UNIVERSITY OF CALIFORNIA, MERCED

Resting-state functional near-infrared spectroscopy: analytical challenges

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

in

Psychological Sciences

by

Pradyumna Lanka

Committee in charge:

Professor Heather Bortfeld, Chair
Professor Haiyan Liu
Professor Kristina C. Backer
Professor Rose M. Scott

2022

The Dissertation of Pradyumna Lanka is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Kristina C. Backer

Rose M. Scott

Haiyan Liu

Heather Bortfeld, Chair

University of California, Merced

2022

To my family

Table of Contents

List of Figures

Acknowledgments

First, I would like to thank my advisor Dr. Heather Bortfeld for her continued support through my graduate program and especially for putting up with my endless rants.

I would also like to thank my committee members Rose Scott, Haiyan Liu and Kristina Backer. Their coursework and guidance helped me become a better researcher.

This dissertation would not have been possible without our collaborator Theodore Huppert, who was helpful in guiding me through my research projects. His advice and feedback supported Chapter 2. The endless zoom conversations with him helped shape the ideas in Chapter 3 of this dissertation.

I am also indebted to the faculty members in the psychological sciences department at UC Merced whose support and guidance helped me understand the multiple facets of my research, some of that knowledge informs the current work.

I would also like to thank my colleagues and peers who were part of my research journey.

I would also like to acknowledge the funding sources and financial support I received from UC Merced and the Department of Psychological Sciences.

Last but not the least, I would like to thank my family for their love, encouragement, and support. I will forever be in your debt.

Pradyumna Lanka

EDUCATION

Ph.D. Psychological Science — Aug 17 — Aug 22

University of California, Merced, CA

- Specialization in Developmental Cognitive Neuroscience with coursework spanning cognitive psychology, neuroscience, and quantitative methods.
- Academic Adviser: Heather Bortfeld

M.S. Electrical Engineering — Jan 14 — Jan 17

Auburn University, Auburn, AL

- Academic and research focus in functional Magnetic Resonance Imaging, signal & image processing and machine learning.
- Graduate Adviser: Gopikrishna Deshpande

B.E. Electrical & Electronics Engineering — Aug 09 — Jul 13

Birla Institute of Technology & Science, Pilani, India

ACHIEVEMENTS

- Developmental Psychology Travel Award at the University of California, Merced for AY 2021-22 ($630).
- Psychology department Graduate Development Support Award at the University of California, Merced for AY 2021-22 ($235).
- 2021 Trainee Professional Development Award (TPDA) by Society for Neuroscience (SFN).
- Donald and Effie Godbold Fellowship at the University of California, Merced for AY 2019-20.
- Developmental Research Award at the University of California, Merced for AY 2019-20 ($550).
- Hatano Cognitive Development Fellowship at the University of California, Merced for AY 2017-18.
- Psychological Sciences Graduate Group Recruitment Fellowship (GGRF) at the University of California, Merced for the AY 2017-18.
- Magna Cum Laude merit award for submitted abstract at the 24th International Society for Magnetic Resonance in Medicine (ISMRM) Annual Meeting in May 2016 in Singapore.
- Graduate student developmental research award by the Psychology department at the University of California, Merced for AY 2020-21 ($150).
- Merit scholarship from BITS Pilani in 2009-2013 for academic achievements.

JOURNAL PUBLICATIONS

Peer-reviewed

- **Lanka, P.**, Bortfeld, H., Huppert, T.J. (2022). Correction of global physiology in resting-state functional near-infrared-spectroscopy. Neurophotonics, in press.

- **Lanka,P.**, Rangaprakash, D., Gotoor, S.S.R., Dretsch, M.N., Katz, J.S., Denney, T.S., & Deshpande, G. (2020). MALINI (Machine Learning in NeuroImaging): A MATLAB toolbox for aiding clinical diagnostics using resting- state fMRI data. *Data in brief*, 29, 105213.
- **Lanka, P.**, Rangaprakash, D., Dretsch, M.N., Katz, J.S., Denney, T.S., & Deshpande, G. (2020). Supervised machine learning for neuroimaging-based diagnostic classification. *Brain Imaging and Behavior*, 14, 2378- 2416.
- **Lanka**, **P.,** & Deshpande, G. (2019). An evaluation of the effectiveness of Prospective Acquisition CorrEction (PACE) for reducing motion-related artifacts in resting state fMRI. *Brain and Behavior*, 9(8), e01341.
- Palaniyappan, L., Deshpande, G., **Lanka, P.**, Rangaprakash, D., Iwabuchi, S., Francis, S., Liddle, P.F. (2019). Effective connectivity within a triple network brain system discriminates schizophrenia from psychotic bipolar disorder at the single-subject level. *Schizophrenia Research*, 214, 24-33
- **Lanka**, **P.**, & Deshpande, G. (2018). Resting state fMRI data from subjects scanned with the EPI-PACE (Echoplanar Imaging – Prospective Acquisition CorrEction) sequence. *Data in Brief*, 20, 2072-2075.
- Jin, C., Jia, H., **Lanka, P.**, Rangaprakash, D., Li, L., Liu, T., Hu, X., Deshpande, G. (2017). Dynamic brain connectivity is a better predictor of PTSD than static connectivity. *Human Brain Mapping*, 38(9), 4479-4496.

Non peer-reviewed

- Deshpande, G., **Lanka, P.**, Youngblood, S.A., (2017). Artificial Intelligence Applications to Military Medicine. *SOAR: Artificial Intelligence and Machine Learning for Defense Applications*,59-70.

PRESENTATIONS

Oral presentations

- **Lanka, P.** (March 2021). Statistical analysis of resting-state fNIRS data. Developmental Psychology Journal Club Spring 2021, University of California, Merced.
- **Lanka, P.** (May 2019). Directional connectivity with fNIRS using Granger causality. Developmental Psychology Journal Club Spring 2018, University of California, Merced.
- **Lanka, P.** (September 2017). Challenges in automated disease classification using Rs-fMRI. Developmental Psychology Journal Club Fall 2017, University of California, Merced.
- **Lanka, P.,** Rangaprakash, D., & Deshpande, G. (May 2016). Perils of Using Cross-validation for Performance Estimation in Neuroimaging-based Diagnostic Classification. Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM), Singapore, 0225.
- **Lanka, P.,** (November 2015). Prospective motion correction for RS-fMRI connectivity. Presented at Imaging Journal Club, Auburn University, AL.

Poster presentations

- **Lanka, P.**, Bortfeld, H. & Huppert, T. (June 2022). Comparison of methods for correcting temporal autocorrelation in resting-state fNIRS data. OHBM 22 Annual meeting.
- **Lanka, P.**, Huppert, T., & Bortfeld, H. (November 2021). Effective correction of physiological noise and head motion artifacts with resting-state fNIRS data. Neuroscience 21 Virtual meeting.
- **Lanka, P.,** Huppert, T., & Bortfeld, H. (October 2021). Effective correction of physiological noise and head motion artifacts with resting-state fNIRS. fNIRS 2021 virtual conference.

- **Lanka, P.,** (January 2021). Correcting for Physiological noise & motion artifacts in resting-state fNIRS data. 4thAnnual Neurophotonics Center Symposium. Virtual event.
- **Lanka, P.**, Huppert, T., & Bortfeld, H. (May 2019). Understanding the neurological basis for shared intentionality and other social interactions using inter-subject connectivity and fNIRS hyperscanning. Berkeley/Stanford Developmental Psychology Symposium, Berkeley, CA. Organization for Human Brain Mapping (OHBM), Singapore, 3723.
- **Lanka, P.,** Rangaprakash, D., Gotoor, S.S.R., Dretsch, M.N., Katz, J.S., Denney, T.S., & Deshpande, G. (June 2018) Machine Learning in NeuroImaging (MALINI): A Matlab-based toolbox for diagnostic classification.
- **Lanka, P.,** Rangaprakash, D., & Deshpande, G. (May 2015). A Comparison of Multiclass Classifiers for Multivariate Pattern Analysis in fMRI. Auburn University-University of Alabama, Birmingham Imaging Retreat, Lake Martin, AL.
- **Lanka, P.,** & Deshpande, G. (October 2015). Artificial Intelligence for Diagnosing Mental Disorders. Graduate Engineering Research Showcase, Auburn University, AL.\

RESEARCH EXPERIENCE

Bortfeld Lab, UC Merced, CA — Aug 17 — Aug 22

Graduate Researcher

- Investigating improved data processing strategies to process resting-state functional near-infrared spectroscopy (fNIRS).
- Investigating shared neural networks of bilingual language control (BLC) and domain-general executive control (EC) with functional magnetic resonance imaging (fMRI).
- Investigating the neural representations of false belief understanding in Spanish-English bilinguals compared to English monolinguals with fNIRS.
- Using Granger Causality to model causal relationships between brain regions with fNIRS

Auburn University MRI Research Center, AL — Aug 14 — Jan 17

Graduate Research Assistant

- Predicting the disease status for neurological disorders such as Alzheimer's, PTSD, Autism and ADHD using fMRI data by implementing several supervised machine learning classifiers in MATLAB.
- Developed classification models for classifying subjects into Controls /PTSD and Bipolar/Schizophrenia.
- Evaluated the impact of prospective motion correction in reducing motion artifacts on MR Signal and functional brain connectivity.
  Coordinated with my research group to scan the dog's brain in the MRI scanner for almost 300 hours to examine its olfactory system.

ACADEMIC TRAINING

- Served as the team lead for the LNLL data science challenge held at UC Merced in exploring the application of machine learning in COVID-19 drug discovery.
- Attended Neurohackademy Summer school in 2020 and 2021 to get trained in reproducible neuroimaging research.
- Visiting Fellowship in Transcranial Magnetic Stimulation, Jan 12-14, 2018, Duke University, Durham, NC.
- fNIRS Advanced Data Analysis Training, May 24-25, 2018, UPMC, Pittsburgh, PA.

- Awarded a certificate of quantitative methods by the department of Psychology at University of California, Merced for completing courses in advanced statistics.

TEACHING EXPERIENCE

Teaching Assistant, University of California, Merced

- PSY 181, Clinical Neuropsychology, Spring 2022
- PSY 130, Developmental Psychology, Fall 2021
- PSY 180, Physiological Psychology, Spring 2020
- PSY 181, Clinical Neuropsychology, Fall 2019
- PSY 180, Physiological Psychology, Spring 2019
- PSY 130, Developmental Psychology, Fall 2018
- PSY 162, Psychology of visual perception, Spring 2018
- PSY 015, Research Methods, Fall 2017

Training in pedagogy

- Certificate of Participation in Pedagogy: Mastering the Classroom with 1st Generation College Students, AY 2017-18.
- Certificate of Participation in Pedagogy: Improving Teaching by Assessing Learning, AY 2018-19. Certificate of Participation in Evidence-Based Teaching Practices, AY 2020-21.

Graduate statistic tutor (AY 2020-21)

- Hold weekly office hours to answer graduate student's questions in statistics about their research and coursework.
- Help graduate students in their research with statistical analyses and provide them with appropriate software tools.
- Provide feedback to graduate students on their study design and research methodology.

SOFTWARE SKILLS

C, Python, R, SPSS, SAS, MATLAB

Abstract

Resting-state functional near-infrared spectroscopy: analytical challenges
Pradyumna Lanka
Doctor of Philosophy in Psychological Science
University of California, Merced, 2022
Professor Heather Bortfeld, Chair

Resting-state functional connectivity has revolutionized neuroimaging in the past 20 years and helped us gain insights into the functional organization of the human brain. Although most studies use functional magnetic resonance imaging (fMRI) for studying functional connectivity, functional near-infrared spectroscopy (fNIRS) has become increasingly popular due to its advantages over fMRI. However, fNIRS data is often corrupted by systemic physiology, extracerebral signal contamination, and motion artifacts which limit its widespread use. These noise sources in resting-state fNIRS data often violate the assumptions of the linear models used to model functional connectivity and may lead to incorrect inferences about statistical significance. This dissertation aims to examine the impact of these noise sources in resting-state fNIRS data and proposes preprocessing strategies and connectivity models that ameliorate their effects and improve the statistical validity of functional connectivity analysis. Chapter 2 examines multiple analysis pipelines using simulated resting-state fNIRS data to find the best strategies to correct for the effects of global systemic physiology on resting-state functional connectivity. Our results indicate that pre-filtering using principal components extracted from short-separation fNIRS channels as part of a partial correlation model was most effective in reducing spurious correlations due to shared systemic physiology. Given the high temporal resolution of fNIRS, modeling the lagged relationships between brain regions is crucial in understanding the flow of information over time. Hence, we also explored methods to model these lagged relationships while controlling for the effects of noise using multivariate Granger causality analysis. Chapter 3 surveys some strategies and techniques that have been proposed to correct for temporal autocorrelation, systemic physiology, and motion artifacts. Together, these two chapters highlight the challenges for resting-state functional connectivity analysis with fNIRS, emphasizing the need for better strategies to remove the effects of noise in the service of obtaining statistically valid inferences about the relationships between cortical regions.

Chapter 1: General introduction

Our understanding of neural correlates of mental processes in the brain often evolves with technological advances in neuroimaging methods, better preprocessing techniques, and advanced statistical analyses. Cognitive neuroscience has evolved from 'blobology' where certain brain regions are associated with complex mental processes, to networks of brain regions involved in such mental processes, to a more recent view that spatio-temporal patterns of neural activity are correlated with cognitive processes. Some of the methods and applications discussed in this dissertation help us build complex spatio-temporal models of cognition and draw statistically valid inferences from such models to help understand cognition. So firstly, I describe the neuroimaging methods that form the basis of this dissertation. Next, I talk about resting-state analysis. Following that, I compare functional magnetic resonance imaging with functional near-infrared spectroscopy on the advantages and disadvantages of these methods with respect to resting-state analysis. I conclude this chapter with the proposed organization of my dissertation studies.

Neuroimaging methods

Functional magnetic resonance imaging (fMRI): FMRI has revolutionized the field of cognitive neuroscience, particularly in allowing us to infer neural correlates of mental processes and behaviors (Poldrack, 2012). fMRI is an indirect measure of neural activity that uses the blood-oxygen-level-dependent (BOLD) contrast to assess changes in local metabolic and hemodynamic parameters to infer the neural activity in specific brain regions. BOLD contrast exploits the differences in magnetic properties of oxygenated and deoxygenated forms of hemoglobin. Hemoglobin has two forms depending on oxygen saturation, oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) which are diamagnetic and paramagnetic, respectively. While HbO does not affect the MR signal, HbR, which is paramagnetic, distorts the surrounding magnetic field (susceptibility), reducing the T2* relaxation time, thereby reducing the MR signal (Buchbinder, 2016). In response to action potentials and metabolic demand in a particular region of the brain, there is increased metabolic oxygen consumption. Increased metabolic demand, in turn, causes an increase in the flow of oxygen-rich blood that outstrips the demand, thus locally increasing the relative proportion of HbO and reducing the concentration of HbR, thus increasing the MR signal a few seconds after the presentation of the stimulus. This measurement of the hemodynamic response to neural activity is the basis for BOLD-fMRI imaging.

Functional near-infrared spectroscopy (fNIRS)*:* Unlike fMRI, which uses the differences in magnetic properties of oxygenated and deoxygenated blood, fNIRS utilizes the differential optical properties of the two forms of hemoglobin to infer brain function. Specifically, fNIRS relies on differences in the absorption of electromagnetic waves by metabolically relevant chromophores such as HbO and HbR in the near-infrared part of the electromagnetic spectrum. The near-infrared (NIR) region of the electromagnetic spectrum (wavelengths of 650-950 nm) is used because of the relative transparency of NIR light to biological tissues, i.e., lower absorption of both water, a major constituent of biological tissue, and hemoglobin, thus allowing light to pass through without significant attenuation (Pinti et al., 2020). Since the concentrations of both forms of hemoglobin

change in response to the metabolic demand, at least two wavelengths of NIR light are needed to resolve the differential absorption for two chromophores, HbO & HbR. Like fMRI, even with fNIRS, we are not directly measuring the neural activity but the relative changes in oxygenated (increase) and deoxygenated (decrease) concentrations in response to the neural activity, termed the hemodynamic response. The hemodynamic response is predominantly caused by the increased regional cerebral blood flow in response to the increased metabolic demands of particular brain regions (Scholkmann et al., 2014).

Continuous wave-fNIRS (CW-fNIRS) is the most used form of fNIRS. It relies on continuously shining NIR light through a light source and a detector placed on either side of brain tissue, and capturing the backscattering of light after probing the neural tissue. Even though the brain is a scattering medium in which attenuation due to scattering is much higher than absorption, a reasonable assumption can be made that attenuation attributable to scattering does not change significantly in response to neural activity, unlike attenuation attributable to absorption. Another common assumption is that the concentration of other chromophores does not change considerably during the session. So, the modified Beer-Lambert law (MBLL) can be used to convert the changes in the light intensity to changes in the relative concentrations of HbO and HbR.

The cortical tissue probed is assumed to be the cortical projection of the midpoint between the source and detectors and is referred to as a channel. Increasing the distance between the source and the detector could lead to considerable light attenuation, whereas shorter distanced channels do not penetrate the cortical tissue. Hence, an optimal range of source-detector distances can be used to probe neural activity from the cortical surface, which is around 3 cm. Unfortunately, CW-fNIRS cannot be used to infer the absolute concentrations of HbO and HbR, but this is not a significant limitation in cognitive neuroscience. We are often interested in changes related to mental processes and task demands compared to a baseline rather than absolute concentrations.

A significant limitation of neuroimaging methods such as fMRI and fNIRS is that the relationship between the brain and a particular mental process is inherently correlational. While these methods allow for forward inference, in which specific brain regions are activated in response to certain cognitive processes induced by certain stimuli and tasks, these methods do not specify whether the activated region is necessary nor sufficient for a particular mental process (Poldrack, 2012).

Resting-state analysis

Resting-state paradigms on functional neuroimaging methods focus on the spontaneous, spatially structured, low-frequency fluctuations (< 0.1 Hz) in the brain when the person is not performing an explicit task. Resting-state functional connectivity was discovered in the mid-90s by Bharat Biswal when he observed the left and right sensorimotor cortices were highly correlated (Biswal et al., 1995). Several sensory (Visual & auditory), higher-order cognitive (default mode network, central executive network, salience network), and several other networks have since been identified (Lee et al., 2013). The brain consumes large amounts of energy at baseline. Despite being just 2% of the body weight, the brain consumes over 20% of the total body energy consumption, with cognitive tasks increasing the local energy consumption by less than 2.5% (Buchbinder, 2016). Thus, the bulk of metabolic demands is from baseline neural

activity. The resting-state analysis focused on the spatial and temporal structure of the low-frequency fluctuations in the baseline neural activity to identify functional brain networks and map the brain's functional architecture.

Processing resting-state data is often more challenging than processing task-based data. Typically, task-based analyses are univariate approaches based on the general linear model (GLM) in which the fNIRS data is fitted as a linear combination of the neural signal (expected neural response based on the stimuli) and the noise is modeled as nuisance regressors (Pinti et al., 2020). Unlike task-based analysis, where the neural response can be approximately modeled, only the noise is often modeled with resting-state analysis since the neural signal is unknown. The modeled noise is then removed/subtracted from the total signal to infer the neural activity. Hence any unmodeled noise will remain and be conflated with the neural signal. This could significantly affect statistical inferences on resting-state connectivity metrics (Bright et al., 2017). Most task-based analyses use univariate approaches rather than bivariate and multivariate approaches used in the resting-state analysis. Hence, methods that work well for removing noise in the task-based analyses do not work well for resting-state analyses.

Comparison of resting-state fNIRS with resting-state fMRI

fMRI had several years head-start compared to fNIRS with resting-state analysis and identifying resting-state networks. However, resting-state analysis with fNIRS is increasing in popularity. Below we compare the advantages and disadvantages of resting-state analysis with fMRI and fNIRS. Though some of these differences result from the nature of the data acquisition with these methodologies, some of the differences are specific to resting-state analyses.

FMRI has high spatial resolution and allows for greater brain coverage, including cortical and subcortical regions; thus, subcortical and cortico-subcortical networks can be inferred from fMRI. Unlike fMRI, fNIRS currently only probes the cerebral cortex and cannot detect neural activity from the subcortical structures; hence resting-state analysis with fNIRS is limited to identifying cortical networks.

Typically, fMRI data is high dimensional with tens of thousands of voxels; often, atlases are used to group the voxels into anatomical or functional divisions ROIs before calculating connectivity measures to reduce the dimensionality of the data. Reducing the dimensionality of the data could make the analysis less computationally intensive, more interpretable, and reduce the statistical errors associated with multiple hypothesis testing. With fNIRS, the number of channels can range from a few channels to at most a few hundred, depending on the number of sources, detectors, and the instrument's configuration. So, fNIRS data is low-dimensional compared to fMRI; thus, we are limited to probing the neural activity from a few cortical areas. Processing fNIRS data is, therefore, less computationally intensive.

FNIRS has a higher temporal resolution than fMRI, with typical sampling rates around 10 Hz, though both methods estimate the slow and lagged hemodynamic response to neural activity. Oversampling of the hemodynamic signals can have certain advantages with task-based fNIRS, such as better estimation of the shape of the HRF, prevention of aliasing of relatively high-frequency cardiac noise (around 1-1.2 Hz), and better statistical power. A higher sampling rate in resting-state data can lead to increased autocorrelation and, consequently, increased false positives (Santosa et al., 2017). The

higher temporal resolution also necessitates using time-lagged connectivity methods such as Granger causality to model the relationship between different regions in the brain as the zero-lag metrics such as functional connectivity may not completely capture the relationship between brain regions.

Similarly, fMRI is extremely sensitive to head motion, especially with resting-state fMRI, as head movements can have lingering effects on the fMRI signal. Uncorrected motion artifacts in the resting-state fMRI lead to distance-dependent artifacts in which connectivity varies as a function of distance since the signal changes in response to head motion are similar across nearby regions and more dissimilar across farther regions (Power et al., 2012, 2015; Satterthwaite et al., 2012). Removal of motion artifacts involves removing the motion corrupted time points and their neighboring time points, thus reducing the degrees of freedom of our data. Though motion artifacts affect the fNIRS signal, they have momentary effects. Even after motion artifact correction, fNIRS can have more degrees of freedom due to its higher sampling rate leading to more data points in fNIRS for the same duration of data collection.

With fMRI, an anatomical image is often also captured along with the rest scans. The anatomical image is co-registered to the functional scan, which allows for better registration later onto a standard atlas such as MNI-152, leading to lower inter-subject variability and the ability to draw better group-level inferences. However, with fNIRS, anatomical images are often unavailable, so either the 10-20 system is used for localizing the probe, or 3D digitizers are used to capture anatomical locations of optodes and the corresponding cortical regions probed. The 10-20 system, due to lack of spatial specificity, may not be appropriate for fNIRS due to its relatively higher spatial resolution. Hence the use of digitizers to capture optodes locations and cranial landmarks and transform them from the subject space to the common standard space for group-level analysis is more appropriate. Given the limited information available from the cranial coordinates, even 3D digitizers may still lead to inaccurate spatial localization, greater inter-subject variability, and difficulty with group-level inferences with fNIRS compared to fMRI.

Finally, with fNIRS, since the light passes through extracerebral layers before and after penetrating the neural tissue, fNIRS signals are more contaminated with extracerebral physiological changes compared to fMRI. While CSF and WM regression are used to remove physiological noise in fMRI, short separation (SS) channels in which the source-detector distance is typically less than 10 mm, just enough to penetrate the extracerebral layers but not enough to penetrate the cortical tissue, can be used to capture the extracerebral signal changes. These SS channels can then be used to partially ameliorate the effects of extracerebral signal changes and systemic physiology in the fNIRS data (Pinti et al., 2020).

Organization

In chapter 2, methods to reduce the impact of physiological noise on resting-state fNIRS analysis are proposed. Although several other methods have been proposed to reduce physiological noise in resting-state fNIRS, no method currently accounts for temporal autocorrelation, global signal, and motion artifacts. However, we argue that the robust AR partial correlation proposed in Chapter 2 does exactly that. Further, as suggested earlier, modeling lagged relationships becomes particularly important with

fNIRS given the higher sampling rate. Thus, we propose using modified multivariate Granger causality (MVGC)—an extension of traditional Granger causality—that can model both lagged and zero-lag relationships. Furthermore, the robust version of the modified MVGC can estimate robust statistical estimates of fNIRS data corrupted by head motion.

Chapter 3 of the dissertation surveys methods and preprocessing strategies for reducing the impact of noise associated with systemic physiology and head motion, which themselves lead to autocorrelation, a global signal, and large fluctuations in resting-state fNIRS data. With over 150 publications currently focused on resting-state fNIRS analyses and its rapidly increasing use, it is critical to assess the current status of research in resting-state analysis with a specific focus on preprocessing strategies. While current preprocessing methods to correct for systemic physiology and head motion are predominantly tested with task-based analysis in mind, their relevance and effectiveness for resting-state analysis is examined. Autocorrelation in resting-state fNIRS data is a huge issue that is further exacerbated by the increased physiological noise and high temporal resolution. Unfortunately, this issue has received scant attention. Thus, this chapter also highlights the effect of temporal autocorrelation on resting-state functional connectivity, along with strategies to remove its influence.

Finally, I conclude the dissertation by highlighting some of the limitations of the current work, suggesting directions for future research that addresses these limitations and contributes to our understanding of functional brain organization.

Chapter 2: Correction of global physiology in resting-state functional near-infrared spectroscopy

Chapter abstract

Resting-state functional connectivity (RSFC) analyses of functional-near infrared spectroscopy (fNIRS) data reveal cortical connections and networks across the brain. Motion artifacts and systemic physiology in evoked fNIRS signals present unique analytical challenges, and methods that control for systemic physiological noise have been explored. Whether these same methods require modification when applied to resting-state fNIRS (RS-fNIRS) data remains unclear. We systematically examined the sensitivity and specificity of several RSFC analysis pipelines to identify the best methods for correcting global systemic physiological signals in RS-fNIRS data. Using numerically simulated RS-fNIRS data, we compared the rates of true and false positives for several connectivity analysis pipelines. Their performance was scored using receiver operating characteristic (ROC) analysis. Pipelines included partial correlation and multivariate Granger causality, both with and without short-separation measurements, and a modified multivariate causality model that included a non-traditional zeroth-lag cross term. We also examined the effects of pre-whitening and robust statistical estimators on performance. Consistent with previous work on bivariate correlation models, our results demonstrate that robust statistics and pre-whitening are effective methods to correct for motion artifacts and autocorrelation in the fNIRS time series. Moreover, we found that pre-filtering using principal components extracted from short-separation fNIRS channels as part of a partial correlation model was most effective in reducing spurious correlations due to shared systemic physiology when the two signals of interest fluctuated synchronously. However, when there was a temporal lag between the signals, a multivariate Granger causality test incorporating the short-separation channels was better. Since it is unknown if such a lag exists in experimental data, we propose a modified version of Granger causality that includes the non-traditional zeroth-lag term as a compromising solution. A combination of pre-whitening, robust statistical methods, and partial correlation in the processing pipeline to reduce autocorrelation, motion artifacts, and global physiology are suggested for obtaining statistically valid connectivity metrics with RS-fNIRS. Further studies should validate the effectiveness of these methods using human data.

Introduction

Spontaneous or resting-state functional connectivity (RSFC) has been an important tool for characterizing the network architecture of the human brain. Biswal and colleagues first noted that synchronized spontaneous fluctuations in the functional magnetic resonance imaging (fMRI) blood-oxygen-level-dependent (BOLD) signal were organized into distinct networks that were conserved across subjects (Biswal et al., 1995). More than twenty-five years later, these results have been replicated in hundreds of research studies and expanded to nearly every aspect of human brain function, development, and disease (Badhwar et al., 2017; Betzel et al., 2014; Hull et al., 2017; Lee et al., 2013; Oldham & Fornito, 2019; van den Heuvel & Hulshoff Pol, 2010; Vértes & Bullmore, 2015). Resting-state networks have also been observed using other neuroimaging modalities, including electroencephalography and

magnetoencephalography (Gaudet et al., 2020; van Diessen et al., 2015), as well as functional near-infrared spectroscopy (fNIRS) (Hu et al., 2020; Niu & He, 2014). A major advantage of the resting-state paradigm is that there is no task compliance necessary, which is especially useful in research with infant, aging, diseased, and other non-compliant populations.

Not surprisingly, several analytical methods have been developed to quantify resting-state connectivity. For both fMRI and fNIRS, these methods often focus on the zeroth-lag relationship between brain regions. Zeroth-lag relationships are ones in which signal fluctuations occur simultaneously (within measurement sampling time) in both signals. Perhaps the most common method for estimating functional connectivity is to calculate the Pearson's correlation coefficient between the time-series from two brain regions. There are several ways to characterize resting-state functional connectivity in the time-domain, including (i) seed-based connectivity, (ii) pairwise connectivity, and (iii) independent component analysis (ICA). In seed-based connectivity, a seed region is selected, and its connectivity with all other brain regions is assessed (Mesquita et al., 2010). In pairwise connectivity, the correlation between all possible pairs of channels is calculated (Homae et al., 2010). ICA decomposes the covariance of fNIRS channel data into several statistically independent spatiotemporal components, with some components representing functional networks in the brain either for each subject individually or for the entire dataset concatenated across subjects (Blanco et al., 2021; Ferradal et al., 2016).

Granger causality is an example of an "effective connectivity" method, which relies on the lagged relationship between fNIRS channels in the time-domain. Granger causality analysis (GCA) is based on the principle of Granger causality, which states that for two signals, $\boldsymbol{x}$ and $\boldsymbol{y}$, if the past values of signal $\boldsymbol{x}$ contain information that helps predict $\boldsymbol{y}$ above and beyond the information contained in past values of $\boldsymbol{y}$ alone, then we can say $\boldsymbol{x}$ Granger-causes $\boldsymbol{y}$ (Bressler & Seth, 2011; Seth et al., 2015). The term "causal" refers to the mathematical predictability that the history of the time course of $\boldsymbol{x}$ has to the current value of $\boldsymbol{y}$, rather than the more linguistic context of "cause and effect". GCA is statistically tested by using a likelihood ratio test comparing the goodness-of-fit of a linear model containing both $\boldsymbol{x}$ and $\boldsymbol{y}$ terms to a restricted model containing only $\boldsymbol{y}$ terms. In the case of a multivariate model, this is a likelihood ratio test of a model containing $\{\boldsymbol{x}, \boldsymbol{y}$, and other confounds $\boldsymbol{Z}\}$ to one containing only $\{\boldsymbol{y}$ and other confounds $\boldsymbol{Z}\}$. GCA is especially suited for modeling time-lagged relationships between brain regions. It benefits from computational simplicity without the need to prespecify the direction of influence, unlike the other effective connectivity methods including structural equation modeling and dynamic causal modeling, which are more confirmatory than exploratory. Although several studies have looked at Granger causality in task-based fNIRS paradigms (Z. Liu et al., 2017; Wan et al., 2018; G. Zhou et al., 2016), there is a paucity of studies that examine GCA in resting-state paradigms.

