

Apprentissage par dictionnaire pour le démixage de signaux temporels en IRMf

Argheesh BHANOT, Céline MEILLIER, Fabrice HEITZ, Laura HARSAN

ICube, UMR 7357, Université de Strasbourg, CNRS – 300 bd Sébastien Brant – 67412 Illkirch Cedex

bhanot@unistra.fr, meillier@unistra.fr, fabrice.heitz@unistra.fr, harsan@unistra.fr

Résumé – Le démixage des données IRMf est un problème difficile lorsque l’on s’intéresse au sujet unique (analyse individuelle), notamment en raison de la faible résolution spatiale des données. Nous proposons une méthode de démixage formulée comme un algorithme d’apprentissage par dictionnaire où la contrainte classique de parcimonie est remplacée par une contrainte de localisation spatiale haute résolution. Cette information est fournie par le recalage d’un atlas finement résolu et de la carte de segmentation associée sur les données IRMf. Les premiers résultats chez la souris, sur données quasi réelles, montrent les performances de la méthode, avec une très bonne estimation des signaux.

Abstract – In this paper, we propose a new dictionary learning formulation for fMRI unmixing adapted to the single-subject by integrating high-resolution spatial information. High-resolution information is provided by the registration of an anatomical atlas on the data set. We show on a quasi-real (mouse) dataset the benefit of using a high-resolution spatial segmentation map in the decomposition of low-resolution fMRI.

1 Introduction et hypothèses de travail

L’étude de la connectivité fonctionnelle cérébrale est l’une des problématiques majeures en neurosciences, ouvrant la voie à une meilleure compréhension de l’organisation et du fonctionnement cérébral, que ce soit chez les individus sains ou pathologiques, chez l’homme ou le modèle animal. La connectivité fonctionnelle est généralement étudiée via des séquences d’IRM fonctionnelle (IRMf). En IRMf de repos, il s’agit principalement de détecter les régions du cerveau présentant une activité spontanée et d’étudier les liens entre les différentes régions anatomiques qui forment alors des réseaux fonctionnels. Les méthodes basées sur l’analyse en composantes indépendantes spatiale (ICA pour *independent component analysis*) ont été largement utilisées pour le démixage des données IRMf de repos [1]. L’ICA peut être vue comme le démixage de signaux temporels de la forme :

$$\mathbf{Y} \simeq \mathbf{U}\mathbf{A}, \quad (1)$$

où la matrice $\mathbf{Y} \in \mathbb{R}^{N \times P}$ représente les données IRMf avec P le nombre de voxels dans le cerveau, N la longueur des signaux temporels enregistrés. Les colonnes de la matrice $\mathbf{U} \in \mathbb{R}^{N \times R}$ contiennent la signature temporelle, appelée aussi décours temporel, des R réseaux fonctionnels et $\mathbf{A} \in \mathbb{R}^{R \times P}$ est la matrice d’abondances codant la contribution des R réseaux dans chaque voxel. Les méthodes basées sur l’ICA posent le problème du choix du nombre de composantes indépendantes. Certaines ne sont que du bruit, elles doivent donc être filtrées selon des critères semi-automatiques. Récemment les approches par apprentissage de dictionnaire [2] ont démontré leur efficacité pour le démixage en reformulant le problème (1) en imposant, non pas une indépendance spatiale des colonnes de \mathbf{A} , mais une parcimonie spatiale :

$$\min_{\mathbf{A}, \mathbf{U}} \frac{1}{2} \|\mathbf{Y} - \mathbf{U}\mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_1. \quad (2)$$

La convergence vers un minimum global n’est pas garantie pour les algorithmes d’apprentissage par dictionnaire car la fonction à minimiser n’est pas convexe. En pratique, une bonne initialisation de l’une des matrices permet une estimation conjointe de \mathbf{U} et \mathbf{A} biologiquement satisfaisante.

