Simulation-based Bayesian inference for high dimensional inverse problems: Application to magnetic resonance fingerprinting

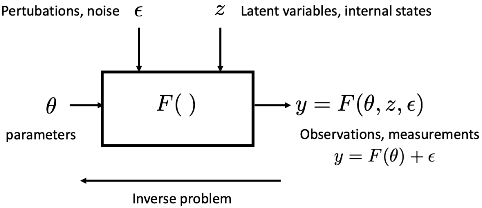
Florence Forbes

Statify team, Inria Grenoble Rhône-Alpes
https://team.inria.fr/statify/
florence.forbes@inria.fr

June 2022

A methodological framework for inverse problems

Common situation: Recover the causes from the observation of the effects



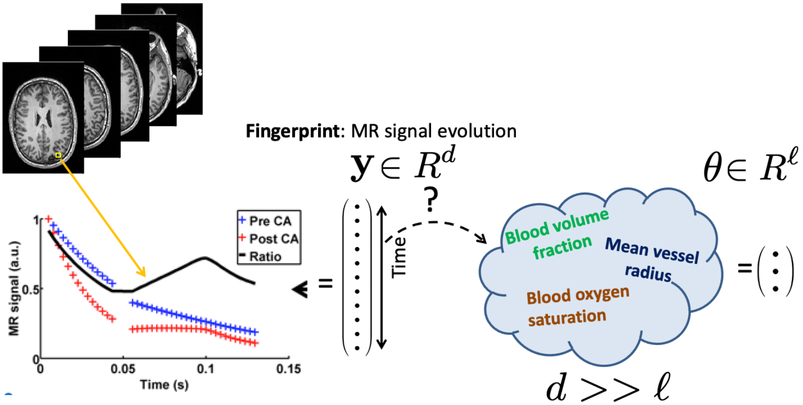
F: analytical formula, mechanistic model, simulator, black box

Inversion: F^{-1} ? instability, non-unicity or existence of the solution
$\longrightarrow$ Ill-posed problem

Example:

- Retrieving brain microvascular properties ($\boldsymbol{\theta}$) from MRI ($\mathbf{y}$)

Retrieving brain microvascular properties from MRI



[Lemasson et al 2016]

One signal per voxel to invert, $d = 32$, $\ell = 3$

A data generating model

- Prior: $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$
- Likelihood: $f_{\boldsymbol{\theta}}(\mathbf{z})$

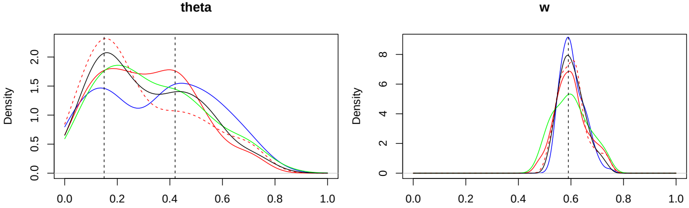
$\longrightarrow \mathbf{z} = \{z_1, \ldots, z_d\}$ can be simulated from $f_{\boldsymbol{\theta}}$

Goal: Estimation of $\boldsymbol{\theta}$ given some observed $\mathbf{y} = \{y_1, \ldots, y_d\}$

Posterior: $\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\mathbf{y})$

Prediction: $\hat{\boldsymbol{\theta}}(\mathbf{y}) = \arg\max_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta}|\mathbf{y})$ or $\hat{\boldsymbol{\theta}}(\mathbf{y}) = \mathbb{E}[\boldsymbol{\Theta}|\mathbf{y}]$

$+$ uncertainty quantification (eg. variance, multiple modes)



Problem characteristics (commonly encountered in applications)

- Modelisation of the phenomenon (F), or direct/forward model
  - Theoretical F: imperfect, approximation $\Rightarrow \mathbf{y} = F(\boldsymbol{\theta}) + \epsilon$ e.g. $f_{\boldsymbol{\theta}} = \mathcal{N}(F(\boldsymbol{\theta}), \boldsymbol{\Sigma}_\epsilon)$ (noise known)
  - Complex F: $\pi(\boldsymbol{\theta}|\mathbf{y})$ known up to a normalising constant. Ok to compute a max, not ok for uncertainty assessment: use MCMC, Variational inference
  - No formula for F: no formula for $f_{\boldsymbol{\theta}}$ or too costly to evaluate, just a simulator (deterministic or random). Likelihood only known implicitly via simulations
  - Evaluation or simulation is costly: limit the number of required simulations. Can we simulate online?
- High dimension of $\mathbf{y}$, of $\boldsymbol{\theta}$
- Recover $\boldsymbol{\theta}$ and its reliability, multiple solutions. Bayesian posterior? Or point estimate sufficient?
- Massive inversion, scalability: invert a large number of $\mathbf{y}$, repeated inference
- Parameter domain knowledge: choice of prior, summary stats available or not?

