

BOSTON UNIVERSITY
SCHOOL OF MEDICINE

Dissertation

**BIOMARKER DISCOVERY AND STATISTICAL MODELING WITH
APPLICATIONS IN CHILDHOOD EPILEPSY AND ANGELMAN SYNDROME**

by

ELIZABETH ROSE STEVENS SPENCER

B.S., University of Maryland, College Park, 2015

B.A., University of Maryland, College Park, 2015

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2021

© 2021 by
ELIZABETH ROSE STEVENS SPENCER
All rights reserved except
Chapter 2 © 2018 Journal of Neuroscience Methods

Approved by

First Reader

Mark A. Kramer, Ph.D.

Professor of Mathematics and Statistics

Boston University

Second Reader

Catherine J. Chu, M.D., M.A., M.M.S.C.

Associate Professor of Neurology and Neurologist

Harvard Medical School and Massachusetts General Hospital

Third Reader

Uri T. Eden, Ph.D.

Professor of Mathematics and Statistics

Boston University

ACKNOWLEDGEMENTS

I would like to thank everyone who has helped me during this process. I collaborated with many people during my time at BU and I could not have done this dissertation without everyone's support.

I have worked with amazing mentors, each of whom shaped my dissertation. Mark Kramer, who was always happy to go down any mathematical wormhole, yet simultaneously always kept me focused and on task. Cat Chu, who kept the science exciting and always inspired me to keep pushing on what might feel like an insurmountable problem. Uri Eden, too, who always seemed to have a solution to any problem I had.

I want to thank everyone in GPN. Those in my cohort: Sam Levy, Sam Shelton, Byron Price, Madhura Baxi, and Marek Kowalski, who in year one, helped teach me basic biology (back when I didn't know what a neuron was) and with whom I spent countless hours in the library. My friends in the lab (past and present): Emily Schlafly, Eric Denovellis, and Manu Martinet, who gave me invaluable feedback on presentations and projects.

I also want to give special thanks to Sandi Grasso and Shelley Russek for creating such a warm and supportive community (and taking a risk on accepting a math student into a neuroscience program).

Finally, I would like to thank my family and friends. My mom and sister who have been lifelong pillars of support and pep-talk providers. Thank you to Aunt Kathleen, Uncle Joe, Eleanor, Dan, Becky, and Molly for helping me maintain balance in my life.

**BIOMARKER DISCOVERY AND STATISTICAL MODELING WITH
APPLICATIONS IN CHILDHOOD EPILEPSY AND ANGELMAN SYNDROME**

ELIZABETH ROSE STEVENS SPENCER

Boston University School of Medicine, 2021

Major Professor: Mark Kramer, Ph.D., Professor of Mathematics & Statistics

ABSTRACT

Biomarker discovery and statistical modeling reveals the brain activity that supports brain function and dysfunction. Detecting abnormal brain activity is critical for developing biomarkers of disease, elucidating disease mechanisms and evolution, and ultimately improving disease course. In my thesis, we develop statistical methodology to characterize neural activity in disease from noisy electrophysiological recordings.

First, we develop a modification of a classic statistical modeling approach - multivariate Granger causality - to infer coordinated activity between brain regions. Assuming the signaling dependencies vary smoothly, we propose to write the history terms in autoregressive models of the signals using a lower dimensional spline basis. This procedure requires fewer parameters than the standard approach, thus increasing the statistical power. we show that this procedure accurately estimates brain dynamics in simulations and examples of physiological recordings from a patient with pharmaco-resistant epilepsy. This work provides a statistical framework to understand alternations in coordinated brain activity in disease.

Second, we demonstrate that sleep spindles, thalamically-driven neural rhythms (9-15 Hz) associated with sleep-dependent learning, are a reliable biomarker for Rolandic

epilepsy. Rolandic epilepsy is the most common form of childhood epilepsy and characterized by nocturnal focal epileptic discharges as well as neurocognitive deficits. We show that sleep spindle rate is reduced regionally across cortex and correlated with poor cognitive performance in epilepsy. These results provide evidence for a regional disruption to the thalamocortical circuit in Rolandic epilepsy, and a potential mechanistic explanation for the cognitive deficits observed.

Finally, we develop a procedure to utilize delta rhythms (2-4 Hz), a sensitive biomarker for Angelman syndrome, as a non-invasive measure of treatment efficacy in clinical trials. Angelman syndrome is a rare neurodevelopmental disorder caused by reduced expression of the UBE3A protein. Many disease-modifying treatments are being developed to reinstate UBE3A expression. To aid in clinical trials, we propose a procedure that detects therapeutic improvements in delta power outside of the natural variability over age by developing a longitudinal natural history model of delta power.

These results demonstrate the utility of biomarker discovery and statistical modeling for elucidating disease course and mechanisms with the long-term goal of improving patient outcomes.

TABLE OF CONTENTS

| | |
|--|------|
| ACKNOWLEDGEMENTS | iv |
| ABSTRACT..... | v |
| TABLE OF CONTENTS..... | vii |
| LIST OF TABLES | xii |
| LIST OF FIGURES | xiii |
| LIST OF ABBREVIATIONS..... | xv |
| 1 CHAPTER ONE Introduction | 1 |
| 1.1 Electrophysiological Recordings | 2 |
| 1.1.1 Background..... | 2 |
| 1.1.2 Relevance for Disease..... | 5 |
| 1.2 Mathematical Approaches to Analyze Electrophysiological Recordings..... | 10 |
| 1.2.1 Linking data to scientific hypotheses: statistical models..... | 10 |
| 1.2.2 Time series analysis | 19 |
| 2 CHAPTER TWO A procedure to increase the power of Granger-causal analysis through temporal smoothing | 29 |
| 2.1 Introduction..... | 30 |
| 2.2 Methods..... | 35 |

| | | |
|--------|--|-----------|
| 2.2.1 | Implementation of the standard AR model and standard-Granger method | 35 |
| 2.2.2 | Implementation of the spline AR model and the spline-Granger method | 39 |
| 2.2.3 | Goodness-of-fit: Grenander and Rosenblatt test..... | 43 |
| 2.2.4 | Goodness-of-fit: Durbin-Watson test..... | 44 |
| 2.2.5 | Generation of synthetic signals: single-node simulations..... | 44 |
| 2.2.6 | Generation of synthetic signals: nine-node simulations | 45 |
| 2.2.7 | Calculation of network accuracy | 47 |
| 2.2.8 | Calculation of computation time..... | 48 |
| 2.2.9 | In vivo recordings from a human subject | 48 |
| 2.2.10 | Simulation and model fitting code..... | 49 |
| 2.3 | Results..... | 49 |
| 2.3.1 | Simulation: Single Node..... | 50 |
| 2.3.2 | Simulation: Nine-Node Network | 57 |
| 2.3.3 | In vivo Data: Single Node..... | 63 |
| 2.3.4 | In vivo Data: Multiple Nodes | 65 |
| 2.4 | Discussion..... | 68 |
| 2.5 | Acknowledgements..... | 74 |
| 3 | CHAPTER THREE Source estimates reveal regional spindle deficits and predict cognitive symptoms in Rolandic epilepsy | 75 |
| 3.1 | Introduction..... | 77 |
| 3.2 | Materials and methods | 79 |

| | | |
|-------|---|-----|
| 3.2.1 | Subject data..... | 79 |
| 3.2.2 | Electrical source imaging and minimum norm estimation | 80 |
| 3.2.3 | Artifact and epileptic spike removal procedures | 83 |
| 3.2.4 | Neuropsychological assessment..... | 83 |
| 3.2.5 | Spectral analysis..... | 85 |
| 3.2.6 | Automated spindle detection..... | 86 |
| 3.2.7 | Assessment of spindle features | 87 |
| 3.2.8 | Statistical analysis..... | 89 |
| 3.2.9 | Data availability | 93 |
| 3.3 | Results..... | 93 |
| 3.3.1 | Spindle rate in the inferior Rolandic cortex is reduced in Rolandic epilepsy compared to control subjects. | 93 |
| 3.3.2 | Spindle features are typical in subjects with Rolandic epilepsy but are less bilaterally synchronous. | 96 |
| 3.3.3 | Spindle deficit extends beyond inferior Rolandic cortices. | 97 |
| 3.3.4 | Regional estimates of spindle rate improve predictions of cognitive function better than focal estimates | 100 |
| 3.4 | Discussion..... | 103 |
| 3.5 | Acknowledgements..... | 107 |
| 3.6 | Funding | 107 |
| 3.7 | Competing interests | 107 |
| 3.8 | Supplementary material | 107 |

| | | |
|--------|--|-----|
| 4 | CHAPTER FOUR Deviations from the natural history of delta power in Angelman syndrome reflect treatment effect size and correlate with UBE3A expression | 109 |
| 4.1 | Introduction..... | 110 |
| 4.2 | Materials and Methods..... | 112 |
| 4.2.1 | Human subject data collection | 112 |
| 4.2.2 | Human EEG data analysis | 114 |
| 4.2.3 | Animals | 115 |
| 4.2.4 | Oligonucleotides | 116 |
| 4.2.5 | ASO in vivo administration | 116 |
| 4.2.6 | SDS-PAGE Western Blotting..... | 117 |
| 4.2.7 | Quantitative Reverse Transcription Polymerase Chain Reaction | 118 |
| 4.2.8 | Surgeries and Local Field Potential (LFP) Recordings | 118 |
| 4.2.9 | Mouse LFP data analysis | 119 |
| 4.2.10 | Statistical analysis for Ube3a mRNA upregulation and model residuals | 121 |
| 4.3 | Results..... | 122 |
| 4.3.1 | Delta power estimates are stable over a 24-hour period within an individual | 122 |
| 4.3.2 | Delta power at a future visit can be reliably predicted from a longitudinal natural history model | 123 |
| 4.3.3 | Power to detect a deviation from the natural history model following treatment.. | 125 |

