A Bayesian Model for Brain Network Functional Connectivity using PyMC3

Rui Wang
M.S. Defense Presentation

Department of Biostatistics

Outline

- Introduction

- Methods
    - Spatiotemporal Structure
    - Hierarchical Structure
    - Double Fusion
    - Prior Distribution
    - PyMC3 and NUTS
    - Optimization and Decomposition

- Simulation and Case Study

Outline

- Introduction

- Methods
  - Spatiotemporal Structure
  - Hierarchical Structure
  - Double Fusion
  - Prior Distribution
  - PyMC3 and NUTS
  - Optimization and Decomposition

- Simulation and Case Study

Previous Study

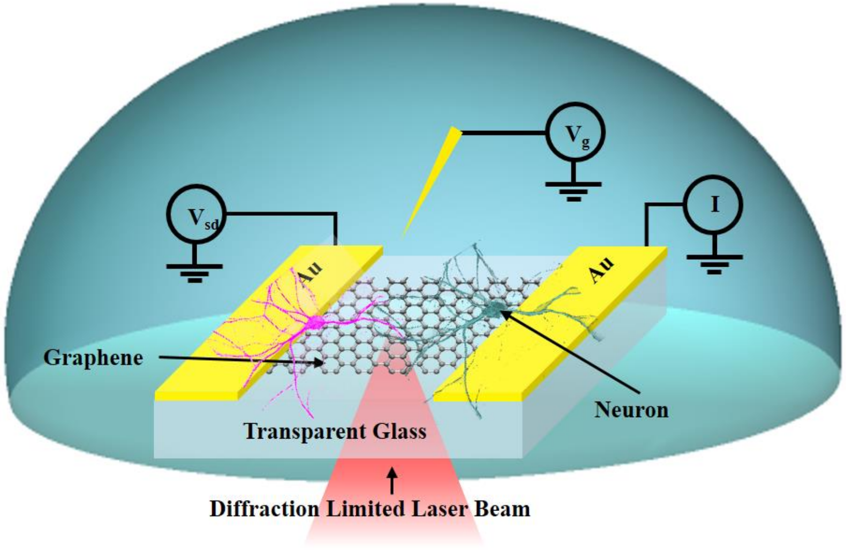
Precise Timing

High Electrical Sensitivity

High Throughput

High Spatial Accuracy

Long-term Duration

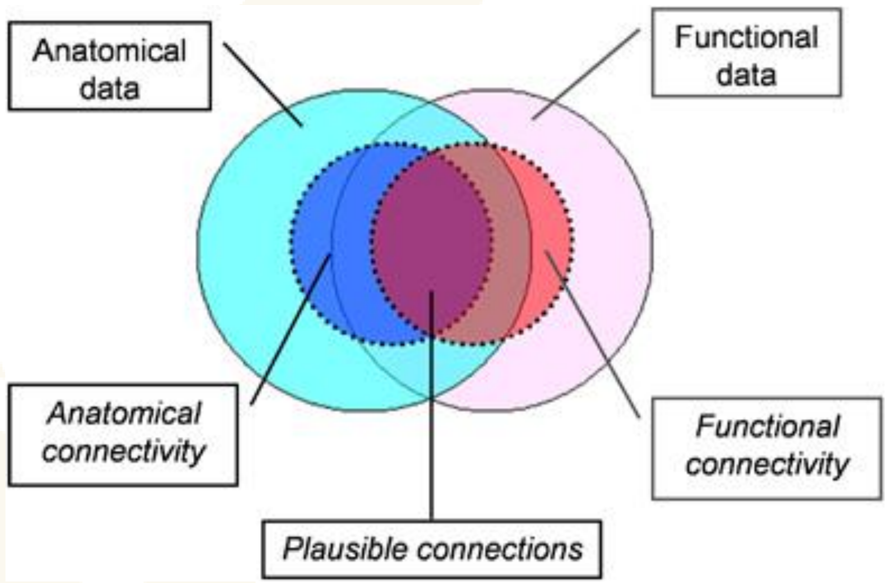


Wang, R et. al, ***Nano Letter*** (in review)

Brain Imaging

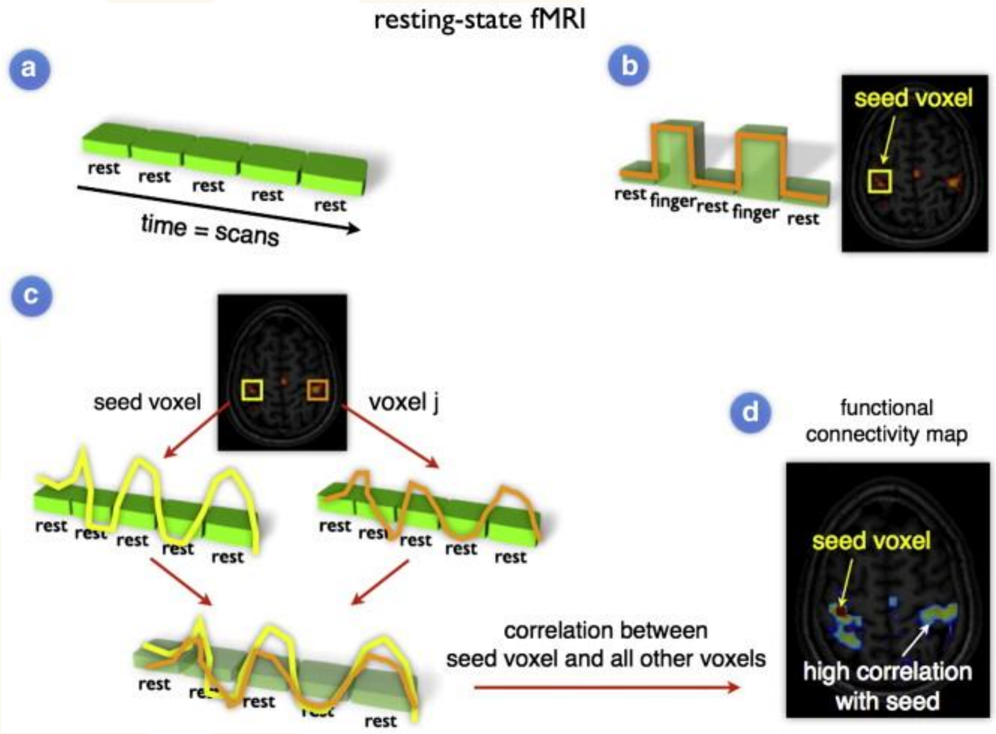
https://www.sciencedaily.com/releases/2016/11/161103141437.htm

