

# Publications du projet ANR SMILES

Université de Caen - UMR LMNO; Université de Rouen - UMR LMRS  
Université de Toulon - UMR LIS; INRIA- Modal

October 26, 2021

## 1 Preprints

1. TrungTin Nguyen, Faicel Chamroukhi, Hien Duy Nguyen, and Florence Forbes. Non-asymptotic model selection in block-diagonal mixture of polynomial experts models. may 2021b. URL <https://arxiv.org/pdf/2104.08>
2. TrungTin Nguyen, Hien Duy Nguyen, Faicel Chamroukhi, and Florence Forbes. A non-asymptotic penalization criterion for model selection in mixture of experts models. 2021c. URL <https://arxiv.org/pdf/2104.0264>
3. TrungTin Nguyen, Faicel Chamroukhi, Hien D Nguyen, and Geoffrey J McLachlan. Approximation of probability density functions via location-scale finite mixtures in lebesgue spaces. 2020a. URL <https://arxiv.org/pdf/2008.09787.pdf>
4. Van Ha Hoang, Gaëlle Chagny, Antoine Channarond, Van Hà Hoang, and Angelina Roche. Adaptive nonparametric estimation of a component density in a two-class mixture model. working paper or preprint, February 2021. URL <https://hal.archives-ouvertes.fr/hal-02909601>
5. TrungTin Nguyen, Hien D Nguyen, Faicel Chamroukhi, and Geoffrey J McLachlan. An  $l_1$ -oracle inequality for the lasso in mixture-of-experts regression models. 2020b. URL <https://arxiv.org/pdf/2009.10622.pdf>
6. T. Huynh and F. Chamroukhi. Estimation and feature selection in mixtures of generalized linear experts models. July 2019. URL <https://chamroukhi.com/papers/prEMME.pdf>
7. A. Roche. Variable selection and estimation in multivariate functional linear regression via the lasso. 2019a. URL <hal.archives-ouvertes.fr/hal-01725351>
8. T-T Nguyen, H D Nguyen, F Chamroukhi, and G J McLachlan. Approximation by finite mixtures of continuous density functions that vanish at infinity. *Submitted*, March 2019. URL <https://arxiv.org/abs/1903.0014>  
Submitted
9. E. Arias-Castro, A. Channarond, B. Pelletier, and N. Verzelen. Minimax estimation of distances in latent graphs. 2018. URL <http://arxiv.org/abs/1804.10611>
10. Hien D. Nguyen, Faicel Chamroukhi, and Florence Forbes. Approximation results regarding the multiple-output mixture of linear experts model. 2018. URL <http://arxiv.org/abs/1704.00946>.  
Submitted

## 2 Articles

1. Hien Duy Nguyen, TrungTin Nguyen, Faicel Chamroukhi, and Geoffrey John McLachlan. Approximations of conditional probability density functions in lebesgue spaces via mixture of experts models. *Journal of Statistical Distributions and Applications*, 8(1):1–15, 2021a. URL <https://arxiv.org/pdf/2012.02385.pdf>

