Multi-Channel Variational Inference for joint analysis of multi-modal imaging and clinical data: application to Alzheimer's Disease

Luigi Antelmi[1], Nicholas Ayache[1] ,Philippe Robert[2,3], Valeria Manera[3,4], Marco Lorenzi[1]

1) UCA, INRIA Sophia-Antipolis, Epione Research Project. 2) CoBTeK, University of Nice Sophia Antipolis.
3) Centre Memoire, CHU de Nice. 4) UCA, INRIA Sophia-Antipolis, Stars Research Project.

The Context

Challenge: the **joint analysis** of biomedical high-dimensional heterogeneous data (structural/functional imaging, biological, clinical) in Alzheimer's Disease **is not straightforward**, and at the same time important for clinical diagnosis and follow up.

Hypothesis: a multidimensional unobserved **latent variable** is the **common cause** of the heterogeneous data observations. The causation happens through independent information routes we call **channels** (e.g. MRI channel, PET channel, etc.)

Goal: estimate the distribution of the latent variable and the deterministic functions that connects it with the observations.

Why Generative Models

Generative Models allow the estimation of the **joint** distribution of: 1) observed data and 2) hidden latent variables that generated them. In the case of C **channels** the causation process takes the form:

$$\begin{aligned} &\text{latent variable: } \mathbf{z} \sim p(\mathbf{z}) \\ &\text{observed channels: } \mathbf{x}_c \sim p(\mathbf{x}_c|\mathbf{z}) \text{ for } c = 1..C \end{aligned}$$

Computing the exact posterior distribution $p(\mathbf{z}|\mathbf{x}_1, \ldots, \mathbf{x}_C)$, to **infer** the latent variable from the observed channels, is not always possible. We look then for an **approximate distribution** to the exact one using the ***Variational*** approach (fast convergence; scalability to high dim. data).

Multi-Channel Variational Model

With the variational approach, is possible to construct C approximate distributions, one for each channel, as close as possible, on average, to the exact posterior one, by solving the following optimization problem:

$$\underset{q \in \mathcal{Q}}{arg\,min}\ \mathbb{E}_c \left[\mathcal{D}_{KL}\left(q\left(\mathbf{z}|\mathbf{x}_c\right) || \, p\left(\mathbf{z}|\mathbf{x}_1, \ldots, \mathbf{x}_C\right) \right) \right]$$

that is equivalent to maximize the data **evidence** through the **lower bound**:

$$\underbrace{\ln p\left(\mathbf{x}_1, \ldots, \mathbf{x}_C\right)}_{\text{Evidence}} \geq \underbrace{\frac{1}{C} \sum_{c=1}^{C} \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)} \left[\sum_{i=1}^{C} \ln p(\mathbf{x}_i|\mathbf{z}) \right] - \mathcal{D}_{KL}\left(q\left(\mathbf{z}|\mathbf{x}_c\right) || p\left(\mathbf{z}\right) \right)}_{\text{Lower Bound}}$$

In the lower bound, the distributions $q(\mathbf{z}|\mathbf{x}_c)$ and $p(\mathbf{x}_i|\mathbf{z})$ are parametrized by functions that can be linear or non-linear.

These functions' coefficients constitute the parameters of the Multi-channel Variational Model.

The dimension of the latent space, dim(**z**), is not generally known *a priori,* but it can be selected as the one that gives the highest lower bound, which means the highest evidence.