Existing likelihood-free or "simulation-based" approaches

- Inverse regression models (*e.g.* SIR, GLLiM, invertible Neural Nets, Normalizing Flows)
  - Pros: high dimensional $\mathbf{y}$ ok, massive inversion (amortization) ok
  - Cons: learning cost vs accuracy ? Theorerical garanties?
- Approximate Bayesian Computation (ABC)
- Bayesian Synthetic Likelihood (BSL)

Remark: Optimization, Markov Chain Monte Carlo (MCMC) ... , not "simulation-based", require F or the likelihood

Inverse regression: dealing with high dimensional data

Inverse problem: Find f (F^{-1}) between $\mathbf{y} \in \mathbb{R}^d$ and $\boldsymbol{\theta} \in \mathbb{R}^\ell$ with $d >> \ell$

$$f: \quad \mathbf{y} \in \mathbb{R}^d \quad \longrightarrow \quad \boldsymbol{\theta} \in \mathbb{R}^\ell \quad \text{(high-to-low)}$$

Regression from $\mathcal{D}_M = \{(\mathbf{z}_m, \boldsymbol{\theta}_m), m = 1:M\}$: d **large** $\Longrightarrow$ **curse of dimensionality**

- **Inverse regression:** learn the low-to-high mapping (F)

Approximate F by $\hat{F}_M$ by learning a regression model on $\mathcal{D}_M$ (put the effort here, offline)

Deduce F^{-1} as $\hat{F}_M^{-1}$ (use an invertible model in the first place) instead of learning F^{-1} directly

- Set $\hat{\boldsymbol{\theta}}(\mathbf{y}) = \hat{F}_M^{-1}(\mathbf{y})$ (straightforward, $\hat{\boldsymbol{\theta}}(\mathbf{y})$ may not be in $\mathcal{D}_M$, no distance computation)

- Add uncertainty modelling by learning the likelihood/posterior instead of F/F^{-1}

GLLiM [Deleforge et al 2015a]: $f_{\boldsymbol{\theta}}$ and $\pi(\boldsymbol{\theta}|\mathbf{y})$ are approximated by mixtures of Gaussians, efficient for both sampling and likelihood evaluation.

Gaussian Locally-linear Mapping (GLLiM)

[Deleforge, Forbes & Horaud, 2015a]

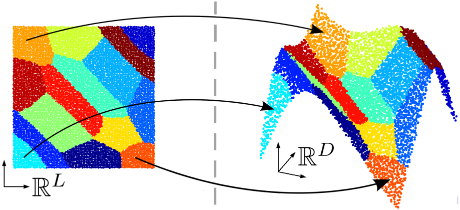
- $\boldsymbol{\Theta} \in \mathcal{X} \subset \mathbb{R}^\ell$ low-dimensional space, $\boldsymbol{Y} \in \mathcal{Y} \subset \mathbb{R}^d$ high-dimensional space,
- A piecewise affine model: Introduce a missing variable Z

$$p(\mathbf{y}, \boldsymbol{\theta}; \boldsymbol{\phi}) = \sum_{k=1}^{K} p(\mathbf{y}|\boldsymbol{\theta}, Z = k; \boldsymbol{\phi})\; p(\boldsymbol{\theta}|Z = k; \boldsymbol{\phi})\; p(Z = k; \boldsymbol{\phi})$$

$Z = k \Leftrightarrow \boldsymbol{Y}$ is the image of $\boldsymbol{\Theta}$ by an affine transformation τ_k

$$\boldsymbol{Y} = \sum_{k=1}^{K} \mathbb{I}(Z = k)(\mathbf{A}_k \boldsymbol{\Theta} + \boldsymbol{b}_k + \boldsymbol{E}_k)$$

$\mathbb{I}$ Indicator function, $\mathbf{A}_k$ $d \times \ell$ matrix, $\boldsymbol{b}_k$ d-dim vector
$\boldsymbol{E}_k$: observation noise in $\mathbb{R}^d$ and reconstruction error, centered, independent on $\boldsymbol{\Theta}$, $\boldsymbol{Y}$, and Z



Hierarchical definition

$\boldsymbol{E}_k$: centered Gaussian independent on $\boldsymbol{\Theta}$, $\boldsymbol{Y}$, and Z

$$p(\mathbf{y}|\boldsymbol{\theta}, Z = k; \boldsymbol{\phi}) = \mathcal{N}_d(\mathbf{y}; \mathbf{A}_k\boldsymbol{\theta} + \boldsymbol{b}_k, \boldsymbol{\Sigma}_k)$$

- Affine transformations are local: mixture of K Gaussians

$$\begin{aligned} p(\boldsymbol{\theta}|Z = k; \boldsymbol{\phi}) &= \mathcal{N}_\ell(\boldsymbol{\theta}; \boldsymbol{c}_k, \boldsymbol{\Gamma}_k) \\ p(Z = k; \boldsymbol{\phi}) &= \pi_k \end{aligned}$$

