



EFIDIR



FOSTER



Fouille de Données pour la Caractérisation du Déplacement par Imagerie Radar et autres résultats du projet EFIDIR

Nicolas Méger

et l'ensemble des participants au projet ANR EFIDIR



UNIVERSITE
CHAMBERY SAVOIE
ANNECY DE



LISTIC
Polytech

Plan

1. Pourquoi EFIDIR ?
 - Un constat
 - Données utilisées
2. EFIDIR : activité scientifique et technique
 - Sous-Projets, Work-Packages
 - Exemples de résultats
 - Actions transverses
 - Diffusion
3. Fouille de données spatio-temporelle : application à la caractérisation du déplacement
 - What is data mining?
 - Problem statement & GFS-patterns
 - Results
4. Perspectives

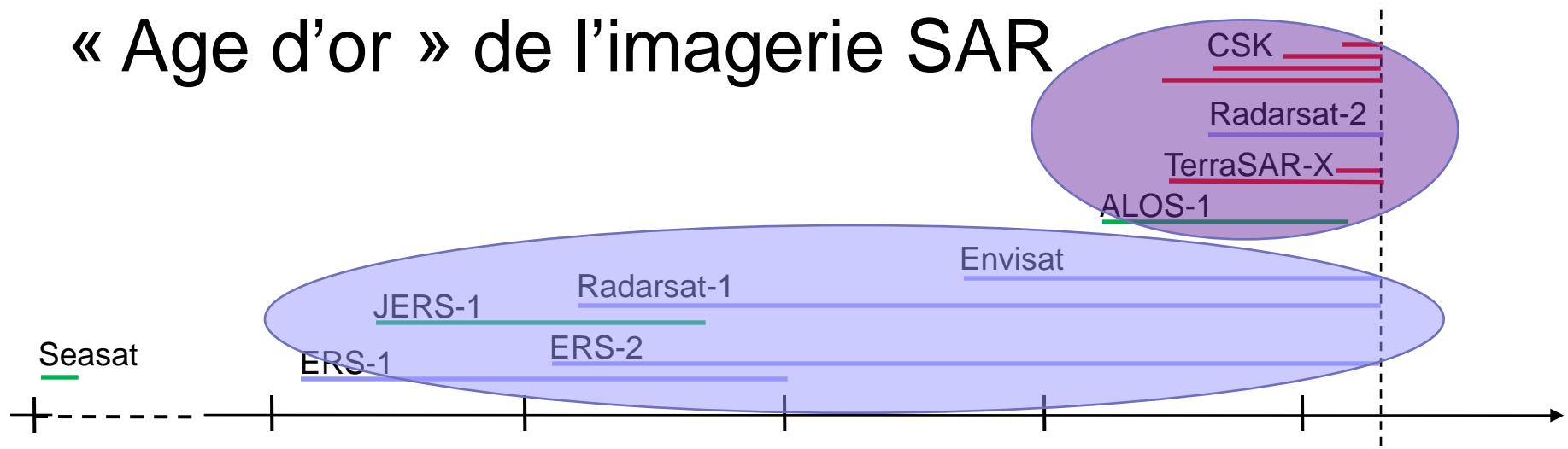
De la masse de données à la géophysique



Type de données	Séisme de Haïti	Séisme de Sendai
SAR	ALOS	✓
	ERS-2,ENVISAT	✓
	RadarSAT	✓
	Cosmo-skymed	✓
	TerraSAR-X	✓
données optiques	✓	✓
GPS	✓	✓
données topographiques	✓	-
mécanisme au foyer	✓	✓
analyses sismologiques	✓	✓

GEO's Event Supersite:
<http://supersites.earthobservations.org>

« Age d'or » de l'imagerie SAR



- 90-95 : 1^{ère} génération:
 - ERS-1, ERS-2, JERS, Radarsat-1 : bande C
 - 95-2005 :
 - ENVISAT (ESA): bande C, 20m, alt-pol, 35 jours
 - 2005-2008 : 2^{ème} génération
 - ALOS (JAXA): bande L, 10m, quad-pol, 46 jours
 - TerraSAR-X (DLR): bande X, ->1m, dual-pol, 11 jours
 - Radarsat-2 (CSA): bande C, ->3m, quad-pol, 24 jours
 - COSMO-SkyMed (ASI): bande X, ->1m, 4 satellites => 4 j.
- } Archives, Interférométrie
- } - Haute Résolution
- Polarimétrie

EFIDIR : ANR Masse de Données et COnnaissances

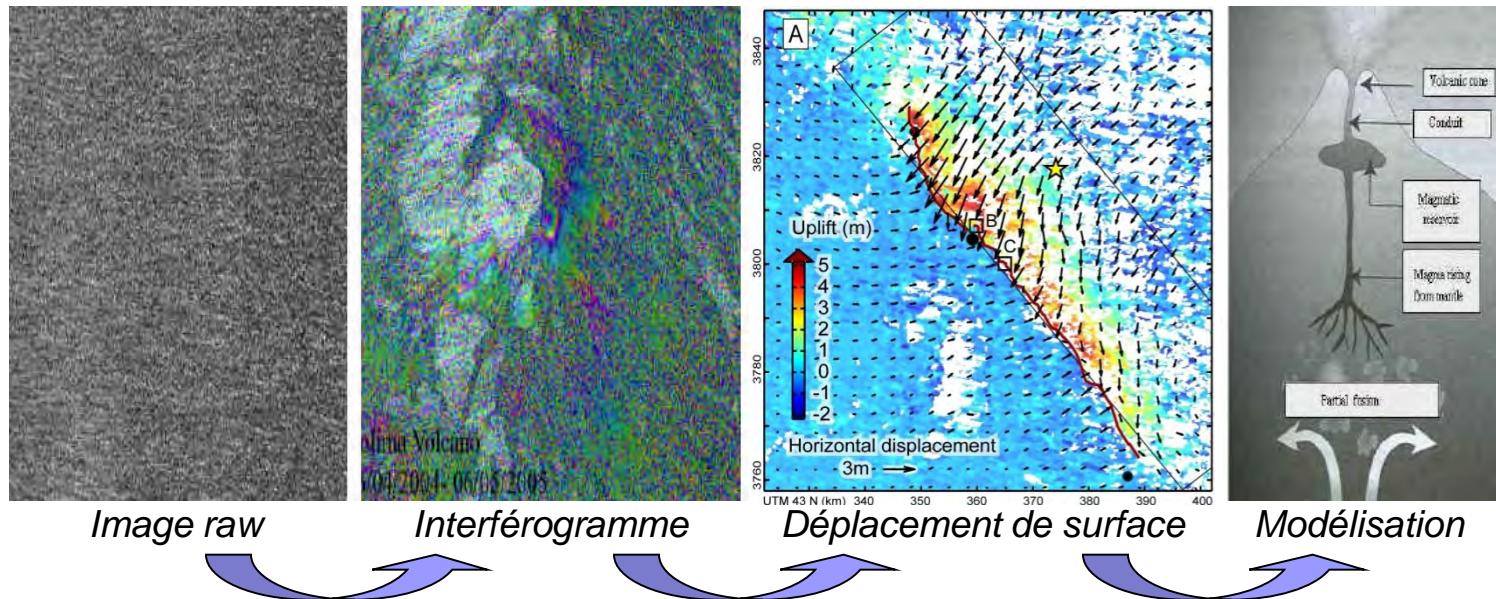
- 2008-2011 (+6 mois)
- 6 partenaires :
 - STIC
 - LISTIC, Université de Savoie
 - IETR, Université de Rennes
 - GIPSA, INP Grenoble
 - LTCI, Télécom ParisTech
 - Sciences de la Terre
 - ISTerre, Grenoble-Chambéry
 - LG, ENS Paris
 - Experts
 - B. Fruneau (U. Marne-la-Vallée)
 - F. Jouanne (U. de Savoie)
 - L. Moreau (U. de Savoie)



Extraction et Fusion d'Informations pour la mesure de Déplacement par Imagerie Radar

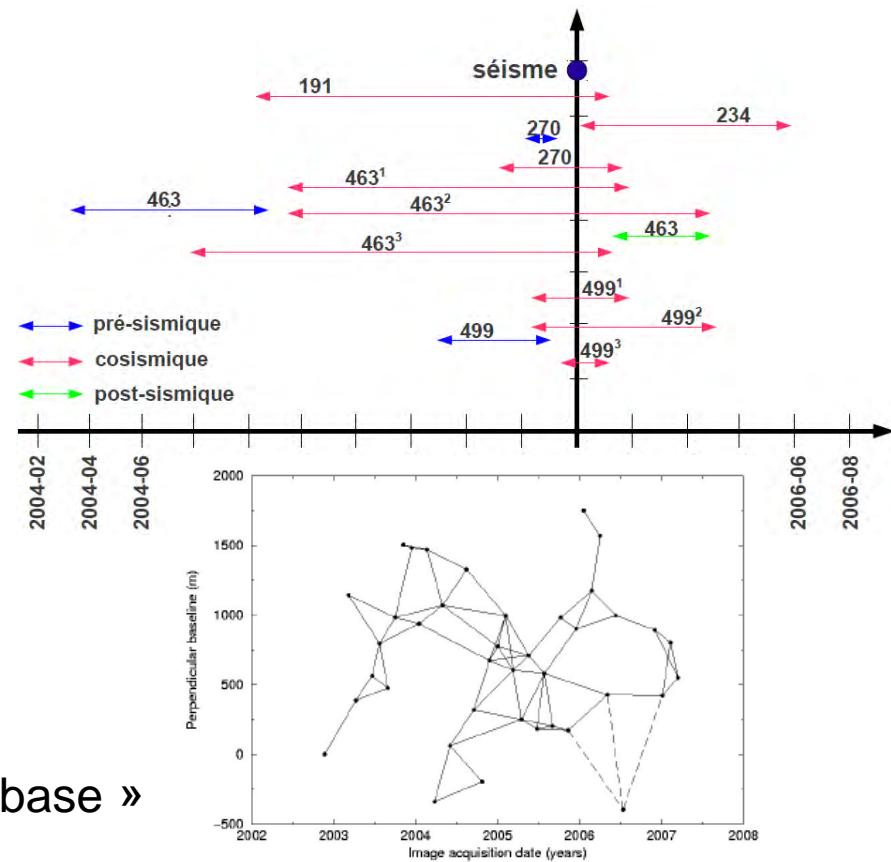
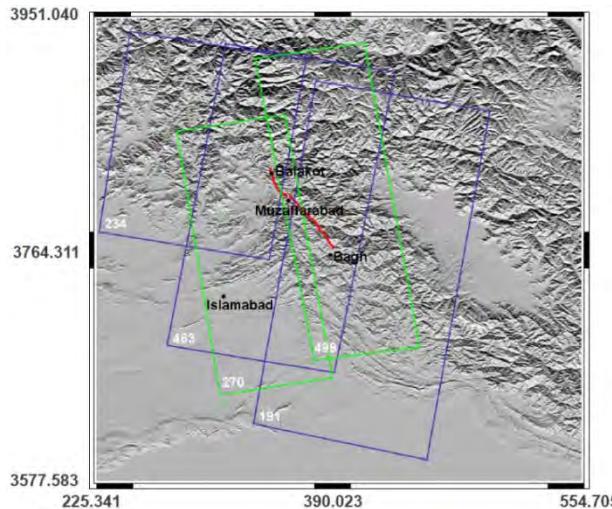
■ Objectifs :

- méthodes de traitement des données SAR :
 - existantes : exploitation des archives, séries temporelles...
 - « nouvelle génération » : haute résolution, polarimétrie...
- mesures de déplacement par imagerie SAR pour différentes applications : glaciers, volcans, séismes...



Exploitation des archives

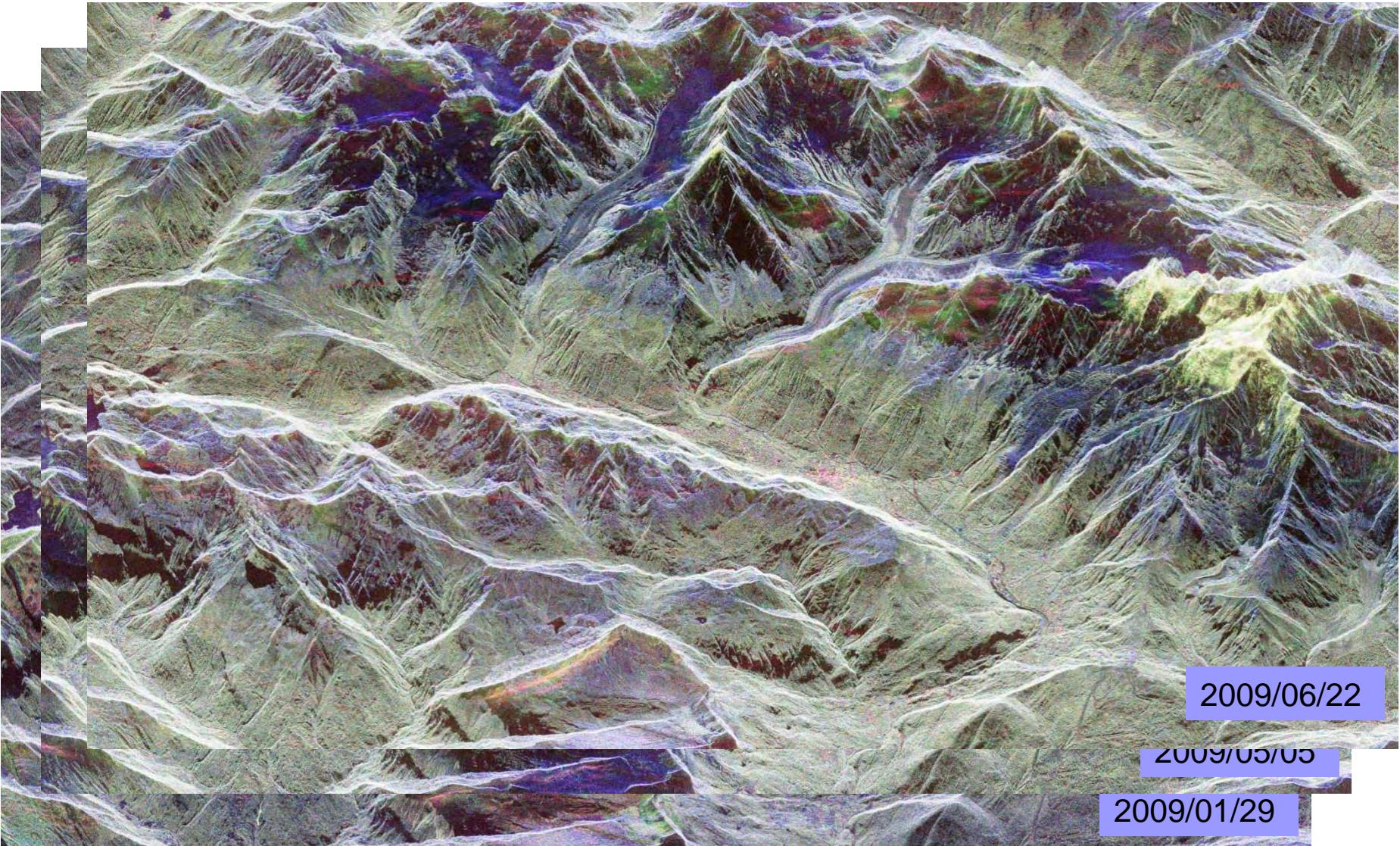
- Séisme du Pakistan (Oct 2005)
 - 1 scène “SLC” = 4 Giga (100000 lignes x 5000 colonnes)
 - ~20 scènes asc. / descendantes → ~10 couples avant/après séisme