2. Ery Arias-Castro, Antoine Channaron, Bruno Pelletier, and Nicolas Verzelen. On the Estimation of Latent Distances Using Graph Distances. *Electronic Journal of Statistics*, 15(1):722–747, 2021. doi: 10.1214/21-EJS1801. URL <https://hal.archives-ouvertes.fr/hal-02376073>
3. Hien D. Nguyen and Faïcel Chamroukhi. Practical and theoretical aspects of mixture-of-experts modeling: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, pages e1246–n/a, Jan 2018. ISSN 1942-4795. doi: 10.1002/widm.1246. URL <https://chamroukhi.com/papers/Nguyen-Chamroukhi-2018.pdf>
4. Faïcel Chamroukhi and Hien D. Nguyen. Model-based clustering and classification of functional data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Dec 2018. URL <https://chamroukhi.com/papers/MBCC-FDA.pdf>. DOI: 10.1002/widm.1298
5. Faïcel Chamroukhi and Bao T. Huynh. Regularized Maximum Likelihood Estimation and Feature Selection in Mixtures-of-Experts Models. *Journal de la Société Française de Statistique*, 160(1):57–85, March 2019. URL [https://chamroukhi.com/papers/Chamroukhi\\_Huynh\\_jsfds-published.pdf](https://chamroukhi.com/papers/Chamroukhi_Huynh_jsfds-published.pdf)
6. G. Chagny, T. Laloe, and R. Servien. Multivariate adaptive warped kernel estimation. *Electron. J. Stat.*, 13:1759–17894, 2019. ISSN 1935-7524. doi: 10.1214/19-EJS1565. URL <https://doi.org/10.1214/19-EJS1565>
7. Christophe Biernacki, Matthieu Marbac, and Vincent Vandewalle. Gaussian-Based Visualization of Gaussian and Non-Gaussian-Based Clustering. *Journal of Classification*, 38(1):129–157, 2021. ISSN 1432-1343. doi: 10.1007/s00357-020-09369-y. URL <https://doi.org/10.1007/s00357-020-09369-y>
8. Marie A. Roch, Scott Lindeneau, Gurisht Singh Aurora, Kaitlin E. Frasier, John A. Hildebrand, Hervé Glotin, and Simone Baumann-Pickering. Using context to train time-domain echolocation click detectors. *The Journal of the Acoustical Society of America*, 149(5):3301–3310, 2021. doi: 10.1121/10.0004992. URL <https://doi.org/10.1121/10.0004992>
9. Ferrari, Hervé Glotin; Marxer, Asch, Exploring time frequency coding of bioacoustical transient, submitted to Journal of American Society of America, 2021

### 3 Actes de conférences

1. F. Chamroukhi and C. Biernacki. Model-Based Co-Clustering of Multivariate Functional Data. In *ISI 2017 - 61st World Statistics Congress*, Marrakech, Morocco, Jul 2017. URL <https://chamroukhi.com/papers/ISI2017.pdf>
2. Faïcel Chamroukhi, Florian Lecocq, and Hien D Nguyen. Regularized estimation and feature selection in mixtures of gaussian-gated experts models. In *Research School on Statistics and Data Science*, pages 42–56. Springer, July 2019. URL [https://chamroukhi.com/papers/HD-MoGGE\\_Submitted](https://chamroukhi.com/papers/HD-MoGGE_Submitted)
3. Faïcel Chamroukhi, Thien Nhat Pham, Van Ha Hoang, and Geoffrey J McLachlan. Mixtures-of-experts with functional predictors. London, UK, December 2021
4. F. Chamroukhi and V-H Hoang. Functional mixtures-of-experts. CRONOS Multivariate Data Analysis workshop. Model-based and multivariate functional data Session, April 2019. URL <https://chamroukhi.com/Talks/FChamroukhi-CRONOS2019.pdf>
5. F. Chamroukhi. Unsupervised statistical learning from heterogeneous time-series and functional data. Keynote note talk at the Research School on Statistics and Data Science-RSSDS 2019, July 2019. URL <https://chamroukhi.com/Talks/FChamroukhi-RSSDS2019.pdf>
6. R Balestrierio, S Paris (LIS), RG Baraniuk. Rencontres Statistiques Lyonnaises. *Analytical Probability Distributions and EM-Learning for Deep Generative Networks*. In NeurIPS 2020. [https://neurips.cc/virtual/2020/public/poster\\_aaf2979785deb27864047e0ea40ef1b7.html](https://neurips.cc/virtual/2020/public/poster_aaf2979785deb27864047e0ea40ef1b7.html)
7. Maxence Ferrari, Hervé Glotin, Ricard Marxer, and Mark Asch. DOCC10: Open access dataset of marine mammal transient studies and end-to-end CNN classification. In *IJCNN*, Glasgow, United Kingdom, July 2020. URL <https://hal.archives-ouvertes.fr/hal-02866091>