- The set of all model parameters is:

$$\boldsymbol{\phi} = \{\boldsymbol{c}_k, \boldsymbol{\Gamma}_k, \pi_k, \mathbf{A}_k, \boldsymbol{b}_k, \boldsymbol{\Sigma}_k, k = 1 : K\}$$

e.g. Mixtures of Linear Experts when $\boldsymbol{\Sigma}_k$ is diagonal [Xu et al 1995]

Regression functions

Both conditional densities are Gaussian mixtures parameterized by ϕ (or ϕ^* easily deduced from ϕ), ϕ is estimated using an EM algorithm.

$$p_G(\mathbf{y}|\boldsymbol{\theta};\boldsymbol{\phi}) = \sum_{k=1}^{K} \underbrace{\frac{\pi_k \mathcal{N}_\ell(\boldsymbol{\theta};\boldsymbol{c}_k,\boldsymbol{\Gamma}_k)}{\sum_{j=1}^{K}\pi_j \mathcal{N}_\ell(\boldsymbol{\theta};\boldsymbol{c}_j,\boldsymbol{\Gamma}_j)}}_{\eta_k(\boldsymbol{\theta})} \mathcal{N}_d(\mathbf{y};\underbrace{\mathbf{A}_k\boldsymbol{\theta}+\boldsymbol{b}_k}_{\boldsymbol{\mu}_k},\boldsymbol{\Sigma}_k)$$

$$p_G(\boldsymbol{\theta}|\mathbf{y};\boldsymbol{\phi}^*) = \sum_{k=1}^{K} \underbrace{\frac{\pi_k^* \mathcal{N}_d(\mathbf{y};\boldsymbol{c}_k^*,\boldsymbol{\Gamma}_k^*)}{\sum_{j=1}^{K}\pi_j^* \mathcal{N}_d(\mathbf{y};\boldsymbol{c}_j^*,\boldsymbol{\Gamma}_j^*)}}_{\eta_k^*(\mathbf{y})} \mathcal{N}_\ell(\boldsymbol{\theta};\underbrace{\mathbf{A}_k^*\mathbf{y}+\boldsymbol{b}_k^*}_{\boldsymbol{\mu}_k^*},\boldsymbol{\Sigma}_k^*)$$

$$\begin{aligned}
\boldsymbol{c}_k^* &= \mathbf{A}_k\boldsymbol{c}_k + \mathbf{b}_k \\
\boldsymbol{\Gamma}_k^* &= \boldsymbol{\Sigma}_k + \mathbf{A}_k\boldsymbol{\Gamma}_k\mathbf{A}_k^\top \\
\pi_k^* &= \pi_k \\
\mathbf{A}_k^* &= \boldsymbol{\Sigma}_k^*\mathbf{A}_k^\top\boldsymbol{\Sigma}_k^{-1} \\
\mathbf{b}_k^* &= \boldsymbol{\Sigma}_k^*(\boldsymbol{\Gamma}_k^{-1}\boldsymbol{c}_k - \mathbf{A}_k^\top\boldsymbol{\Sigma}_k^{-1}\boldsymbol{b}_k) \\
\boldsymbol{\Sigma}_k^* &= (\boldsymbol{\Gamma}_k^{-1} + \mathbf{A}_k^\top\boldsymbol{\Sigma}_k^{-1}\mathbf{A}_k)^{-1}
\end{aligned}$$

Regression functions straightforward via: $\mathbb{E}_G[\boldsymbol{Y}|\boldsymbol{\theta};\boldsymbol{\phi}]$ and $\mathbb{E}_G[\boldsymbol{\Theta}|\mathbf{y};\boldsymbol{\phi}^*]$

Illustration

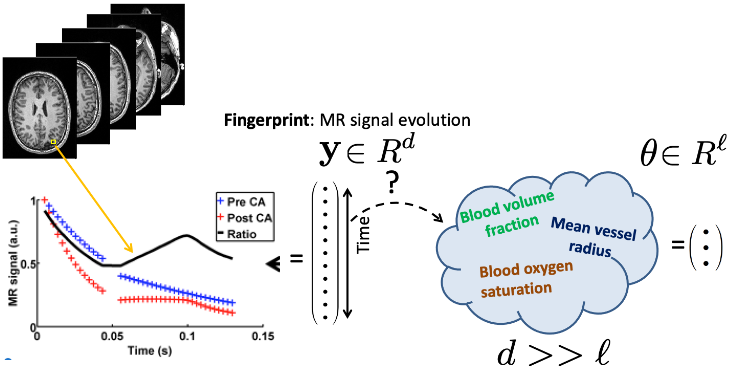
- I. Use GLLiM mixture as an approximation of the posterior
- II. Refine GLLiM approximation with importance sampling
- III. Leverage GLLiM approximation with ABC or BSL

