Kernel PLS Estimation of Single-Trial Event-Related Potentials

Roman Rosipal[1,2]

Leonard J. Trejo[1]

[1] *NASA Ames Research Center*

Computational Sciences Division, Moffett Field, CA

[2] *Department of Theoretical Methods*

Slovak Academy of Sciences, Bratislava, Slovak Republic

Overview

- We use smoothing regression techniques to estimate of ERPs
- $f(\mathbf{x}) = \sum_{d=1}^{D} w_d \phi_d(\mathbf{x}) + w_0$
  We need to estimate :
  D - a number of basis functions
  $\{\phi_d\}_{d=1}^{D}$ - a form of basis functions
  $\{w\}_{d=0}^{D}$ - weighting coefficients
- We address the problem of (temporal) correlated errors (noise)
- We compare kernel partial least squares (PLS) regression, smoothing splines (SS) and wavelet smoothing (WS) techniques on generated and real ERP data

Methods

Kernel PLS Regression

- data sets:

  $\mathbf{X}\ (n_{objects} \times N_{variables})$

  $\mathbf{Y}\ (n_{objects} \times M_{responses})$

  – *zero-mean*

- decomposition:

  $$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}$$

  $$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F}$$

  where:

  $\mathbf{T}, \mathbf{U}$ matrix of score variables (LV, components)

  $\mathbf{P}, \mathbf{Q}$ matrix of loadings

  $\mathbf{E}, \mathbf{F}$ matrix of residuals (errors)

- PLS - bilinear decomposition of $\mathbf{X}$ and $\mathbf{Y}$ with the aim to maximize

$$\begin{aligned} \max_{|\mathbf{r}|=|\mathbf{s}|=1}[cov(\mathbf{Xr},\mathbf{Ys})]^2 &= [cov(\mathbf{Xw},\mathbf{Yc})]^2 \\ &= var(\mathbf{Xw})[corr(\mathbf{Xw},\mathbf{Yc})]^2 var(\mathbf{Yc}) \\ &= [cov(\mathbf{t},\mathbf{u})]^2 \end{aligned}$$

- The weights $\mathbf{w},\mathbf{c}$ can be found using iterative NIPALS algorithm or by solving:

$$\mathbf{XX}^T\mathbf{YY}^T\mathbf{t} = \lambda\mathbf{t}$$
$$\mathbf{u} = \mathbf{YY}^T\mathbf{t}$$

- Nonlinear (kernel) PLS - linear PLS in feature spaces $\mathcal{F}_x, \mathcal{F}_y$

$$\mathbf{X} \rightarrow \mathbf{\Phi} \qquad \mathbf{Y} \rightarrow \mathbf{\Psi}$$