| | | |
|-------|--|-----|
| 4.3.4 | The natural history model of delta activity detects deviations in delta power in ASO-treated AS mice | 126 |
| 4.3.5 | Unsilencing of the Ube3a paternal allele with a Ube3a-ATS ASO in AS mice..... | 129 |
| 4.3.6 | Increased model residuals correlate with increased Ube3a mRNA production | 130 |
| 4.4 | Discussion..... | 132 |
| 4.4.1 | Competing interests | 135 |
| 5 | CHAPTER FIVE Conclusion..... | 136 |
| 5.1 | Innovation and Impact | 136 |
| 5.1.1 | Disease inspired methodology | 136 |
| 5.1.2 | Pathways to understanding disease..... | 140 |
| 5.2 | Future Directions | 142 |
| | BIBLIOGRAPHY..... | 145 |
| | CURRICULUM VITAE..... | 174 |

LIST OF TABLES

| | |
|---|-----|
| Table 1: Percent reduction in spindle rate by being in the active group relative to the control group. | 99 |
| Table 2: Model fits for each neuropsychological assessment using focal and regional spindle rate. | 103 |
| Table 3: Natural history model parameter estimates and p-values. | 124 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1.1: Simulated mixed effects model with correlation between the random intercept and slope. | 15 |
| Figure 1.2: Geometric interpretation of R^2 | 19 |
| Figure 1.3: Sharp edges in electrophysiological recordings induce increases in power at all frequencies. | 25 |
| Figure 1.4: Cartoon figure of functional brain network. | 28 |
| Figure 2.1: Example approximation of history dependence with Cardinal spline basis functions. | 42 |
| Figure 2.2: Illustration of model coefficients for network simulations. | 47 |
| Figure 2.3: Both modeling approaches perform well for a single node with a dominant high frequency spectral peak and long, smooth history dependence. | 56 |
| Figure 2.4: The spline-Granger method is not equivalent to filtering and downsampling the signal. | 57 |
| Figure 2.5: Increases in estimated model order produce accurate functional networks in minimal computation time using the spline-Granger method. | 62 |
| Figure 2.6: The spline AR and standard AR models produce similar results for in vivo data recorded from a single node. | 65 |
| Figure 2.7: The spline-Granger method makes inference of a 26-node network computationally tractable and more sensitive. | 67 |
| Figure 3.1: Example of electrical source imaging procedure. | 82 |
| Figure 3.2: Spindle rate is lower in subjects with Rolandic epilepsy. | 95 |

| | |
|---|-----|
| Figure 3.3: Spindle rate is a more sensitive measure than other spectral estimates. | 95 |
| Figure 3.4: Rolandic epilepsy subjects produce fewer, but healthy, spindles. | 97 |
| Figure 3.5: Spindle rate deficit extends beyond inferior Rolandic cortices..... | 98 |
| Figure 3.6: Spindle rate correlates with neuropsychological assessments. | 102 |
| Figure 4.1: Only a few minutes of EEG data are needed to estimate delta power with high precision. | 123 |
| Figure 4.2: Natural history model overview. | 124 |
| Figure 4.3: Simulation of model implementation. | 126 |
| Figure 4.4: Longitudinal mouse LFP data. | 128 |
| Figure 4.5: Application of natural history model to mouse model of AS..... | 129 |
| Figure 4.6: Unsilencing of the <i>Ube3a</i> paternal allele with a <i>Ube3a-ATS</i> ASO in AS mice leads to increase <i>Ube3a</i> mRNA and UBE3A protein and correlation to model residuals. | 131 |

LIST OF ABBREVIATIONS

| | |
|---------------|---|
| AIC | Akaike information criterion |
| AR | Autoregressive |
| AS..... | Angelman syndrome |
| ASO..... | Antisense oligonucleotide |
| BIC..... | Bayesian information criterion |
| CSF..... | Cerebrospinal fluid |
| ECoG..... | Intracranial electroencephalogram |
| EEG..... | Scalp electroencephalogram |
| ESI..... | Electrical source imaging |
| FDR..... | False discovery rate |
| GLM..... | Generalized linear model |
| ICV..... | Intracerebroventricular |
| IED..... | Interictal epileptiform discharge |
| IVI | Inter-visit interval |
| LFP | Local field potential |
| MEMPRAGE..... | Multi-echo magnetization-prepared rapid acquisition gradient-echo |
| MGH..... | Massachusetts General Hospital |
| MLE..... | Maximum likelihood estimation |
| MRI..... | Magnetic resonance imaging |
| MVAR..... | Multivariate autoregressive |
| NREM..... | Non-rapid eye movement |

| | |
|------------------------|--|
| PV..... | Parvalbumin |
| QMLE | Quasi-maximum likelihood estimation |
| REM | Rapid-eye movement |
| REML..... | Restricted maximum likelihood |
| SEM..... | Standard error of the mean |
| SOM | Somatostatin |
| SSE..... | Sum of squares for error |
| SST..... | Sum of total squares |
| TN..... | True positive |
| TP..... | True negative |
| TRN..... | Thalamic reticular nucleus |
| <i>Ube3a-ATS</i> | <i>Ube3a</i> antisense transcript |
| WISC-V..... | Wechsler Intelligence Scale for Children, 5 th ed |
| WT..... | Wild-type |

1 CHAPTER ONE

Introduction

Translational and statistical neuroscience provides a rigorous framework to characterize and elucidate disease. Broadly, translational neuroscience research aims to develop understanding of disease mechanisms and evolution, and ultimately improve disease course (Engel Jr, 2011). Statistical neuroscience research furthers the aim of understanding brain function and dysfunction via the development of methodology to extract structure from noisy neural data (Kass et al., 2018; Kramer & Eden, 2016). Neural data, either the activity of individual or populations of neurons, provides information on how the brain perceives and processes the world. Statistical models can reveal these mechanisms by characterizing patterns in the brain, *e.g.*, how brain regions couple to perform functions, and uncertainty in those patterns (Kass, Eden, & Brown, 2014). Neurological disorders may disrupt these patterns in stereotyped ways revealing the neurophysiology of disease and providing potential therapeutic targets.

In this dissertation, I develop methodology to discover and leverage biomarkers from electrophysiological recordings of populations of neurons to elucidate neural mechanisms in disease and to measure therapeutic impact. In this introduction, I review the relevant neuroscience, medicine, and statistics background for this work. In Chapter 2, I present a method for which to characterize coupling between brain regions. In Chapter 3, I show how lack of normal sleep patterns can cause cognitive impairment in a common childhood epilepsy. Finally, in Chapter 4, I demonstrate how abnormal brain activity can be used to track disease progression due to efficacious treatment.

1.1 Electrophysiological Recordings

1.1.1 Background

1.1.1.1 Electroencephalogram

The electroencephalogram (EEG) measures brain activity from electrodes placed on the scalp. The electrodes detect voltage activity predominantly from the postsynaptic potentials of cells. Because the cerebral cortex has a columnar structure, the scalp electrodes detect summed activity from the vertically-oriented pyramidal cells rather than cancelled activity (Laureys, Gosseries, & Tononi, 2015; Nunez & Srinivasan, 2006). The advantages of EEG are that it is non-invasive, has a high temporal resolution on the order of milliseconds, and covers the entire cortex (Nunez & Srinivasan, 2006). However, despite covering a broad spatial extent, there is poor spatial resolution due to volume conduction from the signals traveling through the skull and scalp so each electrode detects activity coming from approximately 10 cm² of cortex (Kramer & Eden, 2016; Laureys et al., 2015). Analogously, the scalp electrodes detect a mixture of signals arising from cortex like a microphone in a crowded room detects a mixture of voices. Source-localization is a technique to achieve higher spatial resolution by projecting the activity from the relatively low-dimensional electrodes on scalp onto a high-dimensional source space on cortex and de-mix the signal. The projection is constrained by using biophysical modeling based on known physical properties of the system such as the electrical conductance of the tissues and geometry of the skull (Hamalainen & Sarvas, 1987; Laureys et al., 2015).

1.1.1.2 Oscillations

In contrast to neural spiking in which spike timing encodes information (Kass et al., 2018), for electrophysiological recordings, it is the shape and pattern of the brain's electrical activity, *i.e.*, rhythms, that encode information and communication between brain regions (Buzsáki & Draguhn, 2004; Fries, 2015; Uhlhaas, Roux, Rodriguez, Rotarska-Jagiela, & Singer, 2010). Oscillations represent the interplay between inhibition and excitation in neuronal populations. Due to synaptic and axonal conductance delays, high frequency oscillations recruit local cortical activity whereas low frequency oscillations recruit activity broadly across cortex (Buzsáki, 2006; Buzsáki & Watson, 2012).

