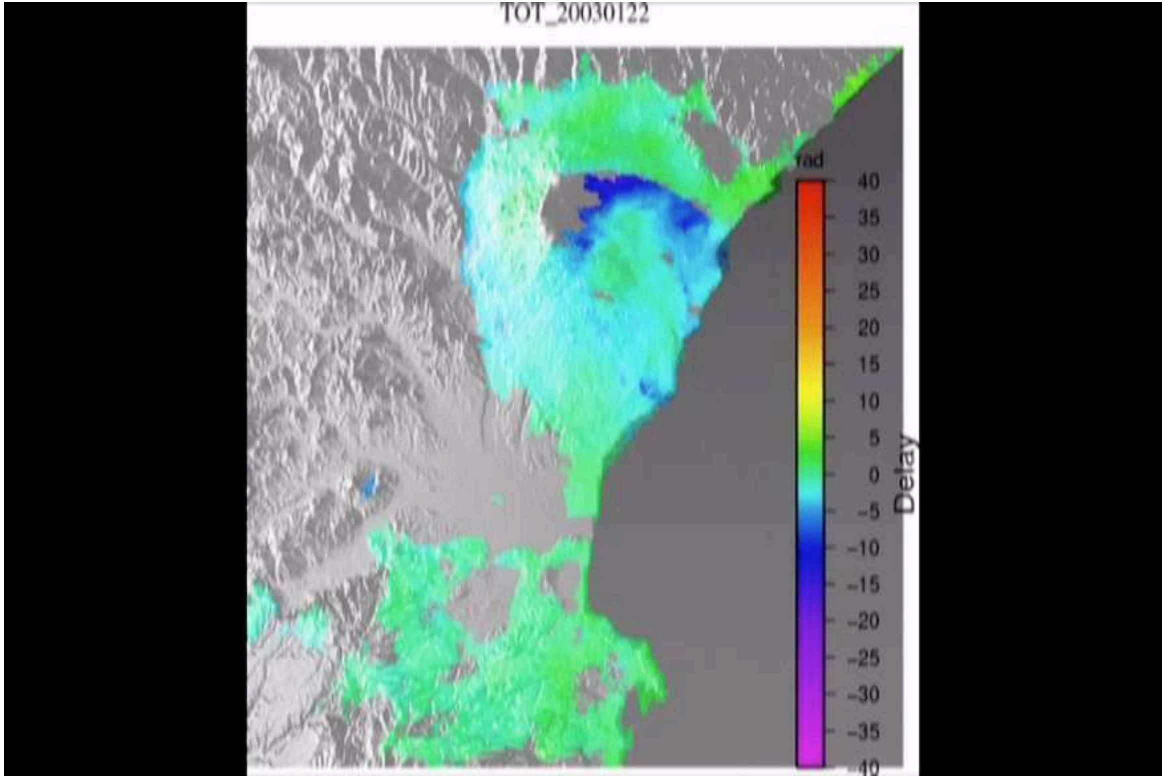


Total phase delays
2003/01/22



Average LOS velocity in rad/yr
(Doin et al. 2011)
$2\pi = 2.8$ cm

The Mount ETNA DFTS



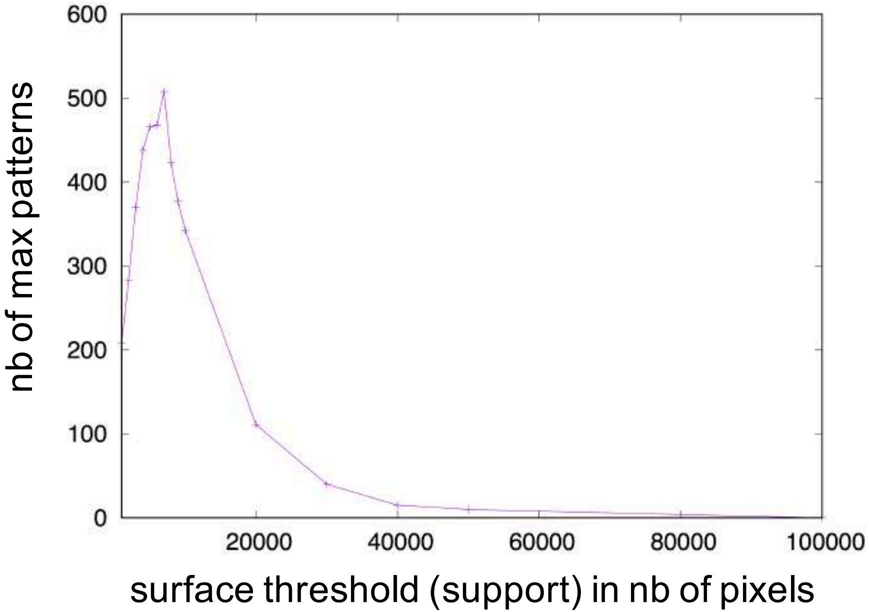
Mount Etna: parameters, number of patterns, ressources consumption

- Parameters :
    - nb of symbols = 3 (equal frequency bucketing)

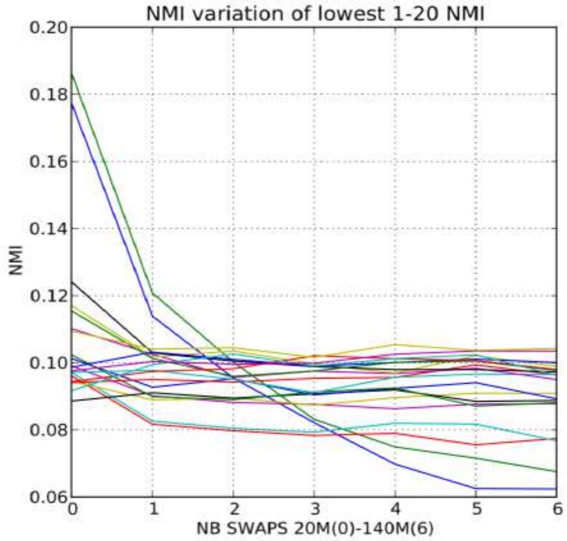
      3: motion away from satellite,
      2: small motion towards satellite,
      1: strong motion towards satellite

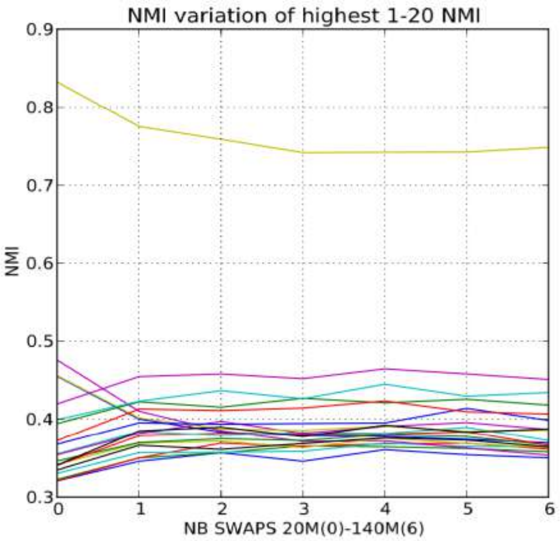
    - σ = 7000 (set to get as many maximal patterns as possible)
    - k = 5
    - nb swap attempts: 100 000 000 (about 20 x nb fields x nb pixels)
- Number of patterns: 2658 GFS-patterns, 508 maximal GFS-patterns
- Space/time requirements: 1.66 GB, 700 s. (single core on a 2.7 GHz Intel Core i7)