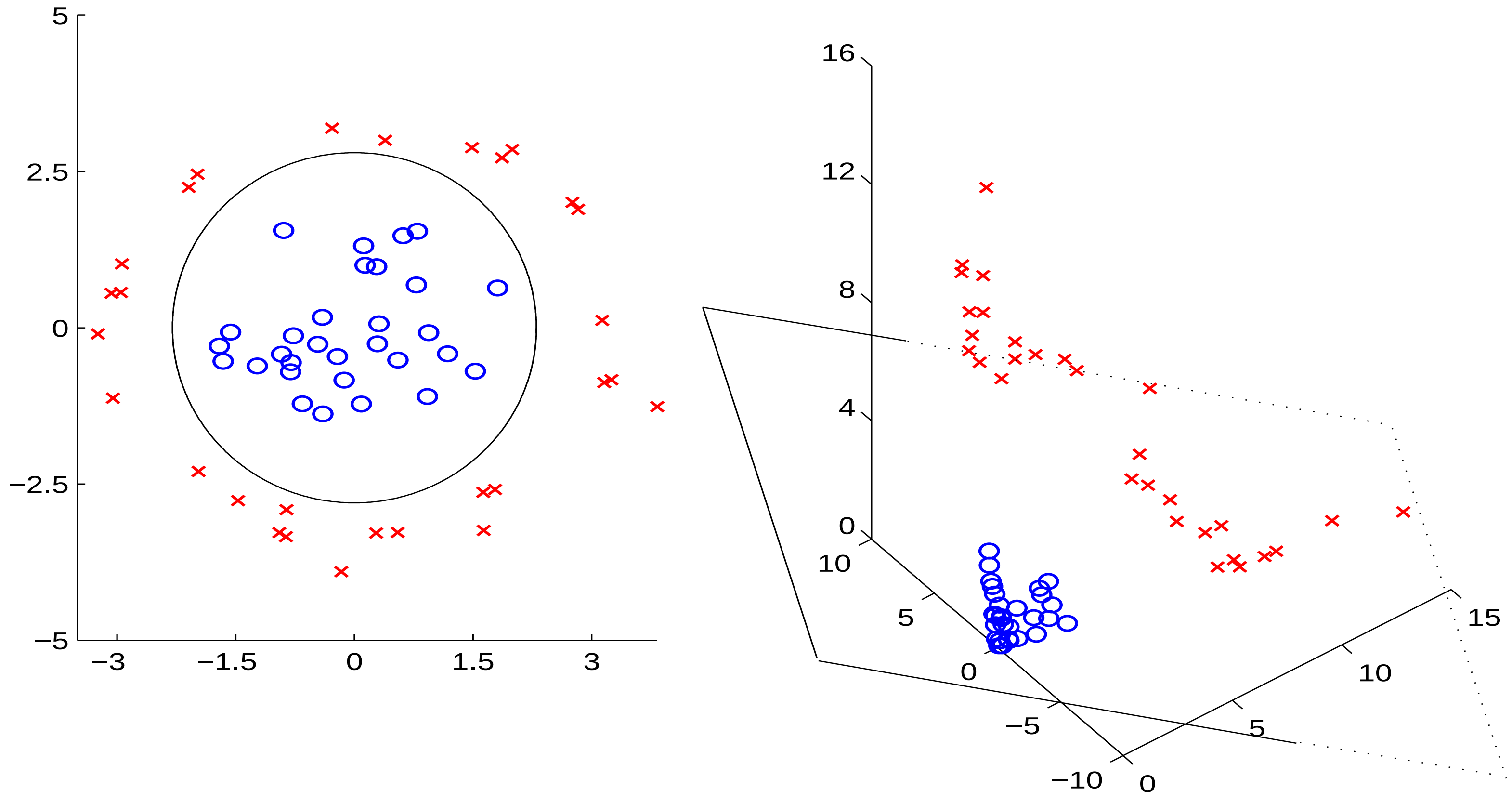
$\Rightarrow$

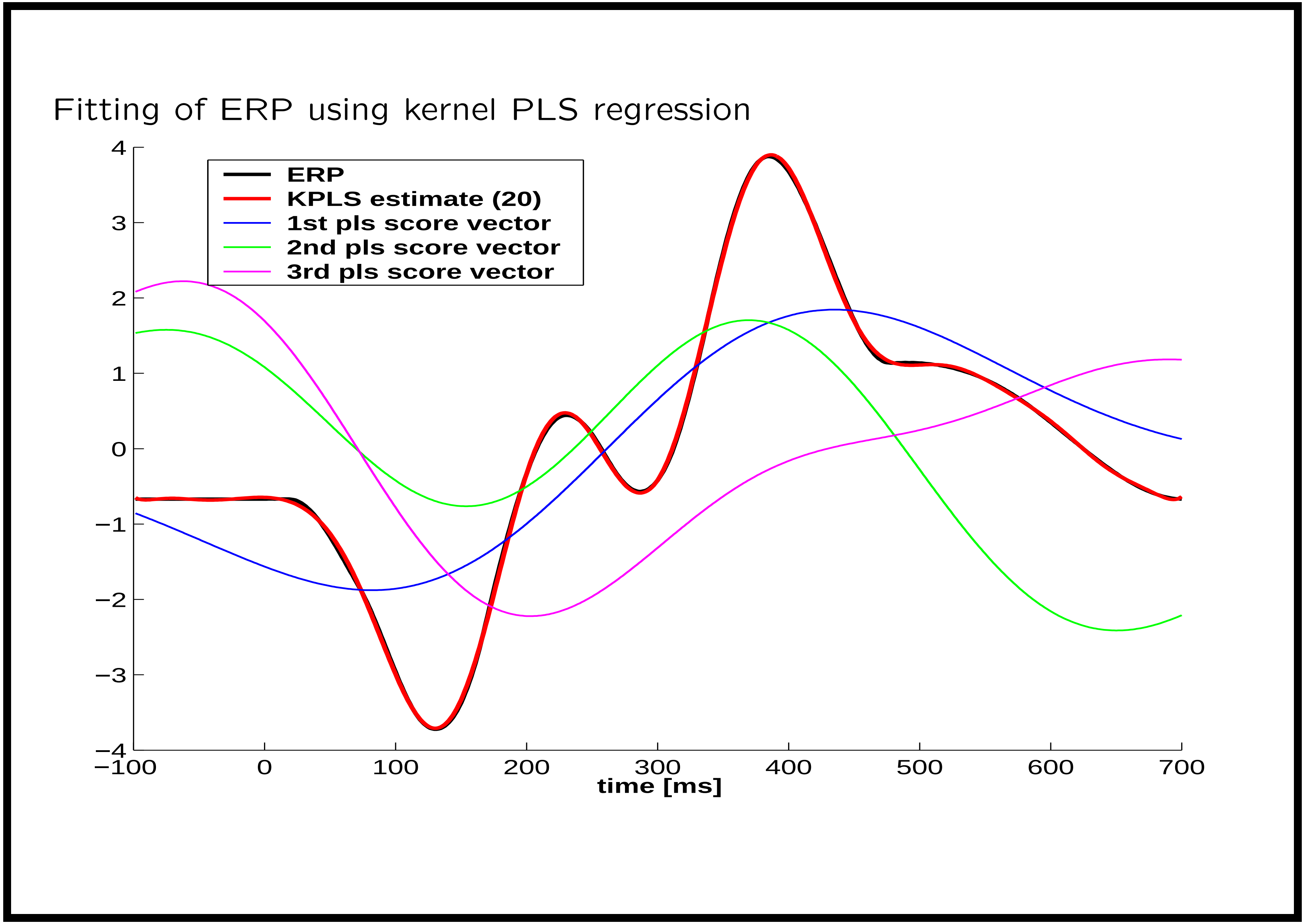
$$\mathbf{\Phi\Phi}^T\mathbf{\Psi\Psi}^T\mathbf{t} = \lambda\mathbf{t} \qquad \mathbf{u} = \mathbf{\Psi\Psi}^T\mathbf{t}$$

$\Rightarrow$

$$\mathbf{K}_\mathrm{x}\mathbf{K}_\mathrm{y} = \lambda\mathbf{t} \qquad \mathbf{u} = \mathbf{K}_\mathrm{y}\mathbf{t}$$

Example of nonlinear (kernel) mapping





Fitting of ERP using kernel PLS regression
ERP
KPLS estimate (20)
1st pls score vector
2nd pls score vector
3rd pls score vector
4
3
2
1
0
−1
−2
−3
−4
−100
0
100
200
300
400
500
600
700
time [ms]

Local Kernel PLS Regression

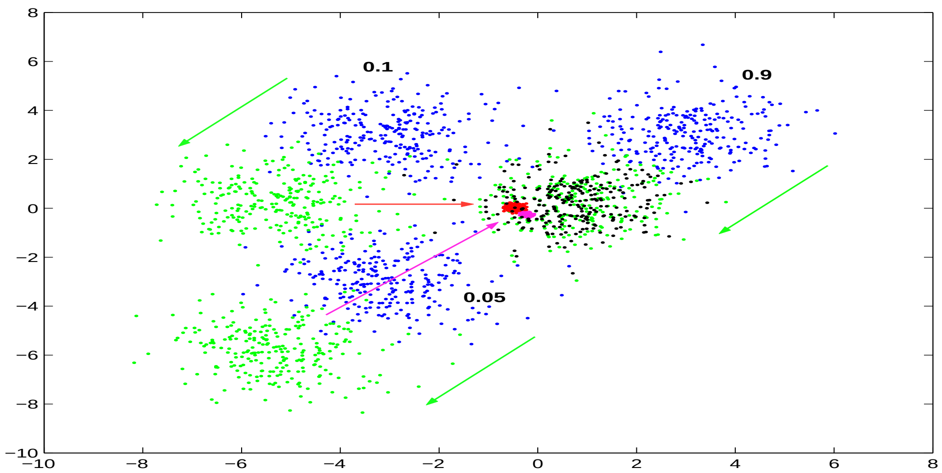
- Soft clustering: **r** - vector of weights

