

+

IN-OVIVE 2015

Extraction de motifs
à partir de données environnementales

Maguelonne TEISSEIRE




+

Plan

1. TETIS, SISO et la fouille de données
2. Le projet Fresqueau
3. Une méthode de fouille de données séquentielles
4. Conclusion, Applications et perspectives

+

UMR TETIS
Territoire, Environnement, Télédétection et
Information spatiale



AgroParisTech

irstea

cirad

+

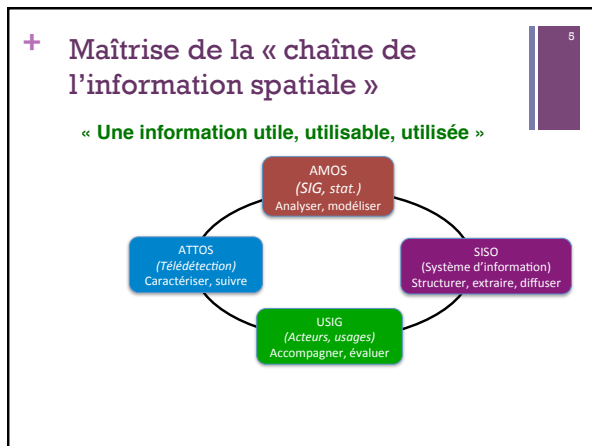
Basée à la maison de la
télédétection



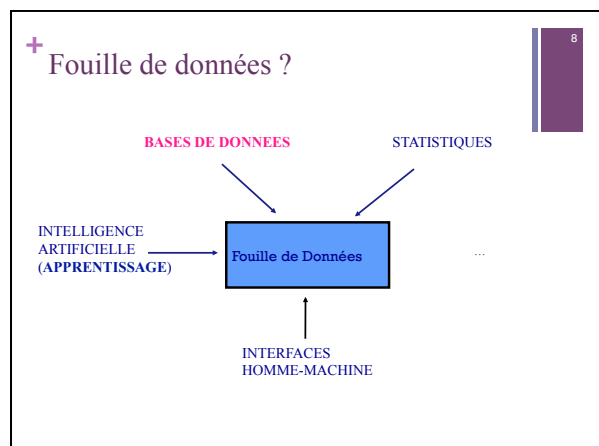
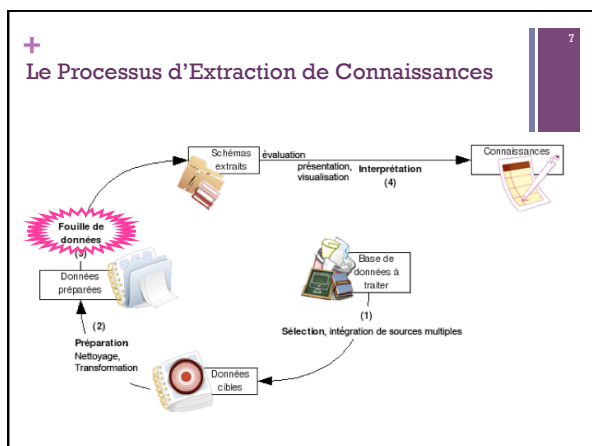
AgroParisTech

irstea

cirad



- ### + Les thématiques
- **Agriculture**
 - Production
 - Pratiques et impacts sur l' environnement
 - **Ressources naturelles**
 - La forêt
 - La biodiversité
 - L' eau
 - **Développement territorial**
 - Le périurbain
 - Le foncier
 - Les activités extractives
 - **Santé animale et Risques naturels**



+ Les tâches de la fouille de données

- **Fouille de données** : de nombreuses tâches possibles (27)
 - **Classification**
Créer une fonction qui classe une donnée élémentaire parmi plusieurs classes prédéfinies existantes
 - **Régression**
Créer une fonction qui donne une donnée élémentaire à une variable de prévision avec des données réelles
 - **Groupement (clustering)**
Rechercher à identifier un ensemble fini de catégories ou groupe en vues de décrire les données
 - **Résumé**
Affiner une description compacte d'un sous-ensemble de données
 - **Modélisation des dépendances**
Trouver un modèle qui décrit des dépendances significatives entre les variables
 - **Détection de changement et déviation**
Découvrir les changements les plus significatifs dans les données

+ Quels algorithmes de fouille de données ?

