

Qu'est-ce que la géomatique?

La géomatique se dit de la science et des technologies d'acquisition, de stockage, d'analyse, d'interprétation et de diffusion de l'information géographique (géographie + informatique). Les données et les informations géoréférencées, de nature qualitative ou quantitative, sont représentées de manière à faciliter leur compréhension et leur utilisation comme instrument d'aide à la gestion des milieux naturels et humains.

Une riche combinaison

La géomatique appliquée à l'environnement est née d'une combinaison de la télédétection, des systèmes d'information géographique, des systèmes de positionnement global et de la cartographie numérique en relation avec les concepts de l'écologie, de géologie, de la géographie, des ressources naturelles, donc de l'environnement au sens large.

L'objectif de la discipline est de créer une représentation détaillée mais compréhensible du monde physique et humain pour un aménagement harmonieux de l'espace que nous y occupons.

Les débouchés

La géomatique appliquée s'implante progressivement dans plusieurs secteurs d'aide à la prise de décision. On peut citer entre autres les domaines tels la planification stratégique, le suivi et la protection de l'environnement, le monitoring écologique, les études d'impact, la gestion de la santé publique, le développement des ressources naturelles, la prévention des géorisques et le fonctionnement de la sécurité civile

Les professionnels travaillant à l'étude de la géomatique et de ses logiciels sont appelés des *géomaticiens* ([chef de projet ou ingénieur en système d'information géographique](#), administrateur SIG, technicien cartographe ou en traitement des données, gestionnaire de bases de données spatiales, architecte SIG...).

Domaines d'application

Les domaines pour lesquels peut s'appliquer la géomatique sont nombreux, avec par exemple :

- Géographie
 - Démographie
 - Environnement
 - Topographie
 - Navigation
 - Géomarketing
 - Hydrographie
 - Foresterie
 - Pêche (halieutique)
 - Aménagement du territoire et Urbanisme
 - Gestion de la périurbanisation et du foncier
 - Agriculture, Pédologie, Agronomie
 - Gestion des risques
 - Gestion des ressources naturelles
 - Sciences de la Vie et de la Terre
 - Sciences humaines et sociales (SHS)
 - Épidémiologie
 - Éco-épidémiologie
 - Gestion cynérgétique
 - Météorologie
 - Observatoires (des transports, du paysage, climat, de l'environnement, etc.)
 - Photo-interprétation
 - Commerce
 - Statistiques
 - Cartographie
 - Occupation du sol
 - Collecte des ordures ménagères
 - Archéologie
-

Apport de la géomatique en géologie

Définitions de base en Télédétection et SIG

La télédétection est une technique qui permet, à l'aide d'un capteur, "d'observer" et d'enregistrer le rayonnement électromagnétique, émis ou réfléchi, par une cible quelconque sans contact direct avec celle-ci. Le traitement et l'analyse des informations véhiculées par le rayonnement enregistré permettent d'accéder à certaines propriétés de cette cible : géométriques (position, forme et dimensions), optiques (réflexion, transmission, absorption, etc.) et physico-chimiques (température, teneur en eau, chlorophylle foliaire, phyto-masse, matière organique du sol,...), etc.

Principes de base de la télédétection

Le principe de base de la télédétection est similaire à celui de la vision de l'homme. La télédétection est le fruit de l'interaction entre trois éléments fondamentaux : une source d'énergie, une cible et un vecteur (figure 1).

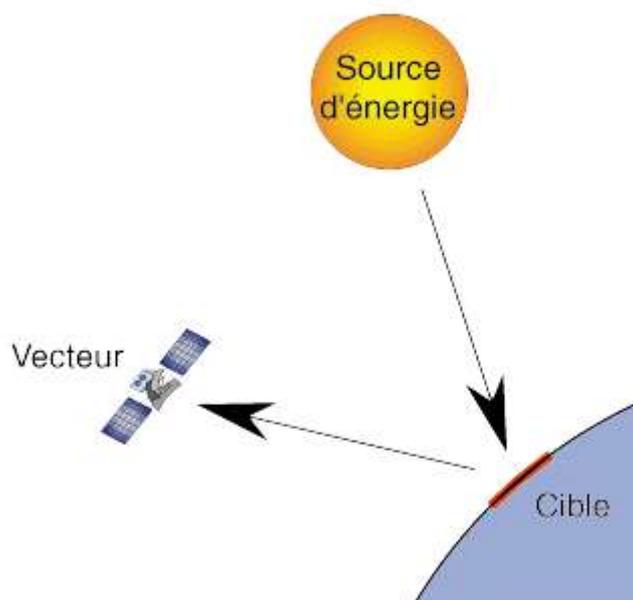


Figure 1 : Principe de base de la télédétection

La cible est la portion de la surface terrestre observée par le satellite. Sa taille peut varier de quelques dizaines à plusieurs milliers de kilomètres carrés.

La source d'énergie est l'élément qui "éclaire" la cible en émettant une onde électromagnétique (flux de photons). Dans l'immense majorité des cas, la source d'énergie est le soleil. Néanmoins, la technologie RADAR nécessite qu'un émetteur soit embarqué sur

le satellite, dans ce cas le satellite lui-même est source d'énergie. Il est également possible de mesurer la chaleur qui se dégage à la surface de la cible (infrarouge thermique), auquel cas c'est la cible qui est source d'énergie (bien qu'il s'agisse d'énergie solaire stockée et réémise).

Le vecteur ou plate-forme de télédétection mesure l'énergie solaire (rayonnement électromagnétique) réfléchi par la cible. Le vecteur peut-être un satellite ou un avion, dominant la cible de quelques centaines de mètres à 36 000 kilomètres. Les capteurs embarqués sur le satellite mesurent le rayonnement électromagnétique réfléchi, puis un émetteur renvoie l'image sur Terre vers des stations de réception.

Lorsque le satellite ne fait que capter le rayonnement réfléchi, on parle de **télédétection passive** et lorsque le satellite émet une onde vers la cible et en mesure l'écho, on parle de **télédétection active**.

Les principales étapes de la télédétection

De manière plus détaillée, on peut schématiser la télédétection comme un ensemble de 7 étapes clés (figure 2) :

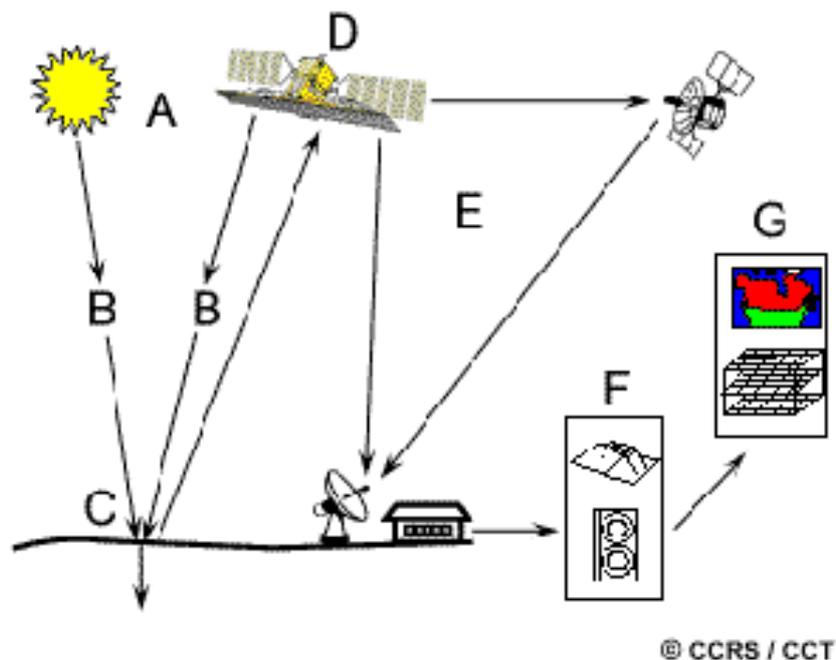


Figure 2 : Les sept étapes de la télédétection

1. Source d'énergie ou d'illumination (A) : À l'origine de tout processus de télédétection se trouve nécessairement une source d'énergie pour illuminer la cible. Le plus souvent, voire dans la presque totalité des cas, cette source d'énergie est le soleil. Mais le satellite lui-même peut être source d'énergie : c'est le cas pour le domaine de la télédétection radar.

2. Rayonnement et atmosphère (B) : Durant son parcours « aller » entre la source d'énergie et la cible, le rayonnement interagit avec l'atmosphère. Une seconde interaction se produit lors du trajet « retour » entre la cible et le capteur.

3. Interaction avec la cible (C) : Une fois parvenue à la cible, l'énergie interagit avec la surface de celle-ci. La nature de cette interaction dépend des caractéristiques du rayonnement et des propriétés de la surface. Chaque objet géographique émet ou réfléchit un rayonnement dans les diverses fréquences du spectre électromagnétique. Cette caractéristique s'appelle le comportement spectral. En télédétection, on suppose que tout objet ou classe d'objet sur la surface terrestre possède sa propre « empreinte digitale » dans le spectre électromagnétique (la signature spectrale), en fonction de la longueur d'onde du rayonnement qui est réfléchi ou émis par lui-même.

4. Enregistrement de l'énergie par le capteur (D) : Une fois l'énergie diffusée ou émise par la cible, elle doit être captée à distance par un capteur qui n'est pas en contact avec la cible mais embarqué à bord d'un satellite ou d'un avion par exemple, pour être enfin enregistrée sous format numérique.

5. Transmission, réception et traitement (E) : Cette information enregistrée par le capteur est transmise, souvent par des moyens électroniques, à une station de réception généralement située au sol où l'information est transformée en images (numériques ou photographiques).

6. Interprétation et analyse (F) : Une interprétation visuelle et/ou numérique de l'image traitée est ensuite nécessaire pour extraire l'information que l'on désire obtenir sur la cible.

7. Application (G) : La dernière étape du processus consiste à utiliser l'information extraite de l'image pour mieux comprendre la cible, c'est-à-dire la portion d'espace étudiée pour nous en faire découvrir de nouveaux aspects ou pour aider à résoudre un problème particulier.

Ces sept étapes couvrent le processus de la télédétection, du début à la fin.

Les bases physiques de la télédétection

a. Le rayonnement électromagnétique

Selon la théorie corpusculaire de la lumière, le rayonnement électromagnétique peut être considéré comme étant un flux de particules élémentaires appelés photons. Selon la théorie ondulatoire, le rayonnement électromagnétique est composé de deux vecteurs champ électrique et magnétique perpendiculaires et se déplaçant à la vitesse de la lumière (300 000 km/s dans le vide). Deux propriétés principales caractérisent une onde électromagnétique : sa longueur et sa fréquence (figure 3).

