

Note à l'intention des lecteurs et utilisateurs des images de séries temporelles MODIS démélangées, produites par le laboratoire LETG Rennes COSTEL dans le cadre du projet CarHAB

Cette note est officieuse, et permet seulement d'interpréter un premier jeu d'images, merci de garder tout ça en interne pour le moment.

Rédigée par Karel JANIK (karel.janik@uhb.fr)

1. Comment fonctionne le démélangage ?

Il s'agit d'un démélangage non-supervisé. La première étape est une classification k-means à 20 clusters sur l'ensemble des séries temporelles de l'image. On a une série temporelle de 30 à 50 données par pixel et par année. Les données sont les valeurs de *lai*, *fcover* ou *fapar*. Seul le *fapar* est fourni dans ce premier jeu d'images (d'après une première analyse rapide, la variable *fapar* fournit de meilleurs résultats que les autres variables biophysiques), mais les autres sont également disponibles.

Ensuite, on calcul les barycentres des 20 clusters issus de la classif k-means. On a donc 20 séries temporelles de dimension identique à celle des pixels.

L'étape la plus intéressante consiste à chercher les 5 barycentres parmi les 20 calculés qui permettent de décrire au mieux les 15 autres (par une simple moyenne pondérée des séries temporelles de ces 5 barycentres). Les 5 séries temporelles qui ressortent gagnantes de cette recherche sont les *endmembers*, les séries temporelles les plus discriminantes.

Le résultat fourni par ce démélangage est donc une image à 5 bandes, chaque bande étant le pourcentage d'appartenance du pixel au cluster respectif. Un pixel pur sera à 100 % dans une bande. On obtient aussi une image d'erreur. En effet, chaque pixel possède une erreur, calculée avec une RMSE entre la moyenne pondérée des 5 séries temp *endmembers* et la série temp du pixel.

Pour le démélangage à 5 clusters, les pourcentages sont exprimés avec un pas de 5 %. Cette approximation change très peu les résultats et fait gagner plusieurs heures de calculs par image.

Tous les détails de calculs sont disponibles ici : http://hal.ird.fr/ird-00693533/file/Benhadj_automatic_Unmixing_MODIS_IJRS_R2_rev_accept.pdf (même méthode, mais dans cette publication, ils recherchent 3 *endmembers* parmi 20 clusters)

2. Que signifient les noms des fichiers?

- Images

Le nom des images (qui sont en Lambert II étendu, EPSG : 27572) permet de connaître d'un coup d'oeil la donnée de l'image, l'année, le département et la variable biophysique.

Exemple : *p04d42fapar* signifie qu'il s'agit d'une image de proportions (en opposition aux images d'erreur, qu'on ne fournira pas pour l'instant pour économiser des octets sauf sur demande), en 2004 dans le département de la Loire et qu'on utilise la donnée fapar. Les erreurs par pixel sont inférieures à 7 %.

Les 5 nombres qui suivent sont les indices des *endmembers* qui définissent l'image. Ils permettent simplement de retrouver les séries temporelles des barycentres correspondant dans un fichier que nous fournirons également sur demande. Enfin, *f5* à la fin du nom signifie qu'on a utilisé un pas de 5 % pour les calculs.

Quand *m90* vient se glisser dans le nom d'une image, cela signifie qu'on a masqué l'image en retirant les pixels recouvert à 90 % ou plus par les couches forêt (IGN) et bâti réunies (sauf pour le département 42, étant donné la gigantesque zone qui apparaît comme bâti à l'est de Bourges, où on a masqué seulement la forêt)

- Fichiers de validation

Pour les années 2006, 2008, 2009 et 2010, des fichiers de validation (qui sont d'ailleurs plutôt des fichiers d'analyse...) sont fournis avec les images (même nom avec *valid_* en amont, dans le dossier *valid*).

Dans ces fichiers, le graphique en haut à gauche rassemblent les séries temporelles des 5 endmembers. Les correspondances entre les bandes de l'image et les couleurs sont :

- + Rouge : bande 1
- + Vert : bande 2
- + Bleu : bande 3
- + Noir : bande 4
- + Cyan : bande 5

Les nuages de points suivant les séries temp des endmembers sont des points de validation. A chaque graphique correspond une occupation du sol différente, extraite de la BD Topo et du RPG (travaux de Alban THOMAS, ingénieur d'étude qui a travaillé sur le projet CarHAB avant l'arrivée de l'ingénieur actuel, Karel JANIK).

On ne considère ici que les pixels purs (recouverts à 90 % ou plus par une seule et même occupation du sol). Pour chacun de ces pixels purs, sur les graphiques, on a inscrit une croix ayant la couleur du endmember correspondant à la valeur de son pourcentage. Par exemple, si un pixel de validation possède les pourcentages 10%, 15%, 15%, 20% et 40 %, sur le graphique, nous aurons 5 croix de couleur différentes les unes au dessus de autres, avec comme valeurs d'ordonnées les pourcentage susdits.

Attention, un pixels de validation pur ne sera pas forcément pur selon le démélangeage !

Cette manière de valider est expérimentale ; il serait bon de l'améliorer, grâce aux retours des utilisateurs, et de trouver d'autres manières d'évaluer l'intérêt et la qualité du démélangeage (travail déjà en cours au laboratoire COSTEL)

NB : pour les autres années, sans données de validation, vous pourrez tout de même retrouver les profils temporels des clusters dans les autres fichiers pdf du dossier de validation. Il faut lire le premier graphique, avec l'erreur $Mk 0$, pour trouver les profils temporels correspondant à l'image (les autres graphiques représentent les combinaisons de endmembers donnant de moins bon résultats)

Vous aurez compris que les résultats de ce démixage et plus particulièrement les profils temporels des endmembers sont libres d'interprétation (même si dans la majorité des cas, on constate facilement une corrélation évidente avec les données de validation). La division « rigide » (pardon pour ce terme) du territoire en classes d'occupation du sol (« prairies », « cultures », « sol nu », etc) est à mettre de côté pour explorer ces images démixées. Etant donné l'échelle du pixel MODIS, énormément de mélanges sont possibles.

3. Possibilités

La suite dépend évidemment des retours que nous aurons de votre part ! Pour l'instant, nous envisageons de construire des arbres de démixage (plusieurs démixages successifs sur des pixels sélectionnés à l'aide de précédents démixage, ce qui permettrait d'introduire plus de variété dans les différentes occupations du sol).

Nous envisageons également d'utiliser les autres variables biophysiques pour l'instant peu utilisées pour déterminer de manière plus exacte les occupations du sol.

Enfin, nous travaillons toujours sur la fusion des données (démixage, BD Topo, RPG) afin de renseigner la segmentation IGN.