Oscillations have been shown to have a strong functional relevance for coordinating and consolidating information (Buzsáki, 2006). The first example of functional oscillatory activity in humans was discovered by Dr. Hans Berger who identified prominent alpha band (8-12 Hz) activity in a human subject with eyes closed during rest (Berger, 1929; Kramer & Eden, 2016). Oscillations are conventionally divided into frequency bands each having specific associations with different functions: delta (1-4 Hz), theta (4-8 Hz), alpha, beta (12-30 Hz) and gamma (>30 Hz) (Buzsáki, 2006; Engel & Fries, 2010). Demonstrated in sleep EEG, brain rhythms evolve over development with regional specificity. Lower frequencies tend to decline while high frequencies gradually increase; lower frequencies are more prominent in the posterior regions in childhood and higher frequencies are more prominent in frontal and centrottemporal regions in adolescence. The temporal and spatial characteristics that vary throughout development indicate that rhythms may be integral during critical periods of cognitive development (Chu, Leahy, Pathmanathan, Kramer, &

Cash, 2014).

1.1.1.3 Appearance in states of consciousness

Different states of consciousness, *e.g.*, sleep and wakefulness, and restfulness and alertness, are associated with different frequency bands (Blumenfeld & Taylor, 2003). Generally, lower frequencies indicate states of diminished consciousness whereas higher frequencies indicate states of arousal (Laureys et al., 2015). Clinically, anesthesia, coma, and epileptic seizures are strongly associated with slow delta rhythms and minimal consciousness (Blumenfeld & Taylor, 2003; Brown, Lydic, & Schiff, 2010; Frohlich, Toker, & Monti, 2021). Delta oscillations are also prominent during deep sleep (Frohlich et al., 2021) whereas gamma oscillations and theta-gamma cross-frequency coupling are associated with rapid-eye movement (REM) sleep (Brankačk et al., 2012; Laureys et al., 2015). Alpha oscillations are associated with restful wakefulness (Berger, 1929) and beta oscillations are associated with the maintenance of cognitive state (Pfurtscheller, Stancák, & Neuper, 1996), while gamma oscillations are associated with alertness (Engel & Fries, 2010). These are only a few examples from the plethora of associations between behavior and function (Buzsáki, 2006) Due to the strong relationship between brain rhythms and state of consciousness, EEG recordings are frequently used in clinical settings, *e.g.*, monitoring sleep stages or degrees of consciousness under anesthesia (Blumenfeld & Taylor, 2003; Brown et al., 2010).

1.1.2 Relevance for Disease

1.1.2.1 EEG as biomarker

A biomarker is a measurement indicating presence or intensity of some normal biological or pathological process (Engel Jr, 2011; Strimbu & Tavel, 2010). For example, blood pressure is a biomarker for heart disease (Desai, Stockbridge, & Temple, 2006). Biomarkers are helpful in diagnosing disease, assessing the efficacy of treatments, and reducing exposure to ineffective therapies (Strimbu & Tavel, 2010); this is especially important when disease progression is difficult to track. For example, in epilepsy, seizures occur variably across patients, varying from once every few days to months to years; therefore, it can be difficult to assess the ability of a treatment in reducing seizures when measuring seizure frequency alone (Engel Jr, 2011). Biomarkers do not necessarily correlate with disease mechanisms however and so few biomarkers can be used as surrogate endpoints in clinical trials, *i.e.*, a marker that can be used as a substitute measurement for patient health (Strimbu & Tavel, 2010).

Electrophysiological recordings, such as EEG, are prime candidates for biomarker discovery. While neuroanatomy can reveal structural abnormalities, neurophysiology reflects the dynamic aspects of the brain and transient brain functioning. This makes electrophysiological biomarkers useful for diagnosing disease, measuring severity of cognitive symptoms, and localizing the circuit mechanisms responsible for generating the abnormal physiology (Frohlich et al., 2019; Kramer et al., 2019, 2021; Ostrowski et al., 2021). A classic example of an electrophysiological biomarker for epilepsy is the interictal epileptiform discharge (IED), sometimes called epileptic spikes, *i.e.*, a large fluctuation of

electrical activity from a population of neurons that occur between seizures (Engel Jr, 2011). IEDs are highly specific to epilepsy and useful for diagnosis. Rate and location of discharges help to differentiate between types of epilepsy (Engel Jr, 2011; Symms et al., 1999). Because EEG is noninvasive and relatively easy to collect, electrophysiological biomarkers are an ideal tool to diagnose and track disease progression.

1.1.2.2 Rolandic epilepsy

Rolandic epilepsy is the most common form of childhood focal epilepsy, accounting for approximately 10% of childhood-onset epilepsies (Astradsson, Olafsson, Ludvigsson, Björgvinsson, & Hauser, 1998; Hauser, 1994). Formerly, Rolandic epilepsy was misnamed Benign Epilepsy with Centrotemporal Spikes, labeled benign because of its excellent prognosis. Seizures and comorbid cognitive deficits, primarily consisting of sensorimotor function, naturally resolve by adolescence (Wickens, Bowden, & D'Souza, 2017), generally two to four years after onset (Panayiotopoulos, Michael, Sanders, Valeta, & Koutroumanidis, 2008). Because of this, many clinicians did not treat Rolandic epilepsy since the cognitive side effects of the antiseizure medications were deemed worse than the cognitive symptoms of the disease (Fernandez, Loddenkemper, Galanopoulou, & Moshe, 2015; Hughes, 2010). However, no seizure is benign; there is a high prevalence of long-term psychosocial problems, including depression and anxiety, in Rolandic epilepsy patients likely due to the presence seizures and cognitive deficits at critical periods in development and social stigma against epilepsy (Camfield & Camfield, 2014). Thus, with growing popularity and advent of medications with fewer side effects, more clinicians are choosing to treat Rolandic epilepsy (Mellish, Dunkley, Ferrie, & Pal, 2015).

Like other epilepsies, an electrophysiological phenotype of Rolandic epilepsy is IEDs; however, uniquely these discharges occur nocturnally and originate from the inferior Rolandic cortices, *i.e.*, the sensorimotor cortex. While useful for diagnosis, IED rate, also called spike rate, does not predict seizure risk in Rolandic epilepsy (Xie, Ross, Kramer, Eden, & Chu, 2018) nor has it been shown to correlate with cognitive symptoms (Kramer et al., 2021). Recently, it has been found that the absence of healthy sleep rhythms, *i.e.*, sleep spindles, correlates with the cognitive deficits that accompany the disease (Kramer et al., 2021), suggesting thalamocortical circuit malfunction (Beenhakker & Huguenard, 2009). Chapter 3 explores the spatial extent of these cognitive deficits using source-localized EEG. Unlike spikes, sleep spindles are a powerful biomarker that directly corresponds to a patient's well-being.

Recently, Rolandic epilepsy has been recognized as an epileptic encephalopathy (Carvill et al., 2013), a disorder in which abnormal epileptic activity, such as spikes, is comorbid with cognitive impairment worse than what would be expected from the underlying pathology (Berg et al., 2010). Rolandic epilepsy is caused by the same mutation (Carvill et al., 2013; Lemke et al., 2013) and exhibits similar abnormal electrophysiological activity (Scheltens-De Boer, 2009) as epileptic encephalopathies. Additionally in a small number of cases, patients develop more severe encephalopathy, such as Landau-Kleffner syndrome and continuous spike-and-waves during slow sleep, and so Rolandic epilepsy is now considered on the spectrum of encephalopathies (Camfield & Camfield, 2014; Fejerman, 2009; Fejerman, Caraballo, & Tenenbaum, 2000; Tovia et al., 2011). Thus,

biomarkers discovered for Rolandic epilepsy may be directly translatable for more severe epileptic encephalopathies.

1.1.2.3 Angelman syndrome

Angelman syndrome (AS) is a rare neurodevelopmental disorder occurring in 1 in 10,000 to 20,000 people and requires lifelong care (Kyllerman, 1995; Mertz et al., 2013; Petersen, Brøndum-Nielsen, Hansen, & Wulff, 1995a; Thibert, Larson, Hsieh, Raby, & Thiele, 2013). AS was first reported by Dr. Harry Angelman in 1965 describing three case studies of patients with now canonical characteristics of AS, such as severe cognitive delay, microencephaly, ataxia and easily provoked bouts of laughter (Angelman, 1965). In fact, AS patients reach a cognitive plateau between 27-30 months of age and are primarily nonverbal (Gentile et al., 2010; Peters et al., 2004). AS is also highly comorbid with epilepsy; approximately 80%-95% have epilepsy though the etiology is unknown (Thibert et al., 2013). Currently, only supportive treatments have been developed, such as antiseizure medications for epilepsy, therapy for aggression and melatonin for sleep disturbances (Buiting, Williams, & Horsthemke, 2016).

AS is caused by decreased expression of the UBE3A protein. In neurotypical subjects, only the maternal UBE3A allele is expressed in neurons, and the paternal allele is silenced through genomic imprinting. In AS, the maternal allele is deleted or mutated leading to reduced expression (Bird, 2014; Thibert et al., 2013). AS caused by a deletion, comprises approximately 75% of AS cases and correlates with a more severe phenotype, *i.e.*, more cognitively delayed and higher susceptibility to seizures, than AS caused by a mutation (Buiting et al., 2016; Gentile et al., 2010).