- Non pas 1 mais n approches ... avec m techniques ...
- 3 approches principales (*R. Agrawal*) **vision Base de Données**
 - Classification*
 - Règles d'association*
 - Motifs séquentiels*

+ Règles d'Association vs Motifs Séquentiels

- **Corrélation entre les produits (RA)**
Les personnes qui achètent des couches achètent de la bière
- **Comportement des clients au cours du temps (MS)**
Les personnes qui achètent des couches achètent trois jours après de la bière


Motifs fréquents, motifs rares, Tendances, anomalies

+ Fouille de données vs Statistiques

The diagram consists of two overlapping blue circles. The left circle is labeled 'Techniques Statistiques' and 'Confirmatoires'. The right circle is labeled 'Techniques de Data Mining' and 'Exploratoires'. A blue arrow points from the left circle to the right, labeled 'Hypothèses'. A blue arrow points from the right circle to the left, labeled 'Nouvelles informations'.

+ Machine Learning vs Data Mining

13



Passage à l'échelle

+ Le projet Fresqueau

14

+ **Projet Fresqueau (ANR MN – MD)**

15

Fouille de données pour l'évaluation et le suivi de la qualité hydrobiologique des cours d'eau

Laboratoires

- LHYGES – ENGEES (Strasbourg)
- TETIS – Cemagref (Montpellier)
- LSIIT (Strasbourg)
- LIRMM (Montpellier)

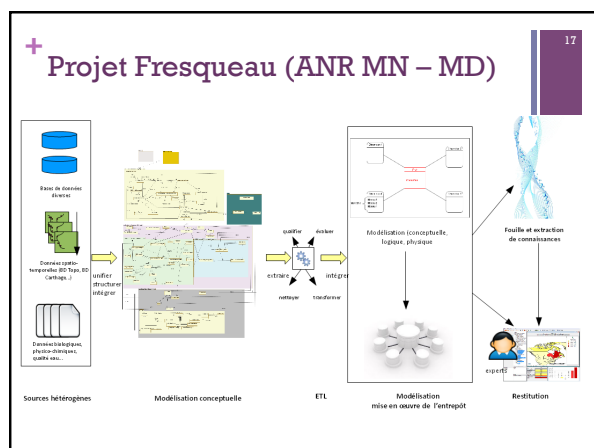
PME

- Aquabio (Région Aquitaine)
- Aquascop (Région Pays de Loire & Languedoc)

+ **Projet Fresqueau (ANR MN – MD)**

16

- > **Les données de qualité d'eau**
Bioindicateurs, paramètres physico-chimiques, ...
→ Qualifier de façon détaillée et complémentaire la qualité des cours d'eau
- > **Les données d'hydrographie**
BD Carthage, BD TOPO, ...
→ Comparer/compléter l'influence de la donnée réseau et ses caractéristiques
- > **Les données des stations de mesure de qualité d'eau**
Stations de suivis nationaux, stations d'études, ...
→ Mettre en avant la complémentarité des informations (thématique, spatiale, temporelle)
- > **Les données de variables de forçage ou de contexte**
Climat, débits, hydrocorégions, ...
→ Caractériser l'environnement des rivières et des points de mesures
- > **Les données liées aux activités humaines**
Occupation du sol, obstacles à l'écoulement, ...
→ Estimer les pressions anthropiques qui s'exercent sur les cours d'eau



+ Une méthode d'extraction d'informations synthétiques à partir de données séquentielles

18



Contexte thématique

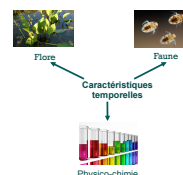
20

Question des experts

Besoins des hydrobiologistes posés sous la forme de questions

Question thématique étudiée :

Peut-on trouver des liens dans le temps entre des ensembles de valeurs de paramètres physico-chimiques et les valeurs de paramètres biologiques ?



21

Contexte thématique

Question des experts
 Besoins des hydrobiologistes posés sous la forme de questions

Question thématique :
 Peut-on trouver des **liens** dans le **temps** entre des ensembles de valeurs de paramètres **physico-chimiques** et les valeurs de paramètres **biologiques** ?