Ces deux approches ont été principalement développées et utilisées pour réaliser des analyses de groupe, *i.e.* pour estimer les réseaux fonctionnels communs à tous les individus d’un groupe. L’estimation est robuste dans le cas des études de groupe du fait de la redondance introduite par l’homogénéité des groupes et par le nombre d’individus. Dans le cas d’études individuelles (sujet unique), l’estimation de ces réseaux par les méthodes classiques est beaucoup plus délicate et engendre, par exemple dans le cas de l’ICA, un grand nombre de composantes spatiales non interprétables d’un point de vue biologique. Classiquement dans les études de groupe, le recalage de tous les individus sur un template commun est nécessaire (de fait cette étape est presque systématiquement appliquée en analyse IRMf). Ce recalage entraîne une interpolation des signaux temporels qui, combinée à la faible résolution spatiale des données IRMf, complexifie le démixage des signaux. Ce point devient critique au niveau de l’individu pour lequel on ne peut bénéficier de la redondance spatiale des réseaux fonctionnels.

Chez la souris, nous disposons d’un atlas et d’une segmentation très fine du cerveau. En faisant l’hypothèse qu’à cette échelle de segmentation, les réseaux fonctionnels sont constitués d’ensembles de petites régions anatomiques, il est possible d’effectuer le démixage au niveau des régions, et de constituer les réseaux *a posteriori*. Afin d’éviter l’interpolation des signaux temporels lors du recalage sur l’atlas, c’est l’atlas qui est projeté sur les données IRMf. Nous proposons, dans ce contexte, une méthode de démixage permettant, contrairement aux approches classiques, le démixage chez le sujet unique et reposant sur un algorithme d’apprentissage de dictionnaire.

2 Modélisation et méthode

Étant donné le modèle (1), le problème de minimisation

$$\min_{\mathbf{A}, \mathbf{U}} \frac{1}{2} \|\mathbf{Y} - \mathbf{U}\mathbf{A}\|_F^2 \quad (3)$$

ne possède pas une unique solution. Afin de restreindre le nombre de solutions, des contraintes de positivité de \mathbf{A} et de somme à un des colonnes $\mathbf{a}_i, i \in \{1, \dots, P\}$ de \mathbf{A} sont introduites dans le modèle. Dans la suite on note S l'ensemble $\{\mathbf{A} \in \mathbb{R}^{R \times P} | \mathbf{1}^T \mathbf{a}_i = 1, i = 1, \dots, P\}$ où $\mathbf{1}$ est le vecteur unité de taille $R \times 1$, et R représente dans notre approche le nombre de régions anatomiques de la carte de segmentation de l'atlas.

Contrairement aux approches classiques d'apprentissage par dictionnaire, nous n'utilisons pas de contrainte de parcimonie spatiale sur la matrice \mathbf{A} sous la forme d'un terme $\lambda \|\mathbf{A}\|_1$, mais nous introduisons une connaissance *a priori* sur la localisation des différentes régions dans le cerveau grâce à l'atlas recalé sur les données. Lors de la projection de l'atlas, la carte de segmentation est également projetée, ce qui nous permet d'obtenir R images binaires correspondant aux masques des R régions de la segmentation dans l'espace de nos données IRMf basse résolution. Cette information est exploitée afin de contraindre la position des coefficients non nuls dans la matrice des abondances \mathbf{A} grâce à une matrice intermédiaire $\tilde{\mathbf{A}}$ contenant cette information spatiale, où l'élément $(\tilde{\mathbf{A}})_{r,i} = 1$ si la région anatomique r de l'atlas est projetée sur le voxel basse résolution i , et 0 sinon. Le problème (3) est reformulé comme :

$$\min_{\mathbf{A}, \mathbf{U}} \frac{1}{2} \|\mathbf{Y} - \mathbf{U}\mathbf{A}\|_F^2 + \mu_\sigma \|\mathbf{U}\|_2^2 + \mathcal{I}_{\mathbb{R}^+}(\mathbf{A}) + \mathcal{I}_{\tilde{\mathbf{A}}}(\mathbf{A}) + \mathcal{I}_S(\mathbf{A}), \quad (4)$$