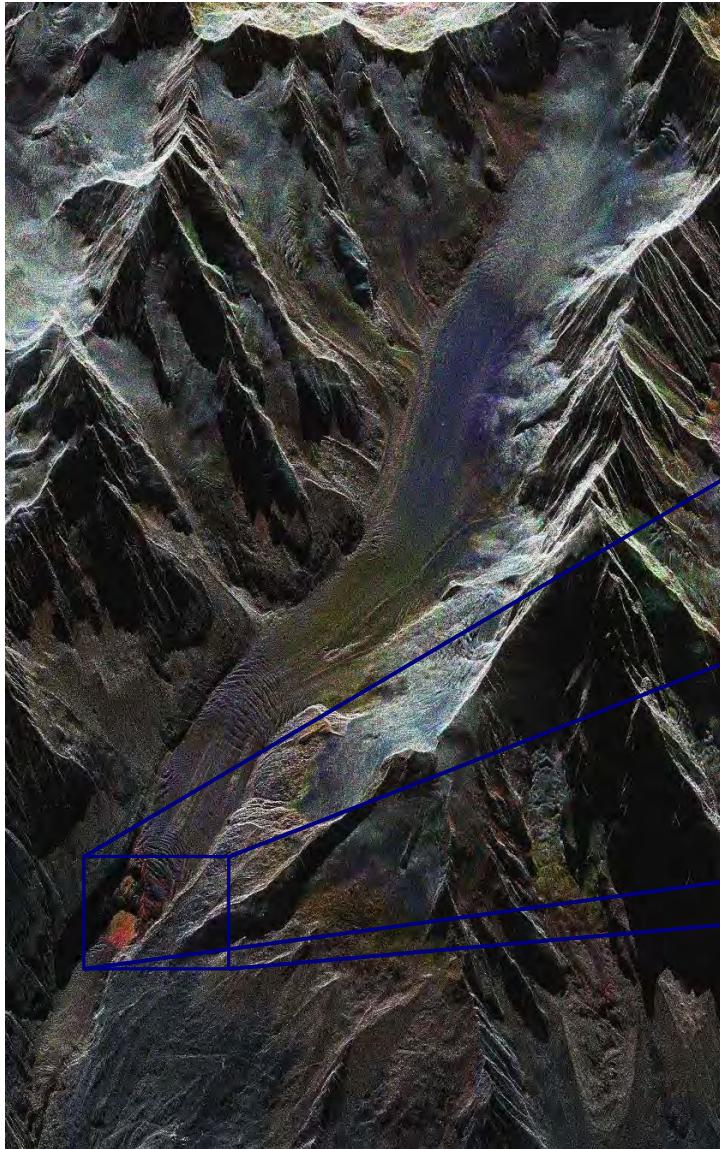
- Subsidence de Mexico
 - 38 images ENVISAT,
 - 71 interférogrammes « petite base »

R:HH-VV,
V:2HV,
B:HH+VV

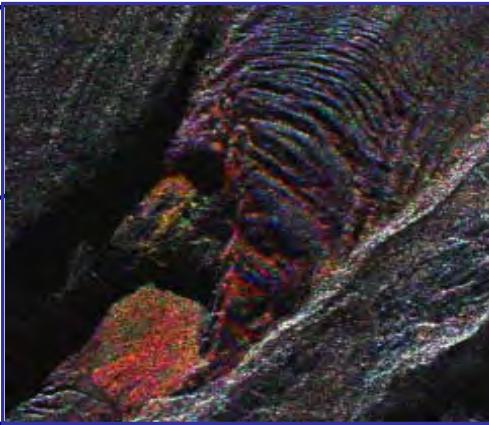
Images SAR polarimétriques – Radarsat-2



Images SAR haute résolution - TerraSAR-X (~2m)

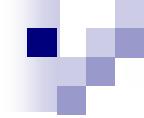


*Glacier d'Argentière,
composition RVB :
29/09/08,
10/10/08,
21/10/08*



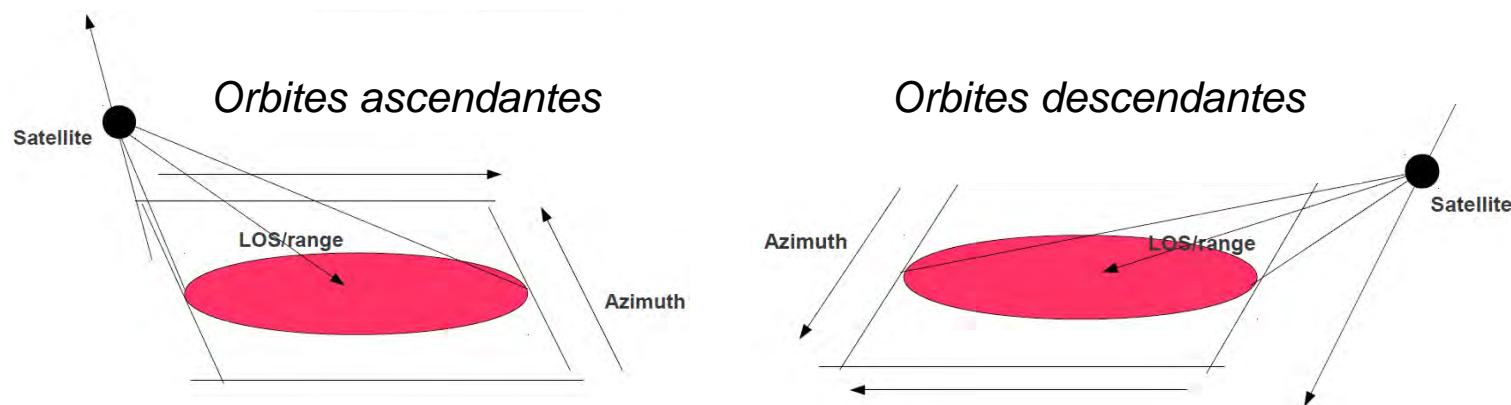
Plan

1. Pourquoi EFIDIR ?
 - Un constat
 - Données utilisées
2. EFIDIR : activité scientifique et technique
 - Sous-Projets, Work-Packages
 - Exemples de résultats
 - Actions transverses
 - Diffusion
3. Fouille de données spatio-temporelle : application à la caractérisation du déplacement
 - What is data mining?
 - Problem statement & GFS-patterns
 - Results
4. Perspectives



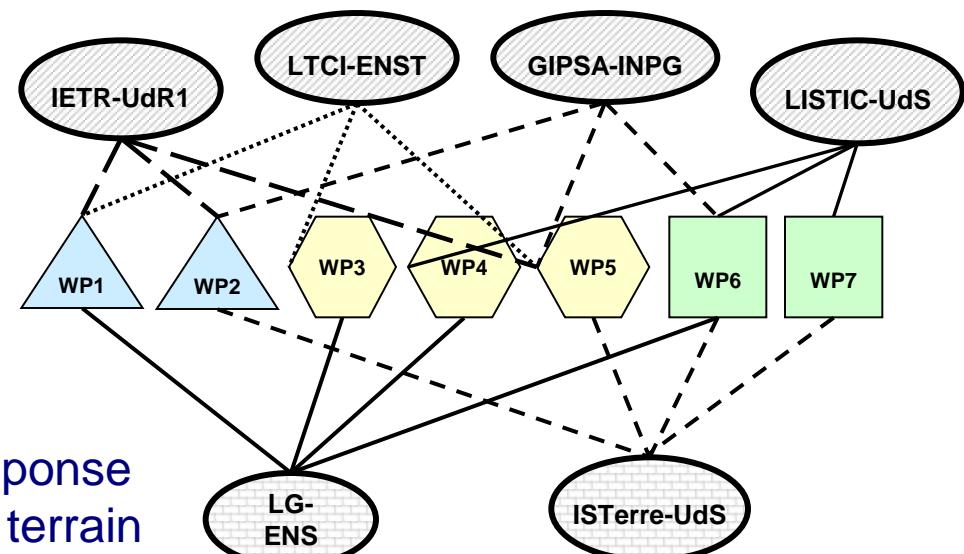
Données SAR → double mesure du déplacement

- Position des cibles dans les images :
 - Déplacement par « offset tracking » : corrélation, max. de similarité...
 - Précision : ~1/10 résolution : ENVISAT => ~1m, TerraSAR => ~10cm
 - Déplacement 2D (LOS et Azimut)
 - Couples ascendants et descendants => 4 projections du déplacement 3D
- Différence de phase après recalage (interférométrie) :
 - Phase ~ temps de trajet ~ distance radar-cible, modulo 2π
 - Précision : fraction de la longueur d'onde (5.6cm en bande C) => cm/mm
 - Déplacement 1D (LOS)





3 Sous-Projets 7 Work-Packages



- SP1: Extraction d'informations : de la réponse électromagnétique aux caractéristiques terrain
 - WP1: Caractérisation des diffuseurs stables
 - WP2: Caractérisation des milieux diffusants
- SP2: Extraction d'informations : des données SAR à la mesure de déplacement
 - WP3: Des rétrodiffuseurs cohérents aux séries temporelles de déplacements
 - WP4: Développement d'une chaîne de traitement NSBAS
 - WP5: Mesure de grands déplacements à partir de données radar multivariées
- SP3: Fusion d'informations : des mesures de déplacement à la connaissance géophysique
 - WP6: Détection des perturbations atmosphériques
 - WP7: Fusion des mesures de déplacement et estimation des paramètres géophysiques



SP1/WP2 : Caractérisation des milieux diffusants

- Développement d'un Ground Based SAR pour acquérir des images 3D très haute résolution (THR)

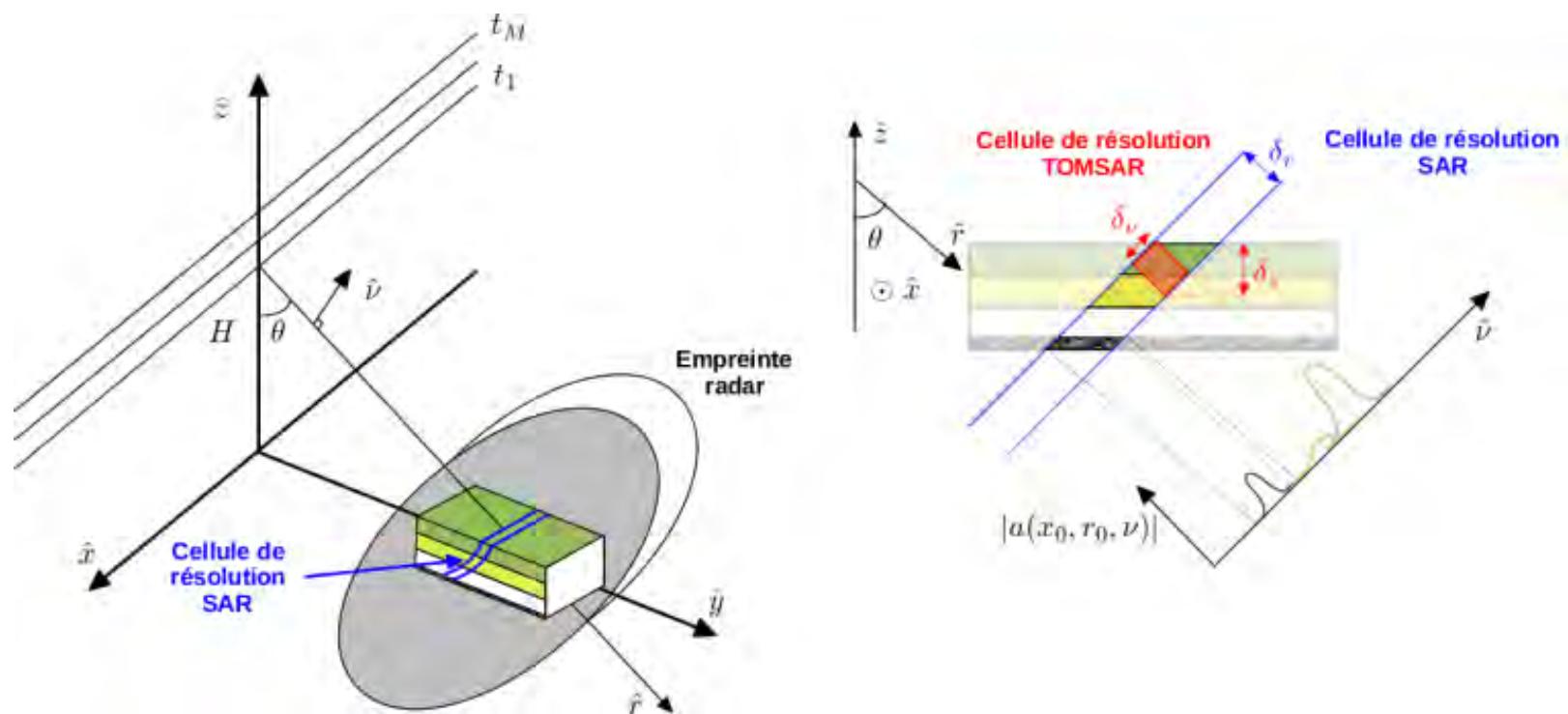


*GB-SAR développé par l'IETR,
expérimentations au CEN de
Météo France, déc. 2010*



