

# Noyau de Mahalanobis pour la Classification d'Images Hyperspectrales : Cartographie du Frêne dans les Hautes-Pyrénées

Mathieu FAUVEL<sup>1</sup>, Alberto VILLA<sup>2,3</sup>, Jocelyn CHANUSSOT<sup>2</sup>, David SHEEREN<sup>1</sup>, Jon Atli BENEDIKTSSON<sup>3</sup>

<sup>1</sup>UMR 1201 DYNAFOR, INRA & University of Toulouse, BP 52627, 31326 Castanet Tolosan Cedex, France

<sup>2</sup>GIPSA-lab, Grenoble Institute of Technology, France

<sup>3</sup>Faculty of Electrical and Computer Engineering, University of Iceland, Iceland

mathieu.fauvel@ensat.fr, alberto.villa@hyperinet.eu,

jocelyn.chanussot@gipsa-lab.grenoble-inp.fr, david.sheeren@ensat.fr, benedikt@hi.is

**Résumé** – Un noyau adapté à la grande dimension spectrale des images hyperspectrales est proposé dans cet article. Ce noyau repose sur la distance de Mahalanobis calculée pour chaque classe. Le calcul est rendu possible en grande dimension par l'utilisation d'un modèle probabiliste parcimonieux permettant de stabiliser l'inversion de la matrice de covariance lors du calcul de la distance. Ce noyau réalise un compromis entre un noyau appliqué sur les données originales et un noyau appliqué les données projetées sur les premiers axes propres de la classe considérée. Cette approche est comparée à d'autres méthodes pour classifier le frêne sur des images hyperspectrales. En termes de précision, le noyau proposé permet d'améliorer les résultats par rapport au noyau classique de Mahalanobis et donne des résultats similaire au noyau gaussien.

**Abstract** – A kernel adapted to the high spectral dimension of hyperspectral images is discussed in this article. The kernel is based on the Mahalanobis distance defined for each class. A parsimonious probabilistic model is used for the inversion of the covariance matrix. The kernel reaches a trade-off between a conventional Gaussian kernel and a Gaussian kernel on the first principal components of the considered class. The proposed approach is compared to other methods for the classification of Ash tree in hyperspectral images. In terms of accuracies, the proposed kernel performs better than conventional Mahalanobis kernel and performs as well as the conventional Gaussian kernel.

## 1 Introduction

Les avancées techniques des imageurs hyperspectraux embarqués permettent l'acquisition de spectres de longueurs d'onde échantillonnés très finement. Par exemple, le capteur aéroporté Hypspx, utilisé pour l'acquisition d'images dans le cadre du projet MADONNA [1], enregistre pour chaque pixel 160 bandes spectrales contiguës, allant du visible au proche infrarouge. En conséquence, chaque pixel  $x$  de l'image est un vecteur de dimension égale au nombre de bandes spectrales :  $x \in \mathbb{R}^{160}$ . Grâce à la précision de l'échantillonnage spectral, il est possible d'accéder à la nature physique et aux paramètres biophysiques des objets visés par le capteur (reconnaissance des espèces d'arbres, densité du couvert végétal, caractérisation des prairies ...) [2]. Cependant, l'accroissement de l'information disponible va de pair avec une augmentation de la complexité à traiter ces données. Ce phénomène est connu sous le nom de *fléau de la dimension* (pour une revue détaillée voir [3, 4, 5]). En particulier, les approches statistiques classiques sont difficiles à mettre en œuvre, principalement en raison de problèmes d'estimation [6], et les résultats obtenus ne sont pas suffisamment fiables.

Indépendamment de ces problèmes, la dernière décennie a vu la démocratisation des *méthodes à noyaux* dans de nombreux domaines scientifiques [7] (biomédical, traitement du si-

gnal et des images, géosciences ...). Contrairement aux méthodes statistiques paramétriques, ces méthodes ne nécessitent pas la définition d'un modèle, seulement la définition d'une fonction noyau calculant une similarité entre deux objets (scalaire, pixel-vecteur, chaîne de caractères ...). Parmi ces méthodes, les Machines à Vecteurs Supports (SVM) sont sûrement les plus employées. Pour le traitement d'images hyperspectrales, on les retrouve principalement pour des problèmes de classification [8], de régression/inversion [9], de démixage spectral [10] ou de réduction de dimensions [9].

