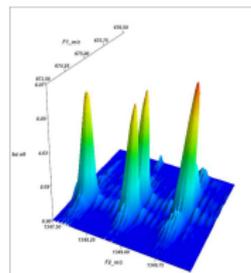


# TABASCO: TRAITEMENT DU BRUIT NON GAUSSIEN EN SPECTROSCOPIE

E. Chouzenoux, M.-A. Delsuc, C. Rolando

*Colloque de restitution - Défi Mastodons*

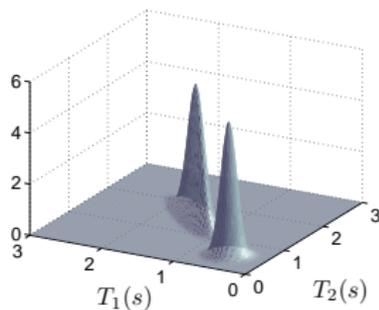
9 février 2017



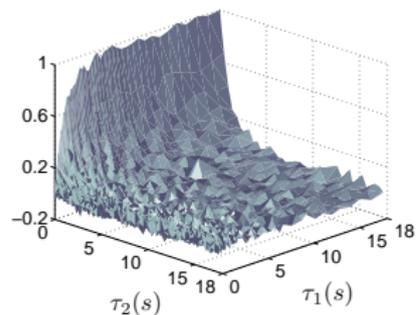
## Interdisciplinarité

- *Laboratoire d'Informatique Gaspard Monge (LIGM)* : INS2I  
E. Chouzenoux (MCF), G. Chierchia (MCF), H. Talbot (PRCE), J.-C. Pesquet (PRCE)
- *Laboratoire Jacques-Louis Lions (LJLL)* : INSMI  
P. Combettes (PRCE)
- *Miniaturisation pour la Synthèse, l'Analyse et la Protéomique, Lille (MSAP)* : CNRS, INC  
C. Rolando (DR)
- *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)* : INSERM et INS2I  
M.-A. Delsuc (DR) et B. Kieffer (PRCE)

# Problèmes inverses en spectroscopie



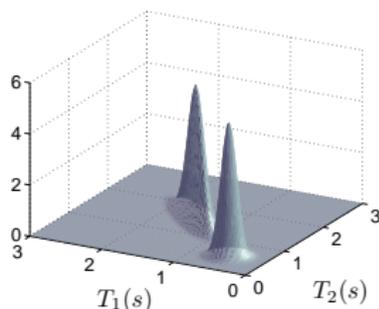
Spectre original



Mesures bruitées

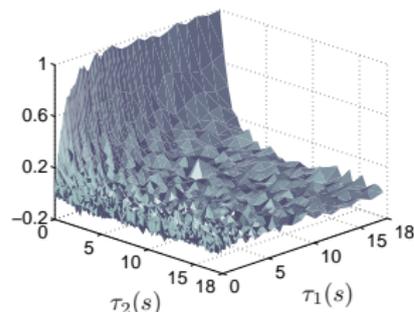


# Problèmes inverses en spectroscopie



Spectre original

$$\bar{x}$$



Mesures bruitées

$$\mathbf{y} = \mathcal{D}(H\bar{x})$$

- ▶  $H$  : matrice modélisant les mesures.
- ▶  $\mathcal{D}$  : modèle de bruit.

Comment restaurer le signal d'intérêt à partir des observations  $\mathbf{y}$  et du modèle  $H$ , dans un contexte de grande dimension ?

## Stratégie de restauration

Comment définir une estimation  $\hat{x}$  de  $\bar{x}$  ?

**Approche naïve** : (en supposant  $H$  inversible)

$$\hat{x} = H^{-1}y$$

↪ Simple mais **amplification du bruit** si  $H$  est mal conditionnée

## Stratégie de restauration

Comment définir une estimation  $\hat{x}$  de  $\bar{x}$  ?

**Approche régularisée** : Minimiser un critère composite  $f + r$ , avec :

$f$  : terme de fidélité aux données, lié au modèle de bruit.

Exemple : moindres carrés si bruit gaussien.

$r$  : terme de pénalisation, liée aux informations a priori.

Exemple :  $\ell_1$  si parcimonie.

### Difficultés :

- ▶ Besoin de **modèles de bruit** fiables.
- ▶ Besoin de **régularisations** efficaces.
- ▶ Besoin de **méthodes d'optimisation** rapides.

⇒ **Projet TABASCO !**

## PALMA : Restauration des signaux en DOSY RMN

\* **Diffusion Ordered SpectroscopY** : modalité d'acquisition en Résonance Magnétique Nucléaire.