Brain Imaging



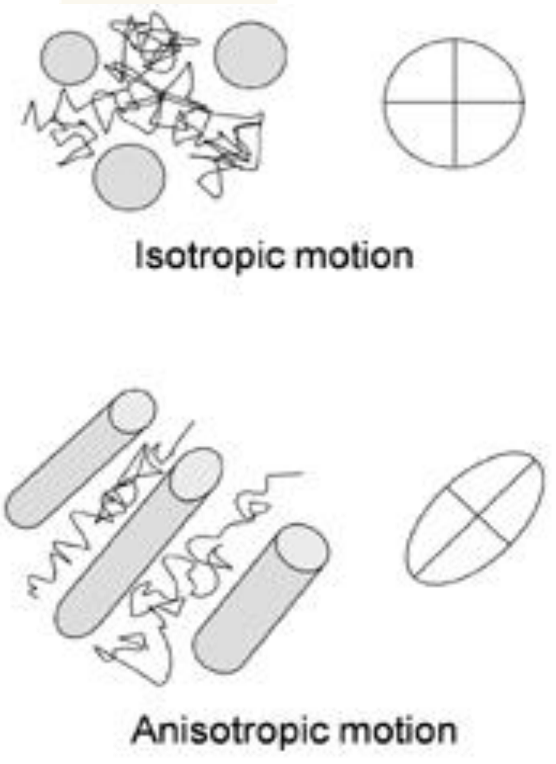
Rykhlevskaia et. al, ***Psychophysiology*** 45, 2 (2008)

Functional Connectivity



van den Heuvel et. al, ***European Neuropsychopharmacology*** 20, 8 (2008)

Structural Connectivity



Rykhlevskaia et. al, ***Psychophysiology*** 45, 2 (2008)

Outline

- Introduction

- Methods
  - Spatiotemporal Structure
  - Hierarchical Structure
  - Double Fusion
  - Prior Distribution
  - PyMC3 and NUTS
  - Optimization and Decomposition

- Simulation and Case Study

Outline

- Introduction
- Methods
  - Spatiotemporal Structure
  - Hierarchical Structure
  - Double Fusion
  - Prior Distribution
  - PyMC3 and NUTS
  - Optimization and Decomposition
- Simulation and Case Study

Spatiotemporal Structure

$$Y_{cv}(t) = \beta_c + b_c(v) + d_c + \epsilon_{cv}(t)$$

- β_c: the grand mean
- $b_c(v)$: the zero-mean voxel-specific random effect
  - Local spatial dependency:

$$Cov\big(b_c(v), b_c(v')\big) = K_c(\|v - v'\|)$$

- d_c: the zero-mean ROI-specific random effect
- $\epsilon_{cv}(t)$: the noise
  - AR (1) temporal structure

Kang, H et. al, ***Brain Connectivity*** 7, 4 (2008)

Kernel Covariance Function

Constant	$K(x, x') = c$
Linear	$K(x, x') = x^T x'$
Gaussian noise	$K(x, x') = \sigma^2 \delta_{x,x'}$
Squared exponential	$K(x, x') = \exp(-\frac{\|x - x'\|^2}{2l^2})$
Exponential	$K(x, x') = \exp(-\frac{\|x - x'\|}{l})$
Matérn	$K(x, x') = \frac{2^{1-v}}{\Gamma(v)} \left(\frac{\sqrt{2v}\|x - x'\|}{l}\right)^v B_v(\frac{\sqrt{2v}\|x - x'\|}{l})$
Periodic	$K(x, x') = \exp(-\frac{2sin^2(\frac{x - x'}{2})}{l^2})$
Rational quadratic	$K(x, x') = (1 + \|x - x'\|^2)^{-\alpha}, \alpha \geq 0$

Kernel Covariance Function

$$r = \|v - v'\|\varphi_c$$

Exponential (Matérn1/2):

$${\sigma_{b_c}}^2 \exp(-r)$$

Gaussian or square exponential (Matérn∞):

$${\sigma_{b_c}}^2 \exp(-\frac{1}{2} r^2)$$

Matérn5/2:

$${\sigma_{b_c}}^2 (1 + \sqrt{5}\mathrm{r} + \frac{5}{3} r^2) \exp(-\sqrt{5} r)$$

Matérn3/2:

$${\sigma_{b_c}}^2 (1 + \sqrt{3} r) \exp(-\sqrt{3} r)$$

Temporal Correlation

AR (1) structure:

$$\epsilon_{cv}(t) = \delta_c + \phi_{cv}\,\epsilon_{cv}(t-1) + w(t)$$

- δ_c: the constant shift
- ϕ_{cv}: the coefficient with $|\phi_{cv}| < 1$
- $w(t)$: the Gaussian random noise

$$\mathrm{E}[\epsilon_{cv}(t)] = \frac{\delta_c}{1-\phi_{cv}}$$

$$\mathrm{Var}[\epsilon_{cv}(t)] = \frac{{\sigma_{cv}}^2}{1-{\phi_{cv}}^2}$$

Outline

- Introduction
- Methods
  - Spatiotemporal Structure
  - Hierarchical Structure
  - Double Fusion
  - Prior Distribution
  - PyMC3 and NUTS
  - Optimization and Decomposition
- Simulation and Case Study

