Multiscale Bayesian State Space Model for Granger Causality Analysis, with Application to 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

ICORS, Wollongong, July 2017
Neuroscience question

- iEEG data recorded during psychological experimental situation
- Study the dynamics of neuronal processes between regions of interest in response to emotional prosody exposure

Presentation of binaural stimuli (Munich database, Banse & Scherer, 1996) using headphones (750ms)
30 trials of each emotional type (angry, happiness, sadness, fearful, neutral) in three blocks, presented in pseudo-random order
Control condition: extracted envelope of the sounds + white noise/mean F0 (static)/F0 dynamic (changes of F0 in function of the time)
Neuroscience question

Depth electrodes of interest

- Recordings localized within the AMY and OFC
The Data
Granger causality

**Definition**

If a signal $Z^{AMY}$ “Granger-causes” another signal $Z^{OFC}$, then past values of $Z^{AMY}$ should contain information that helps to predict $Z^{OFC}$ above and beyond the information contained in past values of $Z^{OFC}$ alone (Granger, 1969).

→ 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

Hypothesis of research

Signal in amygdala help to predict orbitofrontal signal.
Time domain Granger causality

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

$$Z_t = \sum_{j=1}^{p} Z_{t-j} \vartheta_j + \nu_t,$$

where $Z_t = \begin{pmatrix} Z_{t}^{\text{OFC}} \\ Z_{t}^{\text{AMY}} \end{pmatrix}$ and $\vartheta_j = \begin{pmatrix} \vartheta_{11}(j) \\ \vartheta_{12}(j) \\ \vartheta_{21}(j) \\ \vartheta_{22}(j) \end{pmatrix}$.

- Granger causality criterion based on coefficients:
  → Test the significativity of the VAR coefficients under interest (Hamilton, 1994).

- If $Z^{\text{AMY}}$ does not Granger cause $Z^{\text{OFC}}$:

$$H_0 : \vartheta_{21}(1) = \vartheta_{21}(2) = \vartheta_{21}(3) = ... \vartheta_{21}(p) = 0.$$
Time-varying Granger causality

Hypothesis of research

Signal in amygdala at time $t'$ help to predict orbitofrontal signal at time $t$. 
Time-varying Granger causality

- Neuroscience data: intrinsically **nonstationary** \(\rightarrow\) Characteristic of interest!

- Statistic of causality that **catches the dynamic of the causality pattern through time**

- Practically \(\rightarrow\) **VAR model** that evolves in time.

\[
Z_t = \sum_{j=1}^{p} Z_{t-j} \vartheta_{j,t} + v_t, \quad \text{where } Z_t = \begin{pmatrix} Z_t^{OFC} \\ Z_t^{AMY} \end{pmatrix} \quad \text{and } \vartheta_{j,t} = \begin{pmatrix} \vartheta_{11(j,t)} & \vartheta_{21(j,t)} \\ \vartheta_{12(j,t)} & \vartheta_{22(j,t)} \end{pmatrix}
\]

- Granger causality criterion based on coefficients

- If \(Z_t^{AMY}\) does not Granger cause \(Z_t^{OFC}\) at time \(t\):

\[
H_0 : \vartheta_{21(1,t)} = \vartheta_{21(2,t)} = \vartheta_{21(3,t)} = \ldots \vartheta_{21(p,t)} = 0.
\]
Bayesian State Space Model

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

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

where

\[
\begin{align*}
\varphi_t &= \text{vec}[^{\varphi_1(t)} \varphi_2(t) \ldots \varphi_p(t)]' \\
Z_t &= \begin{pmatrix} Z_t^{\text{OFC}} \\ Z_t^{\text{AMY}} \end{pmatrix} \\
C_t\varphi_t &= \sum_{j=1}^{p} \varphi_j(t) \begin{pmatrix} Z_{t-j}^{\text{OFC}} \\ Z_{t-j}^{\text{AMY}} \end{pmatrix}
\end{align*}
\]

⇒

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

⇒ Huge amount of parameters

- \(\varphi_1^T: [pd^2 T]\), \(A: (pd^2)^2\), \(Q: (pd^2)^2\), \(R: d^2\)
- Example: 1000 time points (typical neuroscience series length)/ 2 channels / model order 4 ⇒ 8132 parameters to estimate!!
Variational Bayes

- Bayesian approach $\rightarrow \Omega_1^b = \{A, Q, R\}$ random
- Target quantity: evidence $p(Z_1^T)$ $\rightarrow$ 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

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

□ Can show:

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

\[
= KL(q(\varphi_1^T, \Omega_1^b|Z_1^T)\|p(\varphi_1^T, \Omega_1^b|Z_1^T)) + F(q(\varphi_1^T, \Omega_1^b|Z_1^T)) > 0
\]

where \(\langle . \rangle\) : expectation, subscript : density used for this expectation.

\[
\log p(Z_1^T) \geq F(q(\varphi_1^T, \Omega_1^b|Z_1^T)).
\]

□ Suitable choice for \(q(\varphi_1^T, \Omega_1^b|Z_1^T) \rightarrow\) integral in \(F(q(\varphi_1^T, \Omega_1^b|Z_1^T))\) tractable

□ Assumption underlying the variational Bayes methodology :

\[
q(\varphi_1^T, \Omega_1^b|Z_1^T) = q(\varphi_1^T|Z_1^T) \prod_{j=1}^{b} q(\Omega_j|Z_1^T).
\]
Variational EM algorithm

