Sparse inverse time correlation model for signal identification in fNIRS data

David Causeur
IRMAR, UMR 6625 CNRS
Agrocampus, Rennes, France

Joint work with Ching-Fan Sheu
(Institute of Education, NCKU, Tainan, Taiwan)

The Phonological Neighborhood Density study

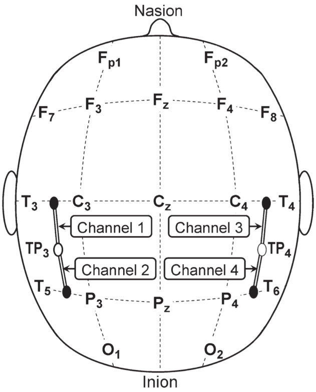
Phonological Neighborhood Density (PND) of a word : number of words that can be generated by replacing a phoneme with another phoneme in the same position.

Examples : SHORT has a high PND, PROOF has a low PND

Words with high PND : [(Chen *et al.*, 2011)]

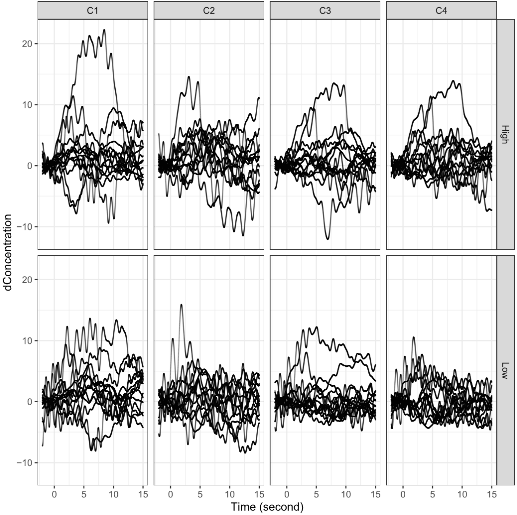
- are recognized more slowly ;
- elicits greater changes in blood oxygenation in the left than in the right hemisphere of the brain.

The Phonological Neighborhood Density study



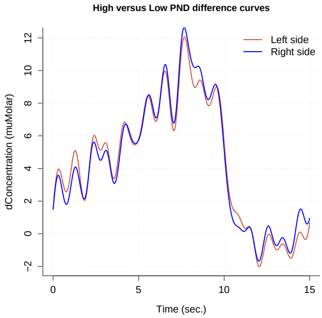
The Phonological Neighborhood Density study

PND $\times$ channels hemodynamic curve data for 14 subjects.



The Phonological Neighborhood Density study

Is the Low PND versus High PND difference curve the same in the left and right side of the brain ?



The Phonological Neighborhood Density study

Is the Low PND versus High PND difference curve the same in the left and right side of the brain ?

Within-subjects functional ANOVA design

- Two factors : PND condition and Brain side
- Test for the PND condition $\times$ Brain side interaction effect
- Responses are high-resolution hemodynamic curves

The Phonological Neighborhood Density study

Two issues

- **Signal detection** : is the effect curve non-zero somewhere within the whole time frame ?
- **Signal identification** : in which time intervals is the effect curve non-zero ?

Functional ANOVA

The linear function-to-scalar regression framework

- Hemodynamic response curve : $\boldsymbol{Y} = \big(Y(t_1), \dots, Y(t_p)\big)'$
- PND condition, Brain side, Subject effects, Interactions : $\boldsymbol{x} = (x_1, \dots, x_m)'$

$$\boldsymbol{Y} = \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{\varepsilon} \sim \mathcal{N}(0; \boldsymbol{\Sigma})$$

In the PND study design

- $p = 3005$ time points (200 samples/sec.).
- $n = 14 \times 4 \times 2 = 112$ response curves
- $m = 43$

Functional ANOVA

Functional ANOVA

- At each time point, a F-test statistic : $\boldsymbol{F} = \left(F_{t_1}, \ldots, F_{t_p}\right)$
- The global test statistic aggregates the $F_{t_k},\ k = 1, \ldots, p$
  - Sum-based aggregations
  - Max-based aggregations

See `R` package `fdANOVA` implementing 12 fANOVA tests for one-way designs

Górecki, T., Smaga, Ł. (2019) fdANOVA : an R software package for analysis of variance for univariate and multivariate functional data. *Comput Stat* **34**, 571–597.