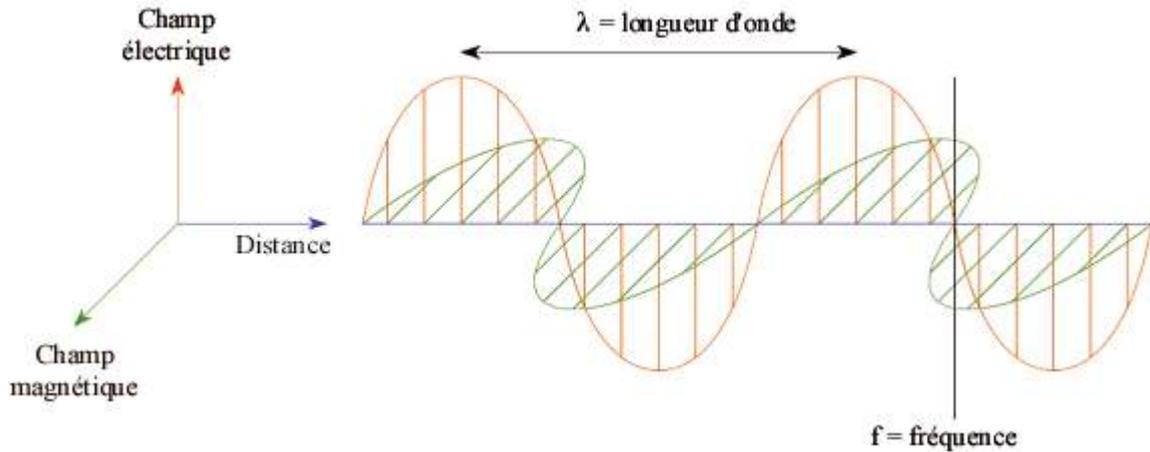


Figure 3 : Représentation de l'onde électromagnétique.

b. Le spectre électromagnétique

C'est le résultat de la décomposition du rayonnement électromagnétique en ses fréquences constituantes. Il s'étend des courtes longueurs d'onde (dont font partie les rayons gamma et les rayons X) aux grandes longueurs d'onde (micro-ondes et ondes radio).

L'ensemble des fenêtres de toutes les longueurs d'ondes est appelé « **spectre électromagnétique** » (figure 4).

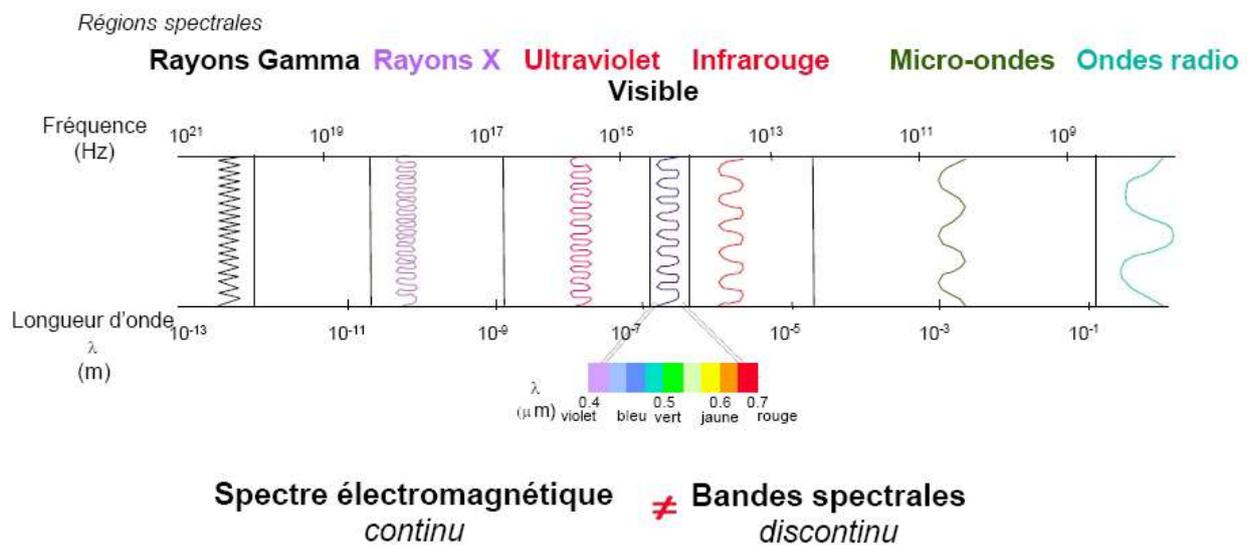


Figure 4 : Détails des domaines spectraux

La télédétection utilise plusieurs régions du spectre électromagnétique. En télédétection, les capteurs des instruments embarqués sur les satellites et avions mesurent la quantité de REM renvoyée par une portion d'espace terrestre, les capteurs embarqués sur les satellites peuvent enregistrer les ultraviolets (rare), le visible, le proche-infrarouge, l'infrarouge moyen, l'infrarouge thermique, l'infrarouge lointain et les hyperfréquences.

c. Signatures spectrales

En mesurant l'énergie réfléchié ou émise par la cible avec une variété de longueurs d'onde, nous pouvons construire la signature spectrale pour un objet. Les signatures spectrales sont donc la réponse radiométrique des objets dans une succession de longueurs d'ondes. Elles sont représentées par des courbes figurant la réflectance en ordonnées et la longueur d'onde en abscisse.

La structure des minéraux est telle qu'il existe de nombreuses bandes d'absorptions dues aux transitions électroniques et aux vibrations moléculaires (figure 5).

La réflectance des roches dépend de leur composition physico-chimique mais également de leur degré d'altération. La signature spectrale d'une roche n'est donc pas uniquement une combinaison des signatures spectrales des minéraux qui la compose (figure 5).

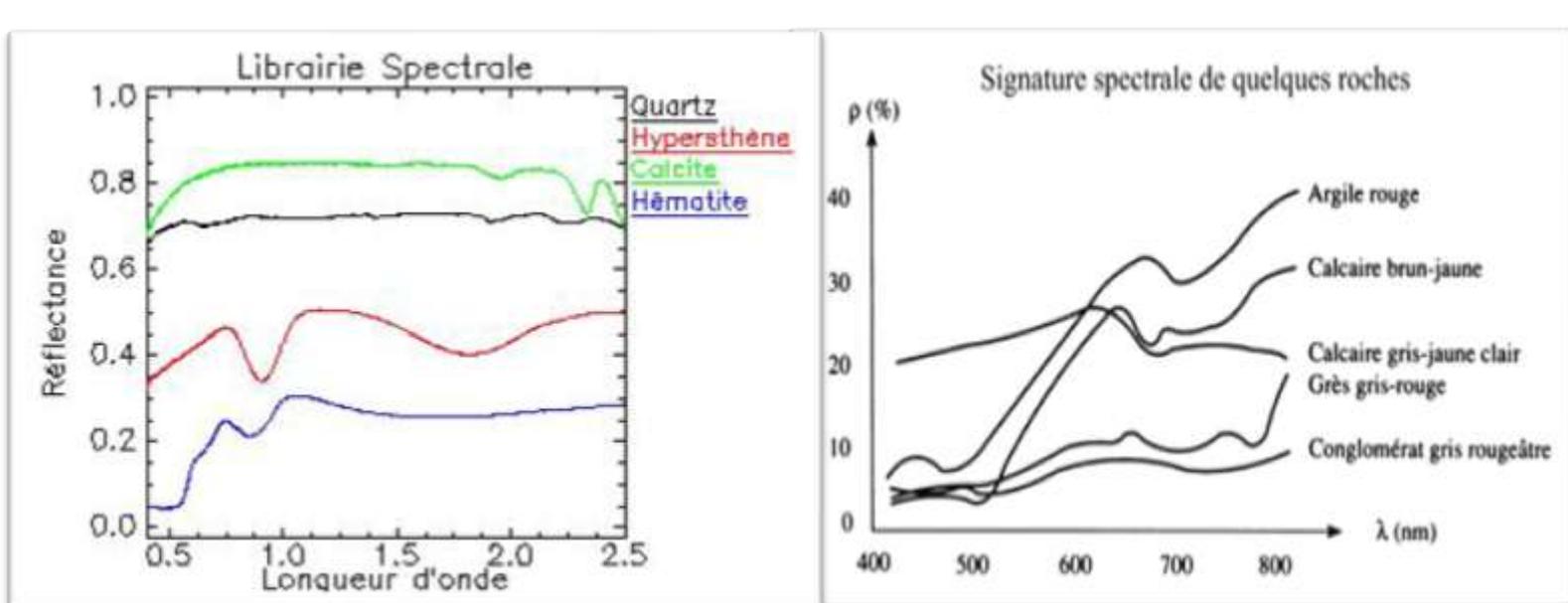


Figure 5 : Signatures spectrales de quelques minéraux et roches.

Les satellites et capteurs Landsat

Les photos satellites utilisées dans ce travail proviennent du satellite Landsat 7 ETM+. Le choix de ces données est dicté par le fait qu'elles sont une source d'informations géologiques importantes et par le fait que ces images soient téléchargeables gratuitement sur le site de la NASA.

La première génération de satellites, composée des satellites Landsat 1, 2 et 3 a été lancée entre 1972 et 1985. Les données enregistrées par cette génération de satellites, conçus pour la cartographie de la végétation, apportent peu d'information dans les domaines du spectre électromagnétique discriminatoire des minéraux.

Actuellement seuls les satellites Landsat 4, 5 et 7 de la seconde génération lancés entre 1982 et 1997, sont encore en fonction. Ces satellites fournissent des données d'une même surface à différentes dates, sur toute la surface terrestre exceptée pour les régions polaires. Les

données des capteurs Thematic Mapper (TM) à bord de Landsat 4 ou 5 sont utilisées pour plusieurs applications comme la gestion des ressources, la cartographie, la surveillance de l'environnement et la détection des changements anthropiques ou naturels à la surface de la terre. La limite de résolution spatiale du capteur TM est de 30 m pour toutes les bandes, sauf l'infrarouge thermique qui est de 120 m. Toutes les bandes sont enregistrées sur une étendue de 256 valeurs numériques (8 octets). La résolution spectrale des bandes individuelles TM et leurs applications sont fournies dans le (Tableau 1).

Landsat 7, dernier satellite de la famille Landsat, possède à son bord le capteur multi-spectral ETM+ (Enhanced Thematic Mapper Plus) (<http://landsat.gsfc.nasa.gov>). Ce capteur enregistre la réflectance de la surface terrestre dans sept bandes spectrales (Tableau 2) quatre dans le Visible et Proche Infrarouge, deux dans l'Infrarouge Moyen, une dans l'Infrarouge Thermique et une bande en mode panchromatique (l'image panchromatique est obtenue à partir de données acquises dans une seule bande spectrale, correspondant à la partie visible du spectre à l'exception du bleu, la bande couvre la région du spectre comprise entre 0,52 et 0,90 μm). Ces bandes spectrales peuvent être utilisées pour la discrimination lithologique.

Tableau 1: Caractéristiques et applications des bandes spectrales du capteur TM.

