Leveraging statistical mixtures and learning techniques in some data science problems

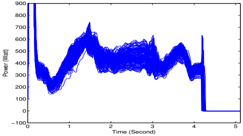
Faïcel Chamroukhi

Talk @ JSDMS 2023, November 12, Hammamet

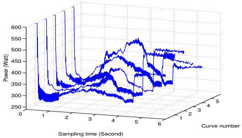
Real-world data are complex

- Heterogenous, Multimodal, High-Dimensional, Unlabeled, Possibly Massive ...
- Need for adapted analysis tools

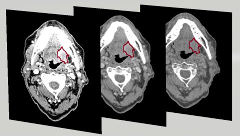
Transport : Railway switch curves diagnostic



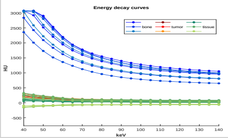
Predictive Maintenance



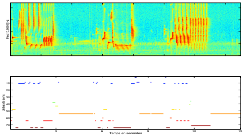
Health : Medical images

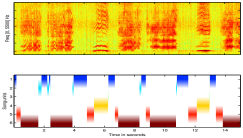


Dual-energy computed tomography

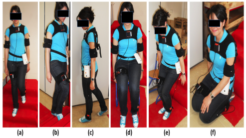


Acoustics : scene listening (marine, terrestrial)

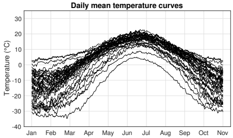


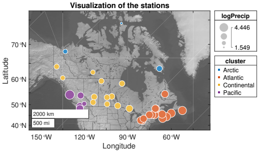


Health & Well Being : Activity recog.



Climate/Environment : meteorological data





- Establish well-principled (with statistical guarantees) predictions in heterogeneous, high-dimensional and decentralized situations,
- Construct efficient algorithms that operate in unsupervised way, provide interpretable solutions and enjoy computational guarantees.

Scientific framework

$\hookrightarrow$ Latent variable **models** : $f(x|\boldsymbol{\theta}) = \int_{\mathcal{Z}} f(x, z|\boldsymbol{\theta})\mathrm{d}z$

$\hookrightarrow$ **Inference**, **Selection** and **representation unsupervised** and at scale

1 Scientific Challenges
2 Latent Variable Models
- Mixture models
- Mixtures of Experts Models
3 High-Dimensional Learning
- Learning with high-dimensional predictors
- Learning with functional predictors
4 Federated Learning

Approximation capabilities of finite mixture distributions

Density approximation in Unsupervised Learning

- **Data** : observations $\{\boldsymbol{x}_i\}$ d'une v.a $\boldsymbol{X} \in \mathbb{X} \subset \mathbb{R}^d$ à densité (multimodale) $f \in \mathcal{F}$
- **Objectif** : approcher la densité cible f (et représenter les données, e.g. *clustering*)
- **Solution :** Approcher f dans la classe $\mathcal{H}^{\varphi} = \bigcup_{K \in \mathbb{N}^\star} \mathcal{H}_K^{\varphi}$ des **mélanges finis** h_K^{φ} (à K-composants) **de translatées dilatées** d'une densité φ (e.g., gaussienne), où

$$\mathcal{H}_K^{\varphi} = \left\{ \boxed{h_K^{\varphi}(\boldsymbol{x}) := \sum_{k=1}^{K} \pi_k \frac{1}{\sigma_k^d} \varphi\left(\frac{\boldsymbol{x} - \boldsymbol{\mu}_k}{\sigma_k}\right)}, \boldsymbol{\mu}_k \in \mathbb{R}^d, \sigma_k \in \mathbb{R}_+, \pi_k > 0 \, \forall k \in [K], \sum_{k=1}^{K} \pi_k = 1 \right\}$$

Théorème : Universal approximation of finite mixtures models (FMM)

(a) Pour toute f.d.p $f, \varphi \in \mathcal{C}$ et un ensemble compact $\mathbb{X} \subset \mathbb{R}^d$, il existe une suite $(h_K^{\varphi}) \subset \mathcal{H}^{\varphi}$, telle que $\lim_{K \to \infty} \sup_{\boldsymbol{x} \in \mathcal{X}} |f(\boldsymbol{x}) - h_K^{\varphi}(\boldsymbol{x})| = 0$.

(b) Pour $p \in [1, \infty)$, si $f \in \mathcal{L}_p$ (f.d.p de Lebesgue) et $\varphi \in \mathcal{L}_\infty$ (f.d.p essentiellement bornée), il existe une suite $(h_K^{\varphi}) \subset \mathcal{H}^{\varphi}$, telle que $\lim_{K \to \infty} \|f - h_K^{\varphi}\|_{\mathcal{L}_p} = 0$.

[J. Communications in Statistics - Theory and Methods, 2022] [PhD, TT Nguyen 2021]

Gaussian mixture models (GMMs)

The finite Gaussian mixture density is defined as :

$$h_K^{\mathcal{N}}(\boldsymbol{x}_i;\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}_i;\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$$

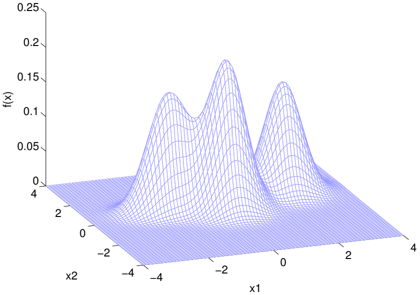


FIGURE – An example of a three-component Gaussian mixture density in $\mathbb{R}^2$.

Finite Mixture Models

$h_K^{\varphi}(\boldsymbol{x}_i;\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \varphi(\boldsymbol{x};\boldsymbol{\theta}_k)$ with $\pi_k > 0 \; \forall k$ and $\sum_{k=1}^{K} \pi_k = 1$.

Maximum-Likelihood Estimation

$$\widehat{\boldsymbol{\theta}} \in \arg\max_{\boldsymbol{\theta}} \ln L(\boldsymbol{\theta})$$

log-likelihood : $\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln \sum_{k=1}^{K} \pi_k \varphi(\boldsymbol{x}_i;\boldsymbol{\theta}_k)$.

The EM algorithm [DLR]

$$\boldsymbol{\theta}^{new} \in \arg\max_{\boldsymbol{\theta}\in\Omega} \mathbb{E}[\ln L_c(\boldsymbol{\theta})|\mathcal{D},\boldsymbol{\theta}^{old}]$$

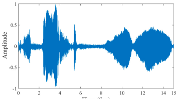
complete log-likelihood : $\ln L_c(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{k=1}^{K} Z_{ik} \log\left[\pi_k \varphi(\boldsymbol{x}_i;\boldsymbol{\theta}_k)\right]$ where Z_{ik} is such that $Z_{ik} = 1$ if $Z_i = k$ and $Z_{ik} = 0$ otherwise.