La fonction noyau joue un rôle clef dans les traitements : elle définit la *similarité* entre deux éléments. Par exemple, avec le noyau gaussien, la similarité est associée à la distance euclidienne alors que pour un noyau polynômial, la similarité est associée à la colinéarité. Plusieurs noyaux ont été définis pour les données hyperspectrales. Dans [11], les auteurs définissent des noyaux prenant en compte le caractère spectral des données. Ils proposent d'utiliser l'angle spectral et/ou la divergence de Kullback-Leibler à la place de la distance Euclidienne dans le noyau gaussien. Dans [12, 13], le noyau gaussien est modifié durant la phase d'apprentissage, de manière à améliorer la séparableté des données et la robustesse au bruit. Plus récemment, dans des problèmes de démixage spectral, un noyau modélisant les phénomènes physiques de mélange a été proposé dans [10]. D'autres noyaux ont été proposés, prenant en

compte la corrélation spatiale entre les pixels mais sans prendre en compte la grande dimension, par exemple [9, Chap. 10]. Actuellement, la définition d'un noyau adapté aux données spectrales de grande dimension est toujours un axe de recherche important.

Dans cet article, l'utilisation du noyau de Mahalanobis est proposée pour la classification d'images hyperspectrales. Pour ce noyau, la distance Euclidienne du noyau gaussien est remplacé par la distance de Mahalanobis. Cette distance nécessite l'inversion de la matrice de covariance des données. En grande dimension, ce problème inverse est mal conditionné et il est nécessaire de le régulariser. Nous proposons l'utilisation du modèle issu de l'analyse en composante principale probabiliste [14] pour la construction du noyau.

La méthode proposée est testée sur des données hyperspectrales réelles, acquises dans le cadre du projet MADONNA.

## 2 Noyau de Mahalanobis

Le noyau de Mahalanobis appartient à la famille des noyaux radiaux [9], il s'écrit ainsi :

$$k(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{(\mathbf{x} - \mathbf{z})^t \hat{\Sigma}^{-1} (\mathbf{x} - \mathbf{z})}{2\sigma^2}\right) \quad (1)$$

où  $\mathbf{x}, \mathbf{z} \in \mathbb{R}^d$ ,  $\hat{\Sigma}^{-1}$  est l'estimation empirique de la matrice de covariance des données et  $\sigma$  est un hyperparamètre strictement positif du noyau. On retrouve deux définitions dans la littérature pour la matrice de covariance : soit elle est définie comme la covariance totale des données, indépendamment des classes ; soit elle est définie classe par classe et le noyau est spécifique à chaque classe. Dans cet article la deuxième définition est considérée car elle permet l'utilisation du modèle probabiliste de l'analyse en composante principale (PPCA). La classe considérée est notée  $c$  et la matrice de covariance associée  $\Sigma_c$ .

### 2.1 Inversion de la matrice de covariance

En dimension modérée,  $d \approx 10$ , l'utilisation de ce noyau a montré des résultats supérieurs en termes de précision de classification par rapport au noyau gaussien classique [15, 16]. Cependant en grande dimension,  $d \gg 10$ , l'estimation de la matrice de covariance est mal conditionnée et son inversion est instable, voir impossible, et cela pour deux raisons :

1. Le nombre important de paramètres à estimer :  $d(d+1)/2$ , soit 12880 pour  $d = 160$ .
2. La faible vitesse de convergence des estimateurs en grande dimension [3].

