A model for a time and frequency Granger causality measure for intracranial electroencephalogram data

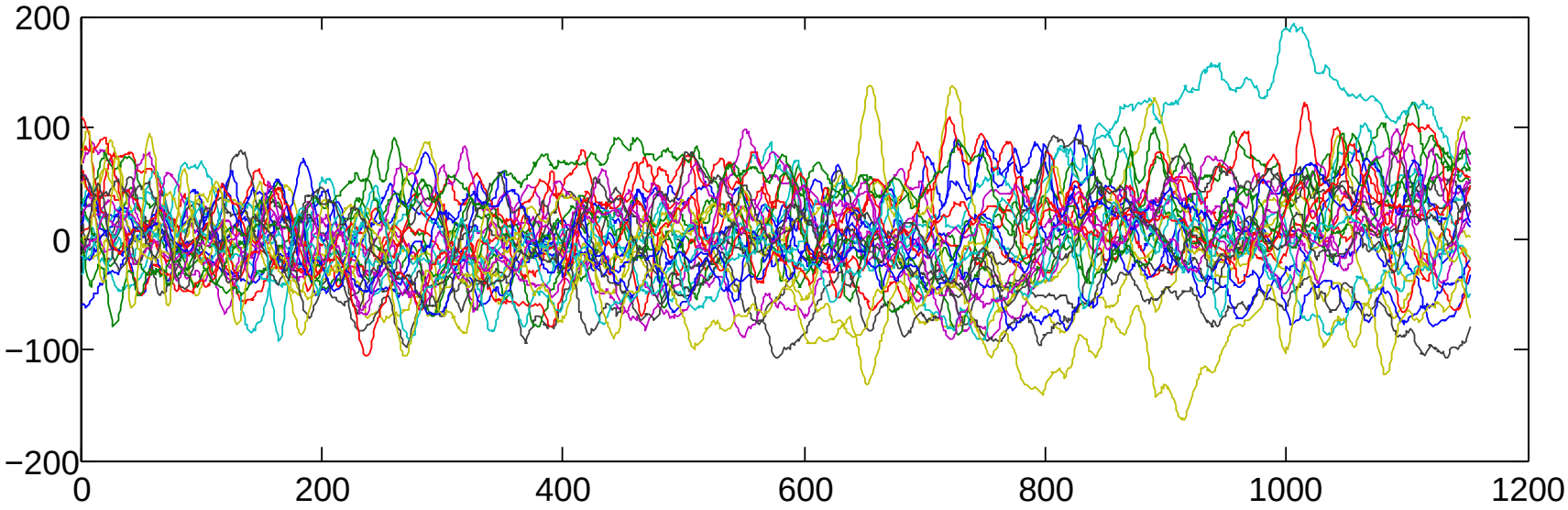
Olivier Renaud

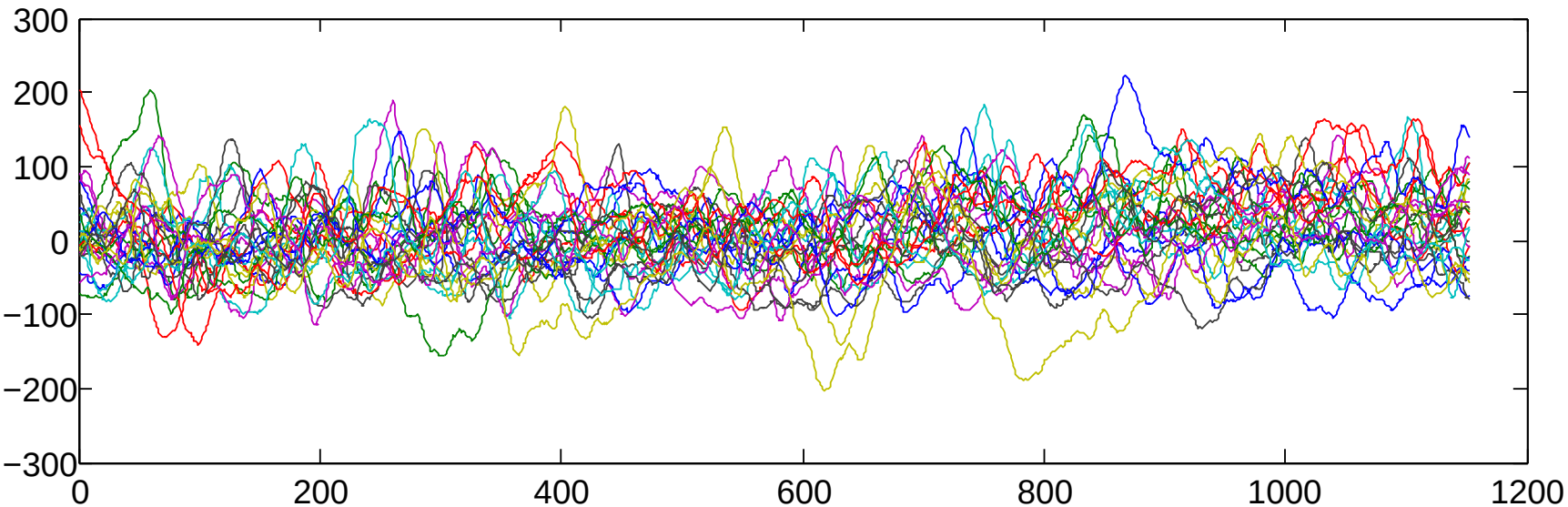
Joint work with Sezen Cekic and Didier Grandjean

http://www.unige.ch/fapse/mad – Dept. of Psychology – University of Geneva

MathStatNeuro workshop, Nice, September 2015

All others must bring data





Substansive question

Hypothesis of research

Oscillations in amygdala at time t and frequency f, (t, f), help to predict orbitofrontal signal at time t.

- Time-varying frequency specific statistic of "causality"/ prediction ability
- Model for multivariate nonstationary time series estimation and prediction
- **Granger causality** : statistical tool to test for dynamical causal relationship (Granger, 1969).

Granger causality

Axiomatic imposition

Cause necessarily precedes its effect in time.

Definition

If a signal X "Granger-causes" another signal Y, then past values of X should contain information that **helps to predict** Y above and beyond the information contained in past values of Y alone ((Granger, 1969)).

$\rightarrow$ Causality in the Wiener-Granger sense is based on the **statistical predictability** of one time series based on knowledge of one other.

- Simple and interpretable method
- Can it be time and frequency specific ?

Time domain Granger causality

- Define a Vector Auto-Regressive model of order p (VAR(p)) :

$$Z_t = \sum_{j=1}^{p} Z_{t-j}\vartheta_j + v_t, \text{ where } Z_t = \begin{pmatrix} Y_t \\ X_t \end{pmatrix} \text{ and } \vartheta_j = \begin{pmatrix} \vartheta_{yy(j)} \vartheta_{xy(j)} \\ \vartheta_{xx(j)} \vartheta_{xx(j)} \end{pmatrix}.$$

- Granger causality criterion based on coefficients
  $\rightarrow$ Test the significativity of the VAR coefficients under interest (Hamilton, 1994).
- If X does not Granger cause Y :

$$H_0 : \vartheta_{xy(1)} = \vartheta_{xy(2)} = \vartheta_{xy(3)} = ...\vartheta_{xy(p)} = 0.$$

Frequency domain Granger causality

- Spectral decomposition of the time domain statistics by **Fourier transform of the VAR** model (Geweke, 1982, 1984).
  - Geweke-Granger causality statistic (Geweke, 1984)
  - Partial Directed Coherence (Baccalà and Sameshima, 2001)
  - Directed Transfer Function (Kaminski and Blinowska, 1991)

Time-varying Granger causality

- Neuroscience data : intrinsically **nonstationary**
  $\rightarrow$ Characteristic of interest !
  $\rightarrow$ Statistic of causality that **catches the dynamic of the causality pattern through time**

- Practically $\rightarrow$ the density $p_t(Z_t|Z_{t-1}^{t-p})$ different for each time.
  $\rightarrow$ **VAR model that evolves in time**.