Another method commonly used to model functional connectivity in fNIRS is the wavelet transform coherence (WTC), which models the relationship between the channels in the time-frequency domain. WTC relies on the continuous wavelet transform (CWT). CWT is the convolution of the scaled and shifted versions of a mother wavelet, and transforms the time series data into the time-frequency domain. Often the Morlet wavelet, a gaussian scaled sine wave, is chosen as the mother wavelet. WTC is estimated

as a cross-correlation between the time series in the time-frequency domain (Chang & Glover, 2010; Grinsted et al., 2004; Tan et al., 2015).

Functional near-infrared spectroscopy

FNIRS is an optical imaging method that uses light in the near-infrared wavelength range (650-950 nm) to estimate temporal changes in the concentrations of oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) (Delpy & Cope, 1997). In this range of red to near-infrared light, the differential optical absorption of HbO and HbR allows for estimation of hemodynamic signals in the brain through a set of optical sources and detectors placed non-invasively on the scalp of participants (Pinti et al., 2020; Scholkmann et al., 2014). Light from these optical sources passes through the skull and scatters through the tissue and can reach the outer few millimeters of the cerebral cortex allowing detection of underlying brain signals. Thus, fNIRS measures the hemodynamic changes in the brain reflecting changes in blood flow, volume, and blood oxygenation and therefore is an indirect measure of neural activity like the fMRI BOLD signal. FNIRS has some advantages compared to fMRI including lower cost, a greater resiliency to head motion (provided the optical sensors remained secured to the scalp), and higher data sampling rates (although still measuring slow hemodynamic signals). FNIRS is also portable and is commercially available in participant-wearable systems, which allows subject movement, application to a wider range of populations that would have contraindications in other modalities, and the ability to perform more flexible or ecologically valid tasks.

However, the disadvantages of fNIRS are its lower spatial resolution, limited depth penetration into the brain, and increased sensitivity to superficial layers of the head relative to the brain itself. The latter issue is particularly confounding for fNIRS measurements, since systemic physiology in the scalp, including cardiac, respiratory, and blood pressure fluctuations, often result in spatially global noise, which is often larger in amplitude than the underlying brain signals of interest (Huppert et al., 2009; Tachtsidis & Scholkmann, 2016). Thus, in the case of RSFC using fNIRS, the separation of the overlying spatial networks due to superficial physiological signals and true underlying neural-related brain networks is challenging.

The first study to demonstrate the feasibility of fNIRS in capturing the very-low frequency and low-frequency oscillations that contribute to resting-state functional connectivity was by Obrig et al., (2000). Following that, several early studies used to investigate the feasibility of RSFC found reliable and expected visual, sensorimotor, and language networks using functional near-infrared spectroscopy (fNIRS) and diffuse optical tomography (DOT) (Lu et al., 2010; Mesquita et al., 2010; White et al., 2009; Y.-J. Zhang, 2010). Since then, researchers have used resting-state fNIRS (RS-fNIRS) and DOT to study infant development (Homae et al., 2010), gender differences in the prefrontal cortex (Chuang & Sun, 2014), preterm birth (Fuchino et al., 2013), language networks (Gallagher et al., 2016), autism spectrum disorder (J. Li et al., 2016; Zhu et al., 2014), affective disorders (Zhu et al., 2017), and aging (Tan et al., 2016). Studies examining the test-retest reliability of RSFC have found good to excellent intraclass correlations across both individual and group level RSFC maps across multiple sessions ranging from an hour to a week (Niu et al., 2013; H. Zhang, Zhang, et al., 2011). More importantly, RSFC with fNIRS (RS-fNIRS) has also been validated with RS-fMRI (Duan

et al., 2012; Sasai et al., 2012). Although extremely popular with fMRI, just a few papers have applied graph-theoretic approaches to RS-fNIRS (Niu et al., 2013; Novi et al., 2016).

Statistical properties of RS-fNIRS signal

RS-fNIRS has some unique challenges compared to RS-fMRI given the signal and noise properties of RS-fNIRS. Due to the way the fNIRS signal is collected, it is sensitive to non-neural extracerebral signal changes in the scalp. RS-fNIRS signal is composed of 3 components: (i) low-frequency neural oscillations (signal of interest), (ii) intracerebral physiological noise originating in the brain, and (iii) extracerebral physiological noise originating in the scalp and non-brain tissues. All the signals are non-evoked, given the nature of the resting-state paradigm. Thus, fNIRS signals are contaminated by both extracerebral signal changes as well as non-neural cerebral signal changes due to systemic physiology attributed to cardiac (around 1-1.2 Hz), respiratory (0.3-0.6 Hz), and blood pressure/Mayer wave (0.1 Hz) fluctuations (Pinti et al., 2019; Tachtsidis & Scholkmann, 2016).

Physiological noise has two effects on the RS-fNIRS signal: (i) inducing temporal autocorrelation, and (ii) increasing spatial covariance between channels across the brain. Both these issues and how they affect the statistical assumptions of the connectivity models used to identify the interactions between time series will be discussed in detail below. The proposed ways to correct for them will be discussed in the subsequent sections.

Temporal autocorrelation

Previous work has demonstrated that the high sampling rate and the systemic physiology (heart rate, respiration and blood pressure), combined with the underlying hemodynamic response, can together lead to temporal dependencies, and hence autocorrelation in RS-fNIRS time series (Santosa et al., 2017). Colored noise reduces the effective degrees of freedom. Autocorrelation may be present for up to several seconds, thus lasting multiple time points depending on the sampling rate (Christova et al., 2011; Santosa et al., 2017). In this case, the assumption of independence of cases is no longer valid and using linear regression/correlation to model the relationship between autocorrelated time series could lead to increased false positives (Granger & Newbold, 1974). The reason for the increased false positives can be attributed to the underestimation of standard errors of the null distribution of the regression coefficients, as well as a possible bias in the estimates (if the autocorrelation structures are different across the time series) due to the temporal autocorrelations in the time series (Arbabshirani et al., 2014), neither of which are considered when drawing statistical inferences.

An effective way to remove the autocorrelation in the time series to obtain more statistically valid estimates of relationships is a procedure called pre-whitening. Pre-whitening works by removing the serial correlations in the time series. Autoregressive (AR) models can be used for pre-whitening the data and removing the serial correlations. Specifically, for fNIRS, using pre-whitening entails using a p^{th} order AR model to model the current time point as a function of the previous time points. The residual term in the AR model after removing the temporal dependencies is termed the innovation term, and is now used to calculate the Pearson's correlation coefficient between channels rather

than using the autocorrelated channel data (Santosa et al., 2017). In a study comparing the use of standard correlation without pre-whitening and AR correlation with pre-whitening on simulated data, Santosa et al. reported a false discovery rate (FDR) as high as 50% at 1 Hz sampling rate and increasing to 70% for a sampling rate of 10 Hz when the standard correlation is used to model the functional connectivity between fNIRS (Santosa et al., 2017). With the use of AR-correlation with pre-whitening incorporated, however, the uncontrolled FDR reduced to the expected 5% at 1 Hz sampling rate and a higher 30% FDR at 10 Hz sampling rate. Similar findings were also reported for experimental fNIRS data, thus demonstrating a reduced type-I error rate when pre-whitening is incorporated into the preprocessing pipeline for estimation of functional connectivity (Santosa et al., 2017). Despite correcting for the temporal autocorrelations in fNIRS, Santosa et al. did not account for the increased spatial covariance due to systemic physiology, which could also lead to increased FDR.

Global systemic physiology

Since systemic physiology is thought to be relatively homogenous across the brain, the signal changes caused by the systemic fluctuations that are shared across all the channels can increase the spatial covariance in the channels (Blanco et al., 2018). This can artificially inflate the strength of the relationship between the time series due to a third term (physiological noise). Figure 1 shows that not correcting for the increased spatial covariance in the fNIRS data could lead to increased type-I errors and reduced specificity, as far more connectivity paths are present than expected based on the "ground truth"

.

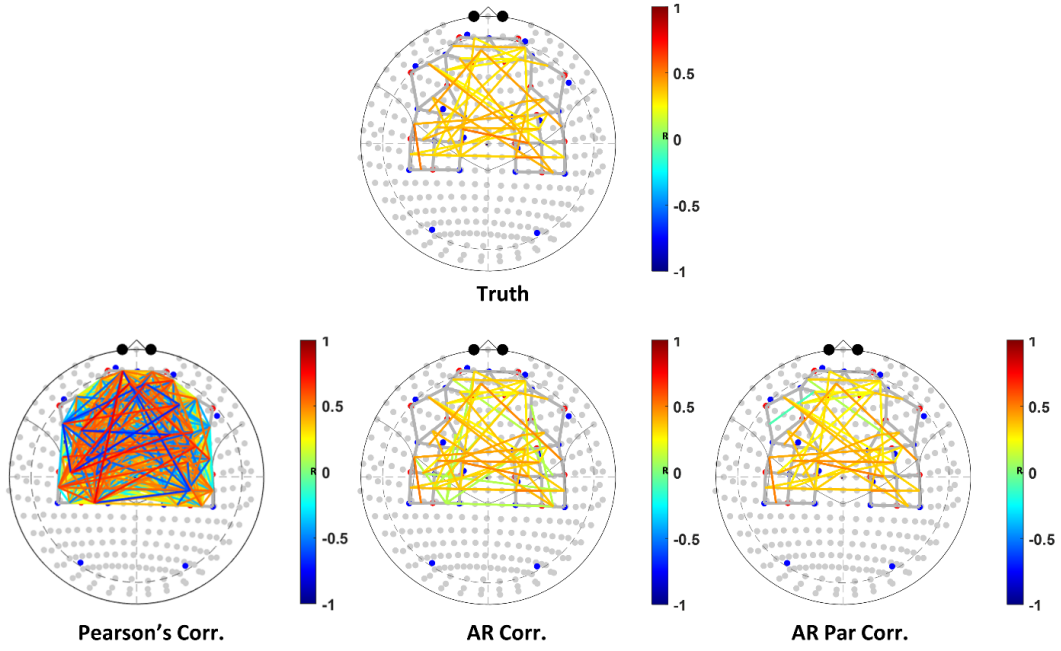

Figure 1. Connectivity maps show the effectiveness of (i) Pearson's correlation coefficient, (ii) autoregressive (AR) correlation, and (iii) AR partial correlation, to

recover true connectivity, in simulated fNIRS data with added physiological noise mimicking AR smoothing and shared covariance. The figure demonstrates the success of AR partial correlation to recover most of the true connectivity patterns in simulated data with fewer false positives.

Proposed solutions to correct for the effects of physiological noise

Several solutions have been proposed to deal with reducing the influence of physiological noise on the violations of assumptions of connectivity due to them. Most methods that have been proposed to remove the effects of physiological data do not account for the autocorrelation in the fNIRS time series, but instead aim to improve the signal-to-noise ratio by removing the influence of physiological noise and non-neural signals.

Generally, the most common method to reduce the interference of physiological noise in fNIRS data is band-pass filtering. Brain-derived signals, exhibit a power spectral density in accordance with the inverse-power law with greater power at lower frequencies and much less power at higher frequencies (Blanco et al., 2018). It must be noted that with fNIRS, we are measuring the hemodynamic/vascular changes associated with neural activity rather than the neural activity itself, so the hemodynamic response function (HRF) acts as a temporal low-pass filter. Given the inverse-power law and the low-pass filtering properties of HRF, there is little high-frequency information present in the RS-fNIRS signal. Finally, the fNIRS signal is also corrupted by high-frequency noise structures such as cardiac and respiratory noise, thus a good signal-to-noise ratio of neural activity compared to noise is often restricted to low-frequency bands in fNIRS. The fNIRS signal also generally has very-low-frequency drift noise which can be another source of autocorrelation. The very-low-frequency drift noise can be modeled and removed by linear, polynomial, or discrete cosine transform (DCT) basis functions, or by filtering the data with an appropriate high-pass filter. Thus, using an appropriate band-pass filter could remove physiological noise, while still preserving the low-frequency neural oscillations (Huppert et al., 2009). It must be noted that low-pass filtering/band-pass filtering often increases the autocorrelation in the time series, whilst reducing the degrees of freedom. Consequently, low-pass/bandpass filtering also increases the false positive rate (Huppert, 2016).

An alternative approach to removing spatial noise is to use spatial principal component (PCA) filtering (Huppert et al., 2009; Y. Zhang et al., 2005). PCA depends on the idea that shared spatial covariance between the signals is due to shared physiology, and the variance is much larger in magnitude compared to the neural signal. So, the first few components obtained from PCA can be used to extract the spatially correlated global signal. Regressing out the PCA components from the fNIRS data provides a way to reduce the shared global signal due to systemic physiology, while still preserving neural-related components.

Still another way to reduce the presence of the physiological signals is to use external recording devices to measure respiration, heart rate and blood pressure, and remove their effects from the RS-fNIRS signal (Kirilina et al., 2012; Scholkmann et al., 2014; Tachtsidis et al., 2010). However, this method is often expensive as it requires using additional instruments to record the physiological signals. Moreover, reducing physiological noise may not entirely remove the autocorrelation in the time series. More

recently, the use of short-separation channels has become a popular approach to estimate local systemic physiology. Because the depth of penetration of light into the tissue is a function of the spacing between the fNIRS light source and detector positions, typically for measuring the brain, a spacing of about 3-4 cm is used, which provides penetration to a few millimeters into the cortex. When short-separation (~5-10mm) spacings are used, the light only reaches into the scalp and can be used as a regressor of superficial physiology (Sato et al., 2016). The use of these short-separation fNIRS measurements has been shown to be an effective way to remove the global effects of systemic physiology for evoked (task-based) functional signal (Santosa et al., 2020).

Previously, several approaches using these short-separation channel measurements been proposed, including as pre-filtering operators or as regressors within linear regression models (Gagnon et al., 2012; Saager et al., 2011; Saager & Berger, 2005; Santosa et al., 2020; Sato et al., 2016; Yamada et al., 2009). However, shorter distance channels can introduce engineering challenges when another detector has to be placed so close to the source, and it can also cause saturation issues (Brigadoi & Cooper, 2015). Placing the detector too far from the source can introduce false negatives in the data because the short-separation channel can capture neural-related variance. Studies have identified the optimum distance for a short-separation channel for adults is 8.4 mm and 2.15 mm for infants using Monte-Carlo simulations (Brigadoi & Cooper, 2015). Meanwhile, hybrid approaches that use short-separation channels together with PCA appear to be more effective in reducing systemic physiology (Wyser et al., 2020; X. Zhou et al., 2021).

One final method for reducing the impact of global systemic physiology on fNIRS-based RSFC data is to use partial correlation instead of Pearson's correlation coefficient. Partial correlation is better at removing the shared covariance between the signals and is more sensitive to the relationship between deep brain signals that reflect true neural connectivity (Sakakibara et al., 2016). Connectivity values obtained from the partial correlation between channels are lower than those obtained from Pearson's correlation due to the removal of extracerebral contributions to the connectivity. However, using partial correlation by controlling for other channels can also often remove neural variance as well, thus increasing the rate of false negatives.

Head motion artifacts

In addition to physiological noise, motion artifacts are another source of error in fNIRS data. Motion artifacts occur when the head movements of the participant cause the optodes to slide or momentarily lose contact with the scalp, leading to spikes in signal intensity or changes in baseline signal intensity (Brigadoi et al., 2014). Motion artifacts are common in datasets from certain populations, such as infants and young children, as well as in experimental paradigms that require movement. Usually, the signal changes induced by head motion are much larger than the signal changes due to neural activity (Barker et al., 2013). Motion artifacts can affect only a few probes or most of the probes, thus increasing or reducing the shared covariance between the channel time series (Santosa et al., 2017).

The simplest way to address the effect of motion artifacts is to remove the motion corrupted time points from analysis, although it has been argued that correcting for motion rather than simply removing the motion corrupted time points is a better strategy

(Brigadoi et al., 2014). For RS-fNIRS, completely removing motion corrupted time points from analysis may reduce the stability of connectivity metrics due to the reduced number of time points/degrees of freedomM. Methods to correct for motion artifacts include spline interpolation (Scholkmann et al., 2010), wavelet filtering (Molavi & Dumont, 2012), PCA (Huppert et al., 2009; Y. Zhang et al., 2005), discrete Kalman filtering (Izzetoglu et al., 2010), and correlation-based signal improvement (CBSI) (Cui et al., 2010). Although these methods are discussed in detail in Brigadoi et al. (2014), in the context of task-based analyses, they are also applicable to RS-fNIRS data. For task-based fNIRS analysis, wavelet filtering was identified as the best performing method relative to the other options (Brigadoi et al., 2014). Since then, several other methods and improvements to existing methods have been proposed with varying degrees of success in correcting for multiple manifestations of motion artifacts in the fNIRS data. Some of those methods include kurtosis-based wavelet filtering (Chiarelli et al., 2015), temporal derivative distribution and repair (TDDR) (Fishburn et al., 2019), and global variance of temporal derivatives (GVTD) -based motion censoring (Sherafati et al., 2020).

Alternatively, motion artifacts in the fNIRS data can be corrected within the context of a generalized linear model by using robust statistics. Head motion present in the resting-state data violates the assumptions of non-normality, as the motion contaminated time points could appear as outliers in the distribution of the innovation terms (Barker et al., 2013; Huppert, 2016; Santosa et al., 2017). Using pre-weighting by downweighting time points considered outliers can reduce the impact of the motion contaminated time points on calculated connectivity metrics such as correlation (Santosa et al., 2017). Santosa et al. proposed a general linear model (GLM) that was robust to the violations of assumptions using methods such as pre-whitening and pre-weighting such that the inferences drawn about the parameters were still statistically valid, unlike traditional methods that do not explicitly account for the violation of assumptions (Santosa et al., 2017).

Aims

In this article, we examine the performance of several approaches to correcting systemic physiology and global signal artifacts for RS-fNIRS as a tool to track functional connectivity between different brain regions. To this end, we quantified the sensitivity/specificity and false positive rates of various numerical methods, including several approaches that use short-separation fNIRS measurements as either pre-filtering/partial correlation methods, as well as within multivariate correlation and causal models. This work is a multi-channel/multivariate expansion of our previous examination of bivariate correlation analysis for fNIRS (Santosa et al., 2017). Santosa et al. demonstrated that the slow hemodynamic signals measured by fNIRS result in spurious correlations, very high false positive rates, and uncorrected type-I error in the standard Pearson correlation model unless corrected for noise autocorrelation and other non-spherical statistical errors (Santosa et al., 2017). However, their work did not address the issue of spatially correlated noise due to superficial systemic physiology, the topic of the current research.

Methods

In order to characterize the sensitivity and specificity of each of the proposed analysis methods, we relied on numerically simulated data with a known ground truth.

We explored the model performance over several classes of simulations ranging from simple normally distributed random noise (where all the statistical assumptions of our model are valid) to "physiological" signals with temporally autocorrelated noise to global spatial physiological noise. Finally, we included the presence of motion artifacts. Each step introduced additional violations to the statistical assumptions of the model, requiring modifications for model generalization. Each data simulation represented a more challenging (and experimentally plausible) statistical problem, which we examined using receiver operator characteristics (ROC) analysis by simulating true and null correlations in the data. All of the methods were implemented in the NIRS AnalyzIR toolbox (Santosa et al., 2018) using MATLAB (R2019b) (MathWorks, Natick, MA).

Data simulation

FNIRS data were simulated for a probe consisting of 16 long-distance channels (27 mm) and 9 short-separation channels (7.1 mm) using a multivariate normal distribution with a specified spatial covariance matrix. The data were simulated assuming near-infrared wavelengths of 690 nm and 830 nm. This known spatial covariance generates the underlying "true" correlations in the data, upon which the various algorithms are assessed. These true correlations were generated with either a zero lag (no time shift) and/or a temporal lag between channels (mathematically causal model). The target "true" correlations were simulated in channel-space for only the long-distance channels. To add additional noise to the model, a semi-infinite homogenous slab model was used to compute the optical forward model (sensitivity of each channel to the underlying "brain" volume). Noise was added to the voxels in the first 2.5 mm modeling a superficial skin layer. For some simulations, superficial noise was temporal and/or spatially smoothed as detailed below. The optical forward model was then used to project these voxels to channel space and added to the base truth correlation signals.

For each simulation, we generated 300 seconds of resting-state data at a 4 Hz sample rate. To generate the ROC curves, a total of 200 data sets were generated for each test. In the ROC analysis, of all possible connections between any two channels in the probe, true connections were simulated for around 10% for the possible connectivity paths, ensuring that global systemic noise was the dominant spatial noise feature. So only 10% of the non-diagonal elements in the covariance matrix used for generating the simulated data were non-zeros. Any significant connections between the remaining 90% of the paths are false positives due to temporal autocorrelation, global systemic physiology or motion artifacts. To generate true and false positive estimates, an equal number of true and null connections were randomly taken from the total adjacency matrix.

Random noise simulations

For the initial simulations, the normally distributed, white noise model described above was used to generate the ROC curves. No additional superficial noise was added to the data. Thus, this set of simulations lacked any realistic "physiological" noise and consisted of both temporally and spatially uncorrelated noise. However, this set of simulations is consistent with the statistical assumptions of the standard Pearson correlation model, which assumes normally distributed, independent, random samples. Results for random noise simulation are not shown.

Physiological temporal noise simulations

In the second set of simulations, serially correlated noise structures were introduced to mimic the temporal structure of physiological noise and the hemodynamic response function. In the voxels of the "skin" layer of the model, a random noise model was convolved with an auto-regressive noise structure of model order 10 to generate physiologically colored noise. These signals were then projected through the optical forward model to channel-space and added to the base true correlation signals. For this set of simulations, only temporally structured noise was added as there was no spatial structure to the noise across channels. The ratio of the simulated neural components to the superficial simulated physiological noise was 1:1 in the simulated data. The signal-to-noise ratio of the neural signal to the random white noise in these simulations was 100:1 (40 dB). The white noise models the shot noise and the signal-to-noise ratio is in the range previously reported (Joseph et al., 2006; Zeff et al., 2007).

Physiological temporal and spatial noise simulations

In the third set of simulations, an additional spatial noise structure was introduced in addition to temporal autocorrelation. In this set of simulations, the skin layer voxels were spatially smoothed using a Gaussian spatial kernel with a full-width half-max of 150 mm. These voxels were projected into channel space to generate a "global" systemic noise model. Since we used the optical forward model, the projection of this noise to the long-distance and short-separation NIRS channels was consistent with the skin origins of these signals. As with earlier simulation, the ratio of the simulated neural signal to the superficial simulated physiological noise was 1:1 in the simulated data and the ratio of neural signal to the white noise was 100:1.

Motion artifact simulations

Two types of motion artifact were added to the time series to mimic the motion artifacts encountered in fNIRS time series: shift artifacts and spike artifacts. These artifacts were added to the base simulations described above (random noise, temporal noise, or spatial and temporal noise). The simulated motion artifacts that were added were similar to the motion effects observed in child imaging studies with moderate levels of head motion (Perlman et al., 2014).

Shift artifacts: The rate of shift artifacts added to the data was around 0.5 per minute or 1 shift artifact for every 480 time points if the data is assumed to be sampled at 4 Hz. The shift artifact was modeled as a scalar shift in amplitude added to the time series sampled from a normal distribution with a zero mean and a standard deviation of 5 times the original data.

Spike artifacts: Similarly, the rate of spike artifacts added to the data was around 2 per minute or 1 spike artifact for every 120 time points of the data for data sampled at 4 Hz. The spike artifact was modeled with a Laplacian distribution function with the peak amplitude sampled from a normal distribution with a mean zero and a standard deviation of 5 times the standard deviation of the original data. The scale parameter was randomly sampled from a uniform distribution of 0.05-5. This ensures that the spike artifact occurs at the same time across multiple channels, but the amplitude of the spike artifact varies across the channels.

Experimental fNIRS data

fNIRS data acquisition

We also wanted to test the effectiveness and applicability of our functional connectivity methods to experimental resting-state data. NIRS data were recorded using a commercial NIRScout-2 (NIRx GmbH, Berlin, Germany) continuous-wave fNIRS system with short-separation measurements. A total of 50 channels (42 channels for long-distance and 8 channels for short separation measurements) were distributed across the bilateral frontal and sensorimotor cortices (see Figure 2). Long-distance channels comprised 16 source optodes (orange circles) and 13 detector optodes (blue rectangles) placed on the scalp, as shown in Figure 2. A detector optode split into 8 detectors (green diamonds) was used for short-separation channels in eight locations across the scalp. The solid line represents long-distance and the green dotted line represents the short-separation channels. The distances between source and detector were 23-50 mm and 7.5 mm for long-distance and short-separation channels, respectively with the light blue solid line in Figure 2 representing the channels. Data for two wavelengths (760 and 850 nm) were recorded at a sampling rate of 7.8125 Hz. After positioning the headcap, signal quality was optimized using the NIRx Aurora software. The ambient light was blocked using an opaque, black shower cap placed on the participant's head during acquisition.

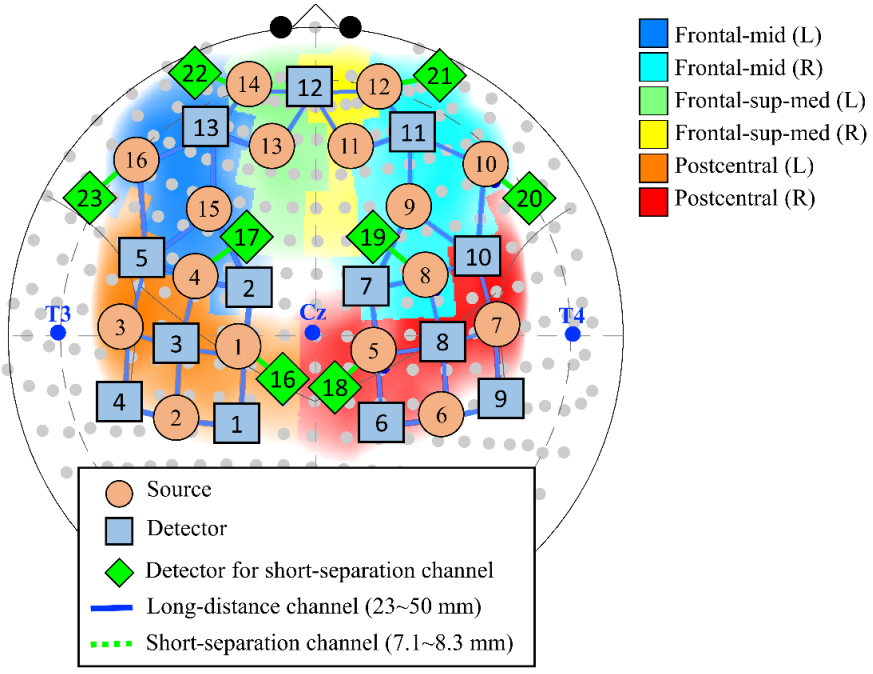

Figure 2. The schematic shows the placement of probes on the bilateral frontal and sensorimotor cortices superimposed on a 10-10 coordinate system for the experimental fNIRS data. 16 sources and 13 detectors optodes formed a total of 42 long-distance channels. Additionally, a detector optode was split into 8 short-separation detectors to give 8 short-separation channels. The data acquired from these probes in 24 subjects was used to compare the effectiveness of zero-lag correlation methods.

Participants & task

About 24 healthy subjects participated in the experiment (9 males and 15 females; mean age 27.4 years, s.d. 8.7 years; 23 right-handed). The subjects were informed about the experimentation and written consent was obtained. The study was approved by the University of Pittsburgh Institutional Review Board. Each subject performed one session of the resting-state scan along with other task scans. The subjects were instructed to minimize body motion and remain relaxed in the sitting position for 5 min without employing any mental effort for the duration of the resting-state scan.

Data pre-processing

We converted simulated fNIRS signal intensities to changes in optical density, and then converted optical density units to HbO and HbR concentrations using the modified Beer-Lambert law (MBLL) with a differential path length factor of 6 and a partial volume correction of 60 for both wavelengths. The data was preprocessed in a similar manner compared to the simulated data except for downsampling the experimental fNIRS time series. After the raw signal intensity was converted to HbO and HbR concentrations using MBLL, the experimental RS-fNIRS data was downsampled to 4 Hz from 7.8125 Hz to match the sampling rate of the simulated data. For the remainder of the article, the analyses focus on just changes in the HbO concentrations and the findings should be applicable for HbR as well.

Figure 3 shows an example fNIRS time series from the simulated data for three conditions, (i) just temporal autocorrelation, (ii) with both temporal autocorrelation and global systemic physiology, and (iii) with both temporal autocorrelation, global systemic physiology, and motion artifacts. Specifically, the left panels of Figure 3 show the time traces of two channels for the three conditions after converting the raw simulated fNIRS signals to HbO signal changes using MBLL. The right panels in Figure 3, show the correlation structure of the channels. As expected, there is increased covariance in the channels with the introduction of global systemic physiology. The covariance further increases drastically with correlation values between channels close to 1, with the introduction of the motion artifacts in the simulated data.