Méthodes de fouille de données temporelles

Méthodes de fouille de motifs discriminants

22

Contexte méthodologique

Processus d'extraction de connaissances

23

Contexte méthodologique

Processus d'extraction de connaissances

24

Contexte méthodologique

Jeu de données temporelles

Du jeu de données temporelles à un jeu de séquences

	Janvier	Février	Mars	Avril	Mai
Station 1	c,d	a		g	d
Station 2		g	c,d,e	f	a,e,g
Station 3	g	d	e	f	

Item : g
 Alphabet : {a,b,c,d,e,f,g}
 Itemset : (c,d,e)

25

Contexte méthodologique

Séquences

Du jeu de données temporelles à un jeu de séquences

	Janvier	Février	Mars	Avril	Mai
Station 1	c,d	a		g	d
Station 2		g	c,d,e	f	a,e,g
Station 3	g	d	e	f	

↓

	Séquences
Séquence 1	< (c,d) (a) (g) (d) >
Séquence 2	< (g) (c,d,e) (f) (a,e,g) >
Séquence 3	< (g) (d) (e) (f) >

26

Contexte méthodologique

Motifs séquentiels

Extraction de motifs séquentiels fréquents (**sous-séquences** fréquentes)

	Séquences
Séquence 1	< (c,d) (a) (g) (d) >
Séquence 2	< (g) (c,d,e) (f) (a,e,g) >
Séquence 3	< (g) (d) (e) (f) >

27

Contexte méthodologique

Motifs séquentiels

Extraction de motifs séquentiels fréquents (**sous-séquences** fréquentes)

	Séquences
Séquence 1	< (c,d) (a) (g) (d) >
Séquence 2	< (g) (c,d,e) (f) (a,e,g) >
Séquence 3	< (g) (d) (e) (f) >

Exemple :
 < (g) (d) > : 3/3

28

Contexte méthodologique

Motifs séquentiels

Extraction de motifs séquentiels fréquents (**sous-séquences** fréquentes)

	Séquences
Séquence 1	< (c,d) (a) (g) (d) >
Séquence 2	< (g) (c,d,e) (f) (a,e,g) >
Séquence 3	< (g) (d) (e) (f) >

Exemple :
 < (g) (d) > : 3/3
 < (c,d) (a) > : 2/3

Un motif M est **fréquent** si $\text{Fréquence}(M) \geq \theta$, une valeur de **fréquence minimale** fixée par l'utilisateur

29

Contexte méthodologique

Motifs séquentiels

Extraction de motifs séquentiels fréquents (**sous-séquences** fréquentes)

Séquences	Séquences
Séquence 1	< (c,d) (a) (g) (d) >
Séquence 2	< (g) (c,d,e) (f) (a,e,g) >
Séquence 3	< (g) (d) (e) (f) >

< (c,d) >
< (c,d) (a) >

30

Contexte méthodologique

Motifs séquentiels clos

Extraction de motifs séquentiels fréquents (**sous-séquences** fréquentes)

Séquences	Séquences
Séquence 1	< (c,d) (a) (g) (d) >
Séquence 2	< (g) (c,d,e) (f) (a,e,g) >
Séquence 3	< (g) (d) (e) (f) >

Un motif séquentiel A est **clos** s'il n'existe pas de motif B tel que Fréquence(A) = Fréquence(B) et A est une **sous-séquence** de B

< (c,d) > n'est pas clos
< (c,d) (a) > est clos

31

Contexte méthodologique


Motifs séquentiels clos

Extraction de motifs séquentiels fréquents (**sous-séquences** fréquentes)

Séquences	Séquences
Séquence 1	< (c,d) (a) (g) (d) >
Séquence 2	< (g) (c,d,e) (f) (a,e,g) >
Séquence 3	< (g) (d) (e) (f) >

Ensemble des motifs clos supportés par S1 et S2 :

< (g) (d) > : 3/3
< (c,d) (a) > : 2/3
< (c,d) (g) > : 2/3



Un **même ensemble** de séquences peut supporter plusieurs motifs séquentiels clos

32

Contexte méthodologique

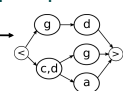
Motifs partiellement ordonnés clos

Extraction de motifs séquentiels fréquents (**sous-séquences** fréquentes)

Séquences	Séquences
Séquence 1	< (c,d) (a) (g) (d) >
Séquence 2	< (g) (c,d,e) (f) (a,e,g) >
Séquence 3	< (g) (d) (e) (f) >

Ensemble des motifs clos supportés par S1 et S2 :

< (g) (d) > : 3/3
< (c,d) (a) > : 2/3
< (c,d) (g) > : 2/3



Motifs partiellement ordonnés clos

33

Contexte méthodologique

Motifs partiellement ordonnés clos

Extraction de motifs séquentiels fréquents (**sous-séquences** fréquentes)