- Two widely used approaches to estimate a time varying VAR model

  - **Windowing approach** - assumes *locally stationarity*
  - **Adaptive estimation approach** - assumes *slowly varying parameters*

Neuroscience data specificities

- Causal interaction may be delayed from a certain **time interval** (υ).
- Past values of Y and X selected at time-points spaced by a certain **time-lag** (τ).
  Frequency behaviour of the data :
  - high frequency : choose a small time-lag
  - low frequency : choose a large time-lag

We need a model that :

- A reliable estimation of the dynamical VAR coefficients with **several trials**
- capture **short as well as long range causal dependencies** between signals
- is **frequency specific** in the modelisation of the causal dependencies between signals
- provides a suitable **order selection criterion**

Bayesian State Space Model

- Rewrite the dynamical VAR model in a **state space** form (Cassidy, 2002).

$$\begin{cases} \varphi_{t+1} = A\varphi_t + w_t & w_t \sim \mathcal{N}_k(0, Q) \\ Z_t = C_t\varphi_t + v_t & v_t \sim \mathcal{N}_d(0, R) \end{cases} \text{ where } \begin{cases} \varphi_t & = vec[\vartheta_{1(t)}, \vartheta_{2(t)}, .., \vartheta_{p(t)}]' \\ Z_t & = \begin{pmatrix} Y_t \\ X_t \end{pmatrix} \\ C_t\varphi_t & = \sum_{j=1}^{p} \vartheta_{j(t)} \begin{pmatrix} Y_{t-j} \\ X_{t-j} \end{pmatrix} \end{cases}$$

$$\Leftrightarrow p(Z_t|C_t, \varphi_t) \sim \mathcal{N}_d(C_t\varphi_t, R), \quad p(\varphi_t|A, \varphi_{t-1}) \sim \mathcal{N}_k(A\varphi_{t-1}, Q), \quad p(\varphi_1) \sim \mathcal{N}_k(\mu_1, \Sigma_1).$$

- Frequentist approach : **EM algorithm**
- **Huge amount of parameters**
  - $\varphi_1^T : [pd^2\,T], \qquad A : (pd^2)^2, \qquad Q : (pd^2)^2, \qquad R : d^2$
  - Example : 1000 time points (typical neuroscience series length)/ 2 channels / model order 4 $\Rightarrow$ 8132 parameters to estimate ! !

Variational Bayes

- Bayesian approach $\rightarrow \Omega_1^b = \{A, Q, R\}$ random
- Target quantity : posterior distribution $p(\varphi_1^T, \Omega_1^b | Z_1^T)$
  $\rightarrow$ Intractable integral of very high dimension

Variational approximation

Posterior density $p(\varphi_1^T, \Omega_1^b | Z_1^T) \rightarrow$ approximated by a *variational posterior density*

$$p(\varphi_1^T, \Omega_1^b | Z_1^T) \approx q(\varphi_1^T, \Omega_1^b | Z_1^T).$$

- Find the distribution $q(\varphi_1^T, \Omega_1^b | Z_1^T)$ that minimizes the KL-divergence between $p(\varphi_1^T, \Omega_1^b | Z_1^T)$ and $q(\varphi_1^T, \Omega_1^b | Z_1^T)$

$$\mathrm{KL}(q(\varphi_1^T, \Omega_1^b | Z_1^T) \| p(\varphi_1^T, \Omega_1^b | Z_1^T)) = \left\langle \log \frac{q(\varphi_1^T, \Omega_1^b | Z_1^T)}{p(\varphi_1^T, \Omega_1^b | Z_1^T)} \right\rangle_{q(\varphi_1^T, \Omega_1^b | Z_1^T)}$$

$\langle . \rangle$: expectation, subscript : density used for this expectation.

Learning rules

- Can show :

$$\log p(Z_1^T) = \mathrm{KL}(q(\varphi_1^T, \Omega_1^b | Z_1^T) \| p(\varphi_1^T, \Omega_1^b | Z_1^T)) - \left\langle \log \frac{q(\varphi_1^T, \Omega_1^b | Z_1^T)}{p(\varphi_1^T, \Omega_1^b, Z_1^T)} \right\rangle_{q(\varphi_1^T, \Omega_1^b | Z_1^T)}$$
$$= \mathrm{KL}(q(\varphi_1^T, \Omega_1^b | Z_1^T) \| p(\varphi_1^T, \Omega_1^b | Z_1^T)) + \mathbf{F}(q(\varphi_1^T, \Omega_1^b | Z_1^T)).$$

  Fundamental equation of the variational Bayesian methodology. $\mathrm{KL}(q\|p) > 0$ $\rightarrow \mathbf{F}(q(\varphi_1^T, \Omega_1^b | Z_1^T))$ (negative Free Energy) : lower bound for the log evidence :

- Suitable choice for $q(\varphi_1^T, \Omega_1^b | Z_1^T) \rightarrow$ integral in $\mathbf{F}\left(q(\varphi_1^T, \Omega_1^b | Z_1^T)\right)$ tractable $\rightarrow$ maximizable in order to obtain an optimal form for the approximation density $q(\varphi_1^T, \Omega_1^b | Z_1^T)$.

Mean-field approximation

☐ Assumption underlying the variational Bayes methodology :

$$q(\varphi_1^T, \Omega_1^b | Z_1^T) = q(\varphi_1^T | Z_1^T) \prod_{j=1}^{b} q(\Omega_j | Z_1^T).$$

☐ Free energy (model evidence lower bound) can be rewritten

$$\mathbf{F}\Big(q(\varphi_1^T | Z_1^T), q(\Omega_1 | Z_1^T), \ldots, q(\Omega_b | Z_1^T)\Big).$$

$\rightarrow$ Mean field approximation may have minor to major impacts on the resulting inference depending on the model at hand

Variational EM algorithm

- **Maximizing the functional** $\mathbf{F}\Big(q(\varphi_1^T|Z_1^T), q(\Omega_1|Z_1^T), \ldots, q(\Omega_b|Z_1^T)\Big)$
  (calculus of variations)
  $\rightarrow$ **Iterative optimal forms** for $q(\varphi_1^T|Z_1^T), q(\Omega_1|Z_1^T), \ldots, q(\Omega_b|Z_1^T)$.

$$q^*(\varphi_1^T|Z_1^T)^{(I+1)} \propto \exp\left\langle \log p(\varphi_1^T|\Omega_1^b, Z_1^T)\right\rangle^{(I)}_{q(\Omega_1^b|Z_1^T)},$$

$$q^*(\Omega_m|Z_1^T)^{(I+1)} \propto \exp\left\langle \log p(\Omega_1^b|\varphi_1^T, Z_1^T)\right\rangle^{(I)}_{-\Omega_m},$$

where $\langle . \rangle^{(I)}_{-\Omega_m}$ is the expectation over all the distributions at iteration (I) except $q(\Omega_m|Z_1^T)^{(I)}$ (Beal, 2003; Ostwald et al., 2014).