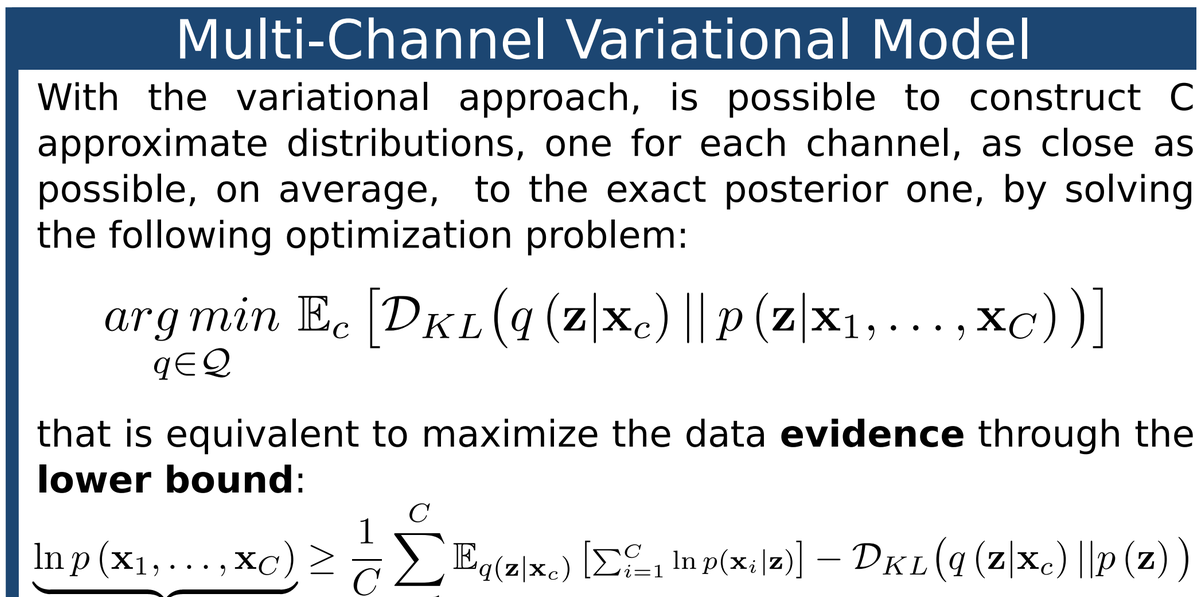


Fig.1: A set of approximate density functions $q\left(\mathbf{z}|\mathbf{x}_c\right)$ one for each channel, are optimized to be, on average, as close as possible to the exact uncomputable posterior $p\left(\mathbf{z}|\mathbf{x}_1, \ldots, \mathbf{x}_C\right)$ Functions f() are used to parametrize $p(\mathbf{x}_c|\mathbf{z})$.

Evidence Lower Bound for Model Selection

Experiment on synthetic data.

Data: multi-channel observations deterministically generated from a latent variable $\zeta \sim \mathcal{N}\left(\mathbf{0}; \mathbf{I}_d\right)$ with d dimensions.

Then we fit multi-channel variational models by choosing the number of latent dimensions in a range containing the true ones.

We found that the best models explaining the data are the ones fitted with a number of latent dimensions that equals the number of the true ones used to generate the observations.

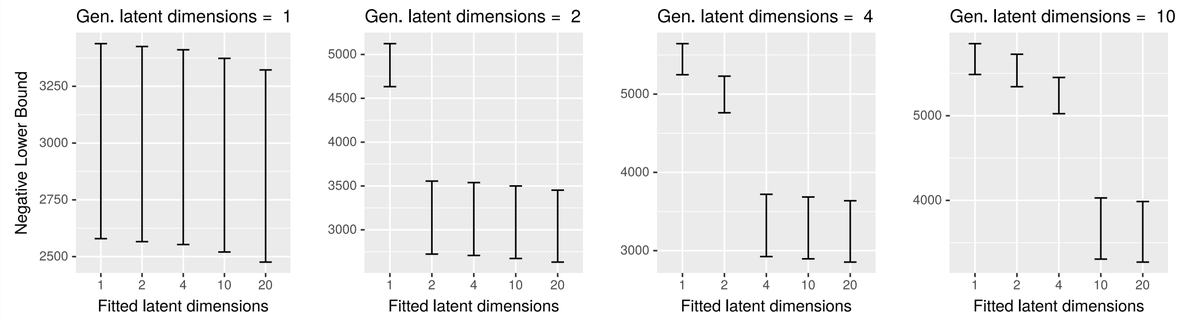


Fig.2: Negative lower bound as a function of generated and flitted latent dimensions. The simplest (=low lat. dims) and more accurate (=lowest cost function) model explaining the data is the one fitted with a number of latent dimensions that equals the number of the true ones used to generate the data.