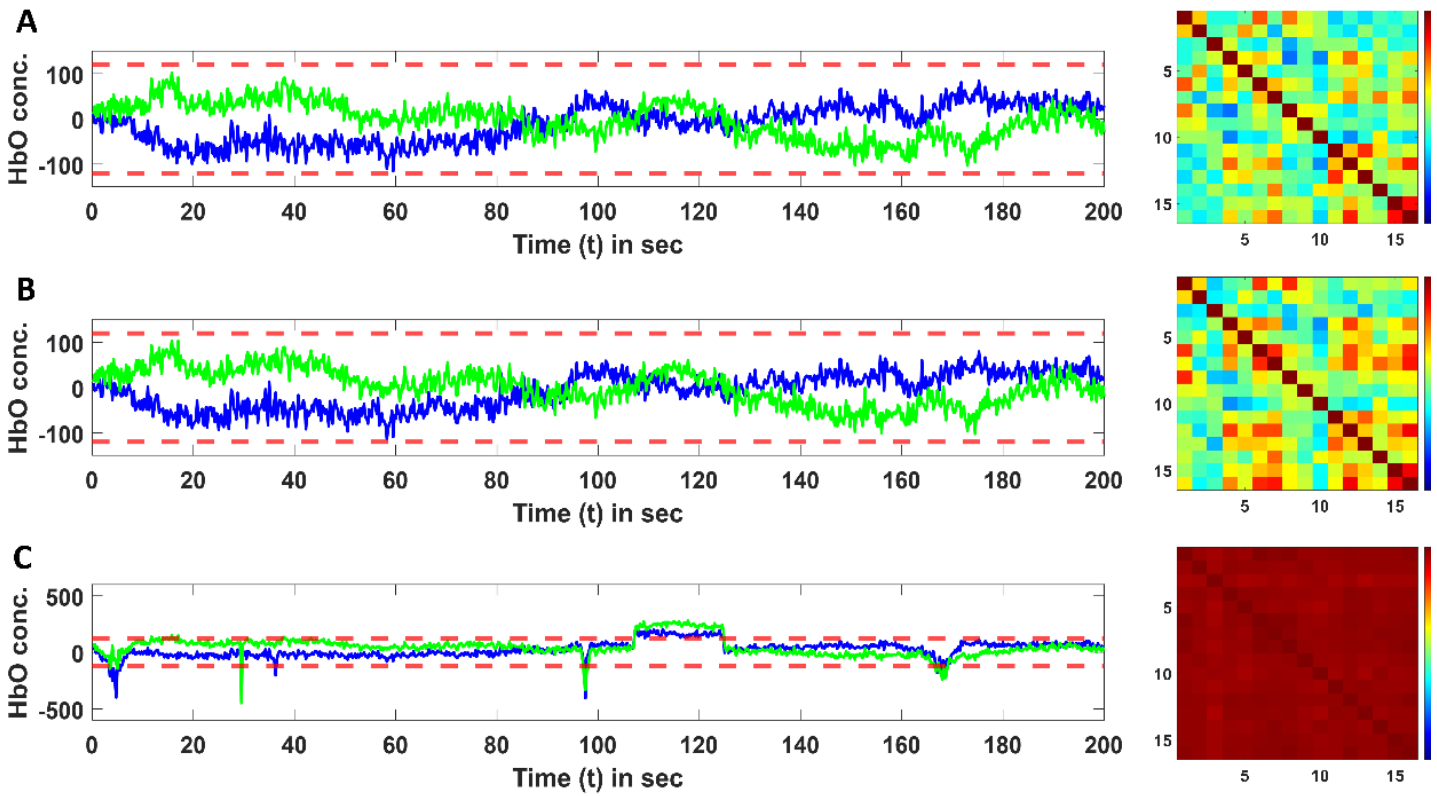

Figure 3. Plots showing the time traces of HbO changes for two representative channels in the simulated data as well as the normalized covariance matrices for the channels. Rows show simulated data with just (A) temporal autocorrelation, (B) with both temporal autocorrelation and global systemic physiology, and (C) with temporal autocorrelation, global systemic physiology, and motion artifacts. The panels on the left show the time series for two representative channels for 200 sec and the panels on the right show the normalized covariance matrix of the HbO signal changes for all the 16 long-distance channels in the data

Pre-whitening

Pre-whitening often uses AR models to remove the autocorrelation in the time series and whitens the frequency content of the signal. Several other articles have explored pre-whitening and its effects in detail (Barker et al., 2013; Santosa et al., 2017). Pre-whitening greatly reduces the false discovery rate in the connectivity models. Here, pre-whitening was implemented using an AR model as follows:
fNIRS time course can be expressed as

$$y_t = \sum_{i=1}^{p} a_i \cdot y_{t-i} + \varepsilon_t \ \ \forall t$$

and $\varepsilon_t \ \in N(0, \ \sigma^2)$

Here y_t is the fNIRS time series, a_i are the autoregressive parameters, p is the model order, which is determined using Bayesian information criterion (BIC), and ε_t is the innovation term. The innovation term is used to calculate the connectivity metrics rather than the time series themselves. Figure 4 shows the (A) power spectral density, and (B) autocorrelation function plots for a representative channel before and after pre-whitening for three cases of simulated data including, (i) with temporal autocorrelation, (ii) with temporal autocorrelation and global systemic physiology, and (iii) with temporal autocorrelation, global systemic physiology and motion artifacts. As expected, the pre-whitening step flattens the power spectral density plot and reduces the temporal autocorrelation in the fNIRS time series.

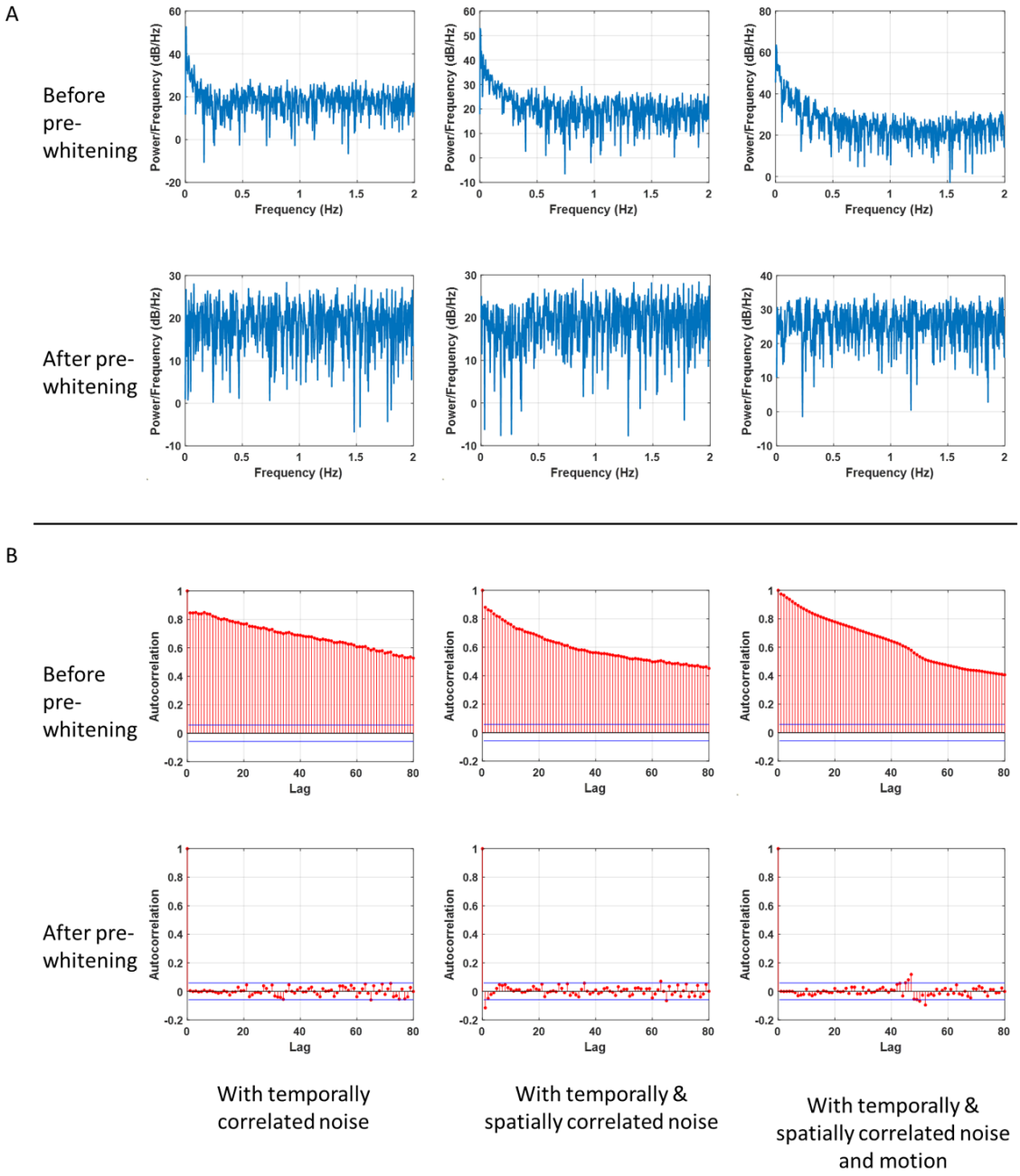

Figure 4. Plots showing the (A) power spectral density, and (B) autocorrelation function plots for a representative channel before and after pre-whitening for three cases of simulated data in each column including, (i) with temporal autocorrelation, (ii) with temporal autocorrelation and global systemic physiology, and (iii) with temporal autocorrelation, global systemic physiology and motion artifacts.

Robust methods

After pre-whitening the time series, the innovation terms contain information added at that time point. The large signal changes due to motion are more prominent as

they came from a different data generating process (motion) relative to the neural data generating process. Because motion artifacts are considered outliers, they can affect assumptions of normality. Often these influential points can have a significant impact on the estimates of connectivity metrics. As mentioned earlier, motion artifacts affect connectivity metrics differently depending on whether the motion is shared across many channels or if it is limited to just a few channels. Motion corrupted time points that are shared across multiple channels can impact the functional connectivity by increasing the covariance between them. Thus, to identify statistical outliers we used a multivariate version of a joint weighting function that was previously introduced in Santosa et al., (2017). The joint weighting function can identify the outliers by computing the geometric length of the time series and downweighting them appropriately depending on how they deviate from the mean, with motion corrupted outliers being assigned lower weights, which results in them having a lower impact on the correlation and regression methods.

A geometric length function r_t is computed from the innovation terms of all the channels (including short-separation channels if available) as follows.

$$r_t = \sqrt{\varepsilon_{1,t}^2 + \varepsilon_{2,t}^2 + \varepsilon_{3,t}^2 \ldots\ldots \varepsilon_{c,t}^2}$$

Where $\varepsilon_{i,t}$ indicates the innovation terms of the i^{th} channel for time point t and c is the total number of channels.

A weighting function (*S*) that downweights outliers depending on how much they deviate from the mean, can be calculated as follows for all time points using the square root of Tukey's bisquare function.

$$S\left(\frac{r}{\sigma}\right) = \begin{cases} 1 - \left(\frac{r}{\sigma \cdot \kappa}\right)^2 & for \quad \left|\frac{r}{\sigma}\right| < \kappa \\ 0 & for \quad \left|\frac{r}{\sigma}\right| > \kappa \end{cases}$$

Here σ is the standard deviation of the distribution and is estimated from the relationship between median absolute deviation (MAD) and the standard deviation of a normal distribution ($\sigma \approx 1.4826*MAD$). MAD is used because it is a more robust dispersion statistic (Hoaglin et al., 1983). The constant κ is set to a value of 4.685 which provides 95% efficiency for a standard normal distribution without any outliers. A weighting matrix $\boldsymbol{S}$ with diagonal elements representing the weighting function for all time points can be multiplied with the innovation terms for all the time series to downweight the statistical outlier points from the normal distribution

$$\boldsymbol{Y}_c = \boldsymbol{S} \cdot \boldsymbol{E}_c$$

Here $\boldsymbol{Y}_c$ is the reweighted innovation terms matrix for all channels, and $\boldsymbol{E}_c$ is the innovation terms obtained after pre-whitening.

So, to reduce the impact of motion artifacts, for regression models, instead of using a linear model and estimating the parameters with ordinary least squares (OLS) estimates, we use robust statistics that use weighted least squares (WLS) to downweight the outliers calculated from the studentized residuals. The weights are calculated using Tukey's bisquare function using the process detailed earlier. The weights calculated are then applied to the original data, and the residuals are re-estimated. The process is repeated till convergence. Robust regression is implemented in the in-built MATLAB function 'robustfit'. The MATLAB function 'robustfit' uses a convergence criterion of

50 iterations or no detectable numerical change in the values of the regression coefficients, whichever is earlier, to assess the convergence of the coefficients. Since MVGC and modified MVGC did not converge in 50 iterations, we changed the maximum iteration limit to 150 iterations, to ensure that the regression parameters converged. The parameters estimated using the WLS regression can be obtained by

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T S\, X})^{-1}\boldsymbol{X^T S y}$$

Here $\widehat{\boldsymbol{\beta}}$ is the parameter estimates and $\boldsymbol{S}$ is the weighting matrix.
More details about the use of robust statistics in fNIRS data processing can be found in other articles (Barker et al., 2013; Huppert, 2016; Santosa et al., 2017).

Principal components analysis-based regression for reducing global signal changes

If the probes are widely distributed across the brain, then the large signal variance shared across multiple channels may mostly reflect physiological noise (Santosa et al., 2020). Hence, the first few components of principal components analysis (PCA) can be used to identify the global signal changes due to physiological noise and regress out its effects from the channel data. The data covariance matrix can be decomposed using singular value decomposition using

$$\boldsymbol{C} = \boldsymbol{U} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V^T}$$

where $\boldsymbol{C}$ is the data covariance matrix given $\boldsymbol{C} = \boldsymbol{Y^T} \cdot \boldsymbol{Y}$, $\boldsymbol{U}$ is a matrix containing left singular vectors. $\boldsymbol{\Sigma}$ is a diagonal matrix containing singular values and $\boldsymbol{V}$ is the matrix containing right singular vectors. Usually, the first few components explaining the most variance are used as a measure of global systemic physiology. We used the minimum number of principal components that explained at least 90% of the channel covariance. The global signal can be extracted from the decomposition of covariance of just the long-distance channels, just the short-separation channels, or both. All three options were examined in our testing. The channel data after PCA regression be obtained by

$$\boldsymbol{y_{PCA\,filtered}} = [\boldsymbol{I} - \boldsymbol{Z} \cdot (\boldsymbol{Z^T} \cdot \boldsymbol{Z})^{-1} \cdot \boldsymbol{Z^T}] \cdot \boldsymbol{y}$$

Here $\boldsymbol{y}$ is either the innovation term or the raw data for the channel depending on the analysis pipeline. $\boldsymbol{Z}$ is the matrix containing the first few components as columns.

Functional connectivity metrics

In this work, we examined several ways to estimate functional connectivity. Here we focus only on time-domain methods including correlation, partial correlation, and several variations of Granger causality. Time-frequency methods like wavelet coherence were not examined in this work for two reasons. First, in our previous work (Santosa et al., 2017), we found that in bivariate connectivity analysis, wavelet coherence outperformed standard Pearson correlation in ROC analysis but did not perform as well as our pre-whitened, robust correlation models. Wavelet coherence also had high false positive rates and uncontrolled type-I errors unless the standard wavelet coherence model (Cui et al., 2012; N. Liu et al., 2016) was modified to include higher-order autoregressive terms and robust (outlier rejection) methods. With these modifications, we found no benefit to wavelet coherence over our pre-whitened robust models. Secondly, to our knowledge, a multivariate and/or partialed version of wavelet coherence has not been described and would require the development of a new statistical model.

Correlation-based analysis models

Pearson's correlation coefficient: Pearson's correlation coefficient between two time series $\boldsymbol{x}$ and $\boldsymbol{y}$ are calculated as

$$r = \frac{\sigma_{x,y}}{\sigma_x \, \sigma_y}$$

where $\sigma_{x,y}$ is the covariance of time series $\boldsymbol{x}$ and $\boldsymbol{y}$ and σ_x and σ_y are the standard deviations of $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively. Standard Pearson's correlation is not a robust statistical estimator and several robust variations have been proposed (Shevlyakov & Smirnov, 2011), and described in the context of fNIRS in Santosa et al., (2017). Robust correlation is calculated as the geometric mean of the standardized regression coefficients on the regression models $\boldsymbol{y}$ on $\boldsymbol{x}$ and $\boldsymbol{x}$ on $\boldsymbol{y}$, to reduce outliers in either $\boldsymbol{x}$ or $\boldsymbol{y}$, which can still influence the correlation estimates. The procedure is described below

$$\boldsymbol{y} = [\mathbf{1} \quad \boldsymbol{x}] \begin{bmatrix} \beta_{0,1} \\ \beta_{x \to y} \end{bmatrix}$$

A weighting matrix is estimated based on the studentized residuals in the above regression model that downweights the outliers in $\boldsymbol{y}$. The weighting matrix $\boldsymbol{S_{x \to y}}$ is then multiplied on both sides of the regression model above and the estimates of the parameters are updated.

$$\boldsymbol{S_{x \to y}}.\boldsymbol{y} = \boldsymbol{S_{x \to y}} [\mathbf{1} \quad \boldsymbol{x}] \begin{bmatrix} \beta_{0,1} \\ \beta_{x \to y} \end{bmatrix}$$

Similarly, we can fit $\boldsymbol{x}$ on $\boldsymbol{y}$,

$$\boldsymbol{x} = [\mathbf{1} \quad \boldsymbol{y}] \begin{bmatrix} \beta_{0,2} \\ \beta_{y \to x} \end{bmatrix}$$

A weighting matrix $\boldsymbol{S_{y \to x}}$ is then estimated based on the studentized residuals to downweight the outliers in $\boldsymbol{x}$. $\boldsymbol{S_{y \to x}}$ is multiplied on both sides of the regression model above and the estimates of the parameters are updated.

$$\boldsymbol{S_{y \to x}}.\boldsymbol{x} = \boldsymbol{S_{y \to x}} [\mathbf{1} \quad \boldsymbol{y}] \begin{bmatrix} \beta_{0,2} \\ \beta_{y \to x} \end{bmatrix}$$

The two regression models are solved alternatively until the parameters $\beta_{y \to x}$ and $\beta_{x \to y}$ converge. The final robust correlation coefficient is estimated using the formula

$$\|r_{robust}\| = \sqrt{\beta_{x \to y} \, . \beta_{y \to x}}$$

The sign of the robust correlation coefficient estimate is determined by the signs of the regression coefficients. The effective degrees of freedom given by $\sum \left(min \left(\boldsymbol{S_{y \to x}}, \boldsymbol{S_{y \to x}}, \boldsymbol{S_{y,x}} \right) \right) - 2$. The effective degrees of freedom are reduced due to the downweighting of the data points especially the outliers.

For both the standard and robust correlation estimates, the probability of observing the data given the null hypothesis (*p-value*) can be estimated from the T-statistic using

$$t = \frac{r \sqrt{df}}{\sqrt{1 - r^2}}$$

with df is the effective degrees-of-freedom with a value of $n-2$ for standard Pearson's correlation, and r is the Pearson's correlation coefficient. Here n denotes the number of time points.

Autoregressive (AR) correlation: The autoregressive correlation was Pearson's correlation coefficient calculated on the pre-whitened time series. As described in Santosa et al., a p^{th} order autoregressive model is estimated for each channel of data and used to pre-whiten the signal (Santosa et al., 2017). The innovations term in the autoregressive model (e.g. the new information added at each time step) is then used in the correlation analysis. We used a maximum model order of 40 (ten times the sample rate) which is higher than the AR filter used in simulating the fNIRS data. Based on empirical observations, the model order is typically less than $10 \cdot F_s$ (here F_s is the sampling frequency) in experimental RS-fNIRS data.

AR partial correlation: Partial correlation involves a two-step process in which the other channels are first projected out of the data of interest and then the correlation is performed on the residuals. We used PCA regression-based projection to partial out the effect of other channels as described in section 'Principal components analysis-based regression for reducing global signal changes.' PCA regression is used to avoid collinearity instabilities of the model. The minimum number of eigenvectors to explain at least 90% of the spatial covariance were used. We examined the case where i) only the other long-distance channels were used in the PCA, ii) only the short-separation channels were used, and iii) both long-distance and short-separation channels were used. For two channels i and j, if the innovation terms after pre-whitening are denoted by $\boldsymbol{\varepsilon_i}$, $\boldsymbol{\varepsilon_j}$, the whitened PCA filtered time series are calculated as:

$$\boldsymbol{\varepsilon_{i,PCA\,filtered}} = [\boldsymbol{I} - \boldsymbol{Z} \cdot (\boldsymbol{Z^T} \cdot \boldsymbol{Z})^{-1} \cdot \boldsymbol{Z^T}] \cdot \boldsymbol{\varepsilon_i}$$

$$\boldsymbol{\varepsilon_{j,PCA\,filtered}} = [\boldsymbol{I} - \boldsymbol{Z} \cdot (\boldsymbol{Z^T} \cdot \boldsymbol{Z})^{-1} \cdot \boldsymbol{Z^T}] \cdot \boldsymbol{\varepsilon_j}$$

Now, the AR partial correlation can be calculated as the Pearson's correlation between the two residuals as the global signal is removed from the channels. The degrees of freedom are further reduced based on the number of principal components retained. The input data to this model was the pre-whitened innovations time series rather than channel data, hence this method is termed AR partial correlation.

Granger causality-based analysis models

Multivariate Granger causality: Multivariate Granger causality (MVGC) tests for lagged relationships between two time series after controlling for the effects of other time series. This is essentially a model-fit test of two regression models. We can test whether $\boldsymbol{x}$ "Granger causes" $\boldsymbol{y}$ after controlling for the channels $\boldsymbol{Z}$ using multivariate autoregressive models (MVAR) as detailed below. For a channel time series, we can model the current time point y_t as a function of previous lags of $\boldsymbol{y}$ and other channels $\boldsymbol{Z}$ in the restricted model.

$$y_t = \sum_{i=1}^{p} \alpha_i \, y_{t-i} + \sum_{j=1}^{nPC} \sum_{i=1}^{p} \gamma_i \, z_{j,t-i} + c_1 + \varepsilon_t$$

(Restricted model)

In the unrestricted model, we add the lag terms of $\boldsymbol{x}$, to model the current time point in $\boldsymbol{y}$

$$y_t = \sum_{i=1}^{p} \alpha_i \, y_{t-i} + \sum_{i=1}^{p} \beta_i \, x_{t-i} + \sum_{j=1}^{nPC} \sum_{i=1}^{p} \gamma_i \, z_{j,t-i} + c_2 + \varepsilon_t$$

(Unrestricted model)

Here p is the number of lag terms included in the model, it is determined by using fit measures using Bayesian information criteria (BIC). To test for causality, we statistically test if adding the additional lag terms in $\boldsymbol{x}$ (unrestricted model) improves the predictability of y_t compared to the model without $\boldsymbol{x}$ (restricted model). If the past of the $\boldsymbol{x}$ contains predictive information about the current time point y_t, then the error in the prediction of $y_t$is improved even after controlling for the lost degrees of freedom. This can be assessed statistically as

Null Hypothesis: $\boldsymbol{x}$ does not Granger cause $\boldsymbol{y}$

H_0: $\beta_1 = \beta_2 = \beta_3 = \ ... \ \beta_p = 0$

Alternative Hypothesis: $\boldsymbol{x}$ does Granger cause $\boldsymbol{y}$

H_A: At least one of β_i is non-zero

Since we have nested models, we can use a nested *F*-statistic to test for Granger causality

$$F = \frac{SSE_R - SSE_U / p}{SSE_U / [n - (2p + 1)]}$$

Here, SSE_R and SSE_U are the sums of squared errors of the restricted and unrestricted models respectively and *n* is the number of time points. So, we can calculate a *p-value* for the *F*-statistic at degrees-of-freedom of p and $[n - (2p + 1)]$. A Granger causality metric G, can also be calculated to assess the strength of the lagged relationship as follows

$$G = \log SSE_R / SSE_U$$

The purpose of using multivariate Granger causality over bivariate Granger causality is to control for the effects of third variables including systemic physiology. The addition of lag terms of several channels in the MVGC model can easily make the regression model unsolvable as the number of variables would exceed the number of time points. So rather than using the channel data itself, we first perform PCA and included the fewest components that explained at least 90% of the variance as we did with partial correlation earlier. We then include components and the time-lagged components rather than the channel data in the restricted and unrestricted models.

Modified multivariate Granger causality: Traditional MVGC includes only the time-lagged history of $\boldsymbol{y}$, $\boldsymbol{x}$, and $\boldsymbol{Z}$ on the right-hand side of the equations. Thus, MVGC only models non-zero lagged relationships (hence the "causality" term). In comparison, the correlation methods described in earlier sections only model zeroth-lag correlation. It is important to recognize that "zeroth-lag" is relative to the sample rate and that at a 4 Hz sampling of the simulated fNIRS data, any synchronous signal changes occurring within

one sample (250ms) would be mathematically zeroth-lag. While neural signaling in the brain is expected to be fairly fast (100s of milliseconds), we did not want to assume that all correlations would be zeroth-lag and, in this paper, we examined multiple scenarios. Here, we introduce a modified MVGC model to model both lagged and zero-lagged relationships between channels. In the modified MVGC we include the zero-lag terms of $\boldsymbol{Z}$ (top principal components) in both the restricted and unrestricted model as well as the zero-lag terms of channels $\boldsymbol{x}$ in the unrestricted model. So, the restricted model can be written as

$$y_t = \sum_{i=1}^{p} \alpha_i \, y_{t-i} + \sum_{j=1}^{nPC} \sum_{i=0}^{p} \gamma_i \, z_{j,t-i} + c_1 + \varepsilon_t$$

(Restricted model)

Similarly, the unrestricted model can be written as

$$y_t = \sum_{i=1}^{p} \alpha_i \, y_{t-i} + \sum_{i=0}^{p} \beta_i \, x_{t-i} + \sum_{j=1}^{nPC} \sum_{i=0}^{p} \gamma_i \, z_{j,t-i} + c_2 + \varepsilon_t$$

(Unrestricted model)

The above equations also include the zero lagged terms of $\boldsymbol{Z}$ and $\boldsymbol{x}$. Now to test for statistical significance of Granger causality, we test a null hypothesis and alternative hypothesis as follows:

H_0: $\beta_0 \; = \; \beta_1 = \beta_2 = \beta_3 = \; \ldots \beta_p = 0$

H_A: At least one of β_i is non-zero

The statistical significance is assessed with the nested *F*-statistic and the Granger causality is calculated similarly to traditional multivariate Granger causality. One caveat of this modified model is that the hypothesis does not test whether the connection between $\boldsymbol{y}$ and $\boldsymbol{x}$ was due to the zeroth or non-zeroth lag terms. One could examine this further by modifying the unrestricted model to also contain the non-zero lagged terms of $\boldsymbol{x}$, such that the only difference between the unrestricted and restricted models is the inclusion of the zeroth-lag term in $\boldsymbol{x}$.

Robust multivariate Granger causality: Finally, we developed a robust statistical version of both the MVGC and modified MVGC models. The definition of outliers in this multivariate model is a bit challenging since, like the bivariate correlation model, outliers can exist in the time courses of $\boldsymbol{y}$, $\boldsymbol{x}$, and/or $\boldsymbol{Z}$. In bivariate correlation, we used the union of outlier weights (e.g. an outlier in $\boldsymbol{y}$ or $\boldsymbol{x}$). However, we found that the extension to many channels of data often resulted in too much of the data being downweighted since an outlier point in any one of the channels caused that time point to be downweighted for all the channels. In particular, the time lags in the MVGC required all lagged columns of the matrix containing that entry to be downweighted as well.

To reduce the complexity of robust regression methods described earlier with multiple terms (channels and their lag terms) in the regression model, rather than using channel data by itself, we use the innovation terms. The procedure for robust multivariate Granger causality is described below. First, the channel data (both short-separation and/or long-distance) is pre-whitened and the innovation terms are calculated. The data are pre-weighted with a weighting matrix calculated from the innovation terms in a procedure described in the section on ‘Robust methods.’ After pre-whitening and pre-weighting, the

innovation terms of the other channels ($\boldsymbol{Z}$) is included in the regression for the restricted model as described below.

$$\varepsilon_t^{\boldsymbol{y}} = \sum_{j=1}^{nPC} \sum_{i=0}^{p} \gamma_i \varepsilon_{j,t-i}^{\boldsymbol{z}} + c_1 + \varepsilon_t$$

(Restricted model)

Here $\varepsilon_t^{\boldsymbol{y}}$ is the innovation term of channel $\boldsymbol{y}$ at time point t after pre-whitening and pre-weighting. $\varepsilon_{j,t-i}^{\boldsymbol{z}}$ are the principal component weights for the j^{th} component at time point $t-i$. nPC denotes the number of principal components used.

Similarly, the innovation terms are also used in the unrestricted model rather than the channel data itself

$$\varepsilon_t^{\boldsymbol{y}} = \sum_{i=1}^{p} \beta_i \, \varepsilon_{t-i}^{\boldsymbol{x}} + \sum_{j=1}^{nPC} \sum_{i=0}^{p} \gamma_i \, \varepsilon_{j,t-i}^{\boldsymbol{z}} + c_2 + \varepsilon_t$$

(Unrestricted model)

Compared to the restricted model, the unrestricted model additionally includes the innovation terms and lags of channel $\boldsymbol{x}$. $\varepsilon_{t-i}^{\boldsymbol{x}}$ is the innovation term of channel $\boldsymbol{x}$ at time point $t-i$.

A weighted least square regression using an iterative robust estimator is used to fit both the restricted and the unrestricted models with the data being downweighted from the estimated studentized residuals. The process is repeated until convergence of the weight matrix and parameter estimates. As with other MVGC models, nested *F*-statistic is used to test for the statistical significance of the lagged relationship between the two time series.

Data analysis plan

To test the effectiveness of each processing method in reducing the effects of temporal autocorrelation, global shared physiology, and motion artifacts, all the connectivity metrics were compared. We expected that all methods which had the pre-whitening step would correct for temporal autocorrelation. To test the need for correcting for global signals due to shared physiological noise between the channels, we expected methods that controlled for the effects of other channels such as partial correlation and MVGC would perform better when the data had systemic physiology. We also expected the robust methods to perform better than the non-robust methods when the simulated data had added motion artifacts. Finally, we expected the zero-lag correlation measures to perform better when the shared simulated neural covariance was at zero-lag and the Granger causality methods to perform better when the information was present in the first lag. The modified Granger causality with zero-lag which models both zero-lag and lagged relationships is expected to perform better at both the zero$^{\text{th}}$-lag and the first lag. In Figure 5, we provide a summary of the different analysis pipelines examined.

We also compared the effectiveness of robust versions of (i) Pearson's correlation, (ii) AR correlation, (iii) AR partial correlation on the RS-fNIRS data from 24 subjects. After group-level analysis, with subject as a random factor, we expected AR partial correlation to show the sparsest network, followed by AR correlation. We expected the Pearson's correlation to show the densest network with almost every channel connected with the other. Since we do not know the ground truth nor do we have

a “gold standard” such as fMRI to compare our results to, we cannot draw broad conclusions about performance of these analyses methods.