Hierarchical Structure

$$\boldsymbol{Y}_c(t) = \boldsymbol{\beta}_c + \boldsymbol{b}_c + \boldsymbol{d}_c + \boldsymbol{\epsilon}_c(t)$$

- $\boldsymbol{Y}_c(t) = [Y_{c1}(t), Y_{c2}(t), \dots, Y_{cV}(t)]^T$
- $\boldsymbol{\beta}_c = \beta_c \boldsymbol{J}_{(1\times V)}$
- $\boldsymbol{b}_c = [b_{c1}, b_{c2}, \dots, b_{cV}]^T$
- $\boldsymbol{d}_c = d_c \boldsymbol{J}_{(1\times V)}$
- $\boldsymbol{\epsilon}_c(t) = [\epsilon_{c1}(t), \epsilon_{c2}(t), \dots, \epsilon_{cV}(t)]^T$

Hierarchical Structure

$$\boldsymbol{Y}_c(t) = \boldsymbol{\beta}_c + \boldsymbol{b}_c + \boldsymbol{d}_c + \boldsymbol{\epsilon}_c(t)$$

- $\beta_c \sim \mathrm{N}\left(0, {\sigma_{\beta_c}}^2\right)$
- $\boldsymbol{b}_c \sim \mathrm{N}\left(0, \Sigma_{b_c}\right)$
- $d_c \sim \mathrm{N}(0, \Sigma_d)$
- $\epsilon_{cv}(t) \sim \mathrm{N}(\frac{\delta_c}{1-\phi_{cv}}, \frac{{\sigma_{cv}}^2}{1-{\phi_{cv}}^2})$

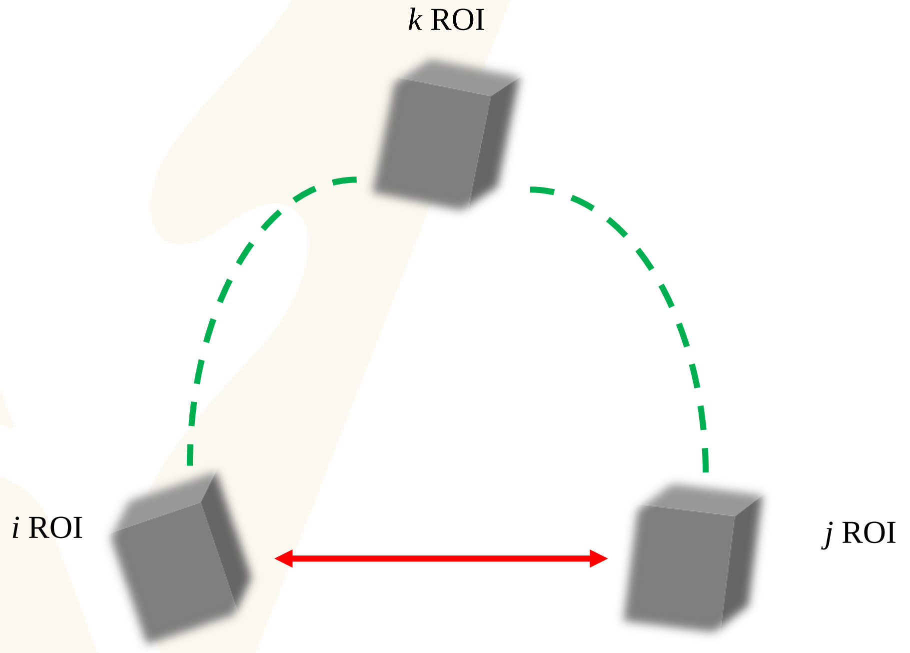
Outline

- Introduction

- Methods
  - Spatiotemporal Structure
  - Hierarchical Structure
  - Double Fusion
  - Prior Distribution
  - PyMC3 and NUTS
  - Optimization and Decomposition

- Simulation and Case Study

Double Fusion



Double Fusion

$$L_d(direct) = \lambda L_{sc} + (1 - \lambda)L_{nfc}$$

$$L_d(indirect) = M_{sc}\lambda L_{sc} + (1 - M_{sc}\lambda)L_{nfc}$$

$$L_d = \omega L_d(direct) + (1 - \omega)L_d(indirect)$$

$$\Sigma_d = L_d \times {L_d}^T$$

Double Fusion

$$\rho_d = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ & 1 & \ddots & \vdots \\ & & 1 & \rho_{(n-1)n} \\ & & & 1 \end{pmatrix}_{n\times n}$$

$$\left[\rho_{12}, \dots, \rho_{1n}, \rho_{23}, \dots, \rho_{2n}, \dots, \rho_{(n-1)n}\right]_{n\_vec}$$

Outline

- Introduction

- Methods
    - Spatiotemporal Structure
    - Hierarchical Structure
    - Double Fusion
    - Prior Distribution
    - PyMC3 and NUTS
    - Optimization and Decomposition

- Simulation and Case Study

Prior Information

$$Y_{cv}(t) = \beta_c + b_c(v) + d_c + \epsilon_{cv}(t)$$

- $\beta_c \sim \mathrm{N}(0, 0.01^2)$

Prior Information

$$Y_{cv}(t) = \beta_c + b_c(v) + d_c + \epsilon_{cv}(t)$$

$$Cov\big(b_c(v), b_c(v')\big) = \sigma_{b_c}{}^2 \exp(-\|v - v'\|\varphi_c)$$

- $\varphi_c \sim \mathrm{Unif}(0, 20)$
- $\sigma_{b_c} \sim \mathrm{Unif}\,(0, 100)$

Prior Information

$$Y_{cv}(t) = \beta_c + b_c(v) + d_c + \epsilon_{cv}(t)$$

- $\lambda \sim \text{Beta}(1,1)$
- $\omega \sim \text{Beta}(1,1)$
- $\sigma_{d_c} \sim \text{Unif}(-8,8)$