A prominent electrophysiological feature of AS is high amplitude slow oscillations in the delta band, present during wakefulness as well as sleep (Frohlich et al., 2020, 2019; Sidorov et al., 2017). Delta power is traditionally associated with states of diminished consciousness making the presence of these oscillations during wake even more striking (see section 1.1.1.3) (Frohlich et al., 2021). Delta power is higher in deletion patients (Frohlich et al., 2019) and correlates with worse cognitive function (Ostrowski et al., 2021). In a mouse model, delta power oscillations similar to AS and increased seizure susceptibility were induced by suppression of UBE3A in GABAergic neurons (Judson et al., 2016). These results suggest that therapeutics correcting the functional pathophysiology may improve the disease phenotype.

AS is a devastating disease with little known about the pathophysiological mechanisms leading to dysfunction (Thibert et al., 2013). However, promising disease-modifying therapies are on the horizon that aim to reinstate UBE3A expression and restore cognitive function (Bi, Sun, Ji, & Baudry, 2016; Meng et al., 2015). In mouse models of AS, it has been shown that reinstating UBE3A expression recovered cognitive function; however, stronger improvement occurred with earlier UBE3A reinstatement (Meng et al., 2015; Silva-Santos et al., 2015). Electrophysiological biomarkers such as delta power are critical for clinical trials to help measure target engagement, *i.e.*, whether a therapeutic sufficiently increased UBE3A in the brain. Additionally, because delta power correlates with cognitive functioning, delta would provide a useful measure of therapeutic impact on patient well-being. Chapter 4 provides a model-based approach to measure changes of delta power due to efficacious treatment in patients with AS.

1.2 Mathematical Approaches to Analyze Electrophysiological Recordings

1.2.1 Linking data to scientific hypotheses: statistical models

1.2.1.1 Statistical models: linear models, generalized linear models, mixed effects models

Statistical models are designed to uncover relationships in data in the presence of noise (Kass et al., 2014). For a set of observations, $(x_1, y_1), \dots, (x_n, y_n)$, statistical modeling can answer the following questions: (i) are x and y related?, (ii) to what extent are x and y related?, and (iii) how accurately can we estimate the effect of x on y ? (James, Witten, Hastie, & Tibshirani, 2013).

The simplest statistical model is a linear regression model, where we consider the random variable Y :

$$Y_i = \beta_0 + \beta_1 * x_i + \epsilon_i,$$

for $i = 1, \dots, n$ observations, where β_0 represents the expected value of Y when $x=0$ and β_1 represents the strength of the relationship between x and Y . This model assumes that the relationship between x and Y is, on average, linear and that the errors are independent and identically distributed, $\epsilon_i \sim N(0, \sigma^2)$ (Kass et al., 2014). This model can be extended to *multiple linear regression*, by including more covariates on the right-hand side of the equation, x_{1i}, \dots, x_{ki} to control for potential confounding influences.

When the data, Y , are not linearly distributed, we use a more general form of regression called *generalized linear regression* (GLM):

$$Y_i \sim f_{Y_i}(y_i | \theta_i)$$

$$\theta_i = f(x_{1i}, \dots, x_{ki}),$$

where f_{Y_i} is a probability density function that is parameterized by θ_i . Linear regression is a special case of GLMs where f_{Y_i} is the normal distribution, parameterized by mean, $\mu_i = \beta_0 + \beta_1 * x_{1i} + \dots + \beta_k * x_{ki}$, and variance σ^2 (Kass et al., 2014). Examples of problems where GLMs are employed include logistic regression in which the dependent variable is described by binary data, *e.g.*, successes and failures, or Poisson regression in which the dependent variable is described by count data, *e.g.*, the number of times an event occurs. While there exist more formal methods to choosing f_{Y_i} , such as the Kolmogorov-Smirnov test and Q-Q plots, one of the most informative tools is visual inspection (Anscombe, 1973; Kramer & Eden, 2016).

One of the primary assumptions of GLMs is that the data observations are independent when conditioned on the predictors. However, in many datasets, there exist correlations between observations violating the assumption of independence. This is particularly prominent in longitudinal data where there are repeat measurements per subject (McCullagh & Nelder, 1989). Naïve approaches to retaining independence include averaging across subjects or analyzing an independent subset of the observed data. An alternative, more powerful, approach that leverages the complete dataset while simultaneously accounting for correlations in the data are *mixed-effects models*, also called hierarchical models in the Bayesian framework (Gałecki & Burzykowski, 2013; Gelman, 2014). Mixed-effects models are extensions of GLMs in which there is a more complex covariance structure (McCullagh & Nelder, 1989). In this framework, the error term (*i.e.*, the difference between the observations and the mean estimate from the model predictors)

is divided into independent errors, *i.e.*, ϵ_i , and correlated errors, named *random effects*. The nonrandom components, the estimated coefficients, are named as *fixed effects*. Suppose we have collected a set of datapoints, $(x_1, y_1), \dots, (x_n, y_n)$, such that the data can be partitioned into j groups. Each group may have a different baseline activity which would be modeled by a *random intercept*, or they may have a different relationship between x and Y which would be modeled by a *random slope*. However, on average, the relationship between x and Y , is described by the fixed effects. A mixed-effects GLM can be expressed as:

$$Y_{i,j}|Group_j \sim f_{Y_{i,j}}(y_i|\theta_i, \nu_j)$$

$$\theta_{i,j} = f(x_{1ij}, \dots, x_{kij}),$$

where ν_j represents the random effects for group j .

Supposing Y is normally distributed, linearly correlated with a single predictor x , and can be divided into j groups, we write the formula for a linear mixed effects model as¹:

$$Y_i|Group_j \sim \beta_0 + \nu_{0,j} + (\beta_1 + \nu_{1,j}) * x_i + \epsilon_i$$

$$\nu_j \sim N(\mathbf{0}, \Sigma), \epsilon_i \sim N(0, \sigma_{model}^2),$$

where, $\nu_j = [\nu_0; \nu_1]$, $\mathbf{0} = [0; 0]$, $\Sigma = \begin{bmatrix} \sigma_{int}^2 & \rho * (\sigma_{int}\sigma_{slope}) \\ \rho * (\sigma_{int}\sigma_{slope}) & \sigma_{slope}^2 \end{bmatrix}$. Here, β_0 and

β_1 represent the fixed effects, and ν_0 and ν_1 represent the random effects (random

¹ Note this is atypical notation. Traditionally, the random intercept and slope are included on the right-hand side of the equations as $(1|Group)$ and $(x|Group)$, respectively, which is a slight abuse of notation. Here, we adopt a Bayesian framework for the model by representing the conditional dependence of group membership on Y on the left-hand side of the equation because we find this more intuitive.

intercept and slope, respectively) due to group membership, j . Σ is the covariance matrix describing the variance of the random intercept, σ_{int}^2 and random slope, σ_{slope}^2 , and their covariance, $\rho * (\sigma_{int}\sigma_{slope})$ (Gałecki & Burzykowski, 2013). Figure 1.1 shows a simulation of a linear mixed effects model with a correlated random intercept and slope ($\beta_0 = 5, \beta_1 = 10$, number of groups = 5, observations per group = 15, $\rho = 0.9, \sigma_{int} = 30, \sigma_{slope} = 5, \sigma_{model} = 250$). Visual inspection reveals that as X increases so does Y (Figure 1.1A); however each group, indicated by color, has slightly different intercepts and slopes (Figure 1.1B) that vary around the mean effect. Additionally, we note that as the random intercept increases so does the random slope (Figure 1.1B). In Chapters 3 and 4, we implement random intercept models. In Chapter 3, we model motor performance as a function of spindle rate; however, we have two measurements per subject: performance in the left and in the right hand (Figure 3.6A). We include a random intercept to account for these correlated measurements, which can be interpreted as baseline differences in motor performance (for example, if a subject plays video game, they may perform better on the task in both hands than a subject who does not). In Chapter 4, we utilize a large dataset of longitudinal EEGs from multiple patients with repeat EEGs to model delta power as a function of time. We again use a random intercept to account for baseline differences in delta power (Figure 4.2).

