Mathematical and Statistical Challenges in Neuroimaging Data Analysis

Brian Caffo (Johns Hopkins University),
Linglong Kong (University of Alberta),
Farouk Nathoo (University of Victoria),
Todd Ogden (Columbia University),
Hongtu Zhu (University of North Carolina at Chapel Hill)

January 31–February 5, 2016

1 Overview of the Field

Nowadays there is great need for mathematical and statistical theory and methods to analyze high dimensional, correlated, and complex neuroimaging data and clinical and genetic data obtained from various cross-sectional and clustered neuroimaging studies. However, the development of such methods for analyzing imaging data itself and integrating imaging data with genetic and clinical data has fallen seriously behind the technological advances on genomics and neuroimaging. To meet this critical and important need and challenges, the main objectives of the proposed workshop are to serve as a platform for bringing the leading figures from different disciplines including statistics, mathematics, computer science, biomedical engineering, and neuroscience, among other related sciences, exchanging new research ideas, and training the next-generation mathematicians and statisticians in the field of neuroimaging data analysis (NDA).

2 Recent Developments and Open Problems

2.1 Image reconstruction, segmentation and registration

Image reconstruction is to use certain iterative algorithms to reconstruct 2D and 3D images in certain imaging techniques. This is a common issue to all structural, neurochemical, and functional images. Mathematical and statistical methods have been widely used to address various technical issues arising from imaging reconstruction. Such methods include functional analysis, sparse methods, inverse problem, Fourier analysis, nonparametric methods, time series, bootstrap, and regression analysis, among many others.

Image segmentation is the process of assigning a label to every location in an image such that locations with the same label share certain visual characteristics. Image segmentation is typically used to locate important features including objects and boundaries (lines, curves, etc.) in images, which leads to a simple representation that is more meaningful and easier to analyze. Advanced segmentation methods are primarily based on mathematical and statistical methods, such as clustering methods, geometry, partial differential equation-based methods, and pattern theory, among others.

Image registration is the process of transforming different sets of image data into a common coordinate system, which is also called template. In NDA, image data may come from different modalities, from different times, from different viewpoints due to head motion, or from different subjects. Registration is necessary in order to be able to compare or integrate the image data obtained from these different measurements. Mathematical and statistical methods, such as functional analysis, geometry, and nonparametric methods, have been the foundation of developing various imaging registration algorithms. As results can vary somewhat depend on the specific template chosen, an important open problem is the development of methods for template choice, or methods to accommodate the uncertainty associated with this choice.

2.2 Statistical group analysis and Shape analysis

Statistical group analysis is the process of analyzing a sample of images across different groups in an effort to make population level inference. For example, the groups may consist of controls and patients with a specific disorder. Imaging meta analysis, which involves combining the results from multiple group studies is a related area where new techniques for synthesizing results require development. The results of group and meta analysis can be used to inform the development of classification rules based on imaging markers for disease diagnostics and prediction. Advanced statistical methods have played a critical role in addressing various issues in statistical group analysis. Such methods include Bayesian analysis, random effects models, multiple comparison methods, meta analysis, classification methods, sparse methods, nonparametric methods, functional data analysis time series, bootstrap, and regression analysis, among many others. As the data obtained from only a single subject can itself be high-dimensional, a major challenge associated with group studies involves integrating ultra-high dimensional data with complex behavioral measures.

Shape analysis involves the representation, analysis, and processing of geometric shapes extracted from medical images across different subjects or groups. Some of the important aspects of shape analysis are to build boundary representations for a shape, to obtain a measure of distance between shapes, to estimate average shapes from a (possibly random) sample and to estimate shape variability in a sample. Shape analysis requires advanced mathematical and statistical methods including geometry, functional analysis, harmonics analysis, and parametric and nonparametric statistics.

2.3 Connectivity analysis and Multimodal analysis

Connectivity analysis is to establish a pattern of anatomical links (”anatomical connectivity”), of statistical dependencies (”functional connectivity”) or of causal interactions (”effective connectivity”) between distinct units within a nervous system. The units correspond to individual neurons, neuronal populations, or anatomically segregated brain regions. Important mathematical and statistical methods for connectivity analysis include graph theory, social network analysis, multivariate analysis, exponential family, Markov chain Monte Carlo methods, neural network, and time series, among others.

Multimodal analysis is to develop systematic approaches for fusing image data across multiple imaging modalities, in order to find any patterns of related change that may be present. Given that any single imaging modality will only provide a partial snapshot of the true underlying neural activity, multimodal neuroimaging studies can yield a more complete picture; however, devising models that can effectively combine information from different modalities (e.g. scalp level electroencephalogram (EEG) and brain level functional Magnetic Resonance Imaging (fMRI)) is a non-trivial task. Important mathematical and statistical methods for multimodal analysis include measurement error models, multivariate analysis, neural network, and time series, among others.

2.4 Imaging genetics

Imaging genetics involves the collection and analysis of a wealthy set of imaging, genetic and clinical data in order to detect susceptibility genes for complex inherited diseases including common mental disorders (e.g., schizophrenia and bipolar disorder) and neurodegenerative disorders. Understanding genetic mechanisms of inheritable mental and neurological disorders, such as autism and schizophrenia, is an important step in the development of urgently needed approaches to prevention, diagnosis, and treatment of these complex disorders. Currently, imaging provides the most effective measures of brain structure and function, and hence imaging data may serve as important endotraits that ultimately can lead to discoveries of genes for these complex disorders. Although there exist few such methods in the statistical literature, the development of tools for analyzing imaging genetics data will require advanced mathematical and statistical methods including sparse methods, multivariate analysis, regression models, and nonparametric analysis, among many others. While many tools for high-dimensional data analysis have been developed for genetic data with scalar valued phenotypes, imaging-based phenotypes are far more complex given their dimension and the inherent spatial correlation that exists in the observations comprising each image. The development of spatial models for imaging genomics is thus an open area of investigation, but one that poses significant challenges for model development, computation, and inference involving multiple testing on a massive scale.