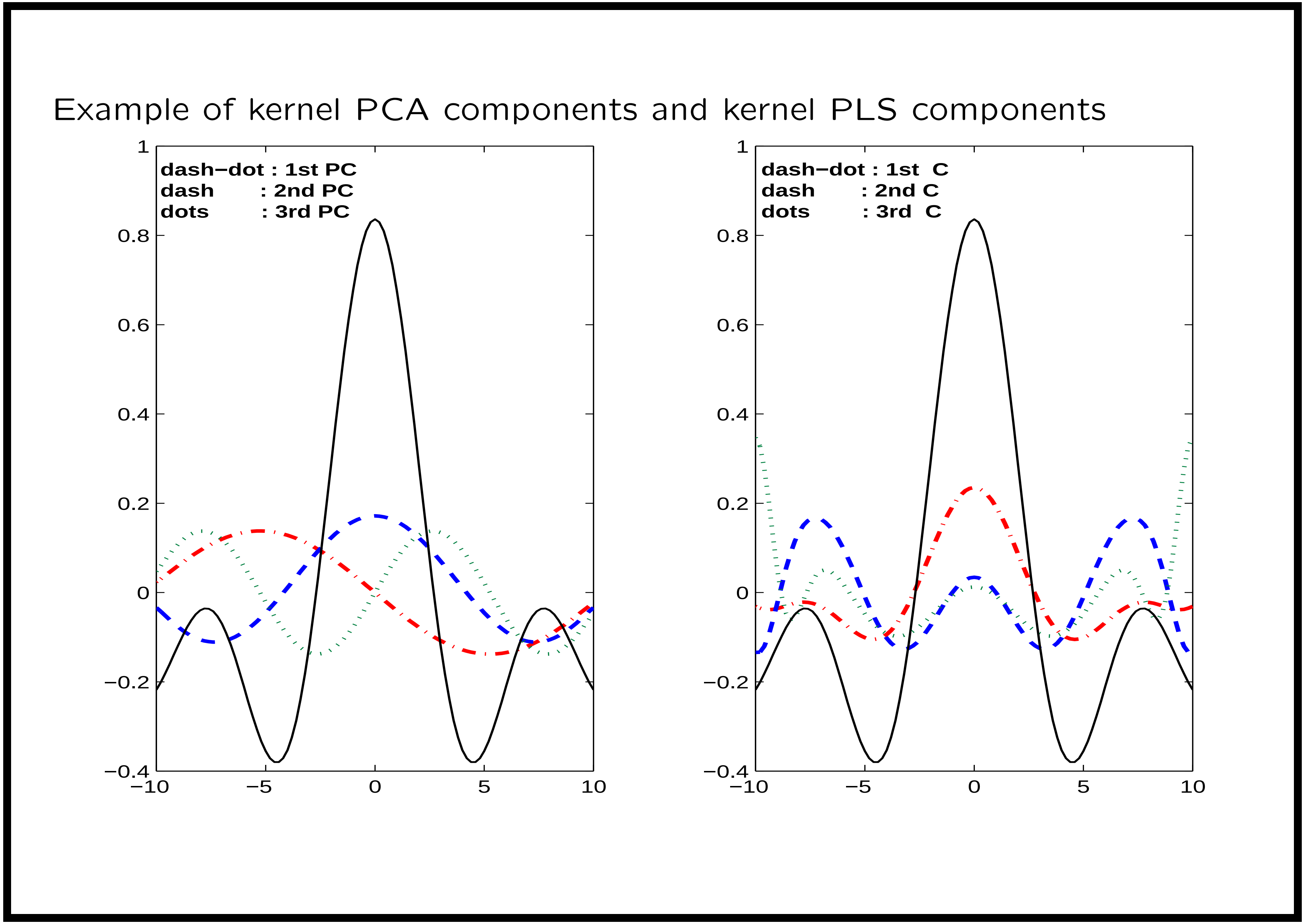
$$r_s = \sum \mathbf{r} \; ; \; \mathbf{R}_d = diag(\mathbf{r}) \; ; \; \mathbf{J} = ones(n,1) \; ; \; \mathbf{I} = eye(n)$$

$$\mathbf{\Phi}_r = \mathbf{R}_d(\mathbf{\Phi} - \mathbf{J}\frac{\mathbf{r}^T\mathbf{\Phi}}{r_s}) \;\; ; \;\; mean(\mathbf{\Phi}_r) = \mathbf{0}$$

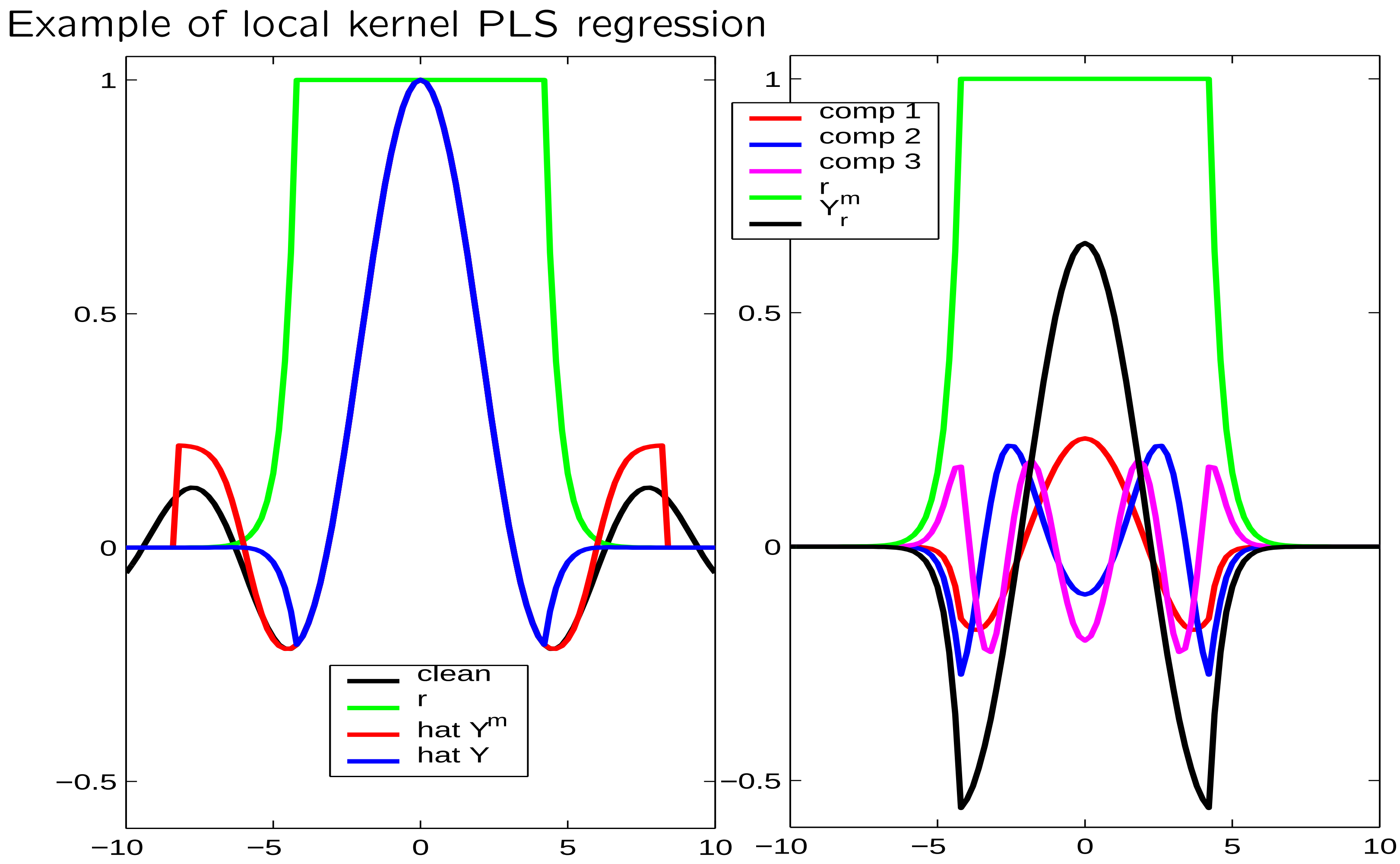
- Kernel variant:

$$\mathbf{K}_r = \mathbf{\Phi}_r\mathbf{\Phi}_r^T = \mathbf{R}_d(\mathbf{I} - \frac{\mathbf{J}\mathbf{r}^T}{r_s})\mathbf{K}(\mathbf{I} - \frac{\mathbf{J}\mathbf{r}^T}{r_s})^T\mathbf{R}_d \quad \mathbf{Y}_r = \mathbf{R}_d(\mathbf{Y} - \mathbf{J}\frac{\mathbf{r}^T\mathbf{Y}}{r_s})$$