## 4 Communications/exposés par les membres du projet

1. Bérard C., Chagny G., Channarond A., Hoang V-H., Roche A., and Vergne N. Journée de la Fédération Normandie Mathématique
2. Faicel Chamroukhi. “Statistical data science and some unsupervised learning problems”. Février 2019. Groupe de travail stat & sciences des données, LMNO.
3. A. Channarond. Clustering dans un modèle de graphe à positions latentes. Séminaire de Probabilités et Statistiques, Laboratoire de Mathématiques d’Orsay (UMR CNRS 8628), Dec 2018
4. J. Gomez. Rencontres Statistiques Lyonnaises. *Fonctionnelles de clusters d’extrêmes de processus et champs aléatoires*. ICJ (Institut Camille Jordan), Villeurbanne, France. 9 janvier 2019. <http://math.univ-lyon1.fr/~fougeres/rs1.html>
5. G. Chagny, A. Channarond, V.H. Hoang et A. Roche. Estimation non-paramétrique dans un modèle de mélange. Journée de la Fédération Normandie Mathématiques, juin 2019, Le Havre.
6. Marius Bartcus. Séminaire au sein du Groupe de travail stat & sciences des données du LMNO.
7. A. Roche. Variable selection and estimation in multivariate functional linear regression. Séminaire de Statistiques du MAP5 (UMR CNRS 8145), Université Paris Descartes, Juin 2019b
8. Christophe Biernacki. MASSICCC: A SaaS Platform for Clustering and Co-Clustering of Mixed Data. In *APSEM 2019 ((Apprentissage et SEMantique)) : éco-systèmes pour la science ouverte et recherche par les données*, Toulouse, France, October 2019. URL <https://hal.archives-ouvertes.fr/hal-02399180>
9. Christophe Biernacki, Matthieu Marbac-Lourdelle, and Vincent Vandewalle. Gaussian Based Visualization of Gaussian and Non-Gaussian Based Clustering. In *SPSR 2019*, Bucarest, Romania, April 2019. URL <https://hal.archives-ouvertes.fr/hal-02400486>

## 5 Organisation de workshops à dimension internationale

- F. Chamroukhi. Co-organisation de l’Ecole internationale RSSD2019: Research School on statistics and Data Science (RSSD 2019), La Trobe University, Melbourne, 24–26 July 2019  
<https://sites.google.com/view/rssds2019/home>

## 6 Talks pléniers invités sur le thème du projet

1. F. Chamroukhi. Statistical data science and some unsupervised learning problems. Invited talk at the International Symposium on Data Science and Computational Intelligence DSCI 2018, December 2018. URL <https://chamroukhi.com/Talks/FChamroukhi-dsci-23122018.pdf>
2. A. Roche. Variable selection and estimation in multivariate functional linear regression. Workshop Advances in Functional Data Analysis: Cluster, Location and Shape, Oct. 2018
3. F. Chamroukhi. Mixtures-of-Experts for Functional Data. Talk at The University of Queensland, School of Mathematics and Physics, Brisbane, Australia, August 2019.  
<https://chamroukhi.com/Talks/FChamroukhi-UQ-08-2019.pdf>

## 7 Exposés de séminaire accueillis sur le thème du projet

1. Suzanne Varet (LMO, Université Paris-Saclay), Numerical performance of Penalised Comparison to Overfitting for bandwidth selection in kernel density estimation, Novembre 2018.
2. A. Channarond (LMO, Université Paris-Saclay), Clustering dans un modèle de graphe à positions latentes, Séminaire Proba-Stat, Décembre 2018.
3. G. Chagny (LMRS, Université de Rouen), Estimation pour des modèles de mélange semi-paramétriques, Journée interne du LMRS, Décembre 2018, Rouen.
4. Jeanne Minh-Lien Nguyen (Ceremade, Université Paris Dauphine), CDRodeo : Sélection gloutonne de fenêtres multivariées pour l'estimation de densité conditionnelle, Janvier 2019.
5. Luc Lehéricy (LMO, Université Paris-Saclay), Estimation non paramétrique pour des problèmes mal spécifiés non i.i.d.), Février 2019.
6. Charlotte Baey (LPP, Université de Lille), Test pour des composantes de la variance dans les modèles à effets mixtes, Avril 2019.
7. Émilie Lebarbier (AgroParisTech), Segmentation of time-series with dependence, Janvier 2019.
8. A. Meynaoui (LMRS, Université de Rouen), Aggregated test of independence based on HSIC measures, GT Statistique, Octobre 2020, Rouen.
9. T. Pham (LMNO, Université de Caen), Functional mixtures of experts. Séminaire de Statistiques Appliquées at LMNO, Novembre 2020, Caen.
10. G. Chagny (AgroParisTech), Estimation non-paramétrique dans un modèle de mélange à deux classes, Séminaire de l'équipe MIA, Novembre 2020, Paris.