où l'on trouve, dans l'ordre, le terme d'attache aux données, une régularisation Tikhonov sur \mathbf{U} , la contrainte de positivité sur \mathbf{A} , avec $\mathcal{I}_{\mathbb{R}^+}(\mathbf{A}) = +\infty$ si au moins un élément de \mathbf{A} est négatif. Le terme $\mathcal{I}_{\tilde{\mathbf{A}}}(\mathbf{A})$ est la fonction indicatrice de l'ensemble des matrices ayant la même structure que $\tilde{\mathbf{A}}$, i.e. $\mathcal{I}_{\tilde{\mathbf{A}}}(\mathbf{A}) = +\infty$ si au moins un coefficient de \mathbf{A} est non nul alors qu'il l'est dans $\tilde{\mathbf{A}}$. De même $\mathcal{I}_S(\mathbf{A}) = \infty$ si la somme des éléments d'au moins une colonne de \mathbf{A} n'est pas égale à 1.

2.1 Estimation des décours temporels

Si on considère \mathbf{A} fixée, le problème (4) devient :

$$\min_{\mathbf{U}} \frac{1}{2} \|\mathbf{Y} - \mathbf{U}\mathbf{A}\|_F^2 + \mu_\sigma \|\mathbf{U}\|_F^2. \quad (5)$$

Du fait de la basse résolution spatiale des données IRMf, lorsque deux régions voisines très fines de la carte de segmentation de l'atlas sont projetées, elles peuvent se retrouver localisées sur les mêmes voxels basse résolution, menant ainsi à une colinéarité des deux colonnes correspondantes dans la matrice \mathbf{A} . La constante μ_σ est fixée à une valeur très faible ($\mu_\sigma = 10^{-4}$) afin de ne pas introduire trop de lissage dans les décours temporels de la matrice \mathbf{U} mais en prévenant un mauvais conditionnement de \mathbf{A} . La solution ridge du problème (5) est :

$$\hat{\mathbf{U}} = \mathbf{Y}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \mu_\sigma \mathbf{I}_R)^{-1}, \quad (6)$$

où \mathbf{I}_R est la matrice d'identité de taille $R \times R$.

2.2 Estimation des abondances

En fixant \mathbf{U} , le problème (4) devient $\min_{\mathbf{A}} f(\mathbf{A})$ où

$$f(\mathbf{A}) = \frac{1}{2} \|\mathbf{Y} - \mathbf{U}\mathbf{A}\|_F^2 + \mathcal{I}_{\mathbb{R}^+}(\mathbf{A}) + \mathcal{I}_{\tilde{\mathbf{A}}}(\mathbf{A}) + \mathcal{I}_S(\mathbf{A}). \quad (7)$$

Notons que cette fonction est séparable selon les voxels $i \in \{1, \dots, P\}$, on peut donc résoudre en parallèle les P problèmes $\min_{\mathbf{a}_i} f(\mathbf{a}_i)$ avec :

$$f(\mathbf{a}_i) = \frac{1}{2} \|\mathbf{y}_i - \mathbf{U}\mathbf{a}_i\|_2^2 + \mathcal{I}_{\mathbb{R}^+}(\mathbf{a}_i) + \mathcal{I}_{\tilde{\mathbf{a}}_i}(\mathbf{a}_i) + \mathcal{I}_S(\mathbf{a}_i), \quad (8)$$