I. Use GLLiM to approximate the posterior distributions

- Simulate/Design a dictionary or learning set: $\mathcal{D}_M = \{(\boldsymbol{\theta}_m, \mathbf{z}_m), m=1:M\}$
- Estimate GLLiM parameters (K Gaussians) $\boldsymbol{\phi}_{K,M}$ from $\mathcal{D}_M$ using EM and deduce $\boldsymbol{\phi}^*_{K,M}$
- Approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$ by $p_G(\boldsymbol{\theta}|\mathbf{y};\boldsymbol{\phi}^*_{K,M}) = \sum_{k=1}^K \eta_k^*(\mathbf{y})\, \mathcal{N}_\ell(\boldsymbol{\theta}; \mathbf{A}_k^*\mathbf{y} + \mathbf{b}_k^*, \boldsymbol{\Sigma}_k^*)$
- Predict parameters as $E_G[\boldsymbol{\Theta}|\mathbf{y};\boldsymbol{\phi}^*_{K,M}] = \sum_{k=1}^K \eta_k^*(\mathbf{y})(\mathbf{A}_k^*\mathbf{y} + \mathbf{b}_k^*)$
- Propose confidence index via $Var_G[\boldsymbol{\Theta}|\mathbf{y};\boldsymbol{\phi}^*_{K,M}]$

Retrieving brain microvascular properties from MRI using MR fingerprinting

[Boux, Forbes, Arbel, Lemasson & Barbier, 2021]

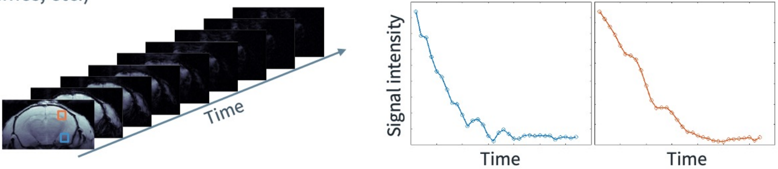


[Lemasson et al 2016]

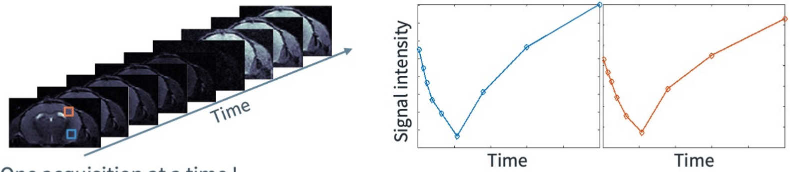
- High dimensional observations
- Massive inversion, amortization required, MCMC not possible
- Time and memory costly
- Exploitation of the approximated posterior pdf for a confidence index

Quantitative MRI: Standard MR signal **acquisition**

- Acquiring images in time according to a specific sequence design (*i.e.* RF pulses, times, etc.)



- Performing other acquisitions with other designs for an other quantity



One acquisition at a time !

Standard biophysical parameter **quantification**

Closed-form expression fitting (CEF): a non-linear least squares solver is used to fit the signals with a biophysical model

- T_2 decay:

$$S_{T_2}(t) = C_{T_2} \exp\left(-\frac{t}{T_2}\right)$$

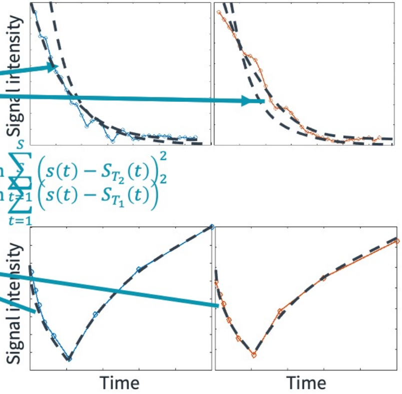
- T_2^* decay:

$$S_{T_2^*}(t) = C_{T_2^*} \exp\left(-\frac{t}{T_2^*}\right)$$

- T_1 decay:

$$S_{T_1}(t) = C_{T_1} \left(1 - \exp\left(-\frac{t}{T_1}\right)\right)$$

- ...

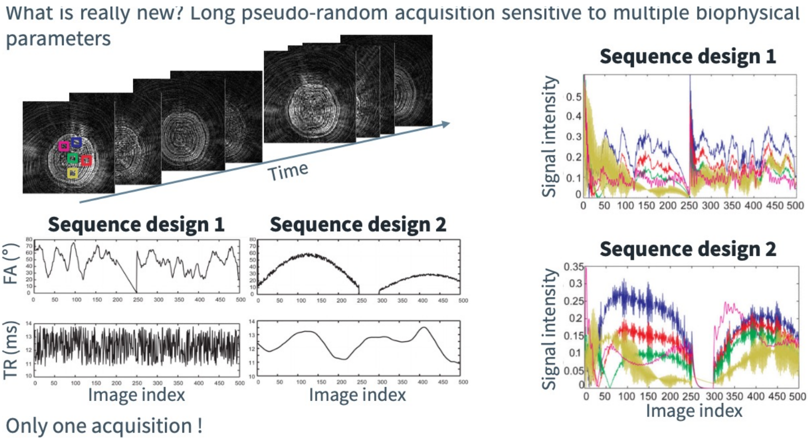


$$\arg\min_{(T_2, C_{T_2})} \sum_{t=1}^{S} \left(s(t) - S_{T_2}(t)\right)^2$$

$$\arg\min_{(T_1, C_{T_1})} \sum_{t=1}^{S} \left(s(t) - S_{T_1}(t)\right)^2$$

One parameter at a time !