Mount Etna: nb of maximal GFS-patterns / surface threshold

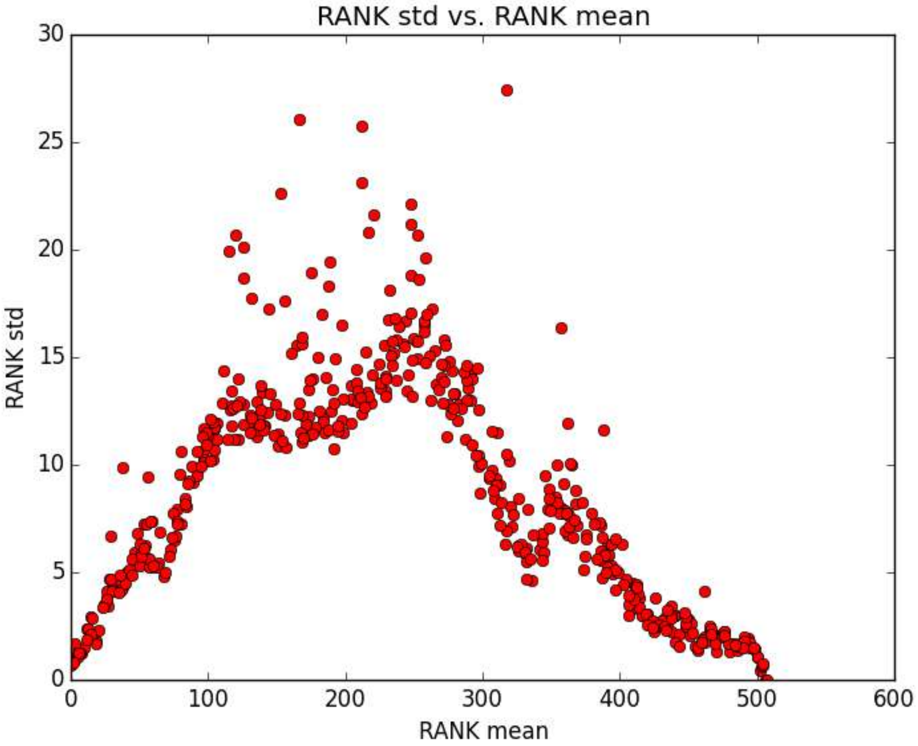


Mount Etna: 100M swap attempts





Mount Etna: ranking stability (over 1000 matrices)



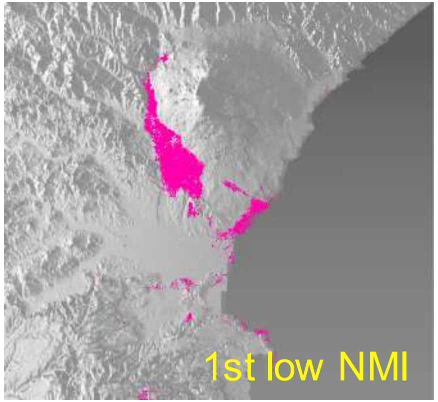
Mount Etna: qualitative results

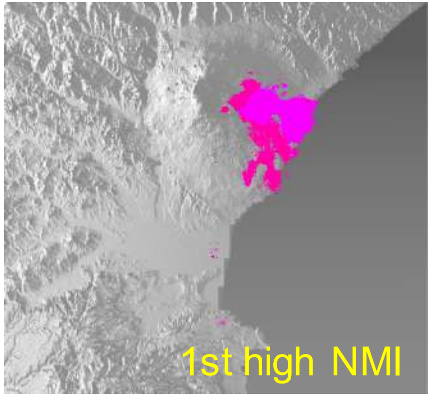
2003  2010

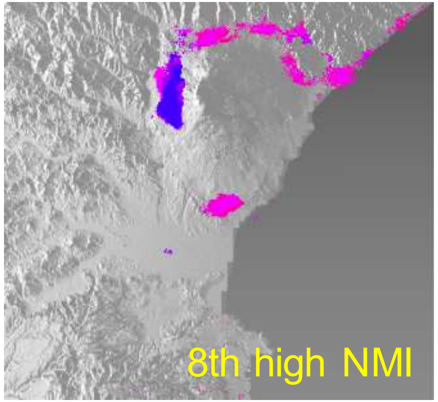
time →

1→1 → 2→ 1 → 1 → 1 → 1 → 3

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

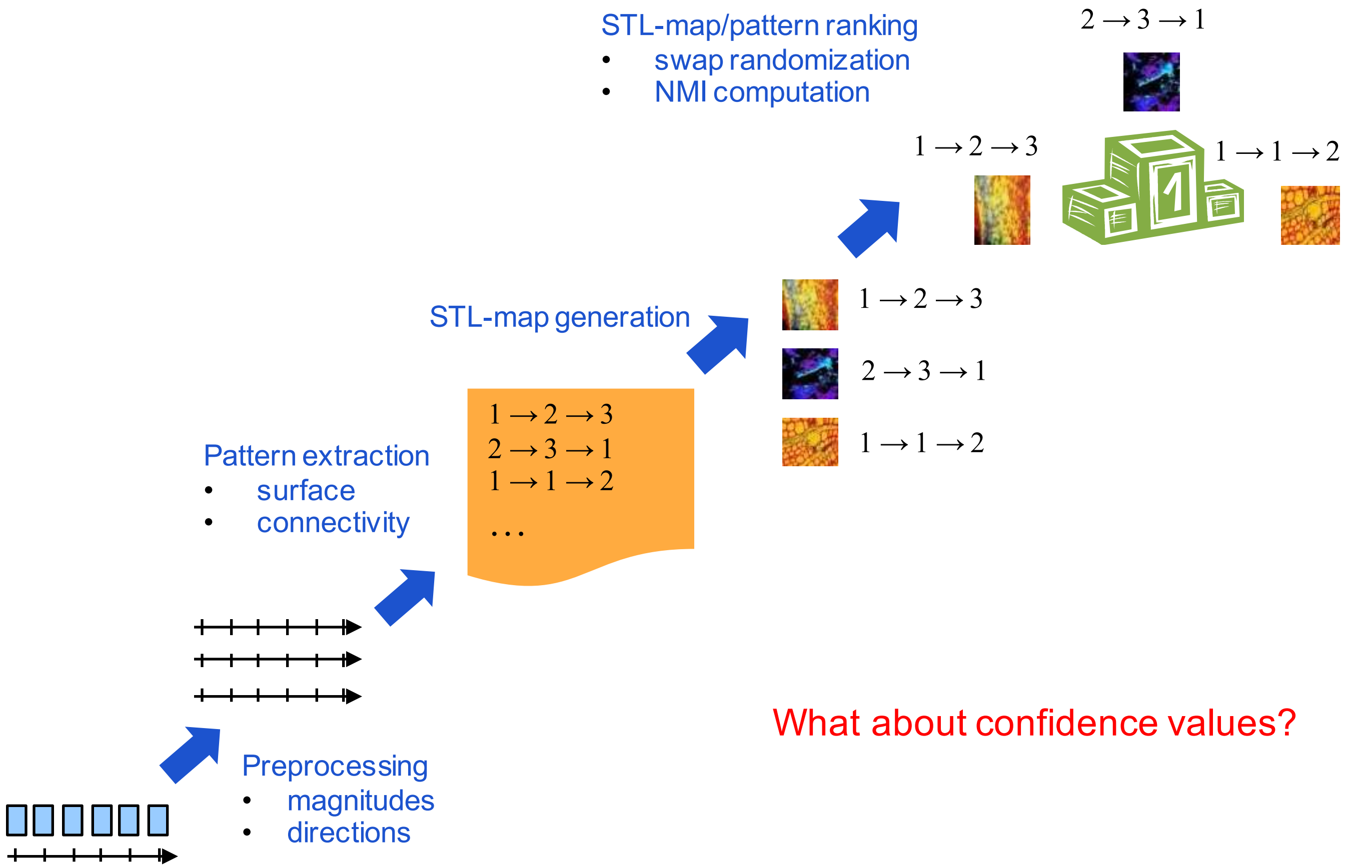






1 → 2 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3 → 3

Workflow

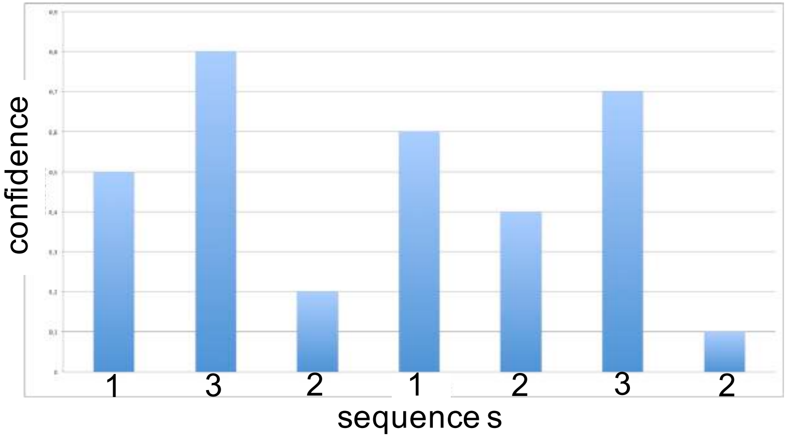


What about confidence values?