We briefly summarize the model fitting procedures here. For GLMs, maximum likelihood estimation (MLE) is used. The model parameters are computed by maximizing

the log-likelihood function, $\ell(\theta)$, usually denoted $\hat{\theta}_{MLE}$. The log-likelihood function is the joint density of the data treated as a function of the model parameters, θ :

$$\ell(\theta) = \log \prod_{i=1}^n f(x_i|\theta). \quad (1.1)$$

The $\ell(\theta)$ estimates the likelihood that the given data are observed given the proposed model. The log-transform of the likelihood allows for easier computations but yields the same result as maximizing the likelihood itself because log is a monotonic transform (Wasserman, 2010). In linear regression, this reduces to minimizing the residual sum of squares (Figure 1.2B) (James et al., 2013). For mixed-effects models, MLE or restricted maximum likelihood (REML) estimation is used. Because the MLE of the variance-covariance parameters are biased, REML corrects the MLE variance-covariance parameters so that they are unbiased estimators. If the goal of the model is to estimate the fixed effects, then MLE is used; if the goal is to estimate the random effects, then REML is used (Gałecki & Burzykowski, 2013). In each mixed effects model implemented in this thesis, we use MLE as we are interested in understanding the main relationship between the dependent and independent variables. Finally, in the case of over-dispersion, quasi-maximum likelihood estimation (QMLE) is used. The data are *over-dispersed* when there is more variability in the data than expected by the model. Over-dispersed data are common, and some argue are the rule rather than the exception (McCullagh & Nelder, 1989). Dispersion is estimated by the variance parameter, σ^2 . In linear regression σ^2 is a free parameter thus QMLE is unnecessary. However, in Poisson regression, for example, the mean and variance relationship is fixed by MLE, $\hat{\mu}_{MLE} = \sigma^2$. If the data has more

variability than expected by the model, then $\mu = c * \sigma^2$, where $c > 1$. QMLE allows for specification of a mean and variance relationship to be used for parameter estimation (McCullagh & Nelder, 1989; Wedderburn, 1974). In Chapter 3, we estimate spindle rate as a function of subject group (active epilepsy, remission epilepsy, healthy control). Upon visual inspection, we observe that spindle rate versus group does not follow a normal distribution because many subjects in the active group had zero spindle rate (Figure 3.2). The slight clustering near zero implies a Poisson model would be useful. Poisson is traditionally used for count data and while spindle rate is a continuous variable, it is derived from count data (the number of spindles). In implementing a Poisson model for continuous data, we encountered over-dispersed data ($\sigma^2 = 95966^2$) and thus used QMLE; though, we obtained qualitatively similar results by using a linear model.

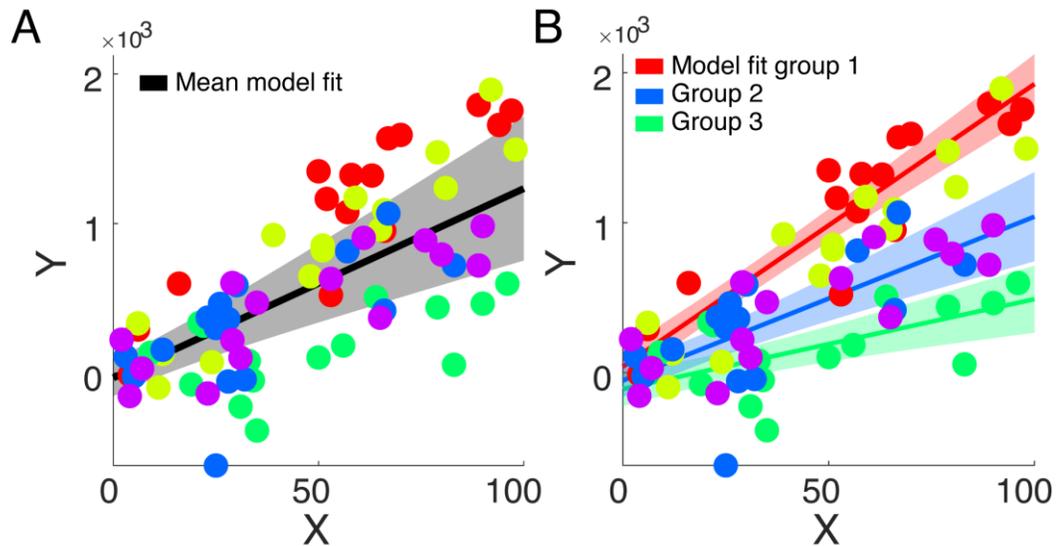


Figure 1.1: Simulated mixed effects model with correlation between the random intercept and slope.

(A) Estimate of the fixed effects independent of group membership on Y .

(B) Example of three conditional model fits with intercept and slope dependent on group membership.

1.2.1.2 Considerations in model selection and interpretation

Statistical modeling aims to express scientific statements about relationships and uncertainty in those relationships in data; however, the ability of the model to reveal such relationships requires careful model selection. Important considerations for selecting a good model are as follows: (i) balancing complexity and parsimony, (ii) hypothesis-driven questions, and (iii) interpretability (Burnham & Anderson, 2004; James et al., 2013; Kass et al., 2014; McCullagh & Nelder, 1989). For (i), data observations are complex and noisy; what we observe is a combination of signal and noise (Brown & Kass, 2009) and models will only ever approximate reality (Burnham & Anderson, 2004). Fitting a model that matches the observations exactly, *i.e.*, *overfitting*, fits to the noise and clouds interpretation of the true signal. Overfitting yields a model that fits the dataset perfectly but will not generalize well. The converse is *underfitting*, in which not enough of the variability in the data is explained, inducing bias (McCullagh & Nelder, 1989). The “ideal” model, though models only approximate reality, lies somewhere in the middle: the most parsimonious model that estimates enough of the variability. For (ii), having multiple working hypotheses rooted in scientific reasoning limits the number of parameters and models selected for analysis. Exploring every possible relationship in the data runs the risk of finding spurious significant results, *i.e.*, Type-1 errors, emphasizing the importance of considering the scientific context before analysis (Burnham & Anderson, 2004; Hochberg & Tamhane, 1987). If criteria (i) and (ii) are met, the model interpretability (iii) follows directly. If the model is parsimonious (excluding unnecessary covariates) and has a strong foundation in science, then interpretation is simple (James et al., 2013) though

interpretation is often omitted from scientific articles. Oftentimes analysis is restricted to the binary question, is this relationship significant or not (*i.e.*, is $p < 0.05$?); however, interpreting model coefficients, *i.e.*, effect sizes, and confidence bounds, yields information about the strength of the relationship and the replicability of the experiment. Additionally, interpretation checks model assumptions; if the interpretation of the model does not make sense in a scientific context, then the model is misspecified (Calin-Jageman & Cumming, 2019).

1.2.1.3 Assessing goodness-of-fit: R^2 and Akaike Information Criterion

There are many ways to assess how well the model fits the data, *i.e.*, goodness-of-fit. Here, we detail two common methods that are used in this thesis: R^2 and the penalized-likelihood-based criterion, the Akaike information criterion (AIC). Ordinary R^2 , also called the coefficient of determination, is for simple and multiple linear regression expressed as:

$$R^2 = 1 - \frac{SSE}{SST},$$

where SST is the sum of total squares, $SST = \sum_{i=1}^n (y_i - \bar{y})^2$ (Figure 1.2A), and SSE is the sum of squares for error, $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ (Figure 1.2B). R^2 is a geometric measure bounded between 0 and 1 representing the proportion of variability in the dependent variable explained by the model (Kass et al., 2014). In the case of simple linear regression, R^2 is equivalent to the square of the correlation between x and y . However, a good value of R^2 is domain-dependent rendering it difficult to interpret across datasets. For example, having a low value does not necessarily imply a poor model fit and having a high value does not necessarily imply a good model fit as R^2 cannot detect model misspecification in

the presence of correlated residuals (Birnbaum, 1973; Kass et al., 2014). Ordinary R^2 improves as more parameters are included in the model and does not prevent against overfitting. A related measure, the adjusted R^2 , is similarly interpreted as R^2 but penalizes for number of parameters by using unbiased estimates of the variance (James et al., 2013) however both ordinary and adjusted R^2 are still very sensitive to Type-1 errors (van Ginkel, 2019).

Another measure of goodness-of-fit is the AIC which balances the log-likelihood of the data, $\ell(\theta)$ (see Equation (1.1)), and the number of parameters fit, k :

$$AIC = 2k - 2\ell.$$

AIC can be used to compare nested or non-nested models fit on the same data, and the model with the lowest AIC is typically preferred. While the raw values of AIC are uninformative, the differences in AIC are interpretable. Typically a difference greater than 3 is considered a meaningful difference (Burnham & Anderson, 2004). In the case of linear regression, the AIC reduces to:

$$AIC = 2k + n * (\log\left(2 * \pi * \frac{SSE}{n}\right) + 1).$$

In the case of linear regression, both R^2 and AIC depend on the SSE (Figure 1.2A).

In Anscombe (1973), the authors present four datasets that estimate approximately the same linear regression model, having the same R^2 and AIC. Visual inspection of the data reveals that only one of the datasets exhibits a linear relationship despite all model fits having equal R^2 and AIC values. This motivates the need for graphical analysis of the data including scatterplots and residual plots. Plotting the residuals versus the fitted values can

demonstrate such lack-of-fit, revealing unexplained trends, heteroscedasticity (*i.e.*, unequal variances in the residuals), and outliers (Kass et al., 2014). There is no gold standard technique for measuring goodness-of-fit; different assessments highlight different features of the models weighing either variance as is the case with R^2 or generalizability as is the case with AIC.

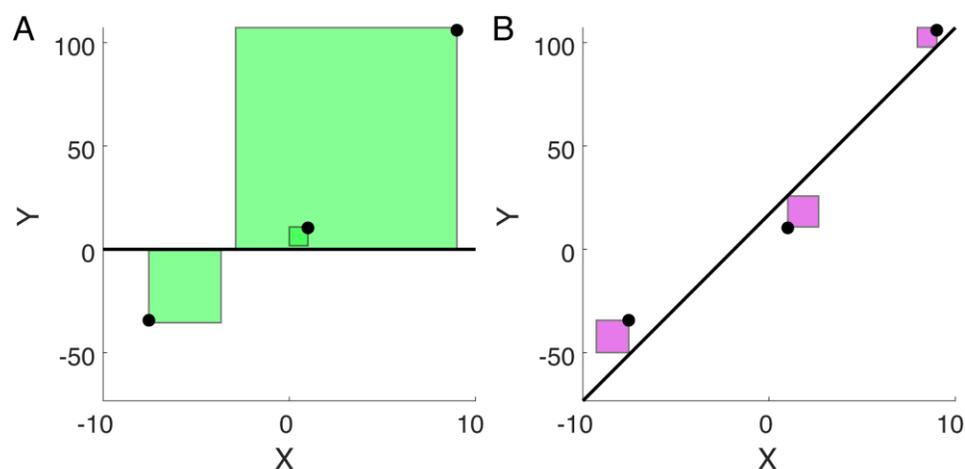


Figure 1.2: Geometric interpretation of R^2 .

