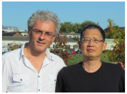
Analysis of Variance for functional data using the R package ERP

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

Collaborators

Ching-Fan Sheu
National Cheng Kung University, Tainan, Taiwan

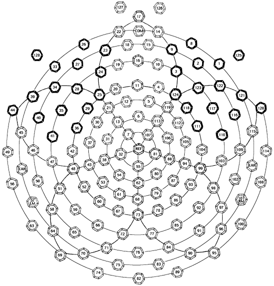


Émeline Perthame
Institut Pasteur, Paris, France

Rencontres R 2018, Rennes

The instrument: a 128-channel geodesic sensor net

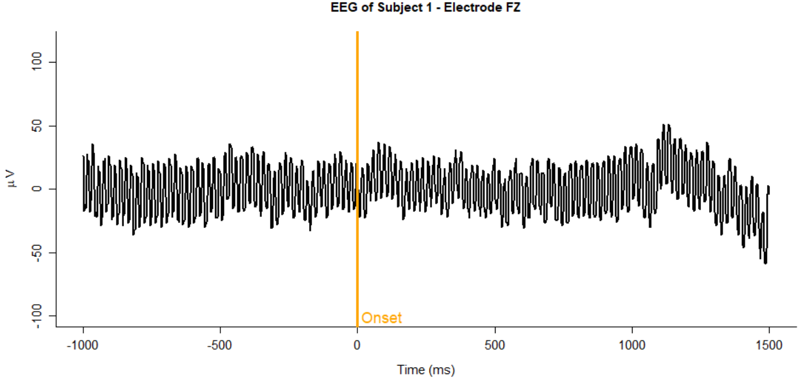
- Electroencephalography (EEG) records electrical activity at scalp locations over time.
- The recorded EEG traces, which are time locked to external events, are averaged to form the event-related (brain) potentials (ERPs).



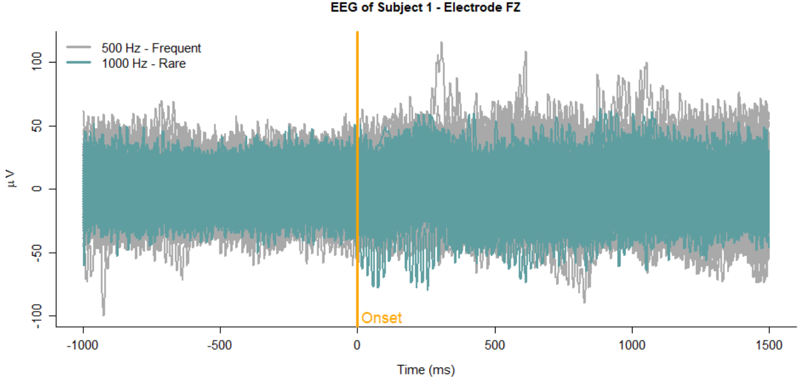
Auditory oddball experiment

- Two auditory stimuli are presented to subjects
  - A stimulus (500Hz) occurring frequently
  - A stimulus (1000Hz) occurring infrequently
- ERPs are recorded on a 1000ms interval after the onset.