L'approche proposée dans cet article a pour but de réduire le nombre de paramètres à estimer sans perdre d'information discriminante. Elle utilise une des propriétés fondamentales des espaces à grande dimension : la parcimonie. Ainsi, les données sont réparties dans un sous-espace  $\mathcal{A}_c$  de  $\mathbb{R}^d$ , le sous-espace supplémentaire  $\bar{\mathcal{A}}_c$  étant composé principalement du

bruit. Pour estimer ces sous-espaces, le modèle PPCA est utilisé ici. Il suppose que les  $d$  variables  $\mathbf{x}[i]$  observées pour la classe  $c$  sont une combinaison linéaire de  $p_c$  variables  $s[i]$  cachées ( $p_c \ll d$ ) :  $\mathbf{x} = \mathbf{W}_c \mathbf{s} + \boldsymbol{\mu}_c + \boldsymbol{\varepsilon}_c$ , avec  $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  et  $\boldsymbol{\varepsilon}_c \sim \mathcal{N}(\mathbf{0}, \varepsilon_c^2 \mathbf{I})$ . Il en résulte que  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_c, \mathbf{W}_c \mathbf{W}_c^t + \varepsilon_c^2 \mathbf{I})$  et que  $\Sigma_c$  peut s'écrire sous la forme suivante :

$$\Sigma_c = \sum_{i=1}^{p_c} (\lambda_{ci}^2 + \varepsilon_c^2) \mathbf{u}_{ci} \mathbf{u}_{ci}^t + \varepsilon_c^2 \sum_{i=p_c+1}^d \mathbf{u}_{ci} \mathbf{u}_{ci}^t \quad (2)$$

où  $\lambda_{ci}$  est la  $i^{\text{ème}}$  valeur singulière de  $\mathbf{W}_c$  et  $\mathbf{u}_{ci}$  le vecteur singulier à gauche correspondant. Le premier terme à droite de l'égalité correspond à la variance/covariance dans le sous-espace signal  $\mathcal{A}_c$  et le second à la variance/covariance dans le sous-espace bruit  $\bar{\mathcal{A}}_c$ . Ces paramètres sont estimés à l'aide de la matrice de covariance empirique de la classe considérée. Il s'ensuit que  $\Sigma_c^{-1}$  peut s'écrire de la manière suivante :

$$\Sigma_c^{-1} = \sum_{i=1}^{p_c} \left( \frac{-\lambda_{ci}^2}{\varepsilon_c^2 (\lambda_{ci}^2 + \varepsilon_c^2)} \right) \mathbf{u}_{ci} \mathbf{u}_{ci}^t + \frac{1}{\varepsilon_c^2} \mathbf{I} \quad (3)$$

Le nombre de paramètres ( $\{\lambda_{ci}, \mathbf{u}_{ci}\}_{i=1}^{p_c}, \varepsilon_c$ ) à estimer est maintenant :  $d(p_c + 1) + 2 - p_c(p_c - 1)/2$ , soit 1717 pour  $d = 160$  et  $p_c = 10$ . La dimension  $p_c$  du sous-espace signal est estimée en minimisant le critère BIC [17] :

$$\text{BIC}(p_c) = -2l + (d(p_c + 1) + 2 - p_c(p_c - 1)/2) \log(n_c) \quad (4)$$

où  $l$  est la log-vraisemblance associée au modèle PPCA pour la valeur de  $p_c$  considérée et  $n_c$  le nombre de pixels d'entraînements de la classe considérée.

### 2.2 Construction du noyau

En insérant (3) dans (1) et en y remplaçant les paramètres par leur estimées, le noyau de Mahalanobis régularisé s'écrit

$$k_M(\mathbf{x}, \mathbf{z}|c) = \exp\left(-\sum_{i=1}^{\hat{p}_c} \left( \frac{-\hat{\lambda}_{ci}^2}{\hat{\varepsilon}_c^2 (\hat{\lambda}_{ci}^2 + \hat{\varepsilon}_c^2)} \right) \frac{(\hat{\mathbf{u}}_{ci}^t (\mathbf{x} - \mathbf{z}))^2}{2\sigma_i^2} - \frac{\|\mathbf{x} - \mathbf{z}\|^2}{2\sigma_{\hat{p}_c+1}^2 \hat{\varepsilon}_c^2} \right) \quad (5)$$

où  $\sigma_i$  sont des hyperparamètres du noyau. Si  $\sigma_i = +\infty, \forall i \in \{1, \dots, \hat{p}_c\}$ , on retrouve le noyau gaussien classique, et si  $\sigma_{\hat{p}_c+1} = +\infty$  on retrouve le noyau gaussien appliqué aux données projetées sur les  $\hat{p}_c$  axes propres de la classe considérée.

