

# CARTES TOPOLOGIQUES ET CLASSIFICATION HIÉRARCHIQUE : APPLICATION A LA CLASSIFICATION DE LA COULEUR DE L’OCÉAN

Meziane YACOUB, Fouad BADRAN et Sylvie THIRIA  
Laboratoire LODYC case 100, Université Paris 6  
4, place Jussieu 75252 Paris cedex 05 France

## 1 Introduction

A la différence des algorithmes de type “nuées dynamiques” [2] ou les méthodes statistiques, la classification ascendante hiérarchique (CAH) est une méthode d’analyse de données permettant d’identifier les groupes de données sans en connaître leur nombre a priori. L’un des inconvénients de la CAH est qu’elle est mal adaptée à la gestion d’un nombre important d’objets à classer.

Les algorithmes de quantification vectorielle sont classiquement utilisés pour trouver un ensemble de vecteurs référents permettant de coder de manière compacte tous les vecteurs d’un ensemble de données. Leur répartition dépend de la fonction densité sous-jacente aux données. Comparées aux méthodes classiques de quantification vectorielle, les cartes topologiques introduisent une relation supplémentaire de voisinage entre les vecteurs référents. L’algorithme topologique le plus connu est celui proposé par Kohonen. D’autres généralisations de cet algorithme ont été proposées, elles proposent des algorithmes d’apprentissage qui permettent en plus d’approximer la fonction de densité de la population observée selon des modèles probabilistes plus complexes, PRSOM est l’un de ces algorithmes [1].

Dans cet article, nous proposons un algorithme de classification mixte qui combine l’algorithme probabiliste des cartes topologiques (PRSOM) et la CAH. Nous testons son efficacité sur un problème réel d’océanographie : la classification des spectres du radiomètre SeaWiFS.

## 2 Algorithme de classification mixte

Lorsque l’ensemble des observations est de grande dimension, une bonne méthode de classification consiste à quantifier l’espace, ce qui réduit le nombre des données, et à poursuivre la classification en effectuant une CAH sur les référents. Les cartes topologiques constituant une méthode de quantification vectorielle, leur utilisation en coopération avec la CAH a été proposée par Murthag [3]. Nous proposons dans cet article une méthode similaire qui consiste à utiliser un algorithme de carte topologique probabiliste (PRSOM), suivie d’une CAH utilisant l’indice de Ward.

L’algorithme de classification mixte que nous proposons procède en deux étapes.

**Etape 1 :** dans cette étape, nous utilisons l'algorithme PRSOM pour trouver un ensemble de référents (neurones) qui résument les observations initiales.

**Etape 2 :** cette étape consiste à effectuer une CAH sur les référents, les feuilles de l'arbre de classification sont alors les référents de la carte obtenue par PRSOM.

Le choix de ces deux algorithmes va permettre une classification prenant en compte, aux deux étapes de la méthode, les variances locales des observations que l'on veut classer. En effet, l'algorithme PRSOM permet de résumer l'information contenue dans les observations sous forme d'une famille de lois normales sphériques dont on a estimé les moyennes et les variances. Dans la carte probabiliste proposée par PRSOM, chaque neurone est une loi normale, l'ordre topologique permet d'en visualiser les proximités. On choisit pour la CAH qui va classer les neurones le critère de Ward qui utilise les variances locales estimées par PRSOM.

Nous présentons dans le paragraphe suivant les résultats obtenus par l'algorithme mixte sur un problème réel d'océanographie: la couleur de l'océan.

### 3 Classification de la couleur de l'océan

L'observation de la couleur de l'océan, qui consiste à mesurer la réflectance spectrale marine dans la partie visible du spectre solaire, permet d'obtenir des indications utiles sur la composition de l'eau, car ce sont les constituants minéraux et biologiques présents dans les premières couches marines qui sont essentiellement responsables des propriétés optiques de l'océan.

L'observation de la mer est effectuée soit in situ, sur des bateaux par exemple, soit depuis un satellite. Une observation satellitale a l'avantage par rapport à une observation in situ que sa couverture est globale. En revanche, pour donner une mesure précise de la luminance au niveau de la mer, dans une mesure satellitale, il faut tenir compte de la correction des effets optiques produits par la dispersion de la lumière dans l'atmosphère. Des capteurs embarqués sur satellite mesurent cette luminance et nous proposent une image de la surface de la mer pour différentes longueurs d'onde dans le domaine du visible et proche infrarouge.