Handling time-dependence in Functional ANOVA

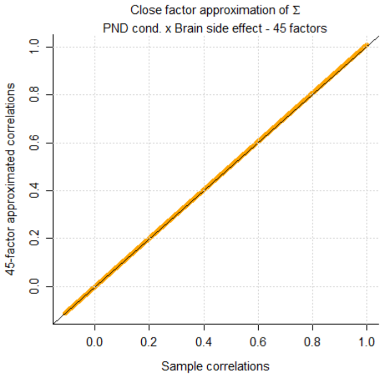
Wald-type testing in Functional ANOVA (see `R` package `ERP`)

- $\hat{\boldsymbol{\delta}} = \left(\hat{\delta}_{t_1}, \ldots, \hat{\delta}_{t_p}\right)$, OLS estimate of the effect curve ;
- Wald-type test statistic : $F = \hat{\boldsymbol{\delta}}' \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\delta}}}^{-1} \hat{\boldsymbol{\delta}} = \left(\hat{\boldsymbol{\delta}}' \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\delta}}}^{-1/2}\right) \left(\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\delta}}}^{-1/2'}\right) \hat{\boldsymbol{\delta}}$;
- $\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\delta}}}^{-1} = n\hat{\boldsymbol{\Omega}} \otimes \boldsymbol{S}_{xx}$, where $\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\Sigma}}^{-1}$;
- $\hat{\boldsymbol{\Sigma}}$: low-rank q-factor approximation of the sample variance-covariance matrix of $\hat{\varepsilon}$.

Handling time-dependence in Functional ANOVA

Low-rank factor model for Σ

$$\Sigma_{p\times p} = \Psi_{\text{diag}} + B_{p\times q} B'_{p\times q}$$



Handling time-dependence in Functional ANOVA

"Lightening" of pointwise F-tests

- $q = 0$: sum of correlated (coloured) pointwise F-tests
- q as large as possible : sum of whitened pointwise F-tests

PND condition x Brain side interaction effect

	$q = 0$	$q = 17$	$q = 45$
p-value	0.869	0.022	0.209

Handling time-dependence in Functional ANOVA

Optimal handling of time-dependence depends on the interplay between the dependence structure and the pattern of association signal

Causeur, D., Sheu, C. F., Perthame, E. and Rufini, F (2020). A functional generalized F-test for signal detection with applications to event-related potentials significance analysis. *Biometrics*. **76**(1), 246-256.

Searching for peaks

Time points with nonzero regression parameters

- Multiple testing viewpoint : strong dependence induces unstability [Sheu *et al.*, AoAS, 2016]
- ℓ_1-penalized estimation of the effect curve

ℓ_1-penalized estimation [(Rothman *et al.*, 2010)]

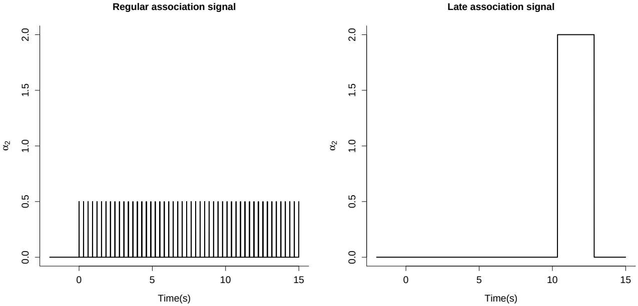
$$\mathcal{D}(\boldsymbol{\beta}; \boldsymbol{\Omega}, \kappa) = -n \log \det(\boldsymbol{\Omega}) + \sum_{i=1}^{n} (\boldsymbol{Y}_i - \boldsymbol{x}_i' \boldsymbol{\beta})' \boldsymbol{\Omega} (\boldsymbol{Y}_i - \boldsymbol{x}_i' \boldsymbol{\beta}) + \kappa ||\boldsymbol{\beta}||_1,$$

where $\kappa > 0$ is a penalty parameter and $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$.

