

SUIVI DE L'EROSION PAR IMAGERIE SATELLITE ET FOUILLE DE DONNEES SPATIO-TEMPORELLES

EROSION MONITORING BY SATELLITE IMAGE ANALYSIS AND SPATIOTEMPORAL DATA MINING

Catherine Pothier ⁽¹⁾, Rémi Andreoli⁽¹⁾, Nicolas Méger ⁽³⁾, Christophe Rigotti ⁽⁴⁾,

1 : Univ Lyon, INSA-Lyon, LGCIE - SMS ID, F-69621, France

2 : Bluecham S.A., 101 rue Roger Laroque, 98807 Nouméa, Nouvelle-Calédonie, France

3 : LISTIC, Université Savoie Mont Blanc, 74940 ANNECY LE VIEUX Cedex - France

4 : Univ Lyon, INSA-Lyon, CNRS, INRIA, LIRIS, UMR5205, F-69621, France

RÉSUMÉ – Cet article présente une méthode de fouille de données permettant d'analyser des séries temporelles d'images satellite. La méthode est utilisée ici pour la mise en évidence de phénomènes liés à l'érosion dans une zone minière de Nouvelle-Calédonie. Les résultats sont ensuite interprétés à l'aide du hub géospatial Qëhnelö™.

ABSTRACT – This paper presents a data mining method to analyze satellite image time series. The method is applied to exhibit phenomena related to erosion over a mining area in New Caledonia. The results are interpreted using the Qëhnelö™ geospatial hub.

1. Introduction

Les Séries Temporelles d'Images Satellites (STIS) contiennent des grandes quantités d'information multidimensionnelles et présentent un fort potentiel, source de compréhension de la dynamique de surface de la Terre. Notamment en ce qui concerne le suivi de l'érosion des sols, ainsi que sa prévention. La Nouvelle-Calédonie est un archipel mélanésien de l'ouest de l'Océan Pacifique. Elle est caractérisée par son patrimoine et sa richesse naturelle (lagon inscrit au patrimoine mondial de l'UNESCO, végétation endémique) mais concentre les principales problématiques des pays insulaires du Pacifique : aléas cycloniques annuels, risque de tsunami locaux, régionaux et lointains (Mimoura, 1999). De plus, des épisodes de précipitations intenses récurrents entraînent l'érosion des sols dont la vulnérabilité est accentuée par les feux et les activités humaines dont l'activité minière passée et récente. Riches en métaux lourds, les produits de l'érosion transitent jusqu'aux mangroves et au lagon en causant parfois, lors d'événements paroxystiques, une dégradation de la qualité des eaux des rivières et le blanchiment du corail.

Afin d'extraire des informations pertinentes de ces ensembles de données que constituent les STIS, différentes techniques de fouilles de données peuvent être envisagées. Récemment, une approche provenant des techniques dites par "extraction de motifs", basée au niveau des pixels et sans *a priori* a été (Julea *et al.*, 2011) proposée. Elle consiste à extraire les évolutions ou sous- évolutions affectant des pixels suffisamment connectés les uns aux autres et couvrant une surface minimum de la scène. Ces évolutions pixelaires sont nommées motifs séquentiels fréquents groupés (GFS-patterns). Cette méthode a donné des résultats prometteurs à la fois pour des séries temporelles d'images satellites optiques et radar. Les résultats sont présentés sous forme de cartes de localisation spatio-temporelle (STL-maps), une carte étant construite pour chaque GFS-pattern. Dans ces cartes, l'information spatiale est donnée par des positions de pixels et l'information temporelle est affichée en utilisant une palette de

couleurs. Une sélection des cartes les plus intéressantes s'appuyant sur une technique de randomisation (Méger *et al.*, 2015) permet de construire in fine un aperçu concis de l'information contenue dans les séries d'images satellitaires. Cette méthode a été appliquée sur une série de 16 images Landsat 7 prises entre 2000 et 2011 dans une zone au sud-est de la Nouvelle-Calédonie caractérisée par un site minier d'extraction et de transformation de classe mondiale afin de mettre en évidence les zones d'érosions dues à l'exploitation de la mine ainsi que les évolutions temporelles des dolines et du maquis minier afin de surveiller et prévoir le risque érosif. Dans cet article, après une présentation de la méthode d'extraction de connaissance, la préparation des images utilisées pour analyser l'érosion sera décrite, puis les résultats obtenus et les interprétations des motifs effectuées à l'aide du hub géospatial Qèhnelö™ seront détaillés.