Black circles indicate example three example data points, (x, y) . In (A), the total variability in Y is shown relative to the mean. The horizontal black line indicates the mean of Y and the green area represents the total sum of squares. In (B), the remaining error in Y is shown after fitting a linear regression model to the data. The black line is the linear fit, and the pink area represents the sum of squares for error.

1.2.2 Time series analysis

1.2.2.1 Random processes and autoregressive models

A process, x_t , is *random* if for each t , the value of x_t is a random variable such that each value can be presented as a probability distribution. Thus, every observation of a random process, *i.e.*, a *realization*, is drawn from a collection of possible sequences (Priestley, 1981). Neural time series, data collected over time, are considered realizations

of a random processes thus we can use statistical tools to describe temporal dependencies, *i.e.*, in the temporal domain, or frequency components, *i.e.*, in the frequency domain (see 1.2.2.2) (Kass et al., 2014). In analyzing time series, we often require that the process be *stationary*, loosely defined here to mean that the statistical properties of the signal stay constant in time (Priestley, 1981). Brain activity from electrophysiological recordings is highly nonstationary so often data are analyzed by windowing the data into shorter observation periods (Cohen, 2014).

A common model of stationary neural activity is an autoregressive (AR) model in which a signal is modeled as a linear function of its own history (Cohen, 2014; Priestley, 1981). For example, a signal, x_t , can be written as a function of its past activity up to lag p to form an $AR(p)$ model as follows:

$$x_t = \sum_{i=1}^p \beta_i * x_{t-i} + \varepsilon_i.$$

AIC is generally used to justify the model order; however, there are a few considerations when choosing model order. Small orders require fewer parameters so are more robust to noise but cannot capture longer lagged effects; large orders capture longer lagged effects but are less robust to noise and require more data to estimate (Cohen, 2014).

Chapter 2 presents an application of this type of model, Granger causality, to networks of connected neural signals in which x_t may be additionally driven by the history of other connected signals, such as y_t . Granger causality estimates if x_t is being driven by y_t , *i.e.*, if the activity of signal x_t can be modeled as a function of the history of y_t in addition to its own history. It does so by building nested multivariate autoregressive models (MVAR) expressed as:

$$x_t = \sum_{i=1}^p \beta_{xi} * x_{t-i} + \sum_{i=1}^p \beta_{yi} * y_{t-i} + \sum_{i=1}^p \beta_{zi} * z_{t-i} + \varepsilon_i \quad (1.2a)$$

$$x_t = \sum_{i=1}^p \tilde{\beta}_{xi} * x_{t-i} + \sum_{i=1}^p \tilde{\beta}_{zi} * z_{t-i} + \tilde{\varepsilon}_i, \quad (1.2b)$$

where (1.2a) is a full model including the history of all signals in the network x_t, y_t , and z_t and (1.2b) is a nested model without the history of y_t . Inclusion of y_t in the final model is determined by computing the F-statistic. This method is broadly used in neuroscience with many applications (Barrett et al., 2012; Ding, Chen, & Bressler, 2006; Greenblatt, Pflieger, & Ossadtchi, 2012); however due to the nonstationary of brain data and large number of parameters in building MVAR models, Granger causality is computationally challenging. Additionally, model parameters are rarely interpreted in Granger causality (see section 1.2.1.2). Chapter 2 addresses these challenges.

1.2.2.2 Spectral analysis

Spectral analysis refers to the statistical tools used to analyze random time series data in the frequency domain. The statistical tools are derived from a powerful theorem in mathematics, Fourier's theorem, which states that any well-behaved periodic function can be written as a linear combination of sine and cosine functions (Priestley, 1981). Simply put, such deterministic functions can be exactly represented as a sum of oscillations. The discrete Fourier transform of a discrete time series, x_1, \dots, x_N , is expressed as follows:

$$X_j = \frac{1}{\sqrt{N}} (\sum_{n=1}^N x_n [\cos(2\pi f_j t_n) - i * \sin(2\pi f_j t_n)]) = \frac{1}{\sqrt{N}} \sum_{n=1}^N x_n e^{-2\pi i f_j t_n},$$

where $t_n = \Delta\{0, 1, 2, \dots, N - 1\}$, Δ = the sampling interval, and $f_j = \frac{j}{T}$ where $j = \left\{-\frac{N}{2} + 1, \dots, \frac{N}{2}\right\}$ represents the discrete frequency. When dealing with random functions or processes, as is neural data, we can apply similar analytic techniques to approximate the signal as a sum oscillations (Kass et al., 2014; Priestley, 1981). In neuroscience, spectral analysis is broadly applied to neural data due to the functional role of oscillations in cognition and consciousness (see sections 1.1.1.2, 1.1.1.3).

A tool to identify the frequency content of a signal is the power spectrum which quantifies the strength of an oscillation in a given frequency band (Buzsáki & Watson, 2012; Kramer & Eden, 2016). One approach to estimate the power spectrum is computing the magnitude squared of the Fourier transform of x_n , $|X_j|^2$, giving the amplitude of the oscillations as a function of frequency, $|f_j|^2$. Another way to think about this is if instead we expand x_n into a sum of sines and cosines (via harmonic regression) oscillating at different frequencies. Then, equivalently, each frequency coefficient's contribution to R^2 is correlated with the power spectrum (Kass et al., 2014). Or in other words, the spectrum represents the amount of variability in the signal due to that oscillation.

In Figure 1.3, an example power spectrum over time, *i.e.*, the spectrogram, is shown for example neural activity of a patient with Rolandic epilepsy. A characteristic of brain activity is evident from the spectrogram: low frequencies are more dominant and have higher spectral power than high frequencies. This is called the power law distribution of brain activity in which the power decreases at a rate of $1/f^\alpha$ (Buzsáki & Watson, 2012). The $1/f^\alpha$ background activity is sometimes described as noise since it is aperiodic despite

some evidence showing it may have a functional role itself (He, Zempel, Snyder, & Raichle, 2010). Periodic oscillatory activity appears as spectral bumps in the power spectrum or spectral islands in the spectrogram. Identifying prominent oscillations can be estimated by isolating the bumps or islands via subtracting the $1/f^\alpha$ background activity (Ouyang, Hildebrandt, Schmitz, & Herrmann, 2020).

There are many considerations when analyzing healthy and pathological brain rhythms to avoid inference of spurious oscillatory activity. These include: (i) artifacts in the EEG, (ii) signal tapering, (iii) variation in skull thickness, and (iv) limited data. For (i), nonneural activity (*e.g.*, muscle movements, blinking) and neural activity (*e.g.*, pathological spikes) may appear as sharp edges in the EEG, leading to increased power at all frequencies (Kramer, Tort, & Kopell, 2008; Ostrowski et al., 2021). Intuitively, a single peak in the spectrum at 10 Hz, for example, corresponds to a 10 Hz sinusoid for all time in the time domain. Conversely, a sharp edge or peak in the time domain leads to power at all frequencies, or a spectral smear, in the frequency domain (Kramer et al., 2008). For example, Figure 1.3A and B show brain activity from a patient with Rolandic epilepsy during non-rapid eye movement (NREM) sleep and the corresponding spectrogram. At approximately 140 s, there is an epileptic spike (Figure 1.3C) leading to a simultaneous “smear” in the spectrogram, illustrating the importance in removing artifacts before analyzing the true neural oscillatory activity. For (ii), computing a power spectrum on a finite duration of EEG activity is equivalent to multiplying a hypothetical infinite duration signal by a rectangular window. Doing so introduces sharp edges in the EEG, leading to an important artifact: leakage into other frequencies. To account for this, the data can be

multiplied by a tapered window so that the edges of the signal slowly decay to zero (Kass et al., 2014; Kramer & Eden, 2016). For (iii), skulls thicken over age and vary between individuals. Thicker skulls induce more volume conduction, *i.e.*, a mixing of signals in the scalp electrodes, and lead to reduced power at all frequencies. By computing the normalized or relative power spectrum, *i.e.*, dividing by the summed power over all frequencies, differences in skull thickness and related non-neural variations can be reduced spatial blurring is reduced (Chu et al., 2014). Finally, for (iv), brain rhythms are highly nonstationary meaning that frequency content can only be estimated on short windows (Cohen, 2014). Oftentimes power spectra are averaged over many trials to yield a consistent estimate (Priestley, 1981). However, for spontaneous activity in which there is no trial structure, such as a seizure, the power spectrum provides only an approximate estimate of the oscillatory activity with large uncertainty. In multitaper spectral analysis, brain activity is multiplied by a set of orthogonal taper functions yielding independent samples from which to compute the power spectrum. Noise is reduced in the power spectrum by averaging the estimates across tapers, and 95% confidence bounds can be estimated using a chi-square distribution (degrees of freedom $2 \times \text{number of trials}$) (Kramer & Eden, 2016). In this dissertation, we study how oscillatory activity has been implicated in disease, and so to gain more confidence in our spectral estimates, we employ multitaper spectral analysis throughout.

