NeuroImage 117 (2015) 343–357

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

An automated pipeline for constructing personalized virtual brains from multimodal neuroimaging data

Michael Schirner [a,b,1], Simon Rothmeier [a,b,1], Viktor K. Jirsa [c], Anthony Randal McIntosh [d], Petra Ritter [a,b,e,f,*]

[a] Dept. Neurology, Charité — University Medicine, Berlin, Germany
[b] Bernstein Focus State Dependencies of Learning, Bernstein Center for Computational Neuroscience, Berlin, Germany
[c] Institut de Neurosciences des Systèmes UMR INSERM 1106, Aix-Marseille Université Faculté de Médecine, Marseille, France
[d] Rotman Research Institute of Baycrest Centre, University of Toronto, Toronto, Canada
[e] Minerva Research Group BrainModes, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[f] Berlin School of Mind and Brain, Mind and Brain Institute, Humboldt University, Berlin, Germany

ARTICLE INFO

Article history:
Received 27 October 2014
Accepted 19 March 2015
Available online 31 March 2015

Keywords:
The Virtual Brain
Connectome
Tractography
Diffusion MRI
Multimodal imaging
Computational modeling

ABSTRACT

Large amounts of multimodal neuroimaging data are acquired every year worldwide. In order to extract high-dimensional information for computational neuroscience applications standardized data fusion and efficient reduction into integrative data structures are required. Such self-consistent multimodal data sets can be used for computational brain modeling to constrain models with individual measurable features of the brain, such as done with The Virtual Brain (TVB). TVB is a simulation platform that uses empirical structural and functional data to build full brain models of individual humans. For convenient model construction, we developed a processing pipeline for structural, functional and diffusion-weighted magnetic resonance imaging (MRI) and optionally electroencephalography (EEG) data. The pipeline combines several state-of-the-art neuroinformatics tools to generate subject-specific cortical and subcortical parcellations, surface-tessellations, structural and functional connectomes, lead field matrices, electrical source activity estimates and region-wise aggregated blood oxygen level dependent (BOLD) functional MRI (fMRI) time-series. The output files of the pipeline can be directly uploaded to TVB to create and simulate individualized large-scale network models that incorporate intra- and intercortical interaction on the basis of cortical surface triangulations and white matter tractography. We detail the pitfalls of the individual processing streams and discuss ways of validation. With the pipeline we also introduce novel ways of estimating the transmission strengths of fiber tracts in whole-brain structural connectivity (SC) networks and compare the outcomes of different tractography or parcellation approaches. We tested the functionality of the pipeline on 50 multimodal data sets. In order to quantify the robustness of the connectome extraction part of the pipeline we computed several metrics that quantify its rescan reliability and compared them to other tractography approaches. Together with the pipeline we present several principles to guide future efforts to standardize brain model construction. The code of the pipeline and the fully processed data sets are made available to the public via The Virtual Brain website (thevirtualbrain.org) and via github (https://github.com/BrainModes/TVB-empirical-data-pipeline). Furthermore, the pipeline can be directly used with High Performance Computing (HPC) resources on the Neuroscience Gateway Portal (http://www.nsgportal.org) through a convenient web-interface.

Introduction

In biology, the notion that structure predicts function is widespread. In human neurosciences, different modalities image different structural aspects making their integration imperative to predict function (Sporns, 2013). The Virtual Brain (TVB, thevirtualbrain.org) uses empirical structural and functional data to build full brain models of individual primates—consisting of interacting dynamic local models—that predict individual whole-brain activity on different scales (Ritter et al., 2013; Sanz-Leon et al., 2013; Roy et al., 2014; Woodman et al., 2014). The interactions between neuronal populations in a full brain model are constrained by the anatomical fiber skeleton, i.e., the structural connectome, obtained from diffusion-weighted magnetic resonance imaging (dwMRI) using tractography techniques. The human brain connectome is the set of neuronal connections in the human brain, a concept that crosses spatial brain scales (Sporns et al., 2005; Craddock

* Corresponding author at: Max Planck Institute for Human Cognitive and Brain Sciences Leipzig, Department of Neurology, Charité Universitaetsmedizin, Charitéplatz 1, 10117 Berlin, Germany. Fax: +49 30 450 560 936.
E-mail address: petra.ritter@charite.de (P. Ritter).
[1] Equal contributions.

http://dx.doi.org/10.1016/j.neuroimage.2015.03.055

et al., 2013). The term connectome is used in the literature for functional connectivity (FC; i.e., statistical dependencies of brain activity), structural connectivity (SC; i.e., anatomical connections between brain areas) and effective connectivity (EC; i.e., causal interaction between brain areas). A connectome is often represented as a weighted graph with nodes defining brain regions and edges characterizing the connections between these regions. FC is a highly variable and non-stationary activity pattern (Bassett et al., 2011b; Allen et al., 2014; Hutchison et al., 2013; Zalesky et al., 2014) arising from interactions within the structural skeleton. FC is a statistical concept that estimates correlations between data from simultaneous measurements of different brain areas that does not necessarily reflect the neuroanatomical structures. On the other hand, the anatomical connection pattern or wiring diagram between neurons and neuronal ensembles, dubbed SC, is typically described in terms of distances and connection strengths mediated by synaptic or electric connections between region pairs. In contrast, EC captures the causal relations between neural systems by quantifying the directed influences that one element of a generative model exerts over another (Valdes-Sosa et al., 2011).