#### 3.1 Données

Dans cet article, nous utilisons des mesures satellite provenant du capteur SeaWiFS. SeaWiFS<sup>1</sup> est un radiomètre embarqué sur le satellite SeaSTAR. Il fait un balayage de l'océan sur une zone de près de 1500 km de largeur au-dessus de la trajectoire du satellite. Il prend des mesures depuis septembre 1997. Les données associées à ce capteur existent sous différents niveaux hiérarchiques. Nous utilisons les données de type GAC de niveau 1 au sommet de l'atmosphère : il s'agit des luminances mesurées par un radiomètre pour chaque longueur d'onde; ce sont des mesures non corrigées. SeaWiFS a 8 bandes de fréquence : 412 nm, 443 nm, 490 nm, 510 nm, 555 nm, 670 nm, 765 nm et 865 nm.

L'image que nous allons traiter dans cet article est montrée sur la figure 1(a), c'est une région de l'image prise le 18 Mars 1998. Elle est composée de 194680 pixels.

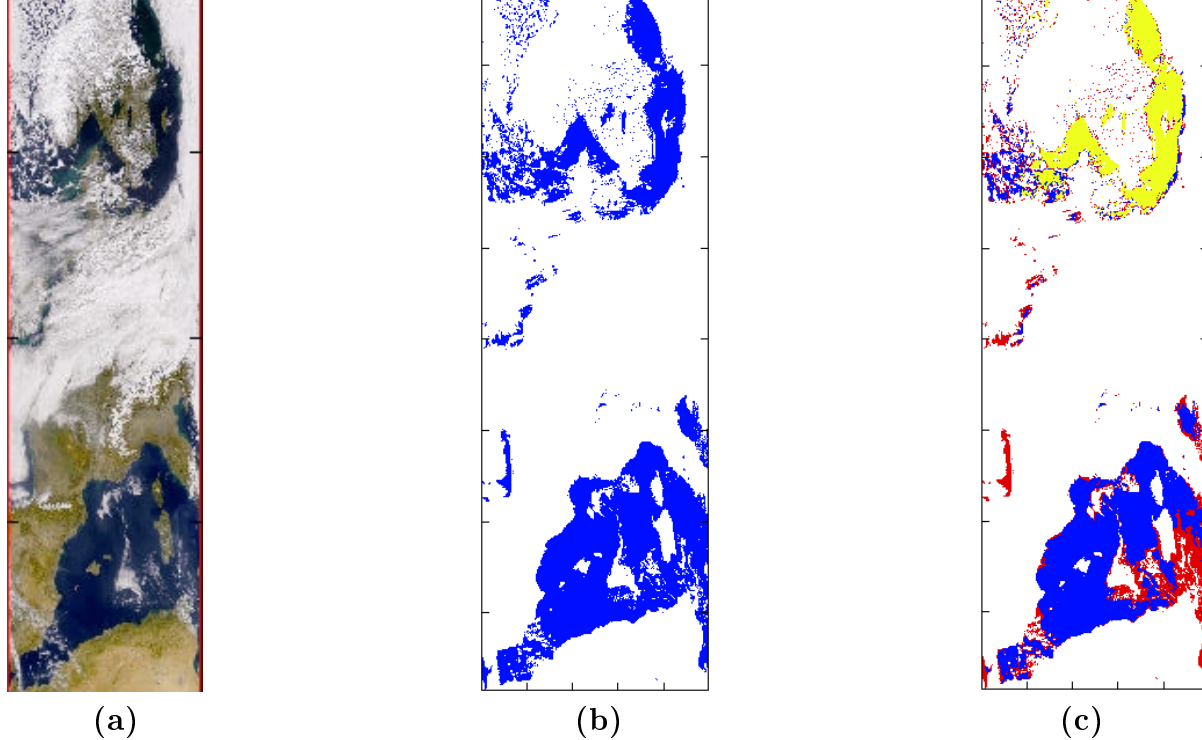


Figure 1: Cette figure montre la région traitée (a), le filtre obtenu en utilisant l'algorithme mixte (b) et les pixels correspondants aux 3 classes obtenues lors de la classification de la mer visible : la classe 1 en jaune, la classe 2 en bleu et la classe 3 en rouge.