2. GFS-pattern, un outil de création de cartes d'intérêts.

L'extraction de GFS-patterns se fait depuis des données symboliques. Il est donc nécessaire de remplacer toutes les valeurs de la STIS par un symbole. Pour ce faire, l'intervalle des valeurs atteintes dans la STIS est découpé en plusieurs sous-intervalles. Chaque sous-intervalle est alors associé à un symbole. Cette opération de quantification/discrétisation permet de construire la STIS symbolique associée à la STIS numérique étudiée. L'utilisation de percentiles est la plus courante, elle est d'ailleurs incluse dans le prototype d'extraction des GFS-patterns. La quantification se fait alors en divisant l'histogramme des valeurs en zones d'aires égales. L'intervalle des valeurs est alors divisé en n sous-intervalles associés à des symboles. La valeur quantifiée peut être une vitesse, une valeur d'intensité lumineuse, une phase radar ou un canal construit depuis les valeurs brutes de télédétection.

L'ensemble de toutes les sous-évolutions possibles au niveau de chaque pixel est ensuite extrait, sans aucune contrainte sur les dates d'occurrences des symboles. En particulier, aucune contiguïté temporelle entre occurrences de symboles n'est exigée. Sont ainsi obtenus plusieurs milliers de motifs séquentiels, et leur nombre doit donc être réduit en ne sélectionnant que ceux qui peuvent présenter un intérêt. Parmi ces motifs sont sélectionnés ceux qui reflètent des formes de signaux les plus détaillées, appelées sous-évolutions maximales (à savoir, les sous-évolutions qui ne sont pas incluses dans les autres sous-évolutions qui ont été retenues). Cependant, nous avons encore souvent à portée de main plusieurs centaines de ces motifs maximaux.

Deux mesures ont été introduites pour sélectionner les sous-évolutions pertinentes : la mesure de surface et la mesure de la connectivité. Une sous-évolution pixellaire n'est conservée que si elle couvre un nombre suffisant de pixels, par exemple, au moins 2% des pixels de la STIS. L'utilisateur peut donc en fonction de la zone étudiée choisir la valeur du paramètre σ de surface minimale devant être couverte par le motif séquentiel par rapport à la surface de l'image. Un tel motif est dit motif séquentiel fréquent. La seconde mesure introduite est la connectivité. Chaque pixel possède (sauf ceux au bord de l'image) 8 pixels voisins. La connectivité moyenne d'une sous-évolution est le nombre moyen de pixels couverts par cette sous-évolution parmi ces 8 voisins. Un motif séquentiel fréquent est alors appelé un motif Séquentiel Fréquent Groupé (GFS-patterns) si et seulement si sa connexité moyenne est supérieure ou égale à un seuil k fixé.

Chaque motif est représenté à l'aide d'une carte de localisation spatio-temporelle colorée (STL-map). Chaque pixel de l'image portant la sous-évolution est coloré sur la carte à son emplacement. Le pixel reste noir si le signal pour ce pixel ne correspond pas à la sous-évolution tracée. Sinon, la couleur indique la date où intervient le dernier élément de la sous-évolution pour ce pixel. L'échelle de couleur est donnée figure 1, du rouge pour la date de la première image de la SITS au magenta pour la date de la dernière image. Les cartes les plus intéressantes, c'est-à-dire celles ne pouvant pas être

retrouvées sur une STIS randomisée ainsi que les cartes n'étant pas modifiées par une telle randomisation sont sélectionnées afin de construire un résumé (Méger *et al.*, 2015) permettant aux experts d'accéder à une version synthétique du contenu informationnel de la STIS étudiée.



Figure 1. Palette de couleurs.

3. Présentation du jeu de données

Dans cette partie, la méthode précédente de fouille de données a été appliquée sur un jeu de données d'images optiques sur une zone côtière du sud de la Nouvelle-Calédonie (communes de Yaté et du Mont-Dore). Ce secteur est caractérisé par une forte nébulosité durant toute l'année. Les ensembles de données et leurs prétraitements seront présentés puis les réglages des paramètres donnés et enfin les résultats obtenus détaillés.