SP1/WP2 : Caractérisation des milieux diffusants

- Développement des méthodes de synthèse 3D pour analyser la réponse volumique de milieux tels que la neige

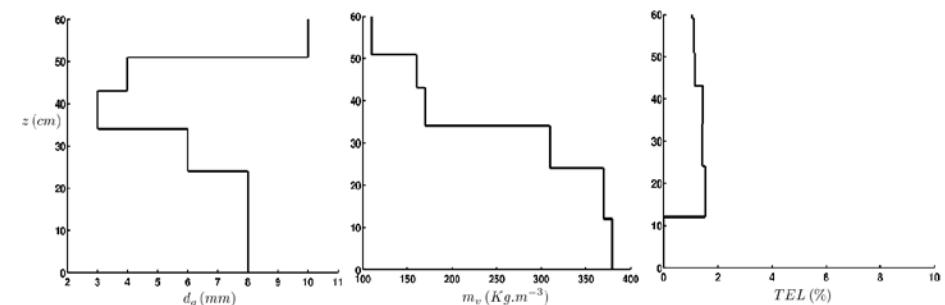
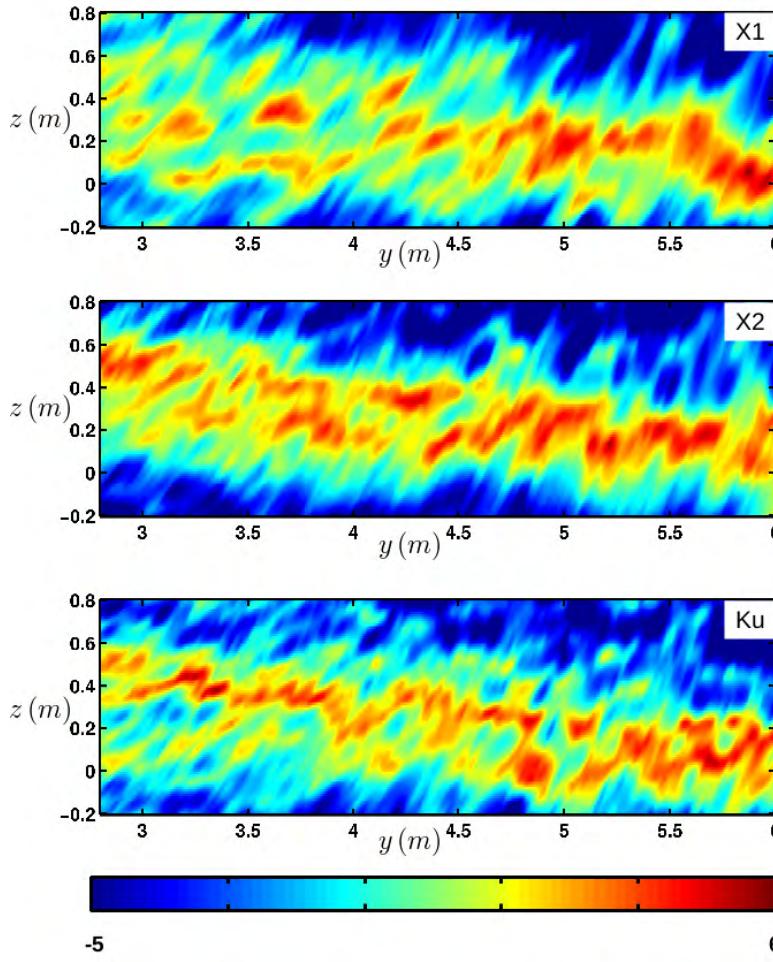


Configuration géométrique d'une mesure MB-inSAR (gauche),
cellules de résolution SAR et TOMSAR (droite)



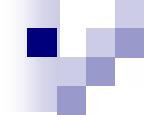
SP1/WP2 : Caractérisation des milieux diffusants

■ Premiers résultats



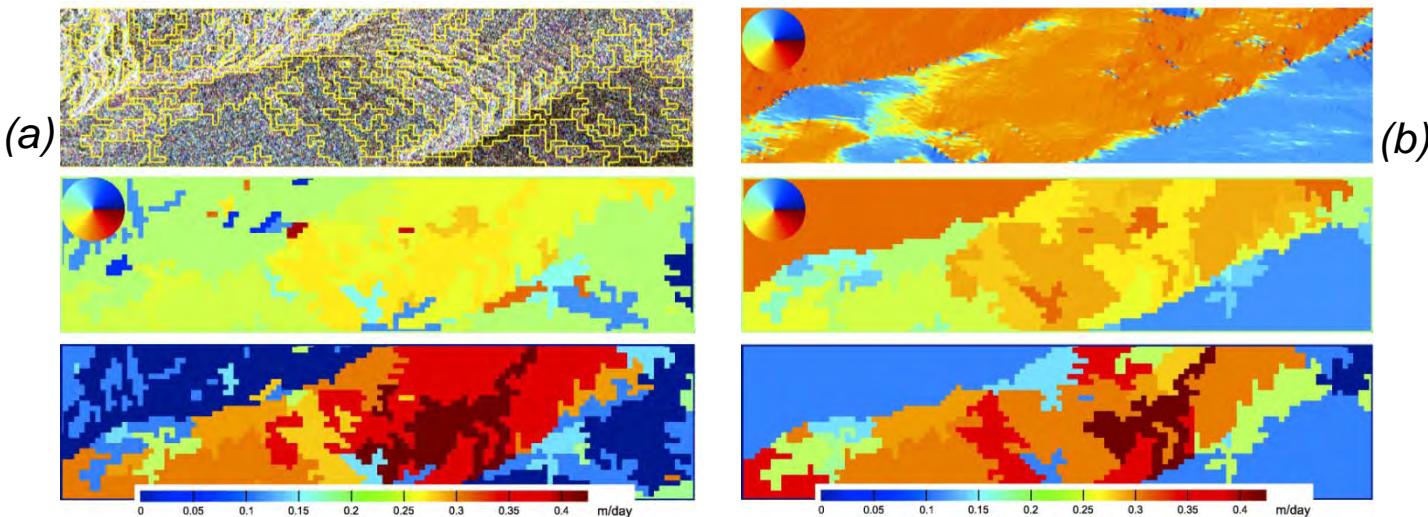
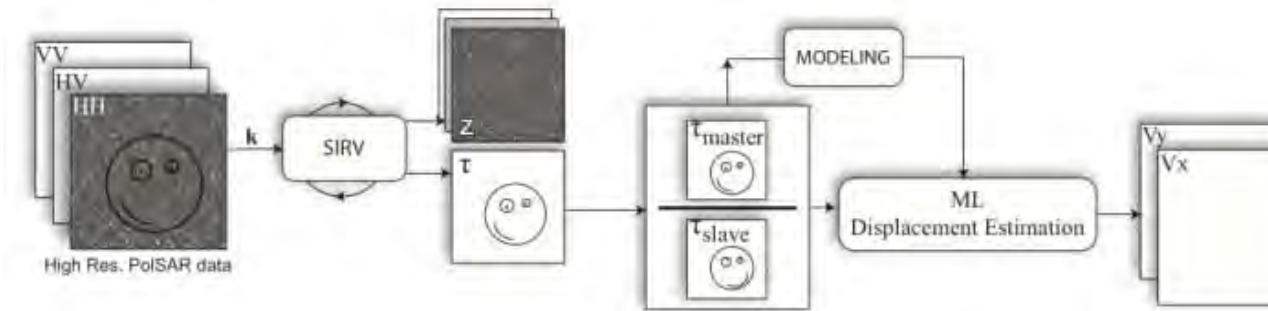
*Paramètres moyens de la structure stratigraphique du couvert.
Diamètre des grains de neige (gauche),
masse volumique (centre),
teneur en eau liquide (droite)*

*Tomogrammes en réflectivité relative,
pour une même coordonnée en azimut x ,
en bandes X (configurations X1, X2) et Ku*

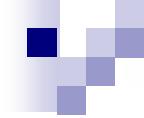


SP2/WP5 : Mesure de grands déplacements

■ Données SAR polarimétriques

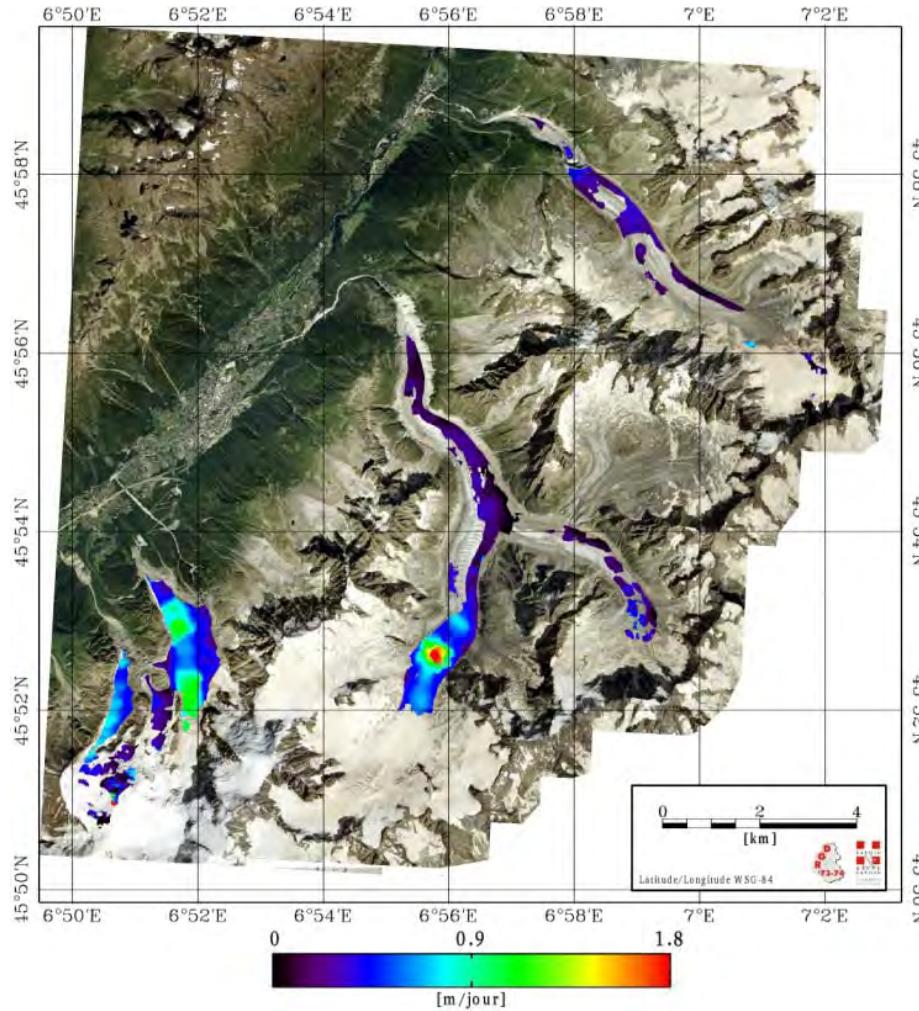


Déplacement 2D d'une partie du glacier d'Argentière, TerraSAR-X dual pol, 06/01-08/02/2009
Utilisation d'une segmentation (a) et introduction d'une contrainte d'écoulement (b)



SP2/WP5 : Mesure de grands déplacements

■ Données SAR haute résolution (HR)

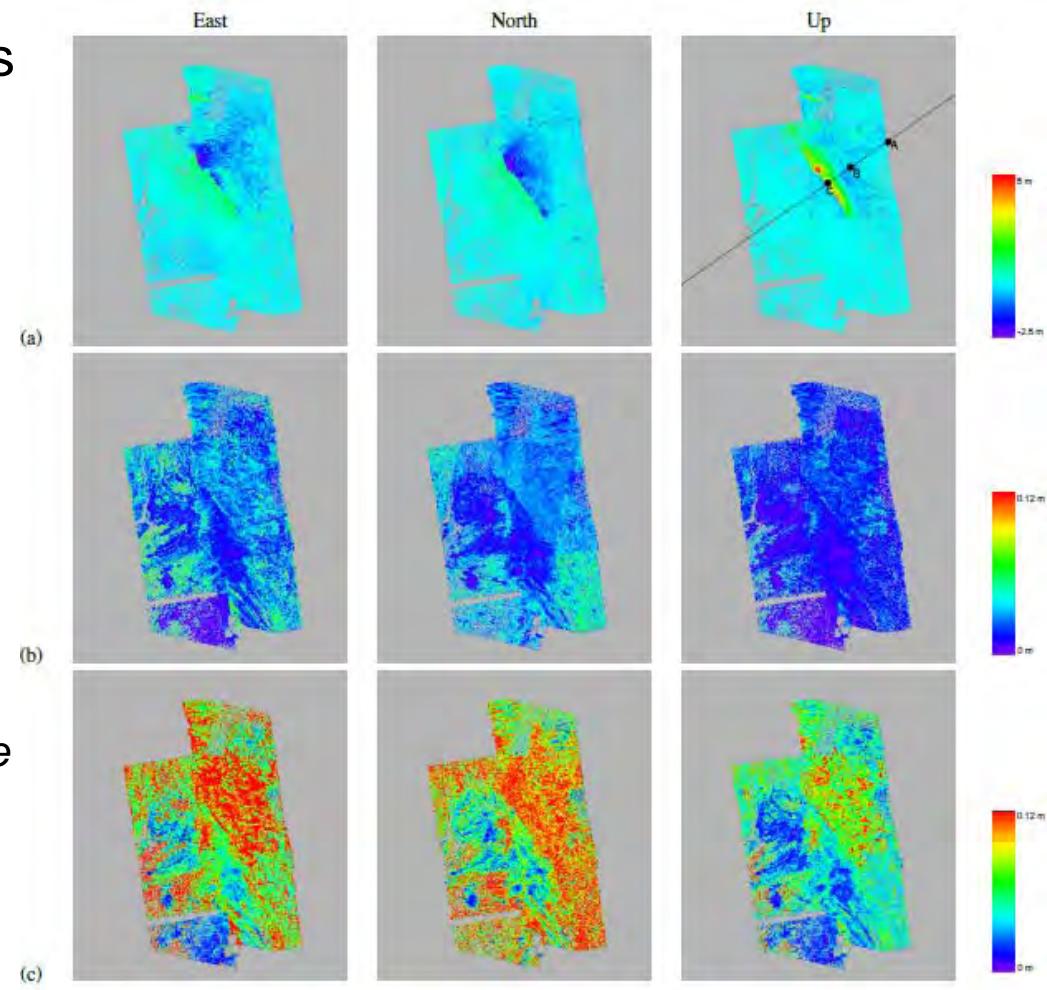


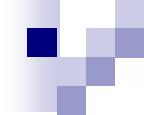
*Glaciers de la vallée de Chamonix
Déplacement 3D (module en m/jour)
Données TerraSAR-X HR (pol. HH)
descendantes du 14 et 25/08/2009
et ascendantes du 16 et 27/08/2009*