RGFS-patterns: symbols and confidence values

- Each symbol occuring at time t in a sequence located at position x,y is associated with a confidence value $\rho(x,y,t)$

$$s = \langle(1,\mathbf{1},0.5),(2,\mathbf{3},0.8),(3,\mathbf{2},0.2),(4,\mathbf{1},0.6),(5,\mathbf{2},0.4),(6,\mathbf{3},0.7),(7,\mathbf{2},0.1)\rangle$$



- Naïve approach: to extract GFS-patterns from high confidence symbols only

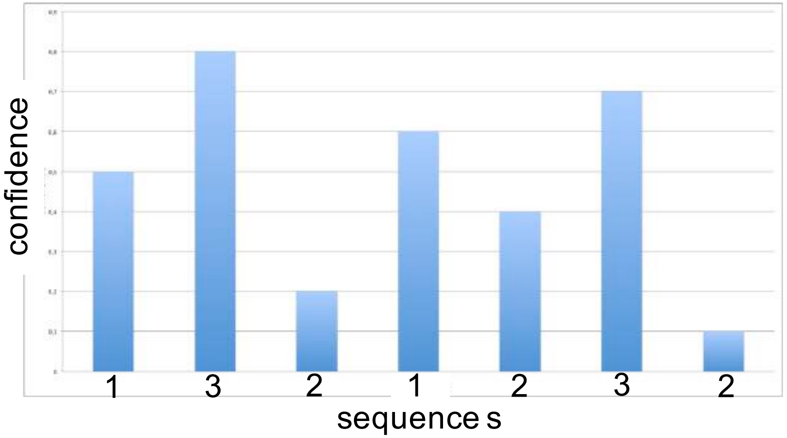
RGFS-patterns: Reliable GFS-patterns – RGFS-patterns

1. Occurrence reliability $$\rho_{occ}(seq(x,y),o) = \min\{\rho(x,y,t) \mid t \text{ } in \text{ } tuple \text{ } o\}$$
2. Pattern reliability at the scale of a sequence $$\rho_{pat}(seq(x,y),\beta) = \max_{o \in \mathcal{O}(seq(x,y),\beta)} \{\rho_{occ}(seq(x,y),o)\}$$
3. Pattern reliability at the scale of a base of sequences $$\rho(\beta) = \frac{\sum_{seq(x,y) \in cover(\beta)} \rho_{pat}(seq(x,y),\beta)}{support(\beta)}$$
4. A GFS-pattern β is reliable if $$\boxed{C_\rho(\beta) \equiv \rho(\beta) \geq \gamma}$$

RGFS-patterns: example

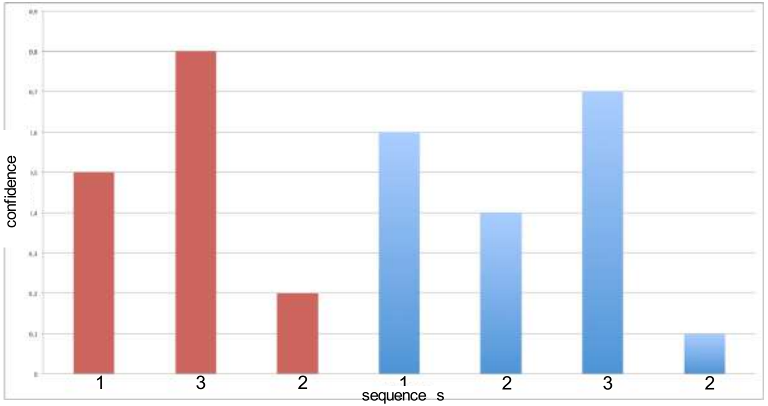
$$s = \langle (1, \mathbf{1}, 0.5), (2, \mathbf{3}, 0.8), (3, \mathbf{2}, 0.2), (4, \mathbf{1}, 0.6), (5, \mathbf{2}, 0.4), (6, \mathbf{3}, 0.7), (7, \mathbf{2}, 0.1) \rangle$$

$$\beta = 1 \rightarrow 3 \rightarrow 2$$



RGFS-patterns: example

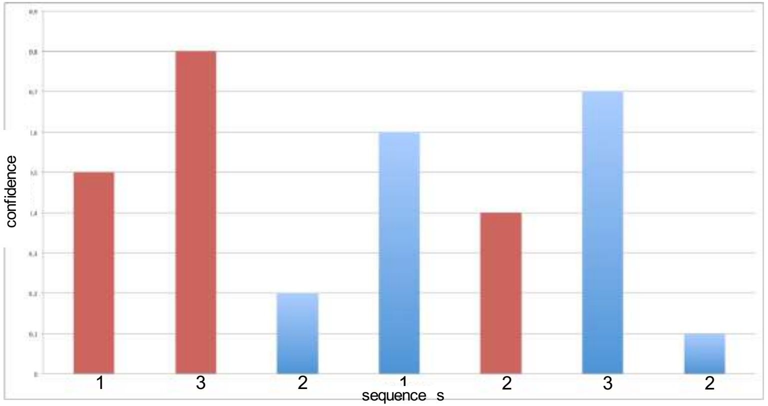
$$s = \langle (1, \mathbf{1}, 0.5), (2, \mathbf{3}, 0.8), (3, \mathbf{2}, 0.2), (4, \mathbf{1}, 0.6), (5, \mathbf{2}, 0.4), (6, \mathbf{3}, 0.7), (7, \mathbf{2}, 0.1) \rangle$$
$$\beta = 1 \to 3 \to 2$$



$$\rho_{occ}(x, y, o_1) = \min\{0.5, 0.8, 0.2\} = 0.2$$

RGFS-patterns: example

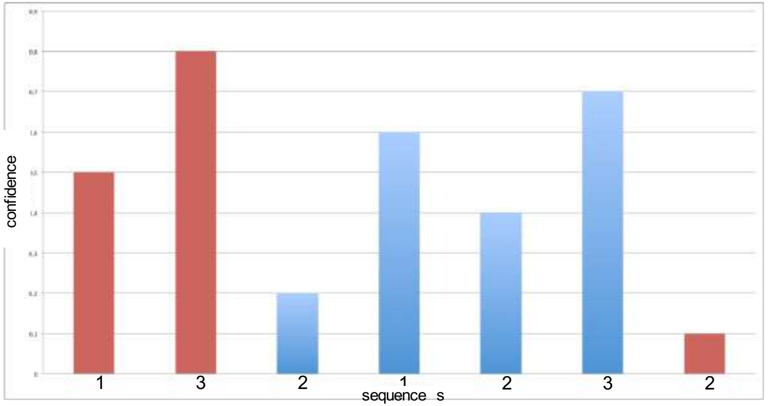
$$s = \langle (1, \mathbf{1}, 0.5), (2, \mathbf{3}, 0.8), (3, \mathbf{2}, 0.2), (4, \mathbf{1}, 0.6), (5, \mathbf{2}, 0.4), (6, \mathbf{3}, 0.7), (7, \mathbf{2}, 0.1) \rangle$$
$$\beta = 1 \rightarrow 3 \rightarrow 2$$



$$\rho_{occ}(x, y, o_1) = \min\{0.5, 0.8, 0.2\} = 0.2$$
$$\rho_{occ}(x, y, o_2) = \min\{0.5, 0.8, 0.4\} = 0.4$$