Model Selection and Inference on ADNI

Experiment on real data: 504 subjects, four channels:

1) Clinical (age, mmse, adas11, cdr-sb, faq, pteducat)
2) MRI (gray matter - AAL atlas parcellation)
3) FDG-PET (Glucose uptake - AAL atlas parcellation)
4) AV45-PET (Amyloid uptake - AAL atlas parcellation)

Model's distributions are parametrized by linear functions.

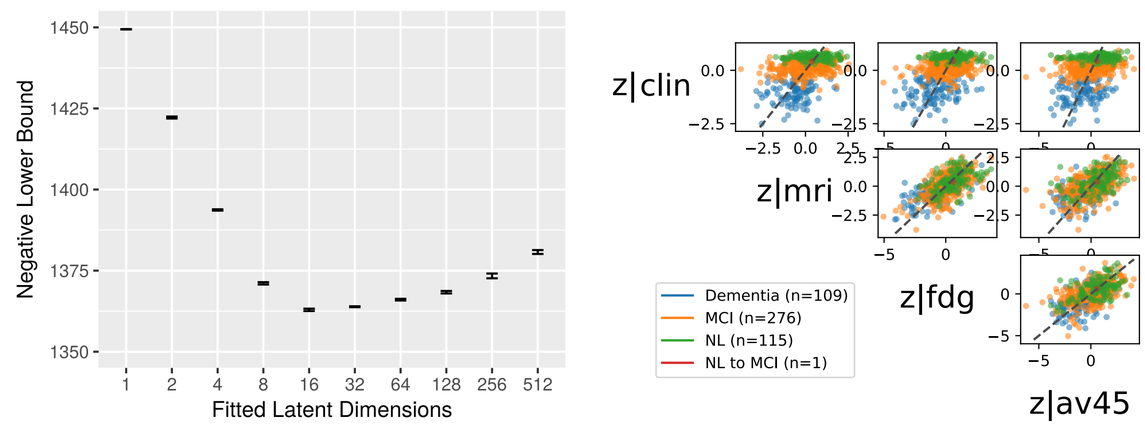


Fig.3: Results. (Left) 16 lat. dims explain the data optimally. (Right) The latent space (only one component shown). Although the optimization is not supervised to enforce clustering, subjects appear stratified by disease status.