Theorem

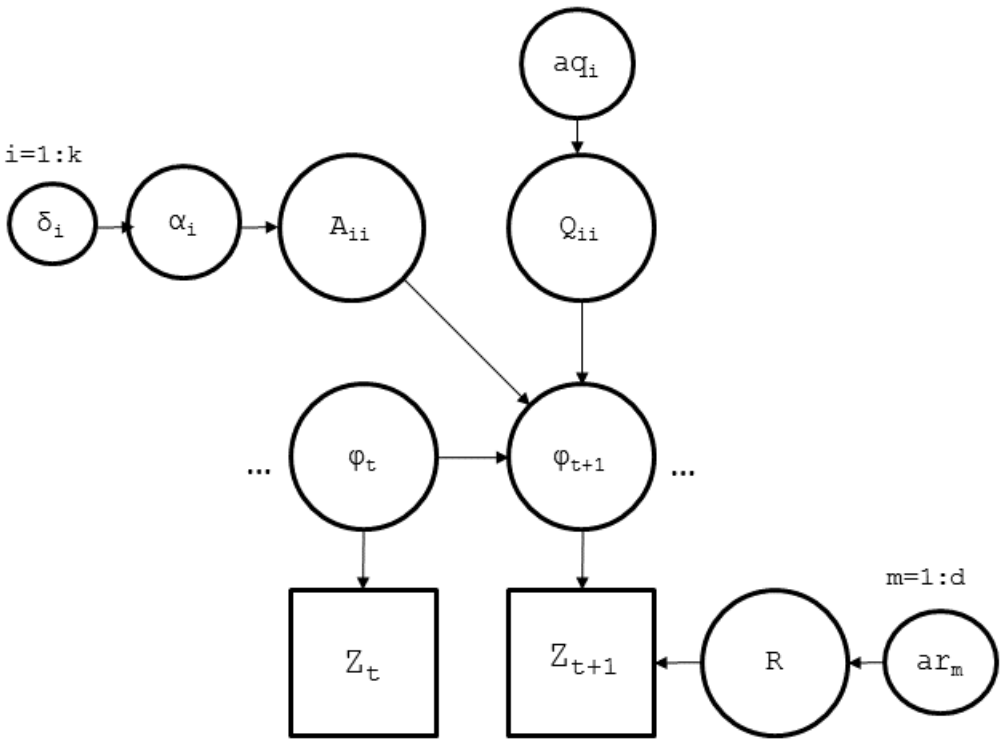
If the complete-data likelihood $p(\varphi_1^T, Z_1^T|\Omega_1^b)$ *is part of the* ***exponential family*** *[...] and if the hidden and parameter priors distributions* $p(\varphi_1^T)$ *and* $p(\Omega_1^b)$ *are* ***conjugate*** *to this complete-data likelihood, the corresponding variational approximate posterior distributions that maximize* $\mathbf{F}$, $q^*(\varphi_1^T|Z_1^T)$ *and* $q^*(\Omega_1^b|Z_1^T)$, *are of the same distributional form than respectively the prior distributions* $p(\varphi_1^T)$ *and* $p(\Omega_1^b)$.

Full model specification

$$\begin{cases} \varphi_{t+1} = A\varphi_t + w_t & w_t \sim \mathcal{N}_k(0, Q) \\ Z_t = C_t\varphi_t + v_t & v_t \sim \mathcal{N}_d(0, R) \end{cases}$$

where

$$\begin{cases} \varphi_t & = vec[\vartheta_{1(t)}, \vartheta_{2(t)}, .., \vartheta_{p(t)}]' \\ Z_t & = \begin{pmatrix} Y_t \\ X_t \end{pmatrix} \\ C_t\varphi_t & = \sum_{j=1}^{p} \vartheta_{j(t)} \begin{pmatrix} Y_{t-j} \\ X_{t-j} \end{pmatrix} \end{cases}$$



The Variational Evidence Lower Bound

- **Evidence** : natural criterion for model order selection (Bayes Factor)
- **Free energy** : lower bound for model evidence $\rightarrow$ Select the model order m_p on the basis of $\mathbf{F}_{m_p}$ quantity
- Intuition that $\mathbf{F}_{m_p}$ is a suitable model order selection criterion :

$$\mathbf{F}_{m_p} = \underbrace{\left\langle \log \frac{p(\varphi_1^T, Z_1^T, \Omega_1^b | m_p)}{q(\varphi_1^T, Z_1^T | m_p)} \right\rangle_{q(\varphi_1^T, \Omega_1^b | m_p)}}_{\text{Av. loglik : Accuracy term}} - \underbrace{\left\langle \log \frac{q(\Omega_1^b | Z_1^T, m_p)}{p(\Omega_1^b | m_p)} \right\rangle_{q(\Omega_1^b | Z_1^T, m_p)}}_{\text{KL divergence prior / variational posterior : Penalty term}}$$

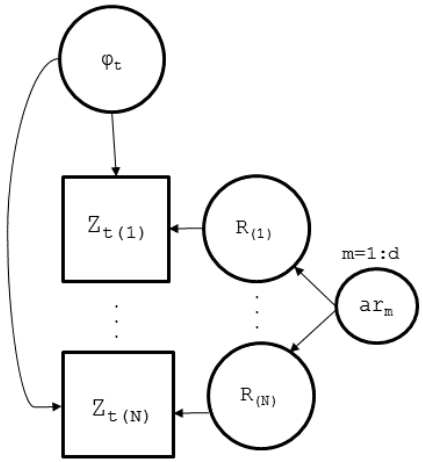
- ! Suppose same tightness of $\mathbf{F}_{m_p}$ for all order p.

Multiple trials

- Can easily be modified to deal with N conditionally independent sequences $\{Z_1^T{}_{(j)}\}_{j=1}^N$
- Usually achieved by running the E-step once for each realization and then averaging all the estimates of the required sufficient statistics for the M-step
- The model can be rewritten

$$\begin{cases} \varphi_{t+1} = A\varphi_t + w_t,\ w_t \sim \mathcal{N}_k(0, Q) \\ \mathbf{Z}_t = \mathbf{C}_t\varphi_t + v_t,\ v_t \sim \mathcal{N}_d(0, \mathbf{R}) \end{cases} \text{ where } \mathbf{Z}_t = \begin{pmatrix} Y_{t(1)} \\ \dots \\ Y_{t(N)} \\ X_{t(1)} \\ \dots \\ X_{t(N)} \end{pmatrix},\ \mathbf{C}_t\varphi_t = \sum_{j=1}^{p} \vartheta_{j(t)} \begin{pmatrix} Y_{t-j(1)} \\ \dots \\ Y_{t-j(N)} \\ X_{t-j(1)} \\ \dots \\ X_{t-j(N)} \end{pmatrix},\ \mathbf{R} = \text{diag} \begin{pmatrix} R_1 \\ \dots \\ R_N \end{pmatrix}.$$

- The conditional dependency on the dataset $\{Z_1^T{}_{(j)}\}_{j=1}^N$ is present merely for variables φ_1^T and R.



Bayesian State Space Multiscale Model

We have

- A reliable estimation of the dynamical VAR coefficients in a several trials situation
- A suitable model order selection criterion

We need

- to capture short- and long-range causal dependencies between signals (better than υ and τ parameters)
- to be frequency specific in the modelisation of the causal dependencies between signals