RGFS-patterns: example

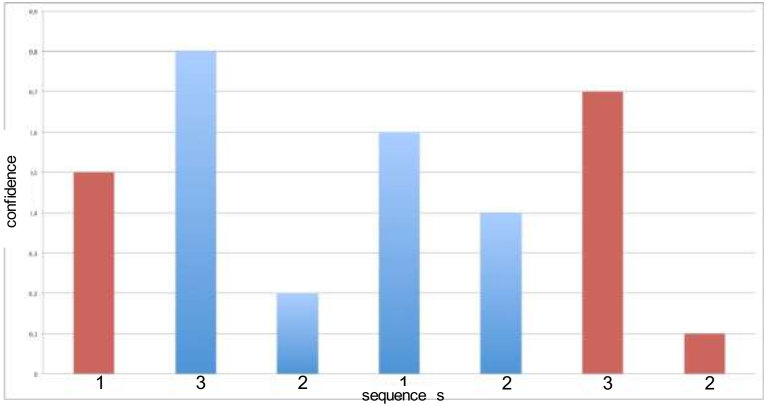
$s = \langle(1, \mathbf{1}, 0.5), (2, \mathbf{3}, 0.8), (3, \mathbf{2}, 0.2), (4, \mathbf{1}, 0.6), (5, \mathbf{2}, 0.4), (6, \mathbf{3}, 0.7), (7, \mathbf{2}, 0.1)\rangle$
$\beta = 1 \rightarrow 3 \rightarrow 2$



$$\rho_{occ}(x, y, o_1) = \min\{0.5, 0.8, 0.2\} = 0.2$$
$$\rho_{occ}(x, y, o_2) = \min\{0.5, 0.8, 0.4\} = 0.4$$
$$\rho_{occ}(x, y, o_3) = \min\{0.5, 0.8, 0.1\} = 0.1$$

RGFS-patterns: example

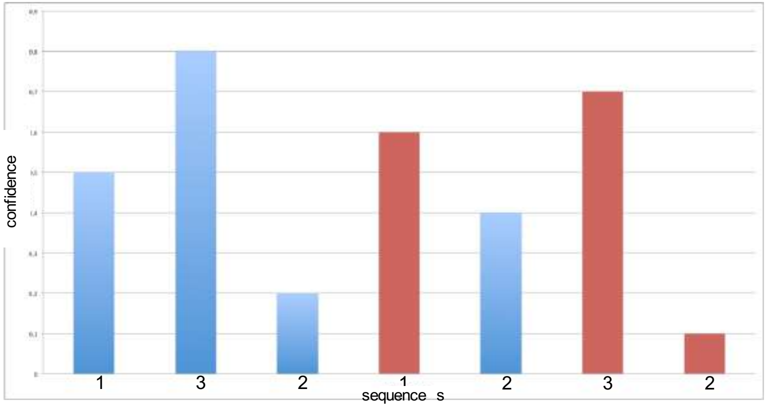
$$s = \langle (1, \mathbf{1}, 0.5), (2, \mathbf{3}, 0.8), (3, \mathbf{2}, 0.2), (4, \mathbf{1}, 0.6), (5, \mathbf{2}, 0.4), (6, \mathbf{3}, 0.7), (7, \mathbf{2}, 0.1) \rangle$$
$$\beta = 1 \to 3 \to 2$$



$$\rho_{occ}(x, y, o_1) = \min\{0.5, 0.8, 0.2\} = 0.2$$
$$\rho_{occ}(x, y, o_2) = \min\{0.5, 0.8, 0.4\} = 0.4$$
$$\rho_{occ}(x, y, o_3) = \min\{0.5, 0.8, 0.1\} = 0.1$$
$$\rho_{occ}(x, y, o_4) = \min\{0.5, 0.7, 0.1\} = 0.1$$

RGFS-patterns: example

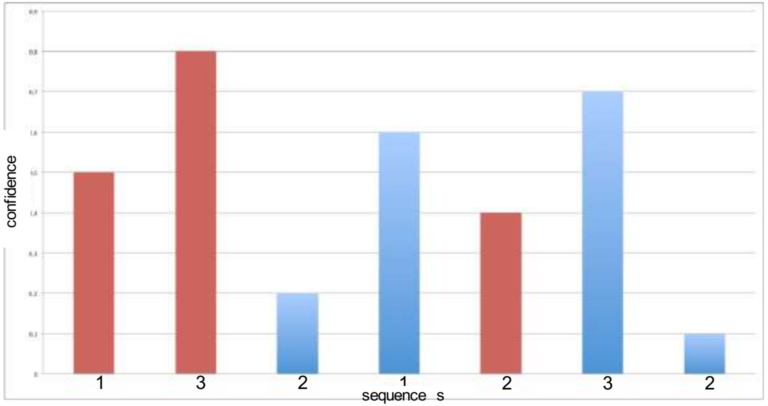
$$s = \langle (1, \mathbf{1}, 0.5), (2, \mathbf{3}, 0.8), (3, \mathbf{2}, 0.2), (4, \mathbf{1}, 0.6), (5, \mathbf{2}, 0.4), (6, \mathbf{3}, 0.7), (7, \mathbf{2}, 0.1) \rangle$$
$$\beta = 1 \to 3 \to 2$$



$$\rho_{occ}(x, y, o_1) = \min\{0.5, 0.8, 0.2\} = 0.2$$
$$\rho_{occ}(x, y, o_2) = \min\{0.5, 0.8, 0.4\} = 0.4$$
$$\rho_{occ}(x, y, o_3) = \min\{0.5, 0.8, 0.1\} = 0.1$$
$$\rho_{occ}(x, y, o_4) = \min\{0.5, 0.7, 0.1\} = 0.1$$
$$\rho_{occ}(x, y, o_5) = \min\{0.6, 0.7, 0.1\} = 0.1$$

RGFS-patterns: example

$$s = \langle (1, \mathbf{1}, 0.5), (2, \mathbf{3}, 0.8), (3, \mathbf{2}, 0.2), (4, \mathbf{1}, 0.6), (5, \mathbf{2}, 0.4), (6, \mathbf{3}, 0.7), (7, \mathbf{2}, 0.1) \rangle$$
$$\beta = 1 \rightarrow 3 \rightarrow 2$$



$$\rho_{occ}(x, y, o_1) = \min\{0.5, 0.8, 0.2\} = 0.2$$
$$\rho_{occ}(x, y, o_2) = \min\{0.5, 0.8, 0.4\} = 0.4$$
$$\rho_{occ}(x, y, o_3) = \min\{0.5, 0.8, 0.1\} = 0.1$$
$$\rho_{occ}(x, y, o_4) = \min\{0.5, 0.7, 0.1\} = 0.1$$
$$\rho_{occ}(x, y, o_5) = \min\{0.6, 0.7, 0.1\} = 0.1$$
$$\rho_{max} = \max\{0.2, 0.4, 0.1\} = 0.4$$

- Dynamic programming

RGFS-patterns: partial pushing of the reliability constraint

- The pattern reliability constraint is not anti-monotone but …
- $\rho(\beta) \leq$ $\widetilde{\rho}(\beta) = \dfrac{\sum_{seq(x,y)\in cover(\beta)} \rho_{pat}(seq(x,y),\beta)}{\sigma}$ (upper bound)
- $C_{\widetilde{\rho}}(\beta) \equiv \widetilde{\rho}(\beta) \geq \gamma$ is anti-monotone

