

# Pôle Transactions Electroniques Sécurisées

Le Pôle de compétitivité numérique de Normandie



Le Pôle TES est soutenu financièrement par :



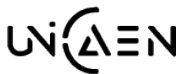
# Réunion INTERLABS 27 Novembre 2017

FAICEL CHAMROUKHI

Professeur, LMNO UMR CNRS

Université de Caen-Normandie

chamroukhi@unicaen.fr



## Cadre de recherche : Sciences des données

- **Contexte** : Données complexes  $\leftrightarrow$  *hétérogènes, temporelles dynamiques, fonctionnelles, incomplètes, de grande dimension, et disponibles en masse*
- **Objectif** : Transformation de telles données en connaissances :  
 $\leftrightarrow$  Reconstruction/révélation de structures cachées, i.e, (hiérarchie de) groupes ;  
prédiction sur données futures ; sélection de variables/réduction de dimension, etc

## Axes de recherche

- 1 Apprentissage statistique sur données hétérogènes et complexes ;
- 2 Apprentissage statistique non-supervisé sur masses de données.

## Savoir faire

- Apprentissage supervisé  $\Rightarrow$  Modèles de régression et de discrimination
- **Apprentissage non-supervisé**  $\Rightarrow$  Modèles de *clustering* et de segmentation
- $\hookrightarrow$  **Modèles** statistiques à **structure latente** :  $f(x|\theta) = \int_z f(x, z|\theta) dz$
- $\hookrightarrow$  **Algorithmes** efficaces sur des données réelles brutes
- **Goût** particulier pour les applications

## Apprentissage sur données temporelles hétérogènes

- Segmentation de données temporelles dynamiques
- Modèle génératif de régression à processus logistique caché
- ↳ Bonnes propriétés et algorithme efficace
- ↳ Permet de détecter une dynamique lente et/ou rapide

[Thèse] Thèse de F. Chamroukhi, 2010, *UTC*

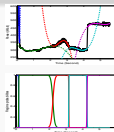
- Segmentation de séries temporelles multidimensionnelles
- ↳ Modèles génératifs de régression multiple à processus latents

[Thèse] Thèse de Dorra TRABELSI, 2010-2013, *Paris 12*

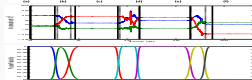
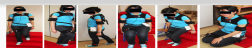
- Classification et visualisation de séquences audiovisuelles
- ↳ Modèles génératifs topographiques en grande dimension

[Thèse] Thèse de Rakia JAZIRI, 2010-2013, *Paris 13*

### Détection d'évènements pour le diagnostic



### Santé : Prédiction de chutes



### Structuration de flux audio-visuel

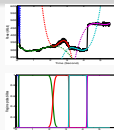


## Apprentissage sur données temporelles hétérogènes

- Segmentation de données temporelles dynamiques
- Modèle génératif de régression à processus logistique caché
- ↳ Bonnes propriétés et algorithme efficace
- ↳ Permet de détecter une dynamique lente et/ou rapide

[Thèse] Thèse de F. Chamroukhi, 2010, *UTC*

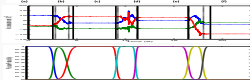
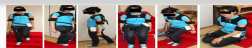
### Détection d'évènements pour le diagnostic



- Segmentation de séries temporelles multidimensionnelles
- ↳ Modèles génératifs de régression multiple à processus latents

[Thèse] Thèse de Dorra TRABELSI, 2010-2013, *Paris 12*

### Santé : Prédiction de chutes



- Classification et visualisation de séquences audiovisuelles
- ↳ Modèles génératifs topographiques en grande dimension

[Thèse] Thèse de Rakia JAZIRI, 2010-2013, *Paris 13*

### Structuration de flux audio-visuel



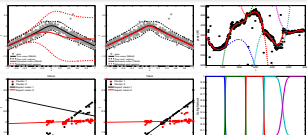
## Modèles de mélanges d'experts robustes

- Données hétérogènes de distributions non-normales
- ↳ Cadre assez récent en classification
- ↳ Mélange d'experts avec *distributions* non-normales

[J] F. Chamroukhi, *Neural Networks*, 79 :20-36, 2016

[J] F. Chamroukhi, *Neurocomputing*, 266 :390-408, 2017

### Modèles robustes de données hétérogènes



## Apprentissage et analyse de données fonctionnelles

- Classification et segmentation de données fonctionnelles hétérogènes
  - Modèles de mélanges hiérarchiques dynamiques à processus latents
- ↳ Cadre probabiliste général et algorithmes d'inférence efficaces

[J] F. Chamroukhi *Journal of Classification*, 33, 2016. doi :10.1007/s00357-.

[J] F. Chamroukhi et al. *Neurocomputing*, 112 :153-163, 2013

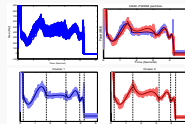
- Régularisation en apprentissage non-supervisé
- ↳ Inférence bayésienne de mélanges à effet mixtes

[J] F. Chamroukhi, *Journal of Statistical Computation and Simulation*, 2015.

- Analyse de données fonctionnelles de grande dimension
  - Problème scientifique récent
- ↳ *Co-clustering* par des modèles à blocs fonctionnels latents

En cours, thèse possible à venir si ANR

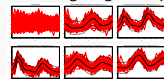
Transport : Clustering pour l'aide à la décision



Images



Clustering en génomique



Transfert : Logiciel MixtComp



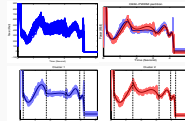
## Apprentissage et analyse de données fonctionnelles

- Classification et segmentation de données fonctionnelles hétérogènes
  - Modèles de mélanges hiérarchiques dynamiques à processus latents
- Cadre probabiliste général et algorithmes d'inférence efficaces

[J] F. Chamroukhi *Journal of Classification*, 33, 2016. doi :10.1007/s00357-.