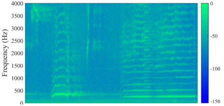
Clustering

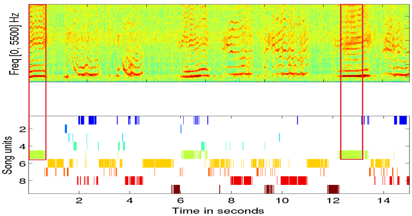
$$\widehat{z}_i = \arg\max_{1\leq k\leq K} \mathbb{P}(Z_i = k|\boldsymbol{x}_i;\widehat{\boldsymbol{\theta}}), \quad (i = 1,\ldots,n)$$

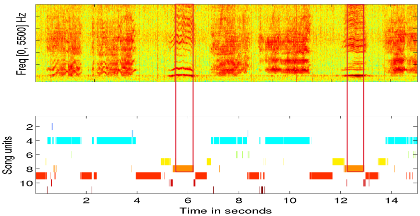
Dirichlet Process Parsimonious Gaussian Mixture for Images clustering

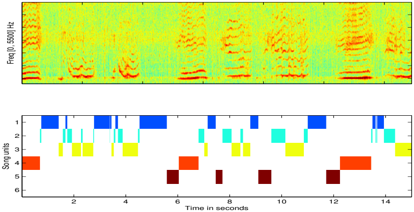
Unsupervised decomposition
of whale song signals

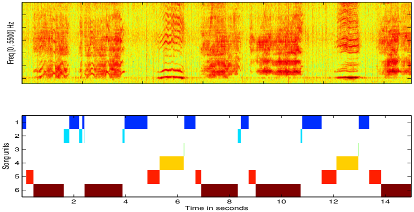








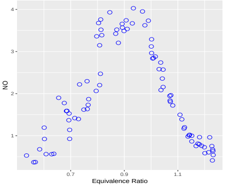


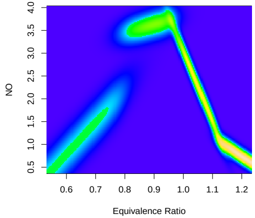


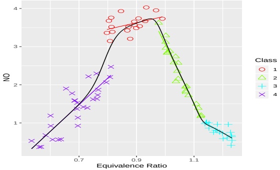
[PhD, M. Bartcus 2015] [Multimedia Tools and Applications, 2018] [wkp ICML 14, IPCR 14, IJCNN 15] [Sabiod.org]
arXiv :1501.03347v2, 2018

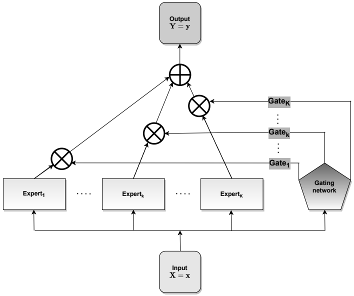
Heterogeneous regression-type data

Mixtures-of-Experts as good candidates to model a response Y given covariarte.s X when governed by a hidden structure accounting for heterogeneity









Apprentissage par Modèles de Mélanges d'Experts (ME)

- **Contexte :** n observations $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$ d'un couple $(\boldsymbol{X}, \boldsymbol{Y}) \in \mathbb{X} \times \mathbb{Y}$ lié via une f.d.p conditionnelle inconnue $f \in \mathcal{F} = \left\{f : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}_+ \middle| \int_{\mathbb{Y}} f(\boldsymbol{y}|\boldsymbol{x}) \, \mathrm{d}\lambda(\boldsymbol{y}) = 1, \forall \boldsymbol{x} \in \mathbb{X}\right\}$
- Scénario de **grande dimension** : $\mathbb{X} \subseteq \mathbb{R}^d$, $\mathbb{Y} \subseteq \mathbb{R}^q$, avec $d, q \gg n$ et **hétérogène**.
- **Objectifs :** Regression ; Clustering ; Sélection de modèle
- **Solution :** Approcher f dans la classe des **mélanges d'experts** :

Soit une f.d.p φ (et un support compact $\mathbb{Y} \subseteq \mathbb{R}^q$), on déifinit les classes suivantes :

- Transaltées-dilatées : $\mathcal{E}_\varphi = \left\{\phi_q(\boldsymbol{y}; \boldsymbol{\mu}, \sigma) := \frac{1}{\sigma^q} \varphi\left(\frac{\boldsymbol{y} - \boldsymbol{\mu}}{\sigma}\right); \boldsymbol{\mu} \in \mathbb{Y}, \sigma \in \mathbb{R}_+\right\}$.
- Mélanges d'experts transaltées-dilatés à réseau d'activation softmax : SGaME :

$$\mathcal{H}_S^\varphi = \left\{\boxed{h_K^\varphi(\boldsymbol{y}|\boldsymbol{x}) := \sum_{k=1}^{K} g_k(\boldsymbol{x}; \boldsymbol{\gamma}) \phi_q(\boldsymbol{y}; \boldsymbol{\mu}_k, \sigma_k)}; \quad \phi_q \in \mathcal{E}_\varphi \cap \mathcal{L}_\infty, g_k(\cdot; \boldsymbol{\gamma}) \in \{\text{softmax}\}\right\}$$

Theorem : Approximation capabilities of isotropic mixtures-of-experts SGaME

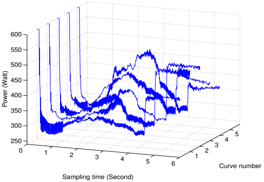
(a) Pour $p \in [1, \infty)$, $f \in \mathcal{F}_p \cap \mathcal{C}$, $\varphi \in \mathcal{F} \cap \mathcal{C}$, $\mathbb{X} = [0, 1]^d$, il existe une suite $\left(h_K^\varphi\right) \subset \mathcal{H}_S^\varphi$ telle que $\lim_{K \to \infty} \left\|f - h_K^\varphi\right\|_{\mathcal{L}_p} = 0$.

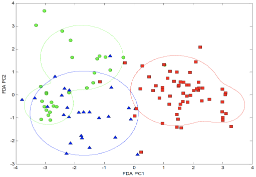
(b) Pour toute $f \in \mathcal{F} \cap \mathcal{C}$, si $\varphi \in \mathcal{F} \cap \mathcal{C}$, $d = 1$, il existe une suite $\left(h_K^\varphi\right) \subset \mathcal{H}_S^\varphi$ telle que $\lim_{K \to \infty} h_K^\varphi = f$ presque uniformément.