Séquences	
Séquence 1	< (c,d) (a) (g) (d) >
Séquence 2	< (g) (c,d,e) (f) (a,e,g) >
Séquence 3	< (g) (d) (e) (f) >

Ensemble des motifs clos supportés par S1 et S2 :

- < (g) (d) > : 3/3
- < (c,d) (a) > : 2/3
- < (c,d) (g) > : 2/3

Motifs partiellement ordonnés clos

34

Contexte méthodologique

Motifs partiellement ordonnés clos

Extraction de motifs séquentiels fréquents (**sous-séquences** fréquentes)

Séquences	
Séquence 1	< (c,d) (a) (g) (d) >
Séquence 2	< (g) (c,d,e) (f) (a,e,g) >
Séquence 3	< (g) (d) (e) (f) >

Ensemble des motifs clos supportés par S1 et S2 :

- < (g) (d) > : 3/3
- < (c,d) (a) > : 2/3
- < (c,d) (g) > : 2/3

Motifs partiellement ordonnés clos

35

Contexte méthodologique

Motifs partiellement ordonnés clos

Extraction de motifs séquentiels fréquents (**sous-séquences** fréquentes)

Séquences	
Séquence 1	< (c,d) (a) (g) (d) >
Séquence 2	< (g) (c,d,e) (f) (a,e,g) >
Séquence 3	< (g) (d) (e) (f) >

Ensemble des motifs clos supportés par S1 et S2 :

- < (g) (d) > : 3/3
- < (c,d) (a) > : 2/3
- < (c,d) (g) > : 2/3

Motifs partiellement ordonnés clos

36

Contexte méthodologique

Méthodes d'extraction de motifs temporels

De nombreux algorithmes développés pour extraire les motifs séquentiels

37

Contexte méthodologique

Méthodes d'extraction de motifs temporels

A horizontal timeline arrow pointing to the right, with vertical lines indicating the years of publication for various algorithms. The algorithms listed are: GSP (1996), PrefixSpan (2001), SPADE (2001), BIDE (2004), FreeSpan (2000), CloSpan (2003), SPAM (2002), and Frecpo (2005). A box labeled 'C. Garriga (2005)' is positioned above the timeline, with a dotted line connecting it to the year 2005.

De nombreux algorithmes développés pour extraire les motifs séquentiels
Mais peu d'approches pour extraire les motifs partiellement ordonnés

38

Processus

Utiliser la notion de structure partiellement ordonnée pour extraire, synthétiser et résumer l'information

Jeu de séquences

39

Processus

Utiliser la notion de structure partiellement ordonnée pour extraire, synthétiser et résumer l'information