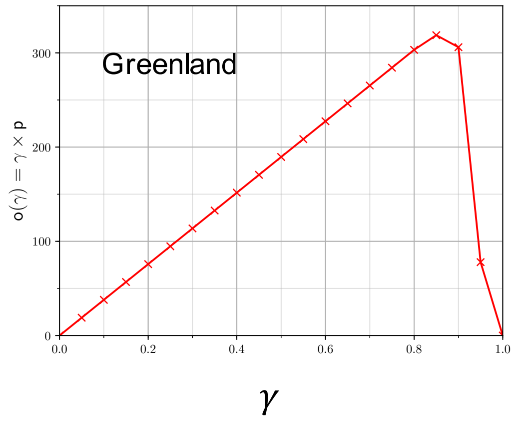
- Partial pushing
  - pruning using the upper bound constraint
  - selection of the reliable GFS-patterns

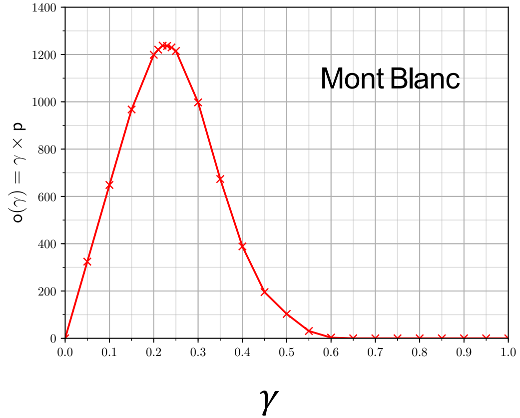
RGFS-patterns: application to glacier monitoring

	Greenland	**Mont Blanc**
Satellites	Landsat (5,7,8) (optical data)	TerraSAR-X (radar data), asc. track
DFTS	20 annual fields (median differential velocity) 1985 – 2014, 458 x 500 pixels, res. 240m x 240m (Tedstone et al. 2015)	25 fields over 11-days each (median differential velocity), May→October, 2009 and 2011, 3x3 reduction, 3494 x 3186 pixels (EFIDIR Tools), res. about 6m x 6m

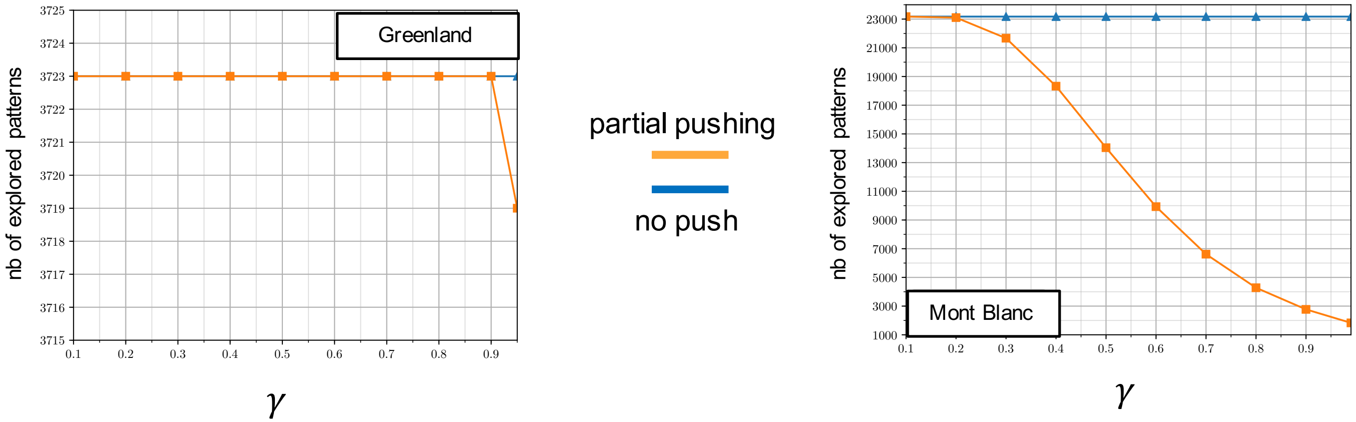
RGFS-patterns: parameters

	Greenland	**Mont Blanc**
symbols (equal frequency bucketing)	1 (low velocity), 2 (close to median), 3 (high)	1 (low velocity), 2 (close to median), 3 (high)
grouping threshold k (average connectivity)	5	5
surface threshold σ (support) (s.t. max. nb of maximal patterns)	7.5%	4%
confidence threshold γ (reliability) (s.t. max. of γ x nb of maximal reliable patterns)	0.85	0.22
ranking	375 max RGFS NMI swap randomization	5625 max RGFS NMI swap randomization





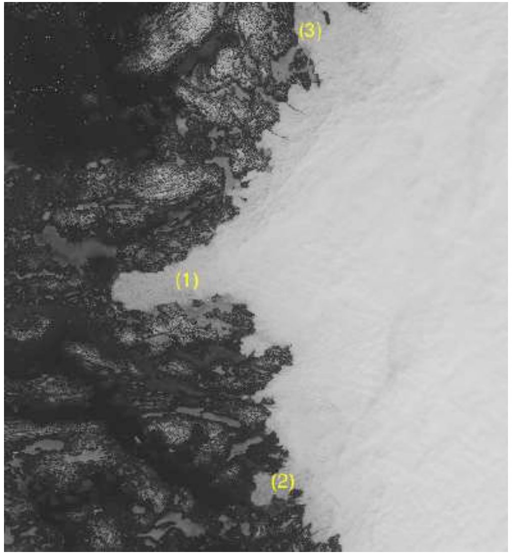
RGFS-patterns: search space reduction



For the retained settings, using an Intel Xeon 3.5 GHz, 1 core:

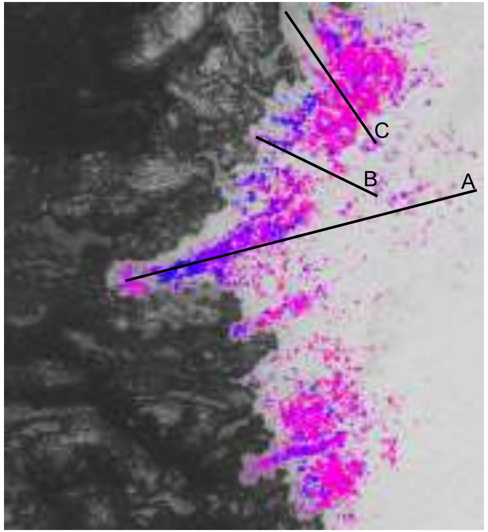
- Greenland - 813 s, 311 Mo
- Mont Blanc - 33 hours 18 minutes, 7470 Mo

RGFS-patterns in the western Greenland Ice Sheet ablation zone

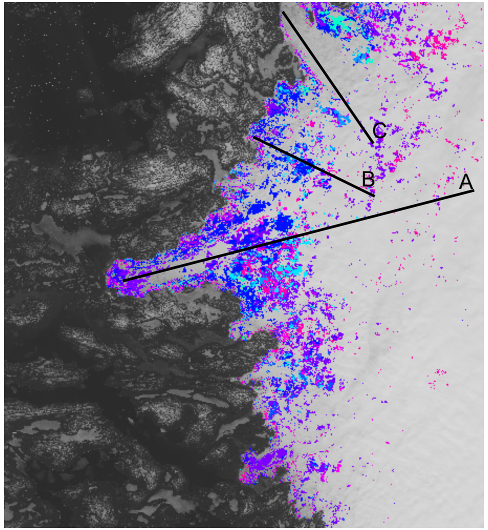


Three of the main glaciers in the area (about 120 km x 120 km)

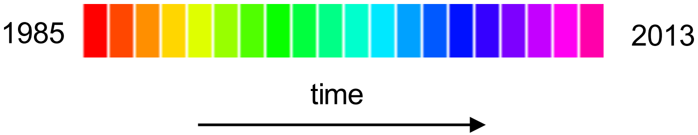
RGFS-patterns over the Greenland Ice Sheet