[PhD TT. Nguyen, 2021] [Journal of Stat. Distributions and Applicat., 2021] [Neurocomputing, 2019] [WIREs DMKD 2018]

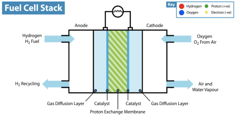
Time-Series Analysis Applications

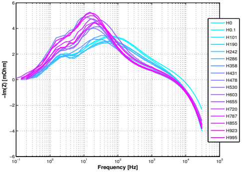
Transport : Railway switch operating state prediction {Collab. avec la SNCF ; Projet Switch-Rdf}

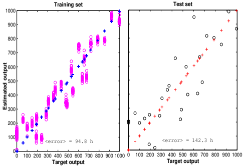




Energy : Fuel cell lifetime prediction {Collab. with Femto-ST, Phd of R. Onanena, 2012}

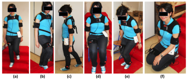


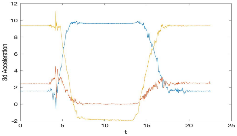


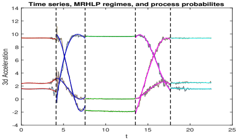


Health & well being : Human activity recognition {Collab. with Paris 12-LiSSi}

[PhD, D. Trabelsi, 2013][Sensors 2015, 748 citations]







Open-Source Toolkits

SaMUraiS : open source software for statistical time-series analysis

SaMUraiS : **St**A**tistical** **M**odels for the **U**nsupe**R**vised segment**A**t**I**on of time-**S**eries

▸ Github ▸ CRAN ▸ Matlab software

Available algorithms and Packages

`RHLP` : Regression with Hidden Logistic Process	▸ R software	▸ Matlab software
`HMMR` : Hidden Markov Model Regression	▸ R software	▸ Matlab software
`PWR` : Piece-Wise Regression	▸ R software	▸ Matlab software
`MRHLP` : Multivariate RHLP	▸ R software	▸ Matlab software
`MHMMR` : Multivariate HMMR	▸ R software	▸ Matlab software
`MPWR` : Multivariate PWR	▸ R software	▸ Matlab software

Include estimation, segmentation, approximation, model selection, and sampling

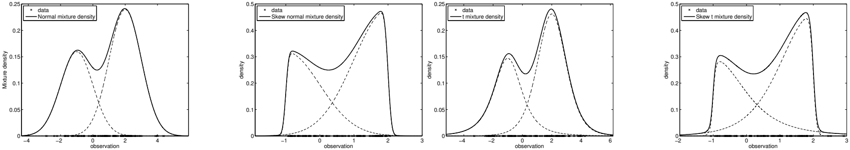
Principled robustness in learning with MoE

Principled robustness in regression and clustering

- **Questionings :** Prediction (non-linear regr., classification) & clustering in presence of **Outliers**, with potentially **skewed**, **heavy-tailed** distributions
- **Answering :** Robust MoE that accommodate asymmetry, heavy tails, and outliers

$$m(y|\boldsymbol{r}, \boldsymbol{x}; \boldsymbol{\theta}) = \sum_{k=1}^{K} \underbrace{g_k(\boldsymbol{r}; \boldsymbol{\alpha})}_{\text{Softmax Gating Network}} \underbrace{\mathcal{ST}(y; \mu(\boldsymbol{x}; \boldsymbol{\beta}_k), \sigma_k, \lambda_k, \nu_k)}_{\text{Skew-}t\text{ Expert Network}}$$

kth expert : has a skew t distribution [Azzalini and Capitano 2003]



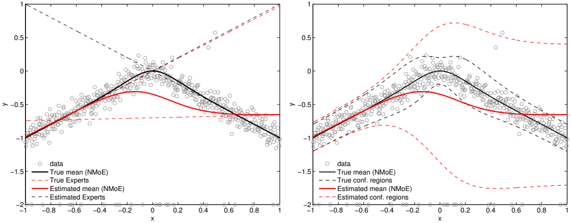
$$\pi_k = [0.4, 0.6],\ \mu_k = [-1, 2]\,;\ \sigma_k = [1, 1]\,;\ \nu_k = [3, 7]\,;\ \lambda_k = [14, -12]\,;$$

Flexible and robust generalization of the standard MoE models

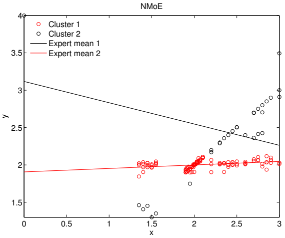
For $\{\nu_k\} \to \infty$, STMoE reduces to SNMoE ; For $\{\lambda_k\} \to 0$, STMoE reduces to TMoE.
For $\{\nu_k\} \to \infty$ and $\{\lambda_k\} \to 0$, StMoE approaches the NMoE.

[*Neurocomputing*, 2017] [*Neural Networks*, 2016c] [IJCNN 2016]

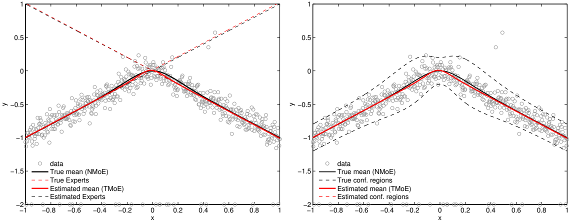
Robust learning with mixtures-of-experts models



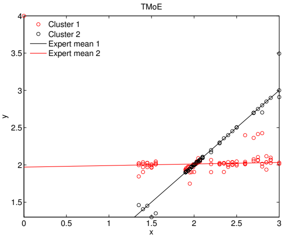
$n = 500$ observations with 5% of outliers $(x; y = -2)$: Normal fit



Tone data with 10 outliers $(0, 4)$: Normal fit



$n = 500$ observations with 5% of outliers $(x; y = -2)$: **Robust** fit

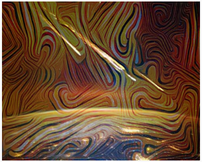


Tone data with 10 outliers $(0, 4)$: **Robust** fit

[*Neurocomputing*, 2017] [*Neural Networks*, 2016c] [IJCNN 2016]

Open-Source Toolkits

MEteorits : open-source soft. Robust learning with mixtures-of-experts models



MEteorits : **M**ixtures-of-**E**xper**T**s mod**E**ling for c**O**mplex and non-no**R**mal d**I**s**T**ributionS

▸ Github ▸ CRAN ▸ Matlab software

Available algorithms and Packages

`NMoE` : Normal Mixture-of-Experts	▸ R software	▸ Matlab software
`SNMoE` : Skew-Normal Mixture-of-Experts	▸ R software	▸ Matlab software
`tMoE` : Robust MoE using the t-distribution	▸ R software	▸ Matlab software
`StMoE` : Skew-t Mixture-of-Experts	▸ R software	▸ Matlab software

- Meteorits include sampling, fitting, prediction, clustering with each MoE model
- Non-normal mixtures (and MoE) is a very recent topic in the field

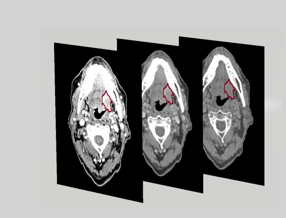
DECT image Clustering (Healthcare/Radiology)

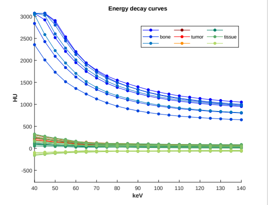
- Learning from Multimodal information in precision medicine
- HNSCC Cancer detection in Radiology : DECT clustering

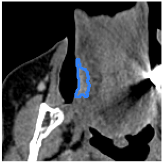
{Collab. with McGill & Florida, College of Medicine} [Diagnostics (AI in medicine), 2022]