In recent years, efforts for multicenter data sharing have increased and several large-scale projects started to collaboratively pool and compile multimodal neuroimaging data, e.g., (Biswal et al., 2010; Van Essen et al., 2012). The Neuroscience Information Framework (http://neuinfo.org/) lists over 2500 different databases with relevance for neuroscience. This high number of heterogeneous resources requires standardized and efficient processing routines in order to (i) extract interpretable and relevant information and to (ii) organize and integrate it in a systematic and unifying structure: "Perhaps the single biggest roadblock to higher order data mining is the lack of standardized frameworks for organizing neuroscience data" (Akil et al., 2011).

We propose to go one step further: In order to get from pure data gathering to knowledge inference we need to connect functional and structural data by means of model-based integration (Jirsa et al., 2002, 2010; Ritter et al., 2013). The formulation of a comprehensive theory of neural computation that allows a qualitative and quantitative mapping between cognitive and neural states is only possible if we close the loop between data-driven inference and model-based prediction. Jirsa et al. (2002) merged geometric and topographic structural information with brain network modeling, but used simplified network connectivity and demonstrated that temporal activation patterns are well captured as observed in human brain imaging. A necessary condition to produce realistic spatiotemporal activations is the additional inclusion of topological information, that is, realistic network connectivity, which poses substantial neuroinformatics challenges. The Virtual Brain is a step into this direction and provides an integrated neuroinformatics platform (Sanz-Leon et al., 2013) for modeling dynamic large-scale brain network models (BNM) constructed from structural data and interacting local dynamic population models. Within its theoretical framework, TVB integrates the relevant information extracted from a variety of empirical sources associating brain network structure with brain function via models of neural activity. By doing so, it abstracts from the high dimensionality of information contained in raw imaging data and unifies relevant structural and dynamical information within a single brain model. The unified theoretical framework provided by TVB together with the processing pipeline for multimodal empirical data opens up new avenues of collective neuroscience. TVB empowers the community to conveniently construct biologically informed brain models, to perform in silico experiments that predict neuronal activity and to expose principles of computation across spatial and temporal scales in a variety of modalities.

Data reduction and fusion are prerequisites for automated data analysis, to ensure interoperability of data structures and for comparability of multicenter acquisitions. One example is the alignment of the spatial and temporal dimensions of recordings from different modalities within and across subjects and their integration into a common reference system. Data turns into information when they are semantically annotated and ontologically aligned. Extracted information gains maximal interpretability when mappings between data sets and their organization into a unified coordinate system can be achieved, e.g., the registration and mapping of anatomical structures between modalities or temporal alignment of simultaneously acquired multimodal data (Calhoun and Lemieux, 2014; Jorge et al., 2014; James and Dasarathy, 2014; Uludağ and Roebroeck, 2014).

The processing pipeline presented in this article provides an efficient and automated way for generating full and self-consistent data sets for TVB model construction integrating anatomical, diffusion weighted and functional MRI scans with EEG recordings. Online supplementary Movie M1 illustrates the involved imaging modalities and estimated source activity along with brain network activity projected onto reconstructed head and cortex models of the exemplary subject QL used throughout this paper. The pipeline runs on standard computers, but also supports a high degree of parallelization for computationally intensive processes, optimized to run on stand-alone workstations and high performance clusters alike. In the following, we describe the functionality of the pipeline by demonstrating each step on the exemplary data set. Up to now we pre-processed 50 full data sets using this pipeline. All data sets were stored in the TVB XNAT (Marcus et al., 2007) database in Toronto where they are made available to the TVB consortium. Along with the processing steps, we illustrate the challenges posed when working with multimodal imaging data and integrating them in a single framework such as provided by TVB. These challenges range from storage requirements due to large amounts of data, interoperability and interfacing between different toolboxes and coordinate systems, fallacies of dwMRI tractography to outcome validation. Each of the imaging modalities serves different purposes during model generation and optimization within TVB:

(i) High resolution *T1-weighted MRI* scans are used to obtain parcellations of cortical and subcortical white and gray matter (WM, GM) into subregions of interest based on anatomical landmarks and to construct anatomically constrained dipole source models for forward modeling and inverse source reconstruction of EEG and magnetoencephalography (MEG). Resulting lead-field matrices and inversion kernels are used to map cortical activity to scalp locations of (M)EEG sensors and vice versa. Furthermore, high-resolution scalp/head, skull and cortex-surface triangulations are used for highly resolved surface simulations and output visualization.

(ii) *fMRI* volumes are parcellated according to the high-resolution atlases derived from T1-weighted data yielding region-wise aggregated BOLD time-series and FC matrices generated from these are used to fit model output by means of parameter tuning.

(iii) *Diffusion-weighted MRI (dwMRI)* data are parcellated according to the high-resolution atlases derived from T1-weighted data yielding estimated white matter fiber tracts and SC matrices. The parcellations are used for defining seed- and stop-locations during tractography.

(iv) *EEG* data is projected to source space and used to optimize parameters of the brain model.

Pipeline results are provided in a format that can be directly imported to TVB and readily integrated into a single full-brain model. As part of this pipeline a novel tractography-based connectome extraction approach is described. The method introduces several concepts that facilitate the standardization of the BNM construction process. Connectomes are embedded at the core of the generic BNM equation to define long-range information transmission thereby linking large-scale network infrastructure with neural mass dynamics. In this context, connectomes are based on a given parcellation of the brain and consist