

# Discrimination périodique à partir d’observations multi-temporelles

Rémi FLAMARY<sup>1</sup>, Khalil HARRANE<sup>1</sup>, Mathieu FAUVEL<sup>2</sup>, Silvia VALERO<sup>3</sup> Mauro DALLA MURA<sup>4</sup>

<sup>1</sup>Laboratoire Lagrange, UMR CNRS 7293, OCA, Université Côte d’Azur, France

<sup>2</sup> Université de Toulouse, INP-ENSAT, UMR 1201 DYNAFOR, France et INRA, UMR 1201 DYNAFOR, France

<sup>3</sup>CESBIO - CNES, CNRS (UMR 5126), IRD, Université de Toulouse, France

<sup>4</sup>GIPSA-lab, UMR CNRS 5216, Grenoble INP, France

remi.flamary@unice.fr, mathieu.fauvel@ensat.fr, mauro.dalla-mura@gipsa-lab.fr

**Résumé** – Cet article s’attaque au problème de discrimination instantanée de données temporelles périodiques. Nous proposons un classifieur linéaire non-stationnaire, régularisé par rapport à sa complexité temporelle. Le modèle est également étendu au cas multi-classe. Cette approche est testée sur des simulations numériques ainsi que des séries temporelles d’images satellites. Les résultats illustrent un très bon comportement.

**Abstract** – In this work, we propose a novel linear classification scheme for non-stationary periodic data. We express the classifier in a temporal basis while regularizing its temporal complexity leading to a convex optimization problem. Numerical experiments show very good results on a simulated example and on real life remote sensing image classification problem.

## 1 Introduction

Depuis l’avènement des séries temporelles d’images satellitaires (STIS), il est possible d’observer des territoires avec une résolution temporelle fine (quelque jours) et une couverture temporelle large (plusieurs années). Ces STIS sont extrêmement utiles pour caractériser l’état des paysages et leur évolution. En particulier, l’extraction de l’occupation et de l’utilisation des sols est une problématique importante.

Les objets présents dans les STIS présentent des comportements évolutifs majoritairement non-stationnaires. Par exemple, la valeur moyenne ainsi que la variance de la réflectance des pixels associés à une parcelle agricole vont varier au cours d’une année, cela en raison de l’évolution du stade phénologique de la culture. Cependant, ces variations ont une nature périodique à l’échelle de plusieurs années, le cycle de végétation se renouvelant périodiquement au fil des saisons.

Malheureusement, les observations sur de longues durées présentent un échantillonnage temporel irrégulier à l’échelle pluriannuelle, notamment en raison des conditions météorologiques (nuages et ombres principalement). Une approche classique en télédétection consiste à pré-traiter les STIS pour interpoler les dates manquantes [1, 2]. La méthode de filtrage fait généralement les hypothèses de périodicité et de *douceur* sur l’évolution temporelle de la réflectance. Une fois ces données filtrées, un algorithme de classification est utilisé pour prédire l’occupation des sols.

Une autre problématique concerne la classification instantanée d’une nouvelle image à partir des STIS du passé. Dans ce cas, il y a très peu de chance d’avoir une correspondance entre la date d’acquisition et les dates disponibles en apprentissage.

Il est alors nécessaire d’adapter les traitements pour pouvoir classifier la nouvelle image sans entraîner le système à nouveau [6].

Dans ce travail, nous proposons d’aborder le problème d’échantillonnage temporel irrégulier de manière différente. Les données ne sont pas pré-traitées mais utilisées directement lors de l’estimation du classifieur. Un classifieur linéaire dont les paramètres évoluent au cours du temps est proposé pour prendre en compte la non-stationnarité des données ainsi que leur échantillonnage temporel irrégulier. Le classifieur périodique est exprimé dans une base fonctionnelle comme par exemple la base de Fourier. Les pondérations dans la base sont estimées en maximisant un critère de discrimination, contrairement aux méthodes classiques de lissage, où les poids de la décomposition sont appris en minimisant une erreur de reconstruction sur les données.