Spatial mixture of functional regressions for dual-energy CT images

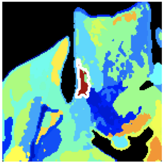
$$m(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{v}; \boldsymbol{\theta}) = \sum_{k=1}^{K} \alpha_k(\boldsymbol{v}; \boldsymbol{\alpha}) f_k(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{\theta}_k) \text{ où } \quad \alpha_k(\boldsymbol{v}; \boldsymbol{\alpha}) = \frac{w_k \phi_3(\boldsymbol{v}; \boldsymbol{\mu}_k, \mathbf{R}_k)}{\sum_{\ell=1}^{K} w_\ell \phi_3(\boldsymbol{v}; \boldsymbol{\mu}_\ell, \mathbf{R}_\ell)}$$





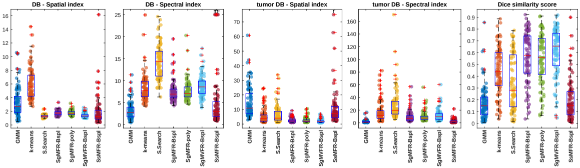


(a) Original slice



(b) Dice=0.84, DB=1.64/6.92

DECT multimodal Data : 3D voxels & energy levels Expert Annotation Automatic Annotation



[A.] Inférence en grande dimension

Questioning : Prediction (non-linear regr., classification) & clustering in presence of

[1.] **High-dimensional** predictors : $\boldsymbol{X}_i \in \mathbb{R}^p$ with $p \gg n$

[2.] **Functional** predictors : $\boldsymbol{X}_i(t)$, $t \in \mathcal{T} \subseteq \mathbb{R}$ {eg. continuous recorded variables}

$\hookrightarrow$ Méthodes d'**Inférence** et **Sélection parcimonieuses**, Soucis d'**interprétabilité**

[1.] HDME : High-Dimensional Mixtures-of-Experts

- Learning : PMLE $\widehat{\boldsymbol{\theta}}_{\mathrm{n}} \in \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{n} \log h_K^{\varphi}(\boldsymbol{y}_i|\boldsymbol{x}_i;\boldsymbol{\theta}) - \mathrm{pen}(\boldsymbol{\theta})$
- $\hookrightarrow$ LASSO penalty : $\mathrm{Pen}_{\lambda}(\boldsymbol{\theta}) = \underbrace{\sum_{k=1}^{K} \lambda_k \|\boldsymbol{\beta}_k\|_1}_{\text{Experts Net.}} + \underbrace{\sum_{k=1}^{K-1} \gamma_k \|\boldsymbol{w}_k\|_1}_{\text{Gating Net.}}$

$\hookrightarrow$ encourages sparse solutions & performs estimation and feature selection

$\hookrightarrow$ computationally attractive (Avoid matrix inversion ; univariate CF updates)

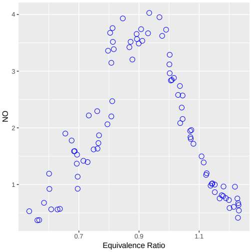
▸ Software Toolbox HDME on Github (GaussRMoE, LogisticRMoE, PoissonRMoE)

[Thèse] Bao Tuyen Huynh. *Estimation and Feature Selection in High-Dimensional Mixtures-of-Experts Modesls* . Thèse de Doctorat de Normandie Université, 2019.

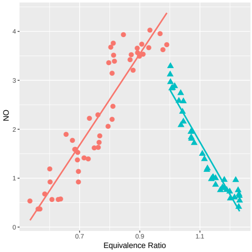
[J] Chamroukhi &Huynh. Regularized Maximum Likelihood Estimation and Feature Selection in Mixtures-of-Experts Models. Journal de la Société Francaise de Statistique, Vol. 160(1), pp :57–85, 2019

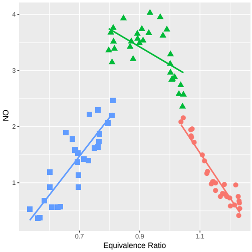
[J] Huynh & C. Estimation and Feature Selection in Mixtures of Generalized Linear Experts Models. arXiv :1810.12161, 2019

(a) Raw Ethanol data set

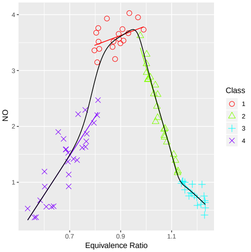


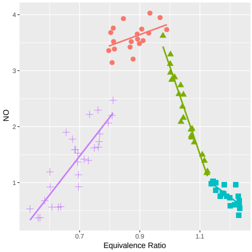
Collection of MoE models with linear mean functions characterized by 2-5 clusters

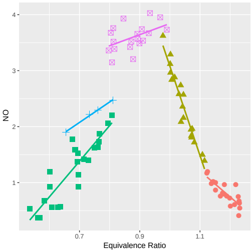




(b) Our best data-driven MoE model







Questioning : Prediction (non-linear regr., classification) & clustering in presence of **High-dimensional** predictors : Data $\mathcal{D}_n = (\boldsymbol{X}_i, Y_i)_{i=1}^n$ where $\boldsymbol{X}_i \in \mathbb{R}^p$ with $p \gg n$

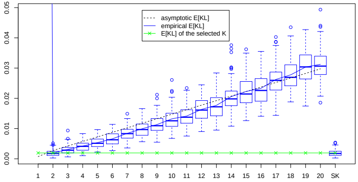
HDME : High-Dimensional MoE : PMLE $\widehat{\boldsymbol{\theta}}_n \in \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^n \log h_K^{\varphi}(\boldsymbol{y}_i|\boldsymbol{x}_i;\boldsymbol{\theta}) - \mathrm{pen}(\boldsymbol{\theta})$

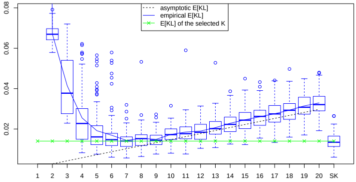
Theorem : Non-asymptotic oracle inequality for collection of MoE models

Résultat : $\exists$ des constantes C et $\kappa(\rho, C_1) > 0$ ($C_1 > 1$) t.q chaque fois que pour $\mathbf{m} \in \mathcal{M}$, $\mathrm{pen}(\mathbf{m}) \geq \kappa(\rho, C_1)\left[(C + \ln n)\dim(\mathcal{H}_{\mathbf{m}}) + z_{\mathbf{m}}\right]$, l'estimateur PMLE $\widehat{h}_{\widehat{\mathbf{m}}}$ satisfait

$$\mathbb{E}\left[\mathrm{JKL}_{\rho}^{\otimes \mathrm{n}}\left(f, \widehat{h}_{\widehat{\mathbf{m}}}\right)\right] \leq C_1 \inf_{\mathbf{m}\in\mathcal{M}}\left(\inf_{h_{\mathbf{m}}\in\mathcal{H}_{\mathbf{m}}} \mathrm{KL}^{\otimes \mathrm{n}}(f, h_{\mathbf{m}}) + \frac{\mathrm{pen}(\mathbf{m})}{n}\right) + \frac{\kappa(\rho, C_1) C_1 \xi}{n} + \frac{\eta + \eta'}{n}.$$

- Résultat non-asymptotique. Si $\mathrm{pen}(\mathbf{m})$ est bien choisie, alors notre PMLE se comporte de manière comparable au **meilleur modèle (oracle)** $h_{\mathbf{m}^\star}$ de la collection, minimisant le risque : $\inf_{\mathbf{m}\in\mathcal{M}}\left(\inf_{h_{\mathbf{m}}\in\mathcal{H}_{\mathbf{m}}} \mathrm{KL}^{\otimes \mathrm{n}}(f, h_{\mathbf{m}}) + \frac{\mathrm{pen}(\mathbf{m})}{n}\right)$ (f est inconnue).