3.1. Présentation de la série temporelle d'image satellite.

La série temporelle d'images satellite est une série d'images optiques, construite à partir des acquisitions faites par le satellite Landsat 7 géré par l'USGS et la NASA (Service géologique et agence spatiale américaines). Elle couvre la côte sud-est de la Nouvelle-Calédonie dans la région de Yaté. Cette région présente d'une part une forte activité minière et d'autre part un site classé au patrimoine mondial de l'UNESCO, le Grand Lagon Sud. Cette série contient 16 images multispectrales (513x513 pixels) acquises entre 2000 et 2011, à une résolution spatiale de 30 m, et avec une présence importante de nuages sur au moins la moitié de la série (figure 2). En plus de la forte présence de nuages, cette série comporte deux autres sources de difficultés. D'une part, les dates d'acquisition des images indiquent des intervalles de temps très différents entre chaque acquisition, et d'autre part, certaines images sont touchées par des défauts de capteurs et par des artefacts (défaut du capteur Landsat-7 à partir de 2003). La méthode de fouille peut toutefois être mise en œuvre sans aucun traitement supplémentaire pour éliminer les nuages, les sources de dégradation de la qualité de l'image et l'intervalle de temps irrégulier.

La cible d'étude de cette zone étant l'évolution de l'érosion et de la végétation, une bande synthétique donnant l'indice de végétation normalisé (NDVI) a été calculée. Son principe repose sur le fait qu'un couvert végétal absorbe l'énergie dans la bande rouge et réfléchit beaucoup dans le proche infra-rouge (Lillesand et Kiefer, 2000). Cet indice a été conçu pour exprimer la présence ou la disparition de biomasse. Il permet aussi dans cette étude de mettre en évidence l'augmentation de la rougeur du sol. Les valeurs de NDVI varient entre -1.0 (absence de la biomasse) à +1,0 (forte présence de la biomasse). Cette bande synthétique a été quantifiée en 3 niveaux basés sur le 33e et le 66e centiles, et ce pour chaque image prise séparément, ce qui permet d'absorber les variations d'exposition/luminosité. Un indice NDVI de faible valeur est alors représenté par le symbole «1», par le symbole «2» pour une valeur moyenne, et par le symbole «3» pour une valeur haute.