### 3.2 Construction d'un masque

Le premier but à atteindre était de vérifier s'il était possible d'établir à partir des mesures prise au sommet de l'atmosphère un filtre permettant de séparer les pixels situés sur l'eau des pixels situés sur la terre et les nuages. Pour cela, nous avons utilisé uniquement les bandes 670 nm, 765 nm et 865 nm dans lesquelles le signal provenant de l'océan est nul ou négligeable. Après avoir entraîné une carte topologique en dimension 2 (10X10) avec PRSOM, nous avons appliqué la CAH et coupé l'arbre afin d'obtenir 2 classes.

La figure 1(b) montre le résultat de la classification en 2 classes. La région correspondant aux pixels de la terre et des nuages est celle montrée en couleur blanche. Nous avons comparé le masque que nous avons établi avec celui proposé par SeaWiFS. Sur la figure 2, une matrice de comportement affiche la répartition des pixels entre les deux classes eau et terre ou nuages, selon les classifications effectuées par SeaWiFS et l'algorithme mixte. 96, 28% des pixels sont classés de la même façon par SeaWiFS et notre algorithme. Tous les pixels que nous avons déterminés comme terre ou nuages sont identifiés de la même façon par SeaWiFS. Les 3, 72% des pixels en différence sont classés eau par notre algorithme et terre ou nuages par SeaWiFS. Ainsi SeaWiFS trouve légèrement plus de terre ou nuages que notre algorithme. Cette différence peut s'expliquer ainsi : SeaWiFS séparerait la mer de la terre et tous les nuages, alors que notre algorithme séparerait la mer de la terre et les nuages forts uniquement. En effet, par la suite, notre algorithme isole dans la mer une catégorie de pixels correspondant aux voiles nuageux. Clairement, l'algorithme mixte permet de retrouver ces résultats de manière non supervisée. On peut donc conclure que dans le cas des données LAC de très haute résolution qui ne contiennent pas l'information

<sup>1</sup>Page web sur le projet SeaWiFS <http://seawifs.gsfc.nasa.gov/SEAWIFS.html>

		SeaWiFS	
		terre ou nuages	mer visible
Algorithme mixte	terre ou nuages	<b>149791</b>	<b>0</b>
	mer visible	<b>7235</b>	<b>37654</b>

Figure 2: Répartition des pixels entre les deux classes eau ou terre et nuages selon les classifications effectuées par SeaWiFS et l’algorithme mixte.

sur la terre et les nuages, il sera possible de définir très finement et automatiquement ce masque.

### 3.3 Classification de la mer visible

Après avoir séparé la mer visible de la terre et des nuages, nous avons appliqué notre méthode uniquement aux pixels de l’eau en utilisant les 8 bandes de fréquence. De même que précédemment, nous avons utilisé une carte 10X10. Le résultat obtenu en demandant à la CAH de classer les pixels de l’image en trois classes est montré sur la figure 1(c). Selon les experts du domaine, la classe 1 correspond aux eaux de la Baltique dont l’optique est influencée par la présence de “substances jaunes” (produits de dégradation végétale), la classe 2 correspond à des eaux claires, majoritaires en méditerranée et la classe 3 correspond aux voiles nuageux au-dessus de la mer. Les trois classes obtenues montrant une grande cohérence avec les connaissances géophysiques de l’expert, nous pouvons maintenant faire confiance à la carte proposée par PRSOM et investiguer les différentes partitions proposées par la CAH. La CAH permet d’obtenir jusqu’à cent classes (puisque notre carte topologique est constituée de 100 neurones). Chaque partition permettant d’obtenir des informations sur la géophysique.

## 4 Conclusion

Dans cet article, nous avons proposé une méthode de classification mixte combinant l’algorithme PRSOM et l’algorithme CAH. Actuellement, nous améliorons cet algorithme en ajoutant une procédure automatique qui permet de suggérer le nombre de classes à retenir en se basant sur le dendrogramme fourni par la CAH.

## References

- [1] Anouar F., Badran F. et Thiria S., (1997) : Self Organized Map, A Probabilistic Approach. proceedings of the Workshop on Self-Organized Maps. Helsinki University of Technology, Espoo, Finland, June 4-6.
- [2] Diday E. et collaborateurs, (1980) : Optimisation en classification automatique. Editeur INRIA
- [3] Thiria S., Lechevallier Y., Gascuel O. et Canu S., (1997) : Statistique et méthodes neuronales. Dunod