Le modèle proposé est décrit dans la section 2, puis les résultats de classification obtenus sur des données simulées et réelles sont discutés dans la section 3. La section 4 présente les conclusions et les perspectives de ces travaux.

## 2 Discrimination périodique

Dans cette section, les données et le modèle sont présentés, puis le problème d’optimisation et la régularisation temporelle sont discutés.

**Données et modélisation** Les données d’apprentissage sont définies par une liste de  $n$  échantillons  $\{\mathbf{x}_i, y_i, t_i\}_{i=1, \dots, n}$ , où  $\mathbf{x}_i \in \mathbb{R}^d$  est un vecteur contenant  $d$  caractéristiques, de classe  $y_i \in \{-1, 1\}$  et  $t_i$  est l’instant temporel de l’observation. Dans

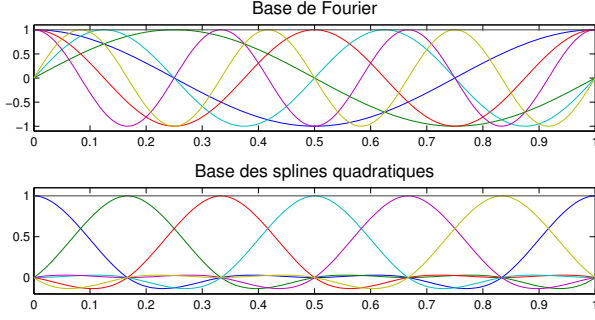


FIGURE 1 – Illustration des fonctions de base de Fourier (haut) et splines quadratiques (bas) pour  $p = 7$ . La courbe noire est la fonction constante  $g_p(t) = 1$ .

une application de télédétection,  $\mathbf{x}_i$  est un pixel contenant  $d$  bandes spectrales,  $y_i$  la classe (culture) et  $t_i$  l’instant temporel de l’acquisition le long de l’année. Il est supposé que la densité jointe entre les trois variables aléatoires est périodique par rapport au temps, c’est-à-dire que  $\mathbb{P}(\mathbf{x}, y, t) = \mathbb{P}(\mathbf{x}, y, t + 1)$ , où la période de 1 a été choisie arbitrairement.

Dans ces travaux, nous nous concentrons sur un classifieur linéaire périodique qui peut être exprimé sous la forme

$$f(\mathbf{x}, t) = \mathbf{w}(t)^\top \mathbf{x} = \sum_{j=1}^d w_j(t) x_j \quad (1)$$

où  $\mathbf{w}(t) \in \mathbb{R}^d$  est le vecteur normal à l’hyperplan séparateur dépendant du temps, ce qui permettra à la frontière de décision d’évoluer le long du temps<sup>1</sup>. Pour construire une fonction discriminante  $f(\mathbf{x}, t)$  périodique par rapport à  $t$ , les paramètres  $\mathbf{w}(t)$  sont exprimés dans une base de fonctions périodiques  $g_k(t)$  :

$$\mathbf{w}(t) = \Psi \mathbf{g}(t) \quad (2)$$

où  $\mathbf{g}(t) = [g_1(t), \dots, g_p(t)]^\top$  est un vecteur de  $p$  fonctions de base périodiques et  $\Psi \in \mathbb{R}^{d \times p}$  est une matrice qui pondère ces fonctions pour chacune des caractéristiques. La fonction de décision peut donc s’écrire :

$$f(\mathbf{x}, t) = \sum_{j=1}^d \sum_{k=1}^p g_k(t) \Psi_{j,k} x_j = \langle \Psi, \Phi(\mathbf{x}, t) \rangle_F \quad (3)$$

où  $\Phi(\mathbf{x}, t) = \mathbf{x} \mathbf{g}(t)^\top = \mathbf{x} \otimes \mathbf{g}(t)$ . Cette formulation permet de reformuler l’apprentissage d’un classifieur périodique en un problème linéaire classique de taille  $d \times p$  paramètres.