[J] F. Chamroukhi et al. *Neurocomputing*, 112 :153-163, 2013

Transport : Clustering pour l'aide à la décision



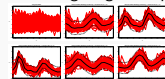
- Régularisation en apprentissage non-supervisé
- Inférence bayésienne de mélanges à effet mixtes

[J] F. Chamroukhi, *Journal of Statistical Computation and Simulation*, 2015.

Images



Clustering en génomique



- Analyse de données fonctionnelles de grande dimension
  - Problème scientifique récent
- *Co-clustering* par des modèles à blocs fonctionnels latents

En cours, thèse possible à venir si ANR

Transfert : Logiciel MixtComp

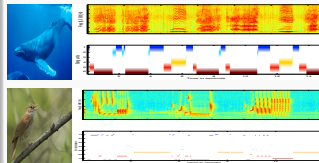


## Apprentissage non-paramétrique

- Représentation et classification non-supervisées de données bio-acoustiques multimodales
- Modèles statistiques et deep learning
- [Thèse] Vincent Roger, 2016-2019, [UTLN/LMNO](#)
- Cadre non-paramétrique : inférence et choix de modèle

[Thèse] Marius BARTCUS, 2012-2015, [UTLN](#)

Biologie - Découverte d'unités de chant



## Contrats de recherche

2018-2020	RIN Asterics : Apprentissage non-supervisé à grande échelle
Partenaires	UMR LMNO (coordinateur) · UMR LMRS
2017-	ANR SMILES ( <i>déposé</i> ) : Stat Modeling and Inference at Large-Scale
Partenaires	UMR LMNO (coordinateur) · UMR LSIS · UMR LMRS · INRIA
2012-2015	SABIOD : Scaled Acoustic BIODiversity, CNRS MASTODONS
Partenaires	UMR LSIS (Coord) · Museum Natural History Paris · UPMC, Orsay
2012-2015	RAPID PHRASE : Augmented Reality and Autonomous Perception
Partenaires	UMR LSIS (Coord) · Inst. Sciences du Mouvement · Prolexia · DGA
2011-2014	ANR COGNILEGO : from Pixels to Semantics
Partenaires	LSIS (Coord) · LNIA · A2IA (entreprise)
2007-2010	SNCF SwitchRdf : Diagnostic de systèmes ferroviaires

## Thèses en cours dans le domaine

2016-2019	Thèse de Huynh Bao Tuyen, UNICAEN, Dir, Bourse : Vietnam
Sujet	<i>High-dimensional mixtures of experts for complex data analysis</i>
2016-2019	Thèse de Vincent ROGER, UTLN, Co-dir, Bourse : Région PACA
Sujet	Bayesian deep networks for signal decomposition and classification
2016-2019	Thèse de Agus RAHARJO, AMU, Co-dir, Bourse : Indonésie
Sujet	Apprentissage multilabel de mélanges d'experts

intérêt au dispositif CIFRE



# Applications possibles en bio, santé

i) Modèles génératifs dynamiques de régression à processus latent [J-1][J-2][J-3][J-4][J-5]

↔ *Analyse de données temporelles pour la robotique assistive*

↔ *Segmentation de séquences génomiques* [Picard et al. *Biometrics* 2007, 2011], de séries temporelles biologiques [Omranian et al. *Scientific Reports* 2015], ..

# Applications possibles en bio, santé

i) Modèles génératifs dynamiques de régression à processus latent [J-1][J-2][J-3][J-4][J-5]

↔ *Analyse de données temporelles pour la robotique assistive*

↔ *Segmentation de séquences génomiques* [Picard et al. *Biometrics* 2007, 2011], de séries temporelles biologiques [Omranian et al. *Scientific Reports* 2015], ..

ii) Modèles génératifs pour l'analyse de données fonctionnelles [J-10][J-12][J-13][J-14]

↔ *Inférence de réseaux de régulations biologiques* [Daudin et al. *Stat&Comp*, (2008)], [Ambroise et al. *EJS*, 2009], [Picard et al. *BMC*, (2009)], ...

↔ *Co-clustering* pour la recherche de gènes différentiellement exprimés, co-exprimés

# Applications possibles en bio, santé

## i) Modèles génératifs dynamiques de régression à processus latent [J-1][J-2][J-3][J-4][J-5]

↪ *Analyse de données temporelles pour la robotique assistive*

↪ *Segmentation de séquences génomiques* [Picard et al. *Biometrics* 2007, 2011], de séries temporelles biologiques [Omranian et al. *Scientific Reports* 2015], ..

## ii) Modèles génératifs pour l'analyse de données fonctionnelles [J-10][J-12][J-13][J-14]

↪ *Inférence de réseaux de régulations biologiques* [Daudin et al. *Stat&Comp*, (2008)], [Ambroise et al. *EJS*, 2009], [Picard et al. *BMC*, (2009)], ...

↪ *Co-clustering* pour la recherche de gènes différentiellement exprimés, co-exprimés

## iii) Apprentissage parcimonieux en grande dimension

↪ *Classification et sélection de variables bio-médicales fonctionnelles de grande dimension* [J-10] [Conf]

Merci de votre attention !