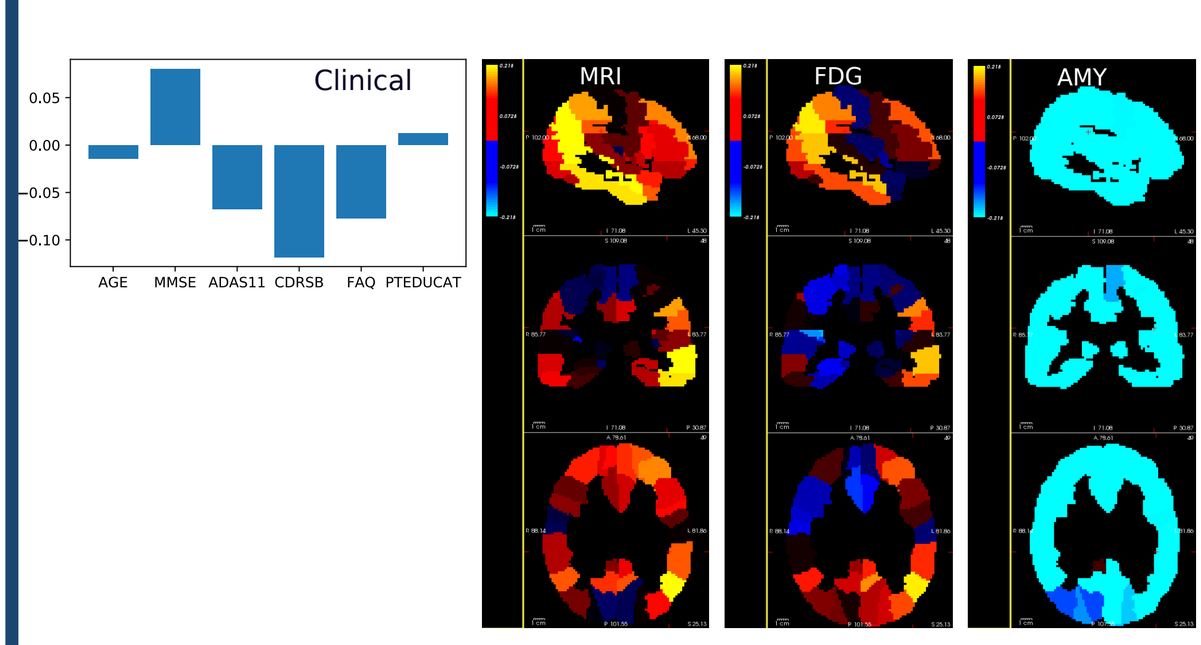


Fig.4: Parameters of the multi-channel variational model. Specifically, these are the coefficients of the (linear) functions f() (cfr. Fig.1) connecting the latent component illustrated in Fig.3 with the multi-dimensional data in the ADNI dataset.

This work has been supported by:
- the French government, through the UCAJEDI Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01;
- the grant AAP Santé 06 2017-260 DGA-DSH, and by the Inria Sophia Antipolis - Méditerranée, "NEF" computation cluster.

Independent Gaussian Process Analysis : Disease staging through extraction of independent patterns of spatio-temporal brain changes

Clément Abi Nader[1], Nicholas Ayache[1], Valeria Manera[2,3], Philippe Robert[3,4], Marco Lorenzi[1]

1 UCA, Inria Sophia Antipolis, Epione Research Project. **2** UCA, Inria Sophia Antipolis, Stars Research Project. **3** CoBTeK, University of Nice Sophia Antipolis. **4** Centre Memoire, CHU de Nice.

Background

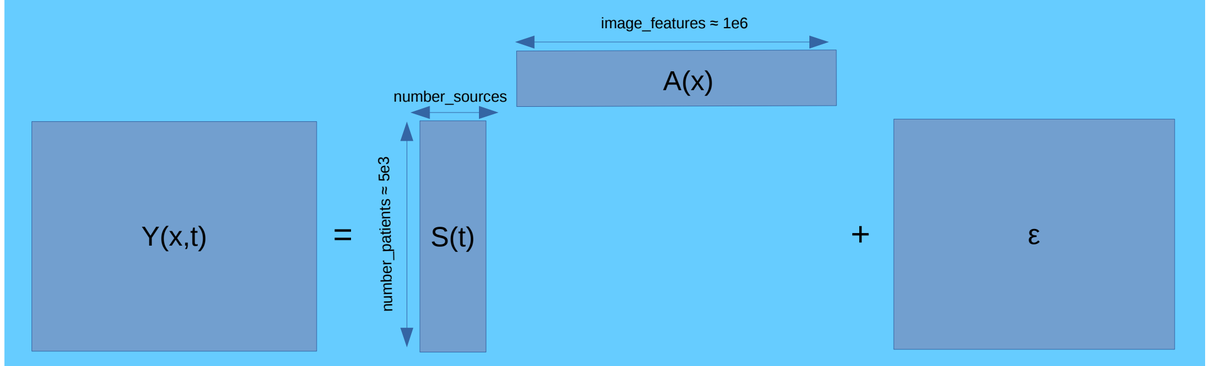
The progression of a biological process (e.g disease, aging, ...) lead to **morphological changes** over time in the brain and affect differently the brain regions. Moreover the timing and the speed of these changes are specific to a subject. To date, we lack of a model allowing to extract these trajectories of brain changes in cohorts of **cross-sectional** brain data, along with individual's position in the trajectory. Therefore, our goal is to decompose **spatio-temporal** brain data, in order to separate different temporal and spatial sources modeling brain progression. But also to automatically estimate an individual **time-shift** that assesses where the subjects lie in these processes.

Method

- Brain changes over time can be characterized by the combination of latent temporal trajectories associated to some relevant structural brain patterns.
- These trajectories are supposed to model realistic biological processes (such as a disease progression), that should exhibit monotonic or constant evolution. Consequently we constrain our model to look only for monotonic temporal sources.
- We introduce a time-shifting parameter for each patient. This is a linear transformation of the time, that estimates a "virtual age" of this patient and positions him on the temporal trajectories.