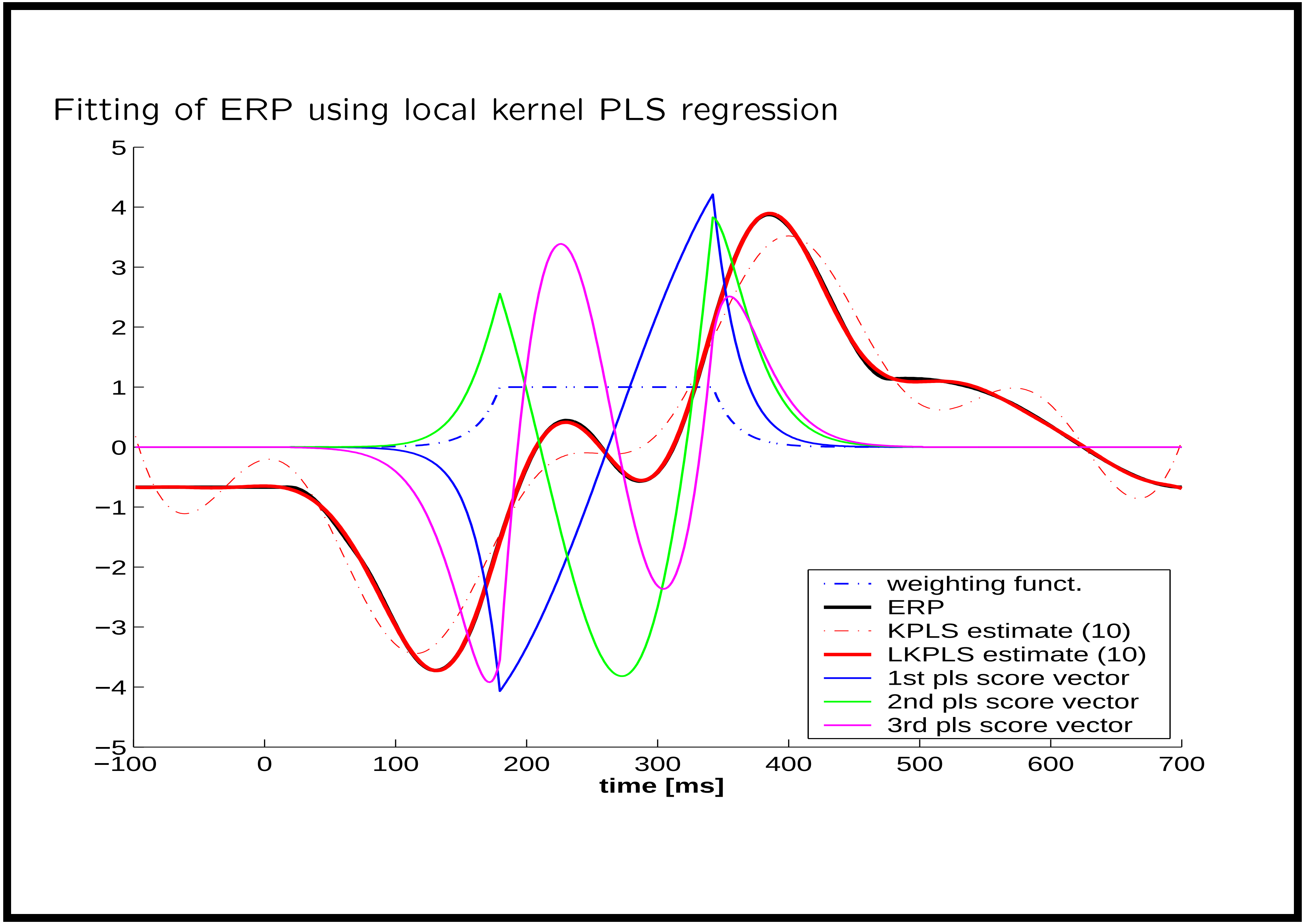




Example of kernel PCA components and kernel PLS components
dash−dot : 1st PC
dash : 2nd PC
dots : 3rd PC
dash−dot : 1st C
dash : 2nd C
dots : 3rd C
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−10
−5
0
5
10



Example of local kernel PLS regression
1
0.5
0
−0.5
−10
−5
0
5
10
clean
r
hat Y^m
hat Y
comp 1
comp 2
comp 3
r
Y^m_r



Fitting of ERP using local kernel PLS regression
5
4
3
2
1
0
−1
−2
−3
−4
−5
−100
0
100
200
300
400
500
600
700
time [ms]
weighting funct.
ERP
KPLS estimate (10)
LKPLS estimate (10)
1st pls score vector
2nd pls score vector
3rd pls score vector

Smoothing Splines

- $min_f(\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int_a^b (f^{(2)}(x))^2 dx \quad \lambda > 0 \Rightarrow$ *natural cubic splines* with knots at $x_i \; ; i = 1, \ldots, n$
- Complete basis $\rightarrow$ *shrink* the coefficients toward smoothing

Wavelet Smoothing

- Complete orthonormal basis $\rightarrow$ *shrink* and *select* the coefficients toward a **sparse** representation
- Wavelet basis is *localized in time and frequency*