Prior Information

$$Y_{cv}(t) = \beta_c + b_c(v) + d_c + \epsilon_{cv}(t)$$

$$\epsilon_{cv}(t) = \delta_c + \phi_{cv}\, \epsilon_{cv}(t-1) + w(t)\varphi_c$$

- $\phi_{cv} \sim \mathrm{Unif}(0, 1)$
- $\sigma_{cv} \sim \mathrm{Unif}(0, 100)$

Prior Information

$$\boldsymbol{Y}_{obs} \sim \mathrm{N}(\boldsymbol{Y}_{cv}, \sigma^2)$$

- $\sigma \sim \mathrm{Unif}(0, 100)$

Outline

- Introduction

- Methods
  - Spatiotemporal Structure
  - Hierarchical Structure
  - Double Fusion
  - Prior Distribution
  - PyMC3 and NUTS
  - Optimization and Decomposition

- Simulation and Case Study

PyMC3 and NUTS

$$Y \sim \mathrm{N}(\mu, \sigma^2)$$

$$\mu = \alpha + \beta_1 X_1 + \beta_2 X_2$$

- $\alpha \sim \mathrm{N}(0, 100)$
- $\beta_1\ or\ \beta_2 \sim \mathrm{N}(0, 20)$
- $\sigma \sim \mathrm{HalfNormal}(0, 1)$

PyMC3 and NUTS

```python
import pymc3 as pm
with pm.Model() as basic_model:

    # Priors for unknown model parameters
    alpha = pm.Normal('alpha', mu=0, sd=100)
    beta = pm.Normal('beta', mu=0, sd=20, shape=2)
    sigma = pm.HalfNormal('sigma', sd=1)

    # Expected value of outcome
    mu = alpha + beta[0]*X1 + beta[1]*X2

    # Likelihood (sampling distribution) of observations
    Y_obs = pm.Normal('Y_obs', mu=mu, sd=sigma, observed=Y)
```

```python
with basic_model:

    # instatiate sampler
    step = pm.NUTS()

    # draw 1000 posterior samples and tune 500 as default
    trace = pm.sample(1000, step = step)
```

Model Diagnostics

- Gelman Rubin statistics:

$$\hat{R} = \frac{\hat{V}}{W}$$

- Effective sample size:

$$\hat{n}_{eff} = \frac{mn}{1 + 2\sum_{t=1}^{T} \hat{\rho}_t}$$

Outline

- Introduction
- Methods
  - Spatiotemporal Structure
  - Hierarchical Structure
  - Double Fusion
  - Prior Distribution
  - PyMC3 and NUTS
  - Optimization and Decomposition
- Simulation and Case Study

Optimization and Decomposition

- Vectorization

- Cholesky decomposition

$$X \sim \mathrm{N}(\mu, \Sigma)$$

$$\Sigma = U^T U$$

$$X = \mu + U^T Z, Z \sim N(0, 1)$$

Outline

- Introduction
- Methods
  - Spatiotemporal Structure
  - Hierarchical Structure
  - Double Fusion
  - Prior Distribution
  - PyMC3 and NUTS
  - Optimization and Decomposition
- Simulation and Case Study

Simulation Study

- Generate time-series data with a length of $\mathrm{T} = 128$ scans using AR (1) (coefficient: 0.6) at 5 ROIs and each ROI contains 100 voxels
- Imposed correlation using a multivariate normal distribution

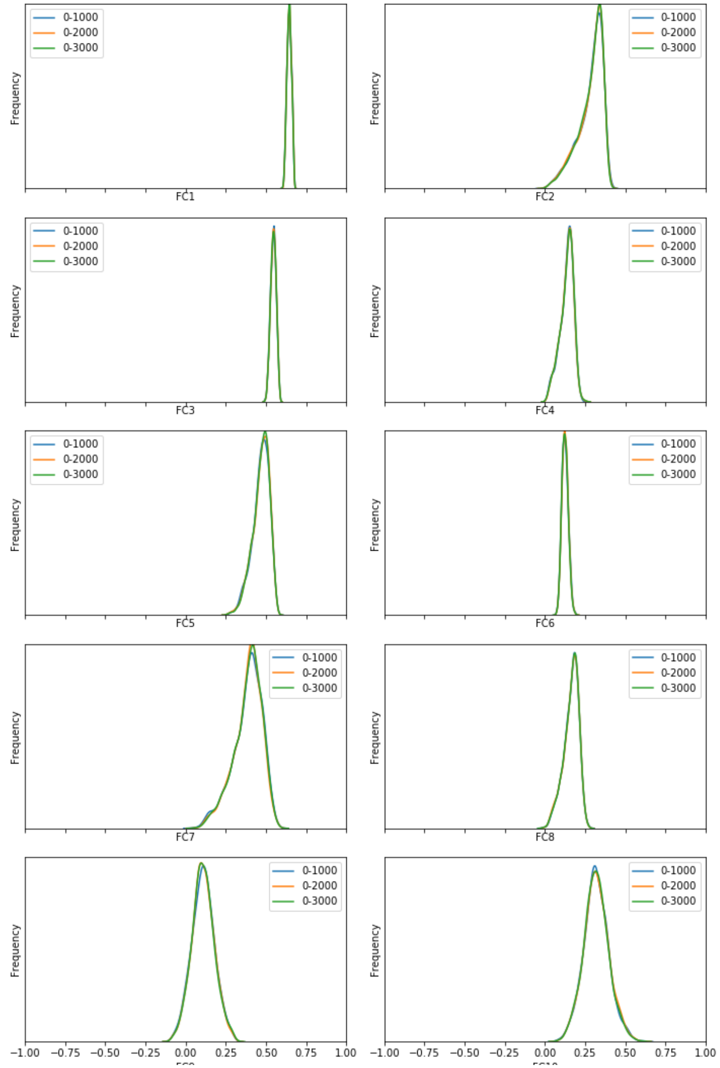
$$\rho_{\mathrm{d}} = \begin{pmatrix} 1 & 0.6 & 0 & 0.5 & 0 \\ & 1 & 0.2 & 0.1 & 0 \\ & & 1 & 0 & 0.1 \\ & & & 1 & 0.2 \\ & & & & 1 \end{pmatrix}$$