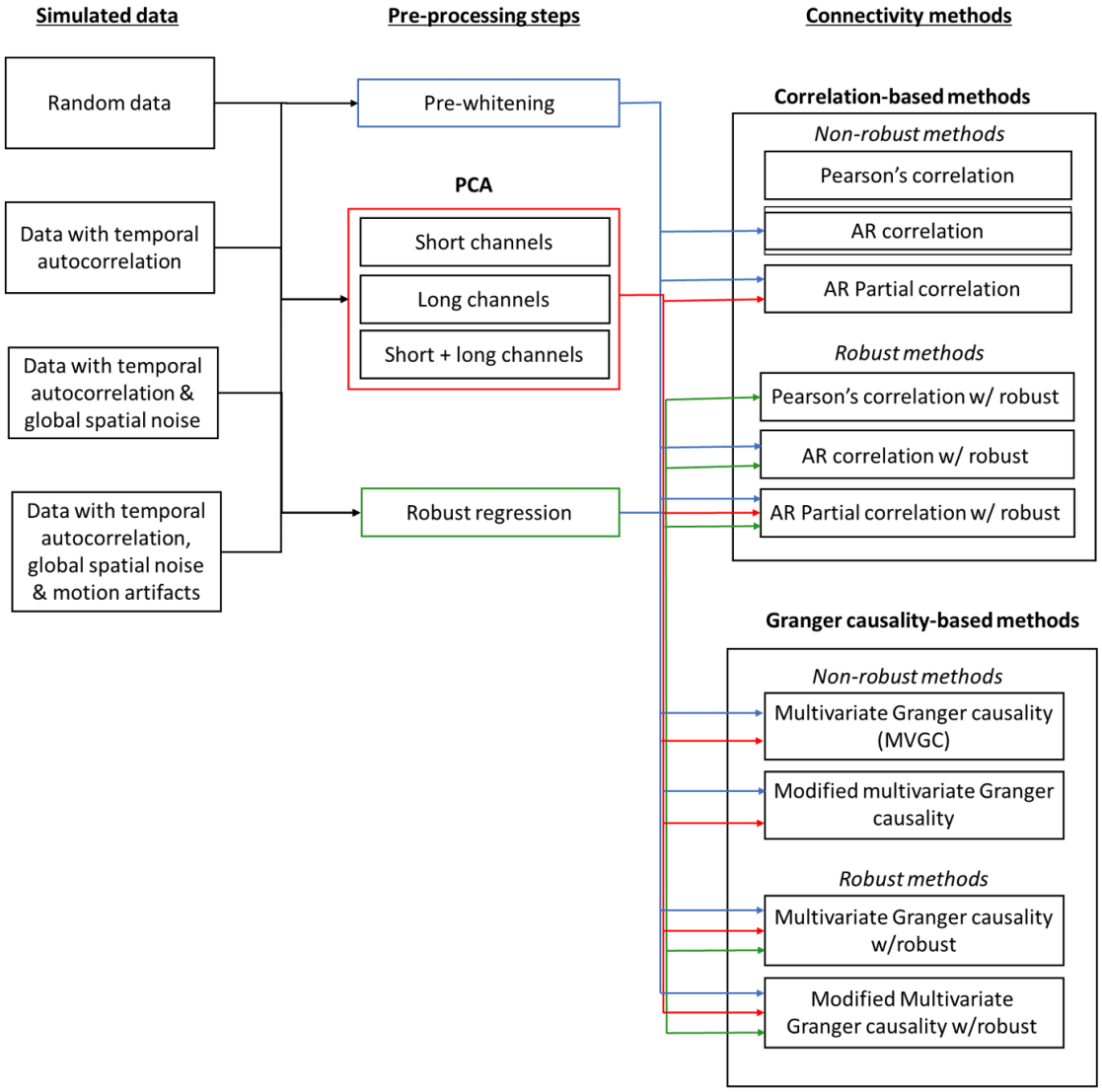

Figure 5. Figure showing different pipelines and connectivity methods that were applied to the four types of simulated data.

Performance metrics

We ran the simulation 200 times, each time generating a different dataset with different covariance structure to mimic shared neural activity. Then we calculated the *p-values* from the ‘correlation coefficient’ for correlation-based methods or an ‘*F*-statistic’ for Granger causality methods. The false positives, true positives, true negatives, and false negatives were tallied up to compare the effectiveness of each of the multiple data processing methods for quantifying connectivity. We used two plots to assess the

performance of the different methods (i) type-I error control plots and (ii) receiver operating characteristic (ROC) curves.

Type-I error control plots

To understand how well the data processing methods control for Type-I error, we plotted the actual false positive rate (from bootstrapping) against the theoretical *p-value* obtained from the test statistic for the method (denoted by p-hat). An ideal method would indicate that the *p-value* from the method (p-hat) would match the actual *p-value,* hence the plot would have a unit slope. Parts of the curve above the unit slope line, indicate the model is likely to have more false positives than expected at that significance level, and the parts under the line indicate that the model is likely to have more false negatives than expected at that significance level.

ROC curves

To investigate the sensitivity and specificity of the data processing methods and the model performance, ROC curves were generated from the known ground truth as some channels were simulated with shared "neural" covariance. The ROC curve plots the true positive rate (sensitivity) vs false positive rate (1-specificity) at different thresholds. Ideal performance of the processing method would result in very high true positives (TP) and very high true negatives (TN) approaching the top left corner of the plot. Similarly, a curve with slope one and an AUC of 0.5 would indicate random guessing. Since the number of true positives was less than the true negatives (only 10% of the non-diagonal elements in the covariance matrix used to generate the simulated data were non-zero) in the simulated data, a random subsample of true negatives equal in size to the true positives was used to generate the ROC curves.

Code implementation

All of the methods were implemented in MATLAB (Natick, Massachusetts). Inbuilt MATLAB functions and the functionality of the NIRS Brain AnalyzIR toolbox (Santosa et al., 2018) were used for preprocessing and implementation of the correlation methods. Other methods including partial correlation, Granger causality, and robust Granger causality were implemented using custom MATLAB scripts. The toolbox and the codes used for this article are available online at https://github.com/pradlanka/rsfc-fnirs.

Results

Comparison of processing methods

As expected, as the progression of simulations and challenges of the noise increased, the basic models such as standard Pearson's correlation quickly failed. For example, Pearson's correlation assumes independent measurements, which is violated by the addition of temporally autocorrelated noise. As shown in Figure 6B, the Pearson's and robust Pearson's correlation models which did not include an autoregressive (pre-whitening) filter had considerably lower performance on ROC curves compared to the pre-whitened and MVGC models. Almost all other methods have an area-under-the-curve (AUC) close to 1. In addition, as shown in Figure 6A, the Pearson's and robust-Pearson's models had substantial false positive rates as indicated by lines above the diagonal in the plot of actual false positive rate versus expected rate (p-hat). For example, at an expected p-hat of 0.05, the actual false positive rate was over 80%. The other analysis models had

type-I error plots much closer to ideal (diagonal) where the expected rate equals the realized rate. The AR-correlation and robust AR-correlation both had the closest to ideal type-I error control. The AR partial correlation model had a slight overestimation of the significance of the test statistic, whereas the robust AR partial correlation overcorrects leading to underestimation of significance of the test statistic (less than expected false positive rate). Both versions of the modified MVGC tended to underestimate the *p-values* slightly, thus leading to slightly higher than expected false-positive rate. Here, we are not showing the results for the standard MVGC model, which does not include a specific zeroth-lag term. Because the data were simulated at the zeroth-lag, which the MVGC does not model, the MVGC shows no significant channels. These results support the necessity of correcting for the autocorrelation in the time series as highlighted in a previous article (Santosa et al., 2017).

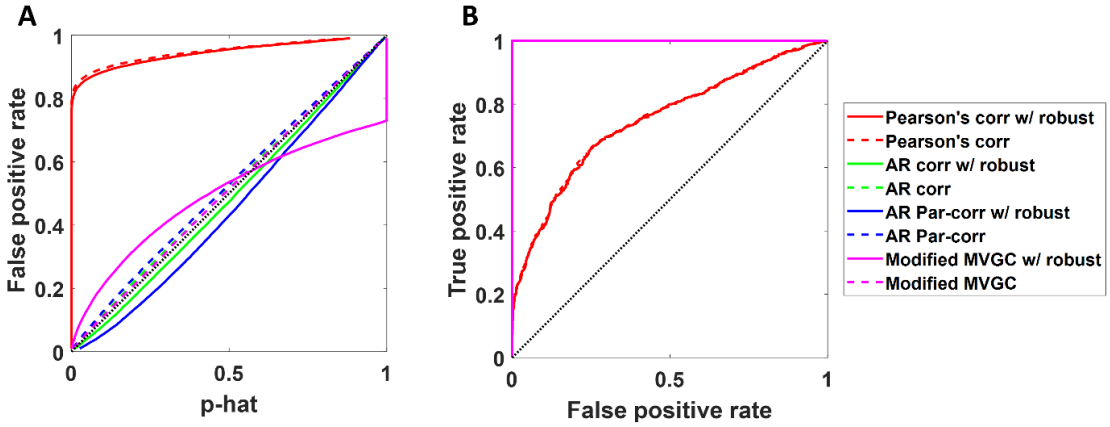

Figure 6. The (A) Type-I error control plots and (B) ROC curves for all the methods with simulated data containing temporal autocorrelation. The dotted lines indicate the non-robust version of the methods, while the solid lines represent their robust counterparts.

The second set of simulations (Figure 7) was generated using both temporal autocorrelation and spatially global systemic physiology. As expected, AR correlation joins the Pearson's correlation coefficient in increased false positives (Figure 7A) and lower AUC in the ROC curves (Figure 7B). While the use of pairwise correlation worked when the noise was independent between channels, spatial noise introduced global false positive connections and large over-reporting of the significance of the model. AR Partial correlation and the modified (zero-lagged) Granger causality still performed well. Again, all the analysis methods except pairwise Pearson's correlation had AUC of near 1 on the ROC curves, which indicates that the true positive channels were still well separated from the true negatives, but that the reporting of the *p-values* (p-hat) was very wrong. In other words, one can find the true connections by setting the right threshold on the correlation coefficient, but that one cannot trust the reported *p-value* for setting that threshold. In the results for both the simulations shown in Figure 6 and Figure 7, there were no motion artifacts added to the simulated data and thus the robust and non-robust methods performed similarly well with their type-I error control curves and the ROC curves. Though the use of robust methods can lead to a loss in degrees of freedom as the weights

of data points can be less than 1, the number of time points is still much larger (>1000), hence the loss in degrees of freedom is almost negligible.

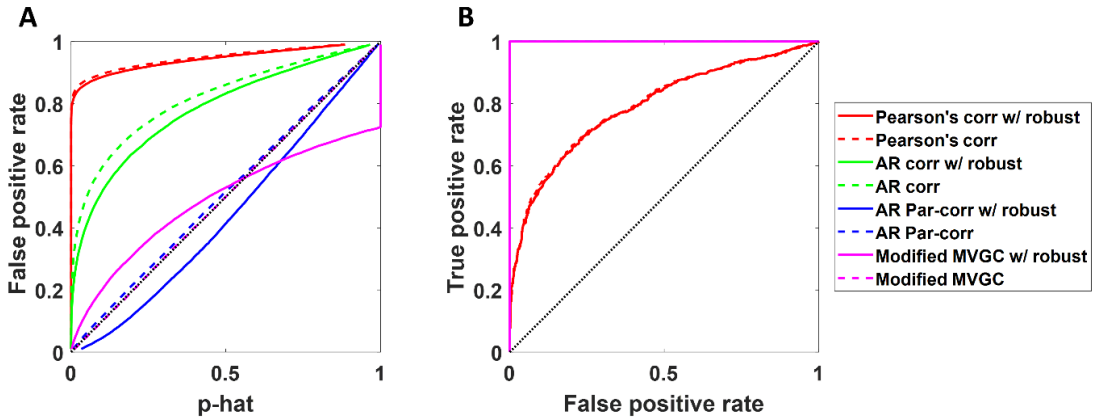

Figure 7. The (A) Type-I error control plots and (B) ROC curves for all the methods with simulated data containing temporal autocorrelation and shared global signal mimicking systemic physiology. The dotted lines indicate the non-robust version of the methods, while the solid lines are their robust counterparts.

For, the third simulation, head motion modeled as spike and shift artifacts was added to the simulated data containing both the temporal autocorrelation and global signal. As shown in Figure 8, the robust methods which downweights outlier points, performed much better than their non-robust counterparts, with robust modified Granger causality and AR partial correlation with short-separation channels having similar AUC, but AR partial correlation had fewer false positives than modified multivariate Granger causality. It is important to note that while robust modified Granger causality was effective at reducing the false positives it still had much higher-than-expected false positives (false positive rate of 0.25 at an expected rate of 0.05).

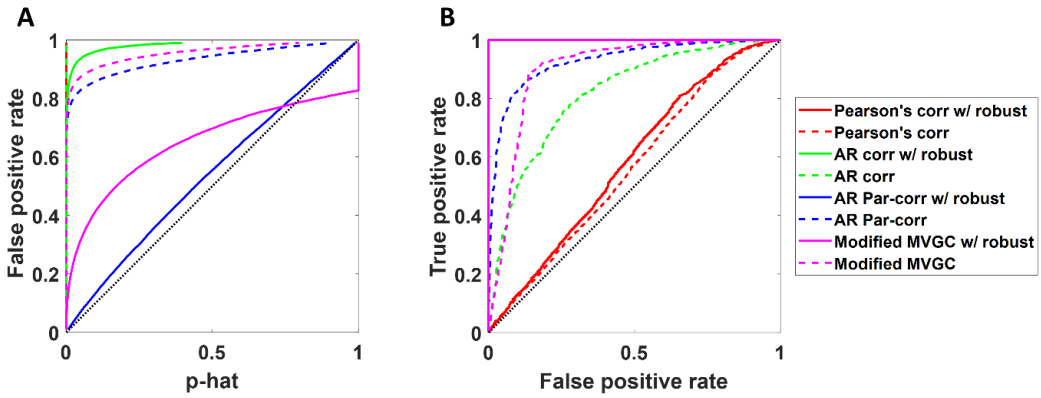

Figure 8. The (A) Type-I error control plots and (B) ROC curves for all the methods with simulated data containing temporal autocorrelation, shared global signal mimicking systemic physiology and head motion artifacts. The dotted lines indicate the non-robust version of the methods, while the solid lines are their robust counterparts.

Comparison of the effectiveness of short-separation channels

In the previous section, we showed AR partial correlation and modified MVGC models that included both the short-separation and other long-distance fNIRS channels in the model. Our results indicated that using these components was important to remove global noise signals. In this section, we explore these models in more depth and compare the use of just short-separation channels, the other long-distance channels, or both short-separation channels and other long-distance channels in the performance of the whitened partial correlation method. In Figure 9A, we show these comparisons for data simulated with both temporal and spatial correlated noise. In the type-I error control plot, the use of only long-distance channels was the worst and likely led to the overestimation of significance of the test statistic (correlation coefficient) indicating that the results still retained too many global connections. The use of only short-separation channels produced the best result, although it slightly under-estimated the significance of the robust estimator and result in a slight introduction of false-negatives in the estimates. Finally, the use of both the short-separation and long-distance channels was slightly worse than the use of the short-separation channels alone, but better than the use of only the long-distance channels. This is probably due to some overfitting in the model.

In Figure 9B, we show the comparisons for data simulated with additional motion artifacts. As expected, the robust statistical estimators greatly outperform the non-robust ones, with the short-separation channels-only model being the most sensitive to these artifacts. Consistent with the motionless simulations, the robust short-separation channels-only and robust version of both short-separation & long-distance gave the best results. The use of only long-distance channels in the robust model still had a substantial over-reporting of the significance of test statistics.

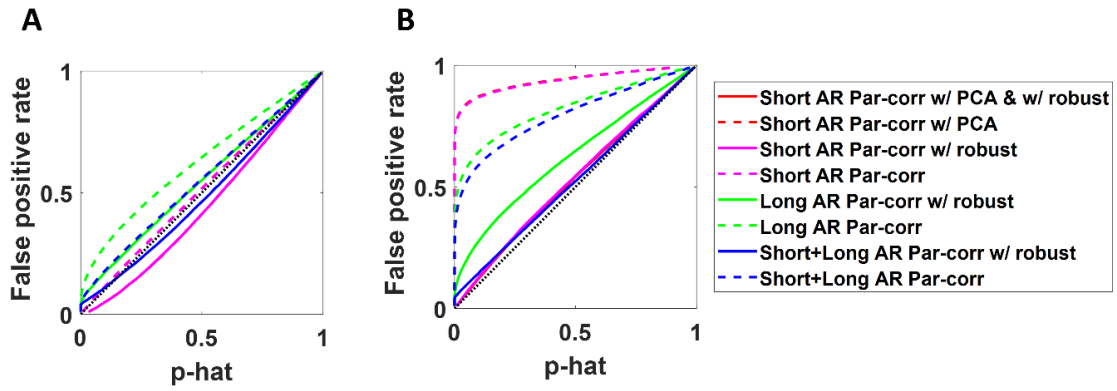

Figure 9. Plots showing Type-I error control plots comparing whitened partial correlation methods using short-separation, long-distance channels or both on simulated data containing (A) temporal autocorrelation and shared global signal mimicking systemic physiology, and (B) temporal autocorrelation, systemic physiology and added head motion. Results are shown for short-separation channels both with PCA and without PCA performed on them before partialing out its effects. The dotted lines indicate non-robust versions of the methods, while the solid lines indicate their robust counterparts.

Finally, in Figure 10, we examined the same analysis using the modified MVGC model in place of the AR partial correlation. Figure 10A shows the results for motion-less simulated data. We found that the MVGC was less sensitive to which channels were used in the model and the short-separation, long-distance, and both models all had similar performance. The robust version of the MVGC was observed to have a substantial over-reporting of the significance of the test statistics even with the motionless data simulations. We believe that this is an uncorrected error in the model concerning the effective degrees of freedom of the estimates. In particular, the definition of statistical outliers in the robust estimator is difficult because the time points in the measurement variables $\boldsymbol{y}$, $\boldsymbol{x}$, and/or $\boldsymbol{Z}$ can all be outliers due to motion. The Tukey's bisquare weight being used in this model assumes the outliers are independent, but our current model does not effectively account for shared outlier time points across multiple channels and thus the degrees of freedom are slightly overestimated. In Figure 10B, the MVGC models are compared for the data simulations with added motion artifacts. Here we found that robust methods out-perform their non-robust versions, but we again found over-reporting in the statistical estimates.

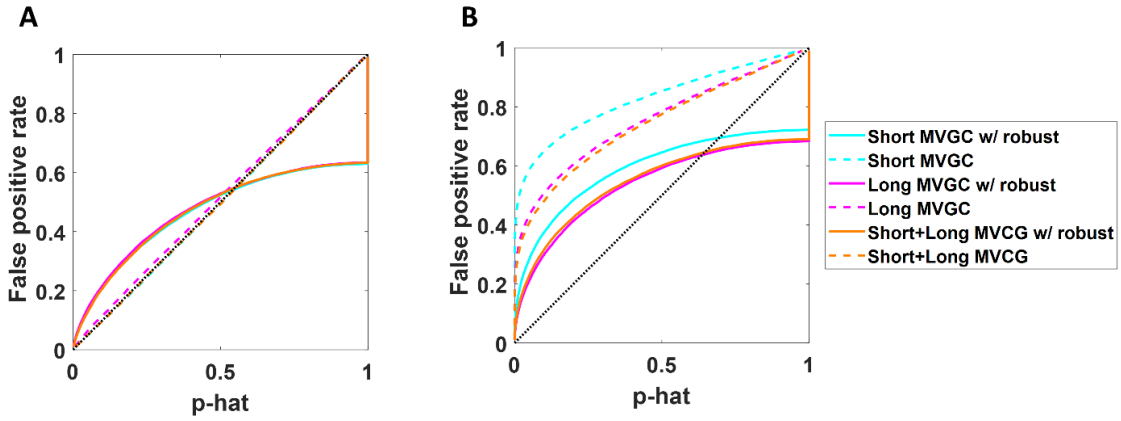

Figure 10. Plots showing Type-I error control plots comparing multivariate Granger causality methods using short-separation, long-distance channels or both on simulated data containing (A) temporal autocorrelation and shared global signal mimicking systemic physiology, and (B) temporal autocorrelation, systemic physiology and added head motion. The dotted lines indicate the non-robust version of the methods, while the solid lines are their robust counterparts.

Modeling lagged connectivity

So far, we have shown the results for the simulated data when the relationships between channels were present at the zero lag. The Pearson's and partial correlation models all assumed zeroth-lag connections whereas the standard MVGC assumed only non-zero lag (causality). The modified MVGC considered both zero and non-zero connections. Thus, when the data were simulated with only zeroth-lag connections, the AR partial correlation was determined to be the best. Standard MVGC (with no zeroth-lag) was not even presented in the previous sections since it failed to find anything in the zeroth-lag simulations (Figure 11A). Likewise, when data were only simulated with

relationships at non-zero lag terms, the AR partial correlation models failed, as shown in Figure 11B. The modified MVGC was the only model that worked in both cases.

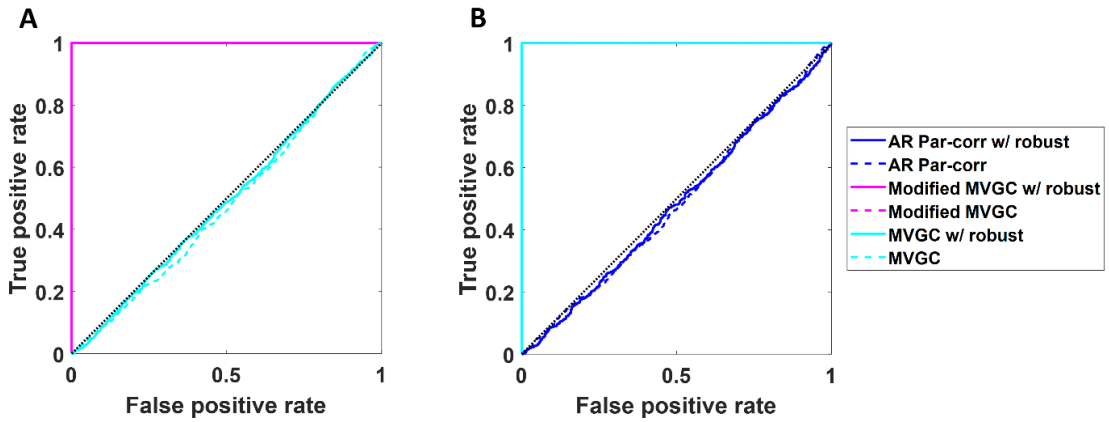

Figure 11. ROC curves comparing whitened partial correlation, multivariate Granger causality, modified multivariate Granger causality with and without robust regression when the linear relationships between channels are simulated at (A) zero-lag, (B) first lag. The simulated data contains temporal autocorrelation and shared global signal mimicking systemic physiology. The dotted lines indicate the non-robust version of the methods, while the solid lines are their robust counterparts.

In Figure 12, with motion artifacts added, the same pattern is shown in that the ROC curves are best for the robust AR partial correlation and robust modified MVGC models in the case of zeroth-lag simulations and best for robust MVGC and robust modified MVGC with non-zero lag simulations. In both cases, the modified MVGC was not quite as good as the AR partial correlation or the MVGC, so if one knew what kind of connections were present, the modified MVGC is not the preferred choice. However, in the absence of this knowledge, this was the only model to work in both scenarios.

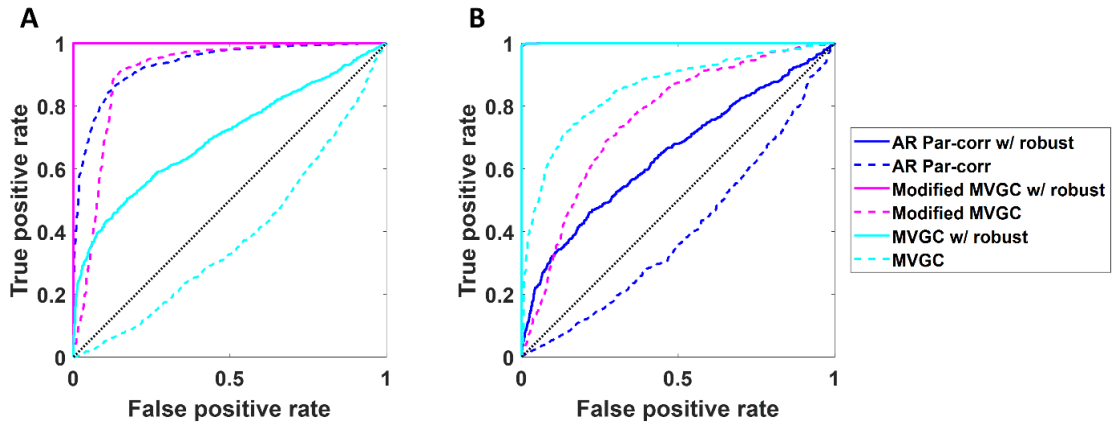

Figure 12. ROC curves comparing whitened partial correlation, multivariate Granger causality, modified multivariate Granger causality with and without robust regression when the linear relationships between channels are simulated at (A) zero-lag, (B) first lag. The simulated data contains temporal autocorrelation, shared global signal mimicking systemic physiology and head motion. The dotted lines indicate the non-robust version of the methods, while the solid lines are their robust counterparts.

Overall simulation results

Figure 13 demonstrates the resulting null distributions produced by each of the simulation types and analysis methods. The dashed red line shows the theoretical distribution assumed by the statistical model. The closer the match between the empirical distribution and the theoretical one, the more accurate the reported statistics and the less bias in the estimates. We see that the standard Pearson's correlation model was quite inaccurate for data with temporally correlated noise. The distribution was too wide compared to the theoretical, which means that *p-values* would be underestimated. Moving to the AR whitened correlation model (second column), we see that this distribution was fixed for the temporally correlated noise (no bias in the estimate), but in the presence of spatially correlated noise (middle row) was shifted from mean zero (bias in the estimate) and the distribution was too wide. The AR partial correlation model did correct for most of the global spatial noise, but then failed when motion artifacts were present. The robust AR partial correlation worked best under these conditions (bottom/right plot), although that distribution is still not ideal.

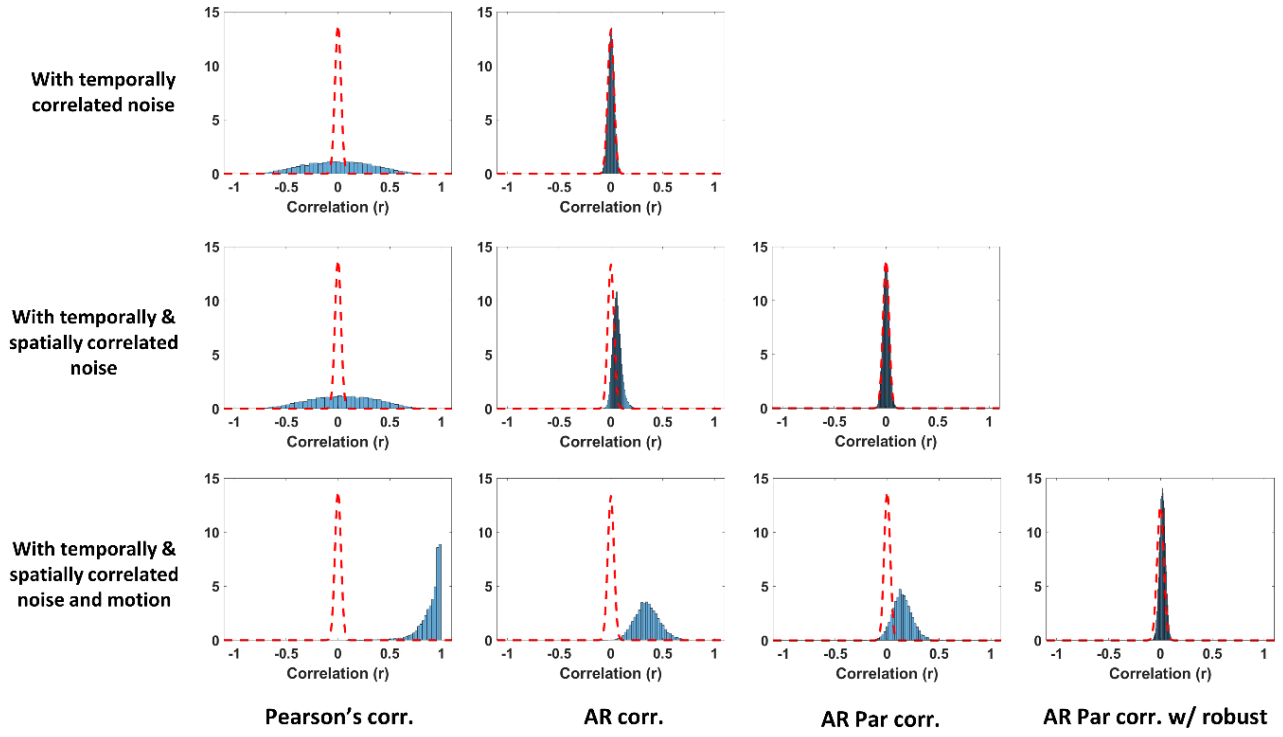

Figure 13. Plots showing the distribution of strength of the relationship between channels (correlation coefficient) quantified by the Pearson's correlation, AR correlation and AR Partial correlation and robust AR Partial correlation. The simulated data contains either temporal autocorrelation (first row), temporal correlation with spatial correlation due to systemic physiology (second row), and temporal correlation, spatial correlation and head motion artifacts (third row). The dashed line indicates the expected distribution without physiological noise or head motion while the histogram shows the distribution of the actual correlation for each combination of simulated data and connectivity method.

Results on experimental RS-fNIRS data

To test the applicability of our functional connectivity methods on experimental RS-fNIRS data, we compared using robust versions of Pearson's correlation, AR correlation and AR partial correlation to model the functional connectivity between the channels on a sample of 24 participants. The group-level results with the significant paths ($q < 0.001$, FDR correction) for the three methods are shown in Figure 14. As the results indicate using AR partial correlation leads to much sparser networks, compared to AR correlation and Pearson's correlation as it corrects for global physiology and autocorrelation. Using a Pearson's correlation revealed a very dense network of connections with most connections likely being false positives. Though, there is no known ground truth or gold standard RS-fMRI networks available for quantitative comparison of the methods, AR partial correlation would probably provide the lowest false positive rate among the three methods.