Correlated Noise Estimate

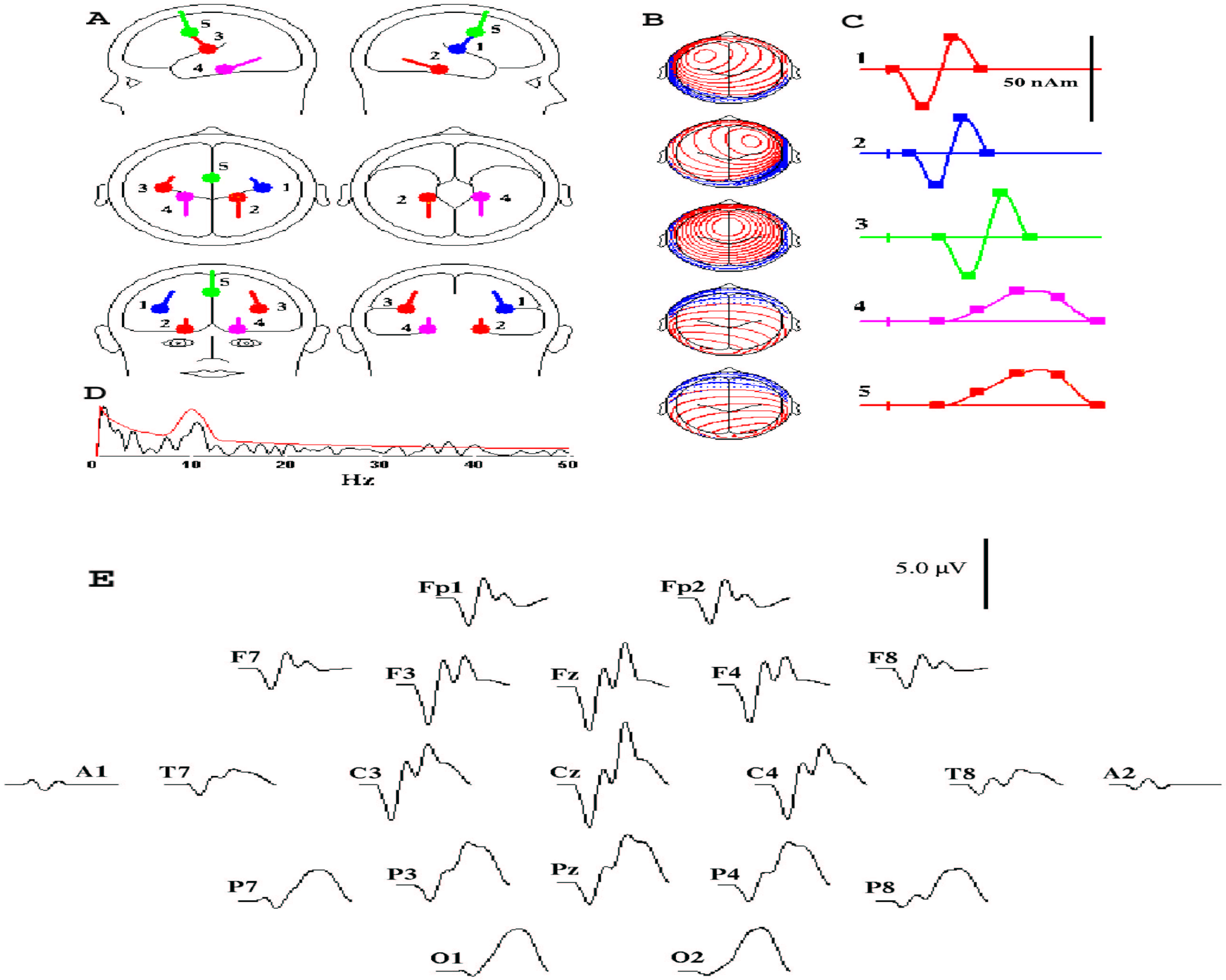
- measured signal$_i$ $=$ ERP$_i$ $+$ (on-going EEG $+$ measur. noise)$_i$
- We compute cov(measured signal$_i$ - avg(measured signal))

Data Construction

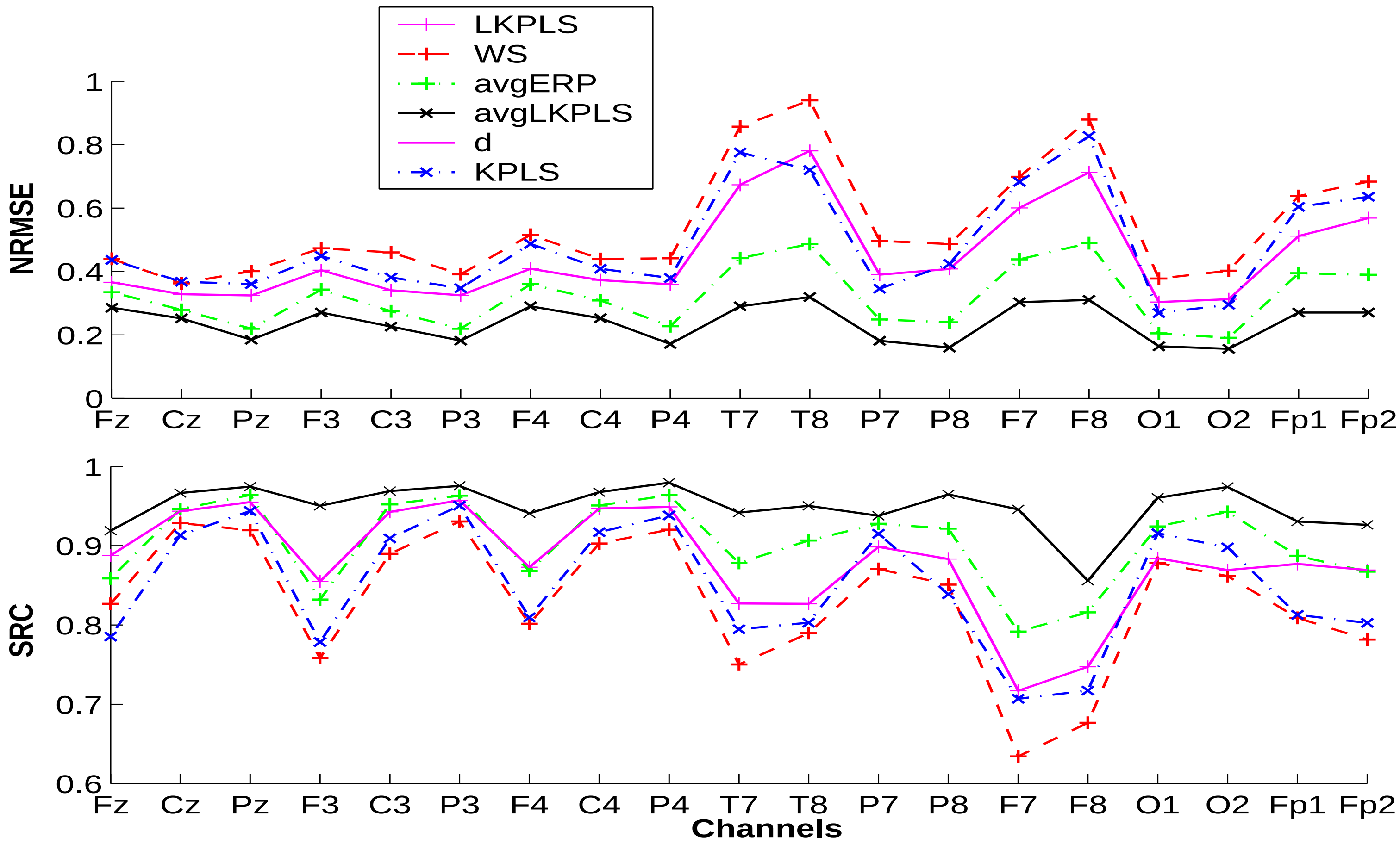
- Generated data:
  Event-Related Potentials (N1,P2,N2,P3)
  $$+$$
  relax state spatially distributed EEG signal + white Gaussian noise

- Real ERP data:
  ERPs recorded in an experiment of cognitive fatigue
  (see Len Trejo et. al., poster no. 36)