MR fingerprinting: MR signal (or fingerprint) acquisition in MRF

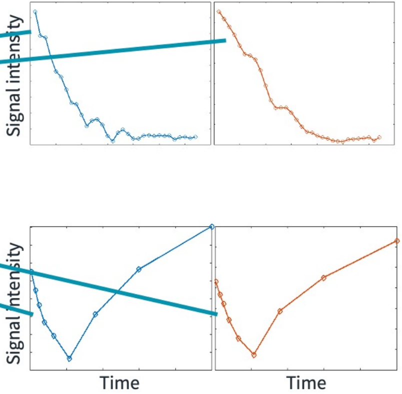


[Ma *et al.*, *Magnetic resonance fingerprinting*, Nature, 2013]

Standard biophysical parameter quantification in MRF

Dictionary-based matching (DBM)*: a set of simulated signals (using a biophysical model) are used to fit the signal

- T_2 decay: $\arg\min_{S_{T_2}} d(s, S_{T_2})$
- T_2^* decay:
- T_1 decay: $\arg\min_{S_{T_1}} d(s, S_{T_1})$
- ...

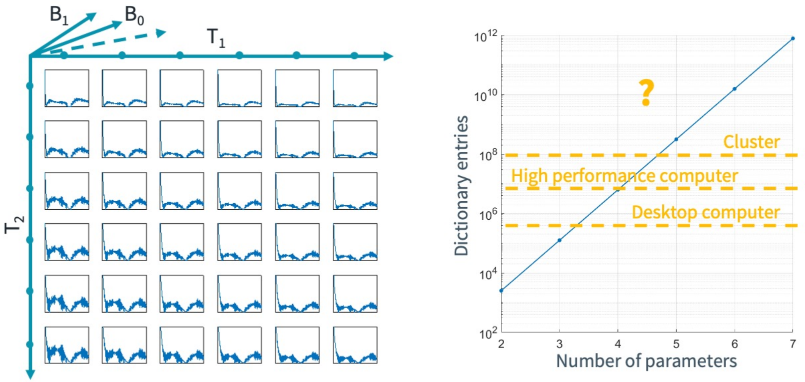


Still one parameter at a time !

*[Ma *et al.*, *Magnetic resonance fingerprinting*, Nature, 2013]

Multi-parametric quantification in MRF

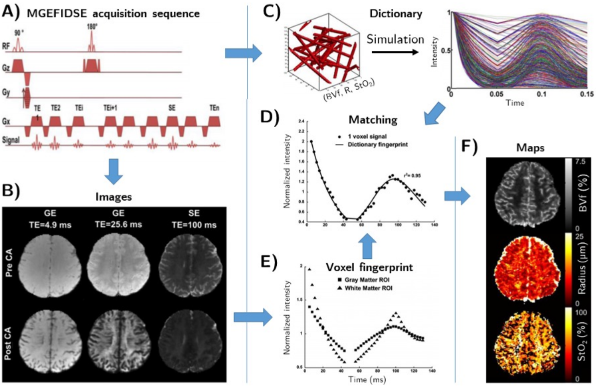
DBM quantification allows multi-parameter quantification from a single fast acquisition



[Ma *et al.*, *Magnetic resonance fingerprinting*, Nature, 2013]

Vascular MRF

Fingerprint: ratio between the MRI signals before and after contrast agent (CA)



[Christen *et al.*, *MR vascular fingerprinting: A new approach to compute cerebral blood volume, mean vessel radius, and oxygenation maps in the human brain*, Neuroimage, 2014]

Simulation tool

Vascular parameters
(*parameters of interest*)

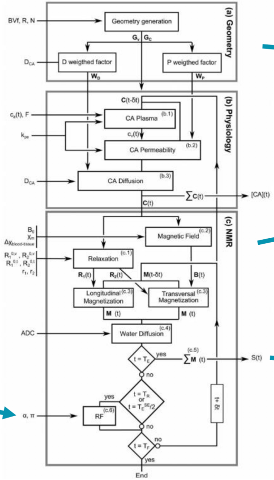
In vascular MRF:

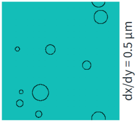
- Blood volume fraction (BVf)
- Vessel size index (VSI)
- Tissue oxygen saturation (StO_2)