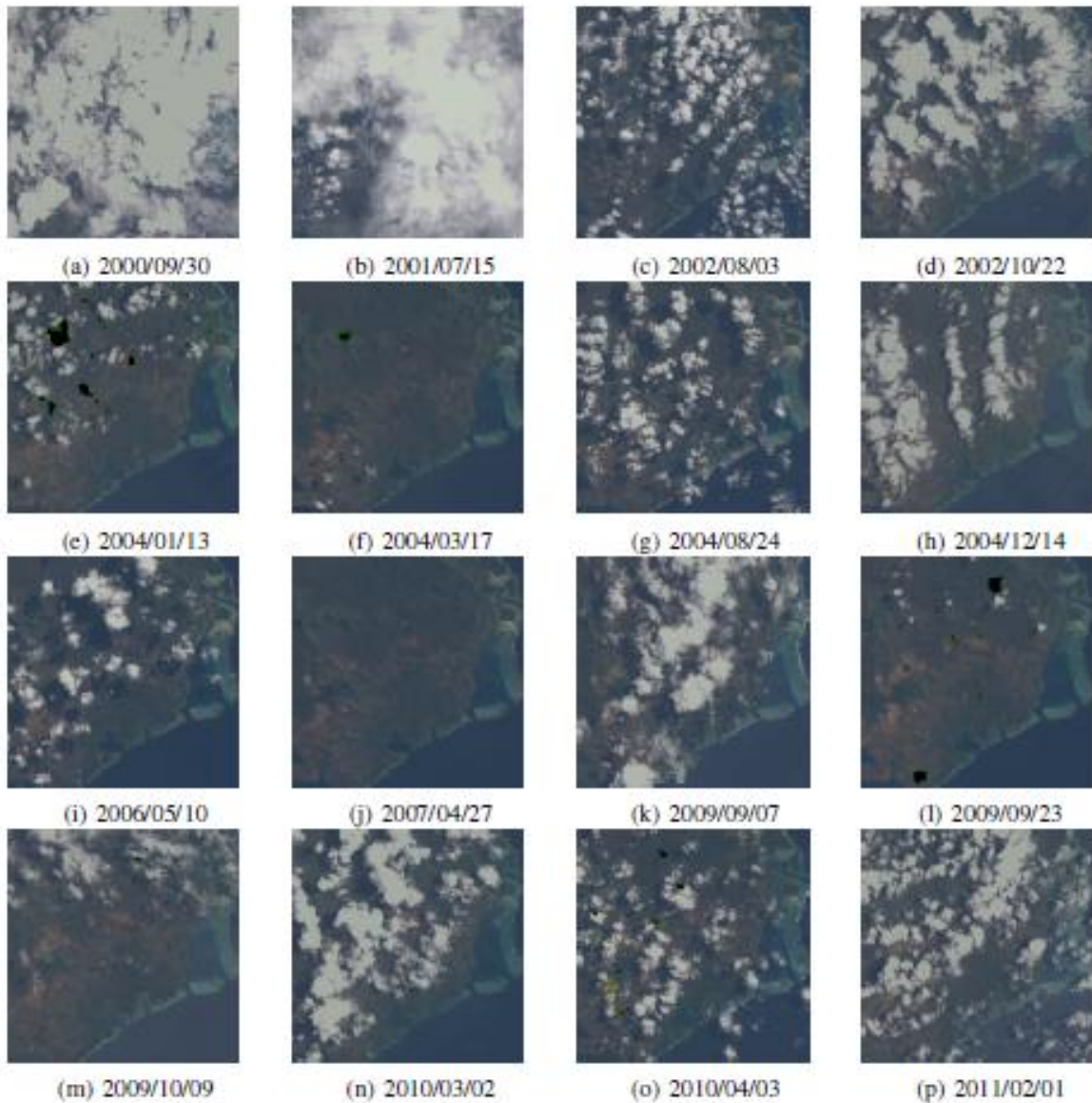


Figure 2. Série temporelle d'images satellite Landsat 7 (représentation RGB).

3.2 Réglage des paramètres

Les paramètres de réglage de l'algorithme sont le seuil de connectivité et la surface minimum, ils sont définis en fonction de la nature des phénomènes d'intérêts recherchés dans la série d'images satellite. Ainsi, afin d'obtenir une certaine cohérence spatiale des phénomènes d'intérêt dans les STL-maps, il a été imposé qu'en moyenne un pixel support d'un motif possède plus de la moitié de ses 8 voisins présentant le même schéma d'évolution. Dans notre cas d'étude, le seuil de connectivité moyenne minimale est égal à 5. Le seuil minimum de couverture est fixé à 2 % afin d'obtenir la plus grande collection possible de STL-map, soit 297 pour notre cas d'étude.

Les différentes étapes du procédé d'extraction ont été développés en C et Python 2.7 et sont exécutables sur une plate-forme informatique standard (Intel Core i7, 2,5 GHz). L'ensemble des traitements comprenant la quantification de la série d'images, l'extraction des GFS-patterns et le tracé des cartes nécessite moins de cinq minutes.

4. Résultats et discussion

Pour cette série d'images satellite, les cartes les plus intéressantes extraites sont présentées figures 3 à 5. En plus des références citées dans le texte, l'interprétation des motifs retenus a été faite au cours de deux visites sur le terrain des auteurs, et en utilisant le hub géospatial Qèhnelö™ de la Mairie de Yaté (utilisateurs finaux des suivis environnementaux sur le secteur).

La première STL-map (figures 3a et 3b) est celle du motif 2,2,1,1,1,2. Ce motif indique une baisse de l'indice NDVI (passage du niveau 2 au niveau 1) durant la série puis sa hausse (niveau 1 au niveau 2). Ce motif décrit une décroissance puis une reprise de la biomasse ou une variation de la rougeur du sol. Les couleurs roses et violettes montrent que ses changements ont lieu plutôt vers la fin de la série, la palette de couleur (figure 1) permettant de savoir la date de fin du motif. La zone (0) correspond à un lac (Grand Lac) potentiellement impacté par des infiltrations de boue et dont le niveau varie au fil du temps (sècheresse, variation du niveau de l'eau du pseudo-karst). Les zones (1) (2) et (3) mettent en évidence la création d'une carrière et une zone minière avec gradient d'extension ainsi que la réalisation du centre technique et nouvelle mine. La zone notée (4) correspond à un décanteur des boues rouges faisant varier la valeur NDVI, et le dégradé de couleur met en évidence l'ouverture puis l'évolution de la construction du décanteur vers le nord-est au fil du temps. Enfin les zones (5) et (6) correspondent aux installations de traitement du minerai et à la base vie construite pendant la période étudiée. Lors de leur utilisation, ces sites d'activité anthropique dégagent de la chaleur faisant aussi varier l'indice NDVI. Le trait de côte apparaît de façon quasi-continue tout le long suite aux variations de niveaux de la mer (ensablement, évolution du platier). Sur la partie droite en haut de la figure se dessine en bleu l'embouchure de la rivière Kwé Binyi, principale rivière du sud de la Nouvelle-Calédonie, où se déposent des boues sédimentaires rouges issues de l'érosion de la zone (Bird *et al.*, 1984). Proche de l'embouchure de la Kwé Binyi est aussi mise en évidence en vert une variation de végétation de type herbiers en eau peu profonde.

