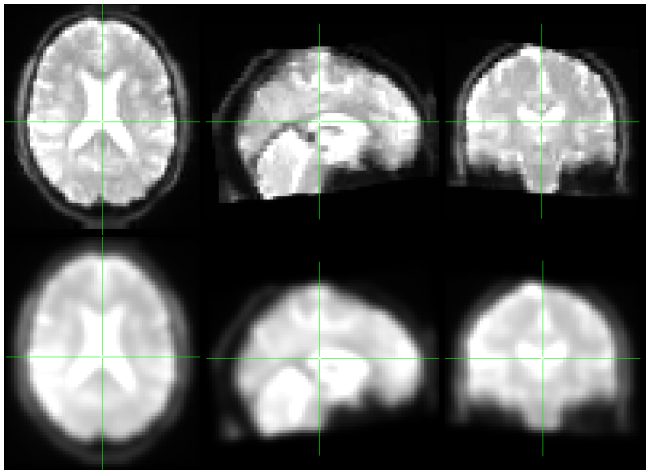


# Structured Features for Bayesian Nonparametric Factor Analysis

Matthew Pearce  
Supervisor: Simon White  
MRC Biostatistics Unit

09/03/2016

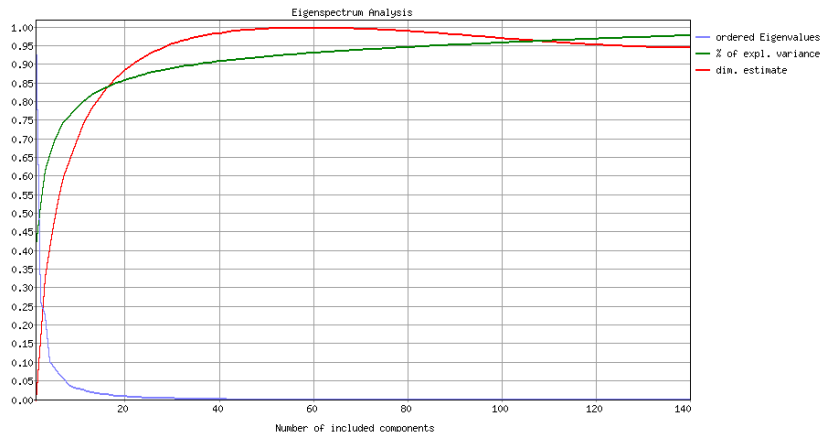
# Smoothing



Smoothing  
1mm FWHM

Smoothing  
1mm FWHM,  
then 10mm  
FWHM

# Dimensionality estimation



# Indian Buffet Process based Factor Analysis

For data  $\mathbf{X}$  ( $T \times V$ ) consisting of  $T$  observations on  $V$  channels.

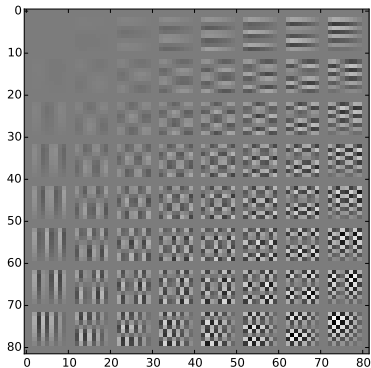
$$\mathbf{X} = (\mathbf{Z} \circ \mathbf{A})\mathbf{S} + \mathbf{E}$$

Sparse matrix  $\mathbf{Z}$  computationally of shape ( $T \times K^+$ ). Weightings  $\mathbf{A}$  ( $T \times K^+$ ). Features  $\mathbf{S}$  ( $K^+ \times V$ ). Extending Knowles & Ghahramani 2007 [1]:

