

PROJET ANR-10-COSI-012

**FOSTER : FOuille de données
Spatio-Temporelles : application à la
compréhension et à la surveillance de
l'ERosion**

PROGRAMME COSINUS 2010

Pôle Pluridisciplinaire de la Matière et de l'Environnement (PPME)
Université de la Nouvelle-Calédonie

Laboratoire d'InfoRmatique des Images et des Systèmes d'information (LIRIS)
Université de Lyon

Laboratoire des Sciences de l'Image, de l'Informatique
et de la Télédétection (LSiIT) - Université de Strasbourg

Laboratoire d'Informatique, Systèmes, Traitement de l'Information
et de la Connaissance (LISTIC) - Université de Savoie

Bluecham SAS

FOSTER - Livrable 2.1
**Etat de l'art sur la fouille de données
spatio-temporelles**

Jean-François Boulicaut¹, Frédéric Flouvat², Andreea Julea^{3,4}, Nicolas Méger⁴,
Christophe Rigotti¹, and Nazha Selmaoui-Folcher²

¹ Université de Lyon, CNRS, INRIA,
INSA-Lyon, LIRIS, UMR 5205, F-69621, Lyon, France
{jean-francois.boulicaut,christophe.rigotti}@insa-lyon.fr
² Université de la Nouvelle-Calédonie, PPME, BP R4, F-98851, Nouméa,
Nouvelle-Calédonie
{frederic.flouvat,nazha.selmaoui}@univ-nc.nc

³ Institute for Space Sciences, P.O. Box MG-23, Ro 077125, Bucharest-Măgurele,
Romania

andreeamj@spacescience.ro
⁴ Université de Savoie, Polytech'Savoie, LISTIC Laboratory, BP 80439, F-74944
Annecy-le-Vieux Cedex, France
{nicolas.meger}@univ-savoie.fr

Table des matières

1	Introduction.....	4
2	Analyse de séries temporelles d'images satellites	4
2.1	Analyse au niveau pixel ou au niveau objet	4
2.2	Méthodes usuelles d'analyse de séries temporelles d'images satellites	8
3	Extraction de motifs dans des bases de données spatio-temporelles (Spatio-temporal Pattern Mining)	12
3.1	Suivi de trajectoires d'objets mobiles	13
3.2	Dynamique spatio-temporelle de phénomènes en fonction de leur environnement	15

1 Introduction

Le projet FOSTER a pour objectif de proposer aux géologues un processus complet permettant de les aider à modéliser et éventuellement prédire les phénomènes érosifs à partir de données satellitaires, de MNT, de données capteurs ou toutes autres sources d'information disponibles sur la zone d'étude. Il s'agit donc d'un processus de découverte de connaissances dans des données. Dans ce contexte, un des principaux objectifs du projet est le développement de nouveaux algorithmes de fouille de données spatio-temporelles.

Plus généralement, les avancées récentes en observation satellitaire ouvrent un grand nombre d'applications : analyse d'écosystèmes profondément affectés par l'activité humaine, évaluation de l'influence d'une sécheresse, d'un feu, d'une inondation, suivi de l'évolution de cultures et de l'occupation des sols. Ainsi, les Séries Temporelles d'Images Satellitaires (STIS) et les Systèmes d'Information Géographique (SIG) sont une mine d'or [64] car ils sont constitués des données à grande échelle temporelle et spatiale contrairement aux données terrain habituellement utilisées. En effet, l'observation régulière de la Terre permet un apprentissage des évolutions et changements normaux et permet donc, par complémentarité, de détecter les changements anormaux. Les développements techniques permettent de profiter de la croissance continue de la résolution spatiale et de la fréquence de revisite d'un même site. Ainsi, les images satellitaires à très haute résolution (THR) sont plus nombreuses, plus riches en information et très complexes, ce qui rend l'interprétation visuelle laborieuse et requiert des analyses automatiques.

La suite de ce document recense des aspects caractéristiques de l'analyse de données (images et/ou bases de données) spatio-temporelles souvent rencontrés dans les travaux des dernières années. Une série d'images satellitaires et les informations adjacentes contenues dans les SIG offrent l'opportunité d'extraire des changements ponctuels ou des évolutions globales au niveau des entités de la scène étudiée et de caractériser les composantes de celle-ci. Différentes approches sont discutées en fonction du niveau de l'entité étudiée (pixel ou objet), de la nature supervisée ou non supervisée des démarches, des méthodes utilisées pour répondre aux diverses tâches (détection de changements, clustering ou classification) et de la nature et complexité de leurs résultats (motifs locaux, modèles globaux).

2 Analyse de séries temporelles d'images satellites

2.1 Analyse au niveau pixel ou au niveau objet

Les facteurs qui influencent le choix entre l'analyse au niveau pixel ou au niveau objet sont :

- le rapport entre les dimensions de pixels et des entités d'intérêt de la scène ;
- le niveau de bruit des données ;
- l'objectif : classification thématique générale ou gestion détaillée des évolutions.

Analyse au niveau pixel Les méthodes classiques d'analyse suivent une approche pixel fondée sur l'hypothèse que chaque pixel d'image est alloué à une seule classe (les pixels sont purs) et que les pixels qui capturent la même classe de couverture de la terre sont proches les uns aux autres dans l'espace des caractéristiques. L'hypothèse sous-jacente de cette approche est que les pixels d'une image se rapportent à des classes de couverture terrestre qui sont relativement séparables par leurs valeurs ou évolutions spectrales. Cette hypothèse n'est pas toujours valable, par exemple dans le cas où le pixel est trop grand par rapport à la variabilité des objets dans le paysage.

En raison de la non-correspondance de la grille de l'image avec les limites de l'objet réel, certains pixels mixtes (mixels) apparaissent dans l'image satellitaire. En cas de pixels mixtes, les réponses spectrales pures des différents objets de la scène sont confondues, menant à un problème de signatures composites. Les pixels mixtes ont été reconnus comme un problème affectant l'utilisation efficace des données de télédétection dans les méthodes de classification et détection des changements [18,23,11].

Fisher [20] a résumé quatre causes du problème de pixel mixte : (1) les limites entre deux ou plusieurs entités cartographiques, (2) la transition à l'intérieur des phénomènes cartographiables, (3) les objets linéaires subpixel (une route) et (4) les petits objets subpixel (une maison, un arbre).

L'*analyse au niveau pixel*, en préservant la résolution d'observation initiale et l'information originale de télédétection, assure une caractérisation plus détaillée des structures étudiées. Elle est une méthode indépendante du domaine d'application qui est très adéquate pour la gestion des évolutions de la couverture terrestre [20,3]. On peut distinguer les modifications à l'intérieur des entités de la scène surveillée (intra-objets) et on peut offrir des informations détaillées en vue d'une aide à la décision. Par exemple, d'une scène agricole peuvent être extraites des informations agronomiques sur les cultures (rendement potentiel, risques de maladie, besoins en eau ou en azote, maturité) qui sont nécessaires pour la conduite optimale des cultures, pour estimer les productions et évaluer la qualité de l'environnement.

L'existence de pixels mixtes conduit à l'élaboration de plusieurs approches pour la classification 'soft' (souvent appelée floue dans la littérature de télédétection) dans laquelle chaque pixel est alloué à toutes les classes dans des proportions variables [71,81,24].

Les résultats de l'approche au niveau pixel sont sensibles à la résolution spatiale par l'intermédiaire de la proportion relative de pixels mixtes, ainsi qu'au bruit. Dans le cas d'une STIS, la méthode exige un recalage parfait des images.

Un exemple de traitement classique d'une STIS est donné dans [29] qui présente une méthodologie et un ensemble de logiciels visant un regroupement pixel par pixel par une stratégie non supervisée et non hiérarchique. Chaque pixel est représenté par la série temporelle de ses valeurs ; les pixels caractérisés par des profils similaires sont affectés au même cluster selon un critère de distance minimale. Les clusters peuvent être regroupés ensuite conformément aux critères choisis par un analyste. Au lieu d'exprimer sa connaissance avant tout regrou-

pement, l'analyste interagit avec une partition exploratoire calculée automatiquement, ce qui offre un certain nombre d'éléments (répartition géographique, relation spatiale et les profils des classes de sortie) susceptibles de faciliter son jugement. Le résultat final est une image classifiée, où tous les pixels d'une même classe ont la même étiquette.

Dans [60], la classification des STIS RSO (Radar à Synthèse d'Ouverture) au niveau du pixel est faite en utilisant des canaux synthétiques dans une étude sur l'estimation précoce des champs agricoles cultivés et non-cultivés. Les canaux synthétiques utilisés pour décrire les évolutions sont des fonctions mathématiques dans le temps comme les moyennes de valeurs de la rétrodiffusion ou les dates de rétrodiffusion maximale (canal qui contient des informations sur la préparation des champs agricoles et sur la phénologie de la récolte). Le même type d'application, toujours sur des STIS RSO, est présentée par [70] où les auteurs utilisent aussi des canaux synthétiques plus raffinés. Le cycle phénologique des cultures est divisé en trois (au début, au cours et à la fin de la saison) et pour chaque période une moyenne de la rétrodiffusion est calculée. Le choix de telles fonctions est un a priori fort qui, pour certaines, efface l'idée même de transition et de changement d'état, mais qui a l'avantage d'être synthétique. On notera aussi qu'ici l'aspect spatial n'est pas pris en compte.

Dans [43,42,41,44] l'évolution des valeurs des pixels est considérée comme essentielle pour caractériser le comportement de la couverture de la terre et les phénomènes météorologiques. L'approche [43,42] introduit l'extraction des motifs fréquents d'évolution à partir de données des STIS en mono et multi-bandes optiques et radar. Les motifs séquentiels fréquents, MSF, sont extraits sous la contrainte de support, c'est-à-dire que le nombre de pixels couverts par un motif dépasse un seuil établi par l'utilisateur.