**Modèle d'observation :**

$$I(q) = \int_{D_{\min}}^{D_{\max}} X(D) \exp(-D\Delta q^2) dD$$

avec :

- ▶  $D$  : Coefficient de diffusion
- ▶  $\Delta$  : Durée de diffusion
- ▶  $q = \gamma\delta g$  : Mesure de la dispersion de phase
- ▶  $\gamma$  : Ratio gyromagnetic
- ▶  $\delta$  : Durée du champs de gradients pulsé (PFG)
- ▶  $g$  : Intensité du PFG

## PALMA : Restauration des signaux en DOSY RMN

\* **Diffusion Ordered Spectroscopy** : modalité d'acquisition en Résonance Magnétique Nucléaire.

**Modèle d'observation :**

$$I(q) = \int_{D_{\min}}^{D_{\max}} X(D) \exp(-D\Delta q^2) dD$$

**Forme discrétisée :**  $\mathbf{y} = \mathbf{H}\bar{\mathbf{x}} + \mathbf{w}$ , avec :

- ▶  $\mathbf{H} \in \mathbb{R}^{M \times N}$ , avec  $H_{m,n} = \exp(-D_n \Delta q_m^2)$
- ▶  $\bar{\mathbf{x}} \in \mathbb{R}^N$  : spectre à restaurer, à valeurs positives
- ▶  $\mathbf{w} \in \mathbb{R}^M$  : bruit additif supposé gaussien i.i.d., d'écart type  $\sigma$

# PALMA : Restauration des signaux en DOSY RMN

## Approche proposée :

$$\underset{x \in \mathbb{R}^N}{\text{minimiser}} \quad \Psi(x) \quad \text{tel que} \quad \frac{1}{2} \| \mathbf{H}x - \mathbf{y} \|_2^2 \leq \sigma \sqrt{M}$$

avec  $\Psi$  fonction de régularisation hybride :

$$(\forall x \in \mathbb{R}^N, \lambda \in [0, 1]) \quad \Psi(x) = \underbrace{\lambda \sum_{n=1}^N \psi_1(x_n)}_{\text{Entropie}} + \underbrace{(1 - \lambda) \sum_{n=1}^N \psi_2(x_n)}_{\text{Parcimonie}}$$

Entropie $\psi_1$	
$x \log(x) + \iota_{[0, +\infty)}(x)$	<i>Shannon</i>
$-\log(x) + \iota_{[0, +\infty)}(x)$	<i>Burg</i>

Parcimonie $\psi_2$	
$ x ^0$	$\ell_0$
$ x $	$\ell_1$
$\log(1 +  x )$	<i>log-sum</i>
$\log(1 + x^2)$	<i>Cauchy</i>

# PALMA : Restauration des signaux en DOSY RMN

Résolution par l'algorithme proximal PPXA+ [Pustelnik *et al.*, 2011] :

### Initialization

$$V^{(0,1)} \in \mathbb{R}^N, V^{(0,2)} \in \mathbb{R}^M$$

$$X^{(0)} = (I_N + \mathbf{H}^\top \mathbf{H})^{-1} (V^{(0,1)} + \mathbf{H}^\top V^{(0,2)})$$

$$\mathbf{B} = (I_N + \mathbf{H}^\top \mathbf{H})^{-1}$$

$$\gamma \in (0, 2),$$

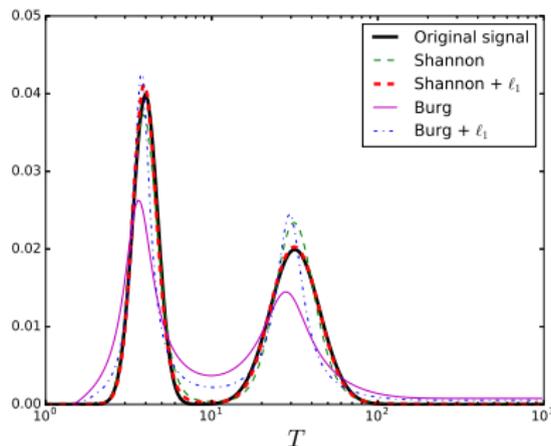
### Minimization

For  $k = 0, 1, \dots$

$$\left[ \begin{array}{l} Z^{(k,1)} = \text{prox}_\Psi(V^{(k,1)}) \\ Z^{(k,2)} = \text{proj}_{\| \cdot - Y \| \leq \eta}(V^{(k,2)}) \\ U^{(k)} = \mathbf{B}(Z^{(k,1)} + \mathbf{H}^\top Z^{(k,2)}) \\ X^{(k+1)} = X^{(k)} + \gamma(U^{(k)} - X^{(k)}) \\ V^{(k+1,1)} = V^{(k,1)} + \gamma(2U^{(k)} - X^{(k)} - Z^{(k,1)}) \\ V^{(k+1,2)} = V^{(k,2)} + \gamma(\mathbf{H}(2U^{(k)} - X^{(k)}) - Z^{(k,2)}) \end{array} \right.$$