où \mathbf{a}_i est un vecteur colonne de la matrice \mathbf{A} . Le vecteur $\tilde{\mathbf{a}}_i$ représente la $i^{\text{ème}}$ colonne de $\tilde{\mathbf{A}}$. Les termes de régularisation dans l'équation (8) peuvent être résumés sous la forme d'une unique fonction indicatrice $g(\mathbf{a}_i) = \mathcal{I}_{\mathbb{R}^+ \cap \tilde{\mathbf{a}}_i \cap S}(\mathbf{a}_i)$. La convexité de l'ensemble $\mathbb{R}^+ \cap \tilde{\mathbf{A}} \cap S$ se vérifie aisément. La minimisation de la fonction (8) peut se faire via l'algorithme FISTA [3]. L'implémentation de FISTA pour résoudre la minimisation de (8) est donnée dans l'algorithme 1.

```

1 pour  $k \leftarrow 1$  à  $K$  faire
2    $\mathbf{a}_i^{(k)} = \text{prox}_g(\boldsymbol{\omega}^{(k)} - \lambda \nabla f(\boldsymbol{\omega}^{(k)}))$ 
3    $t^{(k+1)} = \frac{1 + \sqrt{1 + 4(t^{(k)})^2}}{2}$ 
4    $\boldsymbol{\omega}^{(k+1)} = \mathbf{a}_i^{(k)} + \left(\frac{t^{(k)} - 1}{t^{(k+1)}}\right) (\mathbf{a}_i^{(k)} - \mathbf{a}_i^{(k-1)})$ 
5 fin

```

Algorithme 1 : Algorithme FISTA pour estimer les colonnes \mathbf{a}_i de la matrice \mathbf{A} .

Dans l'algorithme 1, $\nabla f(\mathbf{a}_i)$ est le gradient de $f(\mathbf{a}_i)$, donné par $\mathbf{U}^T(\mathbf{U}\mathbf{a}_i - \mathbf{y}_i)$. Le pas $\lambda \in \mathbb{R}^{+*}$ est choisi égal à l'inverse de la constante de Lipschitz de $\nabla f(\mathbf{a}_i)$, $L = \|\mathbf{U}^T \mathbf{U}\|_F$. Le vecteur $\boldsymbol{\omega}^{(k+1)}$ est défini comme une combinaison linéaire des deux dernières estimations de \mathbf{a}_i grâce à la suite $(t^{(k)})_{k \in \mathbb{N}}$ de coefficients supérieurs à 1. Ici, l'opérateur proximal de la fonction g s'écrit :

$$\text{prox}_g(\mathbf{y}) = \underset{\mathbf{x} \in \mathbb{R}^+ \cap \tilde{\mathbf{a}}_i \cap S}{\text{argmin}} \|\mathbf{x} - \mathbf{y}\|_2 = \mathcal{P}_{\mathbb{R}^+ \cap \tilde{\mathbf{a}}_i \cap S}(\mathbf{y}),$$

où \mathcal{P} est la projection orthogonale sur l'ensemble $\mathbb{R}^+ \cap \tilde{\mathbf{a}}_i \cap S$. Le projeté orthogonal de \mathbf{y} sur $\mathbb{R}^+ \cap \tilde{\mathbf{a}}_i \cap S$ est obtenu par une stratégie de type POCS (projection onto convex sets) en alternant la projection sur le simplexe $S \cap \mathbb{R}^+$ et la projection sur l'ensemble des vecteurs ayant la même structure que $\tilde{\mathbf{a}}_i$. Cette procédure nécessite très peu d'itérations avant convergence.

3 Application à des données quasi-réelles

Afin de tester l'algorithme de démixage proposé, nous utilisons des données acquises dans une étude préclinique sur un modèle de souris Alzheimer. Les données sont constituées d'une image IRMf de repos et d'une image anatomique 3D acquise lors du même examen. L'image anatomique a une dimension $256 \times 256 \times 34$ et une résolution spatiale de $0.08299 \times 0.07812 \times 0.4$ mm. L'IRMf est de taille $147 \times 87 \times 27 \times 500$ avec une

résolution spatiale de $0.1445 \times 0.2299 \times 0.5 \text{ mm}$ et $2s$ pour la résolution temporelle. L'atlas *Allen Brain Atlas* [4] fournit une image 3D du cerveau (template) et une carte de segmentation qui permet d'identifier $R = 595$ structures anatomiques d'intérêt dans le cerveau de la souris. Ces deux jeux de données ont une résolution spatiale de $25 \times 25 \times 25 \mu\text{m}$.