Bandes de fréquence de l'instrument ETM+	Bandes spectrales	Résolution spatiale	Longueur d'onde
Bande 1	Bleu (visible)	30m	0,45-0,5 μm
Bande 2	Vert (visible)	30m	0,52-0,6 μm
Bande 3	Rouge (visible)	30m	0,63-0,69 μm
Bande 4	IR proche	30m	0,75-0,9 μm
Bande 5	IR moyen	30m	1,5-1,7 μm
Bande 6/1	IR Thermique/ lointain	60m	10,4-12,5 μm
Bande 6/2		120m	
Bande 7	IR moyen	30m	2,08-2,35 μm
Bande 8	Panchromatique (vert-rouge-IR)	15m	520-900nm

Tableau 2 : Les caractères des bandes Landsat 7 ETM+

Bandes	Domaine Spectral (μm)	Résolution	Application
TM 1	0,45 - 0,52 (bleu)	30 m	Discrimination entre le sol et la végétation, bathymétrie /cartographie côtière; identification des traits culturels et urbains
TM 2	0,52 - 0,60 (vert)	30 m	Cartographie de la végétation verte (mesure le sommet de réflectance); identification des traits culturels et urbains
TM 3	0,63 - 0,69 (rouge)	30 m	Discrimination entre les espèces de plantes à feuilles ou sans feuilles; (absorption de chlorophylle); identification des traits culturels et urbains
TM 4	0,76 - 0,90 (proche IR)	30 m	Identification des types de végétation et de plantes; santé et contenu de la masse biologique; délimitation des étendues d'eau; humidité dans le sol
TM 5	1,55 - 1,75 (IR de courte longueur d'onde)	30 m	Sensible à l'humidité dans le sol et les plantes; discrimination entre la neige et les nuages
TM 6	10,4 - 12,5 (IR thermique)	120 m	Discrimination du stress de la végétation et de l'humidité dans le sol relié au rayonnement thermique; cartographie thermique
TM 7	2,08 - 2,35 (IR de courte longueur d'onde)	30 m	Discrimination entre les minéraux et les types de roches; sensible au taux d'humidité dans la végétation

Traitements d'images satellitaires sous ENVI

Présentation du logiciel ENVI

Le logiciel ENVI est un logiciel commercial complet de visualisation et de traitements d'images issues de la télédétection. Toutes les méthodes de traitement d'images de corrections géométriques, radiométriques, de démixage radiométrique, de classification et de mise en page cartographique sont présentes. D'autres outils relatifs à la visualisation et à la modélisation de données topographiques sont aussi disponibles.

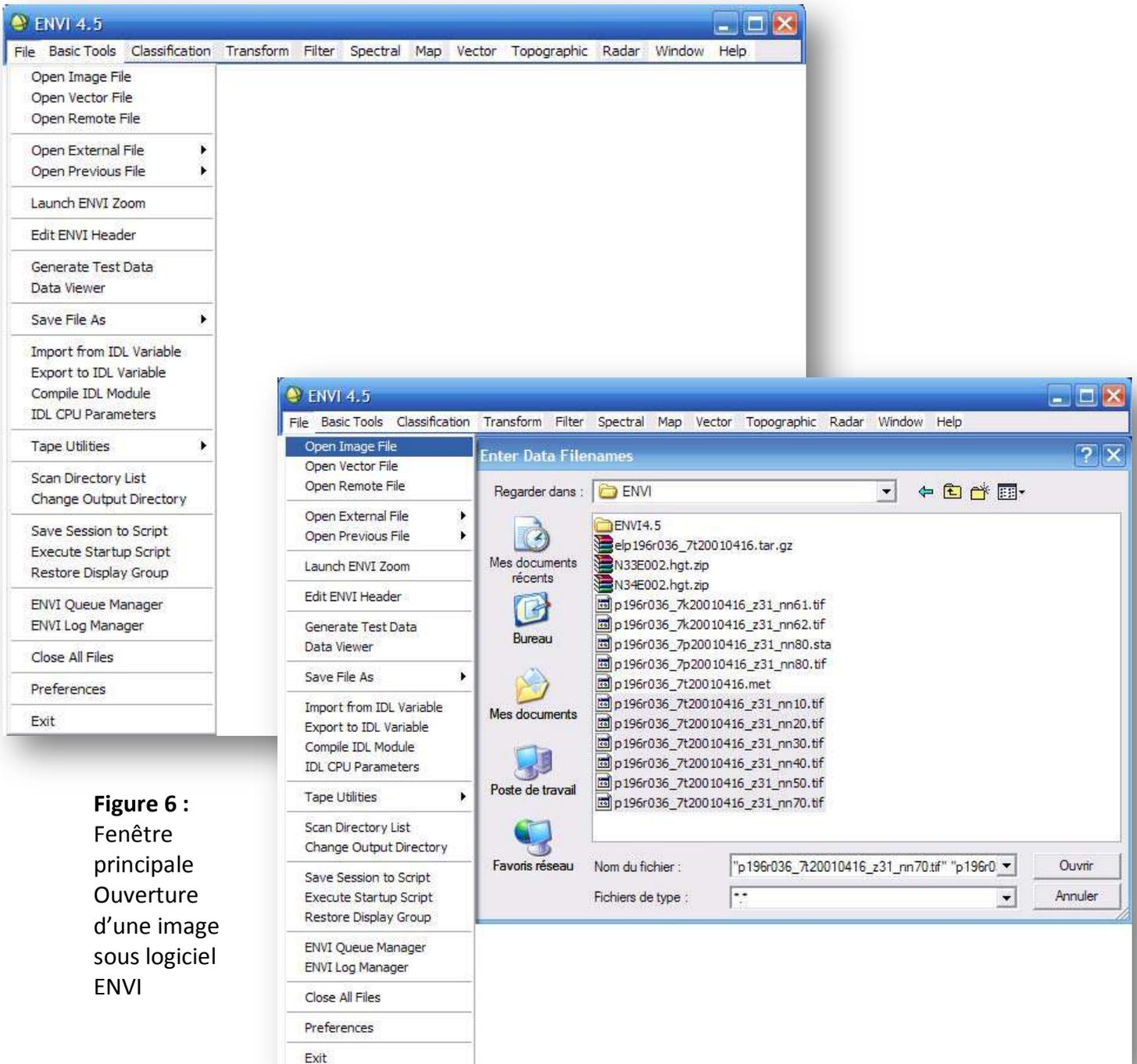
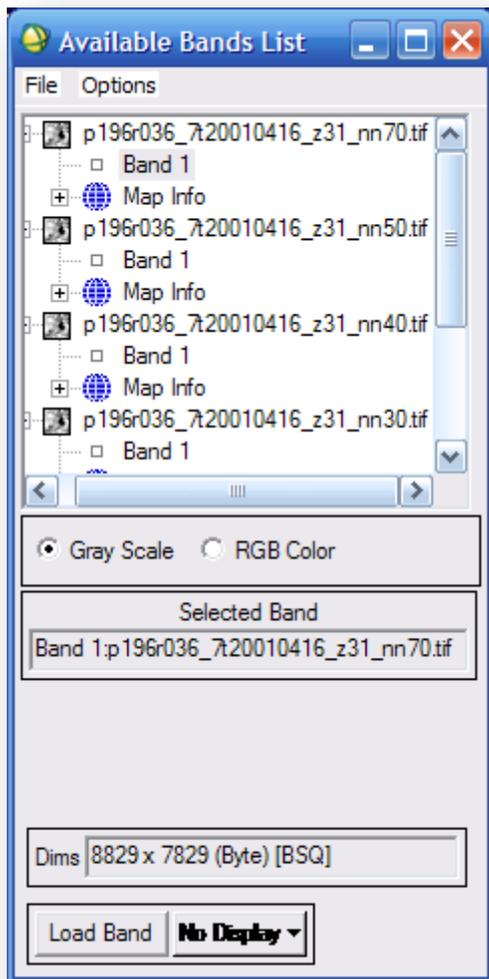


Figure 6 :
Fenêtre principale
Ouverture
d'une image
sous logiciel
ENVI



La boîte de dialogue Available Bands List affiche le nombre de couches que constituent l'extrait de l'image **Landsat**. Ici, on a donc six couches, nommées bande 1, bande 2, bande 3, bande 4, bande 5 et bande 7.

Deux modes de visualisation de l'image sont disponibles : Gray Scale (échelle de gris) et RGB (rouge, vert et bleu). L'image résultante de ce dernier mode d'affichage est dite composition colorée ou image en fausse couleur du fait que les bandes disponibles ne correspondent pas réellement aux couleurs primaires.

Figure 7 : Les deux modes de visualisation des images sous ENVI

Affichage en niveaux de gris

On clique sur la bande 7 et on choisi le mode d'affichage Gray Scale – puis on clique sur le bouton Load Band.

Trois fenêtres sont disponibles : la fenêtre principale Display 1, le Scroll et le Zoom. Le carré rouge sur la fenêtre Scroll permet de délimiter la zone géographique qui s'affichera sur la fenêtre Display. Ces fenêtres peuvent être agrandies ou réduites à l'aide de la souris

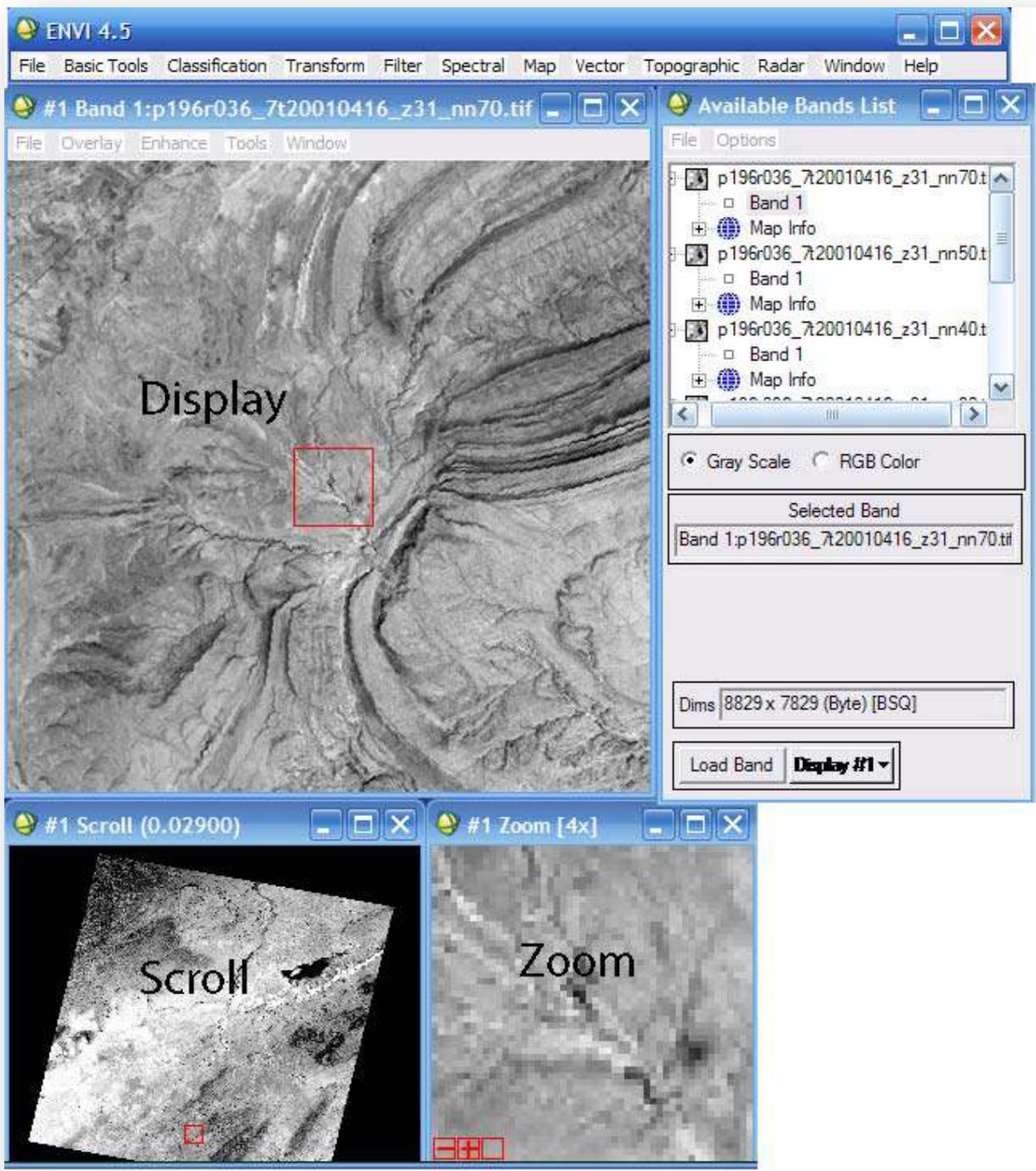


Figure 8 : Les trois fenêtres principales qui apparaissent à l'ouverture d'une image sous ENVI

Affichage en fausses couleurs : composition colorée

affichage en fausses couleurs consiste à attribuer des couleurs fictives aux bandes spectrales.