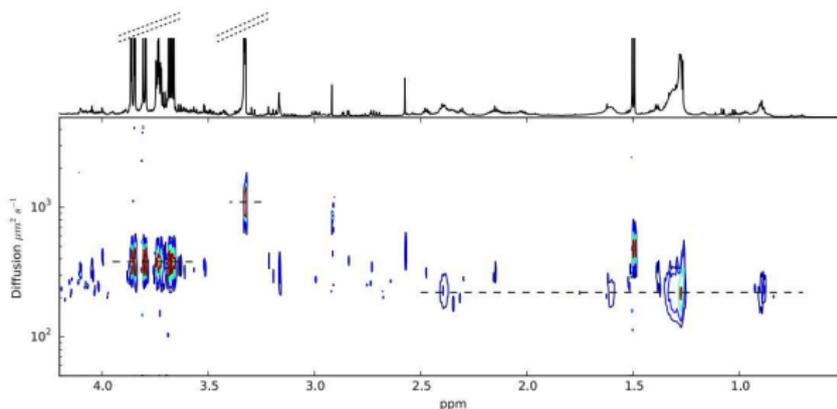
⇒ Proximal Algorithm for L1 combined with MAxent prior (PALMA).

# PALMA : Restauration des signaux en DOSY RMN



⇒ La pénalisation hybride (Shannon +  $\ell_1$ ) permet une nette amélioration de la qualité des signaux reconstruits. ([1 article de conférence internationale](#), LIGM + IGMBC)

## PALMA : Restauration des signaux en DOSY RMN



⇒ Les bonnes performances de PALMA sont confirmées sur des données réelles. (1 [article de journal international](#), LIGM + IGBMC)

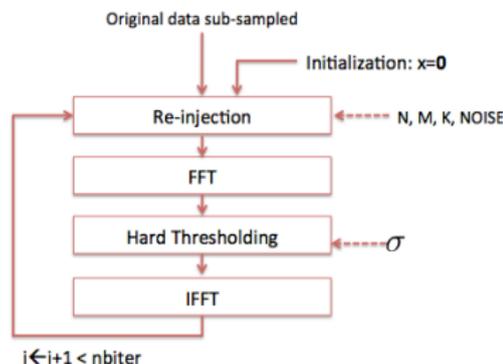
# PG-SANE : Restauration des signaux en MS-FT-ICR

**Modèle d'observation** : Décimation  $\times$  Modulation  $\times$  FFT.

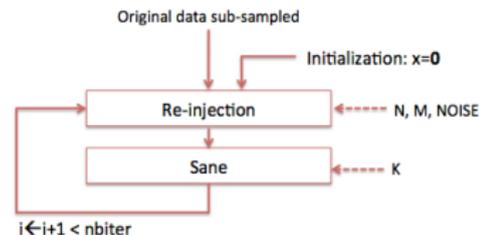
**A priori** : Signal parcomoneux à valeurs complexes.

⇒ Nouvel algorithme efficace qui combine deux approches :

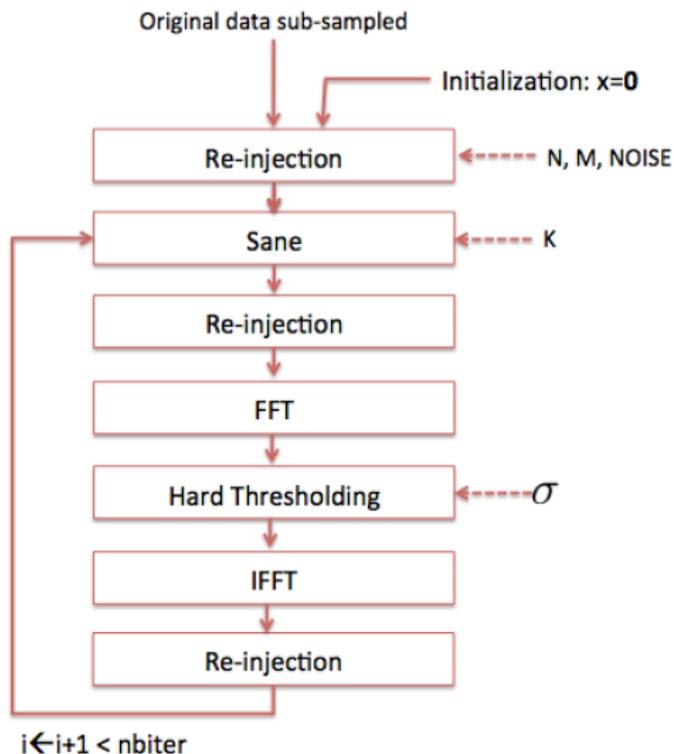
- ▶ Papoulis-Gershberg (analyse spectrale avec contrainte  $\ell_1$ ) ;
- ▶ SANE : Approche de type Cadzow accélérée.