3.1 Recalage d'atlas et prétraitements

Avant le démixage, les étapes classiques de prétraitement des données IRMf (*slice timing*, correction du mouvement au cours de l'examen si besoin) et de recalage sont effectuées. On choisit de recalcr l'atlas avec les données IRM anatomiques qui sont bien mieux résolues spatialement que les données IRMf. On tire parti du fait que les données anatomiques et IRMf sont quasiment parfaitement recalées entre elles (la souris n'a pas bougé) et on effectue d'abord un recalage non rigide des données anatomiques sur les données IRMf pour corriger les distorsions géométriques parfois observées entre les deux modalités. Les données IRMf voient leur résolution spatiale artificiellement augmentée au préalable en subdivisant chaque voxel en $3 \times 6 \times 2$ voxels (la valeur du voxel original est dupliquée dans les $J = 36$ sous-voxels). Ceci permet d'atteindre une résolution spatiale comparable à celles de l'atlas et des données anatomiques, sauf dans la troisième dimension qui correspond à l'épaisseur des tranches lors de l'acquisition IRM (qui ne peut être réduite qu'au détriment de la résolution temporelle). L'atlas est ensuite recalé sur les données anatomiques par recalage non rigide. Les deux étapes de recalages successifs fournissent des champs de déformation que l'on peut appliquer en cascade aux régions de la carte de segmentation de l'atlas haute résolution afin d'obtenir les masques binaires des R régions projetées sur les données IRMf artificiellement augmentées.

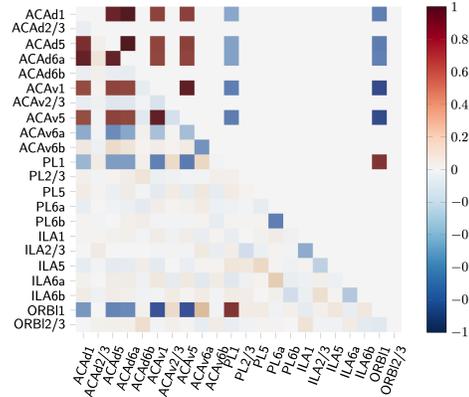
Pour initialiser la matrice $\mathbf{A}^{(0)}$, on utilise l'IRMf augmentée et l'atlas projeté : pour chaque voxel basse résolution i de \mathbf{Y} et toutes les régions $r \in \{1, \dots, R\}$, le coefficient $(A^{(0)})_{r,i}$ contient la proportion de voxels haute résolution occupée par la région r dans le voxel i basse résolution. Si une région r n'est pas présente sur le voxel i alors $(A^{(0)})_{r,i} = 0$.

3.2 Introduction de signaux synthétiques

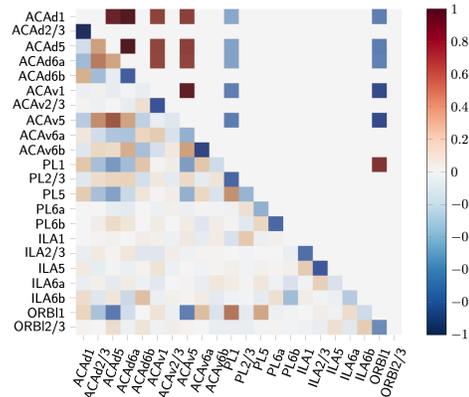
Afin de tester les performances de la méthode proposée en conditions réalistes, on utilise le jeu de données IRMf réelles dans lequel des signaux temporels synthétiques connus sont introduits. Le recalage de l'atlas sur les données IRMf augmentées nous permet de sélectionner avec précision la localisation de sept petites régions du cortex préfrontal : ACAd1, ACAd5, ACAd6a, ACAv1, ACAv5, PL1 et ORB11. Les signaux temporels sont introduits dans l'IRMf à résolution augmentée, puis la résolution est dégradée par moyennage des $3 \times 6 \times 2$ sous-voxels afin de retrouver la basse résolution originale. Cette opération permet de mélanger les signaux synthétiques avec les décours temporels réels des régions voisines. L'estimation des signaux synthétiques ainsi que leurs abondances devient alors aussi complexe que l'estimation des autres régions. Un premier signal synthétique a été obtenu en faisant la moyenne des si-