Une approche d'extraction de motifs d'évolution de séries temporelles d'images satellites avec plusieurs bandes est présentée dans [62] et [63]. Reposant sur l'extraction de motifs séquentiels, la méthode a été spécifiquement conçue afin d'extraire des motifs d'évolutions non-majoritaires qui décrivent des changements. Les auteurs introduisent un seuil maximal de fréquence d'apparition, les motifs découverts ayant donc leur support compris dans un intervalle. Pendant le processus de découverte des motifs, ils éliminent les motifs contenant deux valeurs successives identiques sur une bande, étant intéressés surtout par les changements. Pour tous les exemples présentés ci-dessus, les relations de voisinage spatial des pixels ne sont pas utilisées, les pixels étant traités de façon indépendante.

Analyse au niveau objet Une solution pour pallier les difficultés associées à l'analyse basée sur les pixels peut être de fonctionner à l'échelle spatiale des objets d'intérêt. Un objet est défini comme une entité caractérisée par un ensemble de paramètres dont les valeurs ne se modifient pas dans les différents points qui appartiennent à l'entité considérée. Plus simplement, on peut dire que l'objet a la propriété d'uniformité des paramètres de définition. Un des paramètres les plus simples et les plus utilisés est la valeur du niveau de gris.

Par exemple, une approche basée sur l'objet diminue la possibilité de classer incorrectement les pixels individuels [32,4,65]. En effet, l'analyse au niveau objet est moins influencée par le bruit. Par la focalisation sur les objets du monde réel, les cartes produites de cette façon peuvent être plus directement utilisables par les analystes. L'analyse basée objets d'une séquence d'images exige la segmentation et la classification des images, le problème délicat étant la mise en correspondance des objets individuels dans le temps.

Dans [34] est présenté le cadre de la fouille de séquences d'images satellitaires météorologiques qui combine détection d'objets à partir des scènes et clustering des scènes. Les scènes sont d'abord classées automatiquement à l'aide de SOM (Self-Organizing Map) en deux étapes. Les images incluant des objets similaires en mouvement sont affectées au même cluster. Puis les scènes contenant des objets proéminents sont aussi regroupées. Les groupes sont examinés et étiquetés sémantiquement par l'expert et les séquences d'images sont transformées en une base de données de séquences d'identificateurs de groupes. Ces séquences sont scrutées en utilisant des fenêtres glissantes sous des contraintes temporelles (maximum de temps écoulé entre les signatures) et de fréquence d'apparition pour déterminer des dépendances temporelles fortes de type épisodes comme $A \rightarrow B$. Ceci peut être lu comme "si on observe la signature A une ou plusieurs fois, alors, plus tard, on observe une ou plusieurs fois la signature B". En effet, il y a une condensation de la séquence (par exemple $A \rightarrow A \rightarrow B \rightarrow B = A \rightarrow B$) qui conserve seulement les changements, le rythme étant perdu [33]. Après l'extraction d'images qui incluent des objets proéminents basés sur le résultat du clustering, les positions et les formes des objets sont approximées par un modèle de mélange gaussien à l'aide d'un algorithme de type EM (Expectation-Maximization). Les objets identiques entre les scènes successives sont reconnus et étiquetés. D'autres connaissances telles que les trajectoires d'objets sont extraites à partir des séries temporelles d'identificateurs de groupes et des informations sur les objets. Finalement, les connaissances extraites sont stockées dans une base de données, qui permet des requêtes de haut niveau via l'interface utilisateur, et ainsi la découverte de connaissances par les experts du domaine est facilitée.

Dans [47], est présentée une méthode de segmentation spatio-temporelle d'une STIS à haute résolution qui consiste dans un premier temps en un partitionnement spatial de chaque image et dans un second temps en une sélection temporelle d'instantants pertinents. La scène de la STIS est considérée comme constituée de plusieurs couches, et l'intérêt est concentré sur l'arrière-plan dont les objets sont immobiles mais évoluent radiométriquement. On propose une représentation de la dynamique de cet arrière-plan dans un graphe. Les noeuds de ce graphe représentent les objets de la scène, et les arcs les correspondances temporelles entre ces objets traduisant une relation de type "devenir". Une navigation dans ce graphe permet donc d'accéder à l'histoire des objets. L'analyse de la STIS se décompose en trois étapes : la segmentation des images, la construction du graphe, et un regroupement des évolutions radiométriques similaires. La méthode de multi-segmentation jointe permet de ne pas découpler totalement

la première et la deuxième étape. Une fois les objets extraits, les noeuds du graphe d'adjacence temporelle des objets sont déterminés, et les arcs peuvent alors être trouvés en estimant les correspondances entre objets. Une méthode permettant de regrouper de façon automatique les objets sous-jacents par similarité d'évolution radiométrique est proposée.

Dans [48], les auteurs proposent une méthode qui permet d'exploiter à la fois les structures spatiales des STIS extraites par segmentation spatiale, et l'information temporelle des évolutions des pixels. Les cartes d'évolutions de la STIS, basées sur l'analyse au niveau du pixel, peuvent raffiner des segmentations réalisées sur chaque image de la série. Ainsi, les segmentations peuvent être enrichies par des informations concernant les évolutions temporelles.

2.2 Méthodes usuelles d'analyse de séries temporelles d'images satellites

La raison d'être des méthodes supervisées est d'expliquer et/ou de prévoir un ou plusieurs phénomènes observables et effectivement mesurés. Concrètement, elles vont s'intéresser à une ou plusieurs variables de la base de données définies comme étant les cibles de l'analyse. Parmi les techniques développées dans ce but, on peut citer les techniques à base d'arbres de décision, les techniques statistiques de régressions linéaires ou non linéaires, les techniques à base de réseau de neurones (perceptron mono et multi couches, etc.), d'algorithmes génétiques, d'inférence bayésienne, etc.

Les méthodes non-supervisées permettent de travailler sur un ensemble de données dans lequel aucune des données ou des variables n'a une importance spéciale par rapport aux autres. Cela signifie qu'aucune variable n'est considérée individuellement comme la cible ou l'objectif de l'analyse. Ces méthodes sont souvent utilisées pour dégager un ensemble des groupes homogènes du point de vue des leurs caractéristiques (clustering).

Classification supervisée La classification consiste à examiner les caractéristiques d'une entité et lui attribuer une classe en supposant que certaines classes ou catégories ont été déjà prédéfinies. Dans le cas d'une STIS, l'objectif principal est d'identifier automatiquement pour chaque séquence d'entrée sa classe ou la catégorie correspondante.

L'article [26] présente un processus de classification collaborative multi-stratégie multi-étape de données complexes. L'aspect collaboratif multi-stratégie est basé sur un raffinement automatique et mutuel des résultats de plusieurs classifications. Les auteurs ont défini un concept de résolution des conflits pour représenter les dissensions de classification qui utilise un critère de similitude, basé sur le recouvrement des classes. Le résultat fourni est unique et la méthode peut intégrer différents types d'attributs (numériques, symboliques ou structurés). La méthode proposée est appliquée pour une classification au niveau pixel d'images de télédétection.

Dans [46] est proposée l'utilisation d'un modèle dynamique pour améliorer la classification de la couverture du sol sur une séquence d'images de télédétection.

L'approche consiste à représenter une parcelle de terre comme un système dynamique et à modéliser son évolution (en introduisant des connaissances sur les cycles des cultures) en utilisant le formalisme des automates temporisés. Afin d'affiner les résultats obtenus par un classificateur traditionnel, les observations données par une classification préliminaire des images sont combinées avec les états attendus fournis par une simulation avec un automate.

Clustering Contrairement à la classification supervisée, le clustering est utilisé lorsqu'il n'existe pas de données étiquetées, ce qui signifie que c'est une opération sans supervision.

Dans le cas du clustering, l'objectif est de regrouper les différentes entités dans des classes naturelles, des groupes (ou clusters) de sorte que les objets d'un même groupe soient aussi homogènes que possible et que deux groupes différents contiennent des objets suffisamment différents. Étant donné que chaque instance doit être semblable aux autres instances du même groupe, et dissemblable avec les instances des autres groupes, il est habituel que les méthodes de clustering fassent usage d'une mesure de similarité, afin d'identifier les groupes. Cette mesure de similarité, appelée aussi une fonction de distance, est indispensable pour effectuer le regroupement, mais peut être assez difficile à définir, en particulier en présence de types de données complexes.

Une des principales difficultés est de découvrir le nombre de clusters. Après cela, il est nécessaire d'identifier ces clusters, d'attribuer un nouveau label pour chacun d'eux et de découvrir leurs descriptions lorsque cela est nécessaire. De cette manière, le regroupement est en mesure d'identifier des régions ayant des caractéristiques différentes, ce qui contribue à définir la répartition globale des données. Des corrélations entre les attributs peuvent également être trouvées, ce qui peut aider à la tâche de pré-traitement de sélection de caractéristiques.

Dans [45], les auteurs traitent le problème de segmentation des images satellites multi-dates en identifiant des clusters qui contiennent des pixels qui évoluent similairement dans le temps. Ils introduisent une mesure de similarité des séquences en tenant compte uniquement des changements d'états. L'information sur le rythme n'est donc pas conservée et l'information spatiale n'est pas prise en compte mais ils soutiennent que l'évolution au sol est saisie. L'algorithme utilisé repose sur la stratégie d'alignement des séquences discrètes (distance d'édition). L'intégration de la mesure de similarité dans le cadre d'algorithmes classiques de clustering est également discutée dans l'article.

Dans [25], une approche originale pour le clustering de données multi-dimensionnelles est proposée. La méthode est basée sur l'estimation du nombre de groupes à partir de la construction d'un arbre couvrant minimal (MST, minimum spanning tree) avec l'algorithme de Prim, afin de fournir les paramètres d'initialisation de l'algorithme K-means classique. Les sommets sont supposés être répartis selon une distribution de Poisson et les mesures utilisées pour mesurer la similarité entre les points des données multi-dimensionnelles sont fondées sur des divergences informationnelles symétriques. Deux applications sont présentées en utilisant des mesures de réflectance à différentes longueurs d'onde. L'une porte

sur la classification taxonomique des astéroïdes et l'autre concerne la segmentation dans une image multi-spectrale. L'aspect spatial n'est pas pris en compte.