$\Rightarrow$ **Bayesian State Space Multiscale Model**

Bayesian State Space Multiscale Model

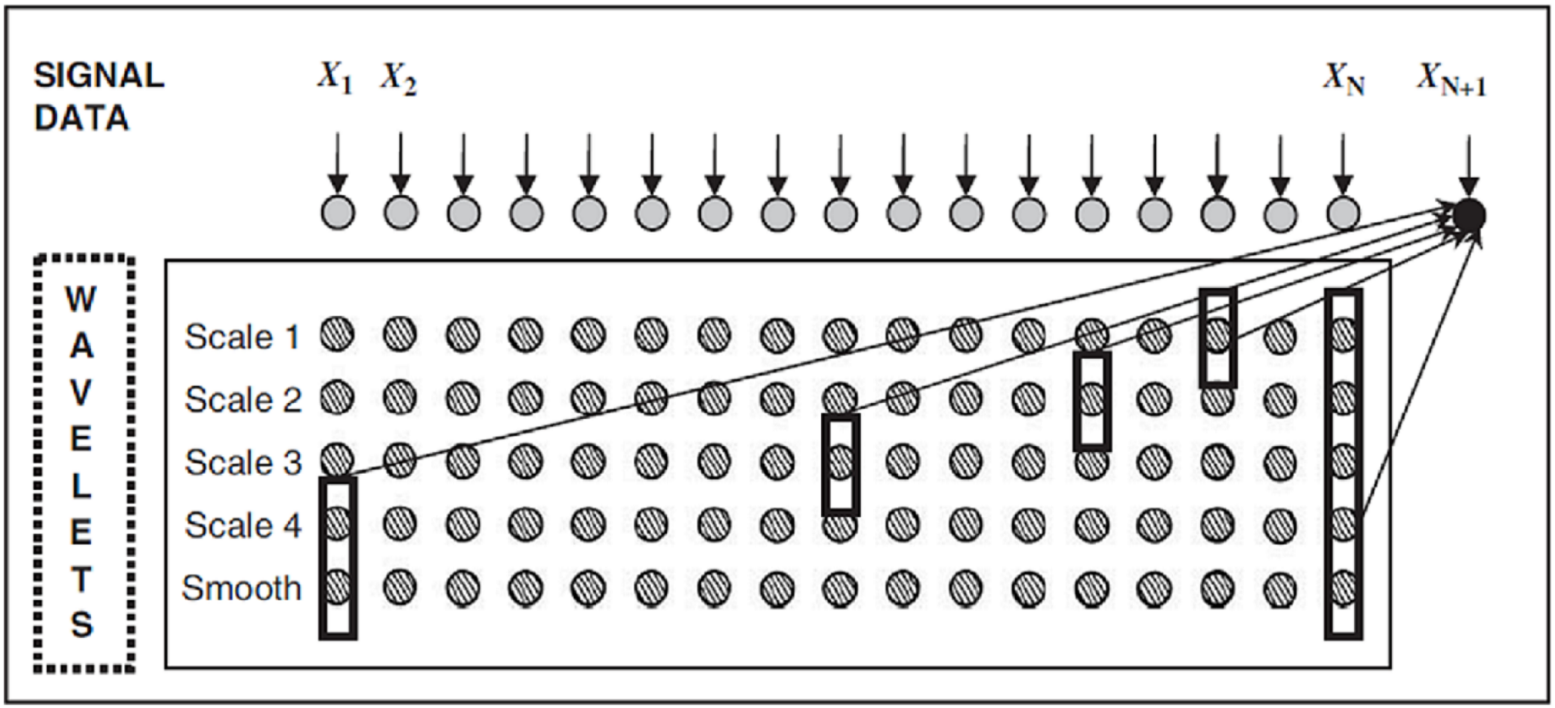
- Idea : decompose the past values of the signals Y_t and X_t contained in $\{C_t\}_{t=1}^T$ in wavelets and use $\{C_t^w\}$ as prediction instead of matrices $\{C_t\}_{t=1}^T$
- only Haar "à trou" wavelet for "pure" prediction
- Construct C_t^w as

$$\begin{aligned} C_t^w = &\{w^y_{j,t-1-2^j(k-1)}\}_{j=1:J,k=1:p_j}, \{s^y_{J,t-1-2^J(k-1)}\}_{k=1:p_{J+1}}, \\ &\{w^x_{j,t-1-2^j(k-1)}\}_{j=1:J,k=1:p_j}, \{s^x_{J,t-1-2^J(k-1)}\}_{k=1:p_{J+1}}. \end{aligned}$$

$$\begin{cases} \varphi_{t+1} = A\varphi_t + w_t & w_t \sim \mathcal{N}_k(0, Q) \\ Z_t = C_t^w \varphi_t + v_t & v_t \sim \mathcal{N}_d(0, R) \end{cases}$$

- All results in VBSSM can be applied
- "À trou" extension $\rightarrow$ generalisation of the VBSSM

Bayesian State Space Multiscale Model



- Wavelet coefficients (scale 1 to 4) and smooth used for prediction
- $J = 4$ and $p_j = 2 \; \forall j$
- only 10 coefficients used, but contains long- and small-range prediction
- Model orders selection : p_j and number of scales J

Bayesian State Space Multiscale Model

Multiresolution methodology benefits :

- Wavelets coefficients w_j directly related to a **specific frequency band**
  $\rightarrow$ frequency specific in the modelisation of the causal relationship
- Capture **short and long-range dependencies** between signals with only **few parameters**
  $\rightarrow$ Avoids the choice of the **time interval** (υ) **time-lag** τ
- Simple and suitably **interpretable** time-varying Granger causality statistic (Chicharro, 2011)

Bayesian Granger-Causality statistic

Bayesian Granger-Causality statistic

A necessary and sufficient condition for X not being Granger-causal for Y_t at specific time t, is if and only if each element of the subset $\widetilde{\varphi_t}$ of φ_t, that contains all the causal coefficients of interest equals zero.

- Sub-vector $\widetilde{\varphi_t}$ of dimension c
- Suitable partition of the posterior parameters $\{\mu_{\mathbf{t}}, \Sigma_t\}$ gives the **marginal posterior density** :

$$q(\widetilde{\varphi_t}|Z_1^T) = \mathcal{N}_c(\widetilde{\mu_t}, \widetilde{\Sigma_t}). \tag{1}$$

- Contribution of $\widetilde{\varphi_t}$ to the prediction of $Y_t \leftrightarrow \widetilde{\varphi_t} = 0$
- Define the **credible region**

$$CR_{\widetilde{\varphi_t}} = \{\widetilde{\varphi_t}|(\widetilde{\varphi_t} - \widetilde{\mu_t})'\widetilde{\Sigma_t}^{-1}(\widetilde{\varphi_t} - \widetilde{\mu_t}) < q_{\chi^2_c}(95\%)\}. \tag{2}$$

Assessment of accuracy : Model order selection

- Bivariate system, lengths $\{500; 1000; 2000\}$, Trials $\{1; 10\}$
- φ_1^T Markovian
- Estimation performed for orders $p = 1 : 10$
- Results based on 100 simulations

Trials	Length	Model order	True model selected
1	500	1	**83%**
1	500	2	95%
1	500	4	96%
1	500	8	97%
1	1000	1	**87%**
1	1000	2	100%
1	1000	4	97%
1	1000	8	99%
1	2000	1	**89%**
1	2000	2	99%
1	2000	4	100%
1	2000	8	98%
10	500	1	**89%**
10	500	2	98%
10	500	4	100%
10	500	8	100%