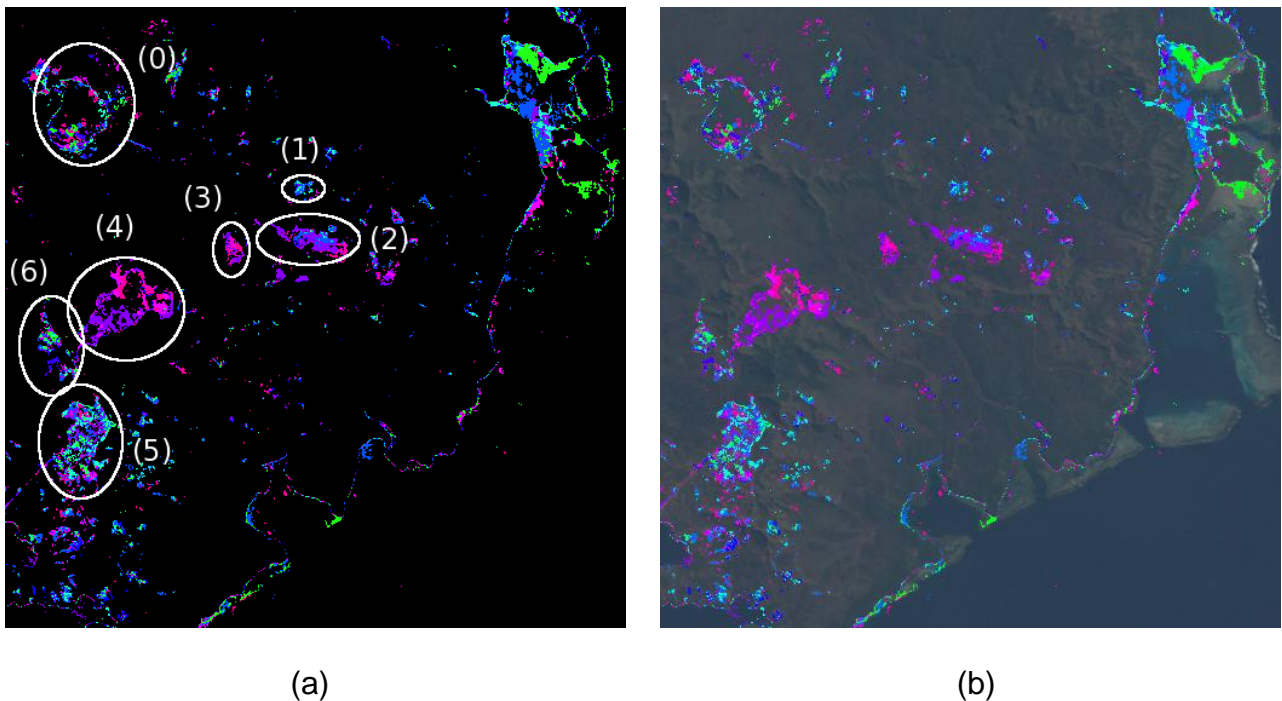


Figure 3. STL-map motif 2,2,1,1,1,2 (a) seul et (b) superposé sur image RGB de la zone.