SP3/WP7 : Fusion de mesures de déplacement et inversion de modèles physiques

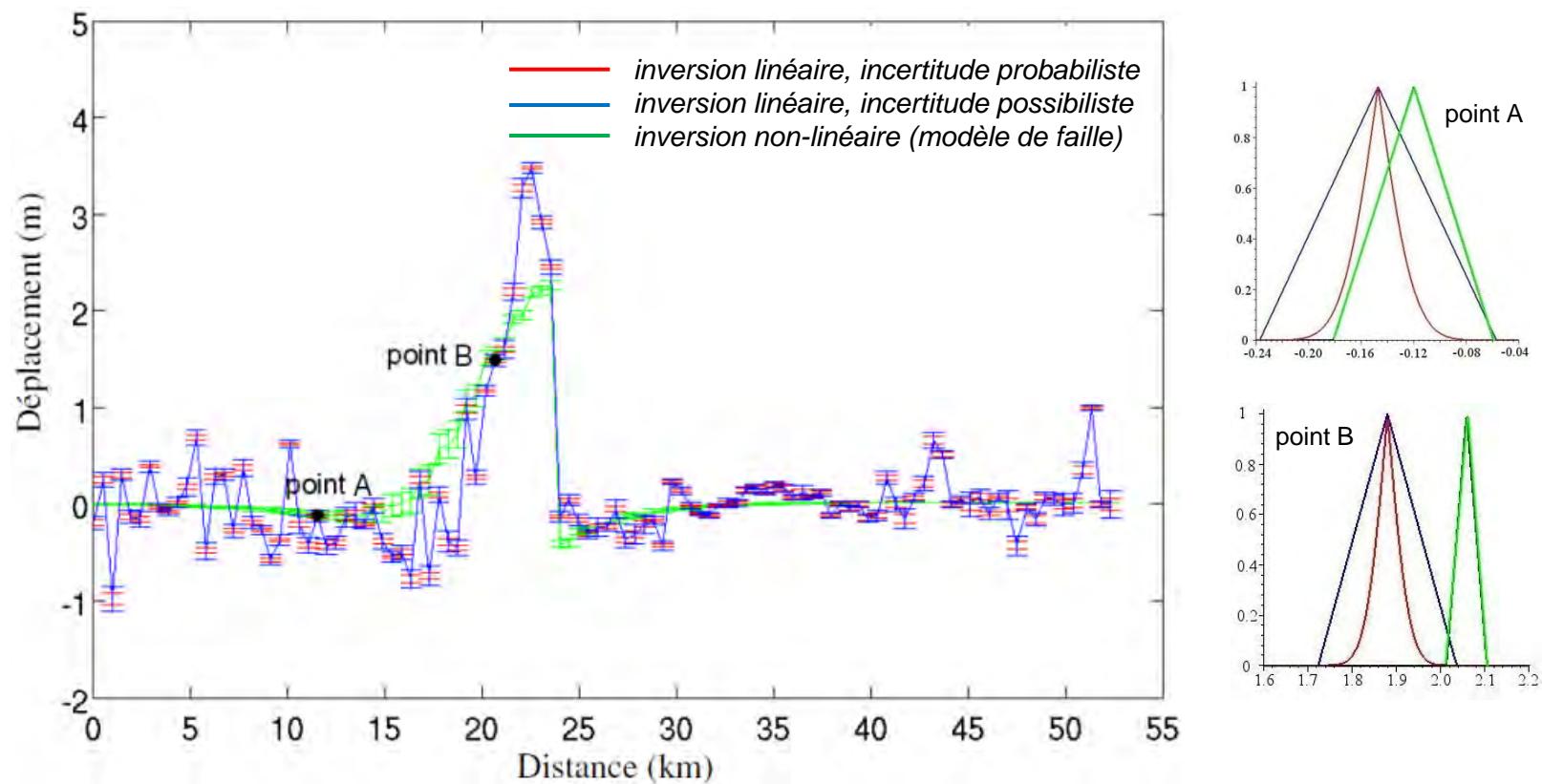
- Différentes stratégies de fusion
- Propagation des incertitudes par approche probabiliste ou possibiliste





SP3/WP7 : Fusion de mesures de déplacement et inversion de modèles physiques

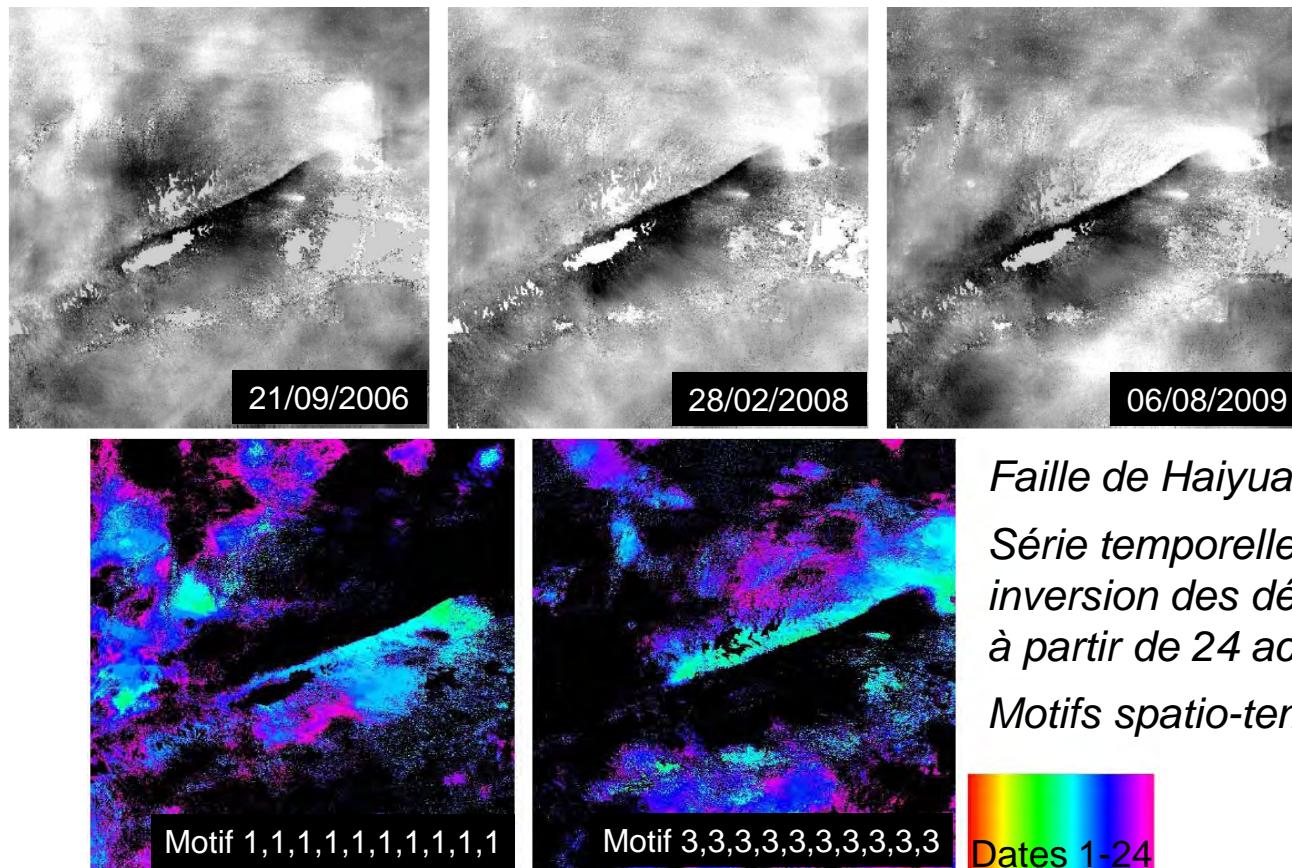
- Comparaison avec les déplacements de surface issus d'inversions d'un modèle de faille



Composante verticale du déplacement de surface et incertitudes associées

SP2/WP4 : mesure de faibles déformations + SP3/WP6 : perturbations atmosphériques

- Traitement de séries temporelles
 - Recherche de motifs spatio-temporels par fouille de données



Actions transverses

- Groupe de travail « logiciel »
 - NSBAS : chaîne de traitement adossée sur le logiciel ROI-PAC
➔ séries temporelles, effets atmosphériques...
 - EFIDIR Tools : bibliothèque/exécutable C/C++, licence GPL
➔ données SAR nouvelle génération, fusion, data mining
 - Xima : IHM analyse/visualisation couplée (X11 → 64 bit)
- Groupe de travail « base de données »
 - Serveur + baie 12 Tera
 - Sites tests : Cachemire, Mont-Blanc, Grenoble...
➔ Données et résultats expérimentaux
- Groupe de travail « site web » : <http://www.efidir.fr>
 - Publications, livrables...
 - Distribution des EFIDIR Tools
 - Evènements : réunions, séminaires, école de printemps...

Diffusion

■ Publications (à T0+42)

		Multipartenaires	Monopartenaires
International	Revues	7	7
	Conférences	44	35
France	Revues	1	
	Conférences	21	10

■ Atelier d'Aussois (1-3 avril 2010) :

- 25 « EFIDIR » + 25 extérieurs invités
- Echanges scientifiques, animation de la communauté SAR/Géophysique

■ Ecole de printemps EFIDIR (1-6 mai 2011) :

- Ecole de Physique des Houches, 70 participants de 22 nationalités
- Cours, travaux pratiques, terrain...

Ecole de printemps EFIDIR – activité terrain



Expérimentation d'un réseau de coins réflecteurs, image TerraSAR-X 05/05/2011

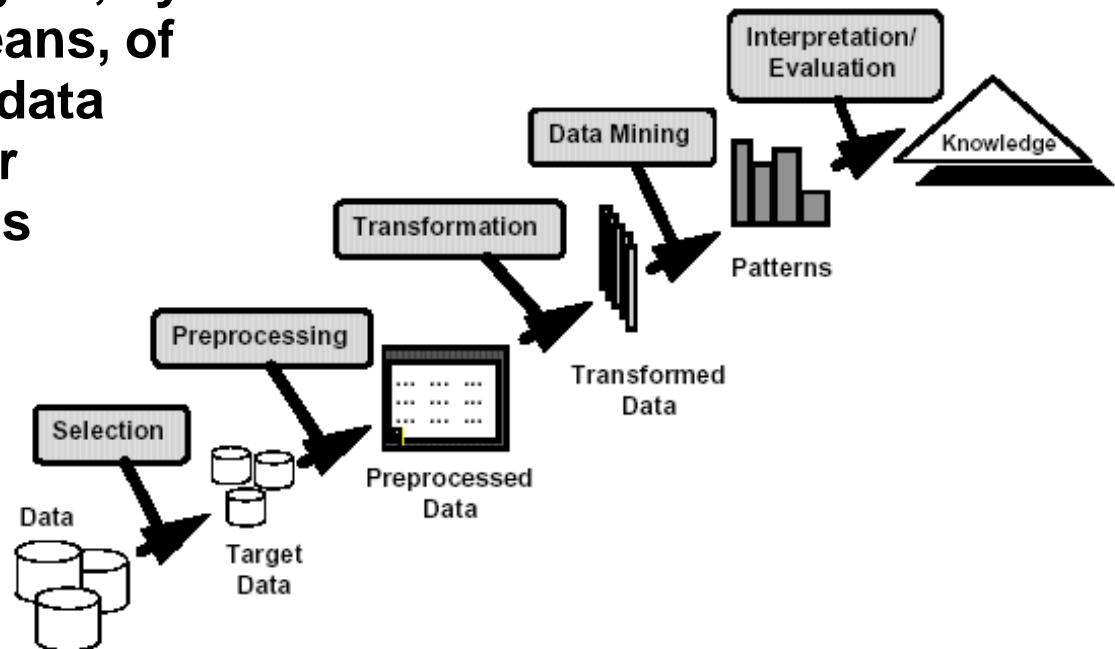
Plan

1. Pourquoi EFIDIR ?
 - Un constat
 - Données utilisées
2. EFIDIR : activité scientifique et technique
 - Sous-Projets, Work-Packages
 - Exemples de résultats
 - Actions transverses
 - Diffusion
3. Fouille de données spatio-temporelle : application à la caractérisation du déplacement
 - What is data mining?
 - Problem statement & GFS-patterns
 - Results
4. Perspectives

What is Data Mining?