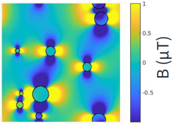
Various parameters
(*fixed in our experiments*)

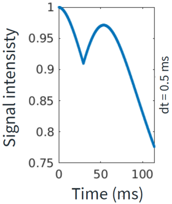
Sequence parameters
(*flip angles, repetition times*)

INPUTS OUTPUTS









Simulation cost

Limitation: simulation time

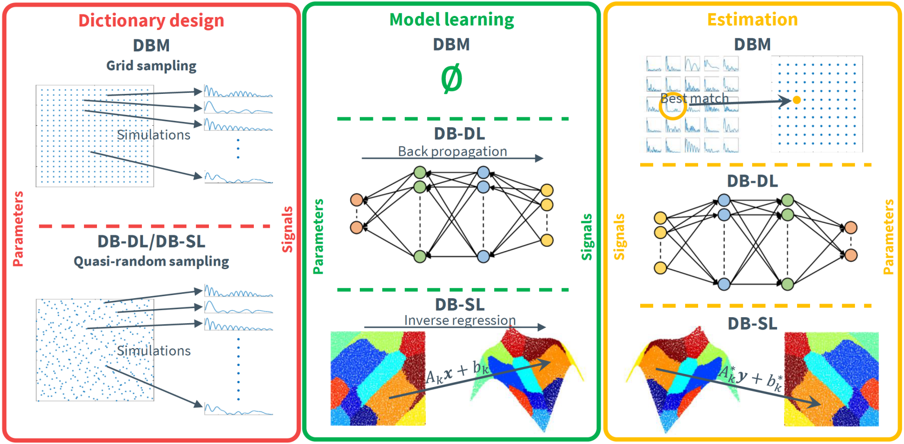
	Simulation tool	Ressources	Time	
Lemasson *et al.**	Previous tool	30-node cluster	10^6 signals ≈ 24 hours	**2.8 sec/signal**
Pouliot *et al.*°	Monte Carlo simulations / realistic angiograms	15-node cluster	4.10^5 signals ≈ 70 hours	**10 sec/signal**
This work	Previous tool	32-core computer	10^5 signals ≈ 67 hours	**2.4 sec/signal**
Christen (ANR MRFUSE)	3D previous tool / realistic angiograms	32-core computer	500 signals ≈ 24 hour	**3 min/signal**

These works suggest that vascular MRF is more accurate than standard vascular approach

How to reduce the need for simulations in vascular MRF?

*[Lemasson *et al.*, *MR vascular fingerprinting in stroke and brain tumors models*, Scientific reports, 2016]
°[Pouliot *et al.*, *Magnetic resonance fingerprinting based on realistic vasculature in mice*, Neuroimage, 2017]

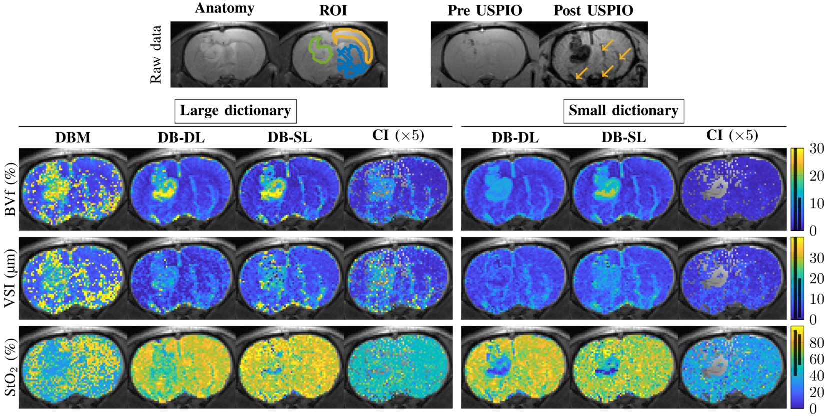
Analysis framework: comparison



DB-DL: [Cohen et al , MR fingerprinting deep reconstruction network (DRONE), Magnetic Resonance in Medicine 2018]

Results: real vascular MRF data

MRI data (4.7T) acquired on a rat bearing a 9L tumor ($M = 167\ 216$ and $M = 4119$)



- bypasses the standard time and memory requirement (Dictionary matching)
- adresses the issue of simulation time
- challenges DL-based quantification (small $\mathcal{D}_M$)
- provides confidence maps for each parameter

Conclusion and Perspectives

A tractable approach to Bayesian inverse problems, that allows "exploration" of the posterior distribution

MR Fingerprinting : a new approach to quantification that

- bypasses the inherent issues in standard MRF (time and memory requirement)
- addresses the problem of simulation time in vascular MRF
- challenges a DL quantification
- provides a confidence map associated to each parameter

Perspectives

- Real standard MRF acquisitions and complex-valued data
- Spatial considerations (neighboring fingerprints)
- Sequential learning easy with GLLiM
- Sequential learning/Amortization compromise
- Other learning scheme than GLLiM (Mixture density networks, Invertible NN, Normalizing flows) [Lueckmann et al 2021]

Thank you for your attention !