Une autre STL-map, correspondant au motif 2,2,3,2,2,2,3 est présentée figure 4. Cette carte reflète un des phénomènes les plus marquants de la zone, la présence du maquis sur les rochers ultramafiques de Nouvelle-Calédonie (Proctor, 2003) qui comprend principalement des conifères à feuilles persistantes. Sur cette partie de la côte, ce maquis se compose essentiellement des espèces endémiques dont des Araucaria, des Gymnostoma, des Arillastrum gummiferum, des fougères arborescentes et un cortège de plantes spécifiques ligneuses et arbustives. La variation de sa biomasse au fil du temps se traduit par la variation de NDVI entre les niveaux moyen et élevé. Les zones en noir cerclées de blanc mettent en évidence des zones sans végétation correspondant à la présence des installations minières décrites préalablement, et à celle du lac. Des repères sous la forme d'étoiles au niveau du trait de côte indiquent des zones où se développe la forêt humide littorale typique de Nouvelle-Calédonie dont la canopée est dominée par les Araucaria columnaris ou pins colonnaires. Ces forêts stables ne présentent pas de variation d'indice de biomasse et ne sont donc pas couvertes par le motif 2,2,3,2,2,2,3. Les autres parties noires à l'intérieur du maquis minier traduisent à la fois des zones de biomasse différentes (comme des repousses après un feu) ou des pertes de végétations par exploitation minières ou érosion.

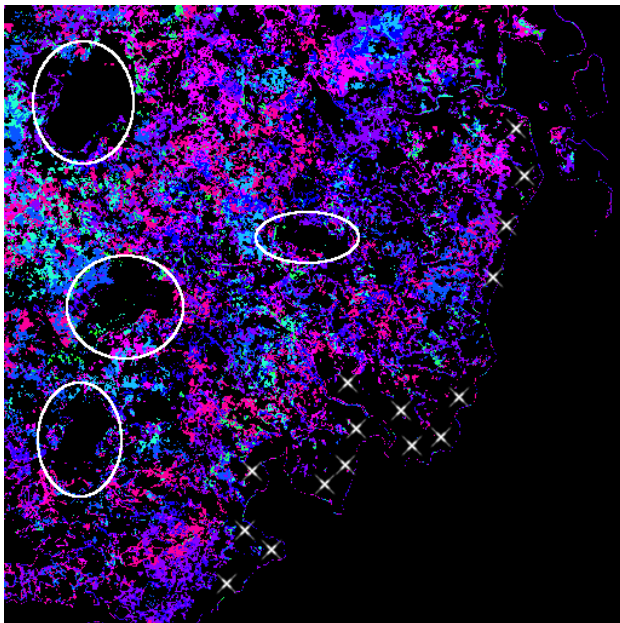


Figure 4. STL-map motif 2,2,3,2,2,2,3.

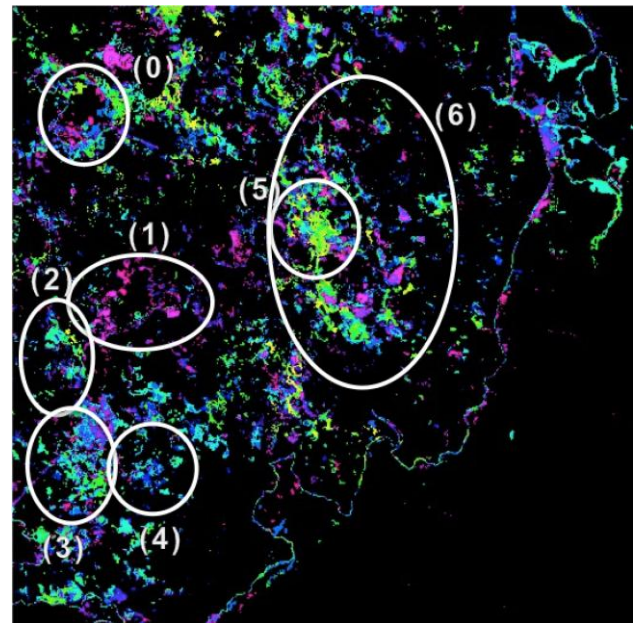


Figure 5. STL-map motif 1,1,2,2.

Enfin la figure 5 présente la STL-map du motif 1,1,2,2 qui correspond à une croissance de la biomasse ou à une modification de la teinte rouge d'une zone. En haut à gauche de la figure, le Grand Lac (0) apparaît avec les variations de son niveau faisant apparaître plus ou moins la végétation au niveau de son bord. Sur la partie gauche de la figure, le déversoir des boues (1), la base de vie (2), l'installation de traitement du minerai (3) sont mis en évidence. En dessous, la reprise de la végétation suite à un incendie le 13/12/2008 apparaît au niveau de la zone (4), sans cette reprise de végétation le climat de cette zone (cyclones) entraîne rapidement l'érosion des sols à nu. Au centre de la figure la zone cerclée (5) correspond au début de l'exploitation de la mine et à l'érosion créée dans les zones périphériques à cette exploitation (modification de la rougeur du sol) et plus au nord de cette zone, des changements de végétation. Les dolines présentes dans la zone cerclée (6) apparaissent différemment suivant leur nature, les dolines temporaires disparaissent en début de série (taches de couleur jaune/vert) quant aux dolines persistantes (couleur rose/violet) elles restent présentes tout au long de la série.