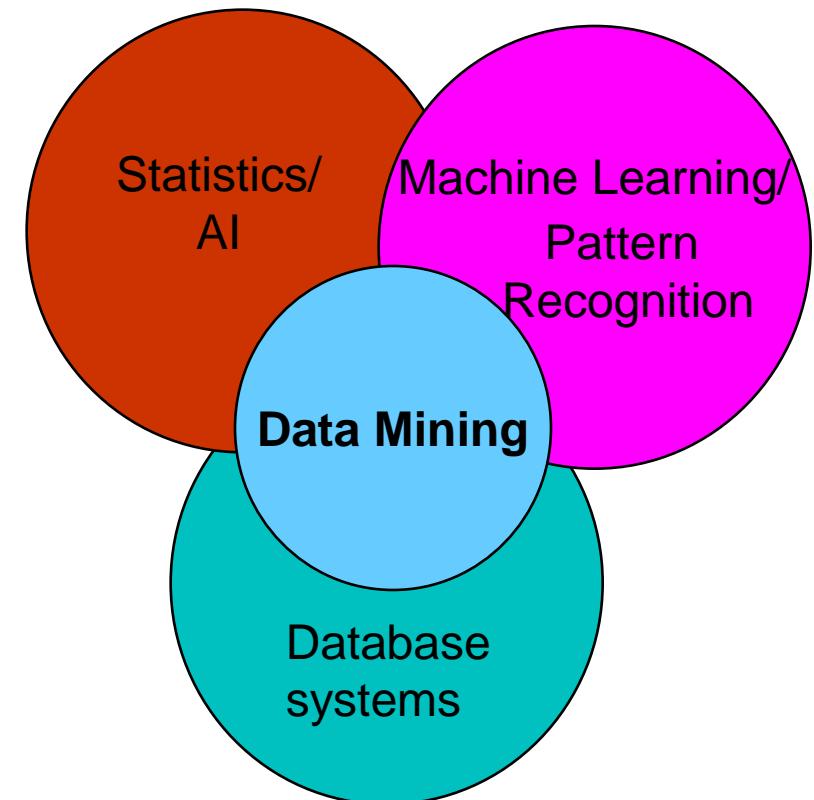
□ Many Definitions

- Non-trivial extraction of implicit, previously unknown and potentially useful information from data
- Exploration & analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns



Origins of Data Mining

- Draws ideas from machine learning/AI, pattern recognition, statistics, and database systems
- Traditional Techniques
may be unsuitable due to
 - Enormity of data
 - High dimensionality
of data
 - Heterogeneous,
distributed nature
of data



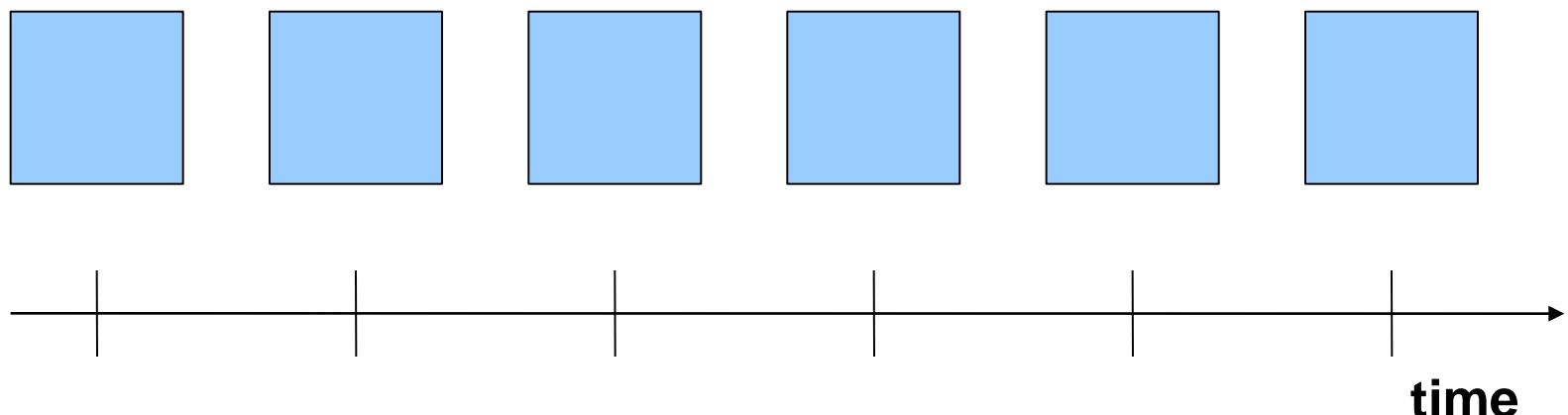


Data Mining Tasks...

- Classification
- Clustering
- Association Rule Discovery
- Sequential Pattern Discovery
- Regression
- Deviation Detection

Problem statement: objectives

- To describe a SITS temporally and spatially.
- Unsupervised techniques.

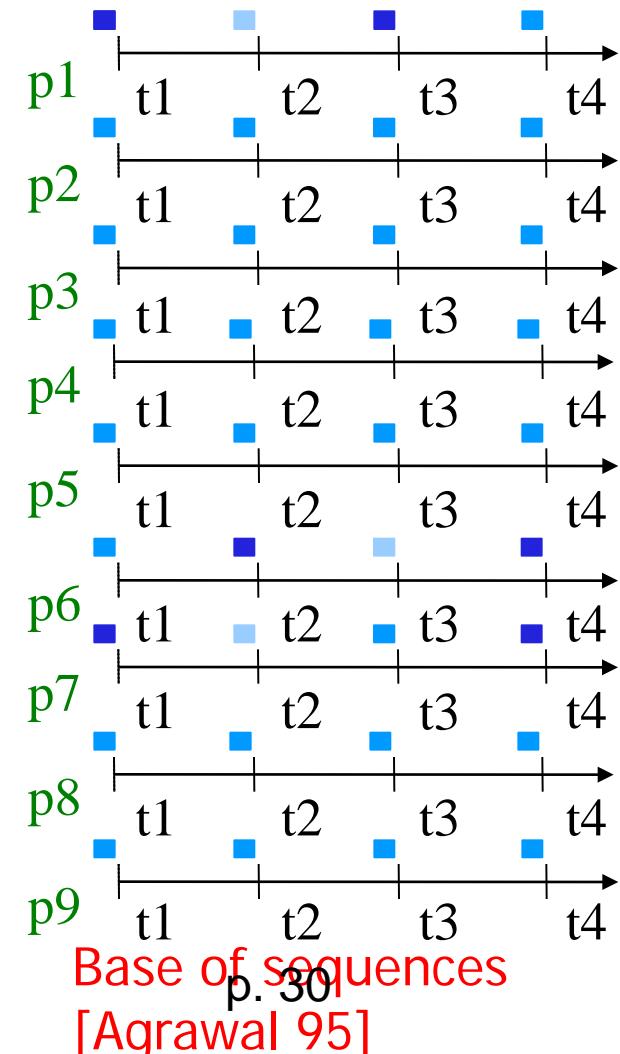
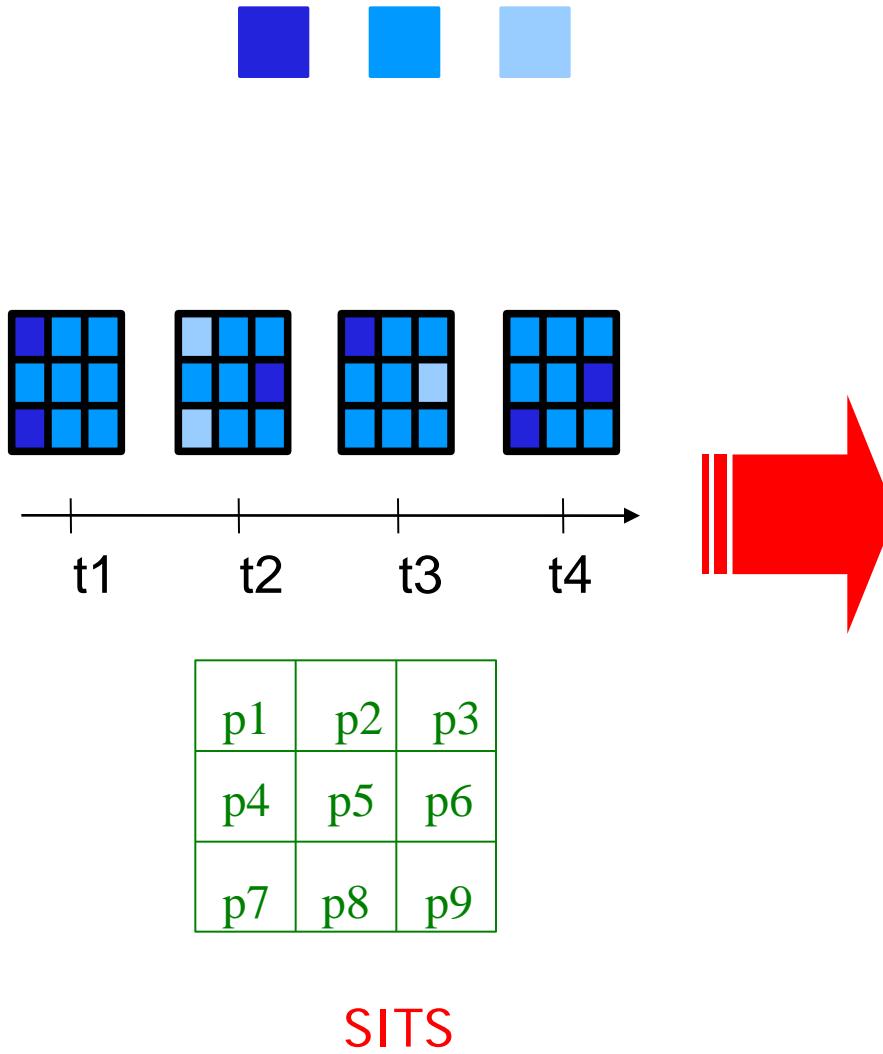


Problem statement: approach

- Assumption: **pixels** relating to a same evolution or sub-evolution should be **numerous** and should be **connected**.
- Constraints
 - **All possible evolutions and sub-evolutions** must be taken into account for all pixels: combinatorial explosion + millions of pixel evolutions to be mined,
 - **Different occurrence dates** of a same sub-evolution should be considered to be robust to noisy acquisitions: clustering methods are discarded.

SITS and base of sequences

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

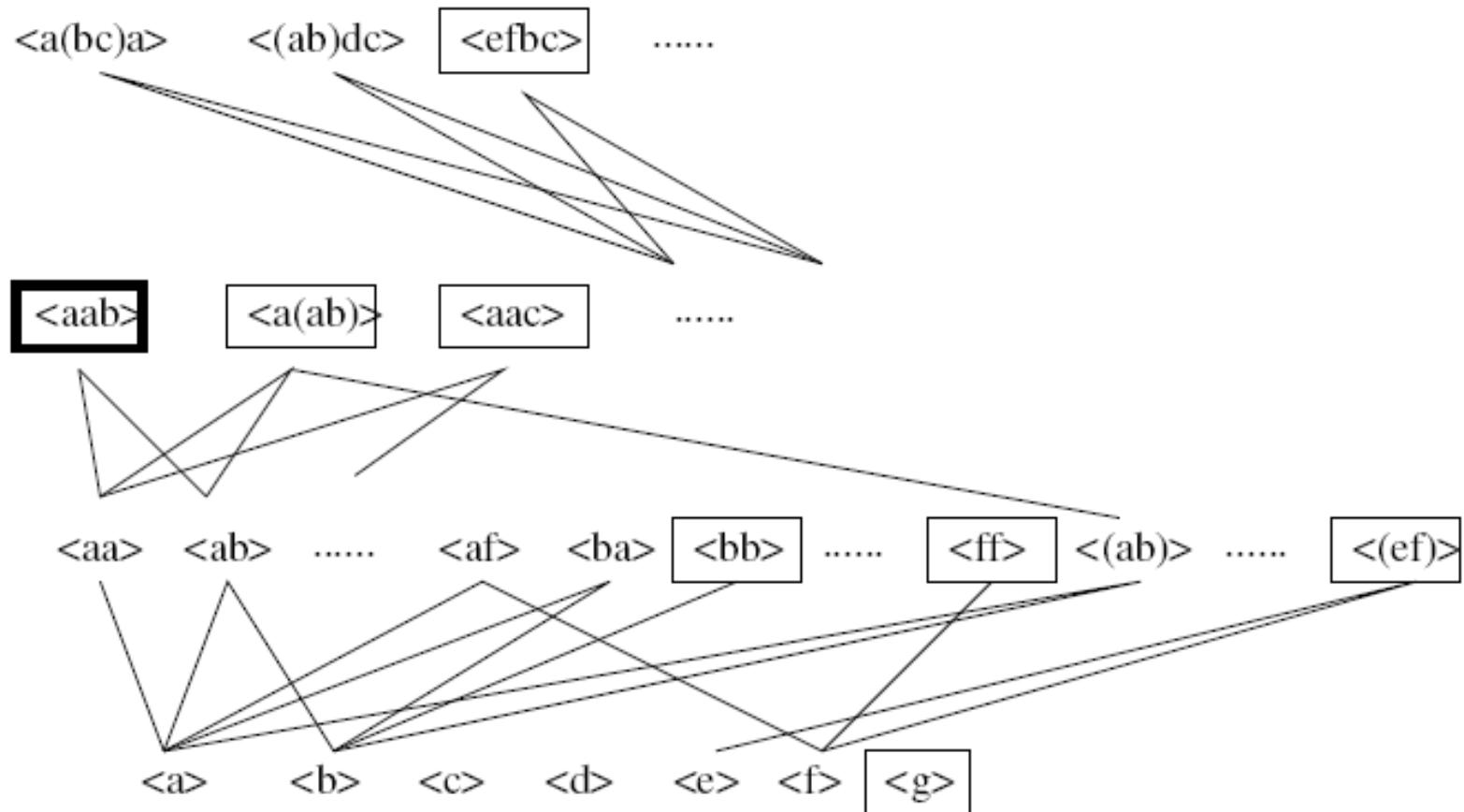


Frequent sequential patterns

- Extracted model: sequential patterns

- Measure: **support**, i.e., the number of pixels supporting an evolution/sub-evolution. It can be used to extract evolutions/sub-evolutions shared by at least σ pixels, the **frequent sequential patterns**.
- σ , termed as the **minimum support**, is used to set an active constraint to prune the search space (anti-monotonicity property).

Pruning the search space using σ



Incorporating spatiality: the GFS-patterns

- A new measure is proposed: the **average connectivity** of evolution α , i.e. the average number of pixels covered by α in the 8-neighborhood of the pixels covered by α .

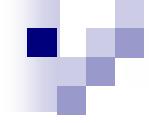
$$LC((x, y), \alpha) = [\sum_{i=-1}^{i=1} \sum_{j=-1}^{j=1} occ((x + i, y + j), \alpha)] - 1$$

$$AC(\alpha) = \frac{\sum_{(x, y) \in cov(\alpha)} LC((x, y), \alpha)}{|cov(\alpha)|}$$

Links(α)

support(α)

- It is used to extract frequent sequential patterns covering pixels that are sufficiently connected: a **minimum average connectivity** is set, K . They are termed as **Grouped Frequent Sequential patterns (GFS-patterns)**.
- This constraint is not anti-monotone but ...