Le choix des fonctions de base est particulièrement important, puisqu’il va encoder l’information sur l’évolution des données au cours du temps. Dans ces travaux, deux types de bases conventionnelles sont proposées : les bases de Fourier et de splines quadratiques périodiques. Un exemple de ces deux types de fonctions de base est présenté dans la Figure 1. Ces bases permettent d’encoder à la fois la périodicité et la régularité temporelle du classifieur.

1. Il est également possible d’ajouter un terme de biais  $b$  au classifieur en ajoutant une caractéristique constante aux observations  $\mathbf{x}$ .

**Estimation du classifieur** Nous pouvons exprimer le problème d’estimation du classifieur de la manière suivante :

$$\min_{\Psi} \sum_i H(y_i, \langle \Psi, \Phi(\mathbf{x}_i, t_i) \rangle_F) + \lambda \Omega(\Psi) + \lambda_t \Omega_t(\Psi) \quad (4)$$

où  $H(y, f(\mathbf{x}, t))$  est un terme d’attache aux données qui mesure l’erreur de prédiction entre l’étiquette  $y$  et la prédiction  $f(\mathbf{x}, t)$ . Dans le cas binaire, nous avons utilisé le terme d’attache aux données de la régression logistique [3, Chap. 4.4] définis par  $H(y, f(\mathbf{x}, t)) = \log(1 + \exp(yf(\mathbf{x}, t)))$ . Les fonctions  $\Omega(\cdot)$  et  $\Omega_t(\cdot)$  sont des fonctions de régularisation convexes et différentiables qui mesurent la complexité globale et temporelle du modèle. On choisit en pratique la norme de Frobenius  $\Omega(\cdot) = \|\cdot\|_F^2$  pour la régularisation globale, qui conduit à la régularisation *ridge* classique : les poids sont pénalisés par leur valeur au carré. La régularisation temporelle  $\Omega_t(\cdot)$  est discutée plus en détail dans la suite.

Le problème d’optimisation présenté dans l’équation (4) est convexe et différentiable pour des fonctions de régularisation différentiables. Il peut cependant potentiellement mener à un grand nombre de paramètres ( $d \times p$ ). Nous proposons donc de le résoudre en utilisant une descente de gradient de type quasi-Newton, comme le L-BFGS décrit dans [4]. A noter que pour de grandes bases de données étiquetées, il est également possible d’utiliser des méthodes d’optimisation stochastique.

**Régularisation temporelle**  $\Omega_t(\cdot)$  pénalise la complexité temporelle du classifieur. Dans cet article, elle est mesurée à l’aide de la norme fonctionnelle de la dérivée temporelle  $m$ ème de chaque composante de  $\mathbf{w}$  :  $\|w_j^m(\cdot)\|^2 = \int_0^1 w_j^m(t)^2 dt$  [8]. La complexité du classifieur est exprimée comme la somme des complexités pour chaque composante :

$$\Omega_t(\Psi) = \sum_{j=1}^d \|w_j^m(\cdot)\|^2. \quad (5)$$

Ce type de régularisation est utilisé en interpolation, pour les splines lissées [8]. Il n’a, à notre connaissance, encore jamais été utilisé dans un cadre de discrimination supervisée. Le terme peut être reformulé en fonction des composantes de  $\Psi$  de la manière suivante :

$$\Omega_t(\Psi) = \sum_{j=1}^d \int_0^1 \left( \sum_{k=1}^p \Psi_{j,k} g_k(t)^{(m)} \right)^2 dt = \langle \mathbf{K}, \Psi^\top \Psi \rangle_F$$

où  $\mathbf{K}$  est une matrice de terme  $K_{k,l} = \int_0^1 g_k^m(t) g_l^m(t) dt$  qui est le produit scalaire fonctionnel entre les dérivées des fonctions de base  $g_k(t)$ .  $\Omega_t(\Psi)$  est une régularisation quadratique par rapport aux paramètres  $\Psi_{j,k}$ .

Pour les fonctions de base de Fourier, qui sont orthogonales deux à deux,  $\mathbf{K}$  est une matrice diagonale. Les termes diagonaux sont de la forme  $(2\pi\nu)^{2m}$  où  $\nu$  est la fréquence de la fonction de base. Pour les fonctions de base de type splines quadratiques la matrice  $\mathbf{K}$  n’est plus diagonale mais reste une matrice circulante.