Extraction de motifs et détection de changements Dans une séquence d'images satellitaires, un grand nombre de travaux se sont focalisé sur la détection des changements entre deux images consécutives ou sur l'extraction des évolutions de la séquence entière. Les changements sont importants dans des applications telles que la surveillance de l'environnement et des forêts, le contrôle et la gestion de l'agriculture et de l'extension des zones urbaines.

La détection de changements est le processus d'identification d'états distincts d'une zone en l'observant à des dates différentes. Ce processus est un type particulier de classification dédiée à la discrimination de deux classes de zones : "avec changements liés à un phénomène d'intérêt" et "autres". Ainsi, l'analyse est supervisée dans le sens que le type de changements doit être précisé et se limite en général à des données provenant de deux dates particulières. Le résultat final est une carte des zones de changement. Les techniques de détection de changement ont généralement besoin de renseignements sur le type de changement qui doit être pris en compte. Par exemple, on peut vouloir chercher des changements brusques, comme les inondations, les tremblements de terre, ou les catastrophes anthropiques (e.g., [39]), ou on peut être intéressé par des changements progressifs tels que l'accumulation de la biomasse (e.g., [78]).

La détection de changements peut se décomposer en deux étapes distinctes : l'obtention d'indices de changements suivie par la discrimination des zones de changements. Ces indices peuvent caractériser des changements ponctuels, dans le voisinage du pixel ou à l'échelle des structures.

Dans la première catégorie la différence des attributs d'intérêt d'un même pixel entre deux dates est utilisée [17], [50]. Bien qu'elles nécessitent un recalage fin et qu'elles soient sensibles au bruit, les techniques de détection de changement appliquées au niveau du pixel sont efficaces, en particulier lorsque des changements de réflectance sont évalués entre deux images optiques. Dans [6], les auteurs utilisent une série temporelle relative à la végétation, et détectent les modifications en assignant à chaque pixel un score de changement. Ils proposent un algorithme récursif de fusion qui exploite les cycles annuels de végétation pour distinguer les points qui ont subi un changement des autres. La capacité de l'algorithme à ignorer les changements saisonniers naturels est particulièrement attractive. Parmi les limitations de l'algorithme, on peut mentionner qu'il n'utilise pas l'information spatiale qui est présente dans les données et qu'il ne découvre pas les motifs dominants dans les données. Une autre étude sur la détection de changements, qui utilise des données MODIS de haute résolution, est présentée dans [51]. La méthodologie de détection est basée sur des sommes annuelles d'un autre canal synthétique (l'indice NDVI) pour un pixel du sol donné. Un changement est détecté si le z-score de la différence des sommes annuelles est supérieur à un seuil.

La deuxième catégorie prend en compte les changements des interactions spatiales des pixels. En général, ce sont dans ce cas des méthodes de caractérisation

de texture sont employées. Par exemple, dans [49] est proposée une technique d'intégration des différences d'intensité et de texture entre deux images. La mesure de texture utilisée est robuste aux variations de bruit et d'illumination.

Concernant la troisième catégorie, les changements sont détectés sur les pixels qui changent de classe d'une image à l'autre. Par cette approche, on peut s'affranchir des problèmes d'étalonnage et de recalage des images mais le problème de l'extraction des structures d'intérêts subsiste (ce qui nécessite généralement des connaissances a priori). Une approche proposée dans [10], consiste à segmenter les deux images et fusionner les segmentations. Cette fusion consiste à obtenir une segmentation commune aux deux images, où chaque parcelle est homogène. Ensuite, chaque parcelle est caractérisée par un vecteur d'indices de changement qui permet d'évaluer les changements en restant à la résolution des parcelles. Cette méthode engendre des résultats détectant des zones de changements compactes et a pour avantage d'être robuste au bruit. Dans [5], les pixels sont regroupés en fonction de leur réflectance et de leur position pour trouver des objets. Les objets dont le comportement ne correspond pas à une référence stable sont sélectionnés.

On peut distinguer trois grandes familles méthodologiques pour l'analyse de changements [62]. Les méthodes bi-temporelles, permettent de situer et d'étudier des changements abrupts ayant lieu entre deux observations d'un phénomène à caractériser. Les méthodes correspondant à des techniques mixtes, principalement statistiques, s'appliquent généralement à deux images mais peuvent être combinées pour en analyser plusieurs. Les méthodes dédiées à l'étude de séries temporelles d'images sont généralement basées sur l'analyse de trajectoires radiométriques de pixels, afin de les comparer ou de détecter des ruptures.

Quel que soit le type de méthode de détection de changement utilisée pour l'analyse de STIS, il existe un décalage entre la quantité d'information que représentent ces séries temporelles, et la capacité des algorithmes à les analyser. Ces algorithmes sont le plus souvent dédiés à l'analyse bi-date d'une scène et se concentrent sur la cartographie des zones de changements et non sur leur caractérisation. Les méthodes bi-date sont de plus liées à des thématiques d'études spécifiques et sont incapables d'appréhender des changements ayant lieu au travers d'une STIS. Quant aux méthodes multi-dates, elles sont souvent difficilement interprétables et ne permettent pas de caractériser le changement.

Ces méthodes sont limitées par la dimension temporelle et ne permettent pas une extraction d'information pour la gestion d'une base de séquences multitemporelles d'images. En outre, ces méthodes sont appropriées pour les changements abrupts mais sont peu performantes en cas de changements progressifs qui s'opèrent sur plusieurs images. Elles n'extraient pas l'information d'évolution disséminée au long d'une STIS. Pour la fouille de données, les intérêts sont plus vastes. Dans ce cas, on peut exploiter la totalité des données fournies par les images, sans une sélection a priori, et on peut obtenir une caractérisation totale des évolutions observées.

Les méthodes d'extraction de motifs séquentiels [2,55,85,61] permettent d'identifier les ensembles de séquences ayant suivi la même évolution. De plus,

elles permettent de caractériser cette évolution, en fournissant le motif partagé par cet ensemble de séquences. L'extraction de motifs d'évolution fréquents à partir des STIS, introduite dans [42,43,41,44] utilise des données mono et multi-bande optiques et radar pour la météorologie et l'agriculture.

L'obtention de motifs montrant l'évolution temporelle des pixels des images peut aussi être utilisé pour la classification. Dans ce cas, les pixels montrant des motifs suffisamment semblables doivent être affectés à la même classe. Cela peut être fait :

- de manière *supervisée* : dans ce cas les classes doivent être définies par l'analyste a priori et un ensemble d'apprentissage doit être formé, constituant un ensemble de pixels pour lesquels la classe d'attribution est spécifiée.
- de manière *non supervisée* : aucune classe n'est définie a priori, les groupes sont générés par la technique elle-même. Les motifs/séquences sont regroupés selon leur similitude globale en groupes et le nombre de ces groupes peut être défini a priori par l'utilisateur ou déterminé par l'algorithme lui-même. La similitude entre motifs/séquences peut être définie de plusieurs façons. L'objectif est de construire une partition (dans laquelle chaque motif/séquence est assigné généralement à une classe) telle que les classes sont en interne aussi homogènes que possible. Leur signification est dérivée par une interprétation appropriée des résultats.

3 Extraction de motifs dans des bases de données spatio-temporelles (Spatio-temporal Pattern Mining)

La découverte de motifs ("Pattern Discovery") dans les données est l'un des problèmes phare en fouille de données. Il a été étudié intensivement avec, par exemple, les efforts en bioinformatique pour l'analyse de données génomiques à large échelle mais aussi dans de nombreux domaines d'applications où des régularités peuvent être porteuses de valeur ajoutée (par exemple, la découverte de règles d'associations dans des données transactionnelles, le calcul de motifs dans des bases de séquences ou des bases de graphes). En général, l'extraction est effectuée sous contraintes et il s'agit alors de ne calculer que les motifs qui satisfont une certaine contrainte dans les données (e.g., des règles d'association fréquentes et valides [1], des motifs séquentiels [2,54,56] satisfaisant éventuellement de plus une expression régulière [75], ou encore des graphes [40] et des arbres [74,86] fréquents). L'une des contraintes les plus étudiées est la contrainte de fréquence minimale mais bien d'autres formes de contraintes plus ou moins complexes ont été étudiées [31,16,72,69]. On sait aujourd'hui comment exploiter efficacement toute combinaison booléenne de contraintes primitives dites monotones ou anti-monotones. On trouve des synthèses sur ce champ d'étude dans les travaux consacrés aux bases de données inductives [38,7] et notamment dans [68] et [8] ainsi que dans les nombreuses publications produites par les projets européens (cInQ dans FP5, IQ dans FP6) ou ANR (e.g., Bingo2) dédiés à ces questions.

Face à l'explosion de la collecte/utilisation de l'information spatiale et des Systèmes d'Information Géographique (SIG), de nombreux travaux ont été réalisés dans le contexte de l'extraction de motifs spatio-temporels. Néanmoins, une majorité de ces travaux traitent souvent les dimensions spatiale et temporelle séparément. L'extraction de séries temporelles vise à repérer des caractéristiques fréquentes dans le temps [2,73,54,85,58] sans prendre en compte l'espace. La recherche de co-localisations [36,87,35,84,77,83] extrait un ensemble de caractéristiques apparaissant fréquemment dans des objets voisins sans prendre en compte le temps. Plus récemment, ces travaux ont été étendus pour intégrer simultanément les dimensions spatiale et temporelle. Une synthèse récente vient d'être éditée par le consortium GeoPKDD [28]. Cependant, dans ces travaux, les domaines de motifs utilisés et donc les motifs extraits ne sont pas à la hauteur de la complexité spatiale des objets à étudier dans des images satellitaires. De même, les contraintes primitives généralement étudiées (typiquement des fréquences minimales) restent insuffisantes pour caractériser les critères d'intérêt des experts comme les géologues.