Pushing the average connectivity constraint

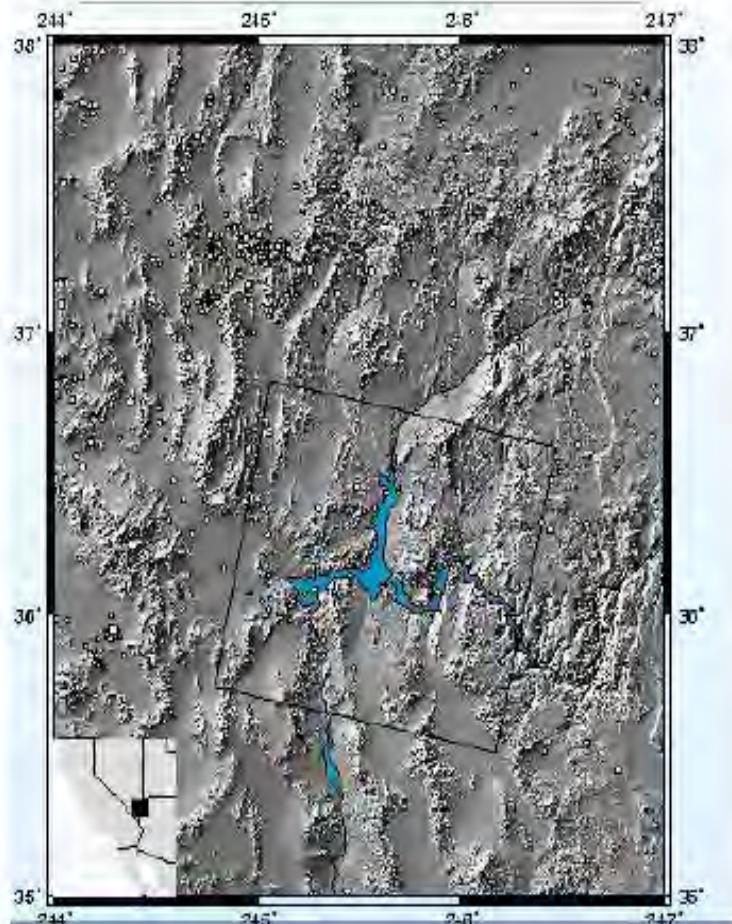
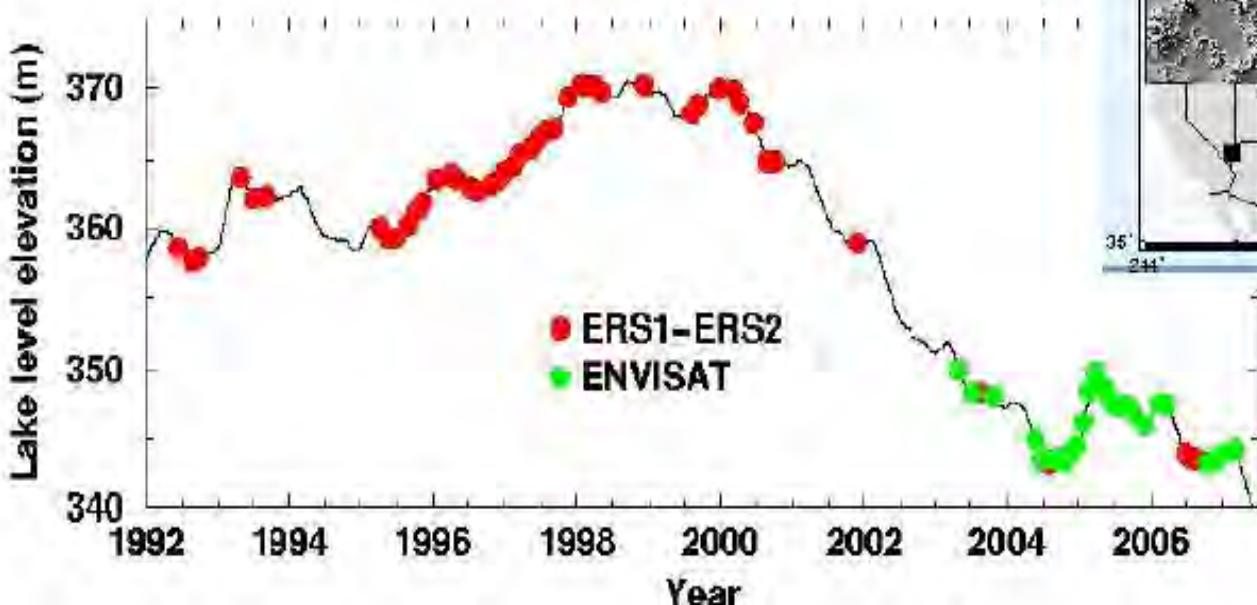
- $AC(\alpha) \leq \text{Links}(\alpha)/\sigma$
- Thus if $\text{Links}(\alpha)/\sigma < k$, then $AC(\alpha) < k$.
- Asking for patterns having $\text{Links}(\alpha)/\sigma \geq k$ is an anti-monotone constraint.
- Strategy: we explore the search space using $\text{Links}(\alpha)/\sigma \geq k$, and we select patterns having $AC \geq k$.
- Speed-up: between 7% and 40%.
- Complementing to support pruning.

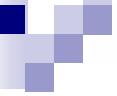
Results

Crustal deformation monitoring
by differential SAR
interferometry

Available data

- Lake Mead.
- About 56 ERS/ENVISAT acquisitions (1996-2008).
- Spatial resolution (after processing): 130m x 130m
- Interferograms: image of displacements/master image.

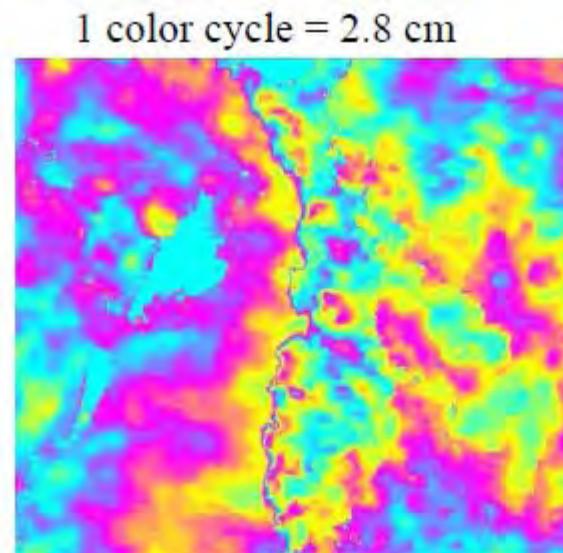




Slow and long wavelength ground motion

Problem: the signal is affected by spatial and temporal random **atmospheric perturbations** ...

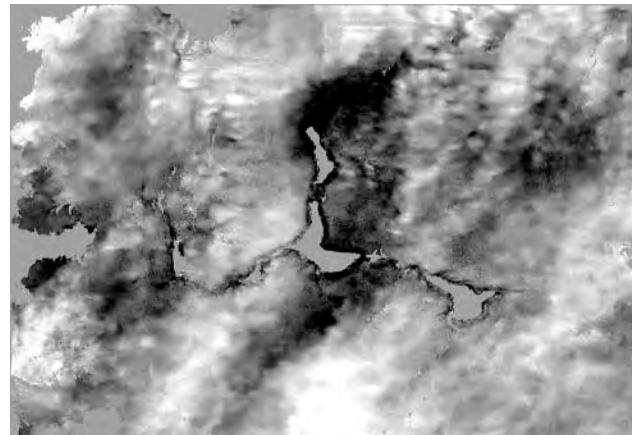
$$\Delta\Phi = \frac{4\pi}{\lambda} (B \sin(\alpha - \theta(z)) + \Delta\rho) + \boxed{\Delta\Phi_{atm}} + \dots$$



Ripples, bubbles,
patches, fronts,

Data selection and preprocessing

- 20 images
- The stratified atmosphere is corrected but the turbulent atmosphere is still present.
- Displacements/master image = f ("real" displacements + atmosphere of the master image + atmosphere of the acquisition).
- Pixel values (phase differences) are quantized into 3 intervals.

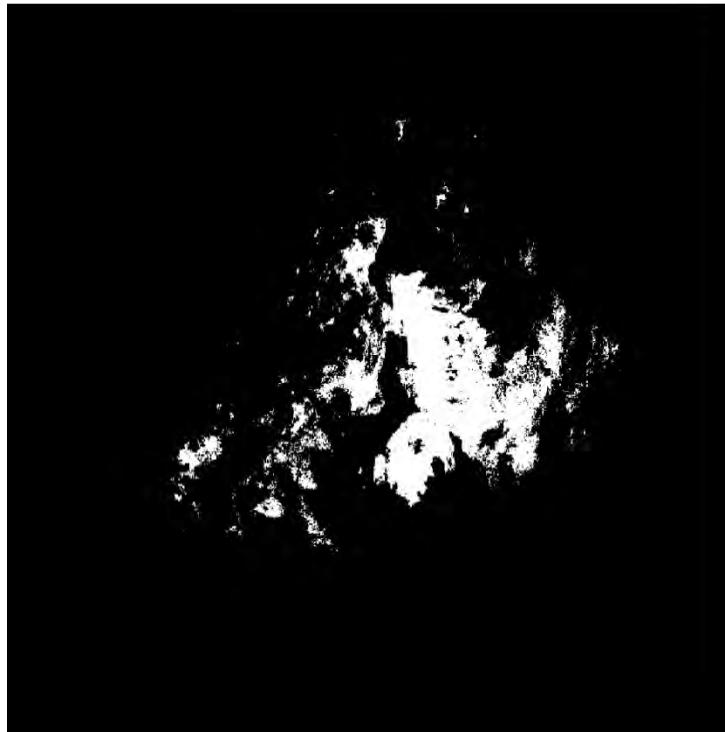


Quantitative results

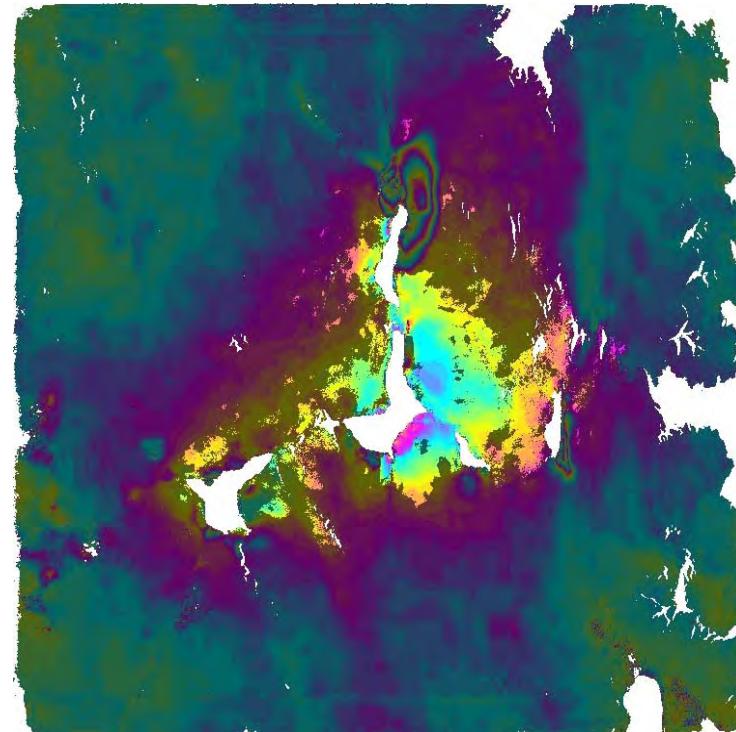
Standard PC: Intel Core 2 @ 3GHz, 4 GB RAM,
linux kernel 2.6.3

- $\sigma = 2\%$
- $k = 6$
- Execution times: 4900 s
- Number of GFS-patterns: 10073
- Longest patterns are selected (5 patterns)

Results ($\sigma=2\%$, $\kappa=6$)



A

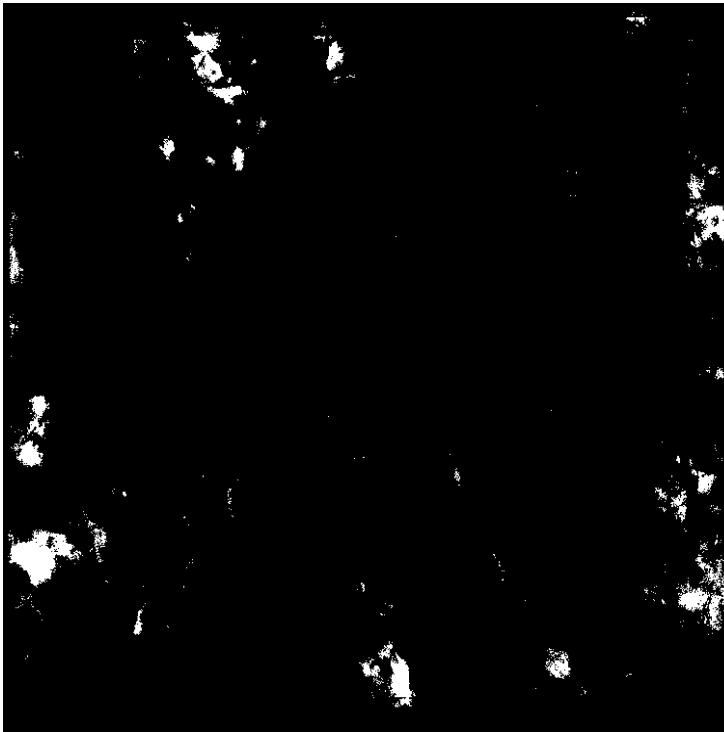


B

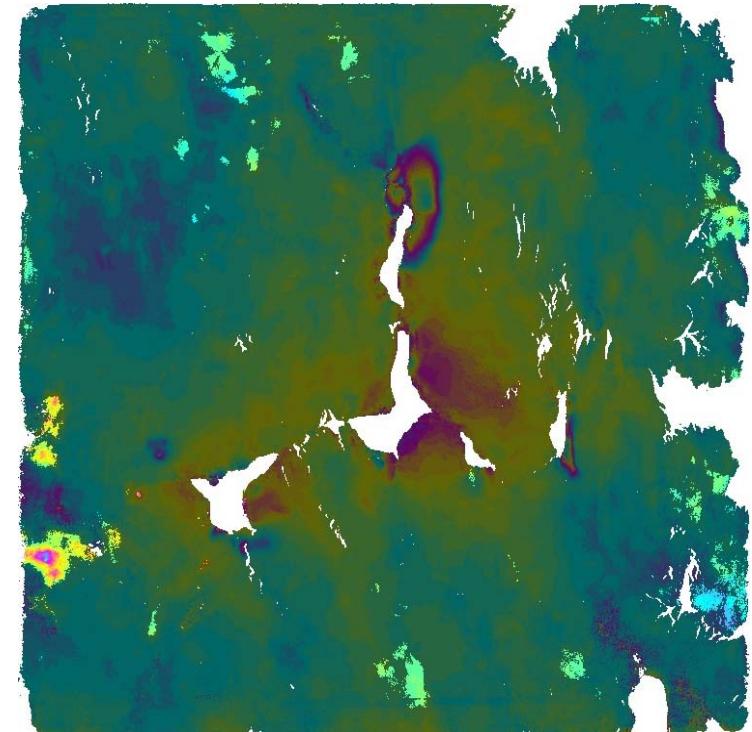
A: the white pixels correspond to the 4 longest evolution (out of 5) stating that these pixels stick to high positive phase difference values before switching to low negative ones (3>3 ... 1>1).