[Thèse, Trung-Tin Nguyen, 2021.] [*Electronic Journal of Statistics*, **2022**] [In revision, *JMVA*. arXiv :2104.08959. 2021b]

FLaMingoS : open source software for learning from functions

FLaMingoS : **F**unctional **L**atent dat**A** **M**odels for cluster**ING** heterogene**O**us time-**S**eries

▸ Github ▸ CRAN ▸ Matlab software

Available algorithms and Packages

`mixRHLP` : Mixture of Regressions with HLPs	▸ R	▸ Matlab
`mixHMM` : Mixture of Hidden Markov Models (HMMs)	▸ R	▸ Matlab
`mixHMMR` : Mixture of HMM Regressions	▸ R	▸ Matlab
`PWRM` : Piece-Wise Regression Mixture	▸ R	▸ Matlab
`uReMix` : Unsupervised Regression Mixtures	▸ R	▸ Matlab

↪ A flexible full generative modeling for FDA

↪ Could be extended to the multivariate case without a major effort

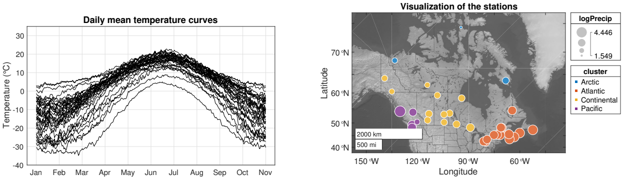


FIGURE – $n = 35$ daily mean temperature measurement curves (X_i's) in different stations (Left) and the $\log$ of precipitation values (Y_i's) visualized with the climate regions (Z_i's) (Right).

- Relate functional predictors $\{X(t) \in \mathbb{R}; t \in \mathcal{T} \subset \mathbb{R}\}$ to a scalar response $Y \in \mathcal{Y} \subset \mathbb{R}$
- Regression and classification of heterogeneous responses given functional predictors

  (1) generative functional modeling, sparsity and feature selection (high-dimension)

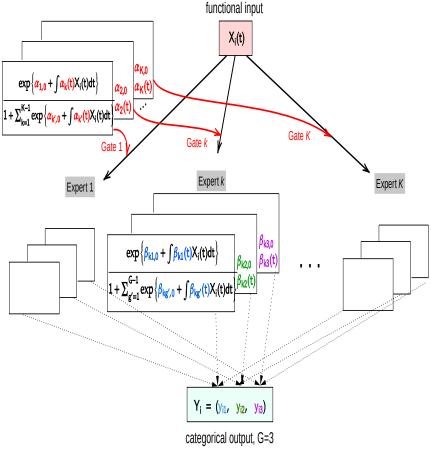
  (2) User guideline : keep an interpretable fit

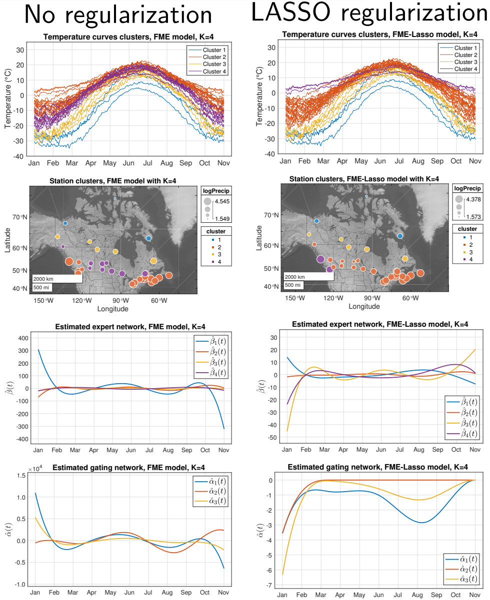
[2.] Functional Mixtures-of-Experts (and Different Learning strategies, in particular)

- $Y_i = \beta_{z_i,0} + \int_{\mathcal{T}} X_i(t)\beta_{z_i}(t)dt + \varepsilon_i$ avec $h_z(X_i(.)) = \alpha_{z_i,0} + \int_{\mathcal{T}} X_i(t)\alpha_{z_i}(t)dt$
- Lasso-type Regularized MLE w.r.t the derivatives of the $\alpha(\cdot)$ and $\beta(\cdot)$ functions

[Conf] Chamroukhi, Pham, Hoang, McLachlan. Mixtures-of-experts with functional predictors. CMStatistics 2021
[Preprint] —. Functional Mixtures-of-Experts. arXiv :2202.02249, Feb, 2022 (**Under Review, Statistics and Computing**)

Mixture-of-Experts Architecture

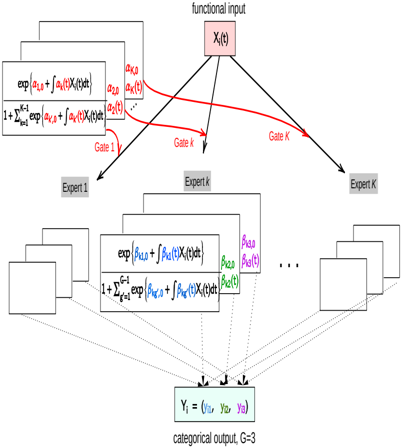




[PhD TN. Pham, 2022] [arXiv :2202.13934] [Submitted, 2023] [**Contract**, ANR SMILES]

Interpretable learning with time-series inputs

Mixture-of-Experts Architecture



produces a meaningful sparse estimates for $\beta_{z_i}(t)$ curves :

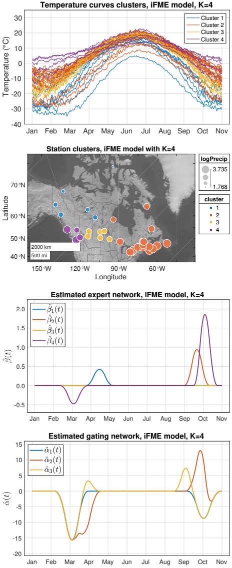
$\beta^{(0)}_{z_i}(t) = 0$ implies that $X(t)$ has no effect on Y at t

$\beta^{(1)}_{z_i}(t) = 0$ means that $\beta_{z_i}(t)$ is constant at t,

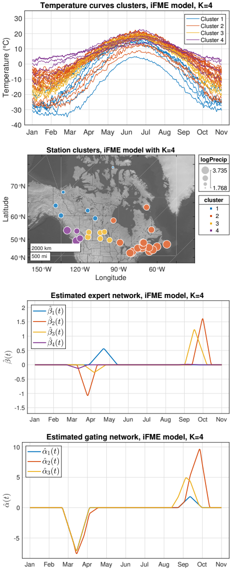
$\beta^{(0)}_{z_i}(t) = 1$ shows that $\beta_{z_i}(t)$ is a linear function of t,

etc.