	ACAd1	ACAd5	ACAd6a	ACAv1	ACAv5	PL1	ORB11
ACAd1	1.00	0.84	0.92	0.67	0.69	-0.45	-0.59
ACAd5	0.84	1.00	0.92	0.67	0.69	-0.47	-0.60
ACAd6a	0.92	0.92	1.00	0.66	0.67	-0.46	-0.58
ACAv1	0.67	0.67	0.66	1.00	0.88	-0.59	-0.82
ACAv5	0.69	0.69	0.67	0.88	1.00	-0.60	-0.82
PL1	-0.45	-0.47	-0.46	-0.59	-0.60	1.00	0.74
ORB11	-0.59	-0.60	-0.58	-0.82	-0.82	0.74	1.00

TABLE 1 – Valeurs des corrélations entre les sept signaux temporels synthétiques introduits dans le jeu de données réelles.



(a) Avec l'information haute résolution



(b) Sans l'information haute résolution

FIGURE 1 – Corrélations des signaux du cortex préfrontal. La matrice triangulaire inférieure représente les corrélations calculées à partir des signaux estimés, la matrice triangulaire supérieure montre les corrélations réelles entre les décours temporels synthétiques.

gnaux des régions ACAd1 et ACAd5 qui étaient déjà fortement corrélés dans les données réelles. Ce signal est ensuite modifié pour créer des signaux fortement corrélés ou anticorrélés pour les sept régions sélectionnées. Leurs corrélations synthétiques sont présentées dans le tableau 1.

3.3 Résultats

La méthode de démixage proposée est testée sur les données quasi-réelles décrites dans le paragraphe 3.2. La figure 1a

représente la matrice de corrélation des sept régions sélectionnées dans le cortex préfrontal ainsi que les régions voisines non modifiées après 500 itérations d’optimisation alternée du problème (4) pour estimer \mathbf{A} et \mathbf{U} . Empiriquement l’algorithme propose une solution acceptable et stable : le gain sur la minimisation du problème (4) est inférieur à 10^{-3} d’une itération à l’autre au-delà de 500 itérations. Pour l’estimation de \mathbf{A} , l’algorithme FISTA est stoppé après un nombre maximum d’itérations ($K = 100$) ou lorsque $\|\mathbf{a}_i^{(k-1)} - \mathbf{a}_i^{(k)}\| < 10^{-8}$.

Afin de souligner l’importance de l’information de segmentation haute résolution dans la méthode proposée, on effectue le démixage en recalant directement l’atlas sur les données IRMf basse résolution. La matrice $\tilde{\mathbf{A}}$ est alors obtenue en projetant directement la carte de segmentation via le champ de déformation estimé lors du recalage, et on constate que cela entraîne un élargissement et un mélange des différentes régions de la carte de segmentation : les méthodes de recalage traditionnelles ne sont pas conçues pour projeter des données dans le sens : haute résolution vers basse résolution, lorsque les résolutions sont très différentes. La figure 1b montre la matrice de corrélation ainsi obtenue à partir des décours temporels et des abondances estimées sans cette information haute résolution.

La figure 2 présente les signaux synthétiques en bleu sur lesquels sont superposés les décours temporels estimés avec la méthode proposée en utilisant l’information haute résolution de l’atlas pour définir la matrice $\tilde{\mathbf{A}}$ (en pointillés rouges) et sans cette information (en vert). Ceci illustre bien l’importance de cette étape de recalage en résolution artificiellement augmentée et la nécessité d’une connaissance *a priori* de la localisation spatiale des régions dont on cherche ensuite à estimer la proportion fonctionnelle dans le décours temporel de chaque voxel. L’erreur quadratique moyenne est indiquée dans les deux cas sur chaque graphe de la figure 2 : EQM(HR) pour l’estimation avec l’information de haute résolution et EQM lorsque cette information n’est pas prise en compte. On constate l’influence de la régularisation basée sur l’information haute résolution puisque EQM(HR) est systématiquement plus faible que EQM.