B: these pixels correspond (clear ones) to strong positive regression coefficients between the phase differences and the water level fluctuations. The atmospheric patterns are thus not considered and the displacements patterns are focused on.

Results ($\sigma=2\%$, $\kappa=6$)



A



B

A: the white pixels correspond to a single long evolution stating that the pixel values stick to low negative phase differences ($1>1\dots>1$).

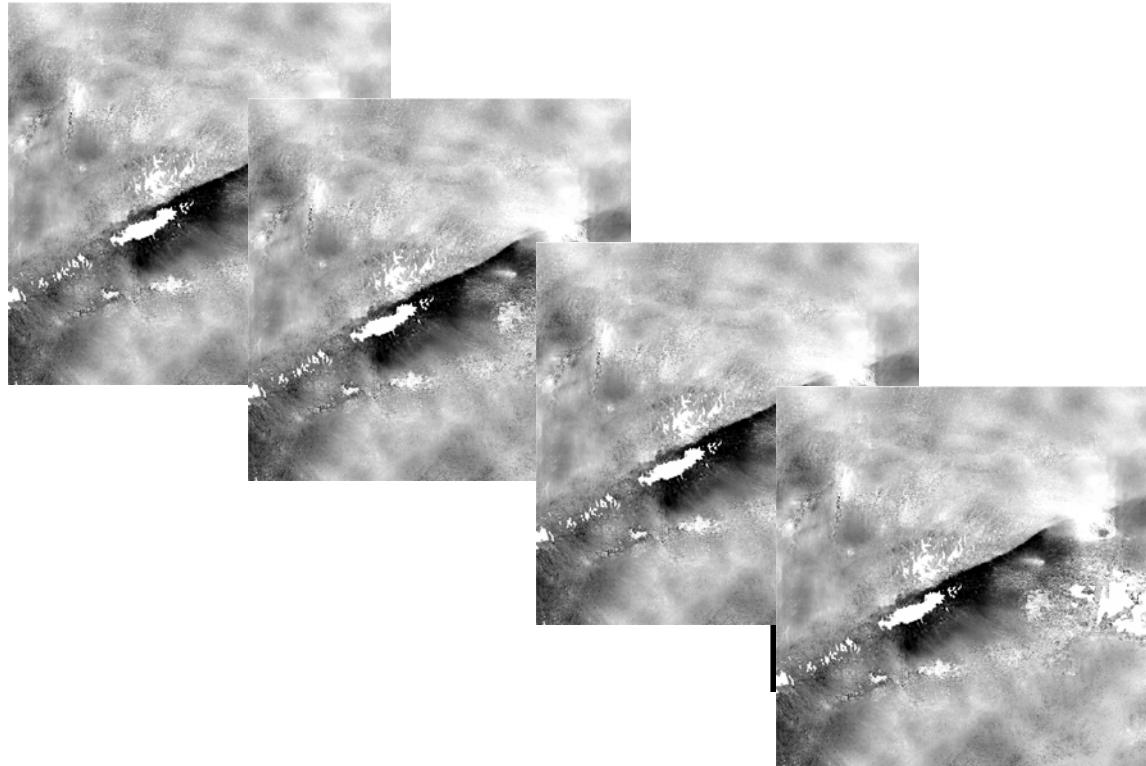
B: these pixels (clear ones) correspond to a negative average speed (uplift), in particular for the Las Vegas area. This can be explained by a decreasing water pumping from the Las Vegas aquifers.



Results

Monitoring the Haiyuan fault

Data



- ENVISAT InSAR time series covering the Haiyuan fault (north-eastern boundary of the Tibetan plateau) over the 2006-2009 period.
- Hit by several major earthquakes at the beginning of the 20th century.
- 24 raw SAR images.

Processing chain

- SAR image coregistration to a single master and interferogram generation.
- Joint inversion of residual orbital and atmospheric delays.

[Cavalié08] Measurement of interseismic strain across the Haiyuan fault by InSar – EPSL

- Validation of atmospheric corrections using the ERA40 global atmospheric model (ECMWF) and correction of each unwrapped interferogram

[Doin09] Corrections of stratified tropospheric delays in SAR interferometry: validation with global atmospheric models - AG

- Inversion of the interferogram series on each track to obtain the increments of smooth LOS radar phase evolution
[Jolivet11] Shallow creep on the Haiyuan Fault, Gansu, China, revealed by SAR Interferometry

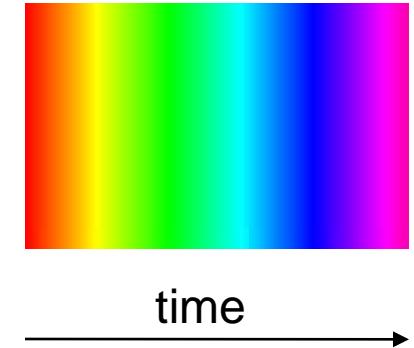
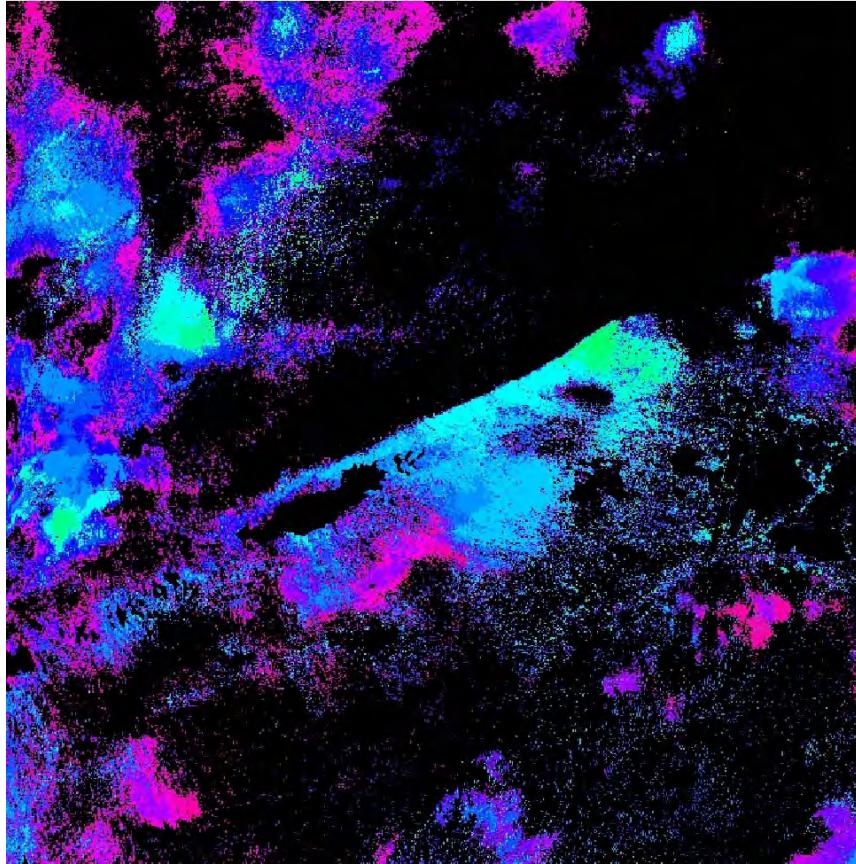
→ A 24 image SITS (701x701 px, resolution 80mx80m) of phase increments.
Contain ground deformation and atmospheric turbulences.

Preliminary results

Standard laptop: 4 GB RAM, Intel Core i5 M520 @ 2.4 GHz, linux kernel 2.6.37

- . $\sigma = 100000$
- . $k = 6$
- . 3 symbols for describing pixel values (/the 33rd and the 66th centiles)
- . Execution times: 1 hour and 8 minutes
- . Memory consumption: 2 GB
- . Number of GFS-patterns: 3413
- . Number of longest GFS-patterns: 11
- . Presented GFS-patterns:
 - 1->1->1->1->1->1->1->1->1 (support=159133, AC=6.15)
 - 3->3->3->3->3->3->3->3->3 (support=180275, AC=6.33)

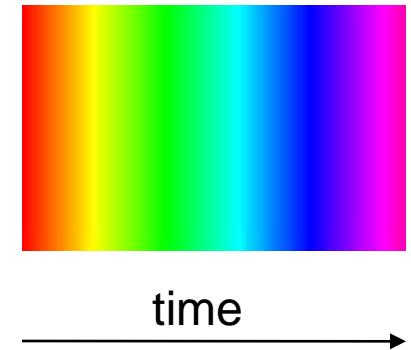
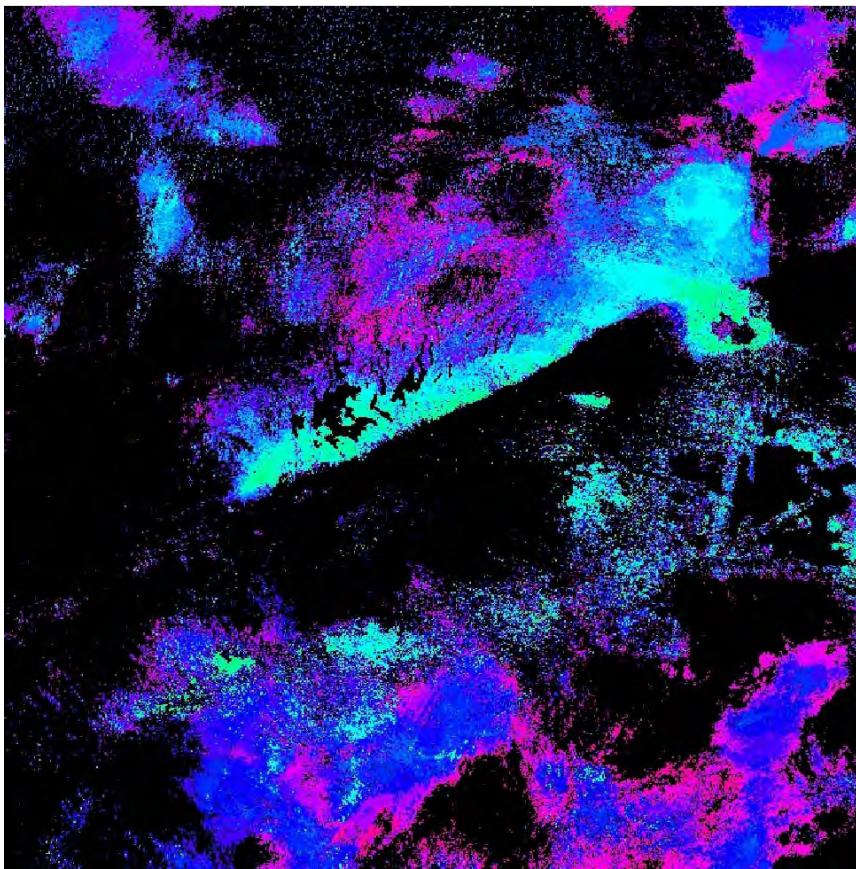
1->1->1->1->1->1->1->1->1



Creep revealed.

Colors show the ending dates of FGS-pattern occurrences.
Not radial w.r.t. to the creeping zone: creep migration.

3->3->3->3->3->3->3->3->3



Creep revealed.
Colors show the ending dates of FGS-pattern occurrences.
Radial w.r.t. to the creeping zone. p. 47



To be continued ...

Financé par
ANR



Perspectives

- Clustering de GFS-patterns
- Comparaison avec des jeux aléatoires de même structure
- SAR 3^{ème} génération :
 - HJ-1C (Chine), bande S, 2012
 - Sentinel 1 (ESA), bande C, 2013 (2nd en 2015), répétitivité :12 -> 6 jours
 - ...
- Nouveaux projets
 - Projet FP7 Marie-Curie « Ground Move »
 - Projet ANR sur les glaciers Himalayens
 - Projets TOSCA (CNES)
 - Proposition d'un GdR « Opti-SAR »
 - ...

References

Main references

A. Julea, N. Méger, P. Bolon, C. Rigotti, M.-P. Doin, C. Lasserre, E. Trouve and V. Lazarescu., Unsupervised Spatiotemporal Mining of Satellite Image Time Series using Grouped Frequent Sequential Patterns, IEEE Transactions on Geoscience and Remote Sensing, vol. 49, issue 4, April 2011, 14 pages.