Jeu de séquences

Algorithmes d'extraction de motifs partiellement ordonnés clos

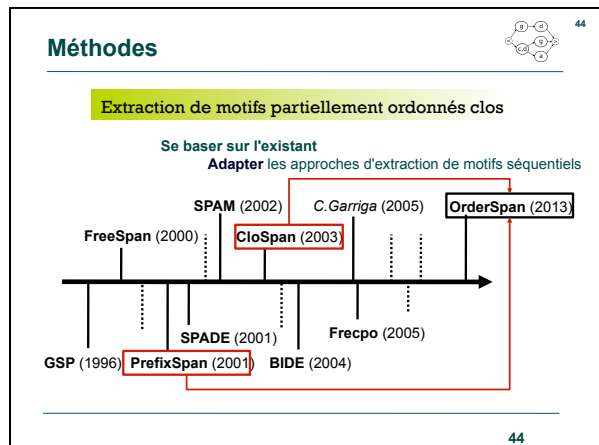
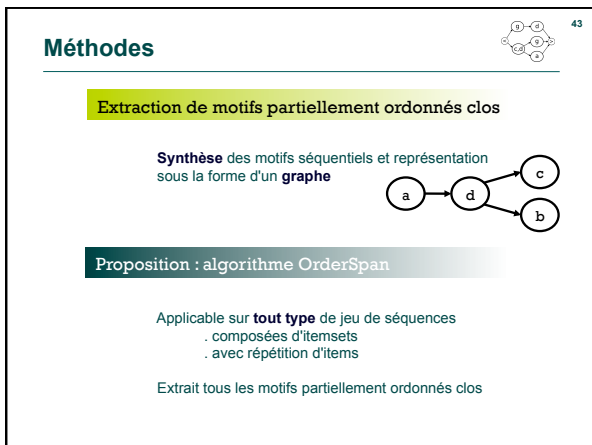
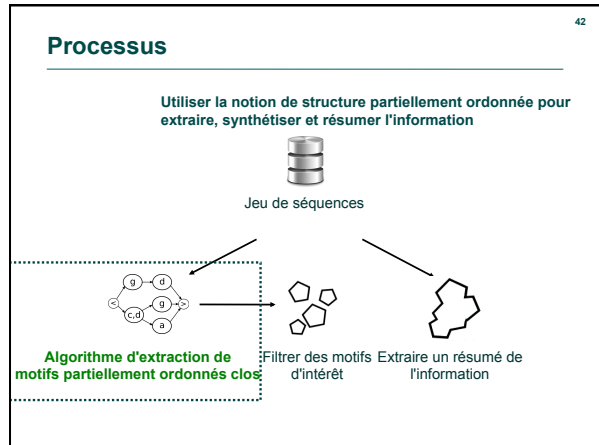
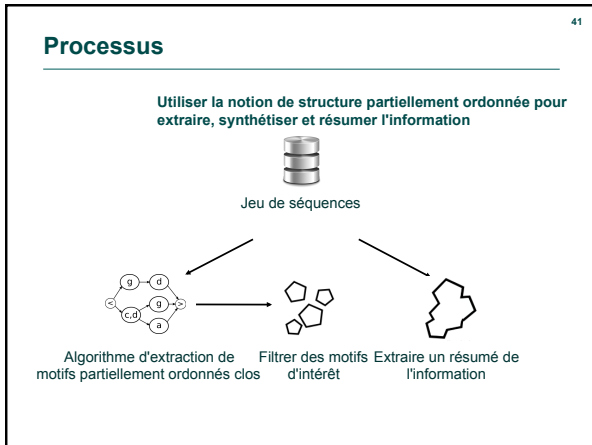
40

Processus

Utiliser la notion de structure partiellement ordonnée pour extraire, synthétiser et résumer l'information

Jeu de séquences

Algorithmes d'extraction de motifs partiellement ordonnés clos Filtrer des motifs d'intérêt



Méthodes

Algorithme

Première étape
 - **Extraction** d'un **sous-arbre** de l'espace de recherche des motifs séquentiels sur un **sous-ensemble** de séquences
 Utilisation des **préfixes fréquents (PrefixSpan)**

<(e)(b)(g)>
 <(e)(a)(g)>
 <(e)(a)(b)>

45

Méthodes

Algorithme

Première étape
 - **Extraction** d'un **sous-arbre** de l'espace de recherche des motifs séquentiels sur un **sous-ensemble** de séquences
 Utilisation des **préfixes fréquents (PrefixSpan)**

<(e)(b)(g)>
 <(e)(a)(g)>
 <(e)(a)(b)>

46

Méthodes

Algorithme

Seconde étape
 - **Fusion** des **sommets** redondants
 Utilisation des **suffixes fréquents**
 - **Suppression** des **arcs** redondants (transitivité)

<(e)(b)(g)>
 <(e)(a)(g)>
 <(e)(a)(b)>

Sommet de fin

47

Méthodes

Algorithme

Seconde étape
 - **Fusion** des **sommets** redondants
 Utilisation des **suffixes fréquents**
 - **Suppression** des **arcs** redondants (transitivité)

<(e)(b)(g)>
 <(e)(a)(g)>
 <(e)(a)(b)>

48

Méthodes

Algorithme

Seconde étape

- Fusion des **sommets** redondants
- Utilisation des **suffixes fréquents**
- Suppression des **arcs** redondants (transitivité)

$\langle e \rangle (b)(g) \rangle$
 $\langle e \rangle (a)(g) \rangle$
 $\langle e \rangle (a)(b) \rangle$

49

Méthodes

Extraction de motifs partiellement ordonnés clos

Synthèse des motifs séquentiels et représentation sous la forme d'un **graphe**

Bilan

Applicable sur **tout type** de jeu de séquences

- . composée d'itemsets
- . avec répétition d'items

Extrait tous les motifs partiellement ordonnés clos

→ Limitation : le nombre de motifs extraits peut être important

50

Processus

Utiliser la notion de structure partiellement ordonnée pour extraire, synthétiser et résumer l'information