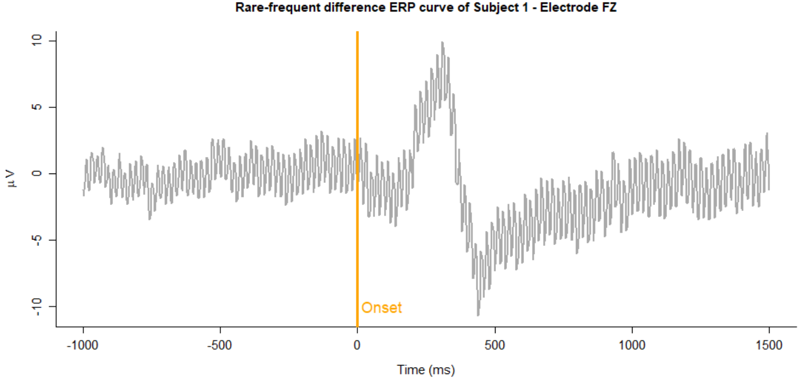
Auditory oddball experiment



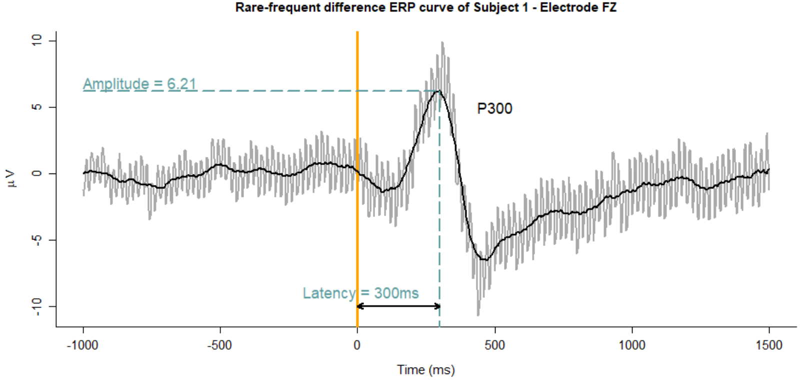
Auditory oddball experiment



Auditory oddball experiment



Auditory oddball experiment



Testing issues in ERP studies

- **Signal detection**

  Any significant difference between experimental conditions?

  Functional Analysis of Variance tests

- **Signal identification**

  Which time intervals are significant?

  Control for false discoveries over timewise tests

Functional ANOVA

Multivariate Analysis of Variance (MANOVA) model

$$Y_{ijt} = \mu_t + \alpha_{it} + \varepsilon_{ijt},$$

where Y_{ijt} is the ERP for the jth subject in condition i at time t

Cuevas, A., Febrero, M., & Fraiman, R. (2004). An ANOVA test for functional data. Computational Statistics & Data Analysis, 47(1), 111–122. - R package `fda.usc`

Functional Analysis of Variance (fANOVA) model

$$Y_{ijt} = \sum_{s=1}^{S} m_s \varphi_s(t) + \sum_{s=1}^{S} a_{is} \varphi_s(t) + \varepsilon_{ijt},$$

where $\varphi_s(.),\ s = 1, \ldots, S$ are B-splines.

Bugli, C. and Lambert, P. (2006). Functional ANOVA with random functional effects: an application to event-related potentials modelling for electroencephalograms analysis. Statistics in Medicine 25(21), 3718–3739.

Functional ANOVA

A 'whole time frame' expanded linear model for ERP curves

Functional ANOVA model

$$Y = (\varphi \otimes \mathbb{1}_n)\mathbf{m} + (\varphi \otimes X)\mathbf{a} + \varepsilon,$$

where $Y = (Y'_{t_1}, Y'_{t_2}, \ldots, Y'_{t_T})'$.

- If $\varphi = I_T$, then this is just the MANOVA model

  (Shen and Faraway, 2004), package `fdANOVA`

- If rank$(\varphi) < T$, this is a 'generalized additive model'

  (Wood, 2017), package `mgcv`

Functional ANOVA

Illustration using the demo dataset in package `ERP`

R script

```
> require(ERP)
> data(impulsivity)                # Illustrative ERP dataset (Shen et al., 2014)
> dim(impulsivity)
[1] 144 505
> head(impulsivity[,1:6])
```

	Channel	Subject	Group	Condition	T_0	T_2
10	FCZ	S11	High	Success	0.08391917	-0.02603725
15	CZ	S11	High	Success	0.33112752	0.27123761
20	CPZ	S11	High	Success	0.71194828	0.72240525
40	FCZ	S11	High	Failure	0.68859053	0.60859489
45	CZ	S11	High	Failure	-0.04983616	-0.15963431
50	CPZ	S11	High	Failure	-0.30644041	-0.49152860

Functional ANOVA

Stack the ERP curves into a single vector

R script

```
> require(reshape2)                # Flexibly reshape data (by H. Wikham, 2017)

> impulsivity.melt = melt(impulsivity,value.name="erp")

> dim(impulsivity.melt)

[1] 72144 6

> head(impulsivity.melt)

  Channel Subject Group Condition variable         erp
1     FCZ     S11  High   Success      T_0  0.08391917
2      CZ     S11  High   Success      T_0  0.33112752
3     CPZ     S11  High   Success      T_0  0.71194828
4     FCZ     S11  High   Failure      T_0  0.68859053
5      CZ     S11  High   Failure      T_0 -0.04983616
6     CPZ     S11  High   Failure      T_0 -0.30644041
```

Functional ANOVA

Add a numeric 'Time' variable

R script

```
> time_pt_char = as.character(impulsivity.melt$variable)
> time_pt = substring(time_pt_char,first=3,last=nchar(time_pt_char))
> impulsivity.melt$Time = as.numeric(time_pt)
> head(impulsivity.melt)
```