Searching for peaks

How does the choice of $\Omega = \Sigma^{-1}$ affect estimation ?

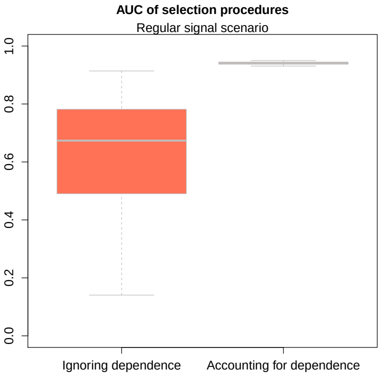
- Two options for High vs Low PND difference curve :

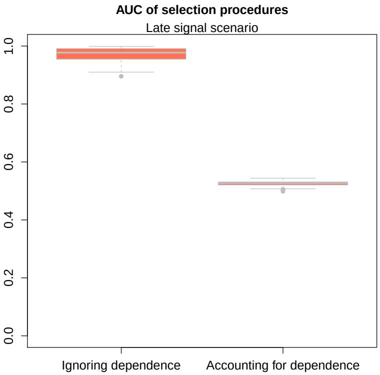


- Two options for Σ in $\mathcal{D}(\beta; \Sigma, \kappa)$:
  - $\Sigma = D_\sigma^2$ diagonal ;
  - Close factor approximation of the sample estimate of Σ.

Searching for peaks

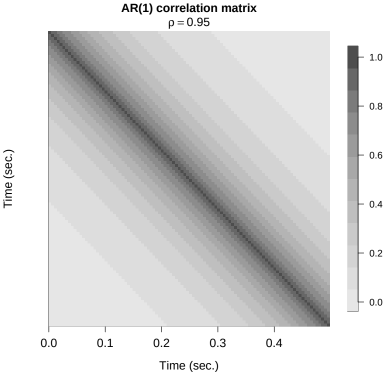
Focus on feature selection

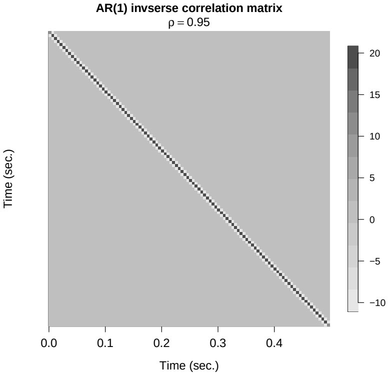




Sparse inverse time-correlation model

Illustration by the AR(1) correlation model





Σ dense

Ω sparse

Sparse inverse time-correlation model

New parametrization

$$\begin{cases} \varphi_{\text{diag}} & = & \Psi^{-\frac{1}{2}} \\ \boldsymbol{\theta}_{p\times q} & = & \Psi^{-\frac{1}{2}}\boldsymbol{B}(\boldsymbol{I}_q+\boldsymbol{B}'\Psi^{-1}\boldsymbol{B})^{-\frac{1}{2}} \end{cases} \quad \Rightarrow \quad \boldsymbol{\Omega}=\boldsymbol{\Sigma}^{-1}=\varphi(\boldsymbol{I}_p-\boldsymbol{\theta}\boldsymbol{\theta}')\varphi$$

Doubly-penalized deviance minimization

$$\mathcal{D}(\boldsymbol{\beta};\boldsymbol{\Omega},\kappa_1,\kappa_2) \quad = \quad \mathcal{D}(\boldsymbol{\beta};\boldsymbol{\Omega},\kappa_1)+\kappa_2\sum_{r=1}^{p}\sum_{s=1}^{q}|\theta_{rs}|^k, \text{ with } k=1 \text{ or } 2,$$

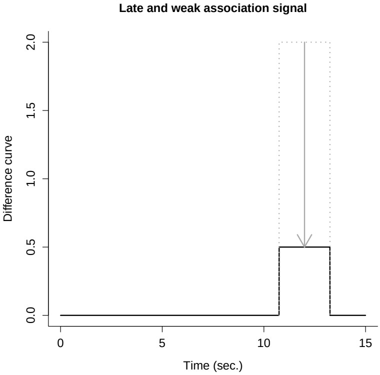
and $\kappa_2 > 0$ is the second penalty parameter.