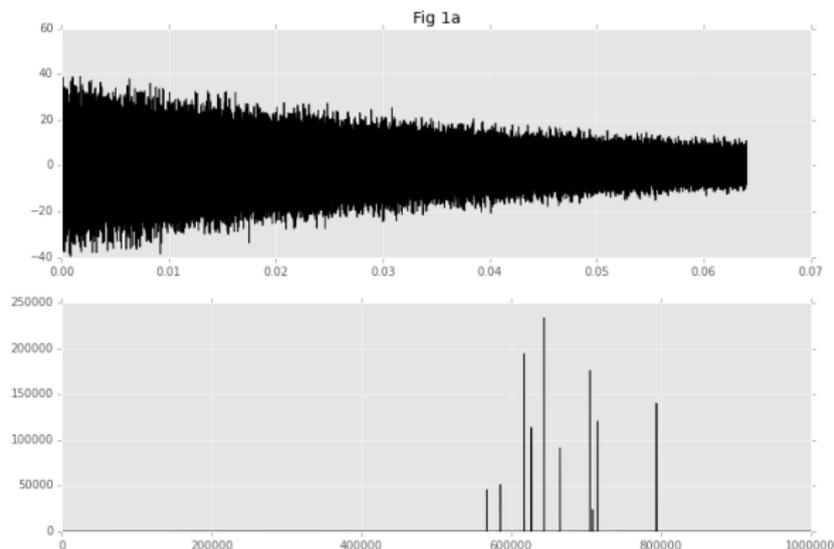
+



# PG-SANE : Restauration des signaux en MS-FT-ICR

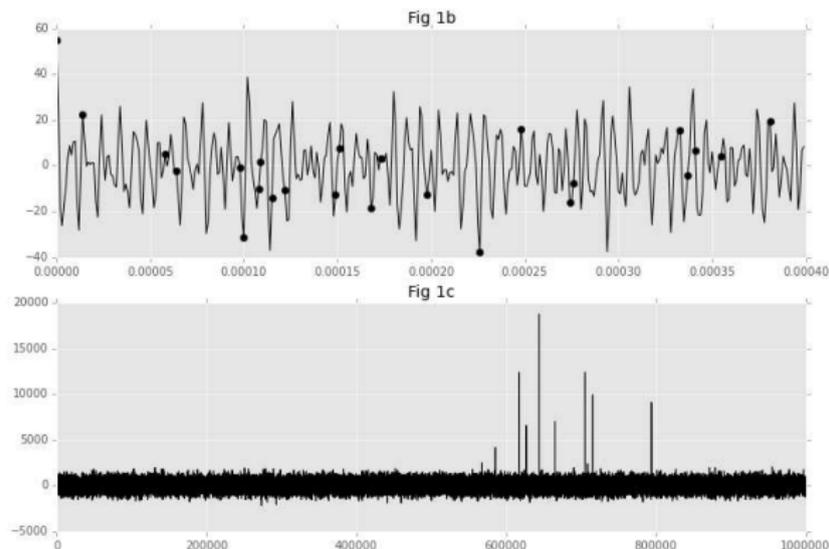


# PG-SANE : Restauration des signaux en MS-FT-ICR



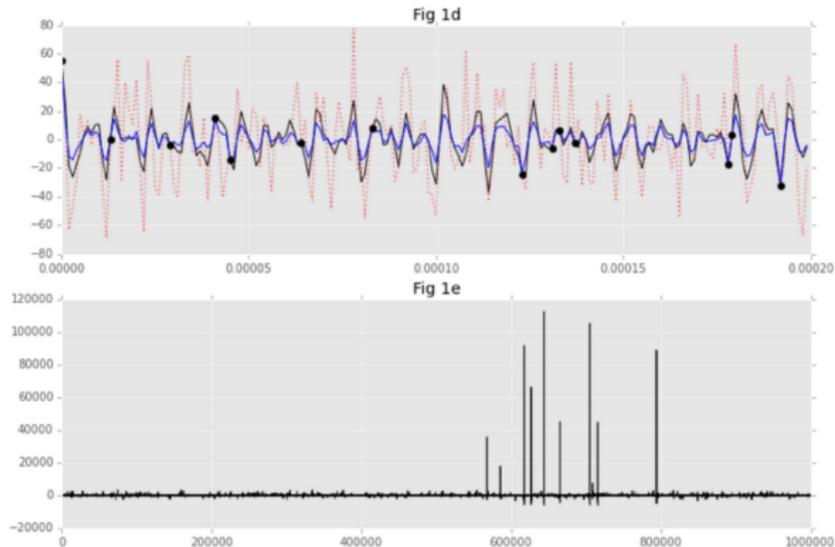
*Données synthétiques simulées*

# PG-SANE : Restauration des signaux en MS-FT-ICR



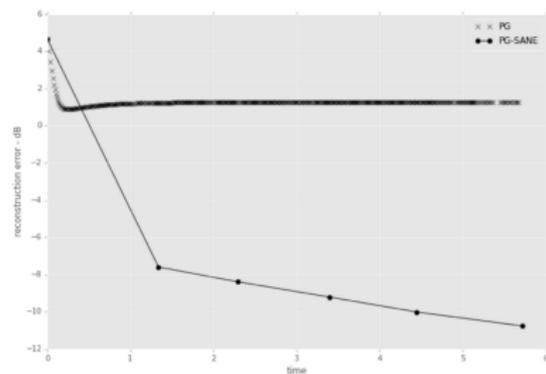
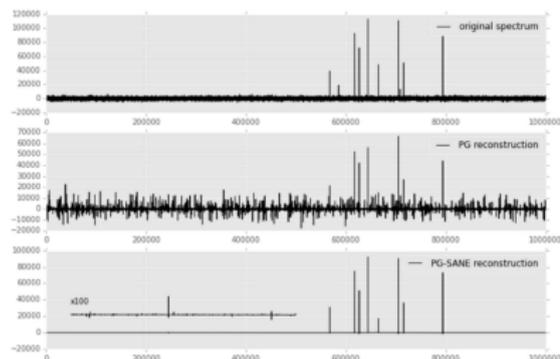
*Données restaurées par simple inversion de Fourier*