	Channel	Subject	Group	Condition	variable	erp	Time
1	FCZ	S11	High	Success	T_0	0.08391917	0
2	CZ	S11	High	Success	T_0	0.33112752	0
3	CPZ	S11	High	Success	T_0	0.71194828	0
4	FCZ	S11	High	Failure	T_0	0.68859053	0
5	CZ	S11	High	Failure	T_0	-0.04983616	0
6	CPZ	S11	High	Failure	T_0	-0.30644041	0

Functional ANOVA

Implement the non-parametric ANOVA test using function `mgcv`

R script

```
> impulsivity.bam = bam(erp ~ s(Time,bs="cr")+s(Time,by=Subject,bs="cr")+
+ s(Time,by=Condition,bs="cr"),data=impulsivity.melt)

> impulsivity.bam0 = bam(erp ~ s(Time,bs="cr")+s(Time,by=Subject,bs="cr"),
+ data=impulsivity.melt)

> anova(impulsivity.bam0,impulsivity.bam,test="F")
```

Analysis of Deviance Table

Model 1: erp ~ s(Time, bs = "cr") + s(Time, by = Subject, bs = "cr")

Model 2: erp ~ s(Time, bs = "cr") + s(Time, by = Subject, bs = "cr") + s(Time, by = Condition, bs = "cr")

	Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)
1	71926	488740				
2	71917	466448	9.004	22293	381.76	$< 2.2e-16$

Functional ANOVA

Can we really trust this very low p-value?

Functional ANOVA

After a random permutation of the 'Condition' labels

R script

```
> impulsivity2 = impulsivity
> impulsivity2$Condition = impulsivity$Condition[sample(1:nrow(impulsivity))]
> impulsivity2.melt = melt(impulsivity2,value.name="erp")
> impulsivity2.melt$Time = impulsivity.melt$Time
```

Functional ANOVA

After a random permutation of the 'Condition' labels

R script

```
> impulsivity2.bam = bam(erp ~ s(Time,bs="cr")+ s(Time, by = Subject, bs = "cr")+
+ s(Time,by=Condition,bs="cr"),data=impulsivity2.melt)

> impulsivity2.bam0 = bam(erp ~ s(Time,bs="cr") + s(Time, by = Subject, bs = "cr"),
+ data=impulsivity2.melt)

> anova(impulsivity2.bam0,impulsivity2.bam,test="F")

Analysis of Deviance Table

Model 1: erp ~ s(Time, bs = "cr") + s(Time, by = Subject, bs = "cr")

Model 2: erp ~ s(Time, bs = "cr") + s(Time, by = Subject, bs = "cr") + s(Time, by =
Condition, bs = "cr")
```

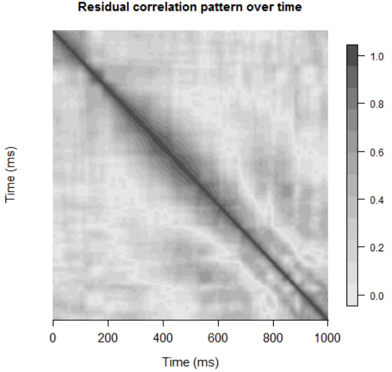
	Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)
1	71926	488740				
2	71917	486554	9.1512	2186	35.31	3.613e-14

Functional ANOVA

Functional ANOVA model

$$Y = (\varphi \otimes \mathbb{1}_n)\mathbf{m} + (\varphi \otimes X)\mathbf{a} + \varepsilon,$$

where $\text{Var}(\varepsilon) = V_\varepsilon = [D_\sigma \mathbf{R} D_\sigma] \otimes I_n$.



Functional ANOVA under dependence

Functional ANOVA model

$$Y = (\varphi \otimes \mathbb{1}_n)\mathbf{m} + (\varphi \otimes X)\mathbf{a} + \varepsilon,$$

where $\text{Var}(\varepsilon) = V_\varepsilon = [D_\sigma \mathbf{R} D_\sigma] \otimes I_n$.

LRT obtained by whitening the residuals: $Y^\star = V_\varepsilon^{-1/2} Y$

based on a $q-$factor decomposition of R:

$$R = \Psi + \Lambda\Lambda', \text{ where } \Psi \text{ is diagonal and } \text{rank}(\Lambda) = q.$$

Note that:

$$V_\varepsilon^{-1/2} = [D_\sigma^{-1/2} \mathbf{R}^{-1/2} D_\sigma^{-1/2}] \otimes I_n,$$

where $\mathbf{R}^{-1/2}$ has a q-factor structure.