**Extension multi-classe** Le modèle introduit dans les équations (1) à (4) est un modèle de classification binaire. Il peut être étendu à un problème de discrimination à  $C$  classes. On cherche alors à estimer  $C$  classifieurs  $\Psi_c$  avec  $c \in 1, \dots, C$  permettant de prédire la probabilité d'appartenance à chaque classe [3, Chap. 4.4]. La fonction d'attache aux données, qui correspond à maximiser la probabilité a posteriori sur les données d'apprentissage, peut s'exprimer par

$$H(y, f(x, t)) = \log\left(\sum_c e^{(\Psi_c - \Psi_y, \Phi(x, t))}\right),$$

où  $\Psi_y$  est le modèle  $\Psi_c$  pour la classe  $c = y \in 1, \dots, C$  [7]. Le cas multi-classe est particulièrement intéressant en terme d'interprétation car il devient possible de visualiser l'évolution temporelle de l'impact de chaque variable ( $w_k$  pour la classe  $k$ ) selon la classe à reconnaître. On notera que, en pratique, grâce à la contrainte de somme à 1 des probabilités, il n'est nécessaire d'estimer que  $C - 1$  modèles, le dernier modèle dépendant linéairement des autres.

### 3 Expérimentations numériques

**Méthodes comparées** Ce travail est à notre connaissance le premier qui s'attaque exactement au problème de classification instantanée de données périodiques. Nous nous comparerons donc à une régression logistique constante le long du temps que nous appellerons LR dans la suite du document. Dans ce cas, le paramètre temporel n'est pas pris en compte : un seul classifieur est construit sur tous les échantillons d'entraînement.

Une approche *naïve* pour pouvoir intégrer le temps dans le processus d'apprentissage est de traiter chaque instant temporel indépendamment, *i.e.*, de construire un classifieur par instant temporel avec les données disponibles pour cet instant. Pour prédire un nouvel instant non appris, une interpolation des classifieurs est faite. Cette approche est notée  $LR_{interp}$ . L'interpolation entre les classifieurs est faite indépendamment pour chaque composante de  $w$  avec une interpolation linéaire.

L'approche proposée est notée  $LR_{basis}$  Fourier et  $LR_{basis}$  Quad. Spline pour l'utilisation de bases de Fourier et des splines quadratiques, respectivement. Le nombre de fonctions est fixé à  $p = 6 + 1$ , c'est-à-dire 6 bases temporelles et une composante constante comme illustré sur la Figure 1. La régularisation se fait sur la dérivée temporelle et l'on a donc fixé  $m = 1$ . La mesure de performance utilisée est le taux de bonne classification sur des données de test à des instants qui n'étaient pas accessibles dans les données d'apprentissage, pour permettre d'évaluer la capacité des différentes méthodes à prédire à des instants temporels non observés.

**Données simulées** Le premier jeu de données consiste en des données simulées, ce qui permet de contrôler chacun des aspects et ainsi d'évaluer différentes situations. Un problème de classification dans  $\mathbb{R}^2$  à 3 classes linéairement séparables est utilisée. La non-stationnarité est ajoutée en effectuant une

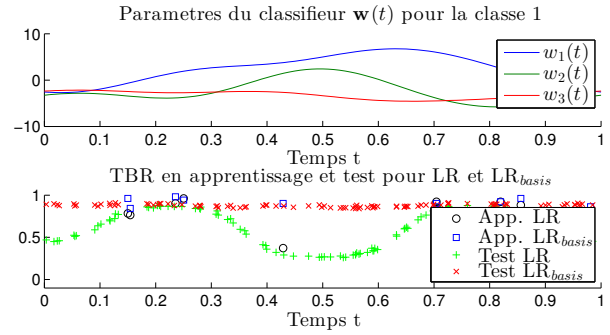


FIGURE 3 – Évolution des paramètres du classifieur au cours du temps (haut). Performance de prédiction (Taux de Bonne Reconnaissance) le long du temps pour LR et  $LR_{basis}$  sur les données d'apprentissage et de test (bas).