Qèhnelö™ est le hub géospatial en ligne (<http://www.bluecham.net/qehnelo/qehnelo-concept/>) qui a été développé pour répondre aux attentes des décideurs et des populations quant au suivi des territoires au sens large. Avec un simple navigateur Internet, Qèhnelö™ donne accès à l'ensemble des informations relatives à un territoire (données satellitaires, données géographiques, documents textes et *crowdsourcing*). Son architecture modulaire permet de s'adapter de manière précise à de multiples demandes. Une plateforme Qèhnelö™ est un ensemble d'applications (ou logiciels/modules) auquel les utilisateurs accèdent de manière sécurisée et transparente par Internet. Ces applications utilisent des ressources en ligne (bases de données géographiques, bibliothèques, annuaires, modèles mathématiques) et offrent des services à haute valeur ajoutée (cartographie rapide, analyse du territoire, suivi environnemental, prévision, études d'impact) (Lille D., 2007). Ce dispositif permet de faciliter l'accès aux données spatiales, tout en leur donnant plus de sens contextuel adapté aux différents experts et utilisateurs (pompiers, techniciens, décideurs, scientifiques) (Andreoli *et al.* 2015).

Les résultats de la fouille de données réalisée à partir de l'extraction de GFS-patterns sont intégrés au hub géospatial Qèhnelö™ utilisé par la mairie de Yaté (figure 6). Le suivi de l'érosion, de la qualité de l'eau, des feux de brousses est primordial pour les populations de la Nouvelle-Calédonie. En effet, le patrimoine naturel se mélange souvent au patrimoine culturel (lieux tabous), l'eau des rivières est directement captée pour la consommation des tribus et la qualité des eaux du lagon influe sur les pêches. Les informations délivrées contribuent ainsi à répondre aux questionnements des décideurs locaux tels que : quel est l'étendue spatiale des impacts ? D'où proviennent les sédiments transportés par les rivières ?