4 Conclusion

Une prise en compte des informations spatiales *a priori* apportées par un atlas haute résolution nous a permis de proposer une solution au difficile problème de démixage des données IRMf individuelles (cas du sujet unique). Nous avons en particulier proposé une nouvelle forme de régularisation dans l’algorithme d’apprentissage par dictionnaire utilisé pour réaliser le démixage. Pour intégrer correctement les informations spatiales haute résolution, nous avons montré l’importance de bien spécifier l’étape de recalage entre l’atlas haute résolution et les données IRMf. L’information fournie par l’atlas est utilisée pour produire une bonne initialisation de la matrice d’abondance, garantissant ainsi la convergence de l’algorithme vers une solution acceptable d’un point de vue anatomique et fonctionnel. La prochaine étape de ces travaux consiste maintenant à modéliser et estimer les relations entre les régions anatomiques, afin de construire les réseaux fonctionnels.

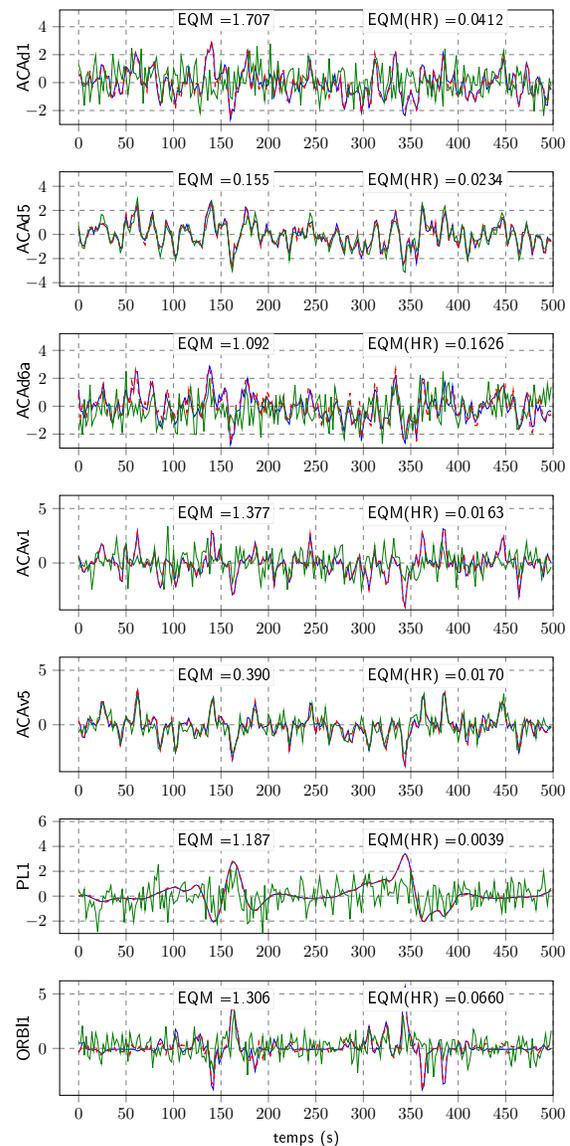


FIGURE 2 – Représentation des 500 premières secondes des décours temporels synthétiques (bleu) ainsi que leurs estimations avec l’information spatiale de haute résolution apportée par l’atlas (pointillés rouges) et sans cette information (vert).

Références

- [1] V. D. Calhoun and T. Adali, “Unmixing fmri with independent component analysis,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 2, pp. 79–90, 2006.
- [2] A. Mensch, G. Varoquaux, and B. Thirion, “Compressed Online Dictionary Learning for Fast Resting-State fMRI Decomposition,” in *International Symposium on Biomedical Imaging*. Apr. 2016, pp. 1282–1285, IEEE.
- [3] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [4] S. W. Oh et al., “A mesoscale connectome of the mouse brain,” *Nature*, vol. 508, pp. 207–214, 2014.