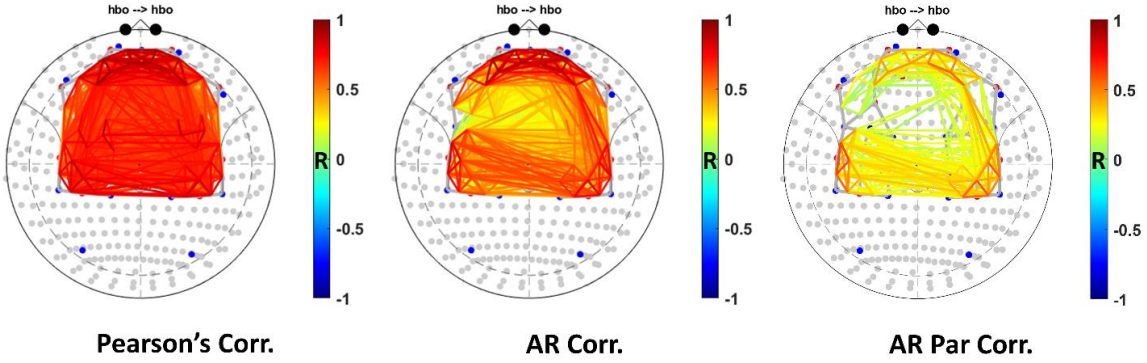

Figure 14. Comparison of the effectiveness of robust versions of Pearson's correlation, AR correlation and AR partial correlation in correcting for temporal autocorrelation, global systemic physiology and motion artifacts in experimental fNIRS data. The maps show significant ($q < 0.001$, FDR corrected) connections between the channels projected on to the 10-5 coordinate system after group analysis using different connectivity methods based on changes in HbO.

Discussion

In the current article, we used simulated data to explore how physiological noise and the sluggish hemodynamic response together lead to temporal autocorrelation and spatial covariance in channel data, which in turn lead to increased false positives and lower AUC for the ROC curves. Although pre-whitening can reduce temporal autocorrelation, spatial covariance due to shared systemic physiology should still be corrected. Our results confirm and expand on previous studies by showing that a combination of pre-whitening with an autoregressive (AR) filter, and partial correlation of short-separation channel data to control for spatial covariance were effective in substantially lowering the false positive rate for resting-state connectivity analysis.

In particular, a shared systemic physiological signal can be captured from the short-separation or both short-separation and long-distance channels, using the first few principal components. We found that long-distance channels alone, while better than no

corrections, were not sufficient to remove the global signals. The use of robust regression was modestly successful in controlling for signal changes due to head motion especially for Granger causality models, although it was very effective for correlation models. One of the challenges in the multivariate models is regarding the definition of motion outliers. One way whether to consider a time point to be an outlier is if it is an outlier in any channel, which tends to lead to over-correction with so many channels and when lagged terms are included (since all lags are also outliers). Another way is to consider a time point as an outlier when it is an outlier in a large enough subset of channels. We followed the latter approach by using the geometric length of the innovations vector (across channels).

In this study, we also introduced a modified MVGC which included both the traditional lagged (causal) terms and a zeroth-lag component. This new model did relieve the assumptions about the zero or non-zero nature of the lag in connectivity and could be used in both scenarios. However, this model scored worse on the ROC analysis for specifically non-zero lag connections compared to the MVGC model and was more sensitive to motion (outliers) compared to AR partial correlation. For both the modified MVGC and MVGC, our robust statistical models were not perfect, and we found that both tended to over-report significance. This as mentioned earlier, may be due to the deficiency of the proposed methods in not fully accounting for the degrees of freedom in the model associated with the Tukey bisquare weighting of outliers in multiple channels of data simultaneously.

Use of pre-whitening models

Temporal autocorrelation in fNIRS and fMRI time series is a known issue that leads to reduced degrees of freedom, increased variance of the correlation coefficient (not accounting causes underestimation of the variance), and thus increased false positives. Although pre-whitening has been suggested to reduce autocorrelation in time series and subsequently false discoveries in both fMRI (Bright et al., 2017; Christova et al., 2011) and fNIRS (Santosa et al., 2017), it must be noted that the use of pre-whitening as a processing step in resting-state analyses is not universally accepted. There are some concerns regarding pre-whitening the time series and its ability to distort the power spectrum which inherently follows the inverse power law (Afyouni et al., 2019; Blanco et al., 2018). Blanco and colleagues argued that if the intrinsic signal is inherently colored, the use of pre-whitening may be inappropriate (Blanco et al., 2018). These researchers observed that the anticorrelation between HbR and HbO signals is diminished by incorporating pre-whitening step in RS-fNIRS processing pipeline (Blanco et al., 2018). However, it must be noted that although the HbO and HbR signals are in antiphase given the cerebral hemodynamics, their peaks are not maximally anticorrelated at zero-lag. Specifically, the positive peak of the HbO and the negative peak in HbR signals do not occur at the same time (Huppert et al., 2006). Zeroth-lag is defined as within the same sample time, which means that this definition depends on the sample rate. As we see from the results in Figures 11 and 12, when there is an intrinsic temporal lag in the underlying connectivity, correlation methods are not designed to capture this. Strictly speaking, HbO and HbR are not actually truly correlated, so much as they are “causal” in the mathematical sense with HbO preceding changes in HbR. Thus, the modified MVGC and MVGC methods are required to model these lagged signals and the finding that HbO

and HbR are no longer correlated once autocorrelation is accounted for, is actually the expected mathematical result.

Recently, other methods than pre-whitening, have been proposed to control for the false positive rate in fMRI. These methods either correct for the reduced degrees of freedom or directly correct for the increased variance of the correlation. Some of these methods were compared on their effectiveness in accounting for the effects of autocorrelation in the time series on functional connectivity in fMRI (Afyouni et al., 2019). However, since fNIRS has a higher sampling rate than fMRI and therefore serial correlations are most prominent, the applicability and the utility of both pre-whitening methods and other methods need to be systematically compared for fNIRS datasets. Our findings differ from previous results from RS-fMRI. Unlike Arbabshirani et al., (2014) who observed some bias in sample correlation coefficient (when the autocorrelation structures were different between the time series), our results were similar to Afyouni et al., (2019) in not finding a bias in the estimate of the sample correlation coefficient for a zero correlation coefficient. Secondly, unlike the results of Arbabshirani et al., (2014) wherein the bias in the estimate canceled out the inflated variance due to autocorrelation, our results show that the inflated variance of sample correlation due to autocorrelation in the time series was much higher. This is probably attributable to temporal dependencies at larger lags due to higher sampling rate of fNIRS data. Hence correction for autocorrelation may be more important for fNIRS than for fMRI, and some of the conclusions from RS-fMRI may not be directly applicable to RS-fNIRS. It should also be noted that the method for correcting for autocorrelations can have important consequences for other steps in the preprocessing pipeline such as motion correction and reduction of spatial covariance using partial correlation. Furthermore, since the autocorrelation in the fNIRS time series is due to a combination of factors including, the hemodynamic response, high sampling rate and presence of physiological noise, removing modeled physiological noise using external measurements, or short-separation filtering, though may reduce the need for higher AR model orders, the residual data is still autocorrelated as has been shown with short repetition times in fMRI (Bollmann et al., 2018). Further validation studies are necessary to support incorporating pre-whitening in the fNIRS data processing pipeline.

Use of filtering methods in the preprocessing pipeline

Filtering methods especially band-pass filtering are often used in the preprocessing pipeline for RSFC as it is assumed that filtering increases the signal-to-noise ratio by reducing the contributions of frequency bands with physiological and other noise sources. However, incorporating both filtering and pre-whitening in the preprocessing pipeline requires some careful thought. First, filtering—especially low pass filtering—often increases the autocorrelation in the time series, worsening the issue of increased false positives in the data. So, any low-pass filtering needs to be performed before pre-whitening (Santosa et al., 2017). Second, filtering removes the high-frequency components in the data, and since pre-whitening tries to whiten the frequency spectrum, it can make the estimates of the AR parameters during pre-whitening unstable. So any filtering should be done within the GLM model (Bright et al., 2017; Santosa et al., 2017). Third, any filtering needs to be applied to both the signal and the noise model (including any physiological signals, short-separation channels) to prevent the reintroduction of the

unwanted frequencies (Bright et al., 2017). Finally, any filtering applied to the data must account for the lost degrees of freedom from filtering, so that the significance of the correlations is not inflated along with increased false positives (Bright et al., 2017; Davey et al., 2013). In the current work, we did not incorporate filtering into the processing pipeline. Future studies must examine if indeed there is an increased sensitivity to the underlying neural activity by incorporating filtering into the preprocessing pipeline, despite some of the issues outlined above.

Use of robust methods to reduce the influence of head motion

Head motion artifacts in fNIRS are often statistical outliers as these changes are much larger in magnitude than the underlying physiological and intrinsic noise in the signal. The introduction of robust statistical estimators is required to avoid the over-leverage of these outlier points on the connectivity estimates. These approaches worked well for the correlation and partial correlation models but only offered a modest improvement for the MVGC models. In particular, these robust methods failed to converge at max iterations of 50 (but did converge when max no. of iterations was increased to 150), especially with MVGC, thus limiting the use of robust MVGC for heavily motion-corrupted data. Future methods should probably explore estimating reliable and robust estimates of multivariate Granger causality in presence of large motion artifacts. The issue with marking and correcting for head motion with multivariate Granger causality is that a motion corrupted time point is downweighted across all lags, hence, there is a large reduction in the degrees of freedom when robust regression is used. Some studies correct for motion artifacts before the estimation of functional connectivity. These methods should explicitly control for the lost degrees of freedom due to motion correction in their analysis of functional connectivity. A related issue to the use of robust methods is to determine what time points are considered outliers and subsequently marked as motion artifacts. Aggressively labeling data as motion-related outliers, could reduce the degrees of freedom of the data and may, unfortunately, lead to downweighting lots of data. However, using lenient thresholds to determine motion-related outliers could lead to increased false positives due to residual motion artifacts uncorrected for in the data. Similarly, robust AR partial correlation though was effective at reducing false discoveries significantly, still had more than expected false positives in the data. Thus, better strategies or multi-step strategies may be needed to reduce the influence of head motion on connectivity.

Impact of imaging duration and sampling frequency on RS-fNIRS

The acquisition rate (sampling frequency, F_s) and the duration of the RS-fNIRS scan determine the number of time points available for the calculation of connectivity metrics and determine the power and sensitivity of these methods. Longer duration scans may lead to stable measures of connectivity, but in infant and other clinical populations, they may present challenges. However, since the temporal autocorrelation also increases with sampling frequency, higher order models may be necessary in data with higher sampling rate to correct for autocorrelation, else it may lead to increased false-positives (Santosa et al., 2017). The methods discussed in this study may need to be modified slightly by using a higher-order AR models for pre-whitening, if the sampling frequency of the data is higher. Nevertheless, we still expect the patterns of results to hold true, with

methods that include pre-whitening outperforming the methods that do not include pre-whitening.

Secondly, different connectivity metrics may require different imaging durations to achieve stability and probably in different populations. The imaging duration may also depend on the effectiveness of the data preprocessing pipeline. Studies show that scan durations as short as 1 minute may be enough to obtain stable and reproducible measures of functional connectivity, and nodal network metrics such as nodal efficiency and nodal betweenness for data acquired at a sampling frequency of 25 Hz in adult populations (Geng et al., 2017). However, longer scan durations (> 5 minutes) may be needed for obtaining stable measures of connectivity measures such as network clustering coefficient, local efficiency and global efficiency (Geng et al., 2017). Similarly, in a sample of children aged 7-8 years with fNIRS data was acquired at 50 Hz, an imaging duration of 2.5-7 minutes was needed to obtain accurate and stable functional connectivity and graph-theoretic network metrics (Wang et al., 2017). Furthermore, lower acquisition rate (~ 4 Hz as in this study) may require longer duration scans to achieve stability of connectivity measures. A thorough analysis on the impact of the acquisition rate and scanning duration on the stability of fNIRS connectivity metrics is needed. Furthermore, the stability of lagged connectivity measures including Granger causality as well as the extensions and variations discussed in this paper with RS-fNIRS data has not been studied and may warrant further research.

Limitations of the current study

This work was primarily a numeric simulation study and, while we introduced structured temporal and spatial noise and motion artifacts as realistically as we could, our findings are nonetheless limited to the properties of our simulations. First, it must be noted that the time series was assumed to be stationary to obtain reliable and valid estimates of connectivity. In real datasets, the use of more complex autoregressive integrated moving average (ARIMA) models may be necessary for pre-whitening. Additionally, the stationarity of the time series can be tested by Augmented Dickey-Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests (Z. Liu et al., 2017). The use of scale-less measures such as WTC could lead to more reliable estimates of connectivity if indeed fNIRS time series are non-stationary. Also, the non-stationarities of the time series could provide interesting insights into the dynamics of functional connectivity across time. Second, in our simulations, we used a Gaussian spatial smoothing kernel to produce spatially correlated noise originating in the skin layer. Spatial heterogeneity in global physiology could reduce the effectiveness of the proposed methods as do the number and location of short-separation measurements. Furthermore, the way the data were simulated, both HbO and HbR contain the same information, so the utility of including both HbO and HbR short-separation channels in the partial-correlation models has not been explored. Analyses with task-based fNIRS data, seem to suggest that adding both HbO and HbR short-separation channels in the model is beneficial (Santosa et al., 2020), future studies should explore this with resting-state analysis. Future studies should replicate the current analyses using experimental RS-fNIRS data and compare the results with the gold standard RS-fMRI to confirm our conclusions. Often, in resting-state analysis, we are concerned more about connectivity differences between population groups or between different conditions. So future studies should also evaluate how the

issues due to autocorrelation, and systemic physiology manifest as group differences or as differences in conditions in functional connectivity and causality.

Conclusion

Complex numerical simulation models may theoretically capture more information about the relationship between fNIRS time channels, but in practice solving them is difficult. In the absence of head motion, use of robust methods may be both computationally inefficient and suboptimal as they can reduce the degrees of freedom. Similarly, although methods such as modified Granger causality can model both the zero-lag and the time-lagged relationships, pre-whitened partial correlation performs best when connectivity information is present at the zero-lag while Granger causality performs best when there is a lagged relationship between time series. The modified MVGC performed well in both scenarios, but at the cost of fewer degrees of freedom and possibly lower sensitivity in situations with low signal-to-noise ratios. Similarly, band-pass filtering, while could probably improve the signal-to-noise ratio and sensitivity, it could also lead to lost degrees of freedom and increased autocorrelation which may offset the gains. Hence a careful consideration of the steps in the preprocessing pipeline is suggested to maximize sensitivity while still reducing false positive rates to near expected levels.

Chapter 3: A review of preprocessing strategies for resting-state functional connectivity analysis with functional near-infrared spectroscopy

Chapter abstract

Functional near-infrared spectroscopy (fNIRS) has become a popular tool to map resting-state networks across the brain. Though functional magnetic resonance imaging (fMRI) is the dominant method for resting-state analyses, fNIRS has certain advantages over fMRI, including higher temporal resolution. Unfortunately, the extracerebral noise and the higher temporal resolution lead to severe temporal autocorrelation in resting-state fNIRS (RS-fNIRS) data. Additionally, noise sources, such as systemic physiology and participant's head movements, can artificially inflate functional connectivity estimates, thereby increasing false positives. Hence correcting for autocorrelation, global signal due to systemic physiology, and head motion artifacts is imperative before estimating functional connectivity and other related measures. Though several methods are proposed to address systemic physiology and motion artifacts, the performance of most of these methods has been evaluated with evoked responses and not on RS-fNIRS data. In this review, we survey methods and strategies used to correct for temporal autocorrelation, systemic physiology, and motion artifacts and discuss their applicability to RS-fNIRS data. Since most articles using resting-state connectivity analyses do not correct for temporal autocorrelation, the issue of increasing variance in the sampling distribution is discussed. We further highlight the increased use of hybrid methods that combine multiple strategies for motion correction. Finally, we discuss some unresolved issues in resting-state connectivity analyses with fNIRS and highlight directions for future research.

Introduction

Functional near-infrared spectroscopy (fNIRS) is a brain imaging method that uses light in the near-infrared window to monitor changes in blood oxygenation, which serves as the basis for researchers to make inferences about changes in neural activity. Oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) have different optical properties, specifically, differential absorption of electromagnetic radiation in the near-infrared window. Moreover, the human tissue is relatively transparent to near-infrared light; hence it can pass through the scalp, skull, and meninges to reach the cortical surface with minimal absorption and scattering. The relative change in the absorption of near-infrared light detected through detectors placed at an appropriate distance on the scalp can be used to infer the changes in concentrations of HbO and HbR in response to neural activity in particular brain regions.

FNIRS has several advantages over other neuroimaging methods. It has higher temporal resolution than functional magnetic resonance imaging (fMRI) and better spatial resolution than electroencephalography. Its portability and adaptability to multiple settings make it a more versatile and ecologically valid neuroimaging modality. It is less susceptible to motion artifacts, thus making it ideal for cognitive neuroimaging applications in developmental and clinical populations. Further, its relatively lower cost and the ability to localize cortical activations make it a useful neuroimaging method to study the human brain across tasks, ages, patient groups, and cultures.

Resting-state functional connectivity (RSFC) or intrinsic functional connectivity characterizes the brain's functional architecture using spontaneous low-frequency fluctuations across anatomically distinct brain regions when the person is not performing an overt task. While predominantly used with fMRI, resting-state analysis with fNIRS is increasingly popular for identifying resting-state networks. A search on the PubMed database with the search term "(resting-state functional connectivity) AND ((functional near-infrared spectroscopy) OR (diffuse optical tomography) OR (fNIRS) OR (NIRS) OR (near-infrared spectroscopy))" revealed a total of 158 publications. The number of publications using resting-state fNIRS (RS-fNIRS) over the years is shown in Figure 1. As the figure illustrates, though the number of publications utilizing RS-fNIRS is minuscule compared to the thousands of publications every year on RS-fMRI, there is still a considerable year-on-year increase in the number of published RS-fNIRS studies. Furthermore, RS-fNIRS is reliable (Niu et al., 2013; Niu & He, 2014; Novi et al., 2016; H. Zhang, Duan, et al., 2011; H. Zhang, Zhang, et al., 2011) and valid (Duan et al., 2012; Sasai et al., 2012) method to study the properties of cortical networks across the brain. A detailed comparison of the advantages of resting-state analysis with fNIRS vis-à-vis fMRI is included in the introduction chapter of this dissertation.

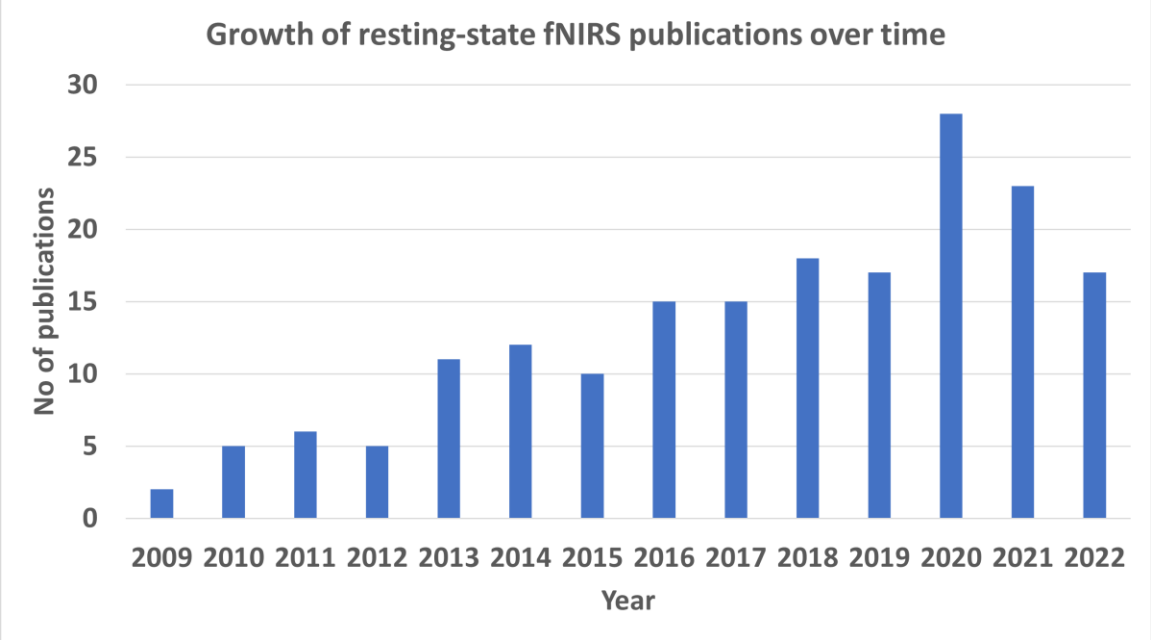

Figure 1. The number of resting-state fNIRS publications over the years.

There are several ways to characterize resting-state connectivity, including in the time domain with Pearson's correlation coefficient (Mesquita et al., 2010; Novi et al., 2016; Sakakibara et al., 2016; Santosa et al., 2017), in the time-frequency domain with wavelet transfer coherence (WTC) (Han et al., 2014; Tan et al., 2015), blind source separation methods such as independent component analysis (Blanco et al., 2021; Ferradal et al., 2016; Zhang et al., 2010), and effective connectivity measures including Granger causality analysis (Medvedev, 2014; Lanka et al., in press). Furthermore, dynamic measures of functional connectivity (Z. Li et al., 2015; Niu et al., 2019) and

graph theoretical measures were also applied to study the properties of resting-state networks with fNIRS (Cai et al., 2018; Niu et al., 2012).

There are some challenges associated with preprocessing RS-fNIRS data and obtaining valid statistical inferences with resting-state connectivity due to the signal and noise properties in the fNIRS data. The hemodynamic signal originating in response to the neural activity in the cerebral cortex is the signal of interest. There are several sources of noise. The global physiological noise in fNIRS mainly arises from the physiologically driven changes in the vasculature that correspond to respiratory, heart rate, and blood pressure changes in the superficial tissue layers, including the scalp and the skull (extracerebral), and from the brain itself. The fNIRS signal is 10-20 times more sensitive to the systemic physiology originating in the blood vessels in the skull and the scalp (extracerebral physiology) than signal changes in the cerebral cortex (Kirilina et al., 2012; Scholkmann et al., 2014; Tachtsidis et al., 2010). The primary sources of systemic physiology are fluctuations attributable to cardiac (around 1.2 Hz), respiratory (0.3-0.6 Hz), and blood pressure (0.1 Hz) cycles (Tachtsidis & Scholkmann, 2016). Another significant source of noise in fNIRS signals is the subject's head movements. The subject's head movements during acquisition can cause the optodes placed on the scalp to slide or momentarily lose contact. These events appear in the fNIRS data as large spikes or changes in the baseline of the fNIRS signal (Brigadoi et al., 2014) and further contribute to the noise in the fNIRS data.

Though the above noise sources are not just limited to RS-fNIRS data, processing RS-fNIRS data can be more challenging than processing task-based fNIRS data. The neuronal signal of interest (typically the experimental design convolved with the hemodynamic response function) and the noise are explicitly modeled in task-based analyses. However, in RS-fNIRS, where there is no expected hemodynamic response, processing often consists of modeling and removing noise sources. Hence any unremoved or statistically unaccounted for noise sources are interpreted as neural fluctuations (Bright et al., 2017). Furthermore, evoked fNIRS responses from task-based paradigms can be averaged across trials to cancel out some types of noises. However, trial-averaging is not feasible with RS-fNIRS. Hence noise has a greater impact on RSFC analyses, and there is a greater need to correct for its effects.

RS-fNIRS signal and the various sources of noise violate the statistical assumptions of Pearson's correlation coefficient, the most commonly used measure to characterize functional connectivity, thus increasing the probability of obtaining incorrect inferences, often leading to increased false positives (Santosa et al., 2017; Lanka et al., 2022). The slow hemodynamic response and the systemic physiology combined with the high sampling rate of fNIRS often induce autocorrelation in the fNIRS time series, thus violating the assumption of independence. Furthermore, motion-induced signal changes appear as outliers in the RS-fNIRS data due to their large magnitude compared to neural fluctuations. Finally, systemic physiology shared across all the channels can increase the spatial covariance between the channels leading to an artificially inflated value of sample Pearson's correlation coefficient between RS-fNIRS time series. These issues have been discussed in detail previously (Santosa et al., 2017; Lanka et al., 2022).

Given the previously described issues, it is imperative to correct systemic physiology and motion artifacts. Several noise correction methods rely on separating the

signal and noise from the RS-fNIRS data based on differences in the spatial, temporal, and spectral differences between the signal and the noise. Often fNIRS data are transformed to clarify/amplify these differences. Many noise correction methods rely on certain assumptions about the properties of the signal and noise. For example, noise originating from head motion and systemic physiology is assumed to be global while neural activity is localized. The neuronal oscillations are assumed to be smooth and localized to a specific frequency band (0.01-0.1 Hz). In contrast, noise is assumed to be abrupt and may (systemic physiology) and may not (motion artifacts) be restricted to specific frequencies. Another class of methods to remove noise from fNIRS data relies on obtaining independent noise measures from external measurements such as accelerometers for head motion, instruments measuring physiology such as HR monitors, pulse oximeters, etc., and short-separation (SS) channels to capture extracerebral physiology. These measurements can then be used to remove the noise from the fNIRS data either with regression or filtering (Santosa et al., 2020).

There is growing concern that correction methods applied to RS-fNIRS data during preprocessing are often insufficient (Hocke et al., 2018; Pfeifer et al., 2018). Furthermore, as mentioned earlier, since methods to correct for noise in fNIRS rely on certain assumptions, these assumptions are often not met, or the methods are not properly applied, leading to suboptimal noise correction (Hocke et al., 2018). The purpose of the current review is three-fold. The first goal is to survey the popular methods used for correcting motion and systemic physiology and examine their efficacy on RS-fNIRS data. The second goal is to highlight the importance of autocorrelation correction for RSFC analyses, as the correction for temporal autocorrelation is often ignored in the preprocessing pipeline with RS-fNIRS analysis. The third and final goal is to highlight the gaps in implementing these preprocessing methods and suggest areas for future research. Accordingly, this review has been organized as follows: In section 2, we describe the issues encountered in drawing valid inferences on temporally autocorrelated fNIRS time series. In section 3, we discuss some strategies, including pre-whitening and variance correction methods, to correct for the effects of autocorrelation on the significance of the correlation coefficient. In the following two sections, sections 4 and 5, we survey some of the methods proposed to correct for systemic physiology and motion artifacts, respectively. We emphasize each method's assumptions, applicability, and pros and cons. In section 6, we discuss some popular toolboxes used for RSFC analyses. Often the toolbox used determines what preprocessing methods and steps are included in an analysis; hence it is vital to understand the methods available in each toolbox to correct for noise. We finally end this review with a discussion in section 7 on some unresolved issues and highlight directions for future research.

Effect of temporal autocorrelation on functional connectivity

FNIRS measures cerebral hemodynamics, a slow and lagged response to the neural activity rather than the neural activity itself. Systemic physiology and hemodynamic response to neural activity combined with the higher sampling rate of fNIRS introduce serial correlations (also termed autocorrelation) into the fNIRS time series. Autocorrelation implies that the individual time points in the fNIRS time series are no longer independent of each other (Santosa et al., 2017). Colored noise can reduce the effective degrees of freedom in the fNIRS time series. The presence of temporal

autocorrelation in fNIRS time series can lead to incorrect statistical inferences with RSFC.

Most studies use Pearson's correlation coefficient to measure the strength of connectivity between two channels. Statistical inferences about the significance of functional connectivity can be drawn from the null distribution of Pearson's correlation coefficient. When there is no autocorrelation (i.e., the observations are independent) in a time series with n time points, and the observations follow a bivariate normal distribution, the sampling distribution of the sample Pearson's correlation coefficient asymptomatically (with large n) follows a normal distribution (James et al., 2019) with mean equal to the population correlation ρ and the variance given by

$$V = \frac{(1-\rho^2)^2}{n}$$

However, since the variance of the sampling distribution of the Pearson's correlation coefficient is dependent on the population correlation ρ, a Fisher r to z transformation is performed as below. Applying this transformation also improves the 'normality' of the distribution.

$$F(r) = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$$

The sampling distribution of the Fisher transformed correlation coefficient asymptomatically follows a normal distribution with a mean given by $F(\rho)$ and a standard error (SE), with an increased sample size as defined below:

$$SE = \frac{1}{\sqrt{n-3}}$$

Thus, a Z-test can be used to test for significance against a null hypothesis of zero correlation. Unfortunately, inferences are valid from the null distribution only in the absence of serial correlation. In serially correlated RS-fNIRS data, the standard error is underestimated if the autocorrelation is not accounted for. Further, differences in the autocorrelation structure of the time series can also introduce bias in the estimates of functional connectivity, i.e., the expected sample correlation differs from the ‘true’ population correlation coefficient (Arbabshirani et al., 2014). Similar results can be obtained if the hemodynamic response is assumed to be a low pass filter acting on the intrinsic neural fluctuations (Davey et al., 2013). Thus autocorrelation could lead to spurious correlations and incorrect inferences regarding the significance of RSFC (Santosa et al., 2017). At a sampling rate of 4 Hz, on simulated fNIRS data with temporal autocorrelation, a false positive rate of a whopping 80% is obtained at an expected false-positive rate of 5% (Lanka et al., in press). Figure 2 illustrates the presence of temporal autocorrelation in the RS-fNIRS time series and its impact on the null distribution of the sample Pearson's correlation coefficient. Furthermore, the validity of inferences drawn with the T-score and the Z-score for sample Pearson's correlation also fails for time series with non-zero cross-correlations (at both zero-lag and non-zero lags) (Afyouni et al., 2019). Thus, the variance of the sampling distribution of the Pearson's correlation coefficient should take into account non-zero cross-correlations and autocorrelations within each time series (Afyouni et al., 2019).