L'image résultante est dite composition colorée et se compose de trois plans de couleur rouge, vert et bleu. Par convention, on attribue la couleur rouge à la bande spectrale proche infrarouge, la couleur verte à la bande rouge et la couleur bleue à la bande verte.

Dans une composition et comme pour toute image couleur, la formation de la couleur est basée sur les trois couleurs primaires rouge, vert et bleu. En mode dit synthèse trichromatique additive, on part du noir (absence de couleur), on ajoute dans des proportions variables les trois couleurs primaires pour former de la couleur. Par exemple, le blanc est formé par les trois couleurs R, V et B à intensités maximales (255).

D'une manière générale, sur une composition colorée, les sols nus apparaissent dans des teintes bleu clair, l'eau dans une teinte bleu foncé ou pratiquement noir, la végétation herbacée en rouge clair.

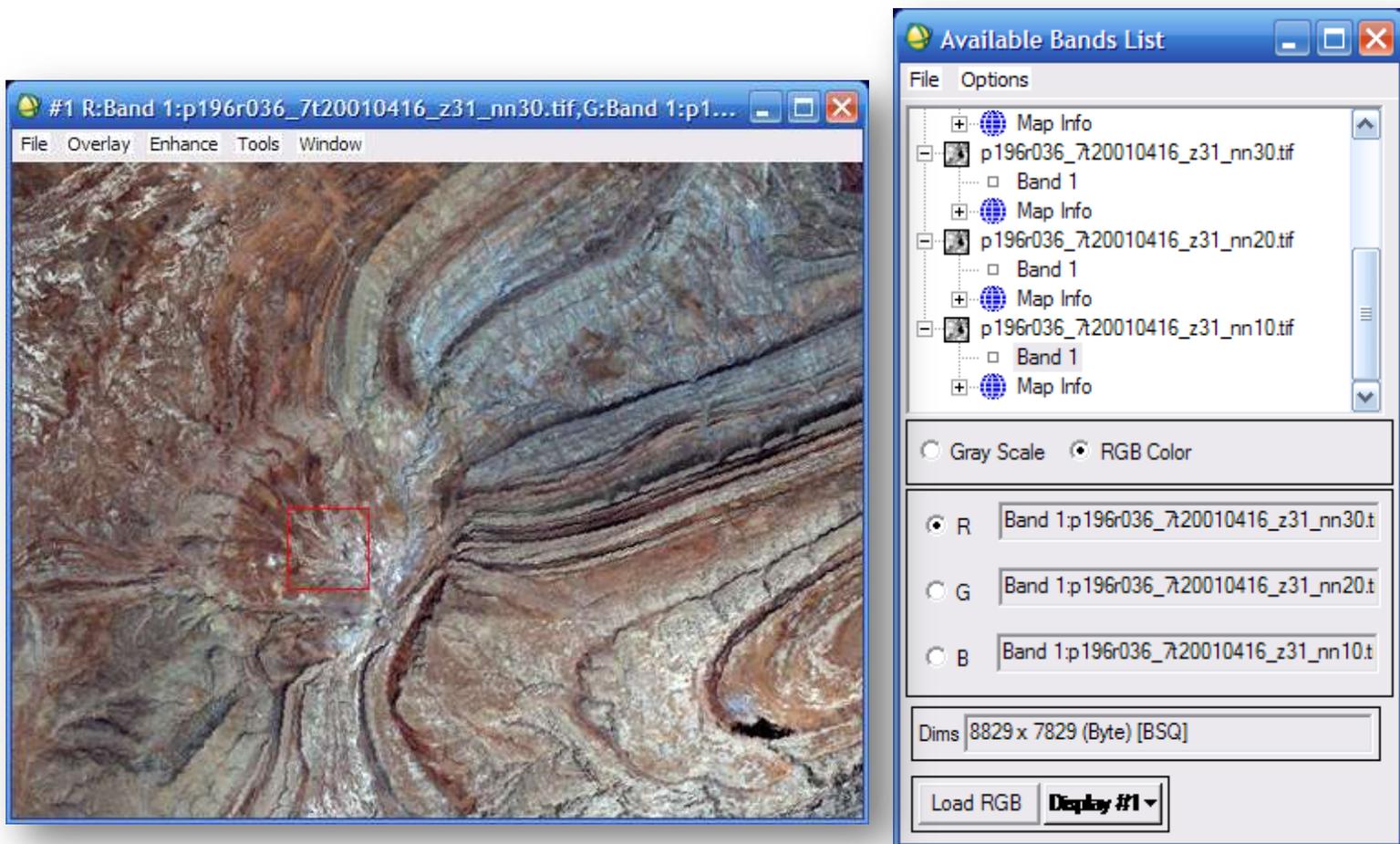


Figure 9 : Visualisation des images en fausses couleurs (ici 321)

Rapports de bandes (ou bandes ratios)

Toute analyse est basée sur la notion de réflectance, c'est le rapport entre la quantité d'énergie réfléchiée par un objet et la quantité d'énergie reçue par le capteur. Cette méthode permet de réduire les effets de la topographie et d'augmenter le contraste entre les surfaces minérales, en pratique il s'agit de définir les bandes (paire) qui entrent dans le calcul des rapports et les introduire dans l'ENVI.

Pour obtenir des néo-bandes qu'on utilise en RGB pour obtenir des images, on clique sur « transform » puis sur « band ratios » (figure 10) :

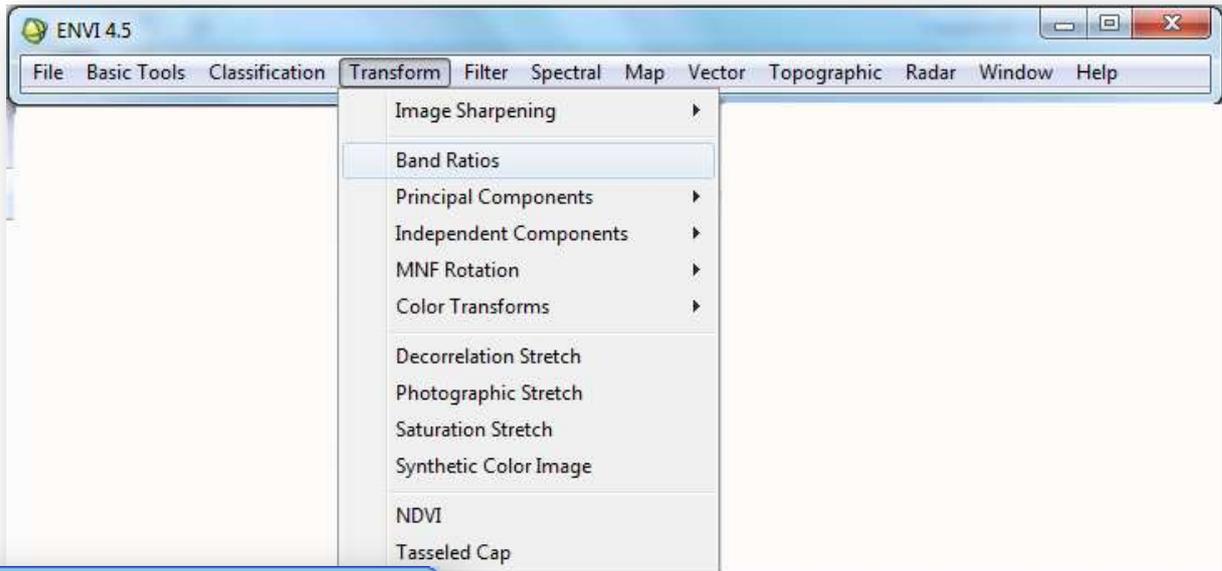


Figure 10 : Fenêtre montrant les étapes à effectuer pour obtenir les « band ratios »

Dans l'exemple qui va suivre on a choisi les rapports des bandes suivantes : 7/3 5/2 et 4/7. A chaque fois on introduit les bandes dans « Numerator » et « Denominator » et on obtient les paires qui sont indiquées dans Selected Ratio Pairs (figure 25). Le calcul d'un rapport de bandes (le rapport de canaux TM correspond à un nouveau canal, dont les valeurs des pixels sont égales au rapport entre les valeurs radiométriques des deux canaux), permet de s'affranchir des variations d'éclairement des pixels, mais aussi de mieux séparer les unités, car les informations complémentaires contenues dans les deux bandes sont prises en compte (tableau 4).