3 Presentation Highlights

3.1 Day 1: February 1, 2016

The morning section I on February 1 featured two presentations about connectivity analysis delivered by Ying Guo from Emory University and Hernando Ombao from University of California, Irvine. Ying Guo's presentation was on "Exploring the brain connectivity: questions, challenges and recent findings" [1]. Brain connectivity analysis based on functional neuroimaging data has drawn significant interest in recent years. A wide range of network modelling tools have been developed for this purpose. The most commonly used methods includes full correlation, partial correlation and Bayes nets. Ying's talk presented some interesting findings in brain functional connectivity and structural connectivity using resting-state fMRI and diffusion MRI. She proposed a measure of the strength of structural connectivity underlying the functional connectivity networks estimated by independent component analysis (ICA) [2]. Hernando's presentation was on "Multi-Scale Factor Analysis of High Dimensional Time Series". Hernando introduced a multi-scale factor analysis model for EEG data [3]. By applying the algorithm he proposed to a single-subject EEG data, he found that even a small number of factors like 3 was able to capture most of the variation within each region. The connectivity between channels/voxels in the same region is generally higher than between channels/voxels from different regions, and Ying also mentioned same finding during her talk.

Jian Kang from University of Michigan and Moo K. Chung from University of Wisconsin-Madison were the presenters in the morning session II on February1. Jian's presentation was on "Posterior Mean Screening for Scalar-on-Image Regression". Neuroimaging data can be used to classify a subject's disease status or predict clinical response or behaviour. There have been many variable selection methods proposed in high-dimensional feature space, and Jian proposed a new approach called posterior mean screening by using the marginal posterior mean of regression coefficients as the screening statistic. Moo's presentation was on "Learning Large-Scale Brain Networks for Twin fMRI" [4]. In many human brain network studies, the number of voxels (p) is usually significantly larger than the number of images/participants (n). Sparse network models are usually used to fix the small n large p problem, however the computational challenge brought by optimizing L1-penalties makes it not practical to learn large-scale brain networks using sparse network models. Moo proposed a model to build sparse brain networks at the voxel level, and the minimization problem can be simply done algebraically instead of using an iterative algorithm. The computational speed gain by doing that makes it possible to use different sparse parameter instead of using a single sparse parameter, which may not be enough or optimal. The method was applied to twin fMRI data to determine the extent of heritability on functional brain networks a the voxel-level for the first time.

In the afternoon session I on February 1, Joerg Polzehl from Weiertrass Institute for Applied Analysis and Stochastics presented "Modeling high resolution MRI: Statistical issues" [5]. Most MRI data has already been through many preprocessing steps before the statistical analysis. A number of new methods have been proposed to increase spatial resolution and reduce acquisition time for MRI, such as multiple receiver coils and subsampling in K-space. However, those more complicated acquisition methods may further diminish the signal-to-noise ratio (SNR), change in the signal distribution and induce spatial correlation. By analyzing the data generating process and the resulting imaging data distribution, Joerg elaborated the effects of typical data preprocessing and the bias effects related to low SNR for the example of the diffusion tensor model in diffusion MRI. Bei Jiang from University of Alberta discussed "Modeling Placebo Response using EEG data through a Hierarchical Reduced Rank Model" [6]. There have been evidence showing that there are individual differences among depression patients on EEG, fMRI and other brain image measurements. The placebo response is a positive medical response due to placebo effect, as if there were an active medication. And it's highly prevalent among antidepressant treatment. By using EEG measurements as a matrix predictor, Bei presented a hierarchical latent class model to differentiate potential placebo responders from non-responders. Given the high dimensionality of the EEG measurements, a reduced rank regression model with a data-driven regularization was used. The application to real data of 96 placebo or drug treated depression patients showed that this model can be used to detect the placebo response and further guide the selection of effective treatment for depression patients in clinical practise.

The afternoon section II on February 1 featured two presentations delivered by Daniel Rowe from Marquette University and Stephen Strother from Baycrest/University of Toronto. Daniel Rowe's presentation was on "Statistical Analysis of Image Reconstructed Fully-Sampled and Sub-Sampled fMRI Data". In order to

accelerate the image acquisition process, methods have been proposed by measuring less k-space data and performing image reconstruction via an estimation of missing data using other image information. Daniel's talk reviewed the measurement and reconstruction of fully-sampled and sub-sampled data in addition to their resulting statistical properties. Daniel presented that the commonly used image reconstruction sensitivity encoding (SENSE) induces long-range through-plane and in-plane correlation [7]. And by showing the potential bias and change brought by the image reconstruction, he suggested that special care needs to be taken when we obtained the preprocessed data and develop models that incorporate processing. Stephen Strother's presentation was on "Metrics for evaluating functional neuroimaging processing pipelines". The typical neuroimaging processing pipelines includes subject selection, experimental design, data acquisition, preprocessing, data analysis and pipeline processing efficacy measuring. Stephen's talk discussed the range of quantitative metrics used in the literature for evaluating the performance of functional neuroimaging processing pipelines [8]. For the preprocessing pipeline, using fixed preprocessing choices across all subjects/sessions is non-optimal and produces a conservative result with reduced SNR and detection power. Instead adapting preprocessing on a subject/session using cross-validation resampling can significantly improve pipeline performance. Also the negative effects of these common pipeline choices are likely to become worse with age and disease. For the processing pipeline, Stephen pointed out small changes within a processing pipeline may lead to large changes in the output, and the results related to human brain function may be obscured by poor or limited choices in the processing pipeline particularly as a function of age and disease.