## 8 Références

- E. Arias-Castro, A. Channarond, B. Pelletier, and N. Verzelen. Minimax estimation of distances in latent graphs. 2018. URL <http://arxiv.org/abs/1804.10611>.
- Ery Arias-Castro, Antoine Channarond, Bruno Pelletier, and Nicolas Verzelen. On the Estimation of Latent Distances Using Graph Distances. *Electronic Journal of Statistics*, 15(1):722–747, 2021. doi: 10.1214/21-EJS1801. URL <https://hal.archives-ouvertes.fr/hal-02376073>.
- Christophe Biernacki. MASSICCC: A SaaS Platform for Clustering and Co-Clustering of Mixed Data. In *APSEM 2019 ((Apprentissage et SEMantique)) : éco-systèmes pour la science ouverte et recherche par les données*, Toulouse, France, October 2019. URL <https://hal.archives-ouvertes.fr/hal-02399180>.
- Christophe Biernacki, Matthieu Marbac-Lourdelle, and Vincent Vandewalle. Gaussian Based Visualization of Gaussian and Non-Gaussian Based Clustering. In *SPSR 2019*, Bucarest, Romania, April 2019. URL <https://hal.archives-ouvertes.fr/hal-02400486>.
- Christophe Biernacki, Matthieu Marbac, and Vincent Vandewalle. Gaussian-Based Visualization of Gaussian and Non-Gaussian-Based Clustering. *Journal of Classification*, 38(1):129–157, 2021. ISSN 1432-1343. doi: 10.1007/s00357-020-09369-y. URL <https://doi.org/10.1007/s00357-020-09369-y>.
- Bérard C., Chagny G., Channarond A., Hoang V-H., Roche A., and Vergne N. Journée de la Fédération Normandie Mathématique.
- G. Chagny, T. Laloe, and R. Servien. Multivariate adaptive warped kernel estimation. *Electron. J. Stat.*, 13:1759–17894, 2019. ISSN 1935-7524. doi: 10.1214/19-EJS1565. URL <https://doi.org/10.1214/19-EJS1565>.
- F. Chamroukhi. Statistical data science and some unsupervised learning problems. Invited talk at the International Symposium on Data Science and Computational Intelligence DSCI 2018, December 2018. URL <https://chamroukhi.com/Talks/FChamroukhi-dsci-23122018.pdf>.
- F. Chamroukhi. Unsupervised statistical learning from heterogeneous time-series and functional data. Keynote note talk at the Research School on Statistics and Data Science-RSSDS 2019, July 2019. URL <https://chamroukhi.com/Talks/FChamroukhi-RSSDS2019.pdf>.
- F. Chamroukhi and C. Biernacki. Model-Based Co-Clustering of Multivariate Functional Data. In *ISI 2017 - 61st World Statistics Congress*, Marrakech, Morocco, Jul 2017. URL <https://chamroukhi.com/papers/ISI2017.pdf>.
- F. Chamroukhi and V-H Hoang. Functional mixtures-of-experts. CRONOS Multivariate Data Analysis workshop. Model-based and multivariate functional data Session, April 2019. URL <https://chamroukhi.com/Talks/FChamroukhi-CRONOS-MDA-2019.pdf>.
- Faïcel Chamroukhi and Bao T. Huynh. Regularized Maximum Likelihood Estimation and Feature Selection in Mixtures-of-Experts Models. *Journal de la Société Française de Statistique*, 160(1):57–85, March 2019. URL [https://chamroukhi.com/papers/Chamroukhi\\_Huynh\\_jsfds-published.pdf](https://chamroukhi.com/papers/Chamroukhi_Huynh_jsfds-published.pdf).
- Faïcel Chamroukhi and Hien D. Nguyen. Model-based clustering and classification of functional data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Dec 2018. URL