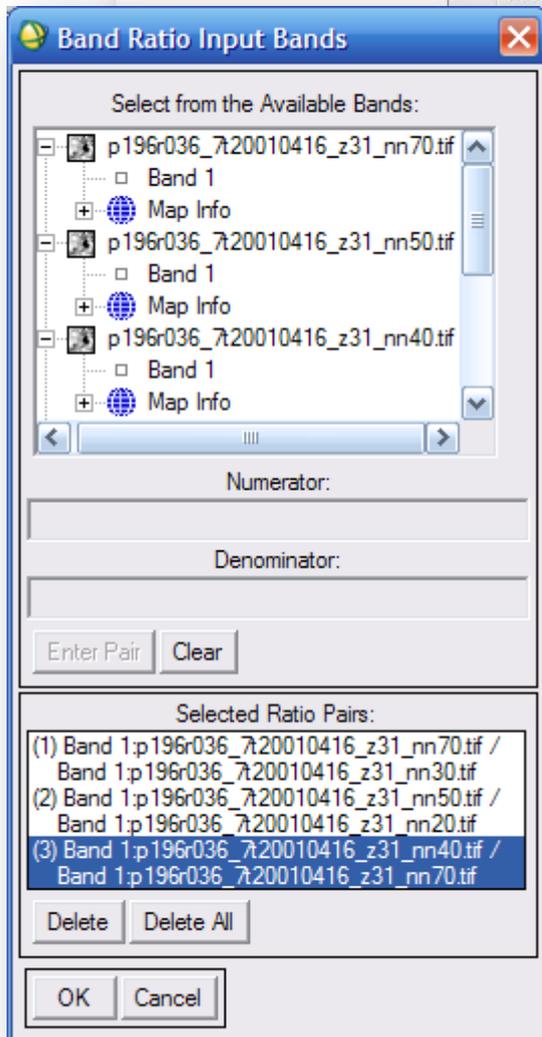


Figure 11 : Rapports de bandes 7/3 5/2 et 4/7

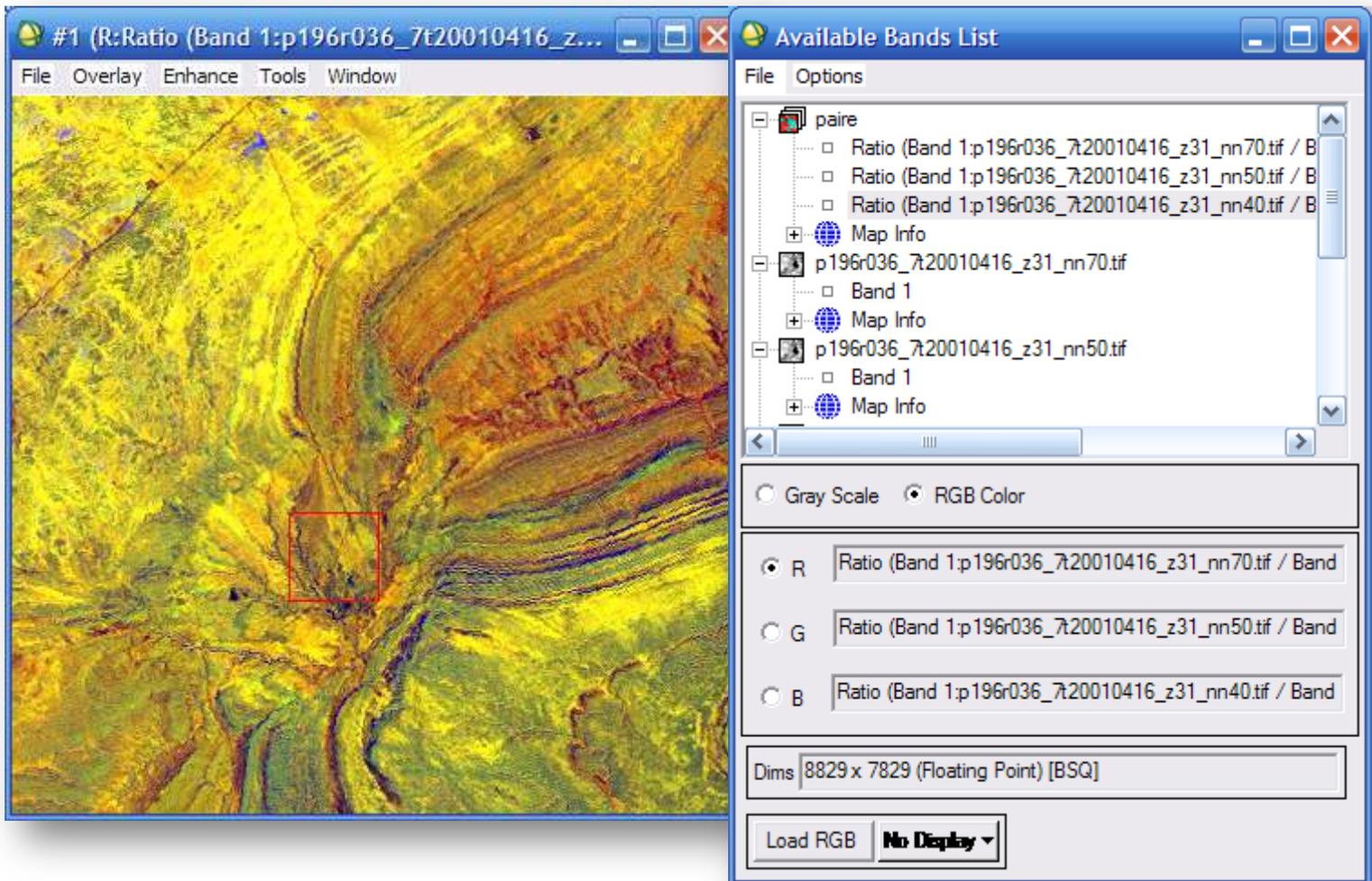


Figure 12 : Exemple d'une image traitée en RVB des rapports de bande 7/3 5/2 et 4/3

Voici quelques exemples de rapports de bandes utilisés en RVB (voir aussi tableau 4) : [3 /1, 5/4, 7/5] mets l'accent sur la composition des matériaux de surface. [3/5, 1/5, 5/7] et [5.7, 4.5, 1.3] bon pour la cartographie des serpentines. [7/3, 5/2, 4/7] permet d'obtenir des informations sur la lithologie des formations. [5/3, 3/1, 7/5] permet d'obtenir des informations sur la lithologie des formations.

RGB	Description
321	Image en couleur naturelle
432	Combinaison standard en fausses couleurs. La végétation apparaît en rouge.
453	La limite entre terre/eau est très nette. Bon pour l'étude des sols et de la végétation.
742	Combinaison souvent utilisée en géologie. Le sable, le sol et les minéraux montrent une multitude de couleurs. Bon pour étudier les régions désertiques.
753	Combinaison souvent utilisée géologie. Le sable, le sol et les minéraux montrent une multitude de couleurs.
541	Combinaison souvent utilisée en agronomie. La végétation apparaît en vert clair.
754	Très utiles en géologie. Fournit la meilleure pénétration dans l'atmosphère. La végétation apparaît en bleu.
731	Combinaison bonne en géologie. Les roches apparaissent dans une multitude de couleurs.

Tableau 3 : Différentes combinaisons utilisées en fausses couleurs et leurs utilités

Tableau 4 : Exemples de rapports de bandes utilisés dans le traitement des images satellitaires

Rapport	Description
5/7	Bonne discrimination pour les roches qui riches en Al-OH
3/1	Bonne discrimination pour les roches ayant subies une altération hydrothermale et pour les oxydes de fer.
1/2, 2/1	Bonne discrimination pour les roches riches en oxydes de fer
3/1, 3/2	Bonne discrimination de la végétation et l'eau
4/1, 4/2	Bonne discrimination de la végétation et l'eau
3/4	Utilisé pour étudier les terres arides et les régions urbaines
7/5	Bonne discrimination des argiles
5/4	Bonne discrimination des oxydes de fer

L'analyse en composantes principales (ACP)

Les données multi-spectrales des différents canaux contenant souvent de l'information similaire, des transformations d'images sont utilisées pour réduire cette redondance de données. Les "néo" bandes qui en résultent sont appelées composantes et permettent de réduire par l'information comprise dans 5 ou 6 canaux en seulement trois composantes tout en conservant plus de 90% de l'information initiale. C'est une technique efficace pour accentuer une image multispectrale pour des fins d'interprétation géologique.

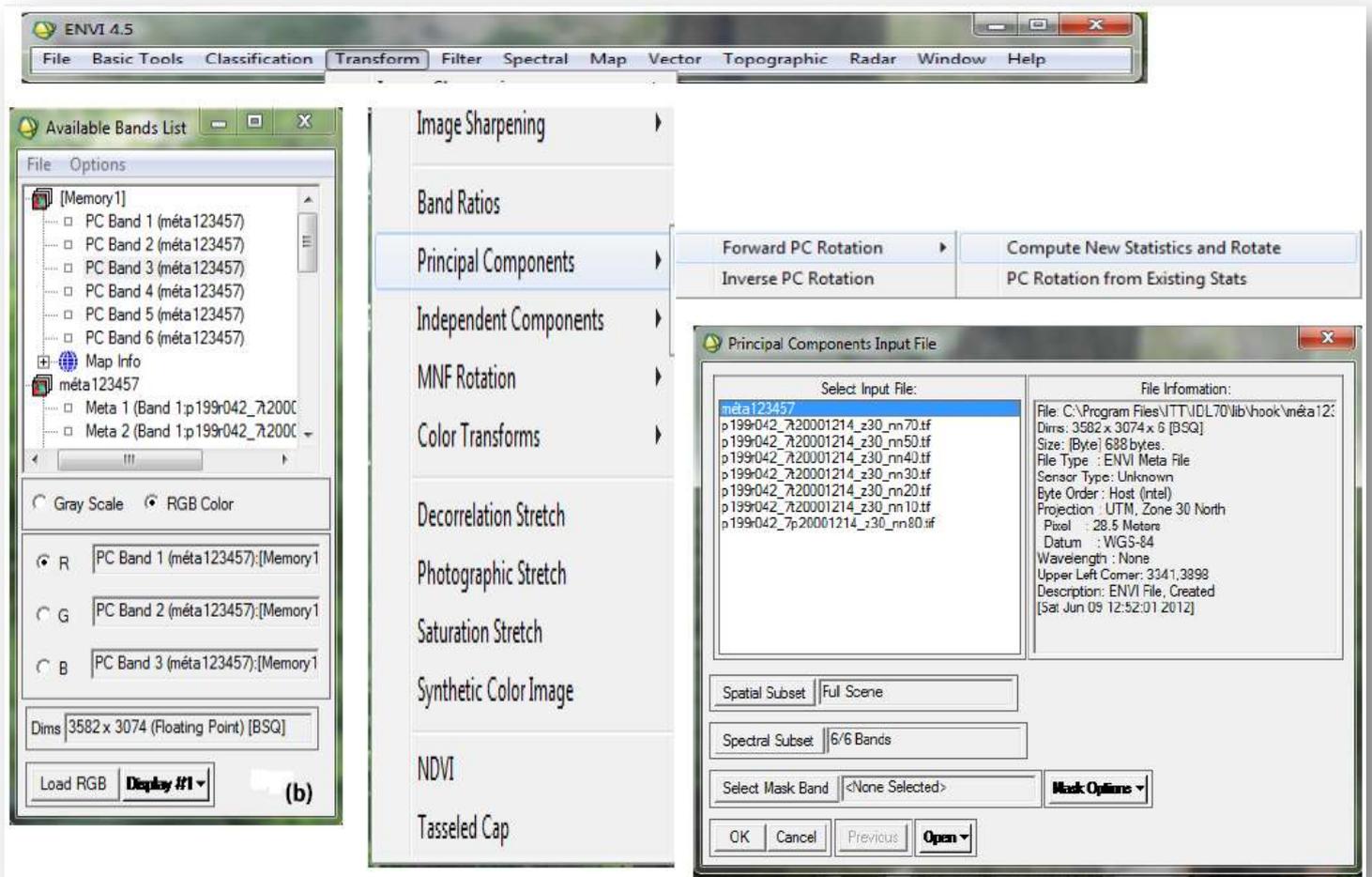


Figure 13: (a) Différentes étapes pour l'obtention des composantes principales à partir du méta-fichier contenant les six bandes TM1, 2, 3, 4, 5 et 7.

(A noter que, lors de la création du méta-fichier, on doit redimensionner la zone d'étude de sorte que les marges noires limitrophes de l'image soient retranchées, ces dernières faussant les calculs en ACP).

(b) réalisation d'une image ACP en composition colorée à partir des trois premières composantes principales CP1, CP2 et CP3.