Jeu de séquences

Algorithme d'extraction de motifs partiellement ordonnés clos

Filtrer des motifs d'intérêt

Extraire un résumé de l'information

51

Méthodes

Filtrer k motifs d'intérêt

Sélection des motifs

Parmi des milliers de motifs extraits, sélectionner quelques motifs d'intérêt

Trois **critères** d'intérêt identifiés pour les utilisateurs :

- la fréquence
- la discrimance
- la redondance

Proposition

Un algorithme qui permet de sélectionner **k motifs en pondérant** les trois critères

52

Application


fresQueau

Question des experts

Besoins des hydrobiologistes posés sous la forme de questions


Question thématique étudiée :
Peut-on trouver des liens dans le temps entre des ensembles de valeurs de paramètres physico-chimiques et les valeurs de paramètres biologiques ?

Poissons



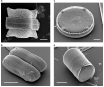
IPR
(Indice Poisson Rivière)

Macroinvertébrés



IBGN
(Indice Biologique Global Normalisé)

Diatomées



IBD
(Indice Biologique Diatomées)

Application

fresQueau

Pré-traitement

Jeu de données initial

Station	Date	Paramètres physico-chimiques					Indice biologiques
		NH ₄ ⁺	NKJ	NO ₂ ⁻	PO ₄ ³⁻	P	
1	02/07	-	-	-	0.123	0.032	-
	06/07	-	0.672	0.026	-	-	-
	07/07	0.088	1.235	0.134	-	0.011	-
	09/07	-	-	-	-	-	17
	12/07	0.154	-	0.246	0.168	0.338	-
	02/08	0.062	0.040	0.091	0.025	0.003	-
	04/08	-	0.023	0.198	-	-	-
	05/08	-	-	-	-	-	12
	07/08	-	-	-	0.046	0.009	-
	01/04	0.043	0.146	0.421	-	-	-
2	04/04	-	-	-	1.325	0.093	-
	07/04	2.331	7.993	0.252	0.132	0.266	-
	08/04	-	1.414	-	-	-	-
	09/04	-	-	-	-	-	8
	11/04	0.117	0.0844	-	0.688	-	-
	12/04	-	-	-	0.067	0.278	-
	03/05	-	0.182	0.0310	0.337	-	-
	06/05	0.004	-	0.012	0.035	0.134	-
	08/05	-	-	-	-	-	10

Application

fresQueau

Pré-traitement

Connaissance experte

Utilisation des classes de qualité

Biologie

	17	13	9	5	
Indice Biologique Global Normalisé	17	13	9	5	
Indice Biologique Diatomées	17	13	9	5	
Indice Poisson Rivière	7	16	25	36	

Physico-chimie Macro-paramètres physico-chimiques basés sur le SEQ

		AZOT				
Ammonium	mg/L	0.1	0.5	2	5	
Azote kjeldahl	mg/L	1	2	4	10	
Nitrites	mg/L	0.03	0.3	0.5	1	
		PHOS				
Orthophosphates	mg/L	0.05	0.15	0.3	0.4	
Phosphore total	mg/L	0.015	0.05	0.1	0.15	

Application

fresQueau

Pré-traitement

Jeu de données discrétisé

Station	Date	AZOT	PHOS	IBGN
1	02/07	-	Vert	-
	06/07	Bleu	-	-
	07/07	Vert	Bleu	-
	09/07	-	-	Bleu
	12/07	Vert	Jaune	-
	02/08	Vert	Bleu	-
	04/08	Vert	-	-
	05/08	-	-	Jaune
	07/08	-	Bleu	-
	01/04	Jaune	-	-
2	04/04	-	Orange	-
	07/04	Orange	Jaune	-
	08/04	Vert	-	-
	09/04	Vert	-	Orange
	11/04	Vert	Jaune	-
	12/04	-	Jaune	-
	03/05	Vert	Vert	-
	06/05	Bleu	Vert	-
	08/05	-	-	Jaune

Regroupement de la physico-chimie

AZOT – Matières azotées hors nitrates (NH₄⁺, NKJ, NO₂⁻)

PHOS – Matières phosphorées (PO₄³⁻, P)

Application fresQueau

Pré-traitement

Découpage des séquences selon la biologie
 Récupération de l'ensemble des variables entre 0 et 6 mois avant une variable représentant une **note d'indice biologique**