De manière générale, nous pouvons distinguer deux types de travaux : ceux étudiant l'évolution spatio-temporelle d'objets mobiles, souvent assimilés au suivi de trajectoires, et ceux étudiant l'évolution spatio-temporelle des propriétés (ou caractéristiques) associées aux objets.

3.1 Suivi de trajectoires d'objets mobiles

L'émergence des nouvelles technologies mobiles (démocratisation des GPS et des téléphones mobiles) a entraîné la collecte de grandes quantités de données spatio-temporelles, permettant ainsi d'entrevoir de nouvelles applications. Par exemple, le projet GeoPKDD [28] a étudié l'aménagement du plan de circulation de grandes agglomérations en fonction des déplacements des véhicules. L'analyse des objets en mouvement a également comme domaines d'applications la géographie socio-économique, le sport (par exemple, les joueurs de football), l'analyse et le contrôle de la pêche, les prévisions météorologiques et l'analyse du mouvement (suivi des ouragans). Ces données relatives aux déplacements, et à la mobilité en général (mobility data), se présentent généralement sous la forme de bases de données de trajectoires. Dans [53,12,82,27,67], les auteurs définissent les trajectoires comme des objets en mouvement représentés par des séquences de tuples (l, t) , où l est la localisation de l'objet au temps t . Autrement dit, une trajectoire est une collection d'arrêts d'un même objet se déplaçant à différentes localisations spatiales. La figure 1 représente schématiquement une base de données de trajectoires.

Le nombre de trajectoires étant important, l'un des objectifs des méthodes d'extraction est de trouver les séquences les plus pertinentes (p.ex. les plus fréquentes). Face à cette problématique, plusieurs approches ont été proposées dans la littérature.

Par exemple, [53,12,13] se sont intéressés à l'extraction de motifs périodiques dans une base de données de trajectoires. Les objets étudiés (p.ex. des bus) ont la particularité de suivre approximativement la même route à intervalles de

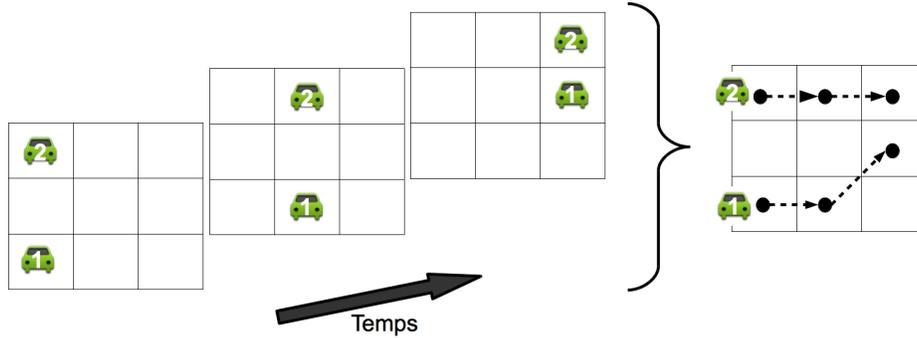


Figure 1. Exemple de base de données de trajectoires de véhicules

temps réguliers. Dans un premier temps, cette approche consiste à résumer l'ensemble des trajectoires d'un même objet par une seule séquence de segments. Les segments de trajectoires similaires sont regroupés en utilisant une fonction de similarité qui tient compte de la proximité spatiale, basée sur l'angle et la longueur spatiale des segments. Cette approche donne une meilleure abstraction des trajectoires et diminue la taille des données pour l'extraction. Contrairement à d'autres travaux, ces motifs intègrent aussi une notion de flou au niveau des localisations, ce qui permet d'extraire un motif même s'il ne se répète pas exactement au même endroit. Dans un second temps, un algorithme par niveaux est utilisé pour extraire les motifs fréquents. Cet algorithme dérivé d'*Apriori* a été optimisé grâce à l'utilisation d'une nouvelle structure de données (substring tree). Ce travail a été validé sur une base de données de trajectoires de bus, où chaque séquence correspondait aux déplacements d'un bus dans une journée.

Dans [21], les motifs étudiés sont des groupes d'objets partageant un type de mouvement (direction, vitesse) à une date donnée dans une certaine région de l'espace. Cinq types de motifs de trajectoire basés sur le mouvement, la direction et la localisation sont proposés (convergence, rencontre, troupeau, leadership et récurrence). Les travaux présentés dans [30] permettent de détecter les 4 premiers types de motifs définis dans [21] en utilisant des algorithmes de calcul approximatif. Les motifs spatio-temporels identifiés sont des sous-groupes d'objets ponctuels mobiles, avec des nombreux éléments localisés dans une région assez petite et présentant un mouvement similaire de point de vue de la direction, du but visé et/ou de la proximité.

Les auteurs présentent dans [59] une adaptation d'un algorithme de clustering basé sur la densité pour les trajectoires d'objets en mouvement. Ils s'appuient pour faire cela sur une notion de distance entre trajectoires. Ils mettent l'accent sur la dimension temporelle - essentiellement en élargissant l'espace de recherche des groupes intéressants en tenant compte des restrictions des trajectoires sources sur des sous-intervalles de temps. L'algorithme proposé vise à chercher les intervalles de temps les plus significatifs qui permettent d'isoler les groupes de qualité supérieure.

Dans [27], les auteurs proposent une extension du paradigme d'extraction de motifs séquentiels à l'analyse des trajectoires d'objets en mouvement. Ils introduisent les motifs de trajectoires comme des descriptions concises de comportements fréquents, en termes d'espace (les régions de l'espace visitées lors des déplacements) et de temps (la durée des déplacements). Ce travail est davantage axé sur des concepts de niveau supérieur (au lieu de découvrir un motif impliquant un endroit spatial précis, une localisation générale est trouvée). Ces localisations générales sont appelées régions d'intérêt (Regions-of-Interest ou RoI). Les motifs fréquents de déplacement entre ces régions sont découverts par la suite.

Les auteurs présentent dans [9] une approche automatique de haut-niveau pour la modélisation des connaissances spatio-temporelles à partir d'images satellitaires. Ils proposent d'utiliser une segmentation multi-approche comportant plusieurs méthodes de segmentation pour améliorer la modélisation et l'interprétation des images. Les expériences, sur deux scènes LANDSAT, montrent que leur approche surpasse les méthodes classiques de segmentation d'image et sont en mesure de prédire des changements spatio-temporels de couverture du sol.

3.2 Dynamique spatio-temporelle de phénomènes en fonction de leur environnement

Dans un certain nombre de travaux [76,80,79,14,15,37], l'objectif est différent de la fouille de trajectoires. En effet, il ne s'agit pas de suivre le déplacement d'objets précis en mouvement, mais d'étudier l'évolution et les interactions globales, dans l'espace et dans le temps, d'ensembles d'évènements en fonction de leur environnement. Un exemple d'application est l'étude de l'évolution des quartiers d'une ville en fonction de différents facteurs tels que le nombre de touristes, le type d'évènements socio-culturels, ou le type d'attractions à visiter à proximité. Dans cet exemple, il est évident que les quartiers ne se déplacent pas. Il ne s'agit donc pas d'étudier des trajectoires mais d'étudier comment évolue l'ensemble des quartiers en fonction de leurs caractéristiques et des évènements qu'il s'y produit. Afin d'illustrer un peu plus cette différence avec la fouille de trajectoires, nous allons prendre un autre exemple : celui de la propagation d'une virus tropical (p.ex. dengue ou chikungunya) dans les différents quartiers d'une ville. Le virus ne peut être associé à une trajectoire car il peut se propager à plusieurs endroits à la fois, tout en restant présent dans le quartier d'origine. Dans ce contexte, l'objectif est plutôt de rechercher comment se propage le virus de quartiers en quartiers en fonction de caractéristiques environnementales statiques (p.ex. le nombre de piscines ou le nombre d'espaces verts) ou dynamiques (p.ex. la météo).

Dans cette catégorie de méthodes, [14,15] ont généralisé le concept de co-localisations à des données spatio-temporelles. Initialement, les co-localisations (colocation) se focalisaient uniquement sur la dimension spatiale. Cette notion a été étudiée dans un grand nombre de travaux [36,84,77,66,22]. Ces motifs spatiaux sont des ensembles de propriétés (ou types d'évènements) fréquemment

associés à des objets spatiaux voisins (cad leurs instances forment des cliques). Le motif $\{mine, erosion, vegetation_faible\}$ pourrait être un exemple de co-localisation associé à des zones où l'érosion serait étudiée. Il représenterait le fait que les propriétés mine, érosion, et végétation_faible sont souvent dans des zones voisines. Autrement dit, il y aurait une corrélation spatiale entre ces éléments. Les co-localisations spatio-temporelles ont été définies comme une extension des co-localisations classiques afin de représenter des ensembles de propriétés associées à des objets voisins dans l'espace et dans le temps. Plus précisément, ces motifs sont tels que leurs instances sont spatialement proches pendant une fraction significative de temps. Ces contraintes ont été intégrées par l'intermédiaire d'une mesure d'intérêt monotone combinant prévalence spatiale et prévalence temporelle. Pour simplifier, seuls les motifs apparaissant un grand nombre de fois dans un grand nombre de temps sont conservés. Par exemple, dans la figure 2, les motifs $\{erosion, piste\}$ et $\{feux, vent, aireRepos\}$ sont des co-localisations spatio-temporelles (les traits en pointillés représentent la relation de voisinage). Toutefois, $\{feux, vent, aireRepos\}$ aura une mesure d'intérêt plus forte que $\{erosion, piste\}$ car il apparaît fréquemment dans un nombre de temps plus important ($\{erosion, piste\}$ n'apparaît qu'une seule fois à chaque temps). Pour extraire ces motifs, les auteurs ont notamment proposés une stratégie générer-tester et un parcours par niveaux de type *Apriori*.