TABLE: Proportion of correct model order selection by $\mathbf{F}_{m_p}$

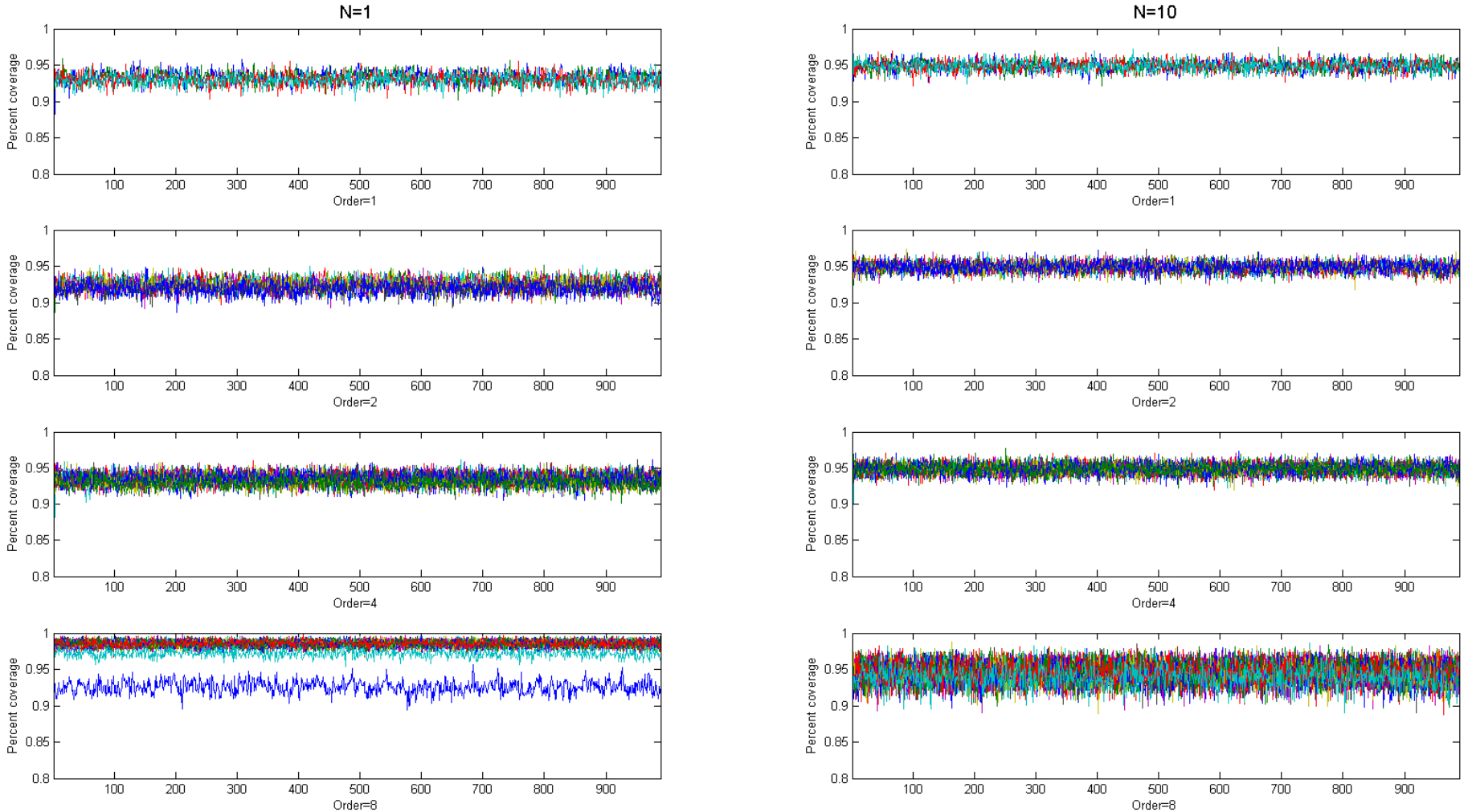
Trials	Length	Model order	Wrong model selected	Occurrence
1	500	1	2 3	**16** **1**
1	500	2	3 4	3 2
1	500	4	5	4
1	500	8	5 6	2 1
1	1000	1	2	**13**
1	1000	2	NONE	NONE
1	1000	4	5	3
1	1000	8	6	1
1	2000	1	2	**11**
1	2000	2	3	1
1	2000	4	NONE	NONE
1	2000	8	4 7	1 1
10	500	1	2 3	**8** **3**
10	500	2	3 4	1 1
10	500	4	NONE	NONE
10	500	8	NONE	NONE

TABLE: Details on wrong model order selections by $\mathbf{F}_{m_p}$.

Assessment of accuracy : Credible Interval Coverage

Hidden variables φ_1^T follow the state model

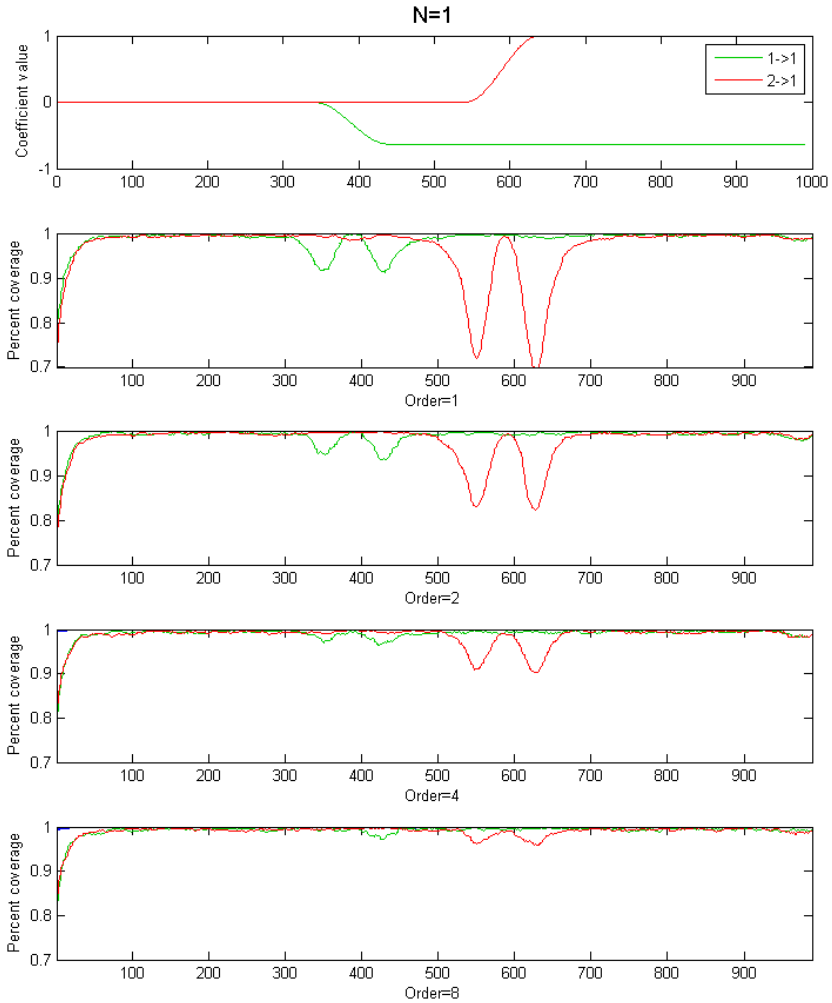
Orders $\{1, 2, 4, 8\}$, length : 1000 time points, 1000 replications.

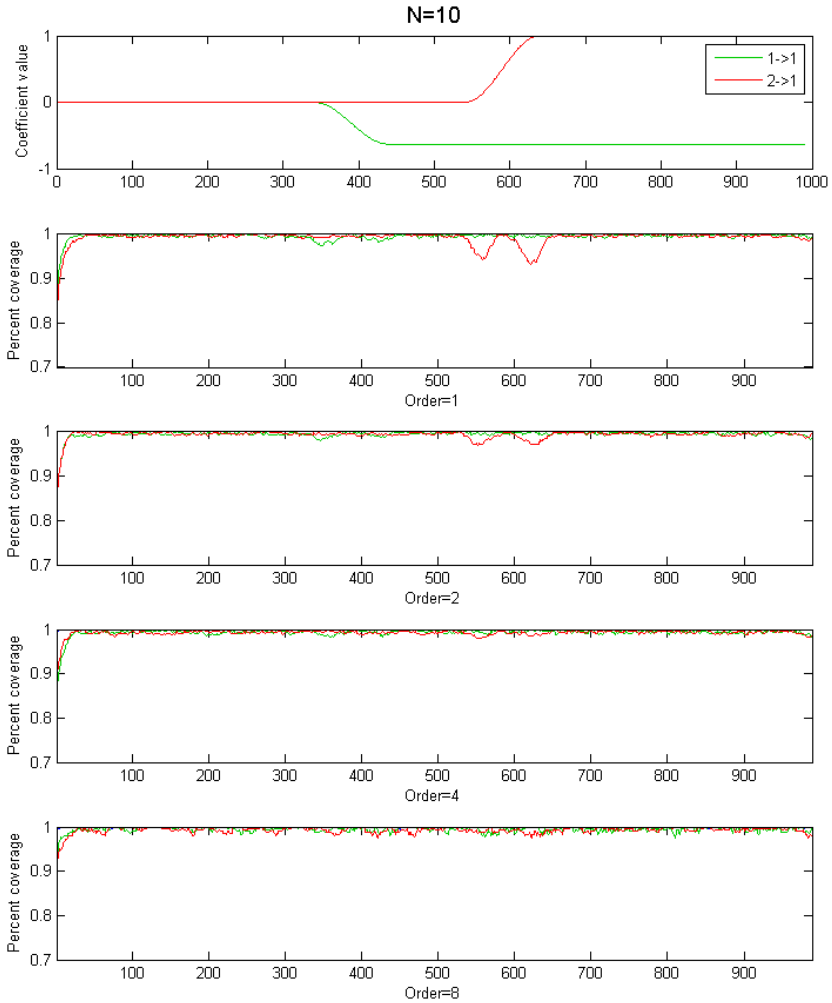


Assessment of accuracy : Credible Interval Coverage

Hidden variables φ_1^T evolve slowly & deterministically

Orders $\{1, 2, 4, 8\}$, length : 1000 time points, 1000 replications.