-F. Boux, F. Forbes, J. Arbel, B. Lemasson, E. Barbier (2021). Bayesian inverse regression for vascular magnetic resonance fingerprinting. IEEE Trans. on Medical Imaging.

-Deleforge, A., Forbes, F. & Horaud, R. (2015)a. High-Dimensional **Regression with Gaussian Mixtures** and Partially-Latent Response Variables. Statistics & Computing.

-Deleforge, A., Forbes, F. & Horaud, R. (2015)b. Acoustic Space Learning for Sound-Source Separation and Localization on Binaural Manifolds. Int. J. Neural Systems.

-Deleforge, A., Forbes, F., Ba, S., Horaud, R. (2015)c. Hyper-Spectral Image Analysis with Partially-Latent Regression and Spatial **Markov Dependencies**. IEEE Journal of Selected Topics in Signal Processing.

-F. Forbes, H. Nguyen, T. Nguyen, J. Arbel (2021). **ABC with surrogate posteriors**. Available on Hal

-B. Kugler, F. Forbes, S. Douté (2021). **Fast Bayesian Inversion** for high dimensional inverse problems. To appear in Statistics & Computing.

-E. Perthame, F. Forbes, A. Deleforge (2018). Inverse regression approach to **robust** nonlinear high-to-low dimensional mapping. Journal of Multivariate Analysis.

-Tu, C.C., Forbes, F., Lemasson, B., Wang, N. (2019), Prediction with high dimensional regression via **hierarchically structured Gaussian mixtures** and latent variables. Journal of the Royal Statistical Society: Series C Applied Statistics.

-B. Lemasson, N. Pannetier, N. Coquery, L. S. Boisserand, N. Collomb, N. Schuff, M. Moseley, G. Zaharchuk, E. L. Barbier, and T. Christen (2016), MR vascular fingerprinting in stroke and brain tumors models, Scientific Reports.

-JM Lueckmann, J Boelts, DS Greenberg, PJ Gonçalves, JH Macke (2021), Benchmarking Simulation-Based Inference, Proceedings of The 24th International Conference on Artificial Intelligence and statistics.

- Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J. L., Duerk, J. L., and Griswold, M. A. (2013). Magnetic Resonance Fingerprinting. Nature, 495(7440):187–192.

CODE

- GLLiM, SLLiM and co: xLLiM R package on the CRAN (also GLLiM in Python, Julia)
  - Perthame, E., Forbes, F., Deleforge, A., Devijver, E., and Gallopin, M. (2017). xLLiM: High Dimensional Locally-Linear Mapping. R package version 2.1.
- MR fingerprinting (Matlab): https://github.com/nifm-gin/MP3

Appendix

Low-to-High or High-to-Low?

If ϕ is unconstrained $\boldsymbol{\theta}$-to-$\mathbf{y}$ or $\mathbf{y}$-to-$\boldsymbol{\theta}$ estimations are equivalent and intractable for large d

Inversion trick: impose a structure on ϕ (ϕ is constrained), *e.g.* $\forall k, \boldsymbol{\Sigma}_k = \sigma^2 \mathbf{I}_d$

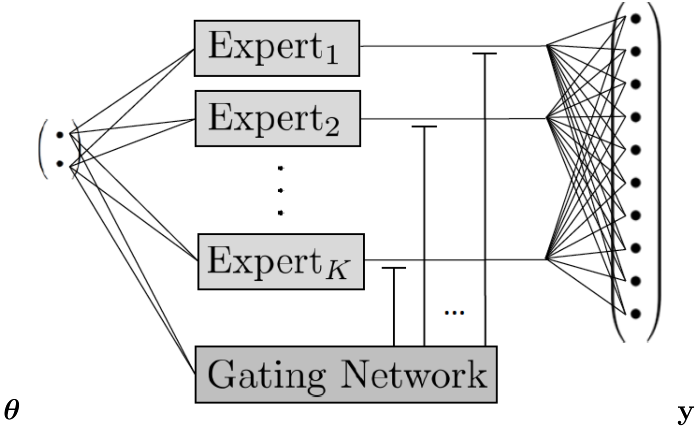
Example: $d = 1000$, $\ell = 2$, $K = 10$

- Low-to-high regression: $K(1 + \ell + d\ell + \ell(\ell+1)/2 + d) = 30,060$ parameters.
- High-to-low regression: $K(1 + d + \ell d + d(d+1)/2 + \ell) =$ **5,035,030** parameters. Requires inversion of 1000×1000 covariance matrices.