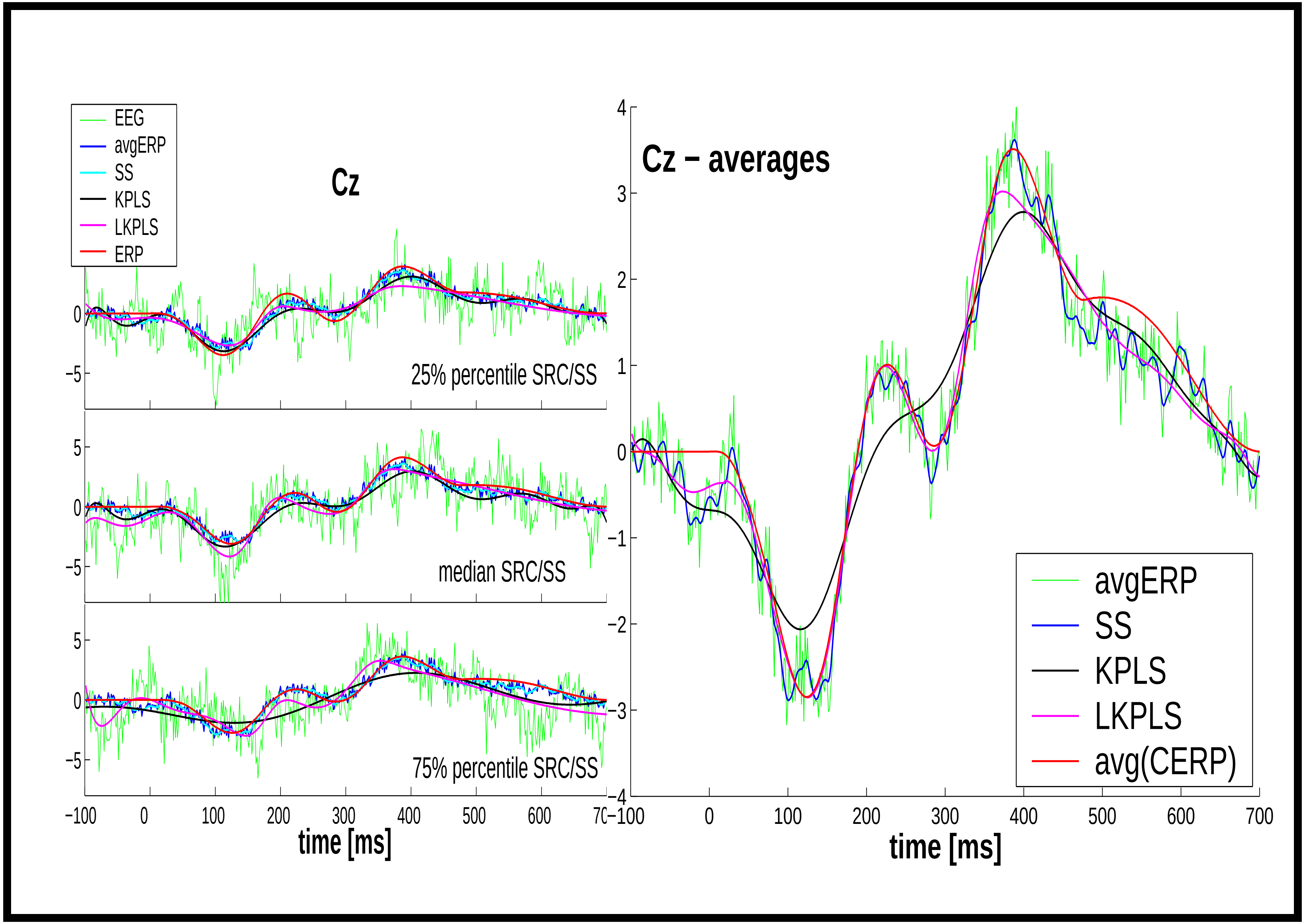
Generation of ERPs using BESA software



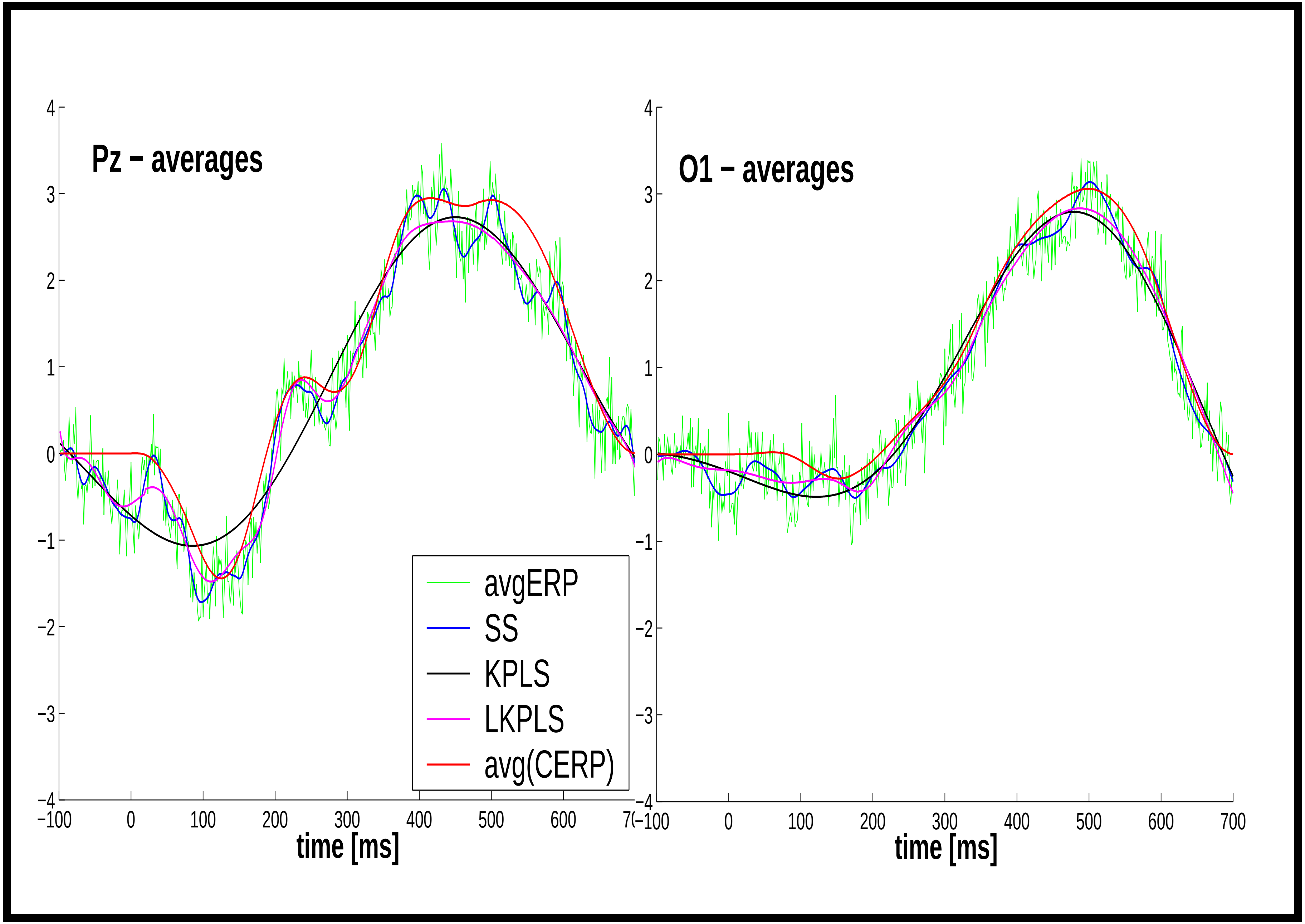
Results



Results on noisy event related potentials (ERPs)–20 different trials were used. Averaged SNR over the trials and electrodes was equal to 1.3dB (min=-7.1dB, max=6.4dB) and 512 samples were used. NRMSE - normalized root mean squared error; SRC - Spearman's rank correlation coefficient.