# PG-SANE : Restauration des signaux en MS-FT-ICR



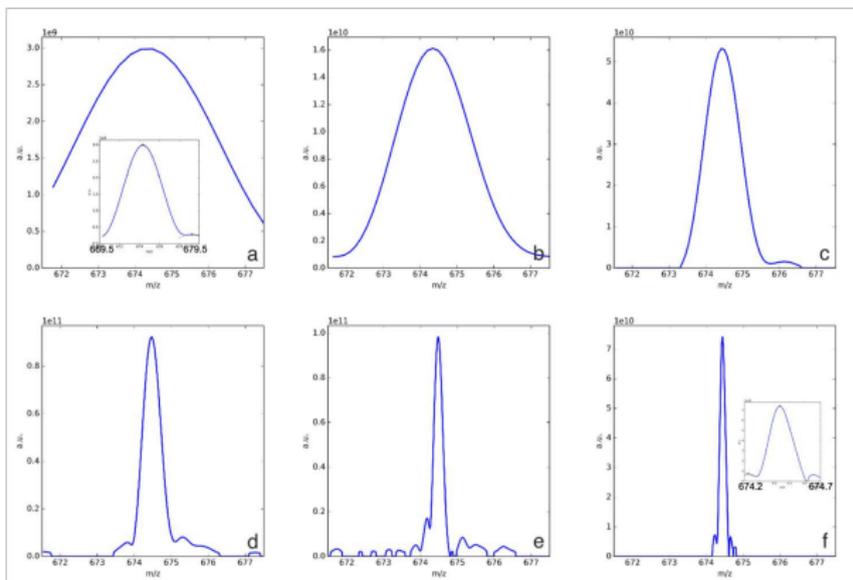
*Résultat obtenu avec PG-SANE*

# PG-SANE : Restauration des signaux en MS-FT-ICR



*Comparaisons entre PG et PG-SANE*

# PG-SANE : Restauration des signaux en MS-FT-ICR



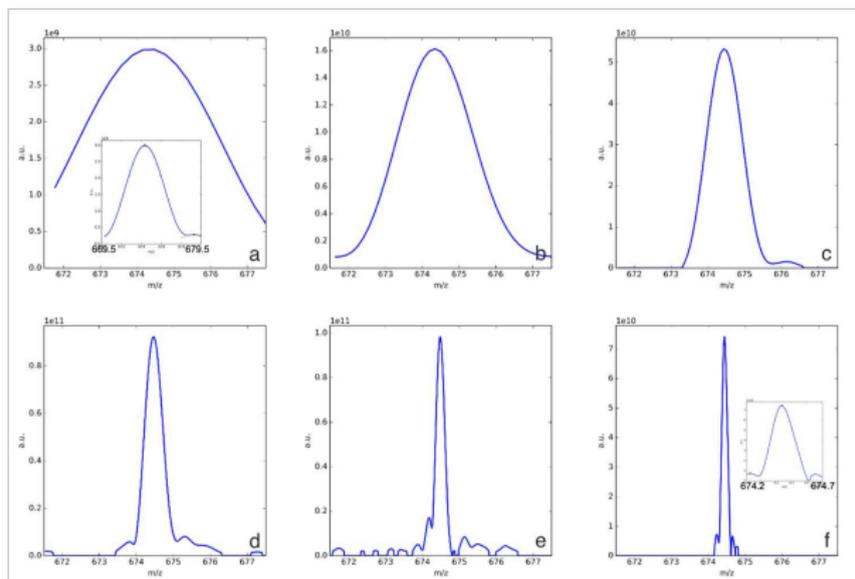
Substance P doubly charged monoisotopic peak 2D FT-ICR spectrum.

Zoom along the F1 (precursor) dimension

(from a to f, from 2k to 64k, non NUS to NUS 1/32).

*Résultats obtenus sur des données réelles*

# PG-SANE : Restauration des signaux en MS-FT-ICR



*Résultats obtenus sur des données réelles*

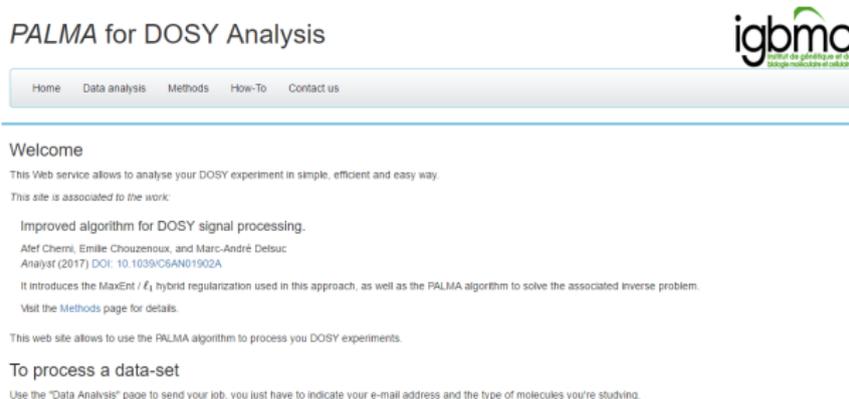
↪ **Un article de journal en préparation (IGBMC + MSAP) :**  
“Analysis of Non Uniform Sampled DataSets in 2D-FTICR-MS”. Lionel Chiron, Fabrice Bray, Afef Cherni, Jean-Philippe Stark, Christian Rolando, Marc-André Delsuc.

## Autres travaux réalisés

- LIGM** Nouvelle méthode de restauration de type Bayes variationnelle. Application à la restauration d'images dégradées par un bruit mixte Poisson-Gaussien. (1 [article de revue soumis](#))
- LIGM** Nouvelle approche de débruitage d'images multi-temporelles dégradées par un bruit gamma multiplicatif dit "speckle". (1 [article de revue soumis](#))
- LJLL** Introduction de la notion de fonction perspective dans le contexte des problèmes de statistiques et inverses en présence de bruit non gaussien. (1 [article de revue soumis](#))
- LJLL** Etude du comportement asymptotique, et des propriétés de consistance, d'estimateurs statistiques régularisés. (2 [articles de revue soumis](#))

# Diffusion et valorisation des résultats

↪ Mise en place d'une plateforme web pour l'utilisation en ligne de la méthode PALMA.



**PALMA for DOSY Analysis**

Home Data analysis Methods How-To Contact us

---

### Welcome

This Web service allows to analyse your DOSY experiment in simple, efficient and easy way.  
This site is associated to the work:

Improved algorithm for DOSY signal processing.  
Atef Cherni, Emile Chouzenoux, and Marc-André Detsuc  
Analyst (2017) DOI: 10.1039/C6AN01902A

It introduces the MaxEnt /  $L_1$  hybrid regularization used in this approach, as well as the PALMA algorithm to solve the associated inverse problem.  
Visit the [Methods](#) page for details.