Notons pour conclure cette partie que pour une classification SVM multi-classe, ce noyau impose une stratégie de type « un contre tous », la matrice de covariance étant estimée classe par classe. De plus, le classifieur construit pour séparer la classe  $i$  de la classe  $j$  n'est pas le même que celui construit pour séparer la classe  $j$  de la classe  $i$  : dans un cas on considère  $\Sigma_i$  et dans l'autre  $\Sigma_j$ .

## 3 Résultats expérimentaux

Le noyau proposé a été testé sur des images hyperspectrales réelles issues du capteur Hypspec. Ces images, acquises dans le

TABLE 1 – Paramètre du modèle PPCA (voir Table 2 pour le nom des classes).

CLASSE	1	2	3	4	5	6	7
$\hat{p}_c$	30	31	32	32	31	31	30
$\lambda_1$	9.30	2.35	4.35	18.3	14.6	2.71	1.41
$\lambda_2$	0.58	0.19	0.84	1.19	0.67	0.24	0.44
$\lambda_3$	0.11	0.01	0.26	0.35	0.26	0.08	0.03
$\varepsilon^2 (\times 10^{-3})$	3.3	1.5	2.4	3.8	6.1	1.6	3.8

cadre du projet MADONNA, doivent permettre une cartographie précise du frêne et plus généralement permettre la reconnaissance des différentes espèces d’arbres présentes sur deux zones d’études du laboratoire DYNAFOR (Canton d’Aurignac et Vallée du Gave de Pau). Elles possèdent 160 bandes spectrales, de 400 à 1000 nm et ont une résolution spatiale de 50 cm. Une vérité terrain a été réalisée à l’aide d’images de la BD ORTHO® (IGN) et de reconnaissance sur le terrain. 32224 pixels ont été labélisés en 12 classes de végétation. Pour l’étape de classification, seulement 7 classes ont été retenues comme ayant suffisamment de pixels ( $> 1400$ ) : Châtaignier, Tilleul, Frêne, Chêne, Fougère, Noisetier, Robinier. Pour la phase d’entraînement, un ensemble d’entraînement composé de 1400 pixels par classe a été sélectionné, le reste des pixels (20248) forme l’ensemble de test.

### 3.1 Analyse des paramètres du modèle PPCA

Les estimations de la taille du sous-espace signal ainsi que des trois premières valeurs propres et de la variance du sous-espace bruit sont données pour chaque classe dans la Table 1.  $\hat{p}_c$  ne diffère pas sensiblement en fonction de la classe considérée et correspond pour chaque classe à environ 98 % de la variance cumulée. On retrouve une valeur élevée pour la première valeur propre de chaque classe, attribuable à l’albédo moyen. Quant à la variance du sous-espace bruit, elle varie significativement entre les classes.

### 3.2 Analyse des résultats de classification

L’approche proposée a été comparée avec un classifieur paramétrique de type mélange de gaussiennes (GMM [2]) avec une régularisation *ridge* et un classifieur SVM utilisant : un noyau gaussien classique, un noyau de Mahalanobis non régularisé, un noyau de Mahalanobis avec une régularisation de type *ridge*. Les hyperparamètres du noyau ont été réglés par minimisation d’une estimation de l’erreur de prédiction (borne rayon-marge) [13]. Les résultats de classification sont reportés dans la Table 2. OA, pour overall accuracy, représente le pourcentage global de pixels bien classés ;  $\kappa$  est le Kappa de Cohen qui mesure l’accord global entre la prédiction et la vérité terrain, en tenant compte de l’accord dû à la chance. La précision utilisateur pour chaque classe est aussi reportée. Plus ces indices sont proches respectivement de 100% et de 1, meilleure est la classification.