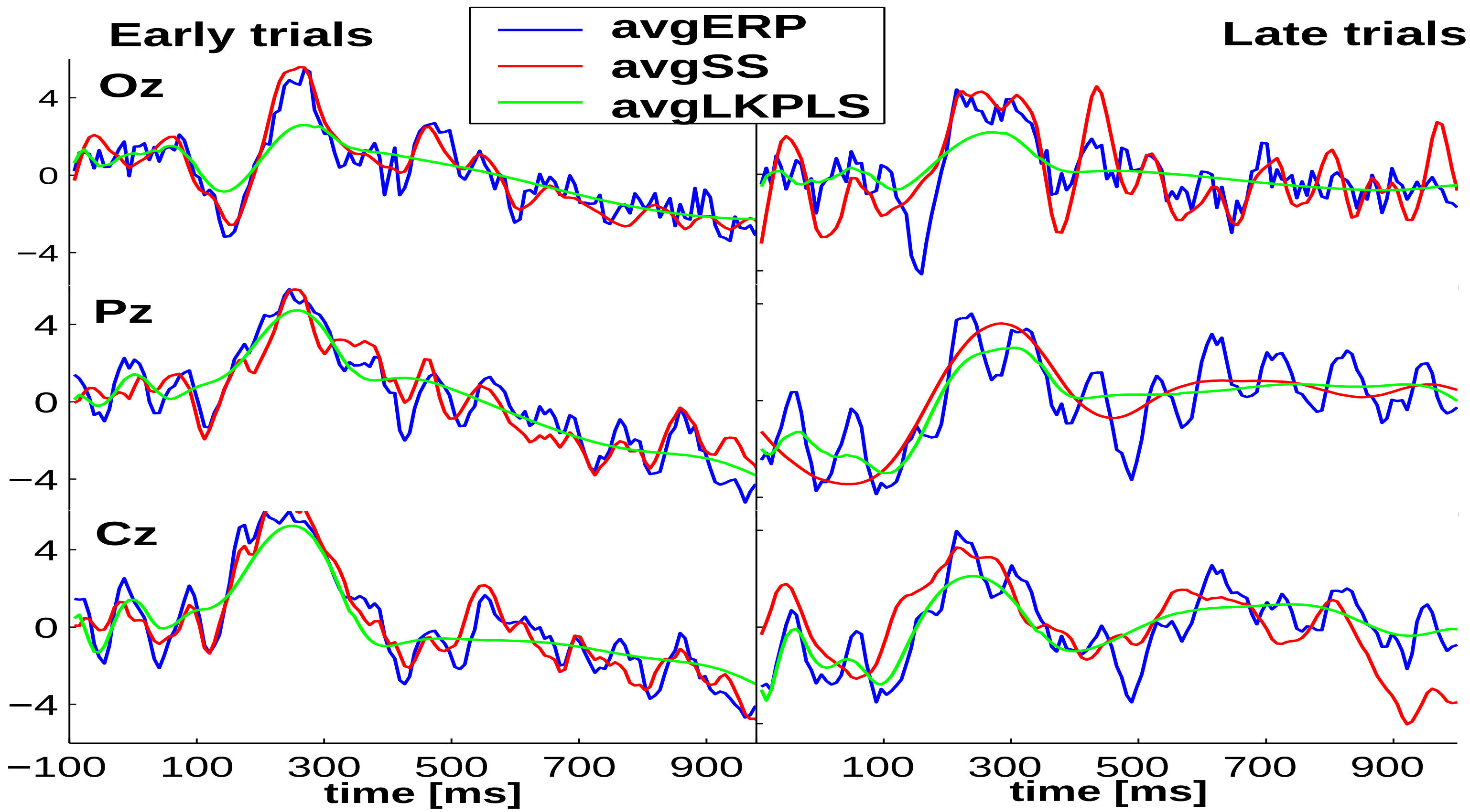


EEG
avgERP
SS
KPLS
LKPLS
ERP
Cz
25% percentile SRC/SS
median SRC/SS
75% percentile SRC/SS
time [ms]
Cz − averages
avgERP
SS
KPLS
LKPLS
avg(CERP)
time [ms]