□ Maximizing the functional \( F \left( q(\varphi_1^T | Z_1^T), q(\Omega_1 | Z_1^T), \ldots, q(\Omega_b | Z_1^T) \right) \)
(calculus of variations)

→ **Iterative optimal forms** for \( q(\varphi_1^T | Z_1^T), q(\Omega_1 | Z_1^T), \ldots, q(\Omega_b | Z_1^T) \).

\[
q^*(\varphi_1^T | Z_1^T)^{(l+1)} \propto \exp \left\langle \log p(\varphi_1^T | \Omega_b^1, Z_1^T) \right\rangle_{q(\Omega_b^1 | Z_1^T)}^{(l)},
\]

\[
q^*(\Omega_m | Z_1^T)^{(l+1)} \propto \exp \left\langle \log p(\Omega_b^1 | \varphi_1^T, Z_1^T) \right\rangle_{-\Omega_m}^{(l)},
\]

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

**Theorem**

*If the complete-data likelihood \( p(\varphi_1^T, Z_1^T | \Omega_b^1) \) is part of the exponential family [...] and if the hidden and parameter priors distributions \( p(\varphi_1^T) \) and \( p(\Omega_b^1) \) are conjugate to this complete-data likelihood, the corresponding variational approximate posterior distributions that maximize \( F \), \( q^*(\varphi_1^T | Z_1^T) \) and \( q^*(\Omega_b^1 | Z_1^T) \), are of the same distributional form than respectively the prior distributions \( p(\varphi_1^T) \) and \( p(\Omega_b^1) \).*
Full model specification

\[
\begin{align*}
\varphi_{t+1} &= A\varphi_t + w_t \quad w_t \sim \mathcal{N}_k(0, Q) \\
Z_t &= C_t\varphi_t + \nu_t \quad \nu_t \sim \mathcal{N}_d(0, R)
\end{align*}
\]

where

\[
\begin{aligned}
\varphi_t &= \text{vec}[\varphi_1(t), \varphi_2(t), \ldots, \varphi_p(t)]' \\
Z_t &= \begin{pmatrix} Z_{t}^{\text{OFC}} \\ Z_{t}^{\text{AMY}} \end{pmatrix} \\
C_t\varphi_t &= \sum_{j=1}^{p} \varphi_j(t) \begin{pmatrix} Z_{t-j}^{\text{OFC}} \\ Z_{t-j}^{\text{AMY}} \end{pmatrix}
\end{aligned}
\]
Bayesian State Space Model

We have

- A reliable estimation of the dynamical VAR coefficients $\varphi^T_I$ for multiple trials
- A suitable model order selection criterion
Multiscale Bayesian State Space Model

We need

- to be frequency specific
- to capture short- and long-range causal dependencies between signals
Multiscale Bayesian State Space Model

BSS model

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

□ Idea : decompose the past values of the signals \(Z_t^{\text{OFC}}\) and \(Z_t^{\text{AMY}}\) contained in \(\{C_t\}_{t=1}^T\) in wavelets and use \(\{C^w_t\}\) as prediction instead of matrices \(\{C_t\}_{t=1}^T\)

□ Haar à trous wavelet for “pure” prediction

□ Construct \(C^w_t\) as

\[
C^w_t = \{w^\text{AMY}_{j,t-1-2j(k-1)}\}_{j=1:J,k=1:p_J}, \{s^\text{AMY}_{J,t-1-2J(k-1)}\}_{k=1:p_{J+1}}, \\
\{w^\text{OFC}_{j,t-1-2j(k-1)}\}_{j=1:J,k=1:p_J}, \{s^\text{OFC}_{J,t-1-2J(k-1)}\}_{k=1:p_{J+1}}.
\]

⇒ MSBSS model

\[
\begin{aligned}
\varphi_{t+1} &= A \varphi_t + w_t & w_t &\sim \mathcal{N}_k(0, Q) \\
Z_t &= C^w_t \varphi_t + v_t & v_t &\sim \mathcal{N}_d(0, R)
\end{aligned}
\]
Multiscale Bayesian State Space Model

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

\( \Rightarrow \) All results in BSS model can be applied
\( \Rightarrow \) à trous extension \( \Rightarrow \) generalisation of the BSS model
Multiscale Bayesian State Space Model

Multiresolution methodology benefits:

- Wavelets coefficients $w_j$ directly related to a **specific frequency band** → frequency specific in the modelisation of the causal relationship
- Capture **short- and long-range dependencies** between signals with only **few parameters** → Avoids the choice of the **time interval** ($\nu$) **time-lag** $\tau$
- Much more robust to arbitrary recording sampling frequency
- Simple and suitably **interpretable** time-frequency Granger-causality statistic (Chicharro, 2011)

![Wavelet Coefficients](image)

![Time-Frequency Graph](image)
Assessment of accuracy: Granger-causality detection

Simulated signals with slowly-varying parameters.
Order 4, length 500, Trials \{1; 10\}, causal parameter 1.

![Graphs showing the assessment of accuracy for Granger-causality detection.](image)
Application

- iEEG data recorded during psychological experimental situations
- 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
- Two experimental conditions: *anger* and *neutral*
Application

Significance levels in $\log_{10}$ scale for the overall causality statistic and for each scale for *anger*.
Application : Discussion

- Multiple testing : the frequency dimension
  - Hierarchically testing
  - usually relaxed in Neuroscience (GCC, PDC, DTF)

- Multiple testing : the time dimension
  - Threshold : $\alpha$ 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)

- Number of scales and model order per scale selection
References


Multiscale Bayesian State Space Model for Granger Causality Analysis, with Application to 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

ICORS, Wollongong, July 2017