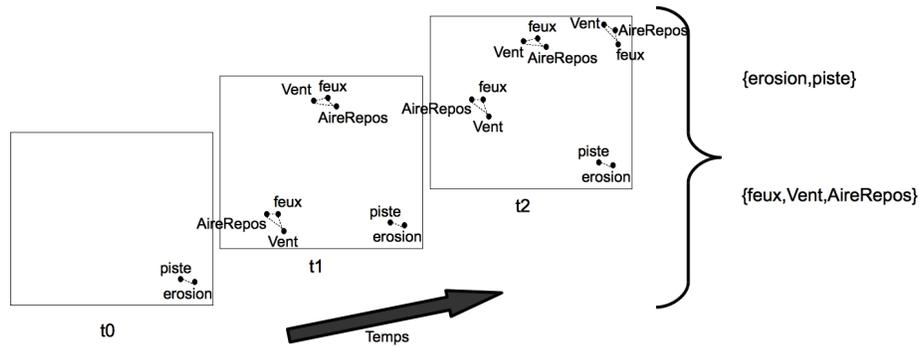


Figure 2. Exemple de co-localisations spatio-temporelles

Le concept de co-localisations a aussi été étudié dans d'autres travaux. Par exemple, les travaux de Qian et al. dans [67] se sont intéressés à l'extraction des SPCOZ (Spread patterns of spatio-temporal co-occurrences over zones). Ces motifs représentent la propagation (la "trajectoire") de co-localisations spatiales. Qian et al. ont ainsi étendu les travaux sur les co-localisations afin d'intégrer une nouvelle dimension temporelle. Plus précisément, ils suivent le déplacement d'éléments de propagation (spread element). Un élément de propagation est une co-localisation "fréquente" localisée et associée à une fenêtre temporelle. Dans la figure 3, la co-localisation fréquente $\{erosion, piste\}$ associée à l'intervalle $[t0, t1]$

est un élément de propagation. Les éléments de propagations combinés deux à deux constituent des arbres représentant la propagation d'un motif (SP-Tree ou Spread Pattern Tree). La figure 3 illustre deux exemples de motifs SPCOZ : le SP-tree de $\{erosion, piste\}$ et celui de $\{feux, vent, aireRepos\}$. L'algorithme d'extraction commence par rechercher toutes les co-localisations fréquentes de taille 2 (par une méthode de type *Apriori*), les utilise pour construire les éléments de propagation pour tous les temps, et génère les SP-Tree correspondants. A cette étape, les motifs construits sont donc des arbres de propagation composés de co-localisations de taille 2. Les autres arbres de propagation (ceux ayant des co-localisations de taille supérieure à 2) sont obtenus par une approche par niveaux de type Apriori, ie les arbres ayant des co-localisations de taille k sont obtenus à partir des sous-arbres composés des co-localisations de taille $k-1$.

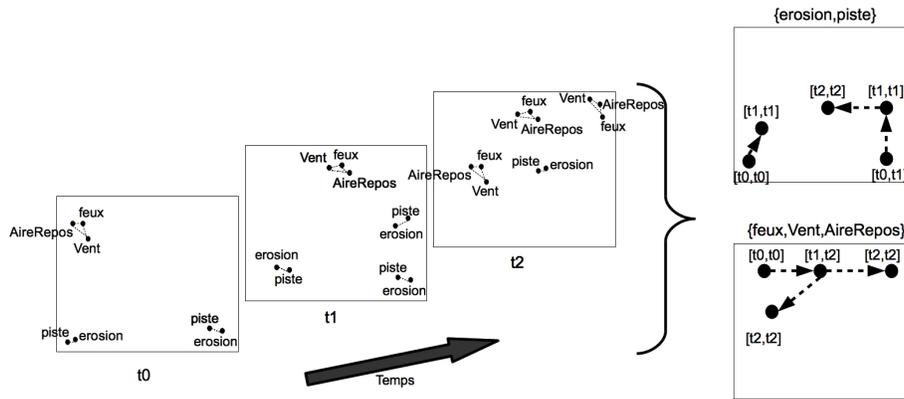


Figure 3. Exemple de SPCOZ

Dans [82], Yang et al. proposent un "framework" pour l'extraction de motifs spatiaux apparaissant fréquemment à différents temps, et une extension permettant de visualiser certaines évolutions. Ils ont validé leur approche sur un jeu de données étudiant l'évolution de molécules et de vortex. Les motifs étudiés, appelés SOAP (Spatial Object Association Pattern), peuvent être représentés sous la forme de graphes. Dans de tels graphes, chaque noeud correspond à une propriété, et chaque arête représente une relation de voisinage. Comme précisé par les auteurs, ce travail partage des similitudes avec les co-localisations car cette approche permet aussi d'extraire des ensembles de propriétés associées à des objets voisins, cad formant des cliques. Toutefois, à la différence de ces dernières, les auteurs considèrent des objets géométriques plutôt que des points pour les calculs de distances et de voisinage. De plus, ils permettent d'extraire trois autres types de configuration (cf exemples de la figure 4) : étoile, séquence et *minLink*. Ce dernier type permet de définir des SOAP plus généraux où seul le nombre minimum d'arêtes (*minLink*) associées à chaque noeud est fixé (cad le degré minimum des noeuds). Par exemple, si $minLink = 1$, tous les SOAP

de type étoile, clique ou séquence sont générés. A noter que pour les séquences, les arêtes représentent une relation de voisinage et de direction. Par exemple, une arête (x, y) représente "x est voisin et au dessus de y". Finalement, les auteurs montrent aussi comment utiliser (en post-traitement) les SOAP fréquents pour visualiser l'évolution d'un même ensemble de propriétés F . Pour cela, ils définissent la notion d'épisodes comme un ensemble d'instances associées à un intervalle de temps où le motif est apparu puis disparu. Les épisodes de tous les SOAP fréquents associé à F sont recherchés (tous types confondus), puis ordonnés en fonction de leur date d'apparition. La séquence ainsi générée permet de visualiser l'évolution spatiale (étoile, clique, ...) de l'ensemble de propriétés étudiés. Toutefois, ce post-traitement ne permet pas de prendre en compte les évolutions de forme des objets étudiés, ainsi que les éventuelles relations de cause à effet. En effet, seuls les changements globaux de "configuration" peuvent être observés. De plus, le nombre d'épisodes peut être important pour un même ensemble de propriétés, ce qui rend très difficile l'analyse de la séquence (comme le montre la figure 4). L'intégration d'informations supplémentaires telle que la météorologie est également délicate.

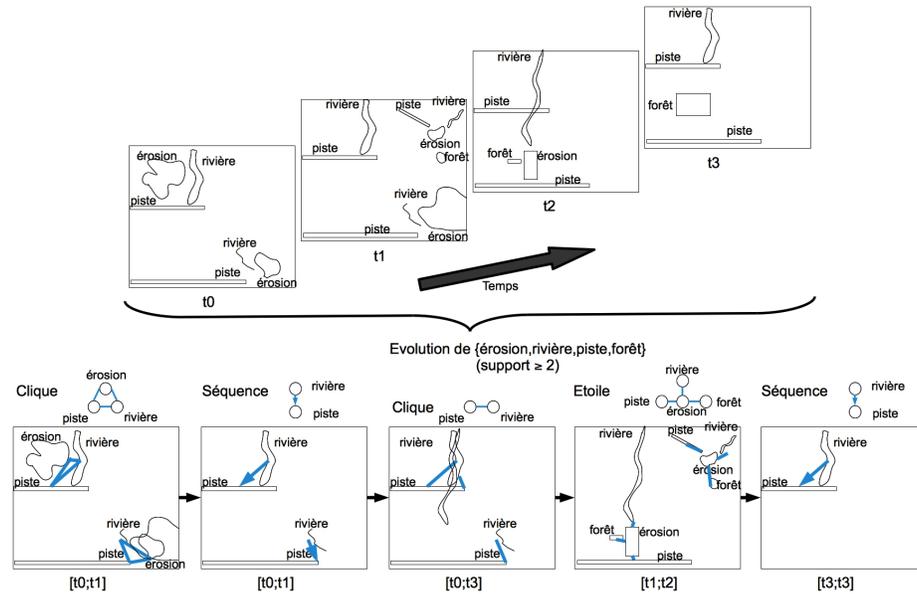


Figure 4. Exemple de SOAP fréquents et visualisation de l'évolution d'un ensemble de propriétés grâce à ces SOAP

De la même manière, les travaux sur l'extraction des séquences, utilisés initialement pour rechercher des régularités temporelles, ont été appliqués et étendus au spatio-temporel. Par exemple, Tsoukatos et al. dans [76] ont

étendu les travaux sur les séquences d'itemsets afin d'extraire des séquences représentant l'évolution dans le temps de zones d'études (p.ex. des quartiers). La base de données considérée est constituée de séquences d'itemsets (ensembles de caractéristiques environnementales) représentant l'évolution temporelle des différentes zones. Un algorithme effectuant un parcours en profondeur de l'espace de recherche est ensuite appliqué pour extraire les séquences les plus fréquentes (ie celles apparaissant dans le plus de zones). La figure 5 illustre un exemple de séquences pouvant être extraites. Les auteurs ont également proposé une approche pour extraire les séquences fréquentes à une granularité spatiale plus élevée (p.ex. région) en exploitant les séquences fréquentes trouvées à une granularité plus faible (p.ex. ville). Ils exploitent pour cela le fait que les séquences extraites à un niveau plus faible resteront fréquentes à un niveau de granularité plus élevé. La méthode recherche alors uniquement de nouvelles séquences fréquentes issues de l'agrégation spatiale.

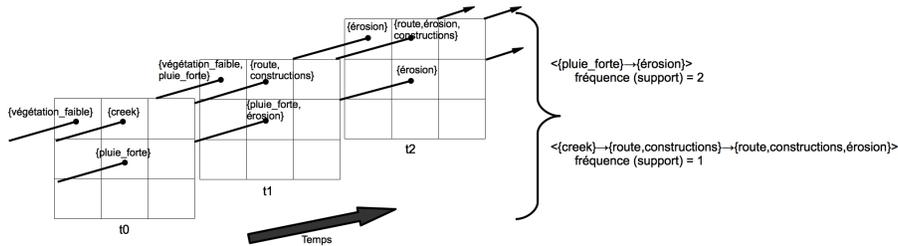


Figure 5. Exemple de séquences représentant l'évolution des zones

Les séquences, et plus généralement les graphes, ont aussi été utilisés pour représenter la *propagation* de phénomènes dans l'espace et dans le temps [52,37,80,79].