Vikas Singh from University of Wisconsin-Madison and Jie Peng from University of California, Davis were the two speakers for the afternoon session III on February 1. Vikas Singh's talk was on "A Multi-Resolution Scheme for Analysis of Brain Connectivity Networks". Vikas presented multi-resolution analysis of shapes and connectivity networks since the multi-resolution methods are sensitive to small changes in the networks [9]. By using wavelet transform on graphs, he applied the method to cortical thickness discrimination and brain connectivity discrimination. Jie Peng's talk was on "Fiber orientation distribution function estimation by spherical needlets". Diffusion MRI (D-MRI) are widely used to reconstruct white matter fiber tracts and to provide information on structure connectivity of the brain. Fiber orientation distribution (FOD) function is a spherical probability density function (p.d.f.) that characterizes the fiber distribution at each voxel of the brain white matter. Jie discussed the estimation of FOD based on a spherical needlets representation. The proposed method leads to much better peak localization compared with existing methods based on spherical harmonics representation, particularly when the separation angles among fiber bundles are small.

3.2 Day 2: February 2, 2016

The morning session I on February 2 featured two presentations on imaging genetics using Bayesian model delivered by Farouk Nathoo from University of Victoria and Michele Guindani from University of Texas M.D. Anderson Cancer Center. Farouk Nathoo's presentation was on "A Bayesian Group-Sparse Multi-Task Regression Model for Imaging Genomics" [10]. Imaging genetics is concerned with finding associations between genetic variations and neuroimaging measures as quantitative traits. Statistically, a multivariate regression analysis can be applied by using potentially interlinked brain imaging phenotypes as response vector and the high-throughput single nucleotide polymorphism (SNP) as covariates. Farouk presented a Bayesian approach based on a continuous shrinkage prior that encourages sparsity and induces dependence in the regression coefficients corresponding to SNPs within the same gene, and cross different components of the imaging phenotypes. The proposed model allows for full posterior inference for the regression parameters using Gibbs sampling. Michele Guindani's presentation was on "Integrative Bayesian Modeling Approaches to Imaging Genetics" [11]. The data used in Michele's presentation has two subgroups, healthy controls and schizophrenic patients. By using the fMRI data and genetic covariates (SNPs implicated in schizophrenia) of all subjects, the goal is to identify brain regions with discriminating activation patterns and SNPs relevant to explain such activations in either (or both) subgroups. A hierarchical mixture model with selection of discriminating features was proposed with 2 components each describing activations in control and case groups. An alternative predictive model for disease status that takes into account direct associations between the SNPs/ROIs information and the disease status, as well as the indirect associations captured by a ROI-SNPs network was also proposed.

Bin Nan from University of Michigan and Jaroslaw Harezlak from Indiana University were the two speakers for the morning session II on February 2. Bin Nan's presentation was on "Tuning parameter selection for

voxel-wise brain connectivity estimation via low dimensional submatrices". The tuning parameter selection accounts for the major computing cost in estimating the voxel-wise brain connectivity. Bin presented a tuning parameter selection procedure using Gap-block cross-validation via low dimensional submatrices. Jaraslaw's presentation was on "Assessing uncertainty in dynamic functional connectivity estimation". Traditional functional connectivity analysis typically assumes that functional networks are static in time, and dynamic functional connectivity analysis tries to analyze the functional network over time. One intuitive and straightforward method is the sliding window technique which performed by conducting analysis on a set number of scans in an fMRI session. This nonparametric approach is easy to implement, however it may also be problematic and not adequate to capture the true dynamic change of the functional network. Jaraslaw presented an algorithm based on multivariate linear process bootstrap, which allows for resample multivariate time series data. A model-free estimation of confidence intervals for the dynamically changing correlation coefficient estimate was also introduced.

Jingwen Zhang from University of North Carolina at Chapel Hill, Zhengwu Zhang from Statistical and Applied Mathematical Sciences Institute and Wei Tu from University of Alberta were the presenters in the afternoon session I on February 2. Jingwen's presentation was on "HPRM: Hierarchical Principal Regression Model of Diffusion Tensor Bundle Statistics". In a diffusion tensor imaging (DTI) study, diffusion properties are observed among multiple fiber bundles to understand the association between neurodevelopment and clinical variables, such as age, gender, biomarkers of subjects. Jingwen proposed Hierarchical Principal Regression Model (HPRM) on functional data to efficiently conduct joint analysis of multiple diffusion tensor tracts on both global level and individual level. The proposed model was applied to genome-wide association study on one-month-old twins to explore important genetic variants related to early human brain development. Zhengwu's presentation was on "Robust brain structural connectivity analysis using HCP data". One of main challenges in structural connectivity analysis is to extract precise and robust connectivity networks from the brain. Zhengwu presented a processing pipeline to reliably construct structural connectivity from the dMRI, including streamline extraction, adaptive streamline compression and robust connectivity matrix construction. Wei Tu's presentation was on "Non-local Fuzzy C-Means Clustering with Application to Automatic Brain Hematoma Edema Segmentation using CT". It is critical to efficiently and accurately segment the hematoma and edema region from computed tomography (CT) scans of patients with intracerebral hemorrhage. However, due to the substantial overlap between the edema and surrounding brain tissue and image artifacts, an accurate and automatic segmentation has been very challenging. Wei presented a two phase clustering algorithm by combing the fuzzy C-Means clustering and non-local smoothing. The first step applied the fuzzy clustering algorithm on the whole brain to find the hematoma tissue, which is also the region of interest (ROI). The second phase will apply the clustering algorithm on the ROI to obtain a more detailed segmentation of edema tissue.