OUR regularization



OUR regularization



[PhD TN. Pham, 2022] [arXiv :2202.13934] [Submitted, 2023] [**Contract**, ANR SMILES]

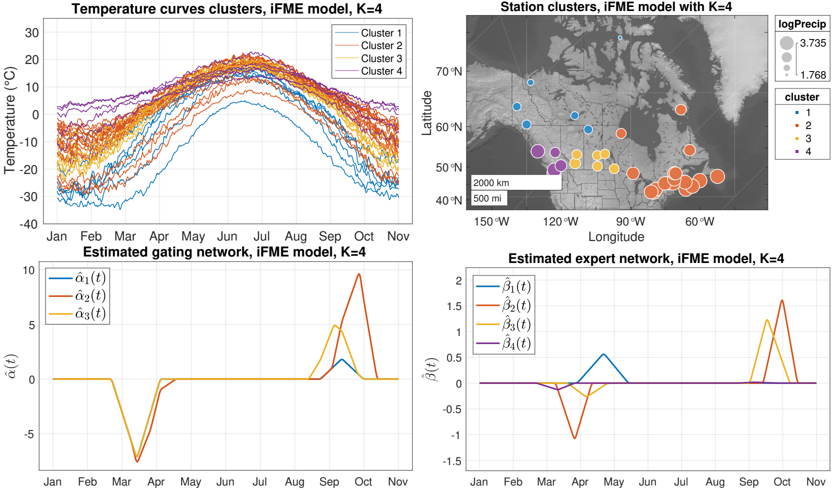
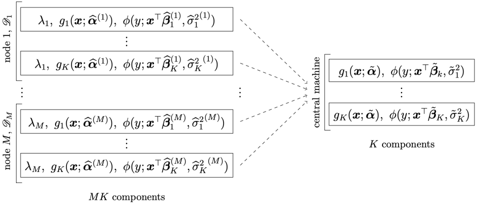


FIGURE – Clustering des températures (g.) et (log)-précipitations prédites (d.) (Atlantique, Pacifique, Continentale et Arctique), et fonctions paramètres (bg - bd.)

[PhD TN. Pham, 2022] [arXiv :2202.13934] [Submitted, 2023] [**Contract**, ANR SMILES]

Aggregating distributed mixtures-of-experts models

- **Clustering** for an informative summary of the data and **MoE** for better prediction
- collaborative mixtures-of-experts for large-scale data



- Local estimators : $\hat{f}_m = f(\cdot|\mathbf{x},\widehat{\boldsymbol{\theta}}_m) = \sum_{k=1}^{K} g_k(\mathbf{x},\widehat{\boldsymbol{\alpha}}^{(m)})\phi(\cdot;\mathbf{x}^\top\widehat{\boldsymbol{\beta}}_k^{(m)},\widehat{\sigma}_k^{2\,(m)})$,
- weighted average : $\bar{f} = f(y|\mathbf{x};\bar{\boldsymbol{\theta}}) = \sum_{m=1}^{M}\lambda_m \hat{f}_m$ where $\lambda_m = \frac{N_m}{N}$ the sample proportion. $\bar{f}$ is good but relates MK components so not our direct target.

$\hookrightarrow$ Reduced estimator : $\bar{f}^R = \underset{h_K\in\mathcal{M}_K}{\arg\inf}\ \rho\left(h_K, \sum_{m=1}^{M}\lambda_m\hat{f}_m\right)$: we seek for a K-component ME h that is closest to the MK-component ME $\bar{f} = \sum_{m=1}^{M}\lambda_m\hat{f}_m$ w.r.t a transportation divergence $\rho(\cdot,\cdot)$, e.g. KL.

{PhD, Pham. *Mixtures of Experts for Distributed Data, 2022*} [to be submitted 2023] [**Contrat**, ANR SMILES]

Federated Learning

Numerical results in Distributed clustering and Prediction

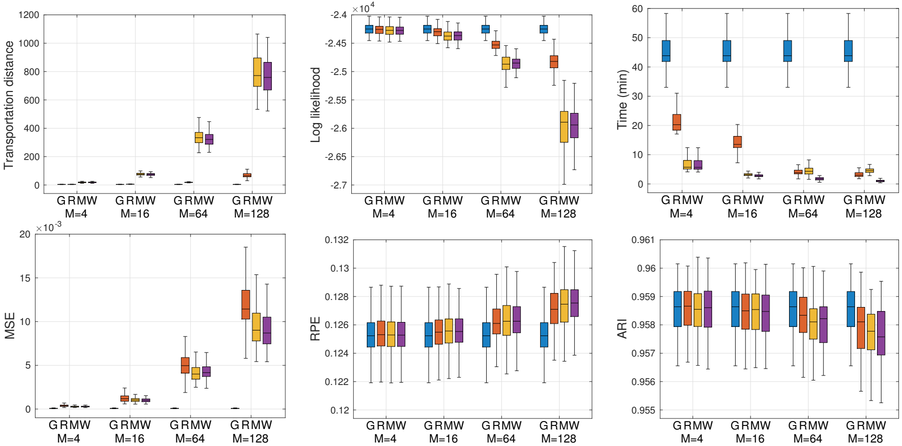


FIGURE – Performance of the Global ME (G), Reduction (R), Middle (M) and Weighted average (W) estimator for sample size $N = 10^6$ and M machines.

{PhD, Pham. *Mixtures of Experts for Distributed Data, 2022*} [to be submitted 2023] [**Contrat**, ANR SMILES]

Some Data Science and AI activities @ SystemX

Science des données, IA et Interaction

Use case-oriented technological research

Designing and interacting with AI systems : from data management to hybrid, robust and explainable AI

Interaction

- Decision support tools
- Interpretability of results
- Context-aware models
- Interactive **reinforcement** learning

Data

- Data lifecycle management
- Data quality and bias characterization
- Operational domain design (ODD)
- Building datasets for industry

Hybrid AI

- **Hybridizing AI** models with...
  ... **expert knowledge**
  ... **physical models**
- Augmentation by **simulation**
- Integrating AI components into complex systems

Trustworthy AI

- **Xplainalbility** for users and designers
- Assessment & improvement of **robustness**
- Assessment and removal of **bias**
- **Monitoring** of AI components