Dans [80], les auteurs se focalisent sur l'extraction de séquences représentant la propagation spatio-temporelle d'événements dans des fenêtres temporelles prédéfinies. Dans cet objectif, ils découpent la dimension temporelle en fenêtres d'une taille donnée (p.ex. 4 jours), divisent l'espace sous la forme d'une grille, et introduisent le concept de flow pattern. Un flow pattern est une séquence d'ensembles d'évènements de la forme $\langle E_1 \rightarrow E_2 \rightarrow \dots \rightarrow E_k \rangle$ où E_i est un ensemble d'évènements de la forme $e(\text{localisation})$, avec e un type d'évènements (p.ex. pluie, vent). Chaque ensemble d'évènements est composé d'évènements spatialement voisins apparaissant au même temps. Deux ensembles d'évènements E_p et E_q sont consécutifs dans la séquence, si leurs évènements appartiennent à la même fenêtre temporelle, s'ils sont tous voisins et qu'ils apparaissent à deux temps consécutifs. L'objectif de ce travail est de trouver les séquences d'évènements apparaissant fréquemment, cad celles apparaissant un grand nombre de fois suivant les mêmes localisations. La figure 6 montre un exemple de jeu de données et quelques flow patterns avec leur support

(dans cet exemple, deux évènements sont voisins si leur distance euclidienne est inférieure ou égale à 1). Les évènements sont localisés par des coordonnées (X,Y) tel que l'évènement pluie(0,1) au temps t0. A titre d'exemple, le flow pattern $\langle \{piste(0,0)\} \rightarrow \{erosion(1,0)\} \rangle$ apparaît 3 fois. A l'opposé, le motif $\langle \{piste(0,0),pluie(0,1)\} \rightarrow \{AirRepos(1,2),feux(1,2)\} \rangle$ à un support de 0 car les évènements $\{AirRepos(1,2),feux(1,2)\}$ ne sont pas voisins de tous les évènements de $\{piste(0,0),pluie(0,1)\}$ (AirRepos et feux du temps t1 ne sont pas voisins de piste du temps t0). Il est aussi intéressant de noter que les motifs $\langle \{piste(0,0),pluie(0,1)\} \rightarrow \{erosion(1,0)\} \rangle$ et $\langle \{piste(0,0),pluie(1,1)\} \rightarrow \{erosion(1,0)\} \rangle$ sont considérés comme deux motifs différents bien que représentant des phénomènes similaires (seule la localisation de l'évènement pluie est légèrement différente). Autre point important, le support du motif $\langle \{AirRepos(1,2)\} \rightarrow \{AirRepos(1,2),feux(1,2)\} \rightarrow \{feux(2,2)\} \rightarrow \{vegetation_faible(1,2)\} \rangle$ est égal à 0 car il n'est pas inclus dans une unique fenêtre temporelle. Pour extraire ces motifs, les auteurs adaptent des techniques d'extraction de motifs séquentiels classiques. Plus précisément, leur algorithme applique une stratégie par niveaux pour trouver les séquences de taille 1 et 2, puis utilisent les motifs fréquents trouvés comme point de départ à un parcours en profondeur de l'espace de recherche.

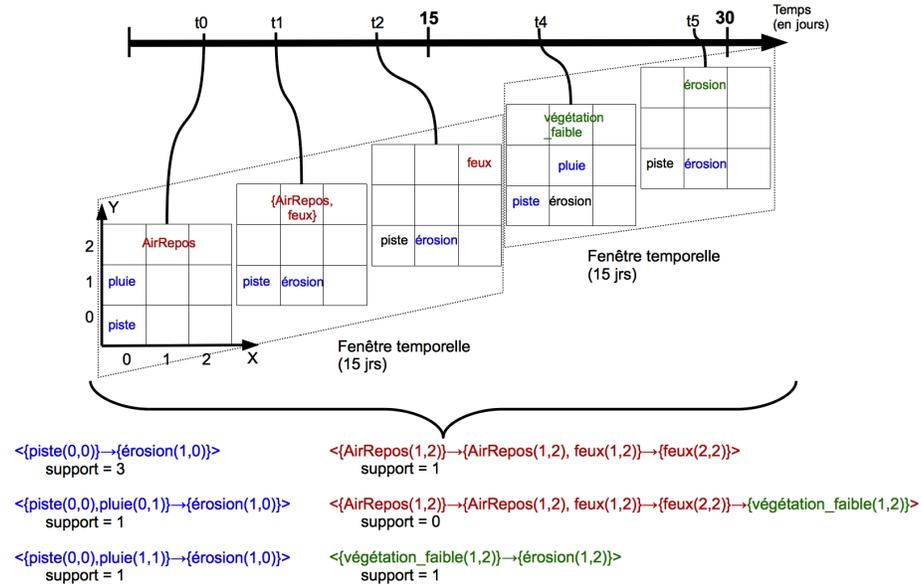


Figure 6. Exemple de flow patterns

Dans un deuxième temps, Wang et al. [79], étendent cette notion et définissent les motifs spatio-temporels généralisés (Generalized spatio-temporal pattern)

comme des séquences de relative eventsets. Un relative eventset est un ensemble d'évènements dont la localisation est remplacée par un positionnement relatif à une localisation de référence. Un motif spatio-temporel généralisé est fréquent s'il a au moins t -minsup (support temporel) occurrences dans le temps et qu'il a au moins s -minsup (support spatial) occurrences dans l'espace (les localisations peuvent être différentes mais la localisation relative doit être identique). Pour extraire ces motifs, les auteurs proposent un nouvel algorithme appelé GenST-Miner qui utilise une approche dérivée de Prefspan [58].

Huang et al., dans [37], se sont concentrés sur le problème d'extraction de séquences de propriétés représentant la propagation de certains types d'évènements. Ces séquences sont de la forme $\langle f_1 \rightarrow f_2 \rightarrow \dots \rightarrow f_k \rangle$, où f_i est un type d'évènements (cf exemple de la figure 7). Cette approche permet donc d'étudier la propagation des évènements pris individuellement (sans prendre en compte leur environnement). Ce modèle considère deux évènements consécutifs s'ils sont spatialement proches (distance euclidienne inférieure à un seuil donné) et apparaissent dans la même fenêtre temporelle. Les auteurs ont également étudié d'autres relations de voisinage dépendant du temps. Ces relations permettent de représenter un rétrécissement de la zone d'influence d'un évènement (son voisinage) au fur et à mesure que le temps passe (c'est ce qui passe lors de la propagation d'une maladie infectieuse). Les auteurs proposent aussi une nouvelle mesure d'intérêt pour ces séquences car pour eux les mesures basées sur le support ne reflètent pas nécessairement un lien de cause à effet entre les évènements. Leur mesure, appelée *sequence index*, s'appuie sur des travaux antérieurs [19] exploitant des statistiques spatiales pour étudier l'indépendance de phénomènes. Ainsi, une séquence est intéressante si le ratio de densité spatial entre deux éléments consécutifs de cette séquence est supérieur à un seuil. Ce ratio de densité est proche de 1 si les deux types d'évènements sont distribués de manière indépendante. Plus la mesure est inférieure à 1, plus les types d'évènements se "repoussent". Au contraire, plus ce ratio est au dessus de 1, plus les chances d'avoir un lien de cause à effet entre ceux-ci sont élevées. La principale limite de cette mesure est de ne pas être anti-monotone. Or, ce type de propriétés est un aspect important pour le passage à l'échelle d'une grande partie des algorithmes d'extraction de motifs. Pour extraire ces séquences, les auteurs proposent donc un nouvel algorithme *Slicing - STS - Miner* basé sur un traitement incrémental des différentes fenêtres temporelles et une extension des séquences à chaque étape. Au lieu de s'appuyer sur l'anti-monotonie, ils exploitent une autre propriété du *sequence index* : si une séquence est intéressante, toutes les sous-séquences ayant le même préfixe sont intéressantes.

Mohan et al. dans [57] étendent ces travaux en étudiant des graphes orientés acycliques de types d'évènements. Tout comme précédemment, leurs motifs, appelés motifs spatio-temporels en cascade, ne permettent pas de prendre en compte l'environnement proche d'un évènement. Dans l'exemple de la figure 7, le voisinage entre *feux* et *piste* n'est donc pas considéré. Par contre, la première et la dernière séquences se retrouvent fusionnées, montrant ainsi que $\langle \text{pluie} \rightarrow \text{erosion} \rangle$ est précédée par *piste* et *feux* (mais ils ne sont pas

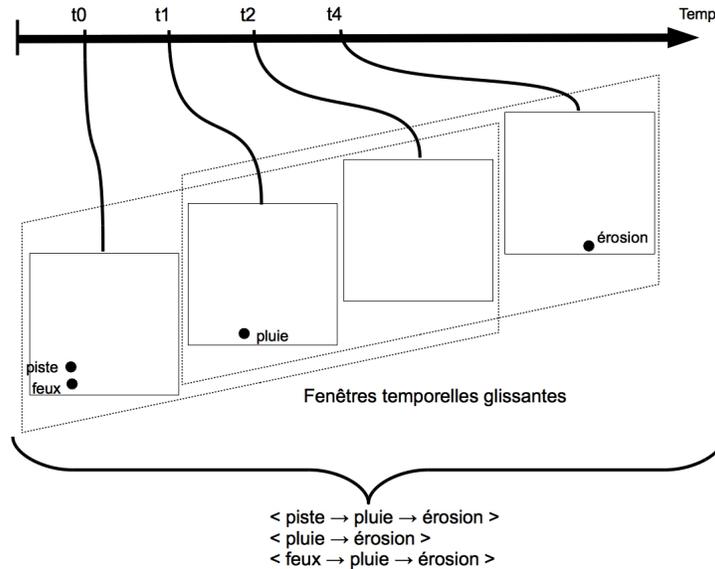


Figure 7. Exemple de séquences représentant la propagation de type d'évènements (sans prendre en compte leur environnement)

nécessairement voisins). Ils introduisent aussi une nouvelle mesure de intérêt appelée cascade participation index construite à partir de la mesure proposée dans [36]. Cette mesure est anti-monotone. Cette propriété est utilisée dans un algorithme par niveaux de type *Apriori* pour extraire les motifs les plus intéressants.