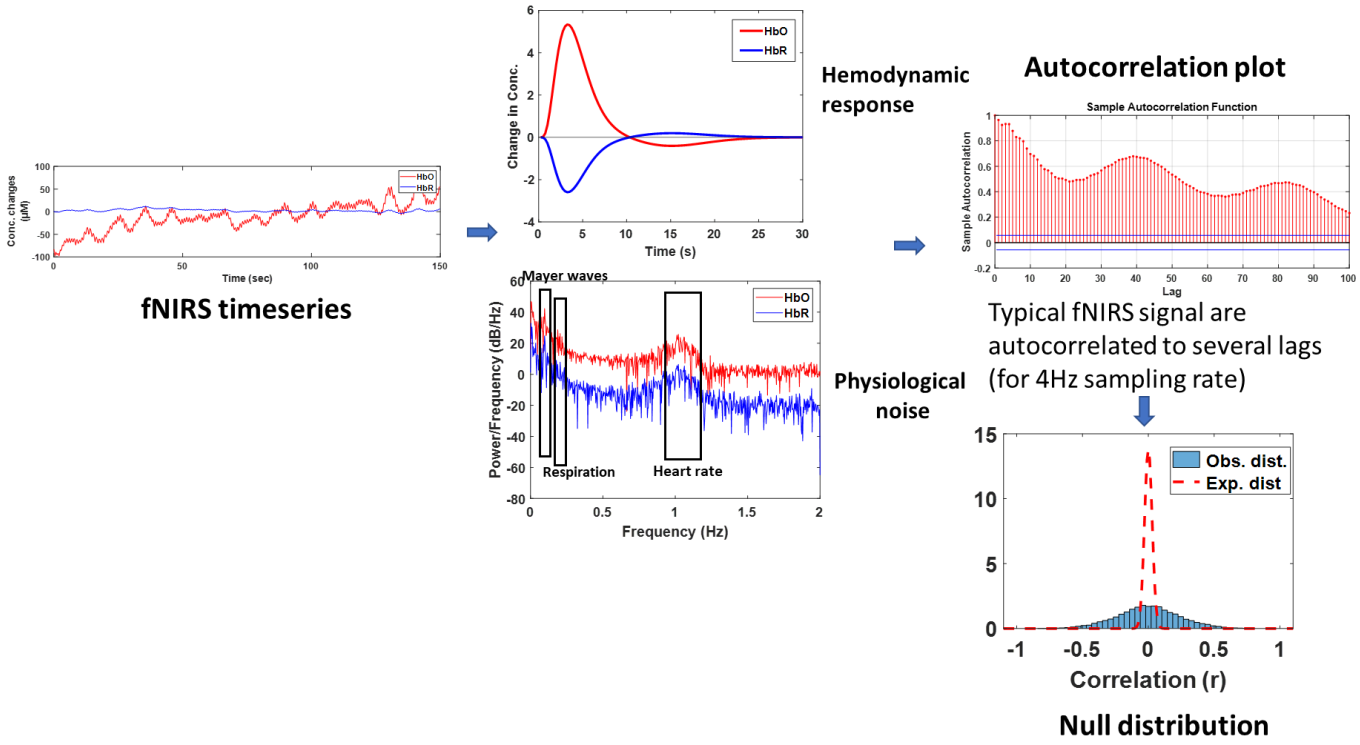

Figure 2. Figure illustrating the sources of temporal autocorrelation in RS-fNIRS time series and the impact of autocorrelation on the sampling distribution of Pearson'correlation coefficient.

This issue is not just limited to RS-fNIRS but also present in RS-fMRI (Arbabshirani et al., 2014; Christova et al., 2011). However, it is much worse with RS-fNIRS due to extracerebral noise contamination and much higher temporal resolution. The false positive rate increases with increasing sampling rate, thus negating the advantage of high temporal resolution of fNIRS over fMRI (Santosa et al., 2017). Therefore, there is a greater need to correct for temporal autocorrelation in RS-fNIRS. Unfortunately, removing the extracerebral noise and systemic physiology does not remove the temporal autocorrelation as one of the significant contributors to the autocorrelation is the hemodynamic response which has an impulse response that peaks after 5-6 seconds and can last for several seconds before returning to baseline (Bollmann et al., 2018; Santosa et al., 2017). Thus, temporal autocorrelation in fNIRS data must be corrected independently of global systemic physiology.

Methods to correct for temporal autocorrelation

One of the two classes of methods proposed to correct for temporal autocorrelation is pre-whitening. Pre-whitening removes temporal autocorrelation and whitens the power in fNIRS time series using autoregressive (AR) models. Pre-whitening on fNIRS entails fitting a p^{th} order AR model to model the current time point as a function of the previous time points (Santosa et al., 2017). After fitting an AR model, the residuals termed 'innovations' which are independent, can be used for estimating functional connectivity. Pre-whitening as a procedure to correct for temporal autocorrelation has been described in previous articles in RS-fMRI (Bright et al., 2017; Christova et al., 2011) and RS-fNIRS (Santosa et al., 2017; Lanka et al., 2022). Typically, a model order for the AR model proportional to the sampling rate of the fNIRS data is chosen to ensure effective correction of autocorrelation (Arbabshirani et al., 2014; Bright et al., 2017). Further, more complex models such as autoregressive-moving-average (ARMA) or autoregressive-integrated-moving-average (ARIMA) can be used to model the fNIRS time series to achieve stationarity (Bright et al., 2017). Pre-whitening is described in detail in other articles (Santosa et al., 2017; Lanka et al., 2022).

Dynamic time-varying connectivity measures are not as frequently used with RS-fNIRS (Z. Li et al., 2015; Niu et al., 2019; Price et al., 2014) compared to RS-fMRI. However, pre-whitening could be helpful with dynamic functional connectivity as it can reduce the variability in connectivity estimates attributable to increased sampling variability due to autocorrelation. This increased variability due to autocorrelation can be misinterpreted as variability in brain states (Honari et al., 2019).

However, there are some concerns about pre-whitening. Firstly, since pre-whitening reverses the effects of autocorrelation by fitting an AR model, the effectiveness of this method is determined by the appropriateness of the model and the ability to obtain accurate parameter estimates of the AR model (Afyouni et al., 2019; James et al., 2019). It has also been pointed out that since the intrinsic signal is inherently colored, pre-whitening may be inappropriate and can complicate the interpretation of pre-whitened time series (Afyouni et al., 2019; Blanco et al., 2018). However, pre-whitening the fNIRS signal in the absence of systemic physiology inverts the filtering effects of the hemodynamic response function. Hence it can be interpreted as a hemodynamic deconvolution filter. Though in the presence of systemic physiology, that interpretation is complicated. Furthermore, pre-whitening reduces the anticorrelation between HbO and

HbR (Blanco et al., 2018), thus limiting the effectiveness of motion correction methods such as correlation-based signal improvement (CBSI).

The second class of methods that correct for autocorrelation compensate for the increased standard error of the sampling distribution of the correlation. Though these methods have been less explored with RS-fNIRS, they have been extensively studied with RS-fMRI (Afyouni et al., 2019; Arbabshirani et al., 2014; Fiecas et al., 2017; James et al., 2019). For a history of methods used for correcting for the increased variance estimates due to temporal autocorrelation, please see Afyouni et al., (2019). Some of the methods that were used to correct for the increased variance are discussed. (i) Bartlett correction factor (BCF) (Bartlett, 1935): Corrects for the inflated variance by assuming an AR(1) time series. Further extensions to BCF proposed by (ii) Assuming a single autocorrelation structure with multiple lags across all nodes (Fox et al., 2005), (iii) Allowing for different autocorrelation structures across multiple lags but not considering the lagged cross-correlations (Bayley & Hammersley, 1946) and (iv) xDF (Afyouni et al., 2019): Correcting for increased variance by considering both autocorrelations and lagged cross-correlations. As expected xDF, had the least bias in estimating the variance of the Pearson's correlation coefficients (Afyouni et al., 2019). Further, the correction proposed by Bayley & Hammersley (1946) was more conservative and overcorrected for the variance due to the confounding of autocorrelation and cross-correlation. However, a limitation of this method is that effectiveness of xDF correction requires knowing auto and cross-correlation estimates in the population. Often with real neuroimaging datasets that is not feasible, so we need to rely on the estimates obtained from the sample.

In RSFC analyses, we are often interested in significance testing and not with point estimates of connectivity (Arbabshirani et al., 2014), so either class of methods can be used. The increased sampling rate can have an proportional impact on lowering the degrees of freedom, underestimating the variance of the sampling distribution, and increasing false positives (James et al., 2019; Santosa et al., 2017), especially with typical sampling rates used in fNIRS. So there is a critical need to evaluate and compare pre-whitening against variance correction methods at higher sampling rates with RS-fNIRS data as only pre-whitening methods have been used (Abdalmalak et al., 2022; Santosa et al., 2017; Lanka et al., 2022) Thus, future studies should compare variance correction methods with pre-whitening methods using RS-fNIRS data at multiple sampling frequencies and in the presence of non-zero cross-correlations.

Method to reduce global systemic physiology

***Band-pass filtering*:** Filtering is probably the most commonly used method to remove systemic physiology in fNIRS processing (Pinti et al., 2019). Band-pass filtering, typically with a lower and higher cut-off frequencies of 0.01 Hz and 0.1 Hz, respectively, is used in RS-fNIRS and RS-fMRI to remove physiology. The success of filtering in removing systemic physiology relies on a few factors. Firstly, the systemic physiology is restricted to a few frequency bands with cardiac fluctuations from 1-2 Hz, respiratory from 0.3-0.6 Hz, blood pressure changes at around 0.1 Hz, and very-low-frequency fluctuations at around 0.04 Hz, as discussed earlier. The low-frequency oscillations typically contributing to resting-state functional connectivity are also restricted to a narrow bandwidth < 0.1 Hz due to the low-pass filtering effects of the hemodynamic response function. So, there is not much frequency overlap between the signal of interest

and systemic noise due to cardiac fluctuations and respiration. Given the relatively high sampling rate of the RS-fNIRS signal, both these noise sources can be filtered out without aliasing. However, the Mayer waves (arterial blood pressure changes) and other very-low-frequency fluctuations overlap with the resting-state connectivity signal and cannot be removed via filtering. Removal of the Mayer wave signal may require using SS channels (Yücel et al., 2016). Despite this, the signal-to-noise ratio of the RS-fNIRS signal is improved due to filtering. Guidelines for choosing optimal parameters, including the filter type and filter order, are provided in Pinti et al., (2019).

However, filtering has some caveats. Though applying a filter can remove some of the systemic physiology, it can reduce the degrees of freedom in the data (Bright et al., 2017) and increase autocorrelation in the RS-fNIRS time series (Christova et al., 2011; Davey et al., 2013; Santosa et al., 2017). Thus if a time series is filtered, then the degrees of freedom in the time series must be corrected appropriately based on a factor that depends on the frequency response of the applied filter (Davey et al., 2013). For an ideal band-pass filter with a lower cut-off frequency of f_l, and the upper cut-off frequency of f_h, and on data with sampling frequency f_s, the effective degrees of freedom can be obtained by a corrective factor given by $2\frac{f_h - f_l}{f_s}$ times the original degrees of freedom (Davey et al., 2013). However, James et al., (2019) have argued that the previously described correction factor overcorrects by overestimating the varia8nce at all sample rates. Alternatively, they propose a more accurate correction factor computed based on the filter's impulse response and the autocorrelation induced by systemic physiology and the hemodynamic response (James et al., 2019).

Principal component analysis (PCA): Blind source separation algorithms such as PCA and independent component analysis (ICA) are also often used for removing systemic physiology from fNIRS (Duan et al., 2018; Virtanen et al., 2009). PCA and ICA decompose the fNIRS signal into a series of orthogonal and statistical independent components, respectively. PCA outperforms ICA in removing systemic physiology (Virtanen et al., 2009), and so PCA is more often used in preprocessing pipeline for removal of systemic physiology, and ICA is used more for identifying cortical networks. PCA-based spatial filters can be used to identify the global physiological signal and partial out its effects when calculating the statistical relationship between channels. PCA depends on the idea that spatial covariance between the signals due to shared physiology is global, whereas the neural variation is local. So, using PCA when the channels are distributed across the head can extract the spatially correlated global signal. PCA-based filtering was as effective as short separation (SS) channel regression in removing the systemic physiology from the data and increasing the reliability of RS-fNIRS metrics (Abdalmalak et al., 2022; Noah et al., 2021). However, PCA requires specification of the number of components to be removed, and PCA cannot be used effectively if the optode locations are confined to a small area on the head (Sato et al., 2016). PCA can also be used on SS channels to capture global physiology. Further, PCA-based filtering with SS can be automated with the best PC component obtained from the PCA decomposition on SS removed rather than the component that explains most variance (F. Zhang et al., 2021).

Short-separation (SS) channels-based methods*:* As fNIRS signal is corrupted by extracerebral signal changes, using SS channels is an effective strategy to remove the

variance associated with extracerebral signal from the fNIRS signal. SS channels in which the source-detector distance is often less than 10 mm are sensitive to superficial physiological noise and not as sensitive to cerebral signal changes as the distance is too short for the detected light to pass through the cerebral cortex (Cooper et al., 2012; Gagnon et al., 2012; Saager et al., 2011; Saager & Berger, 2005). Thus, using SS channels is a very effective way to reduce the influence of superficial physiological noise on the fNIRS signal. Simulation studies have identified the optimum distance for SS channels is 8.4 mm for adults and 2.15 mm for infants (Brigadoi & Cooper, 2015). Although this method is very effective, multiple SS channels may be needed to improve the effectiveness of the methods until an optimal number is reached (Santosa et al., 2020). Increased effectiveness due to more channels can be attributed to the fact that the superficial physiological noise is not homogeneously distributed on the head (Gagnon et al., 2012). Some spatial variability exists in the extracerebral signal changes due to the probe's proximity to major blood vessels. Furthermore., the effectiveness of SS channels increases with good SS channel signal quality and effective distribution of SS channels across the scalp (Santosa et al., 2020). Using SS also removes the effects of Mayer waves in the data that overlap with low-frequency fluctuations of neural origin and hence cannot be easily removed with filtering (Paranawithana et al., 2022).

Partial correlation*:* Partial correlation could be used to reduce the effects of global systemic physiology on functional connectivity. Instead of using standard Pearson's correlation coefficient between time series, partial correlation that controls for the effects of other channels can remove the shared global covariance between the signals due to systemic physiology. Moreover, in the absence of SS channels, partial correlation reflects the relationship between deep brain signals more accurately than Person's correlation (Sakakibara et al., 2016). However, using partial correlation by controlling for other channels can often remove neural variance and increase false negatives. PCA combined with partial correlation is an effective strategy to capture and remove covariance due to global physiology while preserving the neural-related variance. Partial correlation with PCA on long-channels effectively reduced the false-positive rate, though not to the same extent as using SS channels (Lanka et al., in press). However, if PCA is combined with partial correlation, the number of principal components and the variance that needs to be partialed out becomes an important tunable parameter as it balances removing systemic physiology with neural covariance. The optimum value for this parameter would depend on the subject population and the probe layout.

External measuring devices: Another way to reduce the presence of the physiological signals is to use external recording devices to measure respiration, heart rate, and blood pressure and remove their effects from the RS-fNIRS signal (Abdalmalak et al., 2022; Mesquita et al., 2010; Novi et al., 2016; Tachtsidis et al., 2010). These measurements can provide complementary information to those already present in the SS methods in possibly removing systemic physiology obtained from the cerebral cortex (Abdalmalak et al., 2022). However, this method entails additional complexity with setting up and synchronizing these different instruments to record the physiological signals. Furthermore, regressing out physiological noises from these devices from the fNIRS signal may require these measurements to be shifted in time before regression, as

these physiological changes may either precede or succeed fNIRS signal changes (Abdalmalak et al., 2022; Birn, 2012).

Methods to correct for head motion artifacts

Motion artifacts typically occur in the data due to the relative displacement of the optodes with respect to the scalp due to the subject's head movements. So the best way to prevent motion artifacts in the fNIRS data is to ensure that the optodes are tightly coupled to the scalp so that motion artifacts do not arise or are minimal in the first place (Yücel et al., 2014). Often this may not be feasible, especially in infants, elderly or clinical populations. Hence, post-processing correction of motion artifacts is the most common way to deal with motion artifacts (Yücel et al., 2014). As mentioned earlier, fNIRS is less sensitive to motion artifacts than fMRI. However, unlike fMRI, fNIRS does not provide realignment parameters that can be used to characterize subject head movement across the scan duration. Though external motion measurements can be used, they are still relatively rare in fNIRS settings (Virtanen et al., 2011). Motion artifacts are often characterized as generating from a different statistical distribution compared to the neural signal and thus having different spatio-temporal-spectral properties compared to the non-motion corrupted time points. This difference is often exploited in effective motion correction methods. There are usually two steps to correct motion artifacts. In the first step, the motion corrupted time points/segments are identified. In the second step, the motion artifacts are corrected and the motion artifact-corrected fNIRS data is obtained for further analysis. Some of the popular methods to correct for motion artifacts are discussed below.

PCA*:* Just as PCA can be used to correct for global physiology, it can also help correct for motion artifacts. PCA assumes motion artifacts show global, shared, and proportional effects across channels. PCA decomposes the multichannel fNIRS data into orthogonal components ordered by the explained variance. Since motion-related signal changes contribute the most to the data covariance, then by removing the components explaining the largest variance and reconstructing the signal back in the channel space with the remaining components, PCA can be used to correct for head motion. However, as discussed earlier, PCA can overcorrect by removing neural variance, especially if the probe placement is restricted (Jahani et al., 2018; Sherafati et al., 2020).

Targeted PCA (tPCA) *(Yücel et al., 2014)*: Some of the concerns with PCA, including removing neural-related variance, can be ameliorated with tPCA. Unlike standard PCA in which all time points are transformed into an orthogonal basis set, with tPCA only the motion corrupted segments in any channel are included in matrix, that is decomposed into principal components. Motion corrupted segments are identified using a Homer2 function called 'hmrMotionArtifactByChannel' that uses certain criteria to mark motion corrupted segments, including if a segment exceeds a certain peak-to-peak amplitude threshold or if the standard deviation in the segment exceeds a multiple (called STD threshold) of the standard deviation of the entire data. Further, other parameters determine the length of the segment (Huppert et al., 2009). The top principal components that explain the most variance are removed. Following that, the remaining components are retained and projected back to the channel space and joined with the non-motion corrupted segments in a procedure similar to that described in Scholkmann et al., (2010).

This process is repeated multiple times (default value of 3) to remove motion corrupted segments in the data.

TPCA has a few advantages in correcting for head motion artifacts. Firstly it performed better than Spline interpolation and Wavelet filtering for task-based analyses, which were reported to be the best performing methods at that time (Yücel et al., 2014). Unlike standard PCA, there is increased specificity as only the motion corrupted segments are included, so there is minimal concern about loss of neural-related variance. Thirdly, unlike other methods where the motion artifacts are corrected across channels individually, tPCA considers all channels simultaneously. Since motion artifacts are often shared across channel, tPCA as a multivariate approach maybe better suited because even if signal changes are not large enough to be quantified as motion artifacts in one channel, they may be deemed large enough in others. Thus, with tPCA, even the smaller shared signal changes due to motion not deemed large enough to be to be marked and corrected in other methods but that nevertheless impact functional connecitivty metrics, can be corrected.

However, tPCA has a few disadvantages. Firstly, this method requires several tunable parameters to ensure optimal correction, the first set of parameters to identify the motion corrupted segments and the second set of parameters for the PCA decomposition including the number of components to be included and the number of iterations the PCA decomposition process should be repeated. Secondly, this method may not work when head motion is correlated with the task design in task-based paradigm as it may remove neural variance (Yücel et al., 2014). However, it is not a major concern with resting-state analyses.

Correlation-based signal improvement (CBSI) *(Cui et al., 2010):* CBSI relies on the similarity in the hemodynamic response and motion artifacts in HbO and HbR. It assumes that the hemodynamic changes in HbO and HbR associated with neural activity are anticorrelated while the motion artifacts are positively correlated. Since HbO and HbR differ in scale, with HbO having higher amplitude than HbR, a scaled average of HbO & HbR can cancel out the motion artifacts while improving the hemodynamic signal. An advantage of CBSI is that it does not require any free parameters as parameters such as ‘α’ are estimated from the data itself. However, this method has a few drawbacks. First, HBO and HBR are no longer independent measures of the same underlying neural activity. Additionally, since HbR is less corrupted by systemic physiology than HbO, it can increase systemic physiology in the HbR signal. Further, HbO and HbR do not peak at the same time, and there is a temporal lag between their peaks (Huppert et al., 2006; Wolf et al., 2002) that becomes especially prominent at higher sampling rates and if a low-pass and or band-pass filtering step is not performed in the preprocessing. So HbO and HbR are not maximally anticorrelated at zero-lag. Thirdly, this method assumes that the scale of the motion artifacts and the fluctuations due to neurophysiology in HbO and HbR are equal, which may not be justified. Fourth, the method assumes zero or minimal correlation between the motion artifact and the hemodynamic changes, that in certain task-based analyses can be violated as motion could be correlated with the experimental design. However, it is not a cause for concern in RS-fNIRS data.

Spline interpolation-based motion artifact reduction algorithm (MARA) *(Scholkmann et al., 2010):* Like many other motion correction methods, MARA relies on

first identifying the motion corrupted segments by selecting time windows where the standard deviation exceeds the specified threshold based on either signal amplitude or the standard deviation using a sliding window procedure. Then the motion corrupted segments are fitted using cubic smoothing spline interpolation. A tunable parameter p (default 0.99) balances the competing goals of minimizing the error and the curvature (smoothness). Once the data are interpolated, the difference is calculated between the motion corrupted segment and the interpolated data. Then the motion-corrected fNIRS time series is obtained by stitching the difference time series back into the non-motion corrupted segments by shifting the data accordingly by considering the mean value of the previous and current segments. This algorithm also requires several parameters to be tuned, including window length for calculating the sliding window statistics, a threshold to identify motion artifacts, and the smoothing parameter for spline interpolation. This method is effective in correcting for shift artifacts in fNIRS data.

Wavelet filtering (WF) *(Molavi & Dumont, 2012):* WF corrects for motion artifacts by first decomposing the data into wavelet domain using the discrete wavelet transformation (DWT). In the wavelet domain, it is assumed that the motion artifacts, specifically the spike artifacts (large abrupt changes in the amplitude of the signal) present themselves as large wavelet detail coefficients at different levels. A Gaussian distribution is assumed for wavelet coefficients for the background hemodynamic signals, and the wavelet coefficients that exceed a certain probabilistic threshold (α) are considered motion artifacts. The coefficients are zeroed-out for those that exceed the threshold and the data is reconstructed back into the time-domain, thus correcting for the motion artifacts. A parameter 'α' determines which coefficients are considered outliers and is a tunable parameter that balances motion (noise) and signal removal. WF performs well on spike artifacts (large amplitude changes in shorter time intervals). Unfortunately, it does not perform as well against shift artifacts (Yücel et al., 2014).

Kurtosis-based wavelet filtering (kbWF) *(Chiarelli et al., 2015):* A further improvement to the WF was proposed called kbWF which uses a different criterion to identify and remove the influence of outliers in the distribution of wavelet coefficients. Unlike WF, which uses a second-order moment in SD to measure an outlier's departure from normality, the kbWF uses kurtosis, a fourth-order moment. Some of the assumptions of this method are similar to WF, with the notable difference that a motion artifact has a large enough wavelet coefficient to influence the distribution of the wavelet coefficients. The kurtosis of a Gaussian distribution is 3 and the coefficients from neurophysiological data would be Gaussian or sub-Gaussian. So, motion artifacts in fNIRS data can distort a distribution to have a kurtosis greater than 3. KbWF iteratively estimates kurtosis and removes the influence of (zeroes out) the largest wavelet coefficient until the kurtosis below a specified threshold. If the kurtosis is below a certain threshold, the procedure is stopped, and the data is reconstructed back in the time domain with the non-zeroed coefficients. The threshold is chosen to be 3.3 with the rationale that it is closer to 3 but large enough to detect significant deviation in kurtosis. KbWF has similar advantages as traditional WF with a few additional ones. As mentioned earlier, WF relies on threshold 'α' to identify outliers in the distribution of wavelet coefficients. The threshold 'α' can be sensitive to the SNR of the data. In a data with high SNR, a higher α may cause real neurophysiological changes to be removed and using a low α in

low SNR may lead to insufficient removal of motion artifacts (Chiarelli et al., 2015). So, fNIRS data with lower SNR may require a larger α value and vice-versa. Unlike WF, kbWF is less sensitive to the SNR of the data, thus reducing the need to finetune the parameters. Instead, a threshold 3.3 can be used across various SNRs. However, KBWF, similar to WF, also has difficulty removing baseline shift motion artifacts. Similarly, since WF characterizes motion artifacts as large changes in shorter durations, it may be sensitive to the sampling frequency of the fNIRS data (Chiarelli et al., 2015).

Empirical mode decomposition (EMD) *(Gu et al., 2016)**:*** EMD decomposes each channel data into components/basis called intrinsic mode functions (IMF) which are orthogonal to each other and satisfy certain properties such as (a) the number of extrema (minima and maxima) and the number of zero crossings must be the sample or differ at most by one; and (b) at any point in the dataset the mean value of the envelope defined by the maxima and the minima must be zero (Gu et al., 2016). So, this method first identifies data segments contaminated with motion artifacts similar to the previously described methods. Each motion-contaminated data segment is then individually decomposed into IMFs, and the correlation is calculated between each IMF and the original data. The IMF with the largest correlation coefficient is the IMF that captures the motion artifacts. Then this IMF is subtracted from the data, and the remaining IMFs are used to reconstruct the motion-corrected time series. Baseline shifts due to the motion correction are corrected using a similar procedure described in Scholkmann et al., (2010). A few advantages of this method is that it can correct for baseline shifts like some of the previous methods, and does not assume a specific distribution of the neurophysiological signal or the motion artifact. While this method does not require tunable parameters, it does require specifying several parameters for detecting motion contaminated segments.

Temporal derivative distribution repair (TDDR) *(Fishburn et al., 2019)**:*** TDDR makes similar assumptions to other methods that rely on correction in the time domain in that (i) signal fluctuations of neural origin are normally distributed, (ii) fluctuations in the NIRS data due to motion artifacts are much larger in magnitude than those with a neural origin, (iii) motion artifacts occur far less frequently, i.e., just a few time points across fNIRS data are corrupted by motion artifacts. Thus, the signal changes due to motion appear as outliers in the distribution of NIRS signal fluctuations. By down-weighting the time points with large fluctuations, motion-corrected time series can be obtained. TDDR consists of 3 major steps. In the first step, signal fluctuations are obtained by differencing the fNIRS time series. Assessing signal fluctuations also ensures that both spike and baseline shift artifacts can be identified as outliers in the data. In the second step, a weighted mean is calculated and is subtracted from the singal fluctuations to estimate the residuals/variance of the signal fluctuations. Then a robust estimate of the standard deviation of the distribution is estimated from the median absolute deviation. The residuals are scaled by the product of a tuning constant (4.685) and a robust estimate of the standard deviation. The outliers are down-weighted using weights estimated using Tukey's biweight function on the scaled residuals. Since the weights depend on the deviance from the weighted mean, which in turn depends on the weights of the data points, an iterative reweighting procedure is used until all the large fluctuations presumable of non-neural origin are shrunk or eliminated. In the third step, once the

outliers are corrected by downweighing them, the data is integrated to give a motion corrected fNIRS signal.

Unlike some of the methods discussed previously, TDDR has far fewer tuning parameters and works well as long as the assumptions are met. Typically, a tuning parameter of 4.685 achieves a 95% efficiency on a normal distribution. It may depend on the SNR of the data and the proportion of time points corrupted by motion artifacts. Further, the efficacy of this method, like the WF methods, relies on the sampling rate of the data, as a high enough sampling rate is necessary to capture the motion artifact effectively. Also, since TDDR can conflate high-frequency signals with motion artifacts, applying TDDR to the low-frequency part of the data is preferred. Hence the data must be filtered using a low-pass filter, and TDDR must only be applied to the low-frequency component of the fNIRS data. However, this should not be an issue with traditional resting-state preprocessing, as the data is often low-pass filtered. Low-pass filtering, however, can inflate false positives and smear the motion artifact across multiple time points. Since, motion artifacts are not restricted to lower frequencies, TDDR's effectiveness may just be limited to the removal of manifestation of motion artifacts at lower frequencies. While TDDR may be helpful in task-based analyses, its effectiveness may be limited for resting-state connectivity as TDDR corrects for channels individually, while motion is often shared across multiple channels. So TDDR can further be improved by considering and downweighing multivariate outliers in the difference time series across all channels in the multivariate space. Secondly, for RS-fNIRS analysis, rather than temporal differencing, repairing the distribution of the innovation terms after pre-whitening may be more appropriate (Lanka et al., in press).

Motion censoring using global variance of temporal derivatives (GVTD) *(Sherafati et al., 2020):* While a large majority of methods rely on motion detection function 'hmrMotionArtifactByChannel' as specified in Homer2 for identifying motion corrupted time points, a new method called GVTD was proposed as an alternative. Unlike 'hmrMotionArtifactByChannel', which identifies motion artifacts in each channel individually, GVTD relies on all the channels for identifying and marking motion corrupted segments. This method takes its inspiration from framewise displacement (FD), which can be obtained from realignment parameters and derivative variance of signal (DVARS), two correlated measures of signal quality that have been proposed to identify motion corrupted time points for RS-fMRI data (Power et al., 2015). Since we do not have motion measurements without external measuring devices to assess average head motion from one time point to the next, Sherafati et al. proposed GVTD, which relies on the root mean square (RMS) of signal changes/fluctuations/temporal derivatives across all channels, an analogue of DVARS for marking motion corrupted time points in RS-fNIRS data.

Similar to TDDR, GVTD assumes that the fluctuations of neural origin in the fNIRS signal are roughly normally distributed, with motion artifacts appearing as outliers farther from the mean. Thus, GVTD which is computed as the RMS of the temporal derivatives across all channels follow the Chi distribution which is right-skewed when little to no-motion artifacts are present in the fNIRS data. The right skew of GVTD increases with the increased presence of motion artifacts in the fNIRS time series. Thus, a threshold 'c' for GVTD based on the certain standard deviations from the mode (default

value of 10) was chosen to identify motion corrupted time points across all channels. The standard deviation is estimated from the left side of the right-skewed Chi distribution. The motion corrupted time points are then removed and the functional connectivity is calculated from the remaining data. The tuning parameter 'c' optimizes removing too much data and insufficient correction for motion artifacts.