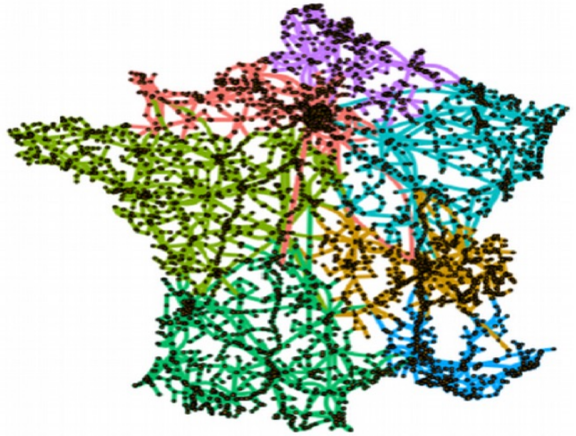
Structuring R&D programs: Confiance.AI (Trustworthy AI) and IA2 (Hybrid AI)

ML for Physical Simulation in Industry

Physics-Informed Machine Learning

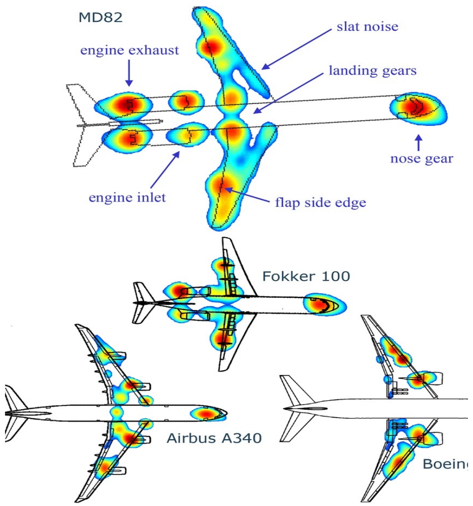
- Covering various fields in physics (mechanics, fluid dynamics, aerodynamics, electromagnetism …)
- In a wide variety of applications in industry, in particular in numerical simulation

Electricity (power grids)



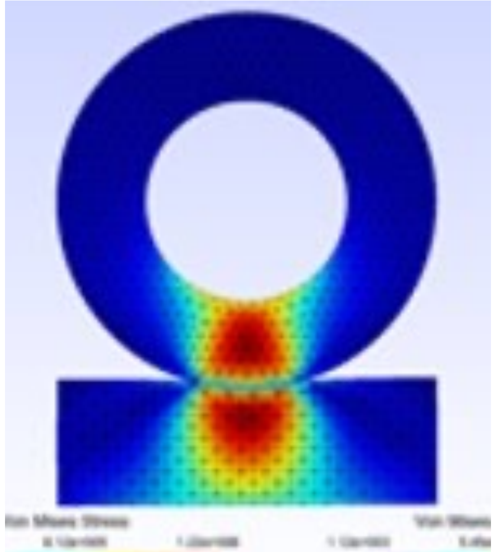
Picture from Marot, A., et al. (2018, October). Guided machine learning for power grid segmentation. In *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)* (pp. 1-6).

Aerodynamics



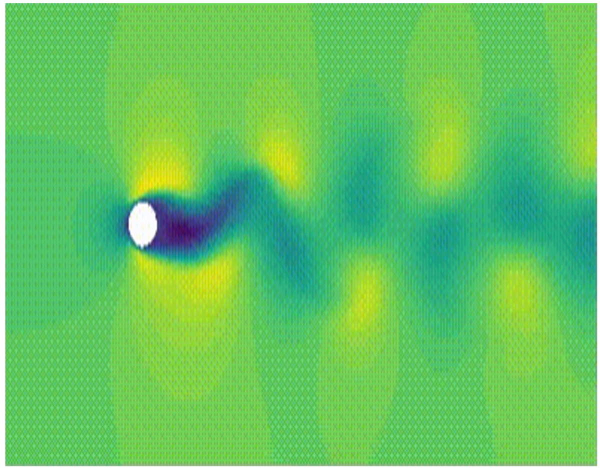
Merino-Martínez *et al. CEAS Aeronautical* Journal (2019).

Solid Mechanics pneumatics



From the internet

Fluid Flows/Dynamics



Picture from Emmanuel Menier

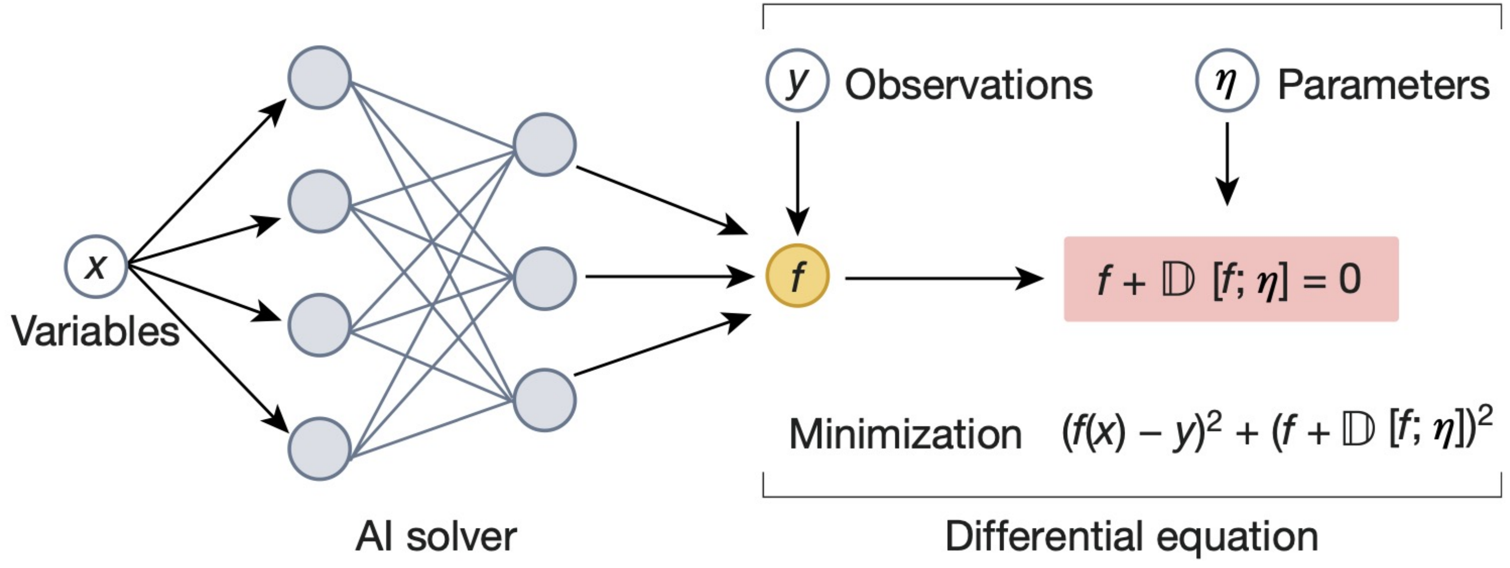
Domain challenges : Physical systems that are
- Complex to model/solve analytically
- Compuationally expensive to solve numerically

eg. , Computational Fluid Dynamics – CFD, Turbulance, Flows

Scientific Challenges

- Problems highly-nonlinear, high-dimensional, with complex structures (eg. organized in graphs…
- Need for adapted NN architectures: Graph NNets, Deep AE ..