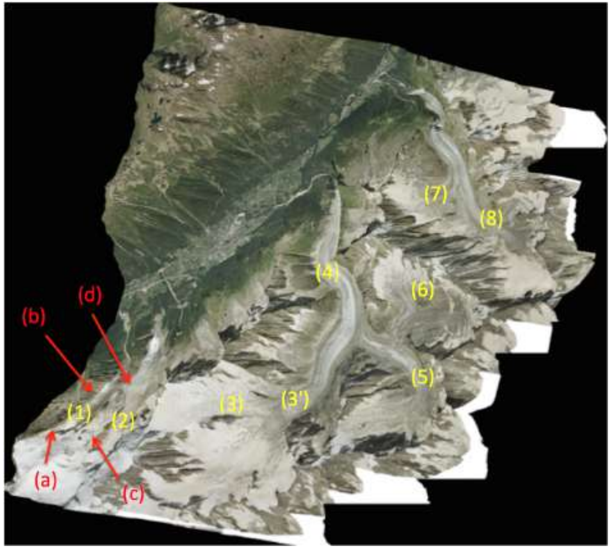
3 → 3 → 2 → 2 → 2 → 2 → 2 → 2 → 1 → 1
progressive slowdown (Tedstone el al. 2015)



3 → 3 → 3 → 3 → 3 → 3 → 3 → 2 → 1
sudden slowdown

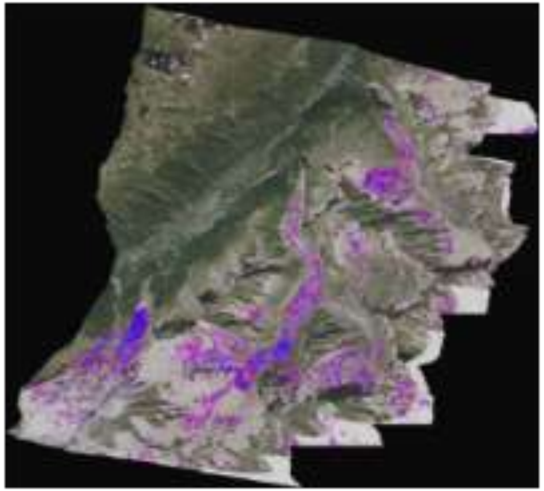


RGFS-patterns in the Mont Blanc area



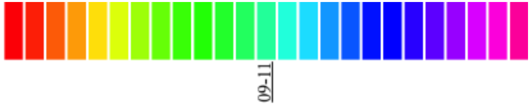
Main glaciers in the area (about 20 km x 20 km) in radar geometry
(1) Taconnaz, (2) Bossons,
(a) head of Taconnaz, (b) 2000m from head, (c) head of Bossons, (d) 2000m from head

RGFS-patterns over the Mont Blanc massif



$3 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 2$

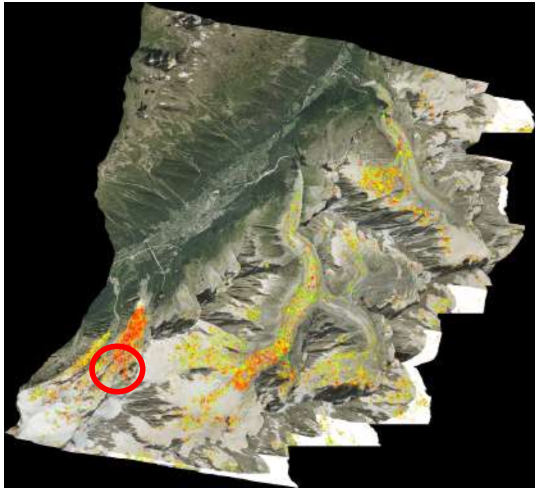
2009



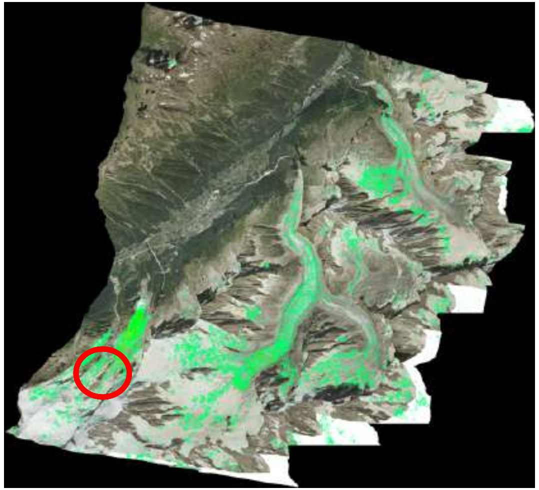
2011

time

RGFS-patterns over the Mont Blanc massif



First symbol of
$3 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 2$
(~ early summer, 2009)

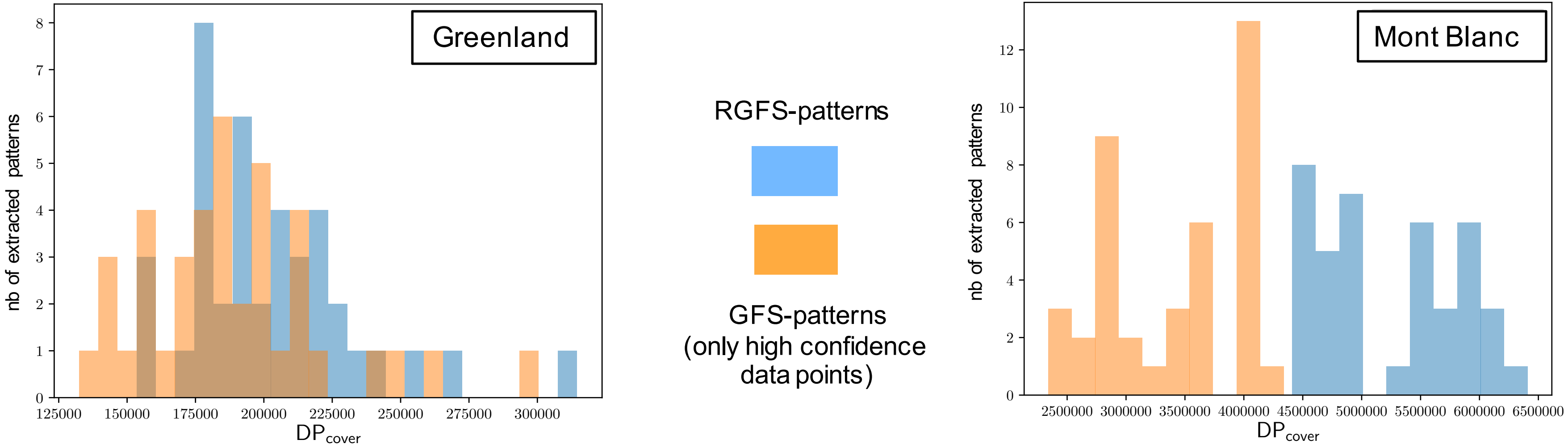


Last symbol 1 of
$3 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 2$
(~ summer and automn, 2009)

- Compatible with [Fallourd 2012]:
  annual cycles (observation on transects) (well known for temperate glaciers)
- Fluctuations of Bossons up to 3000 m, suggest cold based glacier zone is restricted to higher altitude