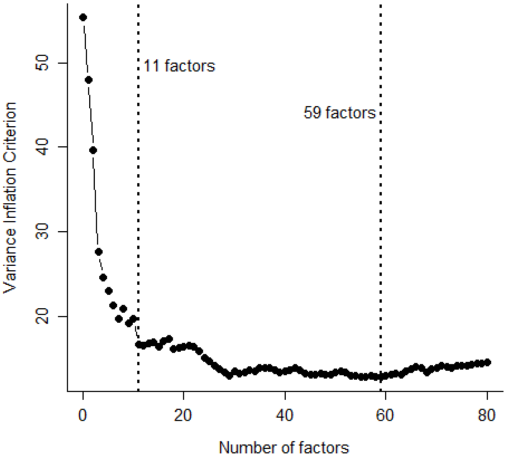
Functional ANOVA under dependence

R script

```
> erpdta = impulsivity[,-(1:4)]
> design = model.matrix( ~ Subject + Condition,data=impulsivity)
> design0 = model.matrix( ~ Subject,data=impulsivity)
> test = erpFtest(dta=erpdta,design=design,design0=design0,nbf=NULL)
```

Sheu, C-.F., Perthame, E., Lee, Y-.S., Causeur, D. (2016). Accounting for time dependence in large-scale multiple testing of event-related potential data. Annals of Applied Statistics. 10(1), 219–245.

Functional ANOVA under dependence



Functional ANOVA under dependence

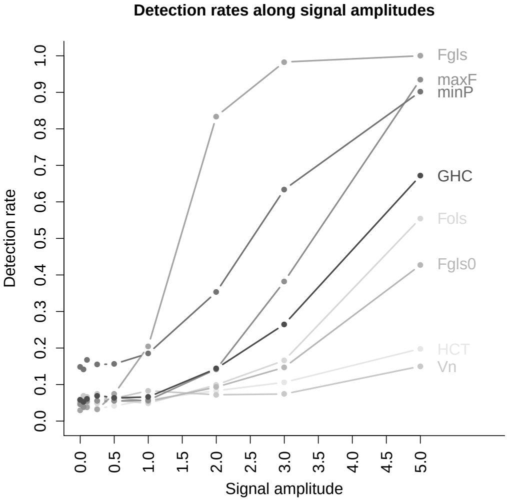
R script

```
> erpdta = impulsivity[,-(1:4)]
> design = model.matrix( ~ Subject + Condition,data=impulsivity)
> design0 = model.matrix( ~ Subject,data=impulsivity)
> test = erpFtest(dta=erpdta,design=design,design0=design0,nbf=NULL)
> test$pval
[1] 3.550709e-15

> erpdta2 = impulsivity2[,-(1:4)]
> design2 = model.matrix(~Subject+Condition,data=impulsivity2)
> test2 = erpFtest(dta=erpdta2,design=design2,design0=design0,nbf=NULL)
> test2$pval
[1] 0.1024562
```

Functional ANOVA under dependence

Power comparison study (based on data-driven simulations)



What else in ERP?

- **Signal identification methods**
  - Usual FDR-controlling multiple testing methods
  - Specific methods handling dependence (as Guthrie and Buchwald, 1991)
  - Factor-adjusted multiple testing
- **A long vignette with a complete demo**

R script

```
> vignette("ERP")
```

What else in ERP?

ERP: Significance Analysis of Event-Related Potentials

David Causeur, Ching-Fan Sheu

2018-06-26

1 Introduction

This vignette aims to demonstrate how to use the 'ERP' package for significance testing of event-related potentials (ERP) through the analysis of an example from a study investigating neural correlates of impulsive behavior (Shen, Lee, & Chen, 2014). ERPs are voltage changes along the scalp time-locked to some physical or mental events in the ongoing electrical brain activity recorded as electroencephalogram (EEG). ERPs are complex waveforms with highly correlated components in time corresponding to brief bursts of synchronized cortical activities. The current version of the 'ERP' package implements the adaptive factor-adjustment (AFA) procedure (Sheu, Perthame, Lee, & Causeur, 2016) and a generalized functional likelihood ratio test developed in Causeur, Sheu, Perthame, & Rufini (under review) for detecting and identifying ERP signals. The procedure proposed by Buthrie & Buchwald (1991) for significance testing of difference potentials is also included in the package for comparison.

Not yet in ERP

- **Data manipulation routines**
  - Averaging over channels in a same ROI
  - Identify peaks
  - Estimate latencies and amplitudes of peaks
- **Specific plotting routines**
  - 'Head plot' of effect curves
  - Map of effect on the scalp

Some helpful functions can be found in packages `erpR` and `erp.easy`

... and we are also working on it.