The afternoon session II on February 2 featured three presentations delivered by Benjamin Risk from Statistical and Applied Mathematical Sciences Institute and University of North Carolina, John Muschelli from Johns Hopkins University and Chao Huang from University of North Carolina, Chapel Hill. Benjamin Risk's presentation was on "Large covariance estimation for spatial functional data with an application to twin studies". A structural estimation model (SEM) can be used to estimate a trait's heritability, and a mass univariate analysis can estimate an SEM at each location in the brain. Extending the model to spatial domains requires an estimation of the covariance functions. Benjamin presented a spatial function SEM using functional principal component analysis (PCA). The proposed model was applied to the imaging data of twin pairs from Human Connectome Project (HCP). John Muschelli's presentation was on "Processing Neuroimaging Data in R: Capabilities". R language is the most frequently used programming language by statisticians, and there have been many different packages/softwares created for neuroimaging data analysis. During the presentation, John discussed the neuroimaging processing pipeline using R, from read/write images, visualization, bias field correction, skull stripping, image registration, tissue-class segmentation and more complex modeling [12]. Chao Huang's talk was on "FFGWAS: Fast Functional Genome Wide Association Study of Surface-based Imaging Genetic Data" [13]. More and more large-scale imaging genetic studies are being widely conducted to collect a rich set of imaging, genetic, and clinical data to detect putative genes for complexly inherited neuropsychiatric and neurodegenerative disorders. Several major big-data challenges arise from testing millions of genome-wide associations with functional signals sampled at millions of locations in the brain from thousands of subjects. Chao presented a Fast Functional Genome Wide Association Study (FFGWAS) framework to carry out whole-genome analyses of multimodal imaging data. FFGWAS consists

of three components including (1) a multivariate varying coefficient model for modeling the relation between multiple functional imaging responses and a set of covariates (both genetic and non-genetic predictors), (2) a global sure independence screening (GSIS) procedure for reducing the dimension from a very large scale to a moderate scale, and (3) a detection procedure for detect significant cluster-locus pairs. The proposed FFGWAS was applied to large-scale imaging genetic data analysis of ADNI data with 708 subjects, 30,000 vertices on hippocampal surface, and 501,584 SNPs.

The afternoon session III on February 2 was roundtable discussion lead by John Aston from Cambridge University, Martin Lindquist from Johns Hopkins University, Hernando Ombao from University of California, Irvine, Joerg Polzehl from Weiertrass Institute for Applied Analysis and Stochastics and Hongtu Zhu from University of North Carolina at Chapel Hill. Discussion topics included the technical challenges in NDA, grant opportunities, grant review criterion, software development of proposed methodologies (Neuroconduct), the training of next generation statisticians.

3.3 Day 3: February 3, 2016

The morning session I of February 3 featured two presentations on functional data analysis. John Aston from Cambridge University presented "Functional Data, Covariances and FPCA of Brain Data". Functional PCA (FPCA) tries to investigate the dominant modes of variation of functional data, such as fMRI, EEG. John introduced a few different approaches to estimate network connectivity by using functional data analysis. Time changing connectivity via functional change point detection and also spatially constrained connectivity based on the used of penalized functional principal components were presented. The FPCA can be defined on the volume or on the surface, and it can be used to detect general mean shifts in image data and potentially connectivity changes. Jeffrey Morris from University of Texas M.D. Anderson Cancer Center presented "Spatial Functional Models for Event-Related Potential Data, with Application to Smoking Cessation Study ". Jeffrey presented a set of functional regression methods to analyze spatially correlated complex functional data such as functional imaging data [14]. Three major strategies for spatial or temporally correlated functional data are presented, 1. Functional spatial or Functional temporal processes, 2. Tensor basis functions, 3. Functional graphical models. And each method has its own benefits and drawbacks and the suitability depends on the data setting and research questions. All these three methods were applied to a smoking cessation study to assess neurological response to different types of visual stimuli.

Brain Hobbs and Jianhua Hu from University of Texas M.D. Anderson Cancer Center were the speakers for the morning session II of February 3. Brain Hobbs's presentation was on "Recent advances in cancer imaging". In many cancer imaging settings, radiologists often identify the presence of solid tumors over a series of a few repeated scans, and often multiple interdependent ROIs are evaluated in isolation. Independent estimation appears limiting for analysis of sparse functional data derived from dynamic imaging techniques that use physiological models to derive multiple interdependent biomarkers acquired from multiple regions of interests (ROI) within the same organ. Brain proposed statistical methods for joint estimation of sparse spatiotemporally correlated imaging-biomarkers using semi-parametric models. Joint prediction is used to identify liver metastases using perfusion characteristics from multiple ROIs acquired using dynamic computed tomography [15]. Jianhua's talk was on "Analysis of spatially correlated functional data in tissue perfusion imaging". Measurements from perfusion imaging modalities provide physiological correlates for neovascularization induced by tumor angiogenesis. Such measurements are often generated repeatedly over time and at multiple spatially interdependent units. To reduce model complexity and simplify the resulting inference, possible spatial correlation among neighboring units is often neglected. Jianhua presented a weighted kernel smoothing estimate of the mean function that leverages the spatial and temporal correlation, particularly, in the presence of sparse observations.