- $SC \sim W_p(6, \rho_d)$

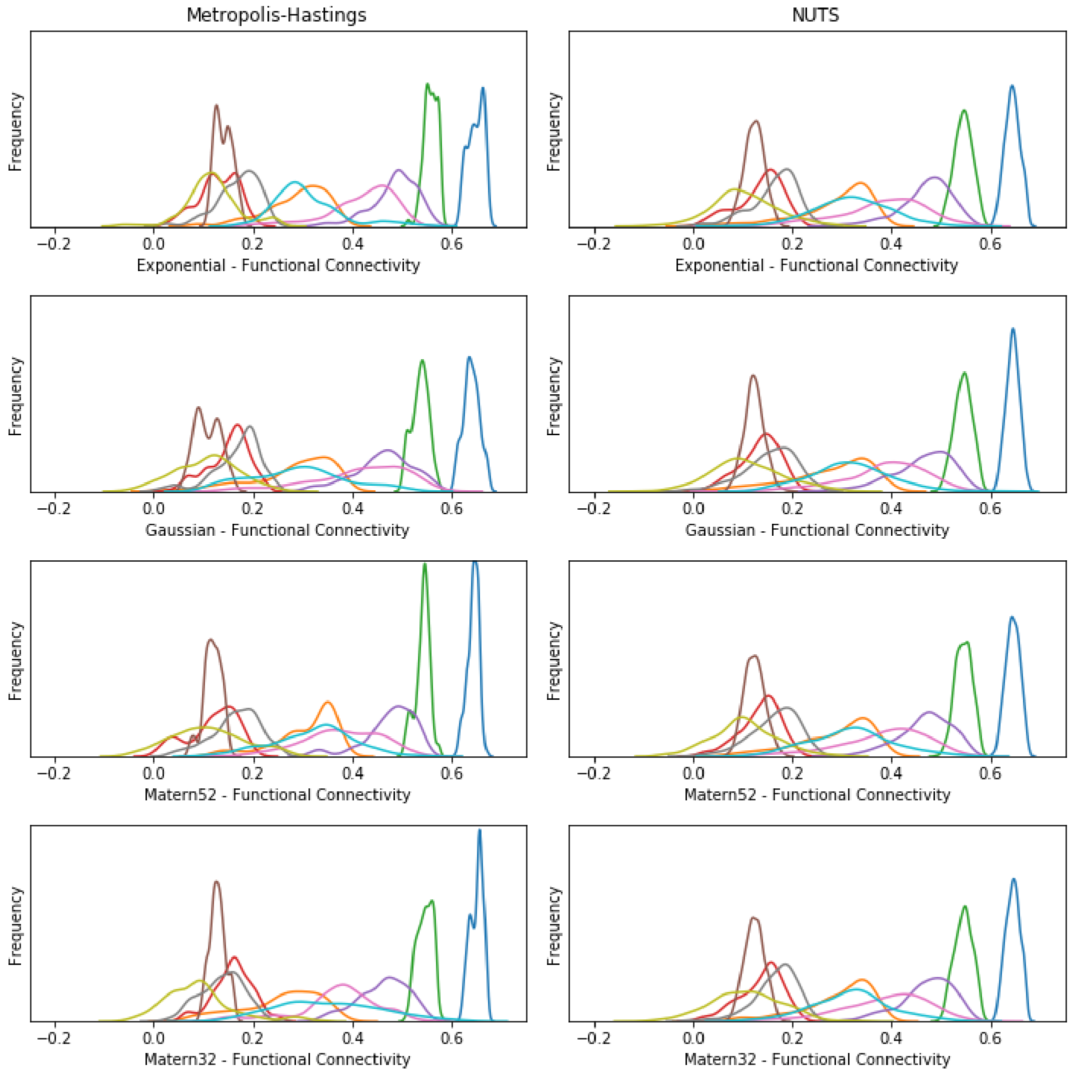
Simulation Study

	Bayesian correct SC				Bayesian independence			
FC	***Median (SD)***	***[2.5% 97.5%]***	$\hat{R}$	$\hat{n}_{eff}$	***Median (SD)***	***[2.5% 97.5%]***	$\hat{R}$	$\hat{n}_{eff}$
ρ_1	0.645 (0.014)	[0.617 0.669]	0.999	1675.438	0.531 (0.125)	[0.170 0.652]	1.002	979.951
ρ_2	0.310 (0.080)	[0.074 0.378]	0.999	1110.936	0.305 (0.078)	[0.091 0.379]	0.999	1303.385
ρ_3	0.546 (0.017)	[0.513 0.576]	0.999	1877.821	0.463 (0.108)	[0.151 0.572]	0.999	1033.009
ρ_4	0.145 (0.042)	[0.039 0.204]	0.999	1341.864	0.146 (0.042)	[0.040 0.204]	0.999	1034.929
ρ_5	0.478 (0.053)	[0.340 0.547]	0.999	1152.252	0.432 (0.090)	[0.197 0.537]	1.001	1285.429
ρ_6	0.123 (0.020)	[0.088 0.165]	1.000	1721.880	0.017 (0.106)	[-0.194 0.216]	1.000	1218.262
ρ_7	0.399 (0.083)	[0.180 0.522]	0.999	1102.607	0.419 (0.098)	[0.174 0.558]	0.999	1381.804
ρ_8	0.173 (0.049)	[0.047 0.235]	0.999	1293.770	0.158 (0.068)	[0.016 0.283]	1.000	1251.840
ρ_9	0.105 (0.070)	[-0.026 0.254]	0.999	1417.283	0.057 (0.063)	[-0.066 0.177]	1.000	1458.304
ρ_{10}	0.316 (0.077)	[0.165 0.484]	1.000	1524.805	0.348 (0.118)	[0.054 0.532]	0.999	1095.445