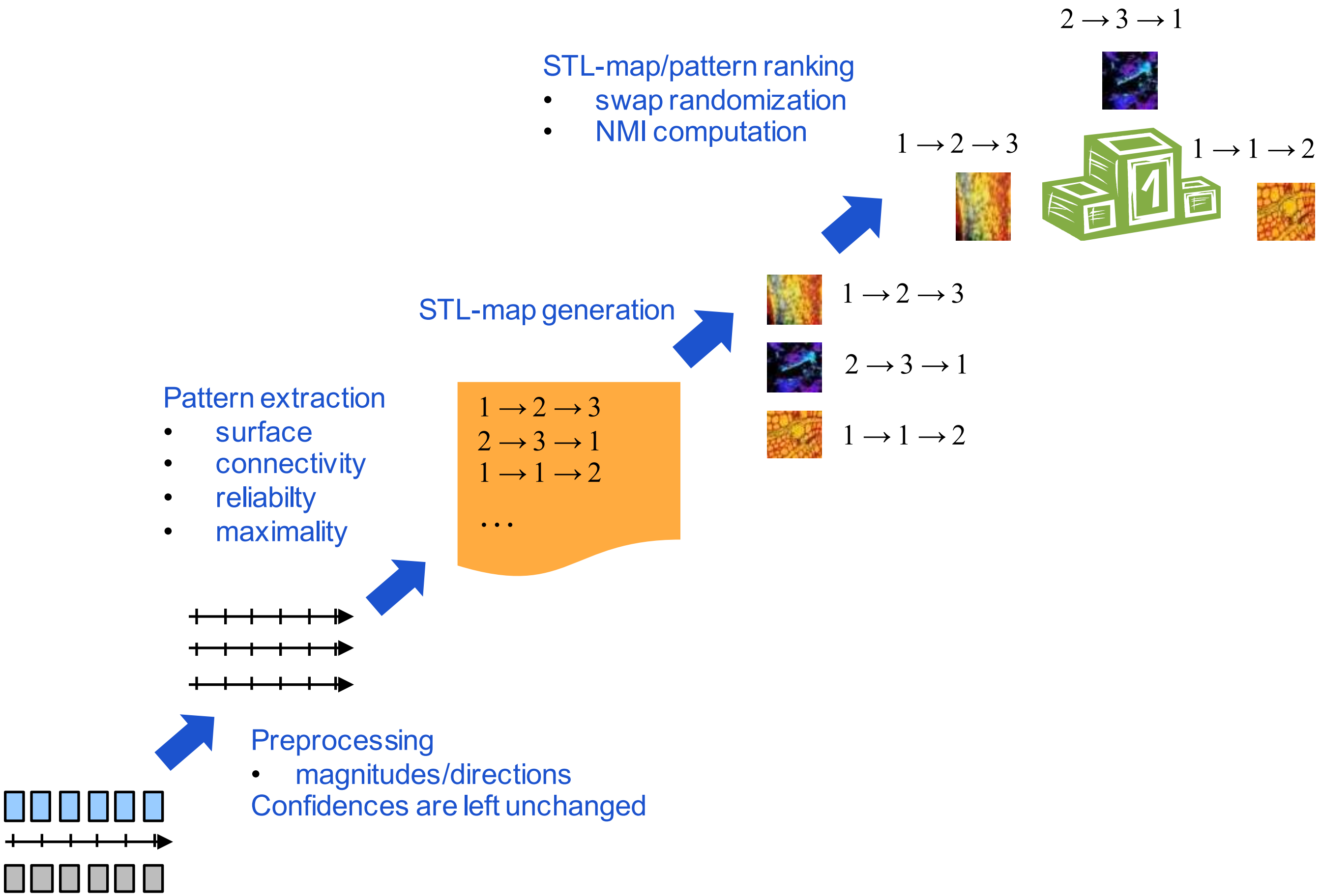
RGFS-patterns: what about the naïve approach?

- Data Point cover of β, a pattern having m symbols: $DP_{cover}(\beta) = support(\beta) * m$



- Mean Data Point cover of R, the set of selected patterns: $MDP_{cover}(R) = \frac{\sum_{\beta \in R} DP_{cover}(\beta)}{|R|}$
  - MDP gain Groenland: 7.2 %
  - MDP gain Mont-Blanc: 53.4%

Workflow


$2 \rightarrow 3 \rightarrow 1$
STL-map/pattern ranking
• swap randomization
• NMI computation
$1 \rightarrow 2 \rightarrow 3$
$1 \rightarrow 1 \rightarrow 2$
$1 \rightarrow 2 \rightarrow 3$
$2 \rightarrow 3 \rightarrow 1$
$1 \rightarrow 1 \rightarrow 2$
STL-map generation
Pattern extraction
• surface
• connectivity
• reliabilty
• maximality
$1 \rightarrow 2 \rightarrow 3$
$2 \rightarrow 3 \rightarrow 1$
$1 \rightarrow 1 \rightarrow 2$
$\ldots$
Preprocessing
• magnitudes/directions
Confidences are left unchanged

When should I use the method?

If only 5 fields of good quality over an area I know well ...

... I do not use the method

If 15 fields of poor quality and I am not an expert of the area ...

... I try it ... it can suggest hypothesis

by finding groups of data points forming regularities over time,

that are, on average, connected over space

and build from "good" quality measures

More information

- RGFS-patterns for DFTS mining / DFTS-P2miner basis: Tuan Nguyen, Nicolas Méger, Christophe Rigotti, Catherine Pothier, Emmanuel Trouvé, Noel Gourmelen & Jean-Louis Mugnier (2018). A pattern-based method for handling confidence measures while mining satellite displacement field time series. Application to Greenland ice sheet and Alpine glaciers. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, n°11, pp. 4390-4402.

- GFS-patterns and pattern ranking (swap randomizaion, NMI, no confidence): Nicolas Méger, Christophe Rigotti, Catherine Pothier, Tuan Nguyen, Felicity Lodge, Lionel Gueguen, Rémi Andréoli, Marie-Pierre Doin & Mihai Datcu (2019). Ranking evolution maps for Satellite Image Time Series exploration: application to crustal deformation and environmental monitoring. *Data Mining and Knowledge Discovery*, vol. 33, n°1, pp. 131-167.

Workflow supported by the DFTS-P2miner plateform

Process decomposed in 5 activities:

1- DFTS quantization (equal frequency bucketting)

2- RGFS-patterns extraction

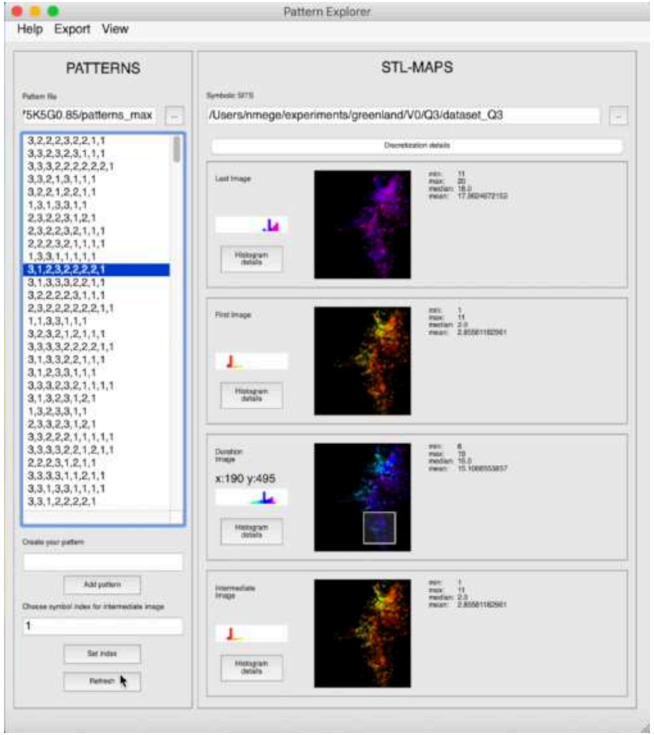
3- maximal RGFS-pattern selection

4- STL-map computation

5- randomization and pattern ranking (select N-highest and M-lowest NMI)

And 6- Use the GUI to explore the patterns

- PatternExplorer Graphical User Interface
- Patterns & pattern variants
- STL-maps (starting, ending, duration intermediate element)
- Temporal statistics
- Subarea selection and tiling mode
- Exploration materials can be exported (statistics, maps)



DFTS-P2miner: technical facts and download links

- Python 2.7 and C (advanced code for time/memory consuming tasks)
- C binaries + Python sources distributed for Mac & Linux (x64) – free for non commercial use.

- DFTS-P2miner tutorial: https://sites.google.com/view/dfts-miner-tutorial
  (a virtual machine ready to install DFTS-P2miner, DFTS-P2miner itself, documentation about the methods and the platform ...)