3.4 Day 4: February 4, 2016

In the morning session I on February 4, Leixi Li from University of California, Berkeley presented "Estimation and Inference for Brain Connectivity Analysis" [16]. Previous studies have demonstrated that brain networks may degrade among Alzheimer's disease (AD) subjects compared to normal aging subjects. Amyloid beta (Aβ) is a form of protein that is toxic to neurons in the brain, and it accumulates outside neurons and forms sticky buildup called Aβ plaques. To understand how Aβ deposition are related to brain connectivity

patterns in cognitively normal elder subjects, Leixi proposed two general framework to tackle the problem. First a comparison of the connectivity networks between the Aβ positive group and Aβ negative group. Second by taking the connectivity network as a predictor, the association between the connectivity network and the Aβ deposition can be modeled. Leixi introduced the symmetric tensor predictor regression model to model the association. Shuo Chen from University of Maryland presented "Brain Connectivity Biomarkers". Many challenges remain for group-level whole-brain connectivity network analyses because the massive connectomics connectivity metrics are correlated and the correlation structure is constrained by the extraordinarily complex, yet highly organized, topology of the underlying neural architecture. Shuo presented several novel machine learning algorithms to automatically detect topological structures, and furthermore construct network "object" oriented statistical inference framework to identify subgraphs as network level biomarkers. Each network biomarker comprises a set of nodes (brain regions) and edges (connectivity metrics), and more importantly the network biomarker is a subgraph with organized topological structures (e.g. clique or multipartite graph).

In the morning session II on February 4, Xiao Wang from Purdue University presented "Optimal Estimation for Quantile Regression with Functional Response". Quantile regression is able to give a full picture of the data by estimating the $100\tau\%$ quantile of the conditional distribution of response Y given X. Quantile regression gives better estimators than mean regression when data are skewed or contain outliers since the appealing robust properties of quantiles, and it also does not require any error distribution. Quantile regression with functional response and scalar covariates has become an important statistical tool for many neuroimaging studies since the variances of errors are varying spatially within the brain. Xiao presented the optimal estimation of varying coefficient functions in the framework of reproducing kernel Hilbert space. Minimax rates of convergence under both fixed and random designs are established. An easily implementable estimator was also presented using alternating direction method of multipliers (ADMM) algorithm. Yimei Li from St. Jude Children's Research Hospital presented "SGPP: Spatial Gaussian Predictive Process Models for Neuroimaging Data" [17]. Yimei presented a spatial Gaussian predictive process (SGPP) model to predict new neuroimaging data by using a set of covariates like age, diagnostic status and existing neuroimaging data set. The SGPP model Yimei presented uses a functional PCA model to capture global dependence, and a multivariate simultaneous autoregressive model to capture local spatial dependence as well as cross-correlations of different imaging modalities. A three-stage estimation procedure was proposed to simultaneously estimate varying regression coefficients across voxels at the global and local spatial dependence structures.

The afternoon session I of February 4 featured three presentations delivered by Marina Vannucci from Rice Universtiy, Todd Ogden from Columbia University and Anuj Srivastiava from Florida State University. Marina Vannucci's presentation was on "A Bayesian Modeling Approach of Multiple-Subject fMRI Data". Marina presented a Bayesian nonparametric regression model for multiple-subject fMRI data [18]. The model incorporates information on both the spatial and temporal correlation structures of the data, and allows for voxel-dependent and subject-specific parameters. It provided a joint analytical framework that allows the detection of regions of the brain that activate in response to a stimulus, while simultaneously taking into account the association, or clustering, of spatially remote voxels within and across subjects. In order to solve the computational challenge brought by the high dimensionality of the data and the large amount of parameters to be estimated, Marina presented a variational Bayes algorithm as an approximate computational technique, and its efficiency was compared to a full Monte Carlo Markov Chain (MCMC) algorithm. Todd Ogden's presentation was on "Functional and imaging data in precision medicine" . A major goal of precision medicine is to use information gathered at the time that a patient presents for treatment to help clinicians determine, separately for each patient the particular treatment that provides the best-expected outcome. Imaging data may also be used in making patient-specific treatment decisions. Todd introduced an ongoing multi-site randomized placebo-controlled clinical trial Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care (EMBARC). The primary goals of EMBARC is selecting measurements that can be made at baseline that will help predict patient response to treatment, and therefore determine a rule, based on these measurements, that will assign the treatment that is best for each patient. Todd presented the general problem of using both scalar and functional data to guide patient- specific treatment decisions and describe some approaches that can be used to perform model fitting and variable selection [19]. Anuj Srivastiava's presentation was on "Elastic Functional Data Analysis for Modeling Shapes of Anatomical Structures" [**?**]. A variety of anatomical structures in human brain can be represented as functions (curves or surfaces) on intervals or spheres. Morphological analysis and statistical modeling of such data faces the following chal-

lenges: the representation spaces are curved, the data is seldom registered, the classical Hilbert structure is problematic, and (nowadays) there is a tremendous amount of data to deal with. Elastic functional data analysis provides a unified framework for dealing with nonlinear geometries and simultaneous registration of function data, and leads to efficient computer algorithms. It has proven to outperform all recent methods in registering functional data. The Functional PCA, resulting from linearized representations under elastic Riemannian metrics, has been used for solving regression and testing under appropriate models. Anuj presented some recent extensions of this work involving morphological analysis of tree-like structures such as neurons.