D’après ces résultats, l’utilisation du noyau de Mahalanobis

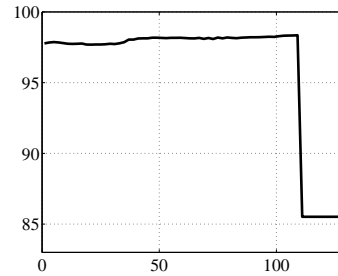


FIGURE 1 – Performance du classifieur « Frêne vs all » en fonction de  $\hat{p}_c$ .

régularisé par l’approche proposée améliore sensiblement la précision de la classification par rapport aux noyaux de Mahalanobis conventionnels. Le noyau gaussien et le noyau proposé donnent des résultats similaires, dont les différences ne sont pas statistiquement significative (test de McNemmar négatif). Le fort rapport signal à bruit (RSB) des données peut expliquer ces résultats ; pour des données dont le RSB est plus faible, le noyau proposé devrait permettre d’obtenir de meilleurs résultats. Ce dernier point est actuellement en considération. Enfin, notons les bons résultats du classifieur GMM, qui sont cependant significativement moins bons que ceux obtenus avec les SVM.

Concernant la classification du frêne, les meilleurs résultats sont obtenus pour les SVM avec le noyau gaussien et le noyau proposé. Ces résultats montrent qu’il est possible de discriminer le frêne de manière précise parmi divers espèces boisées à l’aide des images hyperspectrales.

### 3.3 Analyse de l’influence de l’estimation de $p_c$

La figure 1 montre l’évolution de la performance du classifieur « Frêne vs all » en fonction de  $\hat{p}_c$ . Pour le frêne, la valeur de  $\hat{p}_c$  estimée par le BIC est de 32 (voir Table 1).

Cette figure nous donne deux renseignements importants. Le premier est que le BIC n’arrive pas à sélectionner pour cette classe la valeur optimale de  $\hat{p}_c$ . Le second est que globalement, l’influence de  $\hat{p}_c$  reste limitée sur la performance du classifieur : une large plage de valeurs pour  $\hat{p}_c$  (1 à 35) donne des résultats plus faibles de quelques dixièmes, tandis que les meilleurs résultats sont obtenus pour des valeurs plus élevées ( $> 40$ ). Cependant, l’utilisation d’un trop grand nombre d’hyperparamètres ( $> 110$ ) perturbent l’étape d’entraînement des SVM, l’algorithme d’optimisation est bloqué à la première itération.

## 4 Conclusion

Un noyau de Mahalanobis adapté à la classification de données à grande dimension est proposé dans cet article. Il suppose que les données de chaque classe suivent le modèle PPCA. Pour chaque classe, les sous-espaces signal et bruit sont estimés et sont utilisés pour approximer le calcul de la distance de Mahalanobis. Ce noyau réalise un compromis entre un noyau appliqué sur les données originales et un noyau appliqué sur

TABLE 2 – Résultats de classification. OA : précision globale en pourcentage,  $\kappa$  : coefficient kappa. Pour chaque classe est reportée la précision utilisateur en pourcentage.

	GAUSSIEN	MAHALANOBIS	MAHALANOBIS RIDGE	MAHALANOBIS PPCA	GMM
OA	94,4	54,8	68,5	94,1	91,0
$\kappa$	0.923	0.425	0.576	0.918	0.876
1. CHÂTAIGNER	74.5	38.9	35.4	71.3	85.2
2. ROBINIER	95.3	15.9	32.6	92.9	97.7
3. TILLEUL	94.1	61.6	65.9	98.9	75.2
4. FRÊNE	97.7	36.2	34.4	96.1	84.5
5. NOISETIER	92.4	64.4	63.4	93.4	91.4
6. CHÊNE	98.2	85.3	92.3	98.1	96.7
7. FOUGÈRE	92.8	24.6	40.6	89.9	96.4

les données projetées sur les premiers axes propres de la classe considérée.