The data X has dimensions (number_patients p, image_features f), where each patient has an associated age.

Our model separates the data such that :



Where t represents the time-shift parameter, and x the spatial parameters. We also have $\varepsilon_p \sim N(0, \sigma^2 I)$ for each patient

- Each column of S is treated as an **independent Gaussian Process** to promote smoothness in time and model a plausible biological progression. Moreover we constrain these processes to be monotonic.
- The algorithm also learns a per-patient parameter t_p. This allows the model to find the best ordering of the subjects that explain the data and gives a **"virtual age"** to each subject.
- The spatial sources are modeled as **Gaussian random fields** to encode the spatial continuity of the brain sub-structures.
- Therefore each patient Y_p is described by $Y_p(x,t_p) = S(t+t_p)A$

The model is estimated by minimizing the following cost function :

$$L = -E_{S\sim q_1, A\sim q_2}[\log(P(X|A,S))] + KL(q_1(S)|p(S)) + KL(q_2(A)|p(A)) + Relu(\lambda(\frac{-dS}{dt}+1))$$

Reconstruction term to fit the data

Kullback-Leibler divergences to drive our approximations closer to the priors

Monotonicity term to constrain the algorithm to look only for monotonic temporal sources

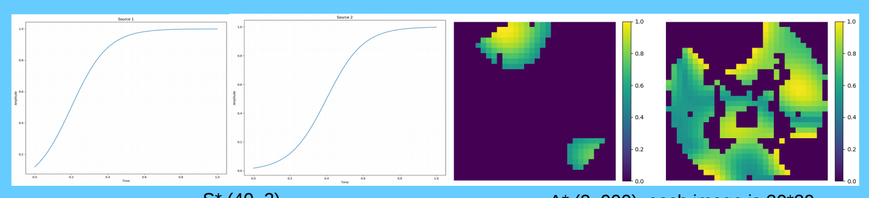
- Where q_1 and q_2 are our approximations for S and A. P(S) and P(A) are the priors on the temporal and spatial sources.
- The algorithm is then applied on the gray matter maps of each subject. These maps are extracted from the T1 brain MRI of each subject and represent an individual high-dimensional biomarker.

Acknowledgments

This work has been supported by the French government, through the UCAJEDI Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01

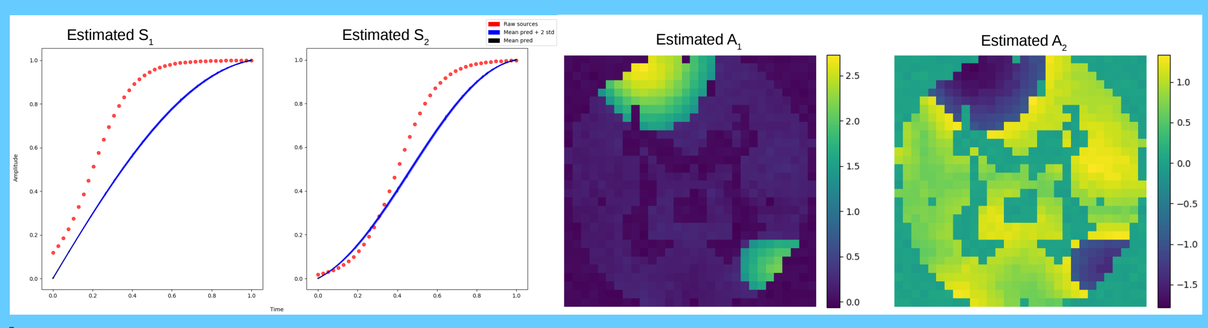
Synthetic Data

Here we present an example with synthetic data where we generate two temporal processes and two associated spatial maps.