- DFTS-P2miner only: https://sites.google.com/view/dfts-p2miner
  (to install it directly on a system, without using the virtual machine, a script for detecting missing python libraries is provided)

DFTS-P2miner: a single parameter file

- select OS (Linux / Mac OS)
- the paths to the Python 2.7 interpreter, the DFTS-P2miner sources, the input DFTS, the output directory,
- the image/field format,
- the preprocessing, extraction and ranking parameters.
- and misc. options: select activities to perform, force recomputation, cleaning, ...

DFTS-P2miner: result directory main structure

- root directory of the results / design to ease exploratory mining and archiving

|

|-Q'a': results for 'a' quantization intervals

    |

    |-RANDOMIZED_DATASETS: randomized datasets computed to rank patterns/STL-maps

    |

    |-S'b'K'c'G'd': directory containing all results for execution with parameters σ=b, κ=c, γ=d

        |

        |-RAND_SWAP*: ranking results. The contents and the full name of the

        | directory depend on the ranking type and on the parameters.

        | "Best" patterns in subdirectories PATTERNS_MAX_HIGH/LOW_NMI*

        |

        |-STLmap_patterns_max: the STL-maps of the all maximal RGFS-patterns (can be cleaned automatically for storage reason, depending on options)

DFTS-P2miner: main result files

- root directory of the results
| files: "log_*" global log of each execution
|-Q'a': results for 'a' quantization intervals
    | files: "dataset_Q*" discretized dataset (the "symbolic" DFST)
    |
    |-S'b'K'c'G'd': directory containing all results for execution with parameters σ=b, κ=c, γ=d
        |
        | files: "log_comp_patterns_max_*" gives the pattern distribution vs pattern size
        | files: "patterns" and "pattern_max" gives low level information about the patterns
        | file: "colorPalette.tiff" gives the color scale used in the maps
        | in subdirectory "RAND_SWAP_*", file "patterns_max_sorted_by_NMI_*"

The result directory contains also copies of the parameter file and of the field list for archiving purpose (and a few other log files).

Practicals

take care, crucial step!

easy if parameter file OK …

easy

https://sites.google.com/view/dfts-miner-tutorial

Run the VM using run VirtualBox (see README FOR UBUNTU DFTS-P2MINER VM)

Download and unzip DFTS-P2miner in the VM (see Tutorial guide)

Check the parameter file of the example test_mb_light contained in DFTS-P2miner archive

Run DFTS-P2miner on the example (see README in test_mb_light)

Explore your results using PatternExplorer (see Tutorial guide)

Install the Greenland dataset and run DFTS-P2miner on it (see Tutorial guide)

References 1/2

- Agrawal, R., Imieliński, T., et Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, *22*(2), 207–216.

- Agrawal, R., et Srikant, R. (1995). Mining sequential patterns. In *Data Engineering, 1995. Proceedings of the Eleventh International Conference on*, (pp. 3–14). IEEE.

- Altena, B., Scambos, T., Fahnestock, M., et Kääb, A. (2018). Extracting recent short-term glacier velocity evolution over Southern Alaska from a large collection of Landsat data. *The Cryosphere Discussions*, (pp. 1–27).

- Doin M.-P., Lodge F., Guillaso S., et al. (2011). Presentation of the small-baseline NSBAS processing chain on a case example: the Etna deformation monitoring from 2003 to 2010 using ENVISAT data. *In Proc. of the European Space Agency Symposium "Fringe“*, Frascati, Italy, pp. 3434-3437.

- Fallourd, R. (2012). Suivi des glaciers alpins par combinaison d'informations hétérogènes : images SAR Haute Résolution et mesures terrain. *Ph.D. thesis, Université de Grenoble*.

- Malet, J-P. (2003). Les glissements de type écoulement dans les marnes noires des Alpes du Sud. Morphologie, fonctionnement et modélisation hydro-mécanique. *PhD thesis, Université Louis Pasteur-Strasbourg I*.

References 2/2

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

- Mannila, H., Toivonen, H., et Verkamo, A. I. (1997). Discovery of frequent episodes in event sequences. *Data mining and knowledge discovery*, *1*(3), 259–289.

- Raucoules, D., de Michele, M., Malet, J. P., et Ulrich, P. (2013). Time-variable 3D ground displacements from high-resolution synthetic aperture radar (SAR). application to La Valette landslide (South French Alps). *Remote Sensing of Environment,* 139, 198–204

- Ryser, H.J. (1957). Combinatorial properties of matrices of zeros and ones. *Canadian Journal of Mathematics* 9, 371–377

- Tedstone, A. J., Nienow, P. W., Gourmelen, N., Dehecq, A., Goldberg, D., et Hanna, E. (2015). Decadal slowdown of a land-terminating sector of the Greenland Ice Sheet despite warming. *Nature*, 526(7575), 692–695.

- Travelletti, J., Malet, J.-P (2012). Characterization of the 3d geometry of flow-like landslides : A methodology based on the integration of heterogeneous multi-source data. *Engineering Geology*, 128 :30 – 48. Integration of Technologies for Landslide Monitoring and Quantitative Hazard Assessment.

Main references related to the method

- Nicolas Méger, Christophe Rigotti, Catherine Pothier, Tuan Nguyen, Felicity Lodge, Lionel Gueguen, Rémi Andréoli, Marie-Pierre Doin & Mihai Datcu (2019). Ranking evolution maps for Satellite Image Time Series exploration: application to crustal deformation and environmental monitoring. *Data Mining and Knowledge Discovery*, vol. 33, n°1, pp. 131-167

- Tuan Nguyen, Nicolas Méger, Christophe Rigotti, Catherine Pothier, Emmanuel Trouvé, Noel Gourmelen & Jean-Louis Mugnier (2018). A pattern-based method for handling confidence measures while mining satellite displacement field time series. Application to Greenland ice sheet and Alpine glaciers. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, n°11, pp. 4390-4402.

- Tuan Nguyen T., Meger N., Rigotti C., Pothier C., and Andreoli R (2016). SITS-P2miner: Pattern-Based Mining of Satellite Image Time Series. *In Proc. of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD'16)*, Riva del Garda, Italie, September, 2016, pp63-66.

- Nicolas Meger, Chrsitophe Rigotti, Catherine Pothier (2015). Swap Randomization of Bases of Sequences for Mining Satellite Image Times Series. *In Proc. of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD'15)*, Porto, Portugal, September, 2015, pp. 190-205.

- Youen Pericault, Catherine Pothier, Nicolas Meger, Christophe Rigotti, Flavien Vernier, Ha-Tai Pham, Emmanuel Trouvé (2015). A swap randomization approach for mining motion field time series over the Argentiere glacier 2015. *In Proc. of th 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MULTI-TEMP 2015*), Annecy, France, July, 2015, pp.1-4.

- Christophe Rigotti, Felicity Lodge, Nicolas Méger, Catherine Pothier, Romain Jolivet & Cécile Lasserre (2014). Monitoring of Tectonic Deformation by Mining Satellite Image Time Series. *Reconnaissance de Formes et Intelligence Artificielle (RFIA) 201*, Rouen (France), July, 2014, pp. 1-6.

- Nicolas Méger, Romain Jolivet Cecile Lasserre, Emmanuel Trouvé, Christophe Rigotti, Felicity Lodge, M.P. Doin, Stéphane Guillaso, Andreea Julea, Philippe Bolon (2011). 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)*, Trento, Italy, July 2011, pp. 1-4.

- Andreea Julea, Nicolas Méger, Philippe Bolon, Christophe Rigotti, Marie-Pierre Doin, Cécile Lasserre, Emmanuel Trouvé, Vasile Lazarescu: Unsupervised Spatiotemporal Mining of Satellite Image Time Series Using Grouped Frequent Sequential Patterns. IEEE Trans. Geoscience and Remote Sensing 49(4): 1417-1430 (2011)

Other collaborators

(not mentioned in the main reference list)

Pauline Faraglia, INSA de Lyon

Jean-Philippe Malet, EOST/ IPGS – Université de Strasbourg


To be continued…

Questions?