Ce noyau a été utilisé pour classifier des images hyperspectrales en vue de la cartographie du frêne dans les Pyrénées. Comparativement au noyau de Mahalanobis conventionnel, les résultats obtenus en termes de précision sont satisfaisants. Cependant, ces résultats restent du même ordre que ceux obtenus avec un noyau gaussien classique. Une perspective de ces travaux concerne l'étape d'optimisation des hyperparamètres du noyau. Actuellement, la méthode utilisée pour les régler est à la fois lourde en temps de calculs et instable (existence de minima locaux) ce qui perturbe le fonctionnement de la méthode. Une autre perspective concerne l'estimation du paramètre  $\hat{p}_c$  : le BIC fournit une valeur exploitable mais non optimale.

Au niveau applicatif, la cartographie du frêne sur l'ensemble des zones d'études reste à réaliser. En raison du volume des données à traiter, l'amélioration de la sélection des hyperparamètres est un préalable.

## Références

- [1] D. Sheeren, M. Fauvel, S. Ladet, A. Jacquin, G. Bertoni, and A. Gibon, "Mapping ash tree colonization in an agricultural mountain landscape : Investigating the potentialities of hyperspectral imagery," in *Geoscience and Remote Sensing Symposium, 2011. IGARSS 2011. IEEE International*, submitted.
- [2] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. New Jersey : John Wiley and Sons, 2003.
- [3] D. L. Donoho, "High-dimensional data analysis : The curses and blessings of dimensionality," *American Math Society on Math Challenges of the 21st century*, pp. 1–32, 2000.
- [4] L. Jimenez and D. A. Landgrebe, "Supervised classification in high dimensional space : geometrical, statistical and asymptotical properties of multivariate data," *IEEE Trans. Syst., Man, Cybern. B*, vol. 28, pp. 39–54, Feb. 1993.
- [5] C. Aggarwal, A. Hinneburg, and D. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *Database Theory — ICDT 2001*, Lecture Notes in Computer Science, pp. 420–434, Springer Berlin / Heidelberg, 2001.
- [6] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. on Information Theory*, vol. IT-14, pp. 55–63, Jan. 1968.
- [7] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *The annals of statistics*, vol. 36, no. 3, pp. 1171–1220, 2008.
- [8] M. Fauvel, J. A. Benediktsson, and J. Chanussot, "Classification en imagerie hyperspectrale haute résolution : utilisation conjointe de l'information spatiale et spectrale," in *Actes du 21ème colloque GRETSI sur le Traitement du Signal et des Images, GRETSI 2007 21ème Colloque GRETSI Traitement du Signal & des Images, GRETSI 2007*, pp. 1–4, 2007.
- [9] G. Camps-Valls and L. Bruzzone, eds., *Kernel Methods for Remote Sensing Data Analysis*. Wiley, 2009.
- [10] J. Broadwater and A. Banerjee, "A generalized kernel for areal and intimate mixtures," in *Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*, pp. 1–4, June 2010.
- [11] G. Mercier and M. Lennon, "Support vector machines for hyperspectral image classification with spectral-based kernels," in *Geoscience and Remote Sensing Symposium*, vol. 1, IGARSS '03. Proceedings, July 2003.
- [12] B. Guo, S. Gunn, R. Damper, and J. Nelson, "Customizing kernel functions for svm-based hyperspectral image classification," *Image Processing, IEEE Transactions on*, vol. 17, pp. 622–629, Apr. 2008.
- [13] A. Villa, M. Fauvel, J. Chanussot, P. Gamba, and J. A. Benediktsson, "Gradient optimization for multiple kernel's parameters in support vector machines classification," in *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, July 2008.
- [14] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.
- [15] S. Abe, "Training of support vector machines with Mahalanobis kernels," in *Artificial Neural Networks : Formal Models and Their Applications - ICANN 2005*, Lecture Notes in Computer Science, pp. 571–576, Springer Berlin / Heidelberg, 2005.
- [16] G. Camps-Valls, A. Rodrigo-Gonzalez, J. Muñoz-Mari, L. Gomez-Chova, and J. Calpe-Maravilla, "Hyperspectral image classification with Mahalanobis relevance vector machines," in *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pp. 3802–3805, July 2007.
- [17] G. Schwarz, "Estimating the dimension of a model," *The annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.