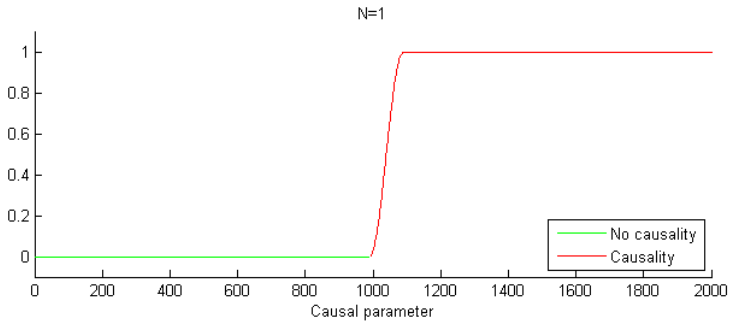


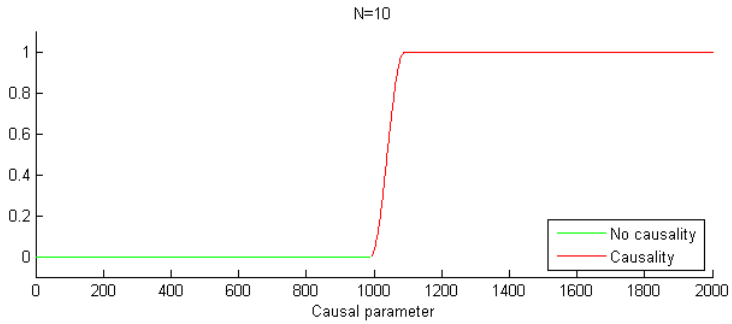


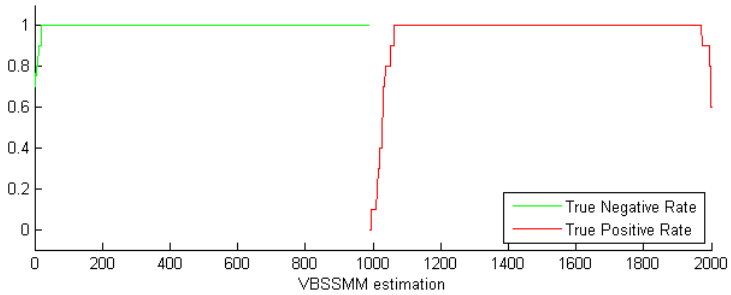
Assessment of accuracy : Granger causality detection

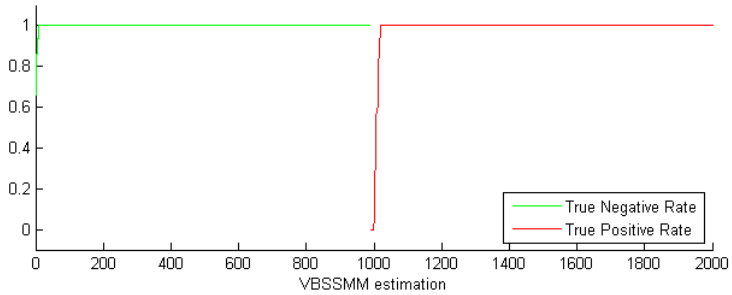
Simulated signals with slowly varying parameters.

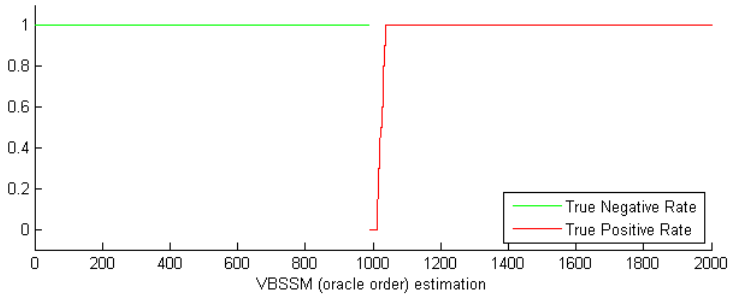
Order 4, length 2000, Trials $\{1; 10\}$.