This web site allows to use the PALMA algorithm to process you DOSY experiments.

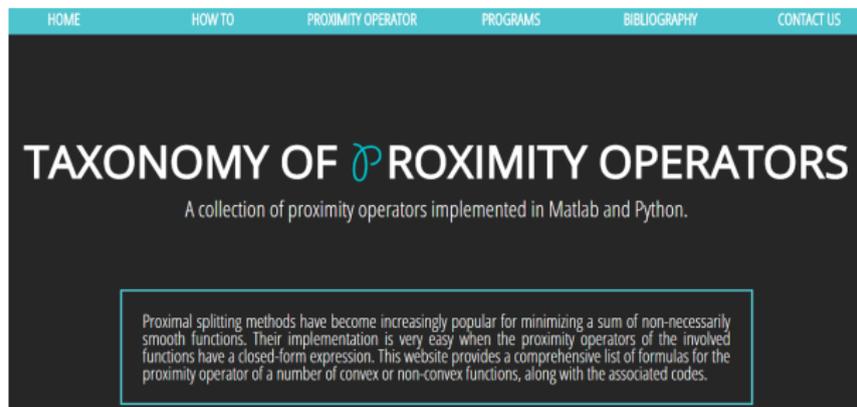
### To process a data-set

Use the "Data Analysis" page to send your job, you just have to indicate your e-mail address and the type of molecules you're studying.

↪ Dépôt de valorisation logicielle au sein de la SATT Alsace pour le code source Python de la méthode PALMA.  
(collaboration LIGM + IGBMC)

## Diffusion et valorisation des résultats

↪ Mise en place d'une plateforme web pour le partage des codes sources Matlab/Python d'opérateurs proximaux.



(projet en cours, collaboration LIGM + LJLL)

## Publications

- ▶ 1 articles de revue accepté co-signé (LIGM + IGBMC)
  - ▶ A. Cherni, E. Chouzenoux and M.-A. Delsuc. *PALMA, an Improved Algorithm for DOSY Signal Processing*. To appear in *Analyst*, 2016.

## Publications

- ▶ 1 articles de revue accepté co-signé (LIGM + IGBMC)
- ▶ 1 article de conférence internationale co-signé (LIGM + IGBMC)
  - ▶ A. Cherni, E. Chouzenoux and M.-A. Delsuc. *Proximity Operators for a Class of Hybrid Sparsity+Entropy Priors. Application to DOSY NMR Signal Reconstruction*. In Proceedings of ISIVC 2016, Tunis, Tunisie, 21-23 novembre 2016.

## Publications

- ▶ 1 articles de revue accepté co-signé (LIGM + IGBMC)
- ▶ 1 article de conférence internationale co-signé (LIGM + IGBMC)
- ▶ 5 articles de revue soumis (LIGM, LJLL)
  - ▶ P. L. Combettes, S. Salzo and S. Villa. *Consistent Learning by Composite Proximal Thresholding*. Soumis, 2016.
  - ▶ P. L. Combettes and C. L. Müller. *Perspective Functions : Proximal Calculus and Applications in High-Dimensional Statistics*. Soumis, 2016.
  - ▶ P. L. Combettes, S. Salzo, and S. Villa. *Regularized Learning Schemes in Feature Banach Spaces*. Analysis and Applications, to appear, 2016.
  - ▶ G. Chierchia, M. El Gheche, G. Scarpa, and L. Verdoliva. *Multitemporal SAR Image Despeckling based on Block-Matching and Collaborative Filtering*. Soumis, 2016.
  - ▶ Y. Marnissi, Y. Zheng, E. Chouzenoux and J.-C. Pesquet. *A Variational Bayesian Approach for Restoring Data Corrupted with Non-Gaussian Noise*. Soumis, 2016.

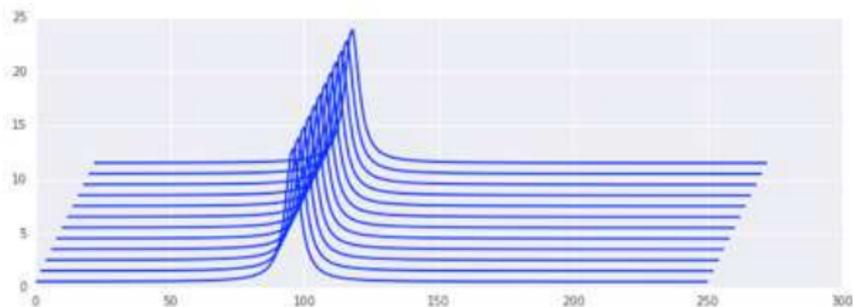


# Budget

- ▶ Subvention globale : 25 keuros
- ▶ Répartition entre équipes équitable
- ▶ Répartition par postes de dépenses
  - ▶ équipements amortissables : 60%
  - ▶ missions : 30%
  - ▶ petits équipements : 5%
  - ▶ autres (prestations, fonctionnement) : 5%