Convergence



Bayesian Correct SC



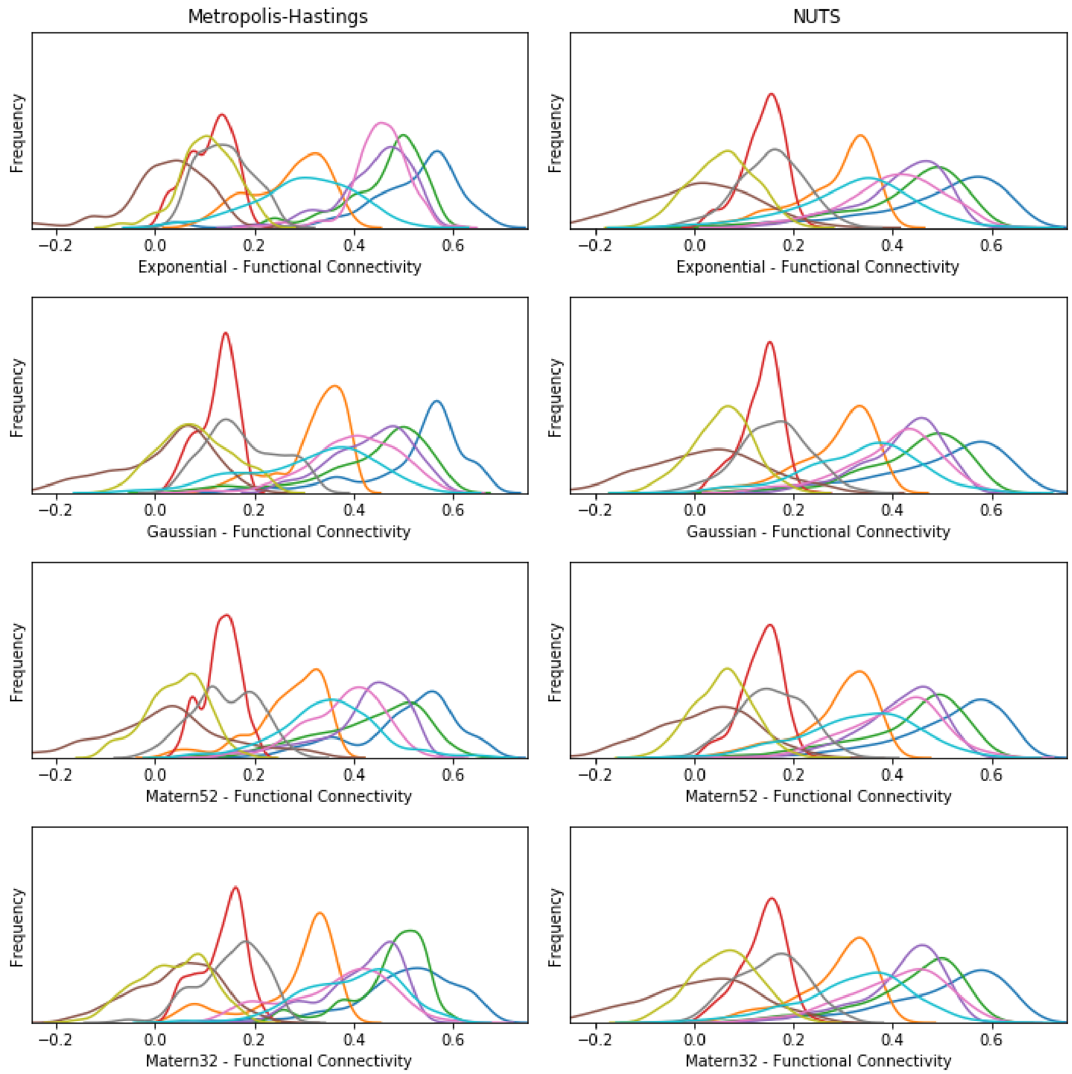
Bayesian Correct SC

	Metropolis-Hastings Median (SD)				NUTS Median (SD)			
FC	***Exponential***	***Gaussian***	***Matérn52***	***Matérn32***	***Exponential***	***Gaussian***	***Matérn52***	***Matérn32***
ρ_1	0.651(0.016)	0.640(0.016)	0.645(0.016)	0.654(0.016)	0.651(0.015)	0.640(0.015)	0.645(0.015)	0.654(0.015)
ρ_2	0.303(0.061)	0.305(0.061)	0.333(0.061)	0.273(0.061)	0.303(0.077)	0.305(0.077)	0.333(0.077)	0.273(0.077)
ρ_3	0.559(0.014)	0.540(0.014)	0.544(0.014)	0.548(0.014)	0.559(0.018)	0.540(0.018)	0.544(0.018)	0.548(0.018)
ρ_4	0.129(0.040)	0.160(0.040)	0.132(0.040)	0.162(0.040)	0.129(0.046)	0.160(0.046)	0.132(0.046)	0.162(0.046)
ρ_5	0.492(0.046)	0.469(0.046)	0.486(0.046)	0.470(0.046)	0.492(0.052)	0.469(0.052)	0.486(0.052)	0.470(0.052)
ρ_6	0.139(0.017)	0.107(0.017)	0.119(0.017)	0.128(0.017)	0.139(0.019)	0.107(0.019)	0.119(0.019)	0.128(0.019)
ρ_7	0.438(0.068)	0.430(0.068)	0.378(0.068)	0.379(0.068)	0.438(0.088)	0.430(0.088)	0.378(0.088)	0.379(0.088)
ρ_8	0.179(0.043)	0.182(0.043)	0.170(0.043)	0.142(0.043)	0.179(0.047)	0.182(0.047)	0.170(0.047)	0.142(0.047)
ρ_9	0.116(0.053)	0.111(0.053)	0.108(0.053)	0.086(0.053)	0.116(0.066)	0.111(0.066)	0.108(0.066)	0.086(0.066)
ρ_{10}	0.293(0.069)	0.293(0.069)	0.336(0.069)	0.350(0.069)	0.293(0.077)	0.293(0.077)	0.336(0.077)	0.350(0.077)