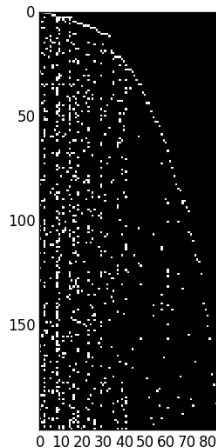
$\mathbf{Z} \sim IBP(\alpha, \beta)$	feature activations
$\mathbf{A}_{t,j} \sim \mathcal{N}(\tau_j, \nu_j^{-1}) \quad j = 1, \dots$	feature weights
$\mathbf{S}_{j,:}^T \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{\phi, \lambda}^{-1}) \quad j = 1, \dots$	features
$\mathbf{E}_{t,:} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \quad t = 1, \dots, T$	measurement error
$\sigma^2 \sim \mathcal{IG}(a, b)$	noise level
$\alpha \sim \mathcal{G}(e_\alpha, f_\alpha)$	IBP strength
$\beta \sim \mathcal{G}(e_\beta, f_\beta)$	IBP repulsion
$\nu_j \sim \mathcal{G}(e_\nu, f_\nu)$	weights' precisions
$\tau_j \sim \mathcal{N}(m_\tau, r_\tau^{-1})$	weights' means
$\ln \lambda \sim \mathcal{N}(m_\lambda, r_\lambda^{-1})$	spatial AR level
$\ln \phi \sim \mathcal{N}(m_\phi, r_\phi^{-1})$	spatial noise level

# Spatial prior for features.

$$\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{\phi, \lambda}^{-1})$$



$$\mathbf{Z} \sim IBP(\alpha, \beta)$$

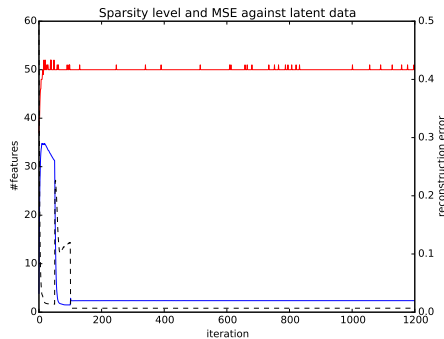
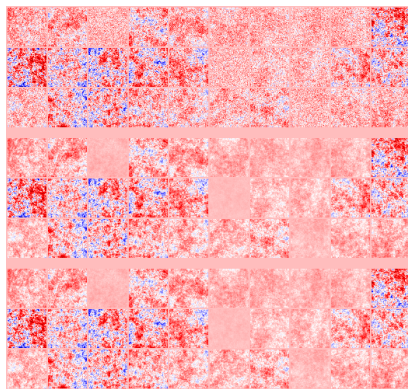


[Example video](#)

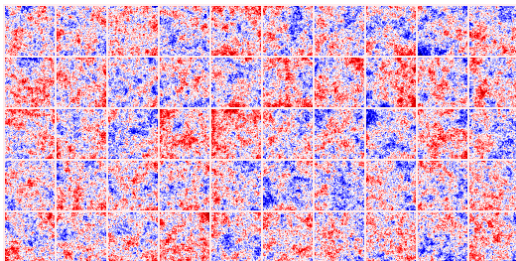
# A simulated dataset

Simulated data, 50 two dimensional features drawn from the prior, with  $\mathbf{Z}_{t,k} \sim B(1/20)$ ,  $\mathbf{A}_{t,k} \sim N(0, 5)$ ,  $\sigma^2 \sim N(0, 1/2)$

LEFT. Top: corrupted data. Middle: latent data. Bottom: ISSFA reconstruction.  
RIGHT. Red: number of active features in model. Blue: mean #activations per observation. Black: reconstruction error.

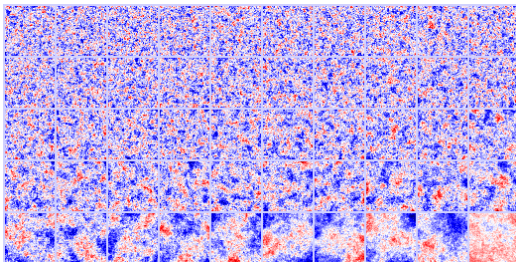


## Comparison of features



Features from final iteration of Infinite Structured Sparse Factor Analysis.

MSE mean reconstruction versus latent data: 0.0034



Eigenvectors from PCA.

MSE reconstruction versus latent data 0.0081.