Séquence station <(... ..) (... ..) (... ..) (... ..) >
 (.. ..) >

Possibilité d'obtenir plusieurs intervalles par séquence si plusieurs prélèvements biologiques ont été effectués

<(... ..) (... ..) (... ..) (... ..) >
 (... ..) >

Application fresQueau

Pré-traitement

Découpage des séquences selon la biologie
 Ensemble de séquences d'intervalles pour chaque classe de qualité de chaque indice biologique

	IBGN	IBD	IPR
●	89	17	62
●	556	108	76
●	1282	532	126
●	2405	1375	162
●	1056	1076	52

Application 59 fresQueau

Résultats

Méthodes implantées dans un logiciel de visualisation

Filter sur les données

Sélection des stations à analyser

Extraction des motifs

Visualisation des motifs

IBGN : 11.49% 15.66% 19.49% 20.81% 2.24%

Application 60 fresQueau

Résultats

Motifs obtenus
 Sélection des 15 motifs les plus équilibrés dans chaque classe de qualité pour chaque indice

Quelques exemples

IBGN rouge

PHOS

IBGN : 1.06% 1.84% 7.11% 17.19% 29.21%

Growth rate de 1,7

IBGN rouge

AZOT PHOS

IBGN : 0% 0.08% 0.61% 4.07% 12.35%

Growth rate de 3,03

La combinaison ou la répétition de plusieurs variables physico-chimiques rendent les motifs plus discriminants

Application fresQueau 61

Résultats

Exemple de consensus discriminants obtenus
Fréquence minimale de 10 %

IBGN jaune

IBGN orange-rouge

Bilan 62

- Motifs partiellement ordonnés clos**
Algorithme pour l'extraction des motifs partiellement ordonnés clos
- Filtrer les motifs d'intérêt**
Algorithme pour filtrer k motifs partiellement ordonnés clos d'intérêt selon plusieurs critères
- Consensus partiellement ordonné**
Utilisation des séquences partiellement ordonnées comme un résumé global
- Méthodes adaptées aux jeux de données découpés en plusieurs classes**

+ Conclusion
Quelles applications ?

63

+ Challenges pour la fouille de données environnementales 64

- **Méthodes bien connues**
 - Clustering
 - Classification
 - Motifs
 - RA spatiales
 - Co-location
 - Trajectoires
 -

- **Questions classiques en Fouille**
 - Passage à l'échelle
 - Qualité des motifs extraits
 - Intégration des connaissances du domaine
 - Metadonnées
 - Ontologies
 - ...

+ À court terme

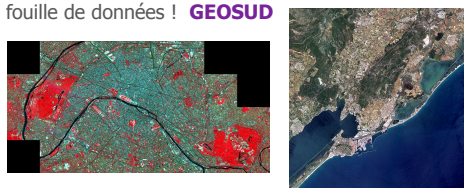
65

- Prendre en compte la **sémantique** pour identifier des motifs "plus" intéressants
 - Événements: inattendus, moindre contradiction temporelle...
 - Trajectoires: les maximales, top-k...
 - Comparer les motifs
 - Événements: inclusion spatiale, temporelle...
 - Trajectoires: similarité de forme, de direction, de proximité...
- Besoin d'une similarité sémantique

+ Et puis ?

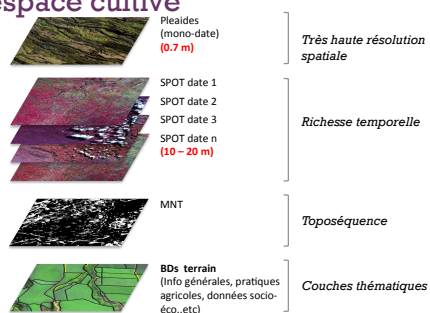
66

- Fouille d'images satellites (Remote Sensing Image Mining) avec une perspective temporelle et spatiale encore peu explorée
- Beaucoup de données,... mais peu de méthode de fouille de données ! **GEOSUD**



+ Cartographie et caractérisation de l'espace cultivé

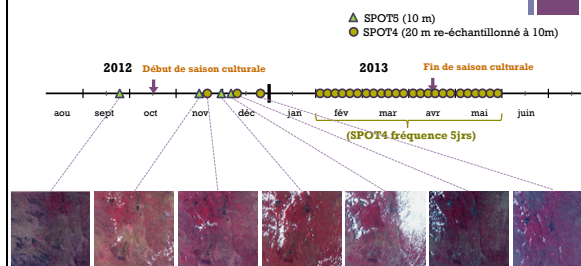
67



+ Granularité temporelle

68

Séries temporelles d'images SPOT 4 & 5 (résolution : 10 m)



+ Quelles informations utilisées?