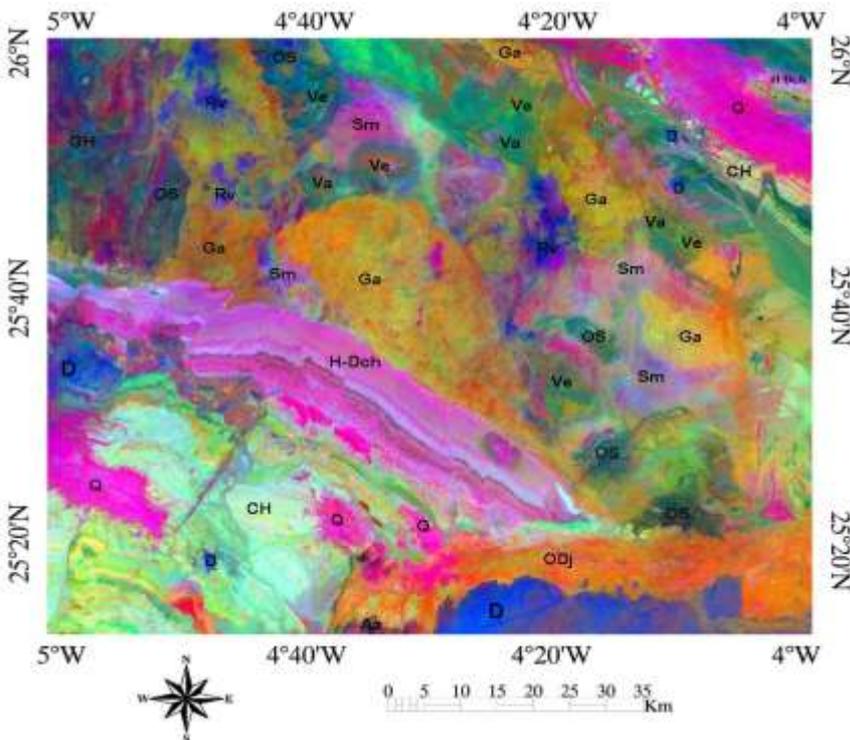


Figure 14: Image ACP (CP1, CP2, CP3) RGB réalisée à partir des 6 bandes TM 1,2, 3, 4, 5 et 7 de la partie de la scène Landsat 7 ETM+, path 199 Row 042 du 14 Décembre 2000 couvrant la région de Mokrid.

Classifications

Classification non supervisée (non assistée) :

La classification non supervisée est utilisée dans le cas où on ne dispose d'aucune information sur la zone d'intérêt, contrairement à la classification supervisée. En premier, les classes spectrales sont formées à partir de l'information numérique des images. Ensuite, elles sont associées, par l'utilisateur, à des classes d'information utile. Des programmes appelés algorithmes de classification sont utilisés pour déterminer les groupes statistiques naturels ou les structures de données (DIOULASSO, 2010). Le classement des pixels est effectué uniquement selon le critère de la ressemblance spectrale, l'étiquetage thématique sera effectué après l'identification des classes spectrales (CALOZ et COLLET, 2001).

On distingue deux méthodes de classification non supervisée sous ENVI (4.7.01) :

- La méthode basée sur les **K-MEANS** : Regroupement des pixels de l'image au centre des classes, régulièrement distribués dans l'hyperespace, les plus probables (statistiques). Elle permet une analyse régionale de grands types homogènes de couverture végétale. L'inconvénient de cette méthode réside dans le risque d'obtenir un nombre moindre de classes finales et donc des classes vides.
- La méthode de classification non supervisée en valeur égale **ISODATA** (Iterative Self-Organizing Data Analysis Technics) : Regroupement des pixels de l'image au centre des classes, régulièrement distribués dans l'hyperespace, les plus proches (distance). L'application de la classification non supervisée entraîne beaucoup de problèmes dans la cartographie de la couverture végétale ; certaines classes peuvent être doubles (même classe spectrale) ou sans signification (mélange de couvertures végétales).

Classification supervisée (assistée) :

Lors de l'utilisation d'une méthode de classification supervisée, l'analyste identifie des échantillons assez homogènes de l'image qui représentent de différents types de surface (Classe d'information) (DIOULASSO, 2010). Lorsqu'on parle de classes, il faut faire la distinction entre des classes d'information et des classes spectrales. Les classes d'information sont des catégories d'intérêts que l'analyste tente d'identifier dans les images. Les classes spectrales sont des groupes de pixels qui ont les mêmes caractéristiques (ou presque) qui ont trait à leurs valeurs d'intensité dans les différentes bandes spectrales des données (KLUSER, 2000). Chaque pixel de l'image sera affecté à la classe thématique correspondante selon son comportement spectral. La classification supervisée est donc conseillée dans le cas où on possède des informations préalables de la zone à cartographier. Le principe de la classification supervisée est basé sur l'identification des échantillons homogènes de l'image. Ces échantillons peuvent représenter une ou plusieurs zones occupées par une catégorie thématique connue, ce qui forme des zones d'entraînements (MAFHOUD, 2009). La procédure de classification supervisée permet une évaluation a posteriori par un test de séparabilité (distance statistique entre les signatures) et une

matrice de confusion (pourcentage des pixels d'une zone d'entraînement distribuée dans le groupe thématique correspondant (D'après CCT, 2008).

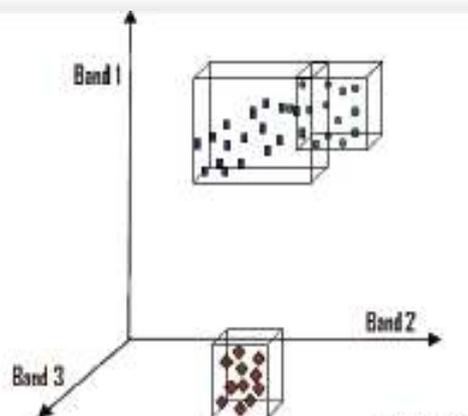
Les étapes de la classification supervisée selon CALOZ et COLLET (2001) sont présentées dans le schéma suivant (figure 15):

1. Avant de commencer la classification supervisée, on doit définir la légende c'est-à-dire choisir les classes souhaitées selon la problématique de recherche.
2. Sélectionnez les échantillons représentatifs pour chaque classe thématique. L'opérateur divise par la suite ces échantillons en deux groupes : un groupe pour l'établissement des signatures spectrales (zones d'entraînement), et un autre servant à la vérification (zones de test).
3. Établissement des signatures spectrales en calculant les paramètres statistiques (moyenne, maximum, minimum, variance, etc.) à partir des zones d'entraînement.
4. Choix de l'algorithme de classification et des règles de décision appropriés à l'objectif (par exemple maximum de vraisemblance). Alors, il faut évaluer les résultats et l'indice de séparabilité par des graphes (histogrammes et courbes des signatures), calculer la matrice de confusion et l'indice de divergence (distance euclidienne entre les centres de classes)
5. Lancement de la classification pour l'ensemble de l'image.
6. Évaluation de la classification en la comparant à des informations de référence (cartes et relevées de terrains).

La figure suivante (figure 15) représente les principales méthodes de classification supervisée qui peuvent être utilisées dans notre recherche. L'algorithme « minimum distance » détermine simplement la distance d'un point à la moyenne du pixel, alors que le « Parallelepiped » algorithme définit un volume rectangulaire autour de la signature pour affiner la classification. L'algorithme « maximum de vraisemblance » (maximum Likelihood) affine la répartition de chaque classe en formant une ellipse qui se rapproche le plus du pixel.

Parallelepiped :

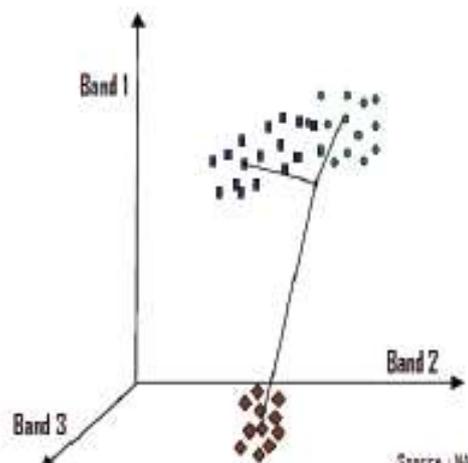
Les pixels sont arbitrairement attribués à la première classe qui se trouve entre la valeur minimale et maximale (pour chacune des bandes spécifiées). Les classes sont rectangulaires (ou l'équivalent multidimensionnel) et peuvent se recouper. Les seuils d'appartenance sont fixes implicitement par les valeurs minimales et maximales de chaque classe. Ce classifieur est l'un des plus simple.



Source : NASA

Minimum Distance:

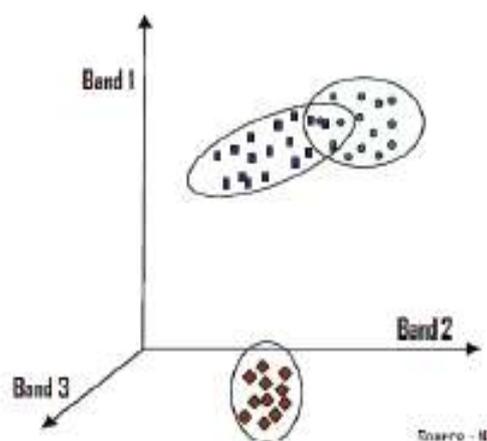
Les pixels sont assignés à la classe ayant la moyenne de classe la plus proche. Un seuil spécifie la distance maximum entre la moyenne des membres d'une classe, en unité d'écart-type. Les pixels éloignés des autres classes ne sont pas classifiés au lieu d'être assignés arbitrairement à une classe peu pertinente. Ceci n'est pas autant flexible que le Maximum Likelihood, mais c'est un meilleur choix, si les échantillons de terrain sont peu nombreux ou peu fiables.



Source : NASA

Maximum de vraisemblance:

Il en existe plusieurs types. Enhanced, Enhanced Neighbor, Standard et Standard Neighbor. Ils prennent en compte, pour dessiner l'ellipse autour des pixels, la distance pondérée par la matrice de covariance de la moyenne, et leur probabilité. Le « Enhanced » produit 2 bandes pour chacune des classes ou régions sélectionnées. Une bande rapporte les probabilités postérieures, ceci indique la vraisemblance d'appartenance du pixel à chacune des classes. L'autre bande rapporte la représentativité des pixels pour chaque classe. Ainsi, en sélectionnant un grand nombre de parcelles d'entraînement augmentera considérablement la taille de l'image de sortie. La « Neighbor classification » va lisser les données en utilisant les pixels voisins pour modifier la probabilité antérieure. Le « Neighbor Filter » fixe le nombre de voisins à utiliser pour le lissage de chaque pixel. Remarque : L'augmentation du nombre de passage augmentera aussi le lissage.