Références

1. R. Agrawal, T. Imielinski, and A. Swami. Mining Association Rules between Sets of Items in Large Databases. In P. Buneman and S. Jajodia, editors, *Proceedings of the ACM SIGMOD International Conference on Management Data*, pages 207–216, Washington, DC, USA, 1993. ACM Press.
2. R. Agrawal and R. Srikant. Mining sequential patterns. In Philip S. Yu and Arbee S. P. Chen, editors, *Proc. of the 11th International Conference on Data Engineering (ICDE'95)*, pages 3–14, Taipei, Taiwan, 1995. IEEE Computer Society Press.
3. P. Aplin, P.M. Atkinson, and P.J. Curran. Fine spatial resolution satellite sensors for the next decade. *International Journal of Remote Sensing*, 18 :1387–1381, 1997.
4. P. Aplin and G.M. Smith. Advances in object-based image classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII, 2008. Part B7.
5. S. Bontemps, P. Bogaert, N. Titeux, and P. Defourny. An object-based change detection method accounting for temporal dependences in time series with medium to coarse spatial resolution. *Remote Sensing of Environment*, 112 :3181–3191, 2008.

6. S. Boriah, V. Kumar, M. Steinbach, C. Potter, and S. Klooster. Land cover change detection : a case study. In *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*, pages 857–865, New York, NY, USA, 2008. ACM.
7. Jean-François Boulicaut, Mika Klemettinen, and Heikki Mannila. Modeling kdd processes within the inductive database framework. In *DaWaK*, pages 293–302, 1999.
8. Jean-François Boulicaut, Luc De Raedt, and Heikki Mannila, editors. *Constraint-Based Mining and Inductive Databases, European Workshop on Inductive Databases and Constraint Based Mining, Hinterzarten, Germany, March 11-13, 2004, Revised Selected Papers*, volume 3848 of *Lecture Notes in Computer Science*. Springer, 2005.
9. Wadii Boulila, Imed Riadh Farah, Karim Saheb Ettabaa, Basel Solaiman, and Henda Ben Ghézala. Spatio-temporal modeling for knowledge discovery in satellite image databases. In *Proceedings of the 7th French Information Retrieval Conference, COnférence en Recherche d'Infomations et Applications - CORIA 2010, Sousse, Tunisia, March 18-20, 2010*, pages 35–49. Centre de Publication Universitaire, 2010.
10. L. Bruzzone and D. Fernández Prieto. Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3) :1171–1182, May 2000.
11. J. B. Campbell. *Introduction to Remote Sensing*. The Guilford Press, New York, USA, third edition, 2002.
12. Huiping Cao, Nikos Mamoulis, and David W. Cheung. Mining frequent spatio-temporal sequential patterns. In *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM '05)*, pages 82–89, Washington, DC, USA, 2005. IEEE Computer Society.
13. Huiping Cao, Nikos Mamoulis, and David W. Cheung. Discovery of Periodic Patterns in Spatiotemporal Sequences. *IEEE Transactions on Knowledge and Data Engineering*, 19(4) :453–467, 2007.
14. Mete Celik, Shashi Shekhar, James P. Rogers, and James A. Shine. Sustained emerging spatio-temporal co-occurrence pattern mining : A summary of results. In *ICTAI*, pages 106–115, 2006.
15. Mete Celik, Shashi Shekhar, James P. Rogers, and James A. Shine. Mixed-drove spatiotemporal co-occurrence pattern mining. *IEEE Trans. Knowl. Data Eng.*, 20(10) :1322–1335, 2008.
16. Loïc Cerf, Jérémy Besson, Céline Robardet, and Jean-François Boulicaut. Data peeler : Constraint-based closed pattern mining in n-ary relations. In *SDM*, pages 37–48, 2008.
17. P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin. Digital change detection methods in ecosystem monitoring : a review. *International Journal of Remote Sensing*, 25(9) :1565–1596, May 2004.
18. P.F. Crapper. An estimate of the number of boundary cells in a mapped landscape coded to grid cells. *Photogrammetric Engineering and Remote Sensing*, 50 :1497–1503, 1984.
19. Noel A. C. Cressie. *Statistics for Spatial Data (Wiley Series in Probability and Statistics)*. Wiley-Interscience, rev sub edition, January 1993.

20. P. Fisher. The pixel : A snare and a delusion. *International Journal of Remote Sensing*, 18 :679–685, 1997.
21. Peter Fisher, Patrick Laube, Marc Kreveld, and Stephan Imfeld. Finding REMO - Detecting Relative Motion Patterns in Geospatial Lifelines. In *Developments in Spatial Data Handling*, pages 201–215. Springer Berlin Heidelberg, 2005.
22. Frédéric Flouvat, Nazha Selmaoui-Folcher, Dominique Gay, Isabelle Rouet, and Chloé Grison. Constrained colocation mining : application to soil erosion characterization. In *SAC*, 2010.
23. G.M. Foody. Fully fuzzy supervised classification of land cover from remotely sensed imagery with an artificial neural network. *Neural Computing and Applications*, 5 :238–247, 1997.
24. G.M. Foody. The Continuum of Classification Fuzziness in Thematic Mapping. *Photogrammetric Engineering and Remote Sensing*, 65(4) :443–451, 1999.
25. L. Gallucio, O. Michel, and P. Comon. Unsupervised clustering on multi-components datasets : Applications on images and astrophysics data. In *16th European Signal Processing Conference EUSIPCO-2008*, pages 25–29, Lausanne, Switzerland, August 2008.
26. P. Gançarski and C. Wemmert. Collaborative multi-strategy classification : application to per-pixel analysis of images. In *Proc. of the 6th International Workshop on Multimedia Data Mining : mining integrated media and complex data*, pages 15–22, 2005.
27. Fosca Giannotti, Mirco Nanni, Fabio Pinelli, and Dino Pedreschi. Trajectory pattern mining. In Pavel Berkhin, Rich Caruana, and Xindong Wu, editors, *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, California, USA, August 12-15, 2007*, pages 330–339. ACM, 2007.
28. Fosca Giannotti and Dino Pedreschi, editors. *Mobility, Data Mining and Privacy - Geographic Knowledge Discovery*. Springer, 2008.
29. Silvio Griguolo. Classification on Sets of Remotely-Sensed Images : a Vegetation Monitoring Model. In E. Binaghi, P.A. Brivio, and A. Rampini, editors, *Soft Computing in Remote Sensing Data Analysis*, pages 235–244. World Scientific, Singapore, 1996.
30. Joachim Gudmundsson, Marc J. van Kreveld, and Bettina Speckmann. Efficient detection of motion patterns in spatio-temporal data sets. In Dieter Pfoser, Isabel F. Cruz, and Marc Ronthaler, editors, *Proceedings of the 12th ACM International Workshop on Geographic Information Systems, ACM-GIS 2004, November 12-13, 2004, Washington, DC, USA*, pages 250–257. ACM, 2004.
31. Jiawei Han and Micheline Kamber. *Data Mining, Second Edition : Concepts and Techniques*. Morgan Kaufmann, January 2006.
32. G.J. Hay and G. Castilla. Object-based image analysis : Strengths, weaknesses, opportunities and threats (SWOT). In *Proceedings of the 1st International Symposium on Object-based Image Analysis (OBIA 2006), Salzburg University, Austria, July 4-5, 2006*, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS), 2006.
33. R. Honda and O. Konishi. Temporal rule discovery for time-series satellite images and integration with RDB. In *Proceedings of the 5th European Conference on*

- Principles of Data Mining and Knowledge Discovery (PKDD '01)*, pages 204–215, London, UK, 2001. Springer-Verlag.
34. R. Honda, Shuai Wang, Tokio Kikuchi, and Osamu Konishi. Mining of moving objects from time-series images and its application to satellite weather imagery. *J. Intell. Inf. Syst.*, 19 :79–93, July 2002.
 35. Yan Huang, Jian Pei, and Hui Xiong. Mining co-location patterns with rare events from spatial data sets. *GeoInformatica*, 10(3) :239–260, 2006.
 36. Yan Huang, Shashi Shekhar, and Hui Xiong. Discovering colocation patterns from spatial data sets : A general approach. *IEEE Trans. Knowl. Data Eng.*, 16(12) :1472–1485, 2004.
 37. Yan Huang, Liqin Zhang, and Pusheng Zhang. A framework for mining sequential patterns from spatio-temporal event data sets. *IEEE Transactions on Knowledge and Data Engineering*, 20(4) :433–448, April 2008.
 38. Tomasz Imielinski and Heikki Mannila. A database perspective on knowledge discovery. *Commun. ACM*, 39(11) :58–64, 1996.
 39. J. Inglada, J.-C. Favard, H. Yesou, S. Clandillon, and C. Bestault. Lava flow mapping during the Nyiragongo January, 2002 eruption over the city of Goma (D.R. Congo) in the frame of the international charter space and major disasters. In *Proc. of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS'03)*, volume 3, pages 1540–1542, July 2003.
 40. Akihiro Inokuchi, Takashi Washio, and Hiroshi Motoda. An apriori-based algorithm for mining frequent substructures from graph data. In *PKDD*, pages 13–23, 2000.
 41. A. Julea, N. Méger, and P. Bolon. On mining pixel based evolution classes in satellite image time series. In *Proc. of the 5th Conference on Image Information Mining : pursuing automation of geospatial intelligence for environment and security (ESA-EUSC 2008)*, ESRIN - Frascati, Italy, March 2008.
 42. A. Julea, N. Méger, and E. Trouvé. On mining METEOSAT and ERS multitemporal images. In *Proc. of the 4th Conference on Image Information Mining for Security and Intelligence (ESA-EUSC 2006)*, Madrid, Spain, November 2006.
 43. A. Julea, N. Méger, and E. Trouvé. Sequential patterns extraction in multitemporal satellite images. In *10th European Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD'06), Practical Data Mining Workshop : Applications, Experiences and Challenges*, pages 94–97, Berlin, Germany, September 2006.
 44. A. Julea, N. Méger, E. Trouvé, and P. Bolon. On extracting evolutions from satellite image time series. In *Proc. of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2008)*, volume 5, pages 228–231, Boston, MA, USA, 2008.
 45. A. Ketterlin and P. Gançarski. Sequence similarity and multi-date image segmentation. In *4th Intl Workshop on the Analysis of Multitemporal Remote Sensing Images*, Leuven, Belgique, July 2007.
 46. C. Largouët and M.-O. Cordier. Improving the landcover classification using domain knowledge. *AI Communications*, 14(1) :35–43, 2001.
 47. C. Le Men. *Segmentation spatio-temporelle d'une séquence temporelle d'images satellitaires à haute résolution*. PhD thesis, Télécom ParisTech, École Nationale Supérieure des Télécommunications, Paris, France, Septembre 2009.