In the afternoon session II of February 4, Wei Pan presented "Testing for group differences in brain functional connectivity" [21]. There have been evidence showing that that altered brain functional networks are associated with neurological illnesses such as Alzheimer's disease. Exploring brain networks of clinical populations compared to those of controls would be a key inquiry to reveal underlying neurological processes related to such illnesses. Standard approaches for comparing networks includes mass-univariate test and deriving some network summary statistics, like clustering coefficient. Mass-univariate tests can be low powered for multiple weak signals since the dimensionality of networks is usually high, and deriving network summary statistics is not easy and over-simplified. Wei proposed a global test. The proposed tests combine statistical evidence against a null hypothesis from multiple sources across a range of plausible tuning parameter values reflecting uncertainty with the unknown truth. The proposed tests are not only easy to use, but also highly powered robustly across various scenarios. The usage and advantages of these novel tests are demonstrated on an Alzheimer's disease dataset and simulated data. Russel Shiohara from University of Pennsylvania presented "Two-Sample Tests for Connectomes using Distance Statistics". Russel proposed statistical methods for quantifying variability in a population of connectomes using general representations. The methods used generalized variances for complex objects based on distance statistics. Methods were developed for two-sample testing at the whole connectome and the subnetwork levels and the asymptotic properties of the test statistics were studied. These methods was applied in a connectomic study of autism spectrum disorders using DTI.

Tingting Zhang from University of Virginia and Martin Linquist from Johns Hopkins University were the two presenters of afternoon session III of February 5. Tingting Zhang's presentation was on "Bayesian Inference for High-Dimensional ODE Models with Applications to Brain Connectivity Studies". Tingting proposed a widely applicable high-dimensional ordinary differential equations (ODE) model to explore connectivity among multiple small brain regions [22]. The new model, called the modular and indicator-based dynamic directional model (MIDDM), uses indicators to represent significant directional interactions among brain regions and features a cluster structure, which consists of modules of densely connected brain regions. A Bayesian hierarchical model was developed to make inferences about the MIDDM and also to provide a new statistical approach to quantify ODE model uncertainty that arises from the inherent inadequacy of the ODE model for a complex system. The proposed Bayesian framework to an auditory electrocorticography dataset to identify significant clusters and directional effects among different brain regions. Martin Linquist's presentation was on "Dynamic Connectivity: Pitfalls and Promises". To date, most resting state fMRI studies have assumed that the functional connectivity between distinct brain regions is constant across time. However, recently, there has been increased interest in quantifying possible dynamic changes in FC during fMRI experiments, as it is thought this may provide insight into the fundamental workings of brain networks. Martin proposed a dynamic conditional correlations (DCC) model to quantify the dynamic change of brain connectivity [23]. DCC is a multivariate GARCH model. The study of dynamic correlations actually increases the number of data points, so there is critical need to use summary statistics that can be used to find meaningful individual differences. The average dynamic correlation and the variability in dynamic correlation in each stage can be used. Also, one can find some connectivity "state" matrices, which are connectivity patterns that subjects tend to return to during the course of an experiment, to compute the dwell time each subject spends in a given state. The standard approach towards determining coherent brain states across subjects is to perform clustering on the results of the dynamic connectivity analysis. Martin evaluated the reproducibility of metrics computed from dynamic FC, and moderately strong reproducibility of the average correlation was found.

3.5 Day 5: February 5, 2016

In the morning session of February 5, there were three presenters from University of Alberta, Giseon Heo, Matthew Brown and Dana Cobzas. Giseon Heo's presentation was on "Persistent homology: an approach for high dimensional data analysis". Topological data analysis (TDA) has been popularized since its development in early 2000, and it has shown its effectiveness in discerning true features from noise in high-dimensional data. Giseon introduced persistent homology, a particular branch of computational topology and discussed how it can be incorporated to classical statistics and techniques in machine learning [24]. Matthew Brown's presentation was on "Opening the analysis black box: Improving robustness and interpretation". One primary purpose of neuroimaging data analysis is to abstract away most of the dimensionality and complexity in the data by extracting just a small number of significant patterns from it.This analysis involves a long chain of steps that interact with the data at various points. In practice, the analysis can fail at various steps due to a host of reasons such as the influence of noise, bad convergence in some optimization algorithm, and so on. However, the final output of the analysis often provides no indication that such failures have occurred. Another important consideration is that the analysis often abstracts away too much of the structure in the neuroimaging data. Matthew discussed several approaches for delving into what the data analysis is doing to allow for improved robustness through quality assurance checking as well as improved interpretation through consideration of important patterns in the data that often go unnoticed. Dana Cobzas's presentation was on "Sparse classification for significant anatomy detection in a group study". Dana presented a new framework for discriminative anatomy detection in high dimensional neuroimaging data [25]. Current methods for identifying significant regions related to a group study typically use voxel-based mass univariate approaches. Those methods have limited ability to identify complex population differences because they do not take into account multivariate relationships in data. High dimensional pattern classification methods aim to optimally perform feature extraction and selection to find a set of features that differentiate the groups. Dana proposed a sparse classification method that identifies anatomical regions that are both discriminative and clinically interpretable. Results on synthetic and real MRI data of multiple sclerosis patients and age- and gender-matched healthy controls show superior performance of our method in detecting stable and significant regions in a statistical group analysis when compared to a generative sparse method and to a voxel-based analysis method.