rotation des données dont l'angle de rotation dépend du temps de la manière suivante

$$\theta = 2\pi \sin(\pi t)^2 A_{max},$$

où  $A_{max}$  est un paramètre qui contrôle la non-stationnarité, *i.e.* l'amplitude de la rotation. Les données sont générées à  $n_{tapp}$  instants temporels tirés à partir d'une loi uniforme et pour chaque instant temporel  $n_{app}$  exemples d'apprentissage sont générés, pour un nombre total d'exemples d'apprentissages  $n = n_{tapp} \times n_{app}$ . Un exemple de répartition des instants temporels est illustré à la Figure 3, où les cercles/carrés représentent les taux de reconnaissances aux instants d'apprentissage.

L'évolution temporelle du classifieur  $LR_{basis}$  avec des bases de Fourier est donnée dans la Figure 3 du haut. Le classifieur ainsi appris évolue bien régulièrement au cours du temps. La Figure 3 bas illustre la capacité de prédiction du classifieur pour des instants non observés durant l'apprentissage (représentés par les croix vertes/rouges). Pour le classifieur LR, de bonnes performances sont obtenues pour des instants proches des instants d'apprentissage les plus représentés dans la base d'apprentissage. Tandis que le classifieur  $LR_{basis}$  garde des performances stables sur l'ensemble de la période et arrive donc à modéliser l'évolution temporelle.

La validation des paramètres est effectuée pour toutes les méthodes en utilisant une 6-fold validation croisée séparant les exemples par instant temporel. Ce type de validation est le plus proche possible de la mesure de performance car la prédiction se fera sur des exemples de test à des instants temporels qui ne sont pas disponibles dans la base d'apprentissage. L'influence des paramètres de simulation sur les résultats sont tracés sur la Figure 2. Les méthodes prenant en compte le temps,  $LR_{interp}$  et  $LR_{basis}$  sont beaucoup plus robustes aux non-stationnarités que LR. De même  $LR_{basis}$  est systématiquement meilleur que  $LR_{interp}$  pour un nombre d'exemples limités. Cette différence de performances est amplifiée sur des données réelles de dimension  $d$  plus grande comme nous le verrons dans la suite.

**Données de télédétection** Nous avons évalué les différentes méthodes sur des données télédétection fournies par le CES-

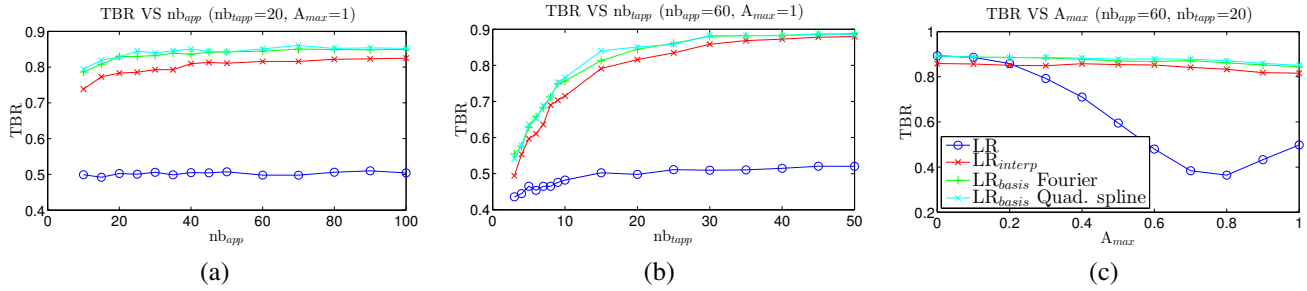


FIGURE 2 – Taux de bonne reconnaissance (TBR) en fonction de (a) nombre d'exemples d'apprentissage par instant temporels  $n_{app}$ , (b) nombre d'instant temporels en apprentissage  $n_{tapp}$  et (c) non-stationnarité  $A_{max}$  pour les différentes méthodes.