Source : NASA

Figure 15 : Principales méthode de classification utilisée par ENVI (4.7.01)
Source : (SHORT, 2000 ; CALOZ et COLLET, 2001 ; ENVI User's Guide, 2009)

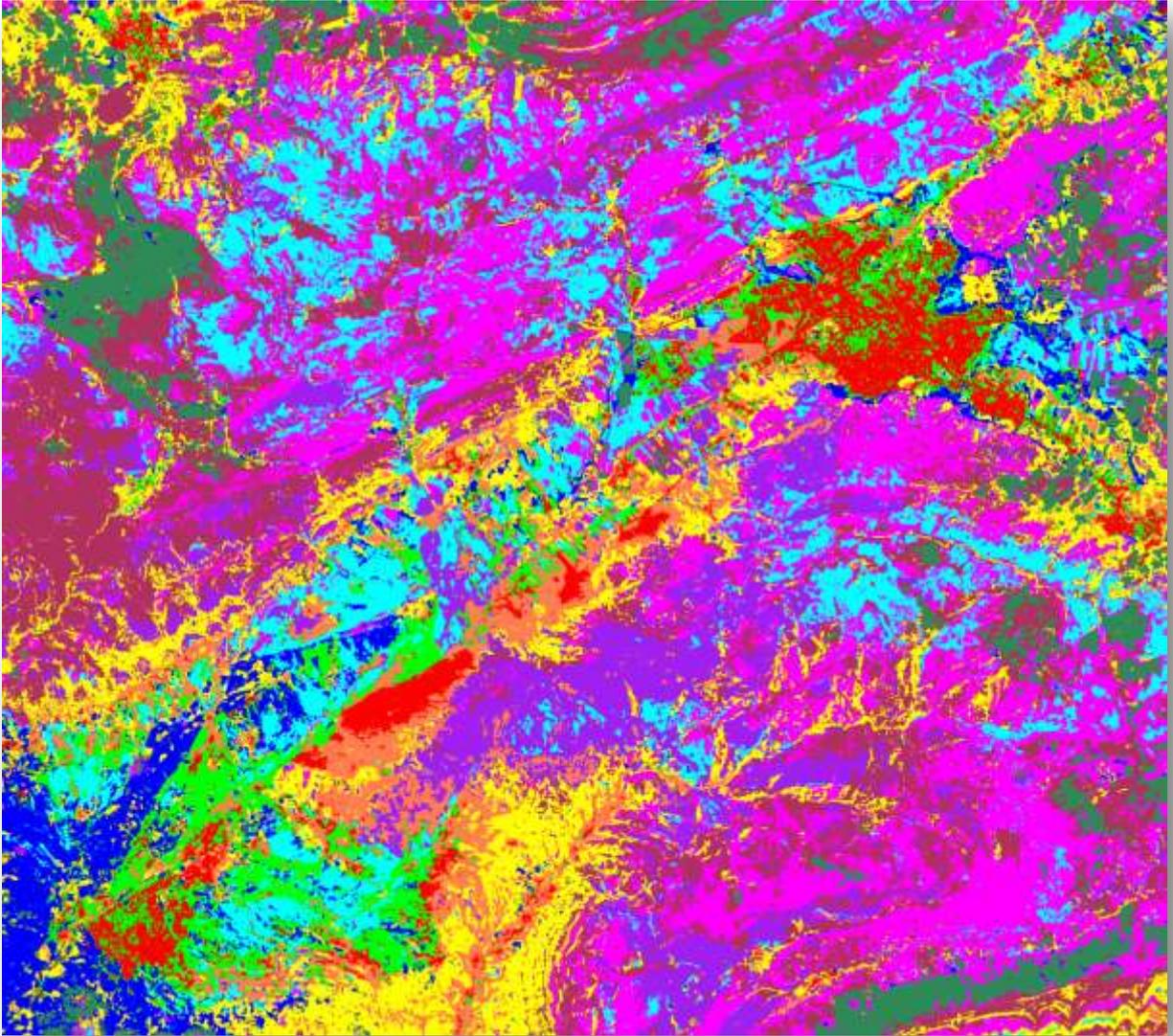


Figure 16 : Extrait d'une scène sur laquelle a été effectuée une classification supervisée, mettant l'accent sur la diversité lithologique (Monts de Batna-Bellezma, 2015)

Indice de Qualité

L'évaluation de la qualité de la classification est une des étapes principales de la cartographie de l'occupation du sol. Pour mesurer la qualité de la classification, il est indispensable de passer par le calcul de la matrice de confusion et de l'indice de séparabilité. La matrice de confusion c'est un tableau affichant les statistiques de la précision de classification d'une image, notamment le degré de classification erronée parmi les diverses classes (D'après CCT, 2008). Elle est calculée avec les valeurs exprimées en pixels et en pourcentage. Cependant, l'indice de séparabilité est exprimé en pixels seulement.

La matrice de confusion est aussi connue sous les termes de : matrice d'erreur, tableau de contingence. Elle permet de calculer une gamme de mesures décrivant la précision de la carte classifiée par rapport à la carte de référence. Pour générer la matrice de confusion, on extrait l'information thématique à partir d'un échantillon de pixels couvrant la même superficie sur les deux cartes. On obtient les données d'étalonnage à partir de la carte de référence et les données de validation de la carte classifiée (D'après CCT, 2008).

L'indice de séparabilité est utilisé pour distinguer la séparabilité des signatures spectrales entre les parcelles d'entraînement (ENVI User's Guide, 2009). Il est basé sur le calcul de la distance spectrale et compare, en général, deux pixels ou deux groupes de pixels mesurés en unité de luminance. Les deux méthodes de calcul de la distance spectrale interclasse fournissent des statistiques qui varient entre 0 et 2.

- $0 < x < 1.0$: séparabilité très faible, 0 implique une confusion totale entre deux classes.
- $< x < 1.9$: séparabilité faible.
- $1.9 < x < 2.0$: bonne séparabilité,
- 2 : séparabilité parfaite avec aucun risque de confusion donc une bonne discrimination des espèces.

Deux classes dont la séparabilité est faible doivent être soit confondue ou redéfinie.

Tableau 5 : Exemple d'une matrice de confusion d'une classification supervisée en Pxl.

Nombre de pixels	Classes thématiques							Totale signatures	Producer's Accuracy	Error's of Omission
	PA	CV	CY	R PA	BR	AM	SN			
Signatures Spectrales										
PA	111	0	16	2	0	0	0	129	111/129	18/129
CV	0	81	1	7	0	13	0	102	81/102	21/102
CY	0	0	13	0	0	0	0	13	13/13	0/13
R PA	1	5	0	19	0	0	0	25	19/25	6/19
BR	2	0	0	2	522	0	0	526	522/526	4/526
AM	0	0	1	0	0	11	0	12	11/12	1/12
SN	6	0	0	0	34	0	737	777	737/777	40/777
Totale Classe	120	87	30	30	556	24	737	1584		
User's Accuracy	111 /120	81/87	13/30	19/30	522/556	11/24	737/737			
Errors of Commission	9/120	5/87	18/30	11/30	34/556	13/24	0/737			

$$\text{Précision globale} = ((111 + 81 + 13 + 19 + 522 + 11 + 737) / 1584) * 100 = 94.31\%$$

Les valeurs de la diagonale de la matrice représentent le nombre de pixels correctement classifiés (tableau 5). De plus, la somme des cellules de la diagonale (1584 pixels) de la matrice représente le nombre total de pixels correctement classifiés. La proportion du nombre total de pixels correctement classifiés par rapport au nombre total de pixels de la matrice donne la « précision globale » de la classification. On constate que la précision globale de la classification est très bonne (94.31%). L'homogénéité des signatures des parcelles d'entraînements a aidés à obtenir la meilleure précision.

L'indice de séparabilité du logiciel de télédétection ENVI (4.7.01) a utilisé les caractéristiques des signatures spectrales des zones d'entraînement par la méthode de *la divergence* et la distance de *Jeffries – Matusita*. Ces dernières recourent aux notions de probabilité d'appartenance d'un pixel à une classe à l'instar des algorithmes paramétriques de classification (ANUTA et al., 1984 ; CALOZ et COLLET, 2001 ; ENVI User's Guide, 2009).

Les valeurs entre parenthèses sont les valeurs de séparabilité des parcelles d'entraînement par les deux méthodes mentionnées ci-dessus (figure 17)

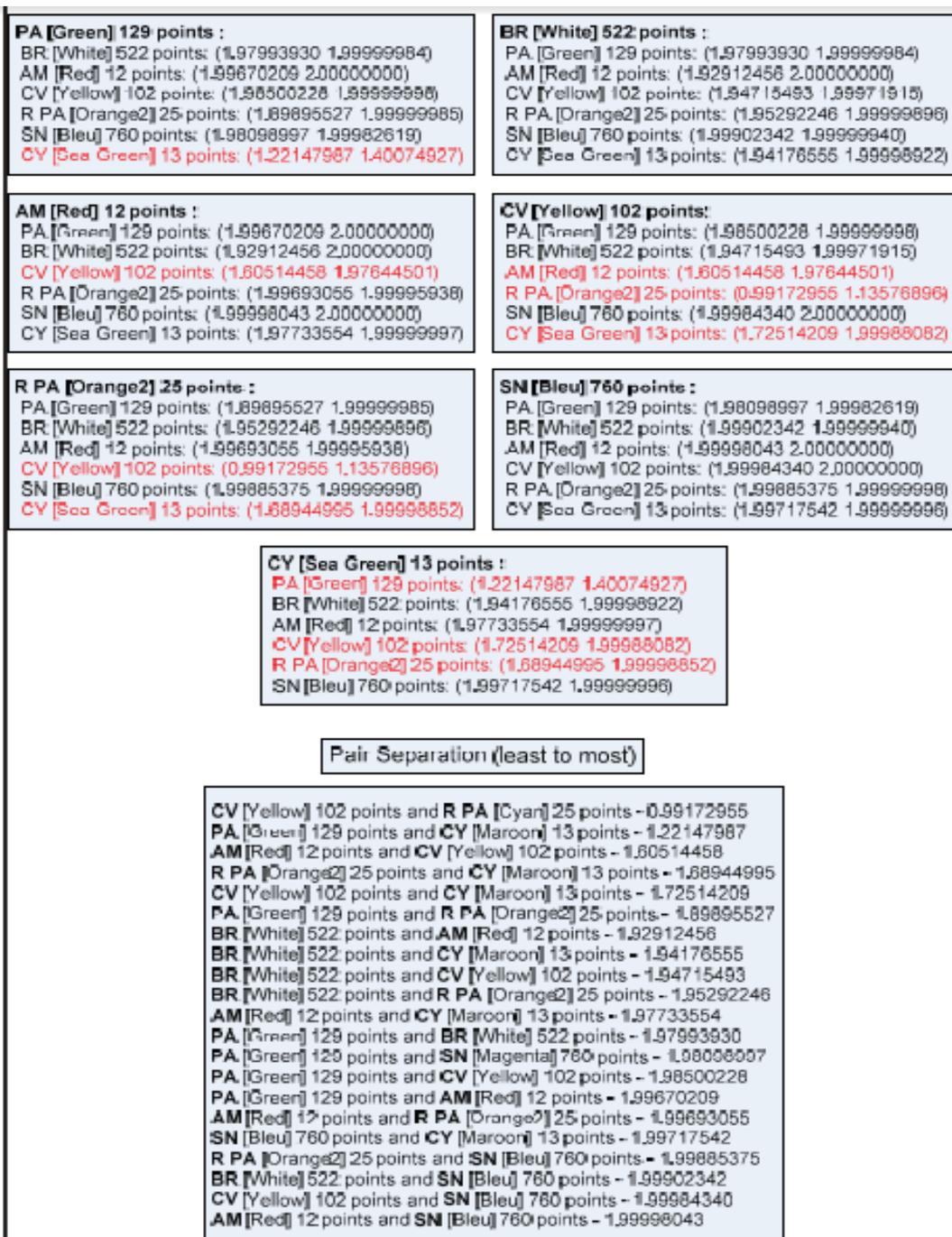


Figure 17 : Indice de séparabilité des parcelles d'entraînement (Pixels)

Validation de la carte produite par classification supervisée

Il est important de savoir que la qualité des cartes thématiques dérivées des données de télédétection peuvent être évaluées et exprimées de façon significative. Ceci est important non seulement en fournissant un guide pour la qualité de la carte et son adéquation à un usage particulier, mais aussi de comprendre l'erreur et ses conséquences probables (FOODY, 2002).

Dans la cartographie thématique des données de télédétection, l'exactitude est un terme utilisé généralement pour exprimer le degré de correction de la carte de classification. L'évaluation de l'exactitude est maintenant acceptée comme une composante fondamentale des enquêtes de la cartographie forestière par télédétection (CZAPLEWSKI, 1994 ; FRANKLIN et al., 2001 ; VALERIA et al., 2009 ; BARIMA et al., 2010). Une erreur de classification est, par conséquent, un certain décalage entre la situation représentée sur la carte thématique et la réalité.