4 Scientific Progress Made

Much progress has been made in this workshop. We summarize the comments from some of the workshop participants on this regard.

John Aston, Cambridge University: Just a quick email to say thanks so much for all your organization last week. The workshop was great, and Banff was really fun.

Brain Hobbs, University of Texas M.D. Anderson Cancer Center: Great conference. Thank you for all of your efforts in effectuating and facilitating my participation.

Clay Holroyd, University of Victoria: Many thanks, [the organizers] for organizing the meeting, and for inviting me. I enjoyed it.

Dana Cobzas, University of Alberta: I felt I like to leave a note also. Thanks a lot to [the organizers] for giving me the opportunity to attend such a good workshop. I learned a lot, and now have a pile of papers to read. The friendly atmosphere encouraged me to talk with researched that I would have never approached otherwise. And of course is always special to be at the Banff centre.

Daniel Rowe, Marquette University: Thanks for organizing the workshop.I really enjoyed it.

Jeffrey Morris, University of Texas M.D. Anderson Cancer Center: Yes, thank you to the organizers for a great meeting!

Joerg Polzehl, Weierstrass Institute for Applied Analysis and Stochastics: thanks also from me. I really enjoyed the program, meeting all of you and of course the fantastic environment BIRS provides.

Leixi Li, University of California, Berkeley: I'd like to echo what Tingting and Hernando said. It was a great workshop, and I enjoyed it a lot. Thanks all the organizers, and particularly, Hongtu and Linglong, for great leadership! And I look forward to the next workshop, and would be happy to contribute in whichever way to make it happen.

Michele Guindani, University of Texas M.D. Anderson Cancer Center: I would also like to thank you for organizing such an interesting workshop. It was the best workshop I also have ever attended.

Hernando Ombao, University of California, Irvine: Thanks to all the organizers especially Hongtu and Linglong for their leadership.It's never too early to plan for the next one! This was the best workshop I haveattended. I like our spirited, honest and respectful discussions.

Stephen Strother, Baycrest/University of Toronto: Dear Nassif - This was one of the best workshops I have attended in quite a few years, particularly for size and the time to discuss the content in some depth, all complimented and enhanced by the facilities and location. All thanks to our organizers, and the excellent BIRS environment. I will definitely keep an eye out for more relevant BIRS meetings.

Tingting Zhang, University of Virginia: Thank Linglong and Hongtu so much for organizing this wonderful workshop. I had a great time there, getting inspired by many great talks, meeting and learning from cheerful friends, while enjoying delicious food and beautiful views there.

Vikas Singh, University of Wisconsin-Madison: I wanted to send an email to congratulate you on such an awesome meeting. To be honest, I've rarely attended a set of sessions that were so interesting with a group of amazing friends and colleagues. Thanks much for asking me to be a part of this get together.

Participants

Aston, John (Cambridge University)
Brown, Matthew (University of Alberta)
Chen, Shuo (University of Maryland)
Chung, Moo (University of Wisconsin-Madison)
Cobzas, Dana (University of Alberta)
Cribben, Ivor (Alberta School of Business)
Guindani, Michele (University of Texas MD Anderson Cancer Center)
Guo, Ying (Emory University)
Harezlak, Jaroslaw (Indiana University)
Heo, Giseon (University of Alberta)
Hobbs, Brian (University of Texas MD Anderson Cancer Center)
Holroyd, Clay (University of Victoria)
Hu, Jianhua (University of Texas MD Anderson Cancer Center)
Huang, Chao (University of North Carolina at Chapel Hill)
Jiang, Bei (Columbia University)
Johnson, Timothy D. (University of Michigan)
Kang, Jian (University of Michigan)
Kong, Linglong (University of Alberta)
Li, Yimei (St. Jude Children's Research Hospital)
Li, Lexin (University of California, Berkeley)
Lindquist, Martin (Johns Hopkins University)
Liu, Kevin (Marquette University)
Morris, Jeffrey (University of Texas MD Anderson Cancer Center)
Muschelli, John (Johns Hopkins University)
Nan, Bin (University of Michigan)
Nathoo, Farouk (University of Victoria)
Ogden, Todd (Columbia University)
Ombao, Hernando (University of California, Irvine)
Pan, Wei (University of Minnesota)
Peng, Jie (University of California, Davis)
Polzehl, Joerg (Weierstrass Institute for Applied Analysis and Stochastics)
Risk, Benjamin (Statistical and Applied Mathematical Sciences Institute and University of North Carolina)
Rowe, Daniel (Marquette University)
Shinohara, Russell (University of Pennsylvania)
Singh, Vikas (University of Wisconsin-Madison)

Song, Yin (University of Victoria)
Srivastava, Anuj (Florida State University)
Strother, Stephen (Baycrest/University of Toronto)
Tu, Wei (University of Alberta)
Vannucci, Marina (Rice University)
Wang, Xiao (Purdue University)
Yu, Yang (University of North Carolina at Chapel Hill)
Zhang, Tingting (University of Virginia)
Zhang, Zhengwu (Statistical and Applied Mathematical Sciences Institute)
Zhang, Jingwen (University of North Carolina at Chapel Hill)
Zhao, Yihong (New York University)
Zhu, Hongtu (University of North Carolina at Chapel Hill)

References

[1] Y. Wang, J. Kang, PB. Kemmer and Y. Guo, An Efficient and Reliable Statistical Method for Estimating Functional Connectivity in Large Scale Brain Networks Using Partial Correlation, *Frontier in Neuroscience* **10** (2016), 123.