# Perspectives

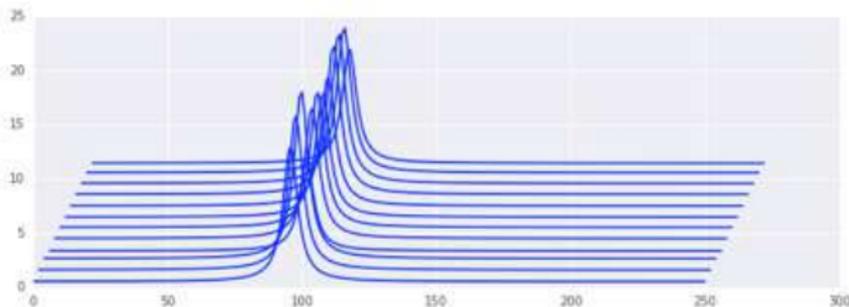
1. Caractérisation/modélisation des différents bruits en DOSY RMN, et intégration dans la méthode PALMA.
2. Régularisation explicite et intégration d'une stratégie de rejection de bruit dans les méthodes PG-SANE et UrQrd, utilisées en MS-FT-ICR.



*Extraits du rapport de stage de N. Bettaieb (financement TABASCO)*

## Perspectives

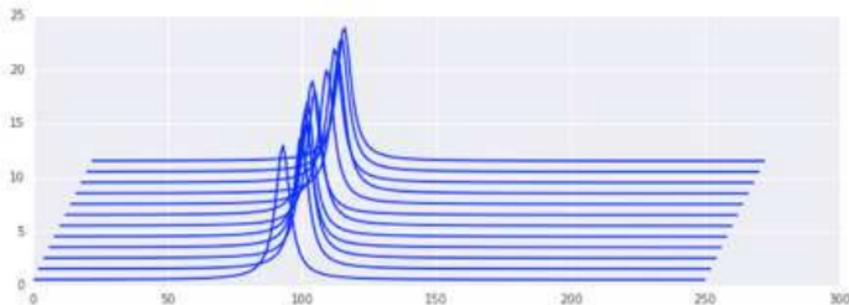
1. Caractérisation/modélisation des différents bruits en DOSY RMN, et intégration dans la méthode PALMA.
2. Régularisation explicite et intégration d'une stratégie de rejection de bruit dans les méthodes PG-SANE et UrQrd, utilisées en MS-FT-ICR.



*Extraits du rapport de stage de N. Bettaieb (financement TABASCO)*

## Perspectives

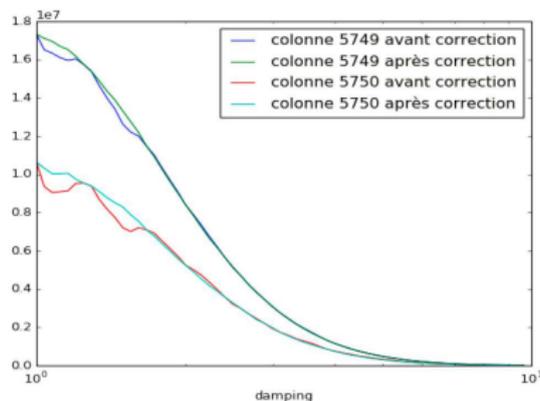
1. Caractérisation/modélisation des différents bruits en DOSY RMN, et intégration dans la méthode PALMA.
2. Régularisation explicite et intégration d'une stratégie de rejection de bruit dans les méthodes PG-SANE et UrQrd, utilisées en MS-FT-ICR.



*Extraits du rapport de stage de N. Bettaieb (financement TABASCO)*

## Perspectives

1. Caractérisation/modélisation des différents bruits en DOSY RMN, et intégration dans la méthode PALMA.
2. Régularisation explicite et intégration d'une stratégie de rejection de bruit dans les méthodes PG-SANE et UrQrd, utilisées en MS-FT-ICR.



# Perspectives

1. Caractérisation/modélisation des différents bruits en DOSY RMN, et intégration dans la méthode PALMA.
2. Régularisation explicite et intégration d'une stratégie de rejection de bruit dans les méthodes PG-SANE et UrQrd, utilisées en MS-FT-ICR.

Merci de votre attention !