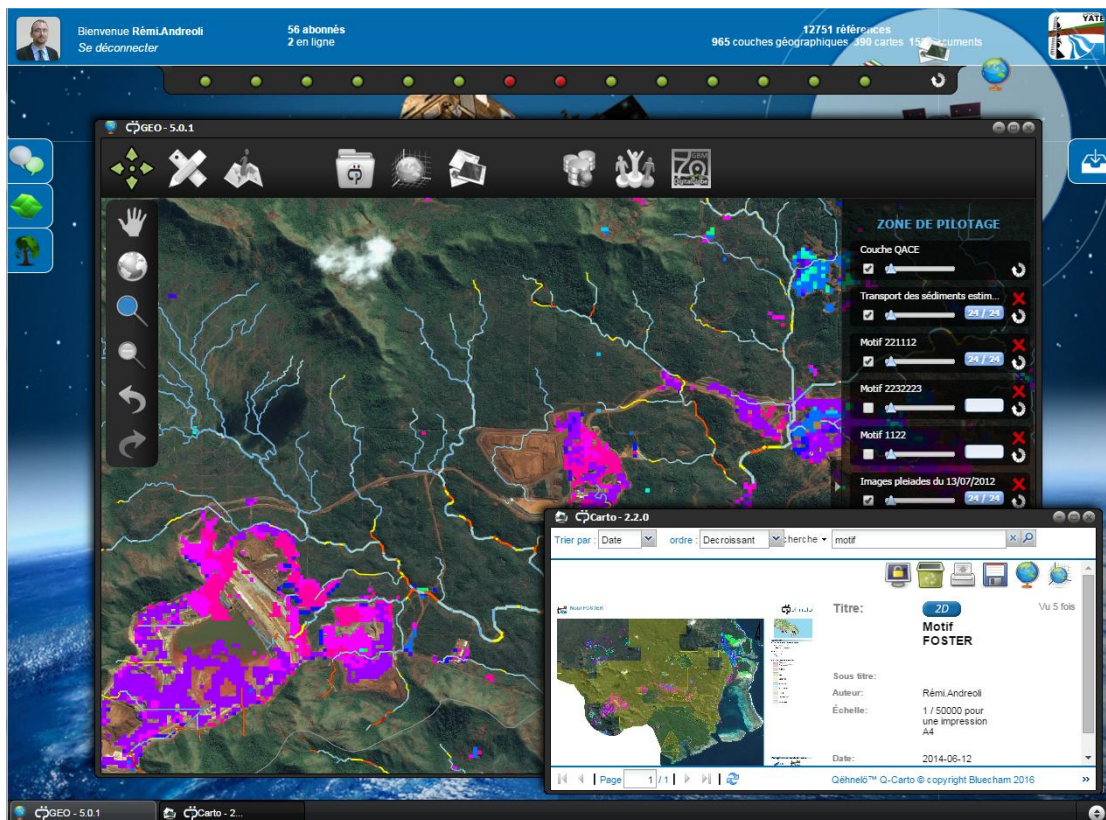


Figure 6 : Motif 2,2,1,1,1,2 superposé à l'analyse du transport sédimentaire des rivières dans le hub géospatial Qèhnelö™ de la mairie de Yaté.

4. Conclusion

Une méthode pour synthétiser sous la forme d'un jeu de cartes représentant des phénomènes remarquables survenus dans une série spatio-temporelle d'images satellite a été exposée. Cette méthode a été utilisée sur un jeu d'images optiques issues des acquisitions du satellite Landsat 7 sur la zone de la côte Sud-Est de la Nouvelle-Calédonie. La bande synthétique NDVI permettant de voir l'évolution de la biomasse et la variation de la rougeur du sol a été employée. Les résultats ont mis en évidence l'évolution de la végétation dont celle du maquis minier ultramafique dans la zone concernée, et des phénomènes liés à l'érosion. Le suivi de la reprise de la végétation post-incendie ou après exploitation minière est primordiale pour limiter l'érosion des sols ultramafiques très sensibles à ce phénomène en raison du climat local. Cette méthode basée sur une technique de fouille de données spatio-temporelles de séries d'images satellite pourra permettre de tracer des cartes de suivre de l'érosion dans des zones difficiles d'accès de la Nouvelle Calédonie (côte oubliée) et de prévenir ce phénomène. Cette méthode est aussi actuellement utilisée pour le suivi de mouvements gravitaires (glissement de terrain...).

5. Remerciements

Catherine Pothier (LGCIE, SMS-ID) et Christophe Rigotti (LIRIS) sont membres du LabEx IMU (ANR-10-LABX-0088).

6. Références bibliographiques

- Bird E. C, Dubois J.-P., Iltis J. A. (1984). *The Impacts of Opencast Mining on the Rivers and Coasts of New Caledonia*. United Nations, University Press.
- Julea A., Méger N., Bolon P., Rigotti C., Doin M.-P., Lasserre C., Trouve E., Lăzărescu V. (2011). Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, n° 4, pp. 1417-1430.
- Lille D. (2007). *Qëhnelö™ : Concept du Geospatial Cloud Computing Environnemental* - 1er Prix au 9ème Concours national d'aide à la création d'entreprises innovantes. MESR, Paris, France.
- Lillesand T., Kiefer R. (2000). *Remote Sensing and Image Interpretation*, 4th ed. John Wiley, New York.
- Mimoura N. (1999). Vulnerability of Island Countries in the South Pacific to Sea Level Rise and Climate Change. *Climate Research*, vol. 12, n° 2-3, pp. 137-143.
- Méger N., Rigotti C., Pothier C. (2015). Swap Randomization of Bases of Sequences for Mining Satellite Image Times Series. *Proceedings, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pp. 190-205.
- Proctor J. (2003). Vegetation and soil and plant chemistry on ultramafic rocks in the tropical Far East. *Perspectives in Plant Ecology, Evolution and Systematics*, vol. 6, n° 1-2, pp. 105-124.
- Andreoli R., Ducarouge B., Maura J., Leopold A., Mougél P.-N., Durand A., Marchand C., Selmaoui-Folcher N., Yesou H., Tinel C., Fontanaz D., Lille D. (2015). L'imagerie spatiale à très haute résolution au coeur du dispositif de Geospatial Cloud Computing Qëhnelö™ : Application aux données Pléiades en Nouvelle-Calédonie. *Revue Française de Photogrammétrie et de Télédétection*, n° 209, pp. 47-57.