Méthode	2009	2010	2011	Moy.
LR	0.439	0.439	0.545	0.475
LR <sub>interp</sub>	0.734	0.635	0.623	0.664
LR <sub>basis</sub> Fourier	<b>0.787</b>	<b>0.768</b>	0.789	<b>0.781</b>
LR <sub>basis</sub> Quad. spline	0.766	0.766	<b>0.799</b>	0.777

TABLE 1 – Résultats de prédiction en taux de bonne reconnaissance pour les données de télédétection. Les résultats en gras correspondent à la meilleur performance pour chaque colonne.

BIO. Ce sont des mesures satellites (SPOT-4 et SPOT-5) comprenant plusieurs dates par an échantillonnées irrégulièrement. La classification est faite à partir des 4 composantes spectrales (B, R, NIR et SWIR). La campagne d'acquisition s'est déroulée sur 3 ans (2009, 2010, 2011) correspondant à 41 échantillons temporels. Nous avons comparé les méthodes sur un problème de classification végétale à 3 classes (Blé, Colza et Maïs) pour lesquelles nous avons accès à une vérité terrain. Les données totalisent  $n = 90461$  exemples répartis sur les 41 échantillons temporels.

Pour évaluer la capacité des méthodes à prédire sur une nouvelle année, les données ont été découpées en  $2ans + 1an$ , deux années pour l'entraînement et une année pour la prédiction. Les paramètres de régularisation ont été sélectionnés par validation croisée en découpant par année sur les données d'apprentissage (2-fold validation croisée). Les résultats sont reportés dans le tableau 1. On voit que l'utilisation de bases temporelles permet d'obtenir des gains importants en terme de taux de bonne reconnaissance, de l'ordre de 33% par rapport à un classifieur constant LR et de 12% par rapport à une interpolation LR<sub>interp</sub>. Ces résultats s'expliquent en partie par l'évolution temporelle des classes de type cultures le long de l'année.

## 4 Conclusion

Nous avons proposé dans ce document un nouveau classifieur non-stationnaire périodique qui a la capacité de prédire instantanément la classe de données variant périodiquement. Notre méthode s'applique parfaitement sur des données multi-temporelles de télédétection, tout en conservant un cadre d'application très large. Le modèle est simple et interprétable et nous proposons une régularisation permettant de limiter la complexité temporelle lorsque peu d'échantillons sont disponibles.

Les simulations numériques ont illustré la meilleure généralisation de notre approche sur des instants temporels non-observés.

Dans des travaux futurs nous comptons étudier comment les modèles périodiques peuvent être utilisés pour classifier une année entière lorsque l'on a la même classe le long de l'année. De même il nous paraît pertinent d'intégrer des connaissances plus poussées dans le processus d'apprentissage et de prédiction comme des modèles de rotation de culture [5].

## Références

- [1] P. M. Atkinson, C. Jeganathan, J. Dash, and C. Atzberger. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment*, 123 :400 – 417, 2012.
- [2] H. Carrao, P. Gonalves, and M. Caetano. A nonlinear harmonic model for fitting satellite image time series : Analysis and prediction of land cover dynamics. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4) :1919–1930, April 2010.
- [3] T. Hastie, J. Friedman, and R. Tibshirani. *The elements of statistical learning*. Springer Series in Statistics, 2001.
- [4] J. Nocedal. Updating quasi-newton matrices with limited storage. *Mathematics of computation*, 35(151) :773–782, 1980.
- [5] J. Osman, J. Inglada, J.-F. Dejoux, O. Hagolle, and G. Dedieu. Crop mapping by supervised classification of high resolution optical image time series using prior knowledge about crop rotation and topography. In *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, pages 2832–2835. IEEE, 2013.
- [6] F. Petitjean, J. Inglada, and P. Gancarski. Satellite image time series analysis under time warping. *IEEE Transactions on Geoscience and Remote Sensing*, 50 :3081, 2012.
- [7] D. Tuia, R. Flamary, and N. Courty. Multiclass feature learning for hyperspectral image classification : sparse and hierarchical solutions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015.
- [8] G. Wahba. *Spline models for observational data*, volume 59. Siam, 1990.