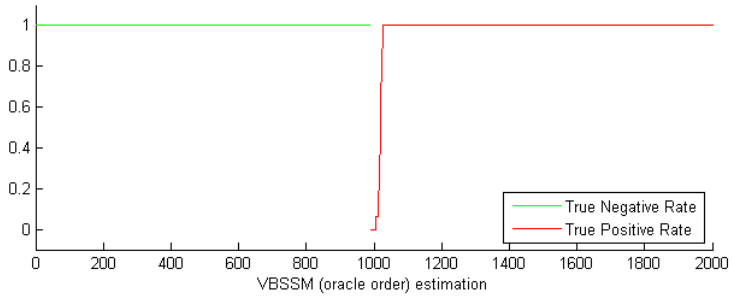




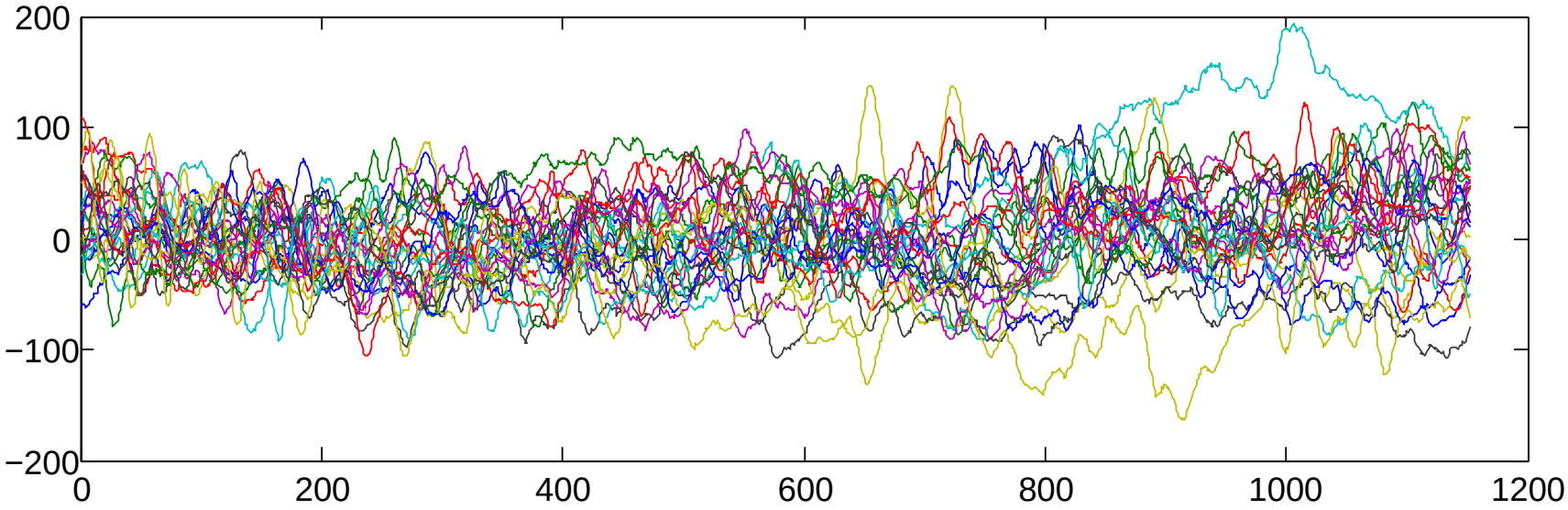


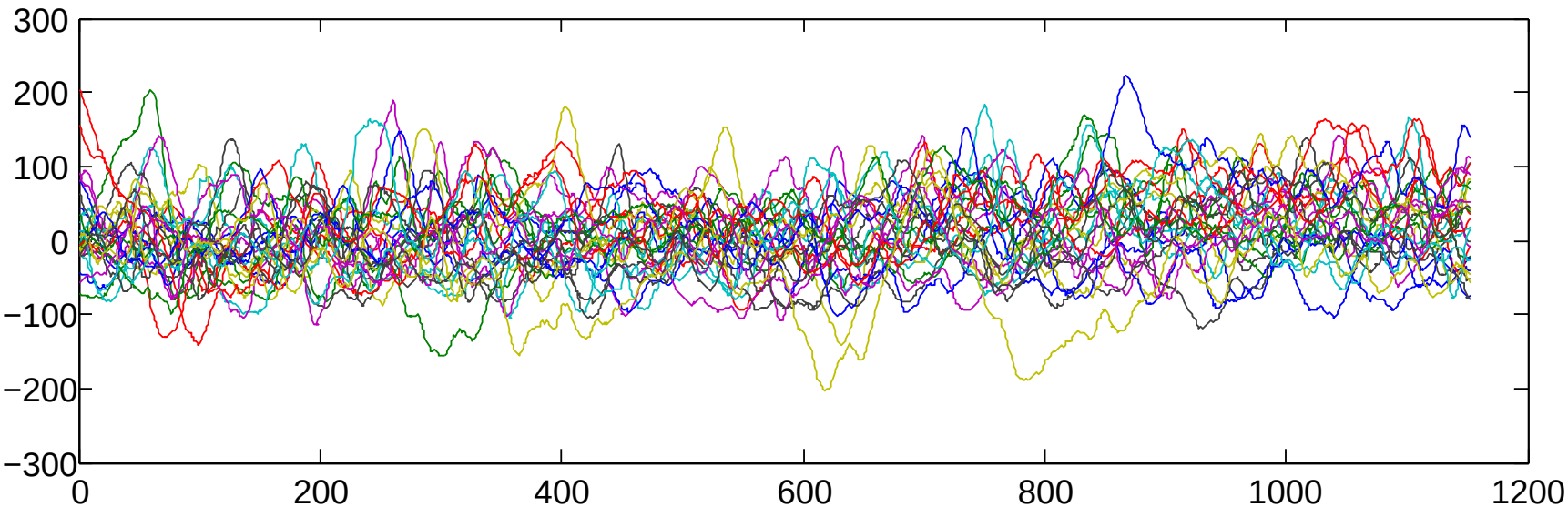






All others must bring data





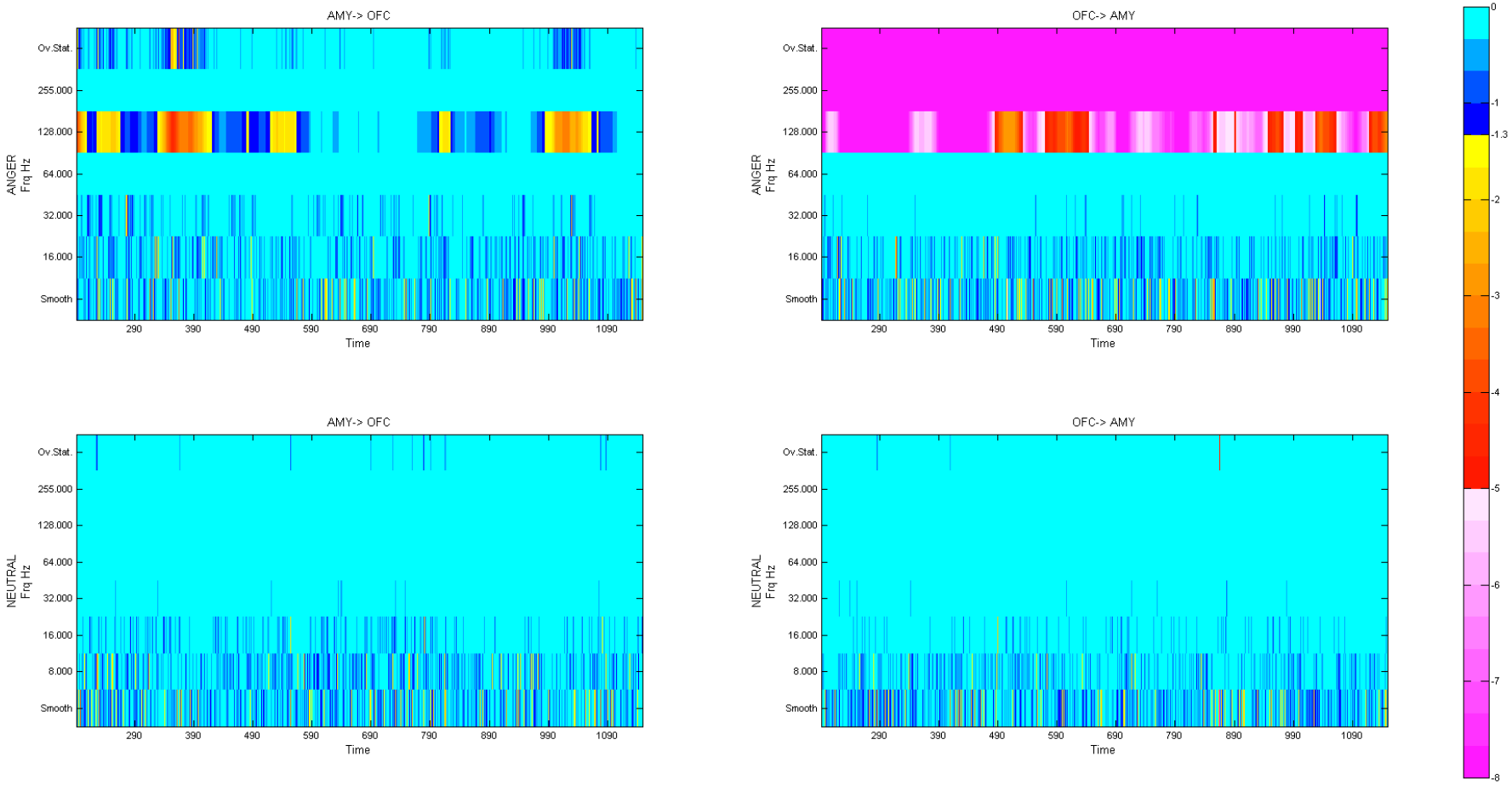
Application

- iEEG data recorded during psychological experimental situations in four epileptic pharmaco-resistant patients
- Recordings localized within the amygdala and medial orbito-frontal cortex
- Study the dynamics of neuronal processes within and between these regions in response to emotional prosody exposure (Christen and Grandjean, 2010)
- Two experimental conditions : *anger* and *neutral*

A significant Granger causality from signal 1 to signal 2 at frequency f and at time t means that the energy of signal 1 at frequency f significantly improves the prediction of the value of signal 2 at time t.