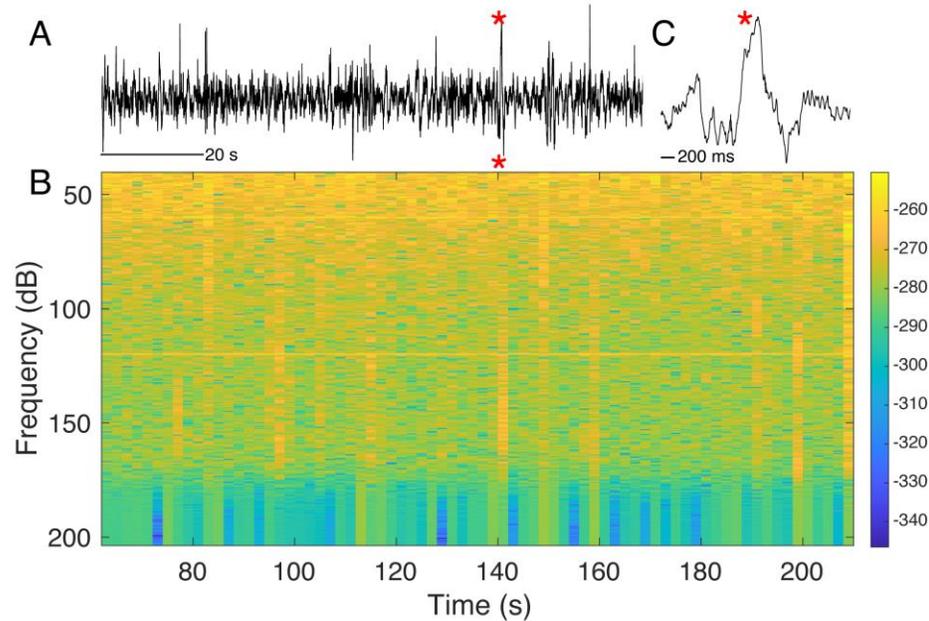


Figure 1.3: Sharp edges in electrophysiological recordings induce increases in power at all frequencies.

- (A) Example electrophysiological activity during sleep from a patient with Rolandic epilepsy. Red asterisk indicates the time of a pathological spike. Scale bar indicates 20 s of activity.
- (B) Corresponding power spectrogram for signal in (A). There is an increase in spectral power at all frequencies, as shown by the vertical stripe, at the time of the corresponding spike in (A; red asterisk).
- (C) Zoomed in signal to visualize the pathological spike. Scale bar indicates 200 ms and red asterisk indicates corresponding time in (A).

1.2.2.3 Functional network models and coherence

Many different cortical areas work together to help process the environment. Often cortical areas work together, or functionally connect, by coupling their activity at the same frequency. Functional connectivity is necessary for healthy brain functioning to process and relay information to the body and mind but can become pathological when brain areas synchronize and desynchronize that should not. Functional network models are a representation of how oscillations interact and synchronize their activity to support

cognition, behavior, and perception (Park & Friston, 2013; Petersen & Sporns, 2015). Many disorders are characterized as network disorders, such as, epilepsy, schizophrenia, and autism because they are well characterized by changes in network topology (Bassett & Sporns, 2017; Kramer & Cash, 2012; Wamsley et al., 2012).

In the language of network science, the human brain consists of two fundamental components: *nodes* (e.g., brain regions) and *edges* that connect node pairs (Figure 1.4). In functional brain networks, edges represent the transient coordination of activity, like traffic that flows over neural highways (i.e., anatomical connections between brain regions) (Englot, Konrad, & Morgan, 2016). The first step in network science research is network inference, i.e., determining which nodes are connected. In social friend networks, where nodes represent individuals and edges represent friendship, this is relatively simple (e.g., by asking are two people friends or not?). In functional brain networks, in which brain data are noisy and nonstationary, there are a plethora of statistical methods to determine the degree to which there is a connection between regions with uncertainty (Bastos & Schoffelen, 2016; Kramer, Eden, Cash, & Kolaczyk, 2009). Each method characterizes different properties of the network and may yield different inferences highlighting the need for strong hypothesis-driven methodology.

Two common complementary approaches to infer functional connectivity between brain regions are cross-correlation and coherence (Bastos & Schoffelen, 2016; Kramer & Eden, 2016). Cross-correlation measures the degree of similarity between two vectors by summing over a sliding dot product of two signals. Cross-correlation has high temporal resolution, but low frequency resolution, though is primarily dominated by delta

oscillations due to the $1/f^\alpha$ property of the brain. The frequency domain equivalent of cross-correlation is coherence which exchanges time resolution for frequency resolution. Coherence assesses the degree of phase coordination between signals for a given frequency band. The coherence between signals x and y at a given frequency, ω , is determined by measuring the consistency of the phase and amplitude between two signals. Expressing the signal x in trial k in polar coordinates, $x(\omega, k) = A_x(\omega, k)e^{i\phi_x(\omega, k)}$, and similarly for signal y , the coherence is computed as follows:

$$coherence_{x,y}(\omega) = \frac{|\frac{1}{n} \sum_{k=1}^K A_x(\omega, k) * A_y(\omega, k) * e^{i(\phi_x(\omega, k) - \phi_y(\omega, k))}|}{\sqrt{(\frac{1}{n} \sum_{k=1}^n A_x^2(\omega, k)) * (\frac{1}{n} \sum_{k=1}^n A_y^2(\omega, k))}},$$

where K is the number of trials. The numerator represents the averaged cross spectral densities of the two signals at frequency ω , and the denominator normalizes the term by the square root of the averaged signal amplitude squared, or power, of x multiplied by that of y (Bastos & Schoffelen, 2016). Consider the special case where the amplitudes of signals x and y are equal (*i.e.*, $A_x = A_y$). If the signals are not coherent, then the phase differences are completely random from trial to trial, and the expected average phase difference and coherence value are zero. If the signals are perfectly coherent, then the phase difference is fixed across trials, and the average phase difference and coherence value are one (Kramer & Eden, 2016). Because different frequency bands are linked to different functional brain states (see sections 1.1.1.2 and 1.1.1.3), coherence can be a useful tool to test state-dependent hypotheses (Fries, 2015).

Functional networks have been proposed to change in Rolandic epilepsy, but with conflicting results (*e.g.*, decreased connectivity in inferior Rolandic cortex (Besseling et

al., 2013; Shamsiri, Sheybani, & Vulliemoz, 2019; Xiao et al., 2015), no difference in connectivity in inferior Rolandic cortex (Clemens et al., 2016), and increased functional connectivity in seizure onset zone for focal epilepsies (Englot et al., 2016; Maccotta et al., 2013; Zaveri et al., 2009). In Chapter 3, we investigate functional networks in Rolandic epilepsy, and find that there no difference in coupling as measured by coherence in the Rolandic cortex (Figure 3.4E,F).

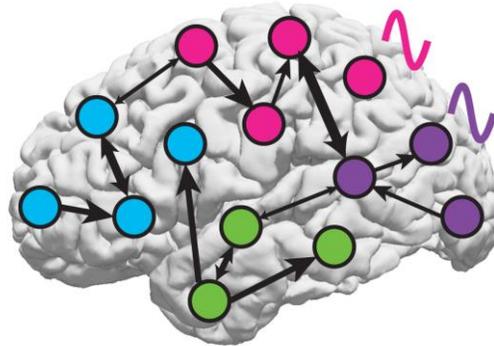


Figure 1.4: Cartoon figure of functional brain network.

Circles indicate nodes, or brain regions emitting electrical signals, and color indicates nodes in the same community that sharing similar coupling properties. Arrows represent the direction functional connectivity between nodes and the width of arrow indicates the strength of the connection.

2 CHAPTER TWO

A procedure to increase the power of Granger-causal analysis through temporal smoothing²

ABSTRACT

Background: How the human brain coordinates network activity to support cognition and behavior remains poorly understood. New high-resolution recording modalities facilitate a more detailed understanding of the human brain network. Several approaches have been proposed to infer functional networks, indicating the transient coordination of activity between brain regions, from neural time series. One category of approach is based on statistical modeling of time series recorded from multiple sensors (e.g., multivariate Granger causality). However, fitting such models remains computationally challenging as the history structure may be long in neural activity, requiring many model parameters to fully capture the dynamics.

New Method: We develop a method based on Granger causality that makes the assumption that the history dependence varies smoothly. We fit multivariate autoregressive models such that the coefficients of the lagged history terms are smooth functions. We do so by

² Chapter 2, in full, is a reprint of the following published article as it appears in press: Spencer, E., Martinet, L. E., Eskandar, E. N., Chu, C. J., Kolaczyk, E. D., Cash, S. S., Eden, U. T., & Kramer, M. A. (2018). A procedure to increase the power of Granger-causal analysis through temporal smoothing. *Journal of Neuroscience Methods*, 308(July), 48–61. <https://doi.org/10.1016/j.jneumeth.2018.07.010>

modelling the history terms with a lower dimensional spline basis, which requires many fewer parameters than the standard approach and increases the statistical power of the model.