Therefore it is better to perform a low-dimensional-to-high-dimensional (forward) regression (ϕ), and then deduce the inverse density (ϕ^*)

$$
\begin{aligned}
\boldsymbol{c}_k^* &= \mathbf{A}_k \boldsymbol{c}_k + \mathbf{b}_k \\
\boldsymbol{\Gamma}_k^* &= \boldsymbol{\Sigma}_k + \mathbf{A}_k \boldsymbol{\Gamma}_k \mathbf{A}_k^\top \\
\pi_k^* &= \pi_k \\
\mathbf{A}_k^* &= \boldsymbol{\Sigma}_k^* \mathbf{A}_k^\top \boldsymbol{\Sigma}_k^{-1} \\
\mathbf{b}_k^* &= \boldsymbol{\Sigma}_k^* (\boldsymbol{\Gamma}_k^{-1} \boldsymbol{c}_k - \mathbf{A}_k^\top \boldsymbol{\Sigma}_k^{-1} \boldsymbol{b}_k) \\
\boldsymbol{\Sigma}_k^* &= (\boldsymbol{\Gamma}_k^{-1} + \mathbf{A}_k^\top \boldsymbol{\Sigma}_k^{-1} \mathbf{A}_k)^{-1}
\end{aligned}
$$

GLLiM includes Mixtures of localized linear experts



Expert k : $p(\mathbf{y}|\boldsymbol{\theta}, Z = k)$, Gating network: $p(Z = k|\boldsymbol{\theta})$

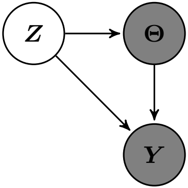
Output: $p(\mathbf{y}|\boldsymbol{\theta}) = \sum_{k=1}^{K} p(\mathbf{y}|\boldsymbol{\theta}, Z = k)\; p(Z = k|\boldsymbol{\theta})$

Extension to partially observed responses

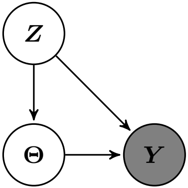
Incorporate a latent component into the low-dimensional variable:

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{T} \\ \boldsymbol{W} \end{bmatrix}$$

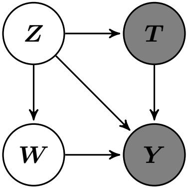
where $\boldsymbol{T} \in \mathbb{R}^{\ell_t}$ is observed and $\boldsymbol{W} \in \mathbb{R}^{\ell_w}$ is latent ($\ell = \ell_t + \ell_w$)



Supervised GLLiM (regression)



Unsupervised GLLiM (dim reduction)



Hybrid GLLiM

Remark: Hybrid GLLiM is supervised GLLiM with $\boldsymbol{\Sigma}'_k = \boldsymbol{\Sigma}_k + \mathbf{A}^w_k \boldsymbol{\Gamma}^w_k \mathbf{A}^{wT}_k$ (factor model, *e.g.* diagonal + low rank matrix)

Particular instances of the hybrid GLLiM model

First three rows: supervised GLLiM methods ($\ell_\mathrm{w} = 0$)

Last six rows: unsupervised GLLiM methods ($\ell_\mathrm{t} = 0$)

Model	$\boldsymbol{c}_k$	$\boldsymbol{\Gamma}_k$	π_k	$\mathbf{A}_k$	$\boldsymbol{b}_k$	$\boldsymbol{\Sigma}_k$	ℓ_t	ℓ_w	K
MLE [Xu et al 95]	-	-	-	-	-	diag	-	0	-
MLR [Jedidi et al 96]	$\mathbf{0}_L$	$\infty\mathbf{I}_L$	-	-	-	iso+eq	-	0	-
JGMM [Qiao et al 09]	-	-	-	-	-	-	-	0	-
PPAM [Deleforge et al 12]	-	|eq|	eq	-	-	diag+eq	-	0	-
GTM [Bishop et al 98]	fixed	$\mathbf{0}_L$	eq.	eq.	$\mathbf{0}_D$	iso+eq	0	-	-
PPCA [Tipping et al 99a]	$\mathbf{0}_L$	$\mathbf{I}_L$	-	-	-	iso	0	-	1
MPPCA [Tipping et al 99b]	$\mathbf{0}_L$	$\mathbf{I}_L$	-	-	-	iso	0	-	-
MFA [Ghahramani et al 96]	$\mathbf{0}_L$	$\mathbf{I}_L$	-	-	-	diag	0	-	-
PCCA [Bach et al 05]	$\mathbf{0}_L$	$\mathbf{I}_L$	-	-	-	block	0	-	1
RCA [Kalaitzis et al 11]	$\mathbf{0}_L$	$\mathbf{I}_L$	-	-	-	fixed	0	-	1

Other extensions

- SLLiM: Student mixtures for more robustness: [Perthame et al 2018]
- BLLiM, Structured GLLiM, for specific covariance structures: [Tu et al 2019]
- GLLiM iid: for *i.i.d.* observations: [Forbes et al 2021]
- Markovian GLLiM, for dependent observations: [Deleforge et al 2015c]
- Missing observations: [Deleforge et al 2015b]