Les méthodes conventionnelles d'évaluation de la précision thématique d'une carte ont été employées. Il s'agit de la matrice d'erreur et de l'indice Kappa (CONGALTON, 1991). Cette méthode s'appuie sur la stratégie aléatoire d'échantillonnage. La matrice d'erreur est formulée pour cataloguer des anomalies entre la carte classifiée et la référence (STORY et CONGALTON, 1986 ; LUNETTA et al., 1991 ; STEHMAN et CZAPLEWSKI, 1998). La matrice d'erreur donne une évaluation de la précision globale de la cartographie et des résultats de classification pour chacune des classes thématiques. L'indice Kappa évalue dans la matrice de confusion l'accord entre les résultats obtenus et la vérité sur le terrain. Il s'étend de 0 à 1 et il se divise en cinq catégories : accord très faible de 0 à 0.20, accord faible de 0.21 à 0.40, accord modéré de 0.41 à 0.60, accord substantiel de 0.61 à 0.80 accord presque parfait de 0.81 à 1 (BLUM et al., 1995).

L'indice Kappa (équation 3) est exprimé en terme de précision globale observée (équation 4) et de précision prévue (équation 5) (CONGALTON, 1991 ; MAMA et OLOUKOI, 2003 ; CHALIFOU et al., 2006 ; D'après CCT, 2008).

$$\hat{\kappa} = \frac{a-b}{1-b} \quad (3)$$

$$a = \frac{1}{N} \sum_{i=1}^{N_c} x_{ii} \quad (4)$$

$$b = \frac{1}{N^2} \sum_{j=1}^{N_c} (x_{+j} \cdot x_{j+}) \quad (5)$$

N_c = Nombre de classes

N = Nombre total d'observations

X_{ii} = Nombre d'observations dans la colonne **i**

Ligne **i** (diagonale de la matrice)

X_{+i} = Total des observations dans la colonne **i**
(Total à droite dans la matrice)

X_{i+} = Total des observations dans la ligne **i**

N_c = Nombre de classes

Systemes d'Information Géographique

Pour définir un SIG, il existe de nombreuses définitions. Deux options sont possibles, la première est d'insister sur les finalités de l'outil, la deuxième est de mettre en exergue ses fonctionnalités. La première est une définition issue des utilisateurs, l'autre est la définition des producteurs. Si l'on se focalise sur les fonctionnalités, selon Denégre et Salgé, un SIG doit pouvoir répondre à cinq questions : Où ? Les objets doivent y être localisés ; Quoi ? Les objets y sont définis (bâtis, forêts, cours d'eau...) ; Comment ? Les relations entre les objets doivent pouvoir être déduites (c'est de l'analyse spatiale) ; Quand ? Les données doivent pouvoir être mises à jour et les historiques conservés ; Et si ? Il doit être possible d'y effectuer des simulations.

Un SIG doit comporter toutes les informations relatives aux objets ou phénomènes concernés, de près ou de loin, par les questions qui pourront lui être posées. Ces informations sont généralement organisées par couches « homogènes », rassemblant un même type d'objets (topographie, cadastre, géologie, population, etc.). La notion de couche est donc centrale : une couche est un ensemble d'objets partageant certains attributs (un tronçon route est défini par : un identifiant, une catégorie, une largeur, une longueur, une capacité, une limitation de vitesse...) et une même forme (point, ligne, surface).

Plus précisément, les SIG comportent des fonctions de saisie des données géographiques (Acquisition), de gestion de ces données (Archivage), de manipulation, de croisement et de transformation de ces données spatiales (Analyse) et de mise en forme des résultats (Affichage).

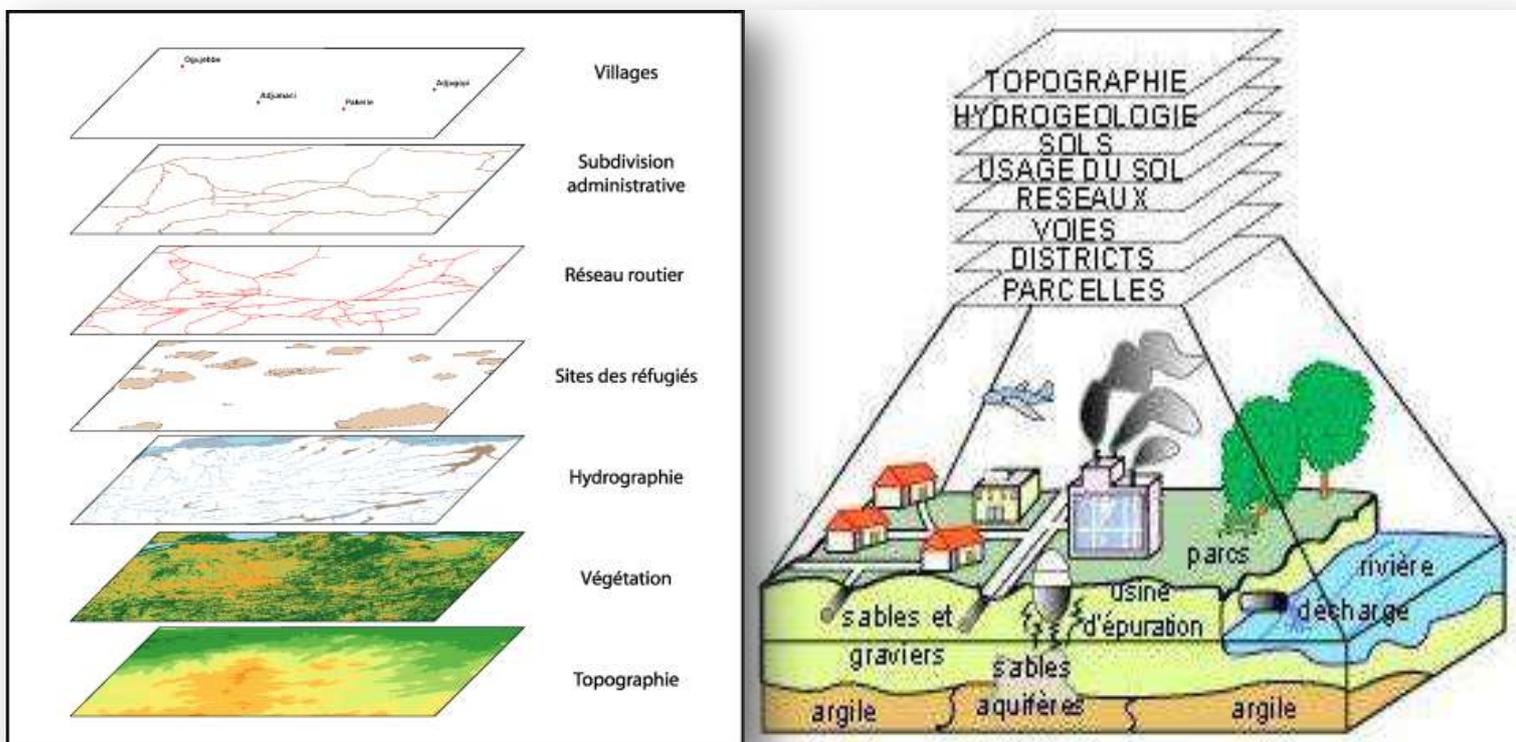


Figure 18: Le principe de couches dans un SIG.

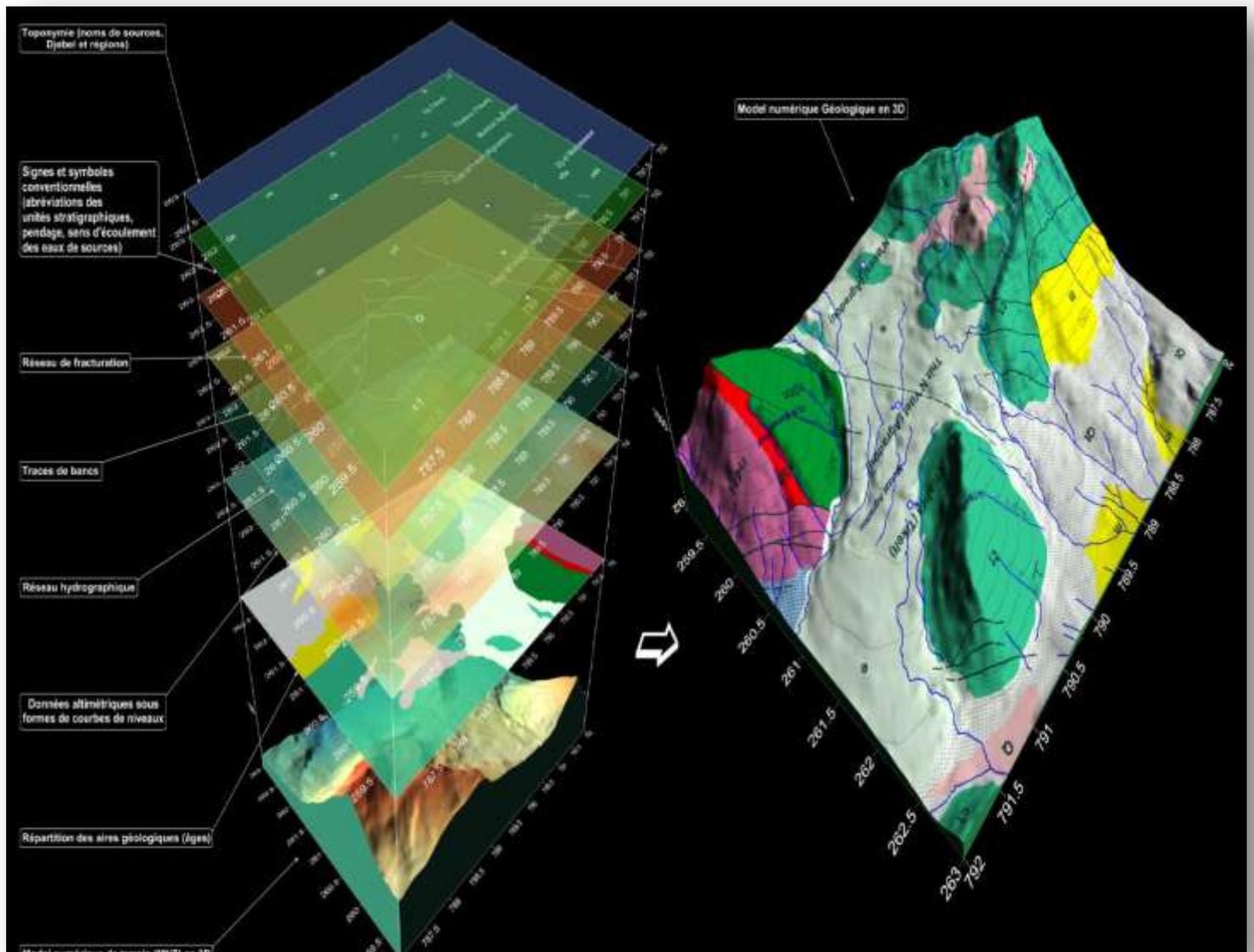


Figure 19 : Exemple d'un SIG multicouches tridimensionnelle dont la finalité est de bâtir un modèle géologique 3D (F.Benagoune 2014).