[2] A. Hyvrinen, and E. Oja, Independent component analysis: algorithms and applications. *Neural networks* **13(4)**, (2000), 411-430.

[3] C. Kirch, B. Muhsal and H. Ombao, Detection of Changes in Multivariate Time Series With Application to EEG data, *Journal of the American Statistical Association* (2014).

[4] M.K. Chung, V.G. Vilalta, P.J. Rathouz, B.B. Lahey, and D.H. Zald, Linear embedding of large-scale brain networks for twin fMRI. *arXiv:1509.04771* (2016).

[5] J. Polzehl, K. Tabelow, Modeling high resolution MRI: Statistical issues with low SNR, *Preprint no. 2179, WIAS, Berlin* (2015).

[6] B. Jiang, E. Petkova, T. Tarpey, and T. R. Ogden. Use of Latent Class Models to Identify an Early Placebo Responder Subgroup with Matrix-valued EEG Measurements. submitted (2015).

[7] M.C. Kociuba, D.B. Rowe: Complex-Valued Time-Series Correlation Increases Sensitivity in FMRI Analysis. *Magn. Reson. Imaging*, **34(6)** (2016), 765-770.

[8] N. W. Churchill, R. Spring, B. Afshin-Pour, F. Dong, and S.C. Strother, An Automated, Adaptive Framework for Optimizing Preprocessing Pipelines in Task-Based Functional MRI. *PloS one*, **10(7)**, (2015).

[9] W. H. Kim, V. Singh, M. K. Chung, N. Adluru, B. B. Bendlin and S. C. Johnson, Multi-resolution statistical analysis on graph structured data in neuroimaging, *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)* (2015), 1548-1551.

[10] F. Nathoo, K. Greenlaw, M. Lesperance, Regularization Parameter Selection for a Bayesian Multi-Level Group Lasso Regression Model with Application to Imaging Genomics, *arXiv:1603.08163* (2016).

[11] F.C. Stingo, M. Guindani, M. Vannucci, V.D. Calhoun, An Integrative Bayesian Modeling Approach to Imaging Genetics. *Journal of the American Statistical Association* **108**(2013), 503.

[12] J. Muschelli, E. Sweeney, M. Lindquist, and C. Crainiceanu, fslr: Connecting the FSL Software with R, *The R Journal* **7(1)**(2015), 163-175.

[13] M. Y. Huang, T. Nichols, C. Huang, Y. Yang, Z.H. Lu, Q.J. Feng, R.C. Knickmeyer and H.T. Zhu, FVGWAS: Fast Voxelwise Genome Wide Association Analysis of Large-scale Imaging Genetic Data, *NeuroImage* **118** (2015), 613-627.

[14] J.S. Morris, Functional Regression, *Annual Review of Statistics and Its Applications* (2015).

[15] B. P. Hobbs and C. S. Ng, Inferring stable acquisition durations for applications of perfusion imaging in oncology, *Cancer Informatics* (2015).

[16] H. Zhou, L. Li, and H. Zhu, Tensor regression with applications in neuroimaging data analysis, *Journal of the American Statistical Association* **108** (2013), 540-552.

[17] J. Hyun, Y. Li, J. Gimore, Z. Lu, M. Styner, H. Zhu, SGPP: Spatial Gaussian Predictive Process Models for Neuroimaging Data, *Neuroimage* **89** (2014), 70-80.

[18] L. Zhang, M. Guindani, J.M. Engelmann, F. Versace, M. Vannucci, A Bayesian spatio-temporal model for multi-subject fMRI data. *Annals of Applied Statistics* (2015).

[19] A. Ciarleglio, E. Petkova, R.T. Ogden, and T. Tarpey, Treatment decisions based on scalar and functional covariates, *Biometrics* **71** (2015), 884?894.

[20] J.S. Marron, J.O. Ramsay, L.M. Sangalli and A. Srivastava, Functional Data Analysis of Amplitude and Phase Variation, *Statistical Science* **30(4)** (2015), 468-484.

[21] J. Kim, W. Pan and Alzheimer's Disease Neuroimaging Initiative, Highly adaptive tests for group differences in brain functional connectivity, *NeuroImage: Clinical* **9** (2015), 625-639.

[22] T. Zhang, J. Wu, F. Li, B. Caffo, D. Boatman-Reich, Dynamic directional model for effective brain connectivity using Electrocorticographic (ECoG) time series, *Journal of the American Statistical Association*, 2015.

[23] M. Linquist, Y. Xu, M.B. Nebel and B. Caffo, Evaluating Dynamic Bivariate Correlations in Resting-state fMRI: A comparison study and a new approach, *Neuroimage* **101** (2014), 531- 546.

[24] V. Kovacev-Nikolic, B. Bubenik, D. Nikoli?c, and G. Heo, Using persistent homology and dynamical distances to analyze protein binding, *arXiv preprint arXiv:1412.1394* (2014).

[25] D. Cobzas, L. Kong, M. Schmid, G. Blevin, and A. Wilman, Sparse classification for significant anatomy detection in a group study, (2015).