Bayesian Correct SC

	Metropolis-Hastings				NUTS			
MSE	***Exponential***	***Gaussian***	***Matérn52***	***Matérn32***	***Exponential***	***Gaussian***	***Matérn52***	***Matérn32***
Total FC	0.043	0.042	0.040	0.037	0.040	0.039	0.040	0.041
Zero FC	0.083	0.084	0.075	0.066	0.076	0.073	0.076	0.078
Low FC	0.024	0.020	0.025	0.024	0.023	0.022	0.023	0.024
High FC	0.003	0.002	0.002	0.003	0.002	0.002	0.002	0.002

Bayesian Independence



Bayesian Independence

	Metropolis-Hastings Median (SD)				NUTS Median (SD)			
FC	***Exponential***	***Gaussian***	***Matérn52***	***Matérn32***	***Exponential***	***Gaussian***	***Matérn52***	***Matérn32***
ρ_1	0.547(0.016)	0.56(0.016)	0.529(0.016)	0.517(0.016)	0.547(0.015)	0.560(0.015)	0.529(0.015)	0.517(0.015)
ρ_2	0.283(0.061)	0.339(0.061)	0.283(0.061)	0.317(0.061)	0.283(0.077)	0.339(0.077)	0.283(0.077)	0.317(0.077)
ρ_3	0.488(0.014)	0.481(0.014)	0.453(0.014)	0.493(0.014)	0.488(0.018)	0.481(0.018)	0.453(0.018)	0.493(0.018)
ρ_4	0.116(0.040)	0.134(0.040)	0.136(0.040)	0.145(0.040)	0.116(0.046)	0.134(0.046)	0.136(0.046)	0.145(0.046)
ρ_5	0.461(0.046)	0.449(0.046)	0.443(0.046)	0.434(0.046)	0.461(0.052)	0.449(0.052)	0.443(0.052)	0.434(0.052)
ρ_6	0.032(0.017)	0.047(0.017)	0.025(0.017)	0.053(0.017)	0.032(0.019)	0.047(0.019)	0.025(0.019)	0.053(0.019)
ρ_7	0.456(0.068)	0.407(0.068)	0.391(0.068)	0.391(0.068)	0.456(0.088)	0.407(0.088)	0.391(0.088)	0.391(0.088)
ρ_8	0.136(0.043)	0.162(0.043)	0.139(0.043)	0.167(0.043)	0.136(0.047)	0.162(0.047)	0.139(0.047)	0.167(0.047)
ρ_9	0.108(0.053)	0.079(0.053)	0.048(0.053)	0.042(0.053)	0.108(0.066)	0.079(0.066)	0.048(0.066)	0.042(0.066)
ρ_{10}	0.307(0.069)	0.339(0.069)	0.356(0.069)	0.411(0.069)	0.307(0.077)	0.339(0.077)	0.356(0.077)	0.411(0.077)