Hybrid ML modeling for solving Partial Differential Equations



A neural framework for solving PDEs, where

- the AI solver is a PINN trained to estimate target function ***f***.
- The derivative of ***x*** is calculated by automatically differentiating the NN's outputs.
- When the differential equation **D(*f;η*)** is unknown (parameterized by ***η***), it can be estimated by solving a multi-objective loss that optimizes both the functional form of the equation and its fit to observations ***y***.

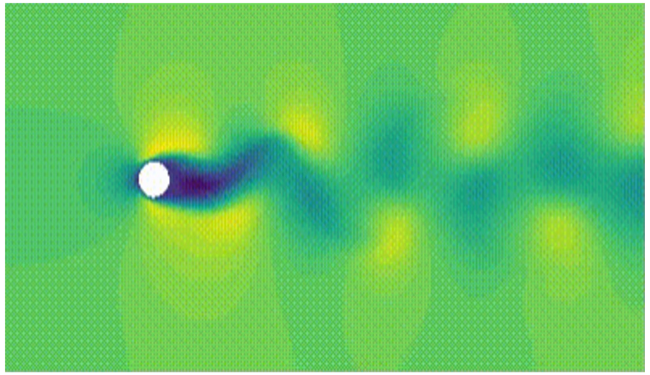
Wang & al. (2023). Scientific discovery in the age of artificial intelligence. Nature, 620. Read Online

- Eg. Learning Computational Fluid Dynamics

- Navier-Stokes Equations: fundamental partial differentials equations (PDE) that describe the flow of incompressible fluids.

C.L. M. H. Navier, Memoire sur les Lois du Mouvements des Fluides, Mem. de l'Acad. d. Sci.,6, 398 (1822)
C.G. Stokes, On the Theories of the Internal Friction of Fluids in Motion, Trans. Cambridge Phys. Soc., 8, (1845)

- Challenge: High-Dimensional non-linear Physical Equations



Simulation from Emmanuel Menier

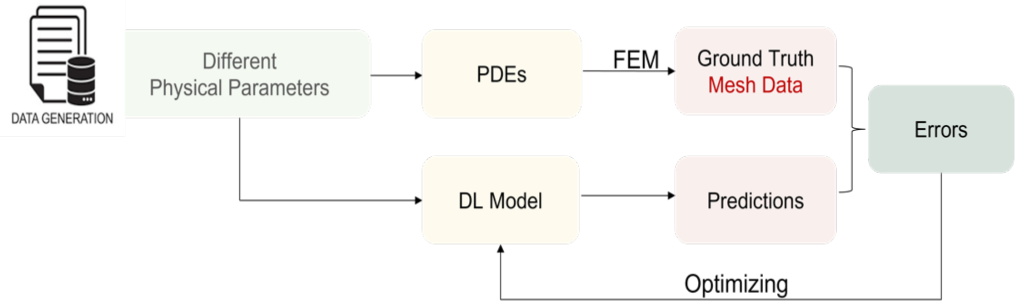
HSA Project : Simulation/machine learning hybrid modeling

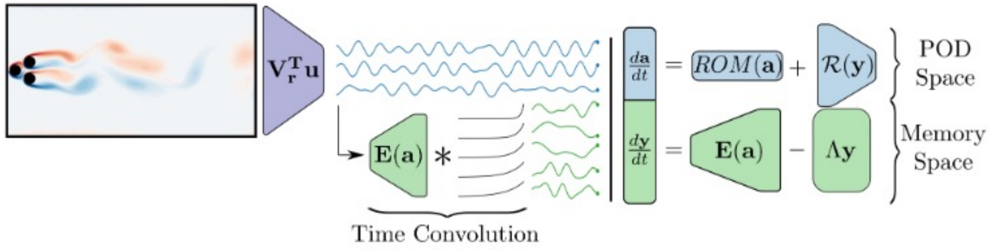
IA² Intelligence artificielle et ingénierie augmentée

Augmenting physical solvers with data-driven models that integrate physical constraints

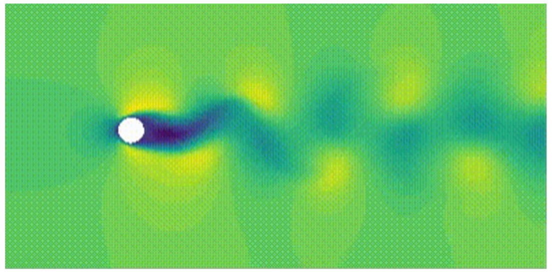
- Hybrid Machine Learninrg as surrogate model for physical simulation
- Questionings: how to deal with high-dimensional, non-linear, and complex structures in theses systems (e.g reduced modeling, ..)

https://www.irt-systemx.fr/projets/HSA/

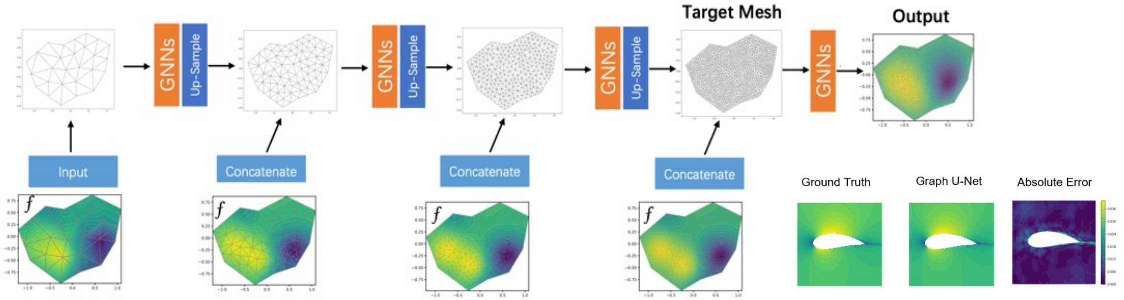




High-Dimensional non-linear Physical Equations



Graph Neural Nets for 3D **meshes**
=> More suitable, as they operate by construction on graphs



Deep Graph Neural Networks for Numerical Simulation of PDEs
PhD thesis de W. Liu. 2023 (LISN, Inria/SystemX). Read Online

Reduced models and deep learning for PDEs
PhD Thesis of E. Menier (in progress) (LISN, Inria/SystemX)

E. Menier et al., 2023. CD-ROM: Complementary Deep-Reduced Order Model. *Computer Methods in Applied Mechanics and Engineering* 410. Read Online

Confiance.ai

www.confiance.ai

A French unique community to design and industrialise trustworthy AI-based critical systems

Key figures

4 YEARS
Duration

9
Large industrial groups

4
Research centers

7
Thematic projects

+300
FTP involved over 4 years

2
Sites: Paris-Saclay and Toulouse

45 M€
Budget

12
Associate partners
(laboratories, SMIs, startups)

Thank you for your attention!