Application

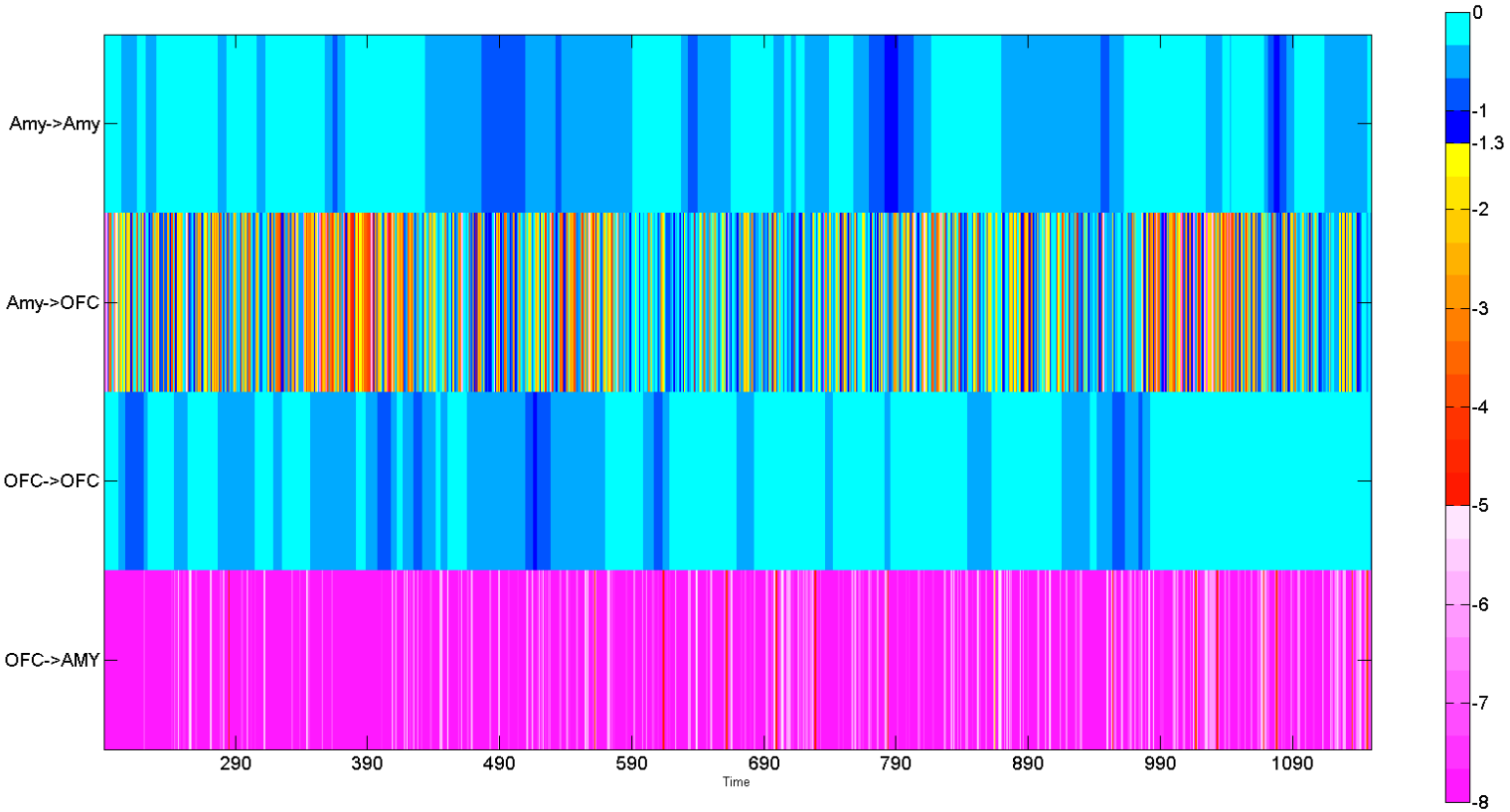
P-values in $\log_{10}$ scale for the overall causality statistic and for each scale.
For *anger* and `neutral`.



Application : Experimental conditions comparison

Difference of causal effect between `anger` and `neutral`

- F statistic : divide overall statistic for A and for N condition $\rightarrow$ (2 independent χ^2)
- P-values on a log_{10} scale (log_{10} scale values $\{0, 1, 1.3, 2, 3, 4\} \rightarrow$ origina p-values $\{1, 10^{-1}, 0.05, 10^{-2}, 10^{-3}, 10^{-4}\}$



Application : Discussion

- Multiple testing : the frequency dimension
  - $\rightarrow$ Hierarchically testing
  - $\rightarrow$ usually relaxed in Neuroscience (GCC, PDC, DTF)
- Multiple testing : the time dimension
  - Threshold : α level but take into account periods of significance only if they are sustained enough
  - Bonferroni correction (too conservative)
  - Cluster mass test (Maris and Oostenveld, 2007)

References

Baccalà, L. A. and K. Sameshima (2001). Partial directed coherence : a new concept in neural structure determination. *Biological cybernetics 84*(6), 463–474.

Beal, M. (2003). *Variational Algorithms for Approximate Bayesian Inference*. Ph. D. thesis, University College London, London.

Cassidy, M.J. Penny, W. (2002). Bayesian nonstationary autoregressive models for biomedical signal analysis. *Biomedical Engineering, IEEE Transactions on 49*(10), 1142–1152.

Cekic, S. (2010). *Lien entre activité neuronale des sites cérébraux de l'amygdale et du cortex orbito-frontal en réponse à une prosodie émotionnelle : investigation par la Granger-causalité*. Master Thesis, Université de Genève.

Chicharro, D. (2011). On the spectral formulation of granger causality. *Biological cybernetics 105*(5-6), 331–347.

Christen, A. and D. Grandjean (2010). Temporal dynamics of amygdala and orbitofrontal responses to emotional prosody using intracerebral local field potentials in humans. In *Speech Prosody*, Volume 100874, Chicago, pp. 1–4.

Ding, M., S. L. Bressler, W. Yang, and H. Liang (2000). Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling : data preprocessing, model validation, and variability assessment. *Biological cybernetics 83*(1), 35–45.

Geweke, J. (1982). Measurement of Linear Dependence and Feedback Between Multiple Time Series. *Journal of the American Statistical Association 77*(378), 304–313.

Geweke, J. F. (1984). Measures of Conditional Linear Dependence and Feedback Between Time Series. *Journal of the American Statistical Association 79*(388), 907–915.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica 37*(3), 424–438.

Hamilton, J. D. (1994). *Time series analysis*, Volume 2. Princeton university press Princeton.

Huang, A., M. P. Wand, et al. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis 8*(2), 439–452.

Kaminski, M. and K. Blinowska (1991). A new method of the description of the information flow in the brain structures. *Biological cybernetics 65*(3), 203–210.

Maris, E. and R. Oostenveld (2007). Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods 164*(1), 177–190.

Menictas, M. and M. P. Wand (2013). Variational inference for marginal longitudinal semiparametric regression. *Stat 2*(1), 61–71.

Ostwald, D., E. Kirilina, L. Starke, and F. Blankenburg (2014). A tutorial on variational bayes for latent linear stochastic time-series models. *Journal of Mathematical Psychology 60*, 1–19.

Renaud, O., J.-L. Starck, and F. Murtagh (2003). Prediction based on a multiscale decomposition. *International Journal of Wavelets, Multiresolution and Information Processing 1*(2), 217–232.

Schlögl, A. (2000). *The electroencephalogram and the adapdative autoregressive model : theory and applications*. Ph. D. thesis, University of Graz, Graz.

Schlögl, A., S. Roberts, and G. Pfurtscheller (2000). A criterion for adaptive autoregressive models. In *Proceedings of the 22 nd IEEE International Conference on Engineering in Medicine and Biology*, pp. 1581–1582.