Méger N., Jolivet R., Lasserre C., Trouvé E., Rigotti C., Lodge F., Doin M-P., Guillaso S., Julea A. and Bolon Ph., Spatio-Temporal Mining of ENVISAT SAR Interferogram Time Series over the Haiyuan Fault in China, In Proc. of the Sixth Int. Workshop on the Analysis of Multitemporal Remote Sensing Images (MULTITEMP'2011), Trento, Italy, July 2011, 4 pages, DOI: 10.1109/Multi-temp.2011.6005067.10

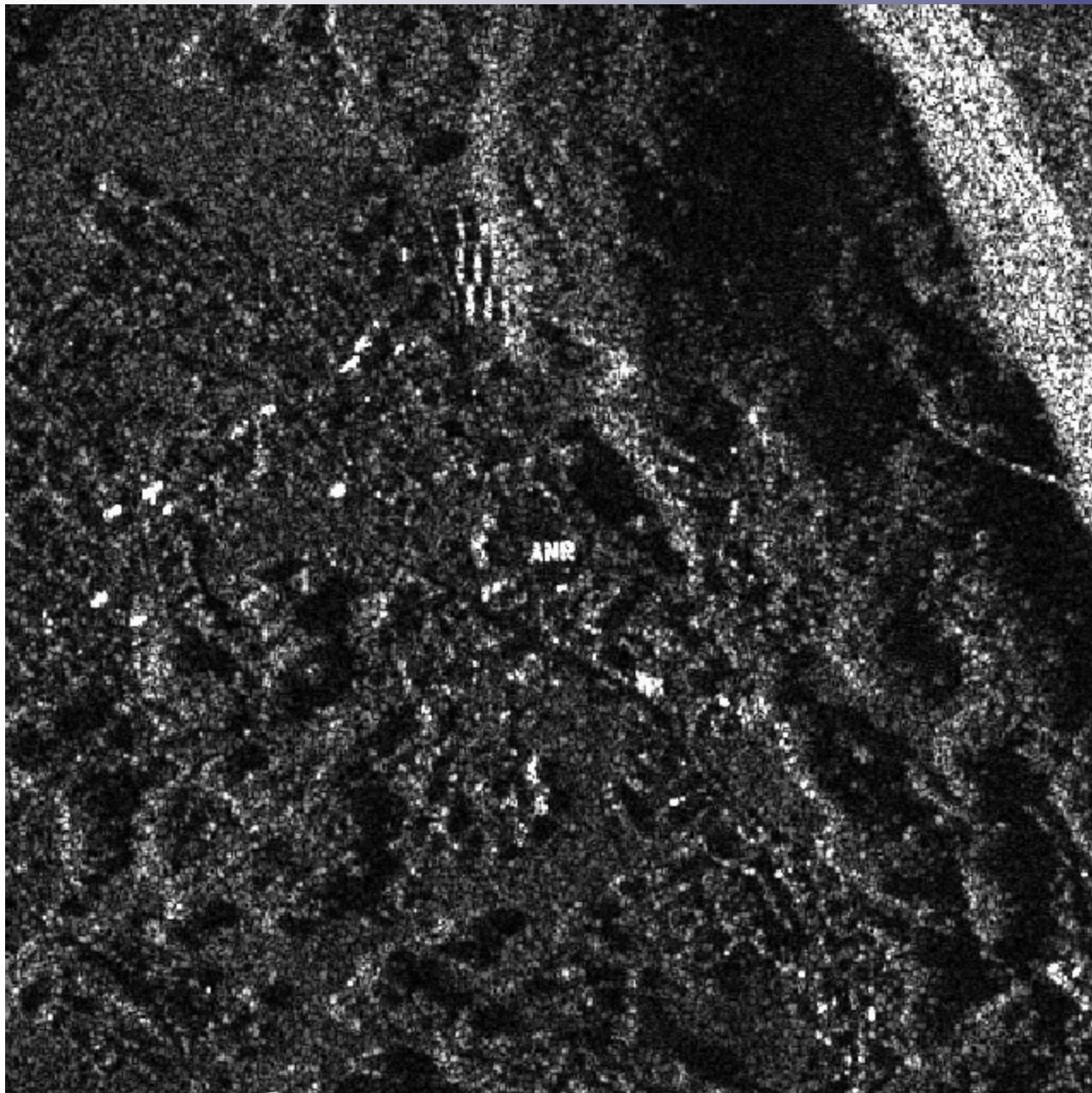
Trouvé E., Nicolas J-M., Ferro-Famil L., Gay M., Pinel, Doin M-P., Méger N., Lasserre C., Mauris G., Vernier F., Fallourd R., Yan Y., Harant O., Jolivet R., EFIDIR : extraction et fusion d'informations pour la mesure de déplacements par imagerie radar, Revue Traitement du Signal (TS), N°28, VOL 3-4, 2011, pp. 375-416.

A. Julea, N. Meger, C. Rigotti, E. Trouve, R. Jolivet, and P. Bolon., Efficient Spatiotemporal Mining of Satellite Image Time Series for Agricultural Monitoring. In journal: Transactions on Machine Learning and Data Mining. 23 pages. Accepted: October 2011. To appear.

Some other applications

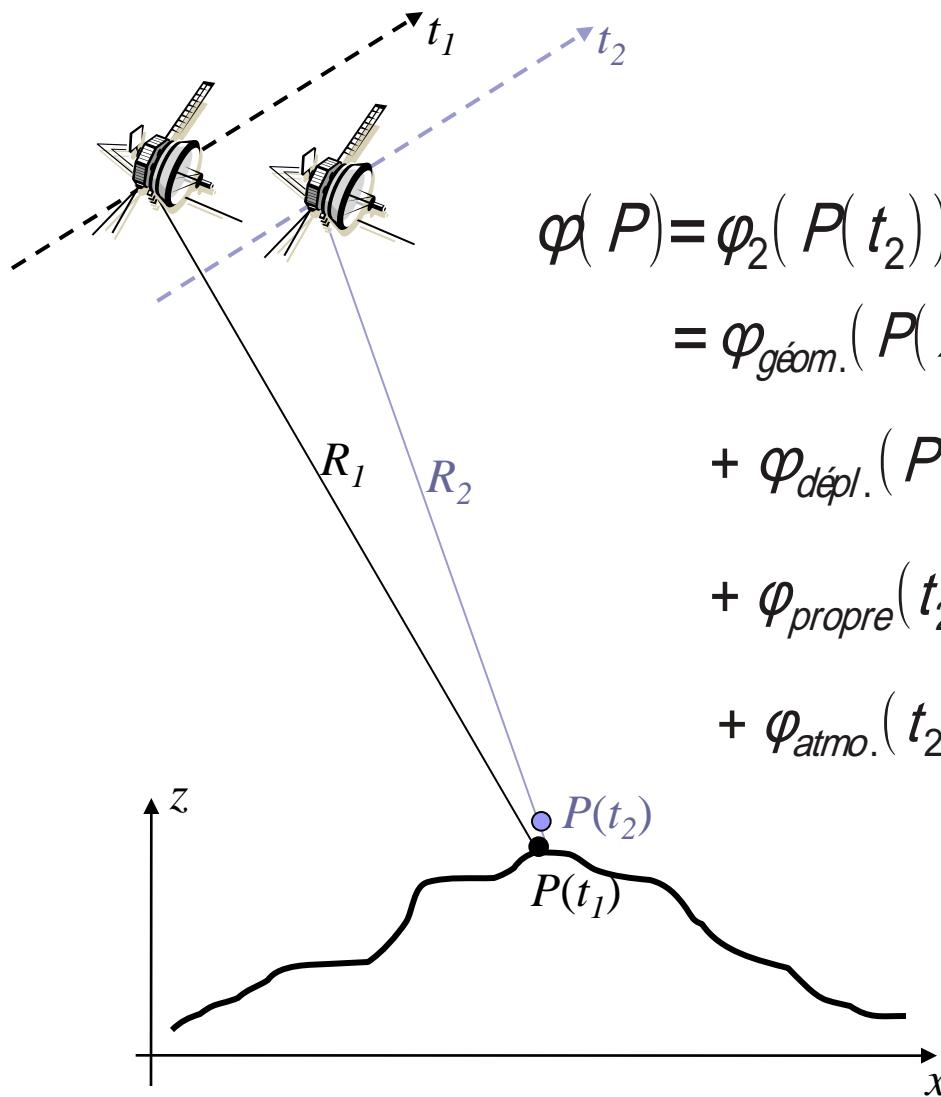
Julea A., Ledo F., Méger N., Trouvé E., Bolon Ph., Rigotti C., Fallourd R., Nicolas J-M., Vasile G., Harant O., Ferro-Famil L., and Lodge F., Polsar Radarsat-2 Satellite Image Time Series Mining Over the Chamonix Mont-Blanc Test Site, IEEE Int. Geoscience And Remote Sensing Symposium (IGARSS 11), Vancouver, Canada, July 2011, pp 1191-1194.

Julea A., Méger N., Trouvé E., On mining METEOSAT and ERS Multitemporal Images, 4th Conf. on Image Information Mining for Security and Intelligence (ESA-EUSC 2006), CD-ROM , Torrejon Air Base - Madrid, Spain, November 2006, 6 pages.





Mesure de déplacement par interférométrie



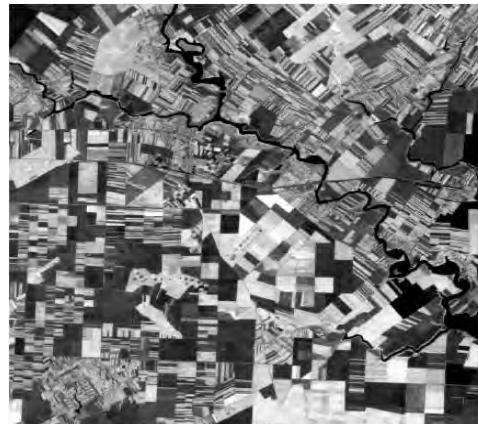
$$\begin{aligned}\varphi(P) &= \varphi_2(P(t_2)) - \varphi_1(P(t_1)) = \frac{4\pi}{\lambda}(R_2 - R_1) + \dots \\ &= \varphi_{\text{geom.}}(P(x, z)) \quad \xrightarrow{\text{Modèle Numérique de Terrain}} \\ &\quad + \varphi_{\text{dépl.}}(P(t_2) - P(t_1)) \quad \xrightarrow{\text{Champ de déplacement}} \\ &\quad + \varphi_{\text{propre}}(t_2) - \varphi_{\text{propre}}(t_1) \quad \xrightarrow{\text{Décorrélation}} \\ &\quad + \varphi_{\text{atmo.}}(t_2) - \varphi_{\text{atmo.}}(t_1) \quad \xrightarrow{\text{Perturbations atmosphériques}}\end{aligned}$$

5. Applications

Crop monitoring by optical remote sensing

Available data

- ADAM SITS: data assimilation through agro-modeling (<http://kalideos.cnes.fr>)
- 60 SPOT images between 2000 and 2008.
- 3 bands: B1 in green (0.5-0.59 μm), B2 in red (0.61-0.68 μm), and B3 in near infrared (NIR 0.78-0.89 μm).
- Spatial resolution: 20m 20m.
- Observed scene: rural area in East Bucharest, Romania.



Data selection and preprocessing

- Sub-scene: 1000 1000 pixels.
- The ground truth is available for the 2000-2001 period (fields belonging to the Romanian National Agricultural Research and Development Institute, 5.9% of the scene).
- 20 images between October 2000 and July 2001.
- A synthetic band B4 is computed. It gives the *Normalized Difference Vegetation Index* (NDVI). $B4 = B3 - B2 / B3 + B2$. Pixel values are quantized into 3 intervals.

Quantitative results

Standard PC: Intel Core 2 @ 3GHz, 4 GB RAM,
linux kernel 2.6.3

- $\sigma = 1\%$
- $k = 5$
- Execution times: 750s
- Number of FGS-patterns: 474
- Maximal FGS-patterns are selected (32 patterns)

Results ($\sigma=1\%$, $K=5$)

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

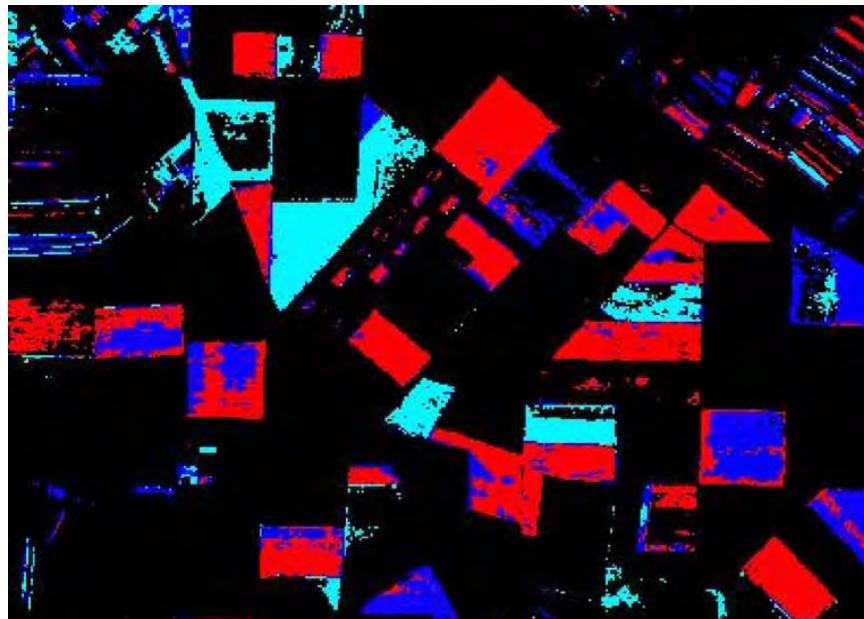


White pixels relate to a single FGS-pattern which characterizes wheat crops.

It covers 61.4% of the wheat crops of the ground truth (MCC) and 91.3% of the white pixels indeed match wheat crops (purity).

Results ($\sigma=1\%$, $K=5$)

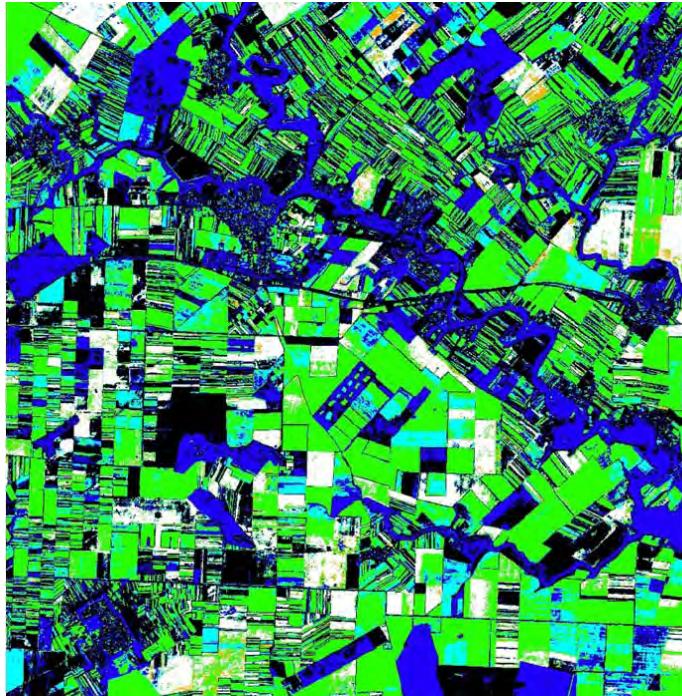
1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 1 → 2 → 2 → 2 → 2



Colored pixels mainly correspond to corn cultures. MCC= 65.7%, purity=76.36%.

Colors represent a temporal discrimination: early acquisitions dates in light blue (pea/sudan grass) and late acquisitions dates in red and dark blue (corn).

Results: towards a single image?



Number of 18-FGS-patterns that cover pixels ($\sigma = 0.5, \kappa = 5, s = 2$).

Black: 0 pattern - Violet: 1 pattern - Blue: 2 patterns - Green: 3 patterns - White: 5 patterns

Results: N_p vs σ , K , S