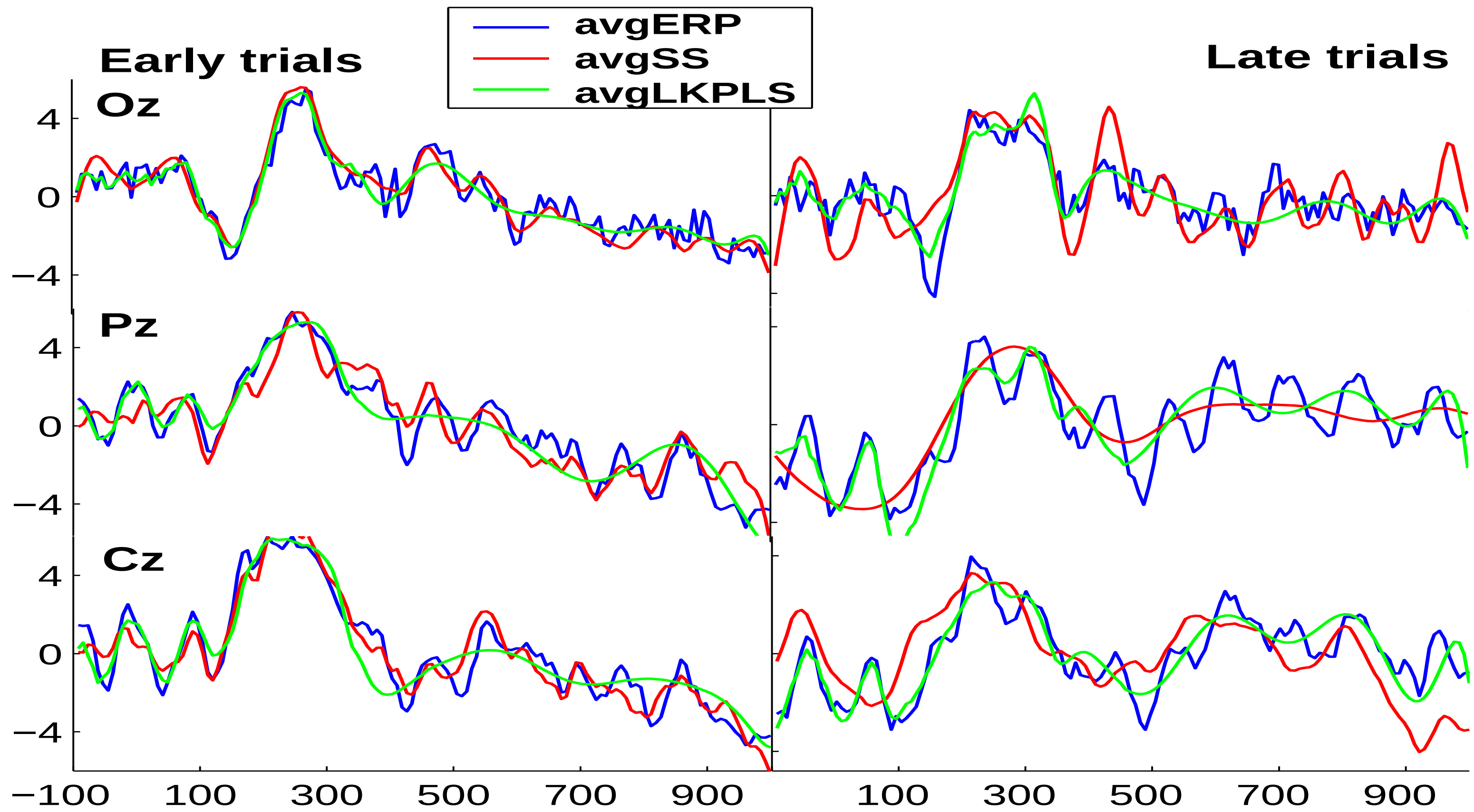


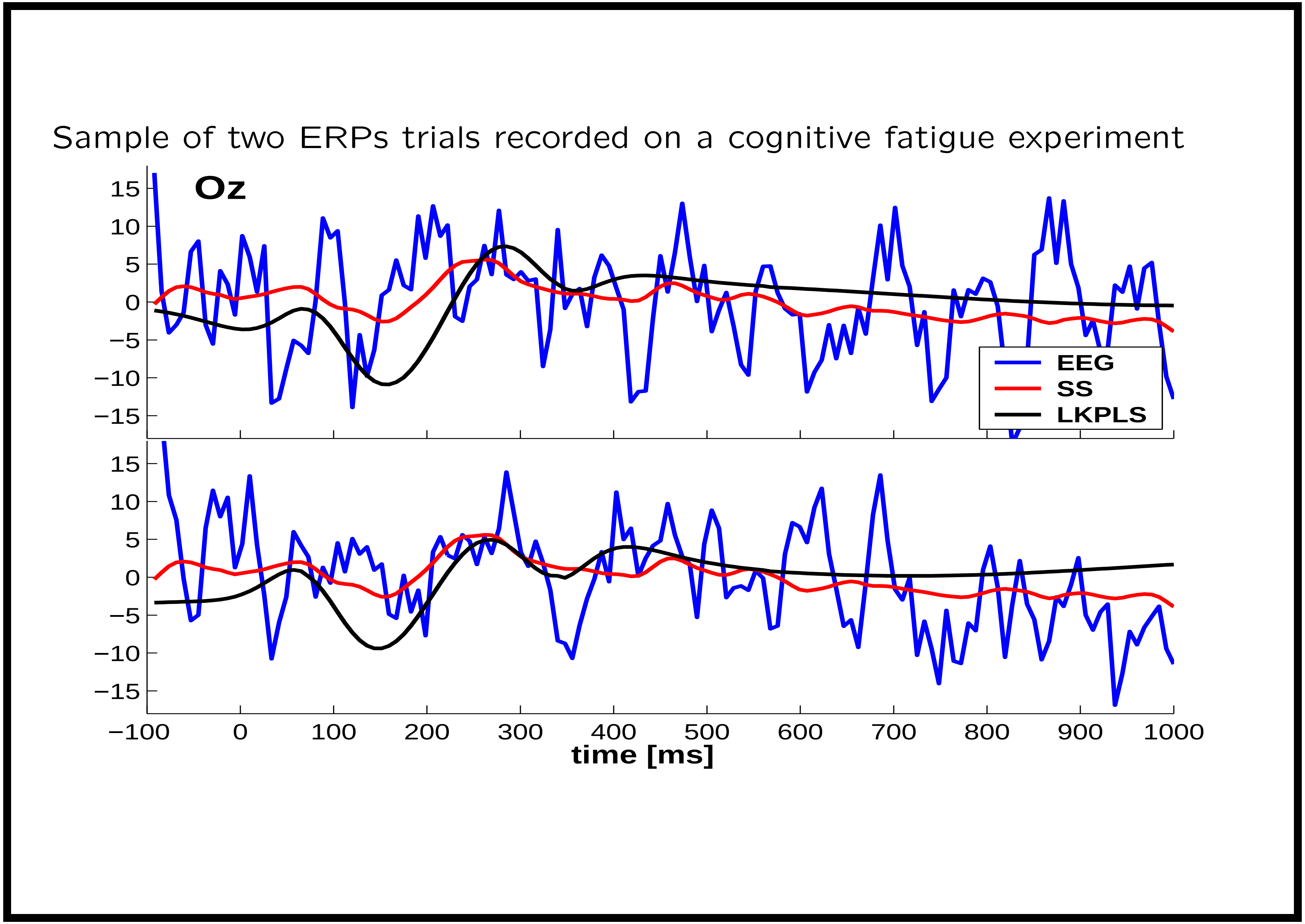
Pz – averages
O1 – averages
avgERP
SS
KPLS
LKPLS
avg(CERP)
time [ms]
time [ms]
4
3
2
1
0
−1
−2
-3
−4
−100
0
100
200
300
400
500
600
700

Results on ERPs recorded on a cognitive fatigue experiment



Results on ERPs recorded on a cognitive fatigue experiment




Sample of two ERPs trials recorded on a cognitive fatigue experiment
Oz
EEG
SS
LKPLS
15
10
5
0
−5
−10
−15
−100
0
100
200
300
400
500
600
700
800
900
1000
time [ms]

Discussion and Future Directions

- Kernel PLS provides comparable results with existing state-of-the-art smoothing and de-noising techniques
- Multivariate (local) kernel PLS allows straightforward extension to estimate of spatio-temporal structure of EEG recordings
- The construction of the (local) kernel PLS regression basis allows to incorporate the prior knowledge about the signal of interest
- Further study of correlated noise structure estimates

References

- Rosipal R., & Trejo L.J.: Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space. Journal of Machine Learning Research, 2:97–123, 2001.
- Schölkopf B., & Smola, A.J ., Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond. The MIT Press, 2002.
- Wahba G.: Splines Models of Observational Data. SIAM,vol. 59, Series in Applied Mathematics, 1990.