As suggested earlier, multichannel approaches like PCA, tPCA and GVTD may be more appropriate for correcting motion artifacts compared to single channel methods used predominantly. Second, scrubbing /censoring motion corrupted data (i.e., removing them entirely) may not be appropriate for specific resting-state connectivity analyses which require temporal continuity, such as dynamic measures of connectivity and Granger causality. However, since GVTD can be used as a motion detection method, rather than removing the data entirely, it could be combined with other motion correction methods such as Spline interpolation with MARA and tPCA that traditionally rely on the Homer2 function 'hmrMotionArtifactByChannel' for motion artifact detection (Sherafati et al., 2020). Further, even when the motion corrupted time points are interpolated to avoid temporal discontinuities, the interpolation error may increase the uncertainty of connectivity metrics.

***Hybrid approaches*:** Since, some methods are more effective in correcting for certain type of artifacts, there is increasing consensus that hybrid approaches combining multiple motion correction approaches may be more effective than any single method. Several hybrid motion correction methods have been proposed and can be formulated. Some of the popular ones are described below.

Spline Savitzky–Golay (Jahani et al., 2018): This method is one of the first hybrid methods to correct for motion. First, the gradient of the data is estimated after passing the data through a Sobel filter to detect edges or large signal changes in the data. Outliers are then identified in the gradient signal as values that are 1.5 times the interquartile range of the signal gradient. The part of the signal not corrupted by motion artifacts is then used for estimating the SNR of the signal. Second, the segments where the time points exceed a certain threshold are identified as baseline shifts and interpolated and removed with spline interpolation (Scholkmann et al., 2010). This effectively removes the baseline shifts and slow spike artifacts. If the SNR of the data is not high enough, then the spline interpolation step is omitted. Finally, the data is passed through the Savitsky-Golay (SG) smoothing filter to further remove the fast spikes in the data. This way, baseline shifts and spike artifacts are corrected with this approach. However, the downside of using a hybrid approach is that tunable parameters now increase to the sum of tunable parameters of each method. So spline and Savitzky–Golay hybrid requires optimizing the threshold for identifying baseline shifts and the parameter p for fitting the smoothing spline interpolation. Further, two additional parameters need to be optimized for the SG part of the algorithm: the frame length and the polynomial order.

Spline interpolation & WF hybrid approach (Di Lorenzo et al., 2019): Spline interpolation and wavelet filtering are some of the best-performing methods for motion correction (Brigadoi et al., 2014; Cooper et al., 2012; Di Lorenzo et al., 2019). Wavelet filtering and spline interpolation can be used in tandem to correct for both shift and spike artifacts encountered in the data. The success of the combination method relies on the fact that spline interpolation is more successful at correcting baseline shifts, and wavelet

filtering is more effective in correcting for spike artifacts. In fact, a combination of Spline interpolation and wavelet filtering outperformed either method alone, tPCA, and Spline Savitzky–Golay in correcting for motion artifacts in multiple infant datasets for task-based analyses (Di Lorenzo et al., 2019). Further, this combination was also effective in correcting for motion artifacts in speech tasks (Novi et al., 2020).

Robust correlation methods *(Santosa et al., 2017)**:*** Unlike the methods discussed earlier that correct for motion during preprocessing, robust estimates of the Pearson's correlation coefficient account for motion artifacts during the estimation of functional connectivity itself. This method is used in tandem with pre-whitening (Santosa et al., 2017) as the innovation terms of neurophysiological origin are assumed to be normally distributed with motion artifacts as outliers in the distribution of the innovation terms. This method consists of two steps. In the first step, the geometric length of both the time courses is calculated with the scaled residuals (innovation terms). Then a pre-weighting function based on the square root of Tukey's biweight function is used to downweigh the influence of outliers, similar to the 'repair' step in TDDR. This step corrects for motion artifacts that are shared across channels. In step 2, robust correlation coefficient estimates are used to estimate functional connectivity from the pre-weighted and pre-whitening fNIRS. Several approaches for calculating robust correlation coefficient estimates can be found in Shevlyakov & Smirnov, (2011). Robust estimates of correlation can be obtained by taking the geometric mean off the robust regression coefficients of the regression of one time series over the other. This method has been described in detail in Santosa et al., (2017). This step further corrects for motion outliers that are further present in just one of the two time series.

Though the performance of robust correlation has not been compared against other methods, it successfully reduced the false positives due to head motion on RSFC estimates (Santosa et al., 2017). Furthermore, it is one of the few methods developed for RS-fNIRS and can be combined with other methods during preprocessing. An additional advantage of robust methods is that not many parameters need to be tuned. A single tuning constant κ determines the weights in the Tukey's biweight function and is typically set to 4.685. Unfortunately, like most motion correction methods, this method can fail when the assumptions of the method fail, including when motion artifacts are not large enough to be classified as outliers. Additionally, it can remove some neural-related variance.

Summary of motion correction methods: Though several methods have been proposed to correct for motion artifacts in fNIRS data, a few issues remain. While similar methods can be applied to correct for head motion artifacts for processing resting-state functional connectivity as with task-based connectivity, it is important to note that methods which perform best with task-based data do not necessarily perform best with RS-fNIRS. This can be attributed to the univariate nature of task-based analysis with a known expectant response and the bivariate and multivariate nature of most functional connectivity methods. This difference not only applies for motion correction but also for motion detection as well. In general, methods like spline interpolation and wavelet filtering perform much better with task-based analysis than tPCA (Chiarelli et al., 2015), while tPCA that relies on multiple channels usually tends to perform better for resting-state analysis (Sherafati et al., 2020). Furthermore, since a lot of methods require specific

parameters to be tuned/optimized for effective motion correction, and often these methods are compared on evoked fNIRS data, it is unclear if these same default parameters also work with RS-fNIRS data. Additionally, optimal parameters often depend on the properties of the fNIRS data and the relative prevalence and type of the motion artifacts. Hence, studies should often report data quality measures such as signal-to-noise (SNR) ratio, number of time points marked as motion outliers, the type of motion artifacts, etc. For example, wavelet filtering and TDDR may perform well at higher sampling rates when the motion artifact is captured in detail, while CBSI may work better and distort the data less at lower sampling frequencies.

Many methods characterize motion artifacts as outliers deviating from the normal distribution either on the time series itself or after transformation into a different domain. So, motion correction, is often about finding effective strategies to remove outliers or their influence from the fNIRS time series and subsequent analyses. So different strategies to remove the influence of outliers can be tested and optimized. Often signal fluctuations are not large enough to be considered motion artifacts but can significantly overestimate functional connectivity metrics when shared across multiple channels. Additionally, methods that work for adults may not necessarily work for infants, given infants' propensity to move more and their inability to follow instructions to remain still for the experiment.

Finally, while most data processing methods can be used for most types of analyses, methods that rely on interpolation or motion censoring can alter the degrees of freedom available in the data and may cause discontinuities that may either preclude or increase the error in estimates of certain resting-state analysis including dynamic connectivity and Granger causality analyses. So, there is a greater need for a thorough comparison of the effectiveness of motion correction methods on multiple datasets and validating them with resting-state fMRI.

RSFC analyses toolboxes

Several toolboxes provide users access to tools and methods for processing RS-fNIRS data. These toolboxes often provide greater accessibility to RS-fNIRS analysis and can help with reproducibility when the processing steps are properly documented. Many of these toolboxes also provide tools to visualize data effectively and compare the results using multiple processing pipelines. Below, we list a few popular toolboxes that provide preprocessing methods as well as methods to estimate functional connectivity and obtain graph-theoretical measures. A notable exclusion in the below list is Homer2 (Huppert et al., 2009). Though it contains several preprocessing methods, it does not support resting-state connectivity analyses.

FC-NIRS *(Xu et al., 2015)**:*** FC-NIRS is one of the earliest and most used toolboxes for RS-fNIRS analyses. It is a standalone application based on MATLAB and is compatible with Windows and macOS. It has tools in the preprocessing pipeline that converts optical density to HbO and HbR conc. using modified Beer-Lambert's law (MBLL). Further, it provides band-pass filtering with a default range from 0.01-0.1 Hz to remove systemic physiology. For motion correction, the FC-NIRS offers a PCA-based correction and spline interpolation with MARA (Scholkmann et al., 2010). Earlier versions supported CBSI for motion correction. Further, the toolbox provides options for

detrending (removing linear trends from the data) and global signal regression (to remove global systemic physiology).

For calculating connectivity metrics, FC-NIRS supports lagged cross-correlation and Pearson's correlation. Further, it supports graph theoretical network analysis by interfacing with the GRETNA toolbox (Wang et al., 2015). The network metrics are calculated after thresholding either on the absolute threshold (selects edges if the strength of the connectivity is greater than the threshold) or a sparsity threshold (selects edges based on the specified proportion of top edges).

NIRS-KIT *(Hou et al., 2021):* NIRS-KIT is a MATLAB-based toolbox. It supports multiple fNIRS data formats and instruments. NIRS-KIT can support spatial registration of individual positioning of NIRS optodes and channels to standard MNI space when a 3D digitizer is used to capture optodes locations. Further it has several standard RS-fNIRS preprocessing steps available, including converting signal intensities to optical density and further to HbO and HbR conc. using MBLL. Further preprocessing steps that can be used in a pipeline include time point trimming (removes the first few points for experimental acclimatization), detrending (uses a polynomial regression model to remove linear and higher order trends in the signal), motion correction (CBSI and TDDR) and filtering. The default band-pass filtering of 0.01 to 0.08 Hz is suggested. To remove the systemic physiology from the RS-fNIRS data, NIRS-KIT provides SS regression, or if SS channels are unavailable, decomposition the RS-fNIRS data into functional and systemic signals based on their hemodynamic differences as described in Yamada et al., (2012). It also provides options for incorporating customized preprocessing steps in the pipeline.

Additionally, NIRS-KIT supports multiple RSFC analyses, including ROI to ROI, ROI to the whole brain, pair-wise connectivity, and ICA decomposition. Functional connectivity is assessed using Pearson's correlation coefficient. Like the FC-NIRS, NIRS-KIT also interfaces with the GRETNA toolbox (Wang et al., 2015) to calculate graph-theoretical network measures and supports both absolute and sparsity threshold. Furthermore, the toolbox also supports amplitude of low-frequency oscillations (ALFF) & fractional amplitude of low-frequency fluctuations (fALFF) that measure low-frequency oscillations that are the basis for resting-state signal.

NIRS AnalyzIR toolbox (Santosa et al., 2018)*:* It is a MATLAB-based toolbox that utilizes object-oriented programming paradigms for the representation of fNIRS data at different stages in data processing and provides appropriate methods depending on the object class at those stages. It is the most flexible toolbox among the toolboxes described as it interfaces with Homer2 and can utilize all the preprocessing tools for motion corrections and removal of systemic physiology in Homer2. However, the GUI interface in the NIRS AnalyzIR toolbox is extremely limited in functionality, and running the toolbox requires some prior experience with MATLAB's command line interface and object-oriented programming. Specifically, the NIRS AnalyzIR toolbox supports correction of autocorrelation in the RS-fNIRS time series using pre-whitening (Santosa et al., 2017). Furthermore, along with motion correction methods implemented in Homer2, the NIRS AnalyzIR toolbox further supports robust regression methods in general and robust correlation in particular to reduce the influences of outliers on the correlation coefficient estimate (Santosa et al., 2017).

Furthermore, for connectivity analysis, NIRS AnalyzIR toolbox supports Pearson's correlation coefficient, AR correlation (estimates the correlation after pre-whitening), partial correlation (controlling for global physiology by controlling for either SS channels, other long channels or both), Granger causality analyses. All these correlation methods can also be implemented in robust methods framework. The effectiveness of some of these methods in obtaining statistically valid inferences on functional connectivity has been studied in Lanka et al., (in press). Finally, NIRS AnalyzIR toolbox interfaces with the Brain Connectivity Toolbox (Rubinov & Sporns, 2010) to calculate brain network measures including measures of functional integration, functional segreagation and centrality.

LIONirs *(Tremblay et al., 2022):* LIONirs is also a MATLAB-based toolbox that aims to provide greater visualization and flexibility for processing fNIRS data without the need for programming expertise. It provides options for specifying processing pipelines and also allows for a hierarchical pipeline with branches for comparing alternative preprocessing methods. It provides a GUI and also provides options for visualizing multimodal data, including from external measuring devices (heart rate, respiration) and electrophysiological data (EEG). To assess the signal quality of channels, LIONirs uses the presence of the cardiac signal within a user-specified frequency. It supports tPCA (Yücel et al., 2014) or PARFAC (Hüsser et al., 2022) for motion artifact correction. For physiological noise correction, the toolbox provides options to filter and regress out the measures of physiology obtained either from external devices, SS channels, or the global signal average (GSR). After noise correction, the toolbox provides a step to convert the optical density to HbO and HbR using the MBLL.

For resting-state analyses, it supports multiple functional connectivity (FC) measures, including Pearson correlation, Hilbert joint phase probability (Molavi et al., 2014), and magnitude squared coherence (Kida et al., 2016). Further, individual connectivity matrices are converted to Z-maps using Fisher r to z transformation. Finally, non-parametric tests are used to test for the significance of the connectivity measures.

Discussion

There are several unresolved questions in RS-fNIRS preprocessing. Unlike in RS-fMRI, there are no commonly used preprocessing pipelines with RS-fNIRS analysis (Hocke et al., 2018). While we would not like to recommend any specific methods for preprocessing to obtain statistically valid measures of functional connectivity, we recommend including methods that correct for the systemic physiology of both cerebral and extracerebral, temporal autocorrelation, and head motion artifacts in the preprocessing pipeline. Given the experimental setup and availability of short separation channels and external physiology recording devices, different preprocessing pipelines may be appropriate. While using both SS and external recording devices to remove systemic physiology would improve the fNIRS data quality, the complexity of the setup may offset these benefits in some situations.

Additionally, careful thought must be given to the inclusion of a preprocessing step and its efficacy and consequences downstream in the analysis. For example, as discussed earlier, while filtering removes systemic physiology and improves SNR, at the same time, it reduces the effective degrees of freedom in the fNIRS time series and can increase autocorrelation. Another major issue with RS-fNIRS that is unresolved is the

order in which preprocessing steps need to be applied. For example, filtering prior to pre-whitening may make the AR coefficients unstable. However, filtering after pre-whitening may reintroduce temporal autocorrelation. Also, motion correction may need to be performed prior to filtering as filtering can smear the motion artifacts and can corrupt adjacent time points. The impact of the order of the preprocessing steps needs to be explored in future studies.

Similarly, motion correction methods with several tunable parameters can be optimized for effective motion correction, implying that they perform extremely well if the optimal parameters are tuned and do not perform as well when they are not tuned (Hocke et al., 2018). Conversely, other methods with fewer tunable parameters perform moderately in most cases. While hybrid methods are ideal for motion correction, they also increase the number of tunable parameters.

The impact of autocorrelation and the correction strategies for autocorrelation in RS-fNIRS has been understudied, despite the greater impact of autocorrelation on RS-fNIRS analysis due to the higher sampling rate. Furthermore, autocorrelation correction has received scant attention except for a few publications compared to methods that correct for systemic physiology and motion correction. The impact of autocorrelation on group-level analyses should be studied as rigorous correction for filtering and autocorrelation could reduce previously observed differences in functional connectivity between different study populations (Arbabshirani et al., 2014; Fiecas et al., 2017; James et al., 2019). Similarly, due to the widespread use of graph-theoretical measures with RS-fNIRS and the potential for autocorrelation to significantly impact these measures (Afyouni et al., 2019), further studies should be conducted with autocorrelation correction to validate these findings.

The Mayer waves signal and their contribution to the changes in HbO depend on the posture during data collection with higher blood pressure and contamination with the Mayer waves in supine compared to sitting and standing (Tachtsidis et al., 2004). These differences must be considered when correcting for systemic physiology and comparing across population groups when imaging in different postures, or across imaging modalities as in fMRI the data collection is performed in the supine position.

The impact of the preprocessing methods and sampling rate on the scan duration for obtaining reliable functional connectivity and network measures must be studied. Though a few studies have examined this issue they have come up with variable scan durations ranging from 1-7 minutes to obtain stable and reproducible functional connectivity estimates and graph theoretical network measures (Geng et al., 2017; Wang et al., 2017). Typically, graph theoretical network measures may require shorter duration scans for stable and reproducible measures than correlation metrics, and different measures would require different scan durations (Geng et al., 2017; Wang et al., 2017). There could be a difference in the minimum scan duration for adults and infants because of the differences in signal and noise properties in the RS-fNIRS data. Also, longer scan durations in infants and children may not be feasible. Further, effective preprocessing methods and pipelines that include pre-whitening may require shorter duration scans to obtain reliable measures of functional connectivity as pre-whitening the time series can reduce the variance in the sample Pearson's correlation coefficient. More research is needed to address the knowledge gaps in this area.

Often fNIRS publications do not describe the preprocessing methods in sufficient detail or perform suboptimal processing which can lead to variability in the fNIRS results (Pinti et al., 2019). This is further exacerbated by the different montages and the variability in the placement of probes. Unfortunately, these issues may limit the reproducibility and the generalizability of the findings. Moreover, certain methods such as PCA may work best only with certain probe placements. Similarly, despite the higher spatial resolution, most studies use a 10-10 coordinate system for placing probes and inferring activations about the brain function. Future studies should examine how the variability in probe placement across subjects and the inability to localize the signal properly to a standardized space affect group level analysis by comparing the 10-10 system with other localization methods such as using a digitizer and an anatomical MRI to map optode locations.

Further, unlike with fMRI, given the variability of experimental setups, study populations, and fNIRS devices, no single preprocessing method or pipeline will be appropriate in all contexts. Efforts to develop an automated pipeline have not succeeded (Hocke et al., 2018). However, there is a greater need to understand the relative effectiveness of each methods across various contexts. Thus, there is a need for studies that perform a comparison of multiple preprocessing pipelines across multiple data sets and validate the results with resting-state fMRI as a “gold standard”.

Chapter 4: General conclusion

Though fNIRS has enormous potential for mapping resting-state functional connectivity, that potential has still not been realized, primarily due to some of the issues raised in the preceding two chapters. In Chapter 2, we proposed a strategy to reduce the impact of the global signal due to shared systemic physiology across the brain. More importantly, we proposed methods to model lagged relationships and information flow using multivariate Granger causality analyses. The proposed robust multivariate Granger causality (MVGC) analysis is relatively effective in correcting for global systemic physiology and head motion artifacts. Further, given the need /opportunity to model lagged connectivity, the extension to MVGC called modified MVGC with zero-lag can test for both zero-lag and lagged relationships between cortical regions. Unfortunately, we could not test the effectiveness of these proposed methods on experimental fNIRS data, which is a significant limitation of the study. While efforts were made to make the resting-state fNIRS (RS-fNIRS) data as realistic as possible, given the difficulty in accurately modeling experimental fNIRS data, it is possible that some factors that impact the effectiveness of our methods were not modeled. Thus, the results may not generalize well to experimental RS-fNIRS data. Unfortunately, an objective evaluation and comparison of the efficacy of several noise correction methods on experimental RS-fNIRS data is extremely challenging as we often do not know the "ground truth." While visual inspection of "known networks," an increase in intra-subject reliability or validation with functional magnetic resonance imaging (fMRI), considered the "gold standard," can be used to compare the effectiveness of several noise correction strategies, each of these validation approaches has its limitations.

Several methods, as discussed in Chapter 3, have been proposed to deal with motion artifacts and remove systemic physiology. Unfortunately, most of these methods have been validated on evoked fNIRS data, and their applicability to RS-fNIRS is unclear. We hope that noise correction strategies will be developed based on the specific challenges associated with resting-state functional connectivity analyses. Similarly, correction for temporal autocorrelation during resting-state connectivity processing has largely been ignored, with most studies not accounting for temporal autocorrelation. Unfortunately, not accounting for temporal autocorrelation in the data could lead to spurious correlations and increased false positives.

Choosing the proper preprocessing steps to correct for motion, temporal autocorrelation, and head motion artifacts is subjective and challenging, as the optimal strategy often depends on several factors, including the placement of optodes, study population, experimental setup, data acquisition parameters such as sampling rate, presence/absence of short separation channels, and use of external devices that measure physiology. Hence there is no one effective data processing pipeline that would work with all scenarios and applications. Researchers should carefully consider the strategies they plan to use for noise correction, visualize their data before and after the application of the method, understand the impact of the technique on the data and other preprocessing steps, and finally, the order of this method in the preprocessing pipeline. For example, as discussed in Chapter 3, there is often an added cost associated with filtering. While filtering can increase the signal-to-noise ratio, it reduces the degrees of

freedom. Reduction in the degrees of freedom increases the uncertainty in our connectivity estimates. Also, filtering before pre-whitening can make the AR coefficients unstable, whereas filtering after pre-whitening can reintroduce temporal autocorrelation in the data. Often the preprocessing steps are informed by the type of the RSFC analyses and the subject population. For example, there may be a greater need to correct for physiology in adults due to greater extracerebral contamination of the fNIRS signal than in infants. Similarly, there is a greater need to correct for head motion in infants than in adults.

Unfortunately, many noise correction methods have to balance sensitivity, specificity, and generalizability. While techniques with several parameters can be optimally tuned to work well for a particular dataset, those optimized parameters may be suboptimal for a different dataset. Similarly, in the absence of head motion, robust methods may be computationally inefficient and suboptimal, as they can reduce the degrees of freedom. We hope that future researchers document the properties of their data (including signal-to-noise ratio and data quality metrics), filter characteristics (if filtering was included in the preprocessing pipeline), the range and mean number of time points per subject marked as motion artifacts, the kinds of motion artifacts encountered in the data, and the parameters used for noise correction strategies, to name just a handful of items of importance. Documenting this information should contribute to improved reproducibility of resting-state findings, and can contribute to the development of automated preprocessing pipelines in the future.

While fMRI may be the dominant imaging modality for mapping connectivity, fNIRS can be extremely useful in mapping brain connectivity in hyperkinetic populations, children, clinical populations, populations in developing countries, and subjects in which MRI is contraindicated due to non-removable electronic devices or metal implants. Another exciting area of research for resting-state analyses is multimodal imaging, in which fNIRS can be used concurrently with either fMRI or electroencephalography. FNIRS is particularly suitable for multimodal imaging as it does not interfere with other imaging methods and can synergistically help map the functional organization and information flow in the brain. The first step to increase the utility of resting-state connectivity analyses is to develop optimal noise correction methods as optimal noise correction in resting-state data leads to reliable and generalizable results, thus contributing to scientific progress.

References

Abdalmalak, A., Novi, S. L., Kazazian, K., Norton, L., Benaglia, T., Slessarev, M., Debicki, D. B., Lawrence, K. S., Mesquita, R. C., & Owen, A. M. (2022). Effects of Systemic Physiology on Mapping Resting-State Networks Using Functional Near-Infrared Spectroscopy. *Frontiers in Neuroscience*, *16*, 220. https://doi.org/10.3389/FNINS.2022.803297/BIBTEX

Afyouni, S., Smith, S. M., & Nichols, T. E. (2019). Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation. *NeuroImage*, *199*, 609–625. https://doi.org/10.1016/j.neuroimage.2019.05.011

Arbabshirani, M. R., Damaraju, E., Phlypo, R., Plis, S., Allen, E., Ma, S., Mathalon, D., Preda, A., Vaidya, J. G., Adali, T., & Calhoun, V. D. (2014). Impact of autocorrelation on functional connectivity. *NeuroImage*, *102*(P2), 294–308. https://doi.org/10.1016/j.neuroimage.2014.07.045

Badhwar, A. P., Tam, A., Dansereau, C., Orban, P., Hoffstaedter, F., & Bellec, P. (2017). Resting-state network dysfunction in Alzheimer's disease: A systematic review and meta-analysis. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *8*, 73–85. https://doi.org/10.1016/j.dadm.2017.03.007

Barker, J. W., Aarabi, A., & Huppert, T. J. (2013). Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomedical Optics Express*, *4*(8), 1366. https://doi.org/10.1364/boe.4.001366

Bartlett, M. S. (1935). Some Aspects of the Time-Correlation Problem in Regard to Tests of Significance. *Journal of the Royal Statistical Society*, *98*(3), 536. https://doi.org/10.2307/2342284

Bayley, G. V., & Hammersley, J. M. (1946). The "Effective" Number of Independent Observations in an Autocorrelated Time Series. *Supplement to the Journal of the Royal Statistical Society*, *8*(2), 184. https://doi.org/10.2307/2983560

Betzel, R. F., Byrge, L., He, Y., Goñi, J., Zuo, X. N., & Sporns, O. (2014). Changes in structural and functional connectivity among resting-state networks across the human lifespan. *NeuroImage*, *102*(P2), 345–357. https://doi.org/10.1016/j.neuroimage.2014.07.067

Birn, R. M. (2012). The role of physiological noise in resting-state functional connectivity. In *NeuroImage* (Vol. 62, Issue 2, pp. 864–870). Academic Press. https://doi.org/10.1016/j.neuroimage.2012.01.016

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, *34*(4), 537–541. https://doi.org/10.1002/mrm.1910340409

Blanco, B., Molnar, M., & Caballero-Gaudes, C. (2018). Effect of prewhitening in resting-state functional near-infrared spectroscopy data. *Neurophotonics*, *5*(04), 1. https://doi.org/10.1117/1.nph.5.4.040401

Blanco, B., Molnar, M., Carreiras, M., Collins-Jones, L. H., Vidal, E., Cooper, R. J., & Caballero-Gaudes, C. (2021). Group-level cortical functional connectivity patterns using fNIRS: assessing the effect of bilingualism in young infants. *Neurophotonics*, *8*(02), 025011. https://doi.org/10.1117/1.nph.8.2.025011

Bollmann, S., Puckett, A. M., Cunnington, R., & Barth, M. (2018). Serial correlations in single-subject fMRI with sub-second TR. *NeuroImage*, *166*, 152–166. https://doi.org/10.1016/j.neuroimage.2017.10.043

Bressler, S. L., & Seth, A. K. (2011). Wiener-Granger Causality: A well established methodology. In *NeuroImage* (Vol. 58, Issue 2, pp. 323–329). https://doi.org/10.1016/j.neuroimage.2010.02.059

Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., Gagnon, L., Boas, D. A., & Cooper, R. J. (2014). Motion artifacts in functional near-infrared spectroscopy: A comparison of motion correction techniques applied to real cognitive data. *NeuroImage*, *85*, 181–191. https://doi.org/10.1016/j.neuroimage.2013.04.082

Brigadoi, S., & Cooper, R. J. (2015). How short is short? Optimum source–detector distance for short-separation channels in functional near-infrared spectroscopy. *Neurophotonics*, *2*(2), 025005. https://doi.org/10.1117/1.nph.2.2.025005

Bright, M. G., Tench, C. R., & Murphy, K. (2017). Potential pitfalls when denoising resting state fMRI data using nuisance regression. *NeuroImage*, *154*, 159–168. https://doi.org/10.1016/j.neuroimage.2016.12.027

Buchbinder, B. R. (2016). Functional magnetic resonance imaging. *Handbook of Clinical Neurology*, *135*, 61–92. https://doi.org/10.1016/B978-0-444-53485-9.00004-0

Cai, L., Dong, Q., & Niu, H. (2018). The development of functional network organization in early childhood and early adolescence: A resting-state fNIRS study. *Developmental Cognitive Neuroscience*, *30*, 223–235. https://doi.org/10.1016/j.dcn.2018.03.003

Chang, C., & Glover, G. H. (2010). Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *NeuroImage*, *50*(1), 81–98. https://doi.org/10.1016/j.neuroimage.2009.12.011

Chiarelli, A. M., Maclin, E. L., Fabiani, M., & Gratton, G. (2015). A kurtosis-based wavelet algorithm for motion artifact correction of fNIRS data. *NeuroImage*, *112*, 128–137. https://doi.org/10.1016/J.NEUROIMAGE.2015.02.057

Christova, P., Lewis, S. M., Jerde, T. A., Lynch, J. K., & Georgopoulos, A. P. (2011). True associations between resting fMRI time series based on innovations. *Journal of Neural Engineering*, *8*(4), 046025. https://doi.org/10.1088/1741-2560/8/4/046025

Chuang, C.-C., & Sun, C.-W. (2014). Gender-related effects of prefrontal cortex connectivity: a resting-state functional optical tomography study. *Biomedical Optics Express*, *5*(8), 2503. https://doi.org/10.1364/boe.5.002503

Cooper, R. J., Selb, J., Gagnon, L., Phillip, D., Schytz, H. W., Iversen, H. K., Ashina, M., & Boas, D. A. (2012). A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. *Frontiers in Neuroscience*, *0*(OCT), 147. https://doi.org/10.3389/FNINS.2012.00147/BIBTEX

Cui, X., Bray, S., & Reiss, A. L. (2010). Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *NeuroImage*, *49*(4), 3039–3046. https://doi.org/10.1016/j.neuroimage.2009.11.050

Cui, X., Bryant, D. M., & Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation.