TELEDETECTION

- Vert
- Rouge
- PIR
- Texture Vert
- Texture Rouge
- Texture PIR

Informations radiométriques, texturales et temporelles

- Carte culture / non culture
- Distance au village
- Taille de la parcelle ou ilot cultivé
- Place dans la toposéquence (MNT)

Informations statiques

TERRAIN

BD Terrain :

- Relevés parcellaires
- Contour GPS
- Culture au sol
- Itinéraire technique
- ...

Typologie des systèmes de culture

Classe A : riz irrigué
Classe B : riz pluvial de tavy (colline)
Classe C : maïs
Classe D : manioc...

Quel polygone d'extraction (objet eCognition, ou fenetre de $x * x$ pixels ?? Quid de la différence de résolution spatiale entre les différentes données ?

+ Un grand merci à...

- Valentine Lebourgeois
- Mickaël Fabrègue

+ Références

- M. Fabrègue, A. Braud, S. Bringay, F. Le Ber, M. Teisseire: Mining closed partially ordered patterns, a new optimized algorithm. *Knowl.-Based Syst.* 79: 68-79 (2015)
- M. Fabrègue, A. Braud, S. Bringay, C. Grac, F. Le Ber, D. Levat, M. Teisseire: Discriminant temporal patterns for linking physico-chemistry and biology in hydro-ecosystem assessment. *Ecological Informatics* 24: 210-221 (2014)
- H. Alatrasta-Salas, J. Azé, S. Bringay, F. Cernesson, N. Selmaoui-Folcher, M. Teisseire: A knowledge discovery process for spatiotemporal data: Application to river water quality monitoring. *Ecological Informatics* 26(2): 127-139 (2015)
- I. Tsoukatos, D. Gunopoulos, «Efficient mining of spatiotemporal patterns», In *Advances in Spatial and Temporal Databases*, p. 428-442, 2001
- H. Cao, N. Mamoulis, D. Cheung, «Mining frequent spatio-temporal sequential patterns», In *ICDM 2005*
- Y. Huang, L. Zhang, and P. Zhang, «A framework for mining sequential patterns from spatio-temporal event data sets». In *TKDE*, p. 433-448, 2008
- P. Mohan, S. Shekhar, J. A. Shine, and J. P. Rogers. Cascading spatio-temporal pattern discovery: A summary of results. In *SDM*, p. 327-338, 2010.
- N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D. W. Cheung. Mining, indexing, and querying historical spatio-temporal data. In *KDD '04*, pages 236-245, 2004.

+ Références

- Y. Wang, E.-P. Lim, and S.-Y. Hwang. Efficient mining of group patterns from user movement data. *DKE*, 57(3):240-282, June 2006.
- C.S. Jensen, D. Lin, and B. C. Ooi. Continuous clustering of moving objects. *TKDE*, 19(9):1161-1174, sept. 2007.
- H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen, and H. T. Shen. Discovery of convoys in trajectory databases. In *PVLDB*, 1(1):1068-1080, August 2008.
- M. R. Vieira, P. Bakalow, and V. J. Tsotras. On-line discovery of flock patterns in spatio-temporal data. In *GIS '09*, pages 286-295, 2009.
- H. H. Aung, K.-L. Tan: Discovery of Evolving Convoys. In *SSDBM*, pages 196-213, 2010.
- Z. Li, B. Ding, J. Han, and R. Kays. Swarm: mining relaxed temporal moving object clusters. In *PVLDB*, 3(1-2):723-734, September 2010.
- P. N. Hai, D. Ienco, P. Poncelet, and M. Teisseire. Ming time relaxed gradual moving object clusters. In *GIS '12*, pages 478-481, 2012.
- Y. Pitarch, D. Ienco, E. Vintrou, A. Bégué, A. Laurent, P. Poncelet, M. Sala, M. Teisseire: Spatio-temporal data classification through multidimensional sequential patterns: Application to crop mapping in complex landscape. *Eng. Appl. of AI* 37: 91-102 (2015).
- K. Zheng, Y. Zheng, J. Yuan, and S. Shang. On discovery of gathering patterns from trajectories. In *ICDE*, 2013.