48. C. Le Men, A. Julea, N. Méger, M. Datcu, P. Bolon, and H. Maître. Radiometric evolution classification in high resolution satellite image time series (SITS). In *Proc. of the 5th Conference on Image Information Mining : pursuing automation of geospatial intelligence for environment and security (ESA-EUSC 2008)*, ESRIN - Frascati, Italy, March 2008.
49. L. Li and M.K.H. Leung. Robust change detection by fusing intensity and texture differences. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)*, 2001.
50. D. Lu, P. Mausel, E. Brondizio, and E. Moran. Change detection techniques. *Intl. J. of Remote Sensing*, 25(12) :2365–2407, June 2004.
51. R. S. Lunetta, J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy. Land-cover change detection using multi-temporal MODIS NDVI data. *Remote Sensing of Environment*, 105(2) :142–154, 2006.
52. Loïc Mabit, Nazha Selmaoui-Folcher, and Frédéric Flouvat. Modélisation de la dynamique de phénomènes spatio-temporels par des séquences de motifs. In *EGC*, pages 455–466, 2011.
53. Nikos Mamoulis, Huiping Cao, George Kollios, Marios Hadjieleftheriou, Yufei Tao, and David W. Cheung. Mining, indexing, and querying historical spatiotemporal data. *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*, page 236, 2004.
54. H. Mannila, H. Toivonen, and A. Inkeri Verkamo. Discovery of frequent episodes in event sequences. *Data Min. Knowl. Discov.*, 1(3) :259–289, 1997.
55. F. Massegli, F. Cathala, and P. Poncelet. The PSP approach for mining sequential patterns. In *Proc. of the 2nd European Symposium on Principles of Data Mining and Knowledge Discovery in Databases (PKDD'98)*, volume 1510, pages 176–184, Nantes, France, September 1998. LNAI, Springer Verlag.
56. Florent Massegli, Fabienne Cathala, and Pascal Poncelet. The psp approach for mining sequential patterns. In *PKDD*, pages 176–184, 1998.
57. Pradeep Mohan, Shashi Shekhar, James A. Shine, and James P. Rogers. Cascading spatio-temporal pattern discovery : A summary of results. In *SDM*, pages 327–338, 2010.
58. B. Mortazavi-Asl, H. Pinto, and U. Dayal. PrefixSpan : mining sequential patterns efficiently by prefix-projected pattern growth. *Proceedings 17th International Conference on Data Engineering*, pages 215–224, 2000.
59. M. Nanni and D. Pedreschi. Time-focused clustering of trajectories of moving objects. *J. Intell. Inf. Syst.*, 27(3) :267–289, November 2006.
60. E. Nezry, G. Genovese, G. Solaas, and S. Rémondrière. ERS - Based early estimation of crop areas in Europe during winter 1994-95. In Guyenne T.-D., editor, *ERS Application, Proceedings of the Second International Workshop held 6-8 December 1995 in London*, volume 383 of *ESA Special Publication*, page 13, 1996.
61. J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu. Mining Sequential Patterns by Pattern-Growth : The PrefixSpan Approach. *IEEE Transactions on Knowledge and Data Engineering*, 16(10) :1424–1440, 2004.
62. François Petitjean, Pierre Gançarski, and Florent Massegli. Extraction de motifs d'évolution dans les séries temporelles d'images satellites. In *Spatial Analysis and GEOmatics*, November 2010.

63. François Petitjean, Pierre Gançarski, Florent Masseglia, and Germain Forestier. Analysing satellite image time series by means of pattern mining. In Colin Fyfe et al., editor, *11th International Conference on Intelligent Data Engineering and Automated Learning*, volume 6283 of *Lecture Notes in Computer Science*, pages 45–52. Springer, 2010.
64. N. Pettorelli, J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20 :503–510, 2005.
65. R.V. Platt and L. Rapoza. An evaluation of an object-oriented paradigm for land use/land cover classification. *Professional Geographer*, 60 :87–100, 2008.
66. Feng Qian, Q. He, and Jiangfeng He. Mining Spatial Co-location Patterns with Dynamic Neighborhood Constraint. *Machine Learning and Knowledge Discovery in Databases*, pages 238–253, 2009.
67. Feng Qian, Qinming He, and Jiangfeng He. Mining spread patterns of spatio-temporal co-occurrences over zones. In *ICCSA (2)*, pages 677–692, 2009.
68. Luc De Raedt. A perspective on inductive databases. *SIGKDD Explorations*, 4(2) :69–77, 2002.
69. Luc De Raedt, Tias Guns, and Siegfried Nijssen. Constraint programming for itemset mining. In *KDD*, pages 204–212, 2008.
70. L.S.M. Schouten, H.J.C. van Leeuwen, E.H. van Valkengoed, J.F. Desprats, C. King, N. Baghdadi, L. Prévot, N. Bruguier, M. Dechambre, and R. Valentin. Land use classification based on time series of micro-wave data (ERS, Radarsat). study in framework of the project ReSeDa - Assimilation of Multisensor & Multi-temporal Remote Sensing Data to Monitor Soil & Vegetation Functioning, 2000.
71. R.A. Schowengerdt. Soft classification and spatial-spectral mixing. In E. Binaghi, P.A. Brivio, , and A. Rampini, editors, *Soft Computing in Remote Sensing Data Analysis*, pages 1–6. World Scientific, Singapore, 1996.
72. A. Soulet and B. Crémilleux. Optimizing constraint-based mining by automatically relaxing constraints. In *The Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 777–780, Houston, USA, 2005. IEEE Computer Society.
73. Ramakrishnan Srikant and Rakesh Agrawal. Mining quantitative association rules in large relational tables. In H. V. Jagadish and Inderpal Singh Mumick, editors, *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, Montreal, Quebec, Canada, June 4-6, 1996*, pages 1–12. ACM Press, 1996.
74. Alexandre Termier, Marie-Christine Rousset, and Michèle Sebag. Treefinder : a first step towards xml data mining. In *ICDM*, pages 450–457, 2002.
75. Roberto Trasarti, Francesco Bonchi, and Bart Goethals. Sequence mining automata : A new technique for mining frequent sequences under regular expressions. In *ICDM*, pages 1061–1066, 2008.
76. Ilias Tsoukatos and Dimitrios Gunopulos. Efficient mining of spatiotemporal patterns. In Christian S. Jensen, Markus Schneider, Bernhard Seeger, and Vassilis J. Tsotras, editors, *Proceedings of the 7th International Symposium, Advances in Spatial and Temporal Databases (SSTD 2001), Redondo Beach, CA, USA, July 12-15, 2001*, volume 2121 of *Lecture Notes in Computer Science*, pages 425–442. Springer, 2001.
77. Florian Verhein and Ghazi Al-Naymat. Fast mining of complex spatial co-location patterns using glimit. In *ICDM Workshops*, pages 679–684, 2007.

78. A. Viña, F. R. Echavarría, and D. C. Rundquist. Satellite change detection analysis of deforestation rates and patterns along the Colombia-Ecuador border. *AMBIO : A Journal of the Human Environment*, 33 :118–125, 2004.
79. Junmei Wang, Wynne Hsu, and Mong-Li Lee. Mining generalized spatio-temporal patterns. In *DASFAA*, pages 649–661, 2005.
80. Junmei Wang, Wynne Hsu, Mong-Li Lee, and Jason Tsong-Li Wang. Flowminer : Finding flow patterns in spatio-temporal databases. In *ICTAI*, pages 14–21, 2004.
81. T.A. Warner and M. Shank. An evaluation of the potential for fuzzy classification of multispectral data using artificial neural networks. *Photogrammetric Engineering and Remote Sensing*, 63 :1285–1294, 1997.
82. Hui Yang, Srinivasan Parthasarathy, and Sameep Mehta. A generalized framework for mining spatio-temporal patterns in scientific data. In *KDD*, pages 716–721, 2005.
83. Jin Soung Yoo and Mark Bow. Finding n-most prevalent colocated event sets. In *DaWaK*, pages 415–427, 2009.
84. Jin Soung Yoo and Shashi Shekhar. A joinless approach for mining spatial colocation patterns. *IEEE Trans. Knowl. Data Eng.*, 18(10) :1323–1337, 2006.
85. M.J. Zaki. SPADE : an efficient algorithm for mining frequent sequences. *Machine Learning Journal, Special issue on Unsupervised Learning*, 42(1/2) :31–60, Jan/Feb 2001.
86. Mohammed Javeed Zaki. Efficiently mining frequent trees in a forest. In *KDD*, pages 71–80, 2002.
87. Xin Zhang, Nikos Mamoulis, David W. Cheung, and Yutao Shou. Fast mining of spatial collocations. In *KDD*, pages 384–393, 2004.