- <https://chamroukhi.com/papers/MBCC-FDA.pdf>. DOI: 10.1002/widm.1298.
- Faïcel Chamroukhi, Florian Lecocq, and Hien D Nguyen. Regularized estimation and feature selection in mixtures of gaussian-gated experts models. In *Research School on Statistics and Data Science*, pages 42–56. Springer, July 2019. URL [https://chamroukhi.com/papers/HD-MoGGE\\_Submitted](https://chamroukhi.com/papers/HD-MoGGE_Submitted).
- Faïcel Chamroukhi, Thien Nhat Pham, Van Ha Hoang, and Geoffrey J McLachlan. Mixtures-of-experts with functional predictors. London, UK, December 2021.
- A. Channarond. Clustering dans un modèle de graphe à positions latentes. Séminaire de Probabilités et Statistiques, Laboratoire de Mathématiques d’Orsay (UMR CNRS 8628), Dec 2018.
- Maxence Ferrari, Hervé Glotin, Ricard Marxer, and Mark Asch. DOCC10: Open access dataset of marine mammal transient studies and end-to-end CNN classification. In *IJCNN*, Glasgow, United Kingdom, July 2020. URL <https://hal.archives-ouvertes.fr/hal-02866091>.
- Van Ha Hoang, Gaëlle Chagny, Antoine Channarond, Van Hà Hoang, and Angelina Roche. Adaptive nonparametric estimation of a component density in a two-class mixture model. working paper or preprint, February 2021. URL <https://hal.archives-ouvertes.fr/hal-02909601>.
- T. Huynh and F. Chamroukhi. Estimation and feature selection in mixtures of generalized linear experts models. July 2019. URL <https://chamroukhi.com/papers/prEMME.pdf>.
- Hien D. Nguyen and Faïcel Chamroukhi. Practical and theoretical aspects of mixture-of-experts modeling: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, pages e1246–n/a, Jan 2018. ISSN 1942-4795. doi: 10.1002/widm.1246. URL <https://chamroukhi.com/papers/Nguyen-Chamroukhi-MoE-DMKD-2018>.
- Hien D. Nguyen, Faïcel Chamroukhi, and Florence Forbes. Approximation results regarding the multiple-output mixture of linear experts model. 2018. URL <http://arxiv.org/abs/1704.00946>. Submitted.
- Hien Duy Nguyen, TrungTin Nguyen, Faïcel Chamroukhi, and Geoffrey John McLachlan. Approximations of conditional probability density functions in lebesgue spaces via mixture of experts models. *Journal of Statistical Distributions and Applications*, 8(1):1–15, 2021a. URL <https://arxiv.org/pdf/2012.02385.pdf>.
- T-T Nguyen, H D Nguyen, F Chamroukhi, and G J McLachlan. Approximation by finite mixtures of continuous density functions that vanish at infinity. *Submitted*, March 2019. URL <https://arxiv.org/abs/1903.00147>. Submitted.
- TrungTin Nguyen, Faïcel Chamroukhi, Hien D Nguyen, and Geoffrey J McLachlan. Approximation of probability density functions via location-scale finite mixtures in lebesgue spaces. 2020a. URL <https://arxiv.org/pdf/2008.09787.pdf>.
- TrungTin Nguyen, Hien D Nguyen, Faïcel Chamroukhi, and Geoffrey J McLachlan. An  $l_1$ -oracle inequality for the lasso in mixture-of-experts regression models. 2020b. URL <https://arxiv.org/pdf/2009.10622.pdf>.
- TrungTin Nguyen, Faïcel Chamroukhi, Hien Duy Nguyen, and Florence Forbes. Non-asymptotic model selection in block-diagonal mixture of polynomial experts models. may 2021b. URL <https://arxiv.org/pdf/2104.08959.pdf>.
- TrungTin Nguyen, Hien Duy Nguyen, Faïcel Chamroukhi, and Florence Forbes. A non-asymptotic penalization criterion for model selection in mixture of experts models. 2021c. URL

<https://arxiv.org/pdf/2104.02640.pdf>.

- Marie A. Roch, Scott Lindeneau, Gurisht Singh Aurora, Kaitlin E. Frasier, John A. Hildebrand, Hervé Glotin, and Simone Baumann-Pickering. Using context to train time-domain echolocation click detectors. *The Journal of the Acoustical Society of America*, 149(5):3301–3310, 2021. doi: 10.1121/10.0004992. URL <https://doi.org/10.1121/10.0004992>.
- A. Roche. Variable selection and estimation in multivariate functional linear regression. Workshop Advances in Functional Data Analysis: Cluster, Location and Shape, Oct. 2018.
- A. Roche. Variable selection and estimation in multivariate functional linear regression via the lasso. 2019a. URL [hal.archives-ouvertes.fr/hal-01725351](https://hal.archives-ouvertes.fr/hal-01725351).
- A. Roche. Variable selection and estimation in multivariate functional linear regression. Séminaire de Statistiques du MAP5 (UMR CNRS 8145), Université Paris Descartes, Juin 2019b.