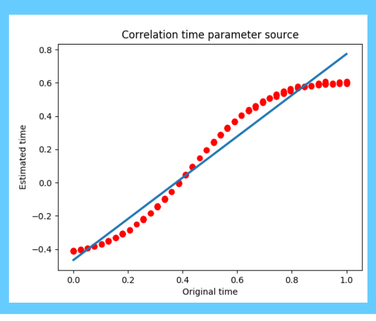
S* (40, 2) A* (2, 900), each image is 30*30

We have 40 time points and we generate for each of them a random number of images X_p such as $X_p = S^*(t_p)A^* + \varepsilon_p$
Next the model is learned using this target matrix X of size (200, 900)



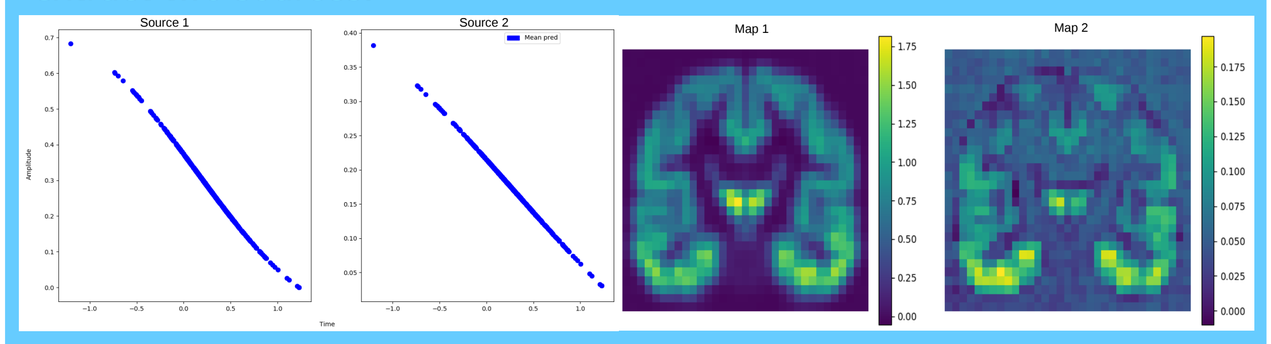
We estimate two smooth Gaussian Processes (in blue) that recover the raw sources (in red). Our model also recovers the original spatial maps.

This figure shows a strong correlation between the original time axis (uniformly distributed between 0 and 1) used to generate the temporal sources and the estimated time of each subject.



Application on Real Data

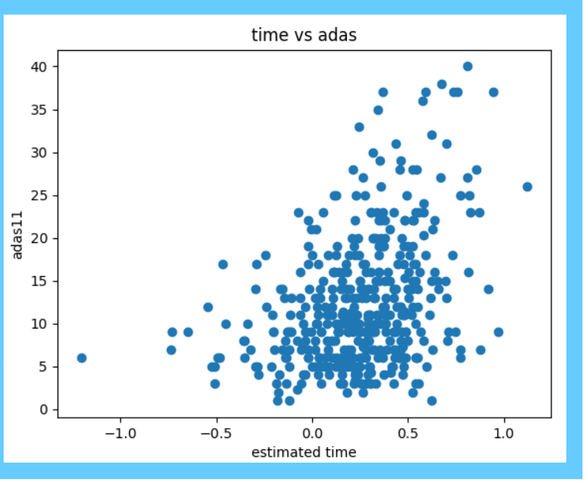
Our input data are the middle coronal slices extracted from the gray matter maps of 555 brains affected by Alzheimer disease from the ADNI database. We sub-sample them and run the algorithm in order to extract two sources.



We get two temporal profiles, each of them related to one spatial map. These profiles show that the gray matter density is decreasing along the estimated time. In particular, subjects who are further on the estimated virtual time axis have a smaller gray matter density than subjects placed earlier on the estimated time scale. We observe a strong activation of the temporal zone on both images, meaning that it's the most affected by this evolution.

The ADAS score is a cognition measure where a low score indicates a healthy person and a high score indicates cognitive impairment. We plot the score of each patient against its estimated time.

We observe that the score increases along the virtual time axis, which means that patients whose estimated time-shift is higher demonstrate lower cognitive abilities.



Conclusion

We presented an original method that combines both the idea of **sources separation** and the estimation of a **virtual individual age** on an absolute time axis, using gray matter maps as high-dimensional biomarkers

The next steps will be to :

- Test this algorithm in a three dimensional setting
- Extend the number of extracted sources
- Once the model is learned, take any particular subject and position it on the temporal trajectories to assess his individual disease stage