NeuroImage, *59*(3), 2430–2437. https://doi.org/10.1016/j.neuroimage.2011.09.003

Davey, C. E., Grayden, D. B., Egan, G. F., & Johnston, L. A. (2013). Filtering induces correlation in fMRI resting state data. *NeuroImage*, *64*(1), 728–740. https://doi.org/10.1016/j.neuroimage.2012.08.022

Delpy, D. T., & Cope, M. (1997). Quantification in tissue near-infrared spectroscopy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*(1354), 649–659. https://doi.org/10.1098/rstb.1997.0046

Di Lorenzo, R., Pirazzoli, L., Blasi, A., Bulgarelli, C., Hakuno, Y., Minagawa, Y., & Brigadoi, S. (2019). Recommendations for motion correction of infant fNIRS data applicable to multiple data sets and acquisition systems. *NeuroImage*, *200*, 511–527. https://doi.org/10.1016/J.NEUROIMAGE.2019.06.056

Duan, L., Zhang, Y. J., & Zhu, C. Z. (2012). Quantitative comparison of resting-state functional connectivity derived from fNIRS and fMRI: A simultaneous recording study. *NeuroImage*, *60*(4), 2008–2018. https://doi.org/10.1016/j.neuroimage.2012.02.014

Duan, L., Zhao, Z., Lin, Y., Wu, X., Luo, Y., & Xu, P. (2018). Wavelet-based method for removing global physiological noise in functional near-infrared spectroscopy. *Biomedical Optics Express*, *9*(8), 3805. https://doi.org/10.1364/boe.9.003805

Ferradal, S. L., Liao, S. M., Eggebrecht, A. T., Shimony, J. S., Inder, T. E., Culver, J. P., & Smyser, C. D. (2016). Functional Imaging of the Developing Brain at the Bedside Using Diffuse Optical Tomography. *Cerebral Cortex*, *26*(4), 1558–1568. https://doi.org/10.1093/CERCOR/BHU320

Fiecas, M., Cribben, I., Bahktiari, R., & Cummine, J. (2017). A variance components model for statistical inference on functional connectivity networks. *NeuroImage*, *149*, 256–266. https://doi.org/10.1016/J.NEUROIMAGE.2017.01.051

Fishburn, F. A., Ludlum, R. S., Vaidya, C. J., & Medvedev, A. V. (2019). Temporal Derivative Distribution Repair (TDDR): A motion correction method for fNIRS. *NeuroImage*, *184*, 171–179. https://doi.org/10.1016/j.neuroimage.2018.09.025

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Fuchino, Y., Naoi, N., Shibata, M., Niwa, F., Kawai, M., Konishi, Y., Okanoya, K., & Myowa-Yamakoshi, M. (2013). Effects of Preterm Birth on Intrinsic Fluctuations in Neonatal Cerebral Activity Examined Using Optical Imaging. *PLoS ONE*, *8*(6), e67432. https://doi.org/10.1371/journal.pone.0067432

Gagnon, L., Cooper, R. J., Yücel, M. A., Perdue, K. L., Greve, D. N., & Boas, D. A. (2012). Short separation channel location impacts the performance of short channel regression in NIRS. *NeuroImage*, *59*(3), 2518–2528. https://doi.org/10.1016/j.neuroimage.2011.08.095

Gallagher, A., Tremblay, J., & Vannasing, P. (2016). Language mapping in children using resting-state functional connectivity: comparison with a task-based approach. *Journal of Biomedical Optics*, *21*(12), 125006. https://doi.org/10.1117/1.jbo.21.12.125006

Gaudet, I., Hüsser, A., Vannasing, P., & Gallagher, A. (2020). Functional Brain

Connectivity of Language Functions in Children Revealed by EEG and MEG: A Systematic Review. *Frontiers in Human Neuroscience*, *14*, 62. https://doi.org/10.3389/fnhum.2020.00062

Geng, S., Liu, X., Biswal, B. B., & Niu, H. (2017). Effect of resting-state fNIRS scanning duration on functional brain connectivity and graph theory metrics of brain network. *Frontiers in Neuroscience*, *11*(JUL), 392. https://doi.org/10.3389/FNINS.2017.00392/BIBTEX

Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, *2*(2), 111–120. https://doi.org/10.1016/0304-4076(74)90034-7

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, *11*(5/6), 561–566. https://doi.org/10.5194/npg-11-561-2004

Gu, Y., Han, J., Liang, Z., Yan, J., Li, Z., & Li, X. (2016). Empirical mode decomposition-based motion artifact correction method for functional near-infrared spectroscopy. *Https://Doi.Org/10.1117/1.JBO.21.1.015002*, *21*(1), 015002. https://doi.org/10.1117/1.JBO.21.1.015002

Han, Q., Zhang, M., Li, W., Gao, Y., Xin, Q., Wang, Y., & Li, Z. (2014). Wavelet coherence analysis of prefrontal tissue oxyhaemoglobin signals as measured using near-infrared spectroscopy in elderly subjects with cerebral infarction. *Microvascular Research*, *95*(1), 108–115. https://doi.org/10.1016/J.MVR.2014.08.001

Hoaglin, D. C. (David C., Mosteller, F., & Tukey, J. W. (John W. (1983). *Understanding robust and exploratory data analysis*. 447.

Hocke, L. M., Oni, I. K., Duszynski, C. C., Corrigan, A. V., Frederick, B. de B., & Dunn, J. F. (2018). Automated Processing of fNIRS Data—A Visual Guide to the Pitfalls and Consequences. *Algorithms 2018, Vol. 11, Page 67*, *11*(5), 67. https://doi.org/10.3390/A11050067

Homae, F., Watanabe, H., Otobe, T., Nakano, T., Go, T., Konishi, Y., & Taga, G. (2010). Development of global cortical networks in early infancy. *Journal of Neuroscience*, *30*(14), 4877–4882. https://doi.org/10.1523/JNEUROSCI.5618-09.2010

Honari, H., Choe, A. S., Pekar, J. J., & Lindquist, M. A. (2019). Investigating the impact of autocorrelation on time-varying connectivity. *NeuroImage*, *197*, 37–48. https://doi.org/10.1016/J.NEUROIMAGE.2019.04.042

Hou, X., Zhang, Z., Zhao, C., Duan, L., Gong, Y., Li, Z., & Zhu, C. (2021). NIRS-KIT: a MATLAB toolbox for both resting-state and task fNIRS data analysis. *Neurophotonics*, *8*(01), 010802. https://doi.org/10.1117/1.nph.8.1.010802

Hu, Z., Liu, G., Dong, Q., & Niu, H. (2020). Applications of Resting-State fNIRS in the Developing Brain: A Review From the Connectome Perspective. *Frontiers in Neuroscience*, *14*, 476. https://doi.org/10.3389/fnins.2020.00476

Hull, J. V., Jacokes, Z. J., Torgerson, C. M., Irimia, A., Van Horn, J. D., Aylward, E., Bernier, R., Bookheimer, S., Dapretto, M., Gaab, N., Geschwind, D., Jack, A., Nelson, C., Pelphrey, K., State, M., Ventola, P., & Webb, S. J. (2017). Resting-state functional connectivity in autism spectrum disorders: A review. In *Frontiers in Psychiatry* (Vol. 7, Issue JAN, p. 205). Frontiers Media S.A. https://doi.org/10.3389/fpsyt.2016.00205

Huppert, T. J. (2016). Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy. *Neurophotonics*, *3*(1), 010401. https://doi.org/10.1117/1.nph.3.1.010401

Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics*, *48*(10), D280–D298. https://doi.org/10.1364/AO.48.00D280

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, *29*(2), 368–382. https://doi.org/10.1016/j.neuroimage.2005.08.065

Izzetoglu, M., Chitrapu, P., Bunce, S., & Onaral, B. (2010). Motion artifact cancellation in NIR spectroscopy using discrete Kalman filtering. *BioMedical Engineering Online*, *9*(1), 1–10. https://doi.org/10.1186/1475-925X-9-16/TABLES/2

Jahani, S., Setarehdan, S. K., Boas, D. A., & Yücel, M. A. (2018). Motion artifact detection and correction in functional near-infrared spectroscopy: a new hybrid method based on spline interpolation method and Savitzky–Golay filtering. *Neurophotonics*, *5*(01), 1. https://doi.org/10.1117/1.nph.5.1.015003

James, O., Park, H., & Kim, S. G. (2019). Impact of sampling rate on statistical significance for single subject fMRI connectivity analysis. *Human Brain Mapping*, *40*(11), 3321–3337. https://doi.org/10.1002/HBM.24600

Joseph, D. K., Huppert, T. J., Franceschini, M. A., & Boas, D. A. (2006). Diffuse optical tomography system to image brain activation with improved spatial resolution and validation with functional magnetic resonance imaging. *Applied Optics, Vol. 45, Issue 31, Pp. 8142-8151*, *45*(31), 8142–8151. https://doi.org/10.1364/AO.45.008142

Kida, T., Tanaka, E., & Kakigi, R. (2016). Multi-dimensional dynamics of human electromagnetic brain activity. *Frontiers in Human Neuroscience*, *9*(JAN2016), 713. https://doi.org/10.3389/FNHUM.2015.00713/BIBTEX

Kirilina, E., Jelzow, A., Heine, A., Niessing, M., Wabnitz, H., Brühl, R., Ittermann, B., Jacobs, A. M., & Tachtsidis, I. (2012). The physiological origin of task-evoked systemic artefacts in functional near infrared spectroscopy. *NeuroImage*, *61*(1), 70–81. https://doi.org/10.1016/j.neuroimage.2012.02.074

Lanka, P., Bortfeld, H., Huppert T.J. (2022) Correction of global physiology in resting-state functional near-infrared spectroscopy. *Neurophotonics, in press.*

Lee, M. H., Smyser, C. D., & Shimony, J. S. (2013). Resting-state fMRI: A review of methods and clinical applications. *American Journal of Neuroradiology*, *34*(10), 1866–1872. https://doi.org/10.3174/ajnr.A3263

Li, J., Qiu, L., Xu, L., Pedapati, E. V., Erickson, C. A., & Sunar, U. (2016). Characterization of autism spectrum disorder with spontaneous hemodynamic activity. *Biomedical Optics Express*, *7*(10), 3871. https://doi.org/10.1364/boe.7.003871

Li, Z., Liu, H., Liao, X., Xu, J., Liu, W., Tian, F., He, Y., & Niu, H. (2015). Dynamic functional connectivity revealed by resting-state functional near-infrared spectroscopy. *Biomedical Optics Express*, *6*(7), 2337. https://doi.org/10.1364/boe.6.002337

Liu, N., Mok, C., Witt, E. E., Pradhan, A. H., Chen, J. E., & Reiss, A. L. (2016). Nirs-

based hyperscanning reveals inter-brain neural synchronization during cooperative jenga game with face-to-face communication. *Frontiers in Human Neuroscience*, *10*(MAR2016). https://doi.org/10.3389/fnhum.2016.00082

Liu, Z., Zhang, M., Xu, G., Huo, C., Tan, Q., Li, Z., & Yuan, Q. (2017). Effective Connectivity Analysis of the Brain Network in Drivers during Actual Driving Using Near-Infrared Spectroscopy. *Frontiers in Behavioral Neuroscience*, *11*, 211. https://doi.org/10.3389/fnbeh.2017.00211

Lu, C. M., Zhang, Y. J., Biswal, B. B., Zang, Y. F., Peng, D. L., & Zhu, C. Z. (2010). Use of fNIRS to assess resting state functional connectivity. *Journal of Neuroscience Methods*, *186*(2), 242–249. https://doi.org/10.1016/j.jneumeth.2009.11.010

Medvedev, A. V. (2014). Does the resting state connectivity have hemispheric asymmetry? A near-infrared spectroscopy study. *NeuroImage*, *85*, 400–407. https://doi.org/10.1016/j.neuroimage.2013.05.092

Mesquita, R. C., Franceschini, M. A., & Boas, D. A. (2010). Resting state functional connectivity of the whole head with near-infrared spectroscopy. *Biomedical Optics Express*, *1*(1), 324. https://doi.org/10.1364/boe.1.000324

Molavi, B., & Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological Measurement*, *33*(2), 259–270. https://doi.org/10.1088/0967-3334/33/2/259

Molavi, B., May, L., Gervain, J., Carreiras, M., Werker, J. F., & Dumont, G. A. (2014). Analyzing the resting state functional connectivity in the human language system using near infrared spectroscopy. *Frontiers in Human Neuroscience*, *7*(JAN), 921. https://doi.org/10.3389/FNHUM.2013.00921/BIBTEX

Niu, H., & He, Y. (2014). Resting-state functional brain connectivity: Lessons from functional near-infrared spectroscopy. *Neuroscientist*, *20*(2), 173–188. https://doi.org/10.1177/1073858413502707

Niu, H., Li, Z., Liao, X., Wang, J., Zhao, T., Shu, N., Zhao, X., & He, Y. (2013). Test-Retest Reliability of Graph Metrics in Functional Brain Networks: A Resting-State fNIRS Study. *PLoS ONE*, *8*(9), e72425. https://doi.org/10.1371/journal.pone.0072425

Niu, H., Wang, J., Zhao, T., Shu, N., & He, Y. (2012). Revealing Topological Organization of Human Brain Functional Networks with Resting-State Functional near Infrared Spectroscopy. *PLoS ONE*, *7*(9), e45771. https://doi.org/10.1371/journal.pone.0045771

Niu, H., Zhu, Z., Wang, M., Li, X., Yuan, Z., Sun, Y., & Han, Y. (2019). Abnormal dynamic functional connectivity and brain states in Alzheimer's diseases: functional near-infrared spectroscopy study. *Neurophotonics*, *6*(02), 1. https://doi.org/10.1117/1.nph.6.2.025010

Noah, J. A., Zhang, X. Z., Dravida, S., DiCocco, C., Suzuki, T., Aslin, R. N., Tachtsidis, I., & Hirsch, J. (2021). Comparison of short-channel separation and spatial domain filtering for removal of non-neural components in functional near-infrared spectroscopy signals. *Https://Doi.Org/10.1117/1.NPh.8.1.015004*, *8*(1), 015004. https://doi.org/10.1117/1.NPH.8.1.015004

Novi, S. L., Roberts, E., Spagnuolo, D., Spilsbury, B. M., Price, D. C., Imbalzano, C. A., Forero, E., Yodh, A. G., Tellis, G. M., Tellis, C. M., & Mesquita, R. C. (2020).

Functional near-infrared spectroscopy for speech protocols: characterization of motion artifacts and guidelines for improving data analysis. *Neurophotonics*, *7*(01), 1. https://doi.org/10.1117/1.nph.7.1.015001

Novi, S. L., Rodrigues, R. B. M. L., & Mesquita, R. C. (2016). Resting state connectivity patterns with near-infrared spectroscopy data of the whole head. *Biomedical Optics Express*, *7*(7), 2524. https://doi.org/10.1364/boe.7.002524

Obrig, H., Neufang, M., Wenzel, R., Kohl, M., Steinbrink, J., Einhäupl, K., & Villringer, A. (2000). Spontaneous low frequency oscillations of cerebral hemodynamics and metabolism in human adults. *NeuroImage*, *12*(6), 623–639. https://doi.org/10.1006/nimg.2000.0657

Oldham, S., & Fornito, A. (2019). The development of brain network hubs. In *Developmental Cognitive Neuroscience* (Vol. 36, p. 100607). Elsevier Ltd. https://doi.org/10.1016/j.dcn.2018.12.005

Paranawithana, I., Mao, D., Wong, Y. T., & McKay, C. M. (2022). Reducing false discoveries in resting-state functional connectivity using short channel correction: an fNIRS study. *Neurophotonics*, *9*(01), 015001. https://doi.org/10.1117/1.NPh.9.1.015001

Perlman, S. B., Luna, B., Hein, T. C., & Huppert, T. J. (2014). fNIRS evidence of prefrontal regulation of frustration in early childhood. *NeuroImage*, *85*, 326–334. https://doi.org/10.1016/J.NEUROIMAGE.2013.04.057

Pfeifer, M. D., Scholkmann, F., & Labruyère, R. (2018). Signal processing in functional near-infrared spectroscopy (fNIRS): Methodological differences lead to different statistical results. *Frontiers in Human Neuroscience*, *11*, 641. https://doi.org/10.3389/fnhum.2017.00641

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., & Tachtsidis, I. (2019). Current Status and Issues Regarding Pre-processing of fNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework. *Frontiers in Human Neuroscience*, *12*(January), 1–21. https://doi.org/10.3389/fnhum.2018.00505

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (Fnirs) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1464*(1), 5–29. https://doi.org/10.1111/nyas.13948

Poldrack, R. A. (2012). The future of fMRI in cognitive neuroscience. In *NeuroImage* (Vol. 62, Issue 2, pp. 1216–1220). Academic Press. https://doi.org/10.1016/j.neuroimage.2011.08.007

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Power, J. D., Schlaggar, B. L., & Petersen, S. E. (2015). Recent progress and outstanding issues in motion correction in resting state fMRI. *NeuroImage*, *105*, 536–551. https://doi.org/10.1016/j.neuroimage.2014.10.044

Price, T., Wee, C. Y., Gao, W., & Shen, D. (2014). Multiple-network classification of childhood autism using functional connectivity dynamics. *Lecture Notes in*

Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8675 LNCS(PART 3), 177–184. https://doi.org/10.1007/978-3-319-10443-0_23

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, *52*(3), 1059–1069. https://doi.org/10.1016/J.NEUROIMAGE.2009.10.003

Saager, R. B., & Berger, A. J. (2005). Direct characterization and removal of interfering absorption trends in two-layer turbid media. *Journal of the Optical Society of America A*, *22*(9), 1874. https://doi.org/10.1364/josaa.22.001874

Saager, R. B., Telleri, N. L., & Berger, A. J. (2011). Two-detector Corrected Near Infrared Spectroscopy (C-NIRS) detects hemodynamic activation responses more robustly than single-detector NIRS. *NeuroImage*, *55*(4), 1679–1685. https://doi.org/10.1016/j.neuroimage.2011.01.043

Sakakibara, E., Homae, F., Kawasaki, S., Nishimura, Y., Takizawa, R., Koike, S., Kinoshita, A., Sakurada, H., Yamagishi, M., Nishimura, F., Yoshikawa, A., Inai, A., Nishioka, M., Eriguchi, Y., Matsuoka, J., Satomura, Y., Okada, N., Kakiuchi, C., Araki, T., … Kasai, K. (2016). Detection of resting state functional connectivity using partial correlation analysis: A study using multi-distance and whole-head probe near-infrared spectroscopy. *NeuroImage*, *142*, 590–601. https://doi.org/10.1016/j.neuroimage.2016.08.011

Santosa, H., Aarabi, A., Perlman, S. B., & Huppert, T. J. (2017). Characterization and correction of the false-discovery rates in resting state connectivity using functional near-infrared spectroscopy. *Journal of Biomedical Optics*, *22*(5), 055002. https://doi.org/10.1117/1.jbo.22.5.055002

Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). The NIRS Brain AnalyzIR Toolbox. *Algorithms*, *11*(5), 73. https://doi.org/10.3390/a11050073

Santosa, H., Zhai, X., Fishburn, F., Sparto, P. J., & Huppert, T. J. (2020). Quantitative comparison of correction techniques for removing systemic physiological signal in functional near-infrared spectroscopy studies. *Neurophotonics*, *7*(03), 1–21. https://doi.org/10.1117/1.nph.7.3.035009

Sasai, S., Homae, F., Watanabe, H., Sasaki, A. T., Tanabe, H. C., Sadato, N., & Taga, G. (2012). A NIRS-fMRI study of resting state network. *NeuroImage*, *63*(1), 179–193. https://doi.org/10.1016/j.neuroimage.2012.06.011

Sato, T., Nambu, I., Takeda, K., Aihara, T., Yamashita, O., Isogaya, Y., Inoue, Y., Otaka, Y., Wada, Y., Kawato, M., Sato, M. aki, & Osu, R. (2016). Reduction of global interference of scalp-hemodynamics in functional near-infrared spectroscopy using short distance probes. *NeuroImage*, *141*, 120–132. https://doi.org/10.1016/j.neuroimage.2016.06.054

Satterthwaite, T. D., Wolf, D. H., Loughead, J., Ruparel, K., Elliott, M. A., Hakonarson, H., Gur, R. C., & Gur, R. E. (2012). Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *NeuroImage*, *60*(1), 623–632. https://doi.org/10.1016/j.neuroimage.2011.12.063

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared

spectroscopy and imaging instrumentation and methodology. *NeuroImage*, *85*, 6–27. https://doi.org/10.1016/j.neuroimage.2013.05.004

Scholkmann, F., Spichtig, S., Muehlemann, T., & Wolf, M. (2010). How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation. *Physiological Measurement*, *31*(5), 649–662. https://doi.org/10.1088/0967-3334/31/5/004

Seth, A. K., Barrett, A. B., & Barnett, L. (2015). Granger Causality Analysis in Neuroscience and Neuroimaging. *Journal of Neuroscience*, *35*(8), 3293–3297. https://doi.org/10.1523/jneurosci.4399-14.2015

Sherafati, A., Snyder, A. Z., Eggebrecht, A. T., Bergonzi, K. M., Burns-Yocum, T. M., Lugar, H. M., Ferradal, S. L., Robichaux-Viehoever, A., Smyser, C. D., Palanca, B. J., Hershey, T., & Culver, J. P. (2020). Global motion detection and censoring in high-density diffuse optical tomography. *Human Brain Mapping*, *41*(14), 4093–4112. https://doi.org/10.1002/HBM.25111

Shevlyakov, G., & Smirnov, P. (2011). Robust Estimation of the Correlation Coefficient: An Attempt of Survey. *Austrian Journal of Statistics*, *40*(1&2), 147–156–147–156. https://doi.org/10.17713/AJS.V40I1&2.206

Tachtsidis, I., Elwell, C. E., Leung, T. S., Lee, C. W., Smith, M., & Delpy, D. T. (2004). Investigation of cerebral haemodynamics by near-infrared spectroscopy in young healthy volunteers reveals posture-dependent spontaneous oscillations. *Physiological Measurement*, *25*(2), 437–445. https://doi.org/10.1088/0967-3334/25/2/003

Tachtsidis, I., Koh, P. H., Stubbs, C., & Elwell, C. E. (2010). Functional optical topography analysis using statistical parametric mapping (SPM) methodology with and without physiological confounds. *Advances in Experimental Medicine and Biology*, *662*, 237–243. https://doi.org/10.1007/978-1-4419-1241-1_34

Tachtsidis, I., & Scholkmann, F. (2016). False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward. *Neurophotonics*, *3*(3), 031405. https://doi.org/10.1117/1.nph.3.3.031405

Tan, Q., Zhang, M., Wang, Y., Zhang, M., Wang, B., Xin, Q., & Li, Z. (2016). Age-related alterations in phase synchronization of oxyhemoglobin concentration changes in prefrontal tissues as measured by near-infrared spectroscopy signals. *Microvascular Research*, *103*, 19–25. https://doi.org/10.1016/j.mvr.2015.10.002

Tan, Q., Zhang, M., Wang, Y., Zhang, M., Wang, Y., Xin, Q., Wang, B., & Li, Z. (2015). Frequency-specific functional connectivity revealed by wavelet-based coherence analysis in elderly subjects with cerebral infarction using NIRS method. *Medical Physics*, *42*(9), 5391–5403. https://doi.org/10.1118/1.4928672

Tremblay, J., Martínez-Montes, E., Hüsser, A., Caron-Desrochers, L., Lepage, C., Pouliot, P., Vannasing, P., & Gallagher, A. (2022). LIONirs: flexible Matlab toolbox for fNIRS data analysis. *Journal of Neuroscience Methods*, *370*, 109487. https://doi.org/10.1016/J.JNEUMETH.2022.109487

van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. In *European Neuropsychopharmacology* (Vol. 20, Issue 8, pp. 519–534). Elsevier. https://doi.org/10.1016/j.euroneuro.2010.03.008

van Diessen, E., Numan, T., van Dellen, E., van der Kooi, A. W., Boersma, M., Hofman, D., van Lutterveld, R., van Dijk, B. W., van Straaten, E. C. W., Hillebrand, A., & Stam, C. J. (2015). Opportunities and methodological challenges in EEG and MEG resting state functional brain network research. In *Clinical Neurophysiology* (Vol. 126, Issue 8, pp. 1468–1481). Elsevier Ireland Ltd. https://doi.org/10.1016/j.clinph.2014.11.018

Vértes, P. E., & Bullmore, E. T. (2015). Annual Research Review: Growth connectomics – the organization and reorganization of brain networks during normal and abnormal development. *Journal of Child Psychology and Psychiatry*, *56*(3), 299–320. https://doi.org/10.1111/jcpp.12365

Virtanen, J., Kotilahti, K. M., Ilmoniemi, R., Noponen, T. E. J., & Virtanen, J. (2011). Accelerometer-based method for correcting signal baseline changes caused by motion artifacts in medical near-infrared spectroscopy. *Https://Doi.Org/10.1117/1.3606576*, *16*(8), 087005. https://doi.org/10.1117/1.3606576

Virtanen, J., Noponen, T. E. J., & Meriläinen, P. (2009). Comparison of principal and independent component analysis in removing extracerebral interference from near-infrared spectroscopy signals. *Https://Doi.Org/10.1117/1.3253323*, *14*(5), 054032. https://doi.org/10.1117/1.3253323

Wan, N., Hancock, A. S., Moon, T. K., & Gillam, R. B. (2018). A functional near-infrared spectroscopic investigation of speech production during reading. *Human Brain Mapping*, *39*(3), 1428–1437. https://doi.org/10.1002/hbm.23932

Wang, J., Dong, Q., & Niu, H. (2017). The minimum resting-state fNIRS imaging duration for accurate and stable mapping of brain connectivity network in children. *Scientific Reports 2017 7:1*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-06340-7

Wang, J., Wang, X., Xia, M., Liao, X., Evans, A., & He, Y. (2015). GRETNA: A graph theoretical network analysis toolbox for imaging connectomics. *Frontiers in Human Neuroscience*, *9*(JUNE), 386. https://doi.org/10.3389/fnhum.2015.00386

White, B. R., Snyder, A. Z., Cohen, A. L., Petersen, S. E., Raichle, M. E., Schlaggar, B. L., & Culver, J. P. (2009). Resting-state functional connectivity in the human brain revealed with diffuse optical tomography. *NeuroImage*, *47*(1), 148–156. https://doi.org/10.1016/j.neuroimage.2009.03.058

Wolf, M., Wolf, U., Toronov, V., Michalos, A., Paunescu, L. A., Choi, J. H., & Gratton, E. (2002). Different time evolution of oxyhemoglobin and deoxyhemoglobin concentration changes in the visual and motor cortices during functional stimulation: A near-infrared spectroscopy study. *NeuroImage*, *16*(3 I), 704–712. https://doi.org/10.1006/nimg.2002.1128

Wyser, D., Mattille, M., Wolf, M., Lambercy, O., Scholkmann, F., & Gassert, R. (2020). Short-channel regression in functional near-infrared spectroscopy is more effective when considering heterogeneous scalp hemodynamics. *Neurophotonics*, *7*(03). https://doi.org/10.1117/1.nph.7.3.035011

Xu, J., Liu, X., Zhang, J., Li, Z., Wang, X., Fang, F., & Niu, H. (2015). FC-NIRS: A Functional Connectivity Analysis Tool for Near-Infrared Spectroscopy Data. *BioMed Research International*, *2015*. https://doi.org/10.1155/2015/248724

Yamada, T., Umeyama, S., & Matsuda, K. (2009). Multidistance probe arrangement to

eliminate artifacts in functional near-infrared spectroscopy. *Journal of Biomedical Optics*, *14*(6), 064034. https://doi.org/10.1117/1.3275469

Yamada, T., Umeyama, S., & Matsuda, K. (2012). Separation of fNIRS Signals into Functional and Systemic Components Based on Differences in Hemodynamic Modalities. *PLoS ONE*, *7*(11), e50271. https://doi.org/10.1371/journal.pone.0050271

Yücel, M. A., Selb, J., Aasted, C. M., Lin, P.-Y., Borsook, D., Becerra, L., & Boas, D. A. (2016). Mayer waves reduce the accuracy of estimated hemodynamic response functions in functional near-infrared spectroscopy. *Biomedical Optics Express*, *7*(8), 3078. https://doi.org/10.1364/boe.7.003078

Yücel, M. A., Selb, J., Cooper, R. J., & Boas, D. A. (2014). Targeted principle component analysis: A new motion artifact correction approach for near-infrared spectroscopy. *Http://Dx.Doi.Org/10.1142/S1793545813500661*, *7*(2), 1350066. https://doi.org/10.1142/S1793545813500661

Zeff, B. W., White, B. R., Dehghani, H., Schlaggar, B. L., & Culver, J. P. (2007). Retinotopic mapping of adult human visual cortex with high-density diffuse optical tomography. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(29), 12169–12174. https://doi.org/10.1073/PNAS.0611266104

Zhang, F., Cheong, D., Khan, A. F., Chen, Y., Ding, L., & Yuan, H. (2021). Correcting physiological noise in whole-head functional near-infrared spectroscopy. *Journal of Neuroscience Methods*, *360*, 109262. https://doi.org/10.1016/J.JNEUMETH.2021.109262

Zhang, H., Duan, L., Zhang, Y. J., Lu, C. M., Liu, H., & Zhu, C. Z. (2011). Test-retest assessment of independent component analysis-derived resting-state functional connectivity based on functional near-infrared spectroscopy. *NeuroImage*, *55*(2), 607–615. https://doi.org/10.1016/j.neuroimage.2010.12.007

Zhang, H., Zhang, Y.-J., Duan, L., Ma, S.-Y., Lu, C.-M., & Zhu, C.-Z. (2011). Is resting-state functional connectivity revealed by functional near-infrared spectroscopy test-retest reliable? *Journal of Biomedical Optics*, *16*(6), 067008. https://doi.org/10.1117/1.3591020

Zhang, H., Zhang, Y. J., Lu, C. M., Ma, S. Y., Zang, Y. F., & Zhu, C. Z. (2010). Functional connectivity as revealed by independent component analysis of resting-state fNIRS measurements. *NeuroImage*, *51*(3), 1150–1161. https://doi.org/10.1016/J.NEUROIMAGE.2010.02.080

Zhang, Y.-J. (2010). Detecting resting-state functional connectivity in the language system using functional near-infrared spectroscopy. *Journal of Biomedical Optics*, *15*(4), 047003. https://doi.org/10.1117/1.3462973

Zhang, Y., Brooks, D. H., Franceschini, M. A., & Boas, D. A. (2005). Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging. *Journal of Biomedical Optics*, *10*(1), 011014. https://doi.org/10.1117/1.1852552

Zhou, G., Liu, J., Ding, X. P., Fu, G., & Lee, K. (2016). Development of effective connectivity during own- and other-race face processing: A granger causality analysis. *Frontiers in Human Neuroscience*, *10*(SEP2016), 474. https://doi.org/10.3389/fnhum.2016.00474

Zhou, X., Sobczak, G., Colette, M. M., & Litovsky, R. Y. (2021). Comparing fNIRS signal qualities between approaches with and without short channels. *PLoS ONE*, *15*(12 December), 1–18. https://doi.org/10.1371/journal.pone.0244186

Zhu, H., Fan, Y., Guo, H., Huang, D., & He, S. (2014). Reduced interhemispheric functional connectivity of children with autism spectrum disorder: evidence from functional near infrared spectroscopy studies. *Biomedical Optics Express*, *5*(4), 1262. https://doi.org/10.1364/boe.5.001262

Zhu, H., Xu, J., Li, J., Peng, H., Cai, T., Li, X., Wu, S., Cao, W., & He, S. (2017). Decreased functional connectivity and disrupted neural network in the prefrontal cortex of affective disorders: A resting-state fNIRS study. *Journal of Affective Disorders*, *221*, 132–144. https://doi.org/10.1016/j.jad.2017.06.024