See Witten, D. and Tibshirani, R. (2009) Covariance-regularized regression and classification for high-dimensional problems, *Journal of the Royal Statistical Society*, Series B **71**(3) : 615-636. For $p=1$, m large, no dimension reduction.

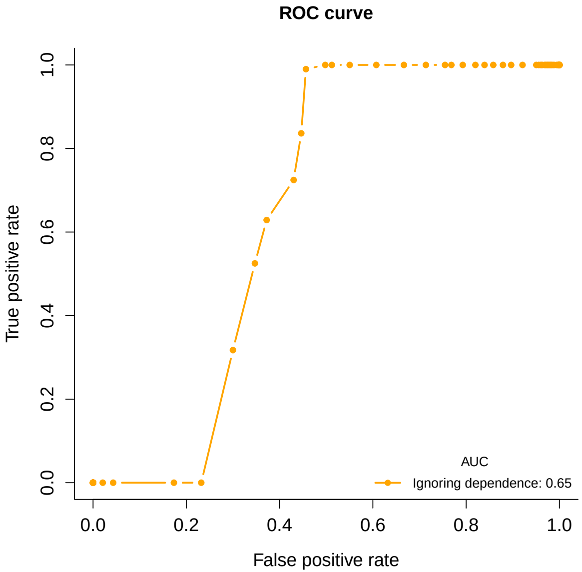
Illustration

PND study data-driven simulation : two-group comparison

- $n = 112$, $m = 3005$ as in the PND study
- Time-correlation : 45-factor approximation of Σ
- Association signal : weak on a late interval

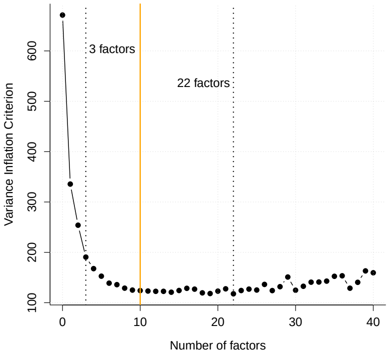


Illustration



Illustration

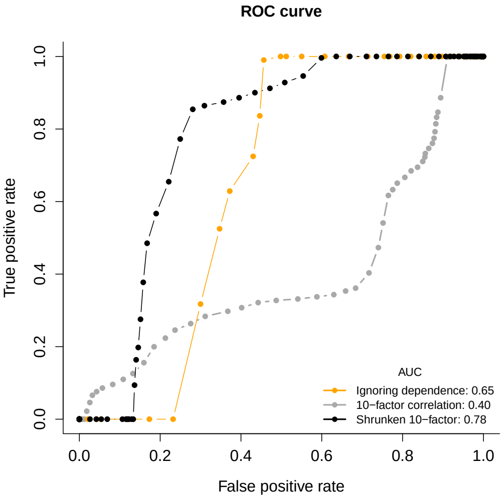
Low-rank model for time-correlation



Friguet, C., Kloareg, M. and Causeur, D. (2009). A factor model approach to multiple testing under dependence. *Journal of the American Statistical Association*. **104** (488), 1406–1415.

Illustration

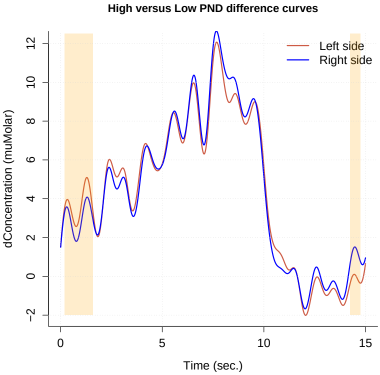
Sparse-inverse time correlation model (ridge)



Illustration

PND study : PND condition x Brain side interaction effect

- 10 factors
- κ_1 and κ_2 minimize CV'd errors



Conclusion

Three take-home messages

- Handling dependence is not a two-option issue
- The best handling depends on the true association signal and the dependence pattern
- Optimizing the handling of dependence is not only a high-dimensional issue