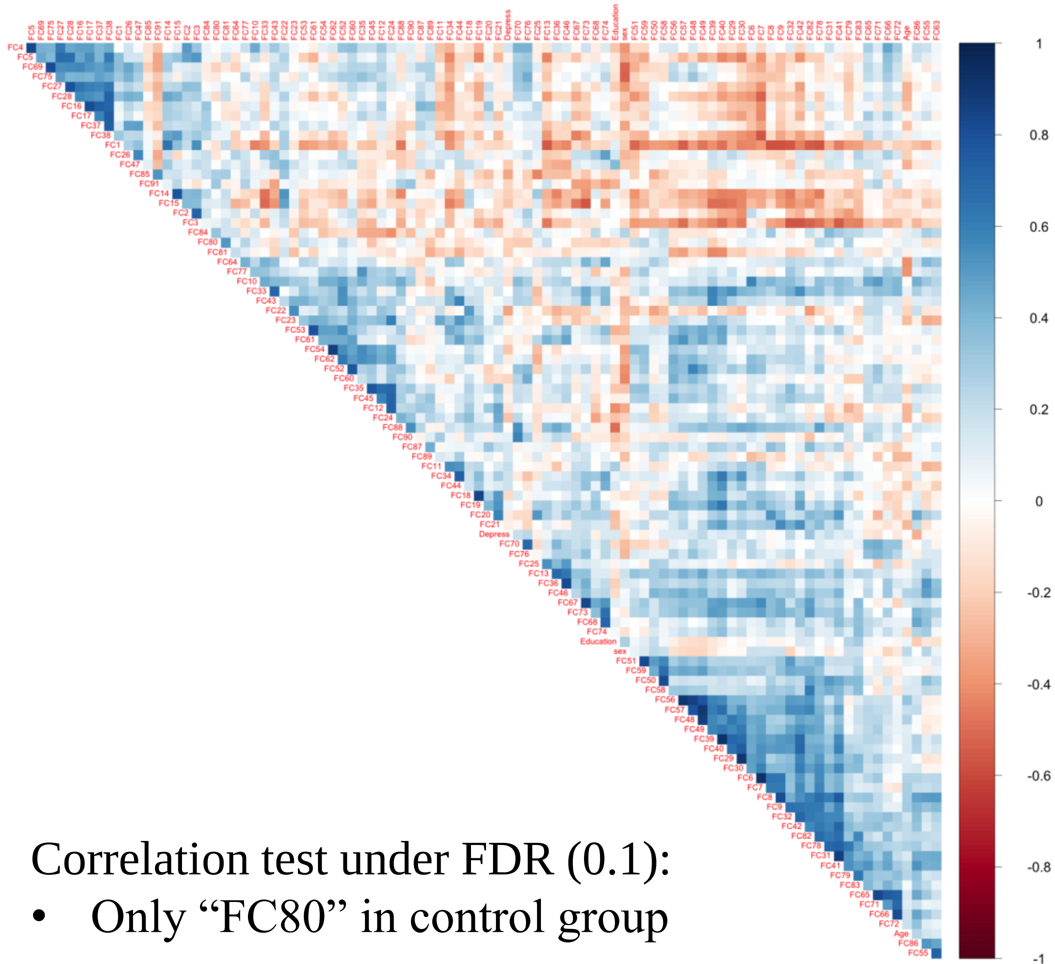
Bayesian Independence

	Metropolis-Hastings				NUTS			
MSE	***Exponential***	***Gaussian***	***Matérn52***	***Matérn32***	***Exponential***	***Gaussian***	***Matérn52***	***Matérn32***
Total FC	0.118	0.119	0.108	0.118	0.111	0.113	0.113	0.113

Case Study

	Control (n=23) Mean (SD)	*MDD (n=18)* Mean (SD)	*Wilcoxon Rank Sum Tests*
Age(years)	31.78 (10.16)	32.06 (8.55)	t = -0.512, p = 0.608
Sex (% female)	65%	50%	t = 0.828, p = 0.408
Education (years)	15.78 (1.73)	16.28 (1.90)	t = -0.512, p = 0.608
Beck Depression Inventory (BDI)	1.90 (2.62)	22.11 (9.38)	t = -4.085, p <0.001
Montgomery–Asberg Depression Rating Scale (MADRS)	0.70 (1.06)	25.29 (3.20)	t = -5.438, p < 0.001
Processing Speed Domain	0.36 (0.67)	0.20 (0.61)	t = 0.841, p = 0.401
Working Memory Domain	0.10 (0.88)	0.02 (0.81)	t = 0.158, p = 0.875
Episodic Memory Domain	0.23 (0.55)	0.07 (0.75)	t = 0.578, p = 0.563
Executive Function Domain	0.20 (0.55)	0.23 (0.59)	t = -0.053, p = 0.958

Case Study



Correlation test under FDR (0.1):

- Only “FC80” in control group

Case Study

$$\text{Cognitive domain} \sim \text{Age} + \text{Sex} + \text{Education} + \text{FC}_i + \text{Depress} + \text{FC}_i * \text{Depress}$$

- Cognitive domain:

Processing Speed Domain

Working Memory Domain

Episodic Memory Domain

Executive Function Domain

- Interaction term $\text{FC}_i * \text{Depress}$: “FC80”

Variable Selection

	Processing Speed Domain	*Working Memory Domain*	*Episodic Memory Domain*	*Executive Function Domain*
Exhaustive	FC4, FC27, FC28, FC48, FC57, FC69	FC26, FC29, FC62, FC64, FC69, FC71	FC6, FC9, FC42, FC57, FC78, FC79	FC6, FC7, FC26, FC27, FC50. FC79
Forward	FC10, FC20, FC29, FC44, FC51, FC58	FC20, FC26, FC33, FC62, FC69, FC85	FC6, FC9, FC11, FC26, FC35, FC65	FC18, FC20, FC26, FC43, FC50, FC77
Backward	FC2, FC4, FC6, FC12, FC20, FC24	FC7, FC18, FC19, FC25, FC26, FC33	FC7, FC9, FC11, FC22, FC29, FC30	FC3, FC6, FC8, FC11, FC13, FC26
Sequential	FC10, FC11, FC17, FC28, FC29, FC78	FC26, FC29, FC62, FC64, FC69, FC71	FC1, FC5, FC6, FC9, FC11, FC35	FC18, FC20, FC26, FC43, FC50, FC77
Lasso	FC10, FC11, FC26, FC29, FC70, FC85	FC20, FC26, FC62, FC64, FC69, FC84	FC6, FC25, FC34, FC57, FC77, FC86	FC11, FC26, FC58, FC70, FC77, FC85

Documentation

wangruinju minor change — Latest commit 304fe04 15 hours ago

accre	add examples	7 days ago
.Rhistory	add html	7 days ago
README.html	minor change	15 hours ago
README.md	minor change	15 hours ago
environment.yml	revise	7 days ago
model.py	update	21 hours ago
slides.pdf	kernel function	21 hours ago

README.md

Double-Fusion

This repository documentation is used to explain the model in the papar by Kang, Hakmook, et al. "A bayesian double fusion model for resting-state brain connectivity using joint functional and structural data." Brain connectivity 7.4 (2017): 219-227.

Since GitHub doest not render the equation in Markdown, you can read the Readme in HTML or slides.

Introduction

Our brain network, as a complex integrative system, consists of many different regions that have each own task and function and simultaneously share structural and functional information. With the developed imaging techniques such as functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI), researchers can investigate the underlying brain functions related to human behaviors and some diseases or disorders in the nervous system such as major depressive disorder (MDD).

We developed a Bayesian hierarchical spatiotemporal model that combined fMRI and DTI data jointly to enhance the estimation of resting-state functional connectivity. Structural connectivity from DTI data was utilized to construct an informative prior for functional connectivity based on resting-state fMRI data through the Cholesky decomposition in a mixture model. The analysis took the advantages of probabilistic programming package as PyMC3 and next-generation Markov Chain Monte Carlo (MCMC) sampling algorithm as No-U-Turn Sampler (NUTS). PyMC3 is new, open-source

Future Work

- Other kernel covariance functions
- 200-300 subjects
- Other machine learning methods of variable selection
- MDD classification

Acknowledgement

Committee Members
Dr. Hakmook Kang (Advisor)
Dr. Qingxia Chen

Faulty and Students
Dr. Jeffrey Blume
Dr. Fonnesbeck Christopher
Dr. Warren Taylor
Sandya Lakkur
David Schlueter
Ya-Chen Lisa Lin
and all the faulty and fellow students in the department

Thank you!