Results: We show that this procedure allows accurate estimation of brain dynamics and functional networks in simulations and examples of brain voltage activity recorded from a patient with pharmaco-resistant epilepsy.

Comparison with Existing Method: The proposed method has more statistical power than the Granger method for networks of signals that exhibit extended and smooth history dependencies.

Conclusions: The proposed tool permits conditional inference of functional networks from many brain regions with extended history dependence, furthering the applicability of Granger causality to brain network science.

2.1 Introduction

The human brain consists of a vast network of interacting elements. Understanding how these elements interact to support cognition, behavior and perception remains a fundamental challenge in neuroscience. One approach to address this challenge is through the analysis of the brain's anatomical networks and functional connectivity - how separate brain regions interact via transient coordination of activity (de Pasquale et al., 2010; Park & Friston, 2013). Anatomical and functional networks, summarized via tools from network analysis (Bassett & Sporns, 2017; Bullmore & Sporns, 2009) provide insight into cognition (Braun et al., 2015; de Pasquale et al., 2010; Kabbara, El Falou, Khalil, Wendling, & Hassan, 2017; Petersen & Sporns, 2015; Telesford et al., 2016), learning (Bassett et al.,

2011; Bassett, Yang, Wymbs, & Grafton, 2015; Singer, 1993) and neurological disorders such as autism, stroke, schizophrenia, and epilepsy (De Vico Fallani, Richiardi, Chavez, & Achard, 2014; Kramer & Cash, 2012; Lynall et al., 2010; Matlis, Boric, Chu, & Kramer, 2015).

There exist many ways to estimate functional connectivity, inspired – in part – by theories of how brain regions communicate. One of the most established theories posits that oscillatory neuronal activity supports communication between brain regions (Bastos & Schoffelen, 2016; Bonnefond, Kastner, & Jensen, 2017; Fries, 2015; Uhlhaas et al., 2010). This theory motivates the application of many functional connectivity measures that characterize coupling between the phase and/or amplitude of rhythmic brain signals (Bastos & Schoffelen, 2016; Greenblatt et al., 2012; Lachaux, Rodriguez, Martinerie, & Varela, 1999). Other popular approaches for measuring functional connectivity - although less physically motivated - have been adopted from other scientific fields. One of the most popular methods is conditional Granger causality, which provides a direct quantification of how much the history of one brain area can predict the activity of another (Granger, 1969; Seth, 2010). There are many advantages to the Granger causality approach. First, it can be used to infer the direction of information flow, also known as directed functional connectivity (Barrett et al., 2012; Ding et al., 2006). Second, it is a model-based approach that is rooted in stochastic process theory (Bastos & Schoffelen, 2016; Cohen, 2014). Third, when the models are conditioned on the whole network and thus use all of the observed data, Granger causality limits the impact of indirect coupling - an important confound in functional network inference (Seth, 2010).

While Granger causality has been successfully applied to analyze multivariate neural activity (Seth, Barrett, & Barnett, 2015), implementation can be computationally difficult on large brain networks (Seth, 2010; Valdés-Sosa et al., 2005). Two features contribute to this computational difficulty: the number of observed brain regions, *i.e.*, nodes, and the duration for brain signal transmission. Modern recordings now support observations from hundreds, or even thousands, of sensors (Jäckel et al., 2017; Viventi et al., 2011). Interhemispheric communication can range between 5-300 ms depending on the myelination of the fiber commissures (Ringo, Don, Demeter, & Simard, 1994), whereas intrahemispheric communication ranges between 5-10 ms (de Pasquale et al., 2010; Smith et al., 2011). Fitting Granger causality models to data with many nodes and long history dependence generates computational challenges due to the large number of parameters to estimate. Because brain signals are highly nonstationary, the amount of data available to fit large brain networks is typically limited (Barnett & Seth, 2014; Cohen, 2014), making networks inferred via conditional Granger causality subject to overfitting (Seth, 2010) and highly underpowered (Kelley & Maxwell, 2003). Furthermore, the more parameters included in the model, the longer the computation time required to estimate the model and deduce the functional network.

Common approaches to address this computational challenge rely on reducing the model size, for example, by only performing analysis on a subset of nodes, by downsampling the time series which removes the high frequency content from the signal, or by using smaller model orders chosen via parsimonious model selection techniques (Barnett & Seth, 2014; Barrett et al., 2012; Seth, 2010). However, these approaches are not

always desirable. For instance, a common approach to infer larger networks is to only implement Granger causality pairwise on nodes (Seth, 2010), which may result in networks confounded by indirect influences (Bastos & Schoffelen, 2016; Ding et al., 2006); conditioning the models on the entire observed network reduces the effects of this confound (Ding et al., 2006). Another approach is to limit the size of the history dependence included in the model. Determining the optimal model order for brain signals is a difficult, but important, problem since over- and underspecified models are not informative (Cohen, 2014; Seth, 2010). It has been proposed that the history dependence of brain signals is on the order of tens to hundreds of milliseconds (Barrett et al., 2012; Kabbara et al., 2017; Ringo et al., 1994); using a smaller model order than the true history dependence may result in a poor representation of the brain functional network (Bressler & Seth, 2011). It has additionally been shown that low model orders do not necessarily capture all complexities in the signal spectrum (Bressler & Seth, 2011; Cohen, 2014). To infer accurate and informative functional networks, it is important to incorporate all nodes that are relevant to the particular phenomenon in question, and to include the appropriate history dependence.

In this paper, we propose a method to reduce the number of parameters needed for model estimation in the Granger causality framework, thus permitting inference of larger brain networks with longer history dependence. Existing approaches involve priors assuming that network connections are sparse (Seth, 2010; Valdés-Sosa et al., 2005). Here, instead of imposing assumptions on the structure of the network, we impose assumptions on the shape of the history dependence between nodes, *i.e.*, the coefficients of the lagged

time indices of the history dependent autoregressive (AR) model. The assumption is suitable for brain signals but is less general than the classic AR model. We express the model coefficients in a spline basis, which imposes a smooth structure on the coefficients of the history dependence. In doing so, we assume the coefficients change smoothly and gradually from one time point to the next. We refer to this model as the *spline AR model*, and when applied to determine functional connectivity, the *spline-Granger method*. These serve as extensions, respectively, of what we will refer to as the traditional *standard AR model* and *the standard-Granger method*.

If the spline AR model is in fact an appropriate representation of the signal, there are many benefits to using this method. First, expressing the data in a spline basis reduces the dimensionality of the model, without removing data (*i.e.*, downsampling) and without ignoring potentially critical nodes in the system (*i.e.*, fitting only pairwise on nodes). Also, by using fewer parameters to estimate the same number of observations, we have more data to fit each parameter and thus can expect more statistical power, in the case that the true history dependence of the signal is well represented by a smooth function. That is, we are more precise in our coefficient estimates and are better able to detect small, nonzero effects between nodes. Thus, the spline-Granger method has more statistical power to detect true connections between nodes. Fewer parameters make computing large functional networks feasible; an appropriate ratio of parameters to observations - given stationarity constraints - can be achieved, resulting in more certainty in our model inference (Kelley & Maxwell, 2003). Therefore, the spline-Granger procedure permits inference of more precise and representative networks that cover a larger spatial and temporal scale.

In what follows we compare the performance and accuracy of the spline AR model to the standard AR model on simulations and real data. Through single-node simulations and *in vivo* recordings, we show that the spline AR model can reconstruct signals with accuracy comparable to the standard AR model, while the spline AR model requires fewer parameters to do so. Then, through simulation of nine-node networks, we show how appropriate choice of model order, or amount of history dependence included, improves the accuracy of the networks inferred. Finally, we implement both measures on a 26-node network of cortical data recorded preceding a seizure. We show that, for the 26-node networks, the spline-Granger method successfully infers functional networks for a smaller time window due to estimating fewer parameters and providing greater confidence in the inferences. We show that, because we have more statistical power and precision in estimating the model coefficients when fitting the spline AR models, the spline-Granger method is more sensitive to detecting true edges, or true positives, between nodes while preserving the same false positives.

2.2 Methods

2.2.1 Implementation of the standard AR model and standard-Granger method

We employ the traditional conditional Granger causality measure for directed functional connectivity, which we refer to as the *standard-Granger method* (Barnett & Seth, 2014). This method is a model-based approach to determine if the activity in one brain region - or node - drives activity in another brain region. If the history of activity at node *A* significantly reduces the amount of unexplained variance in the model of activity

at another node B , then we conclude that node A has predictive power over - or drives - node B (Granger, 1969). For example, consider inference of the directed functional connectivity of a three-node network comprised of signals x_t, y_t, z_t , where t indicates a discrete time index. To test the hypothesis that y_t drives x_t , we build two autoregressive (AR) models of the activity of x_t : a full model including the history of all nodes in the network, and a nested (or restricted) model that includes the history of all nodes except that of y_t . We will refer to these models as the *standard AR model*, and the equations are, respectively:

$$x_t = x_{(t-1:t-p)}^T \beta_{xx} + y_{(t-1:t-p)}^T \beta_{xy} + z_{(t-1:t-p)}^T \beta_{xz} + \epsilon_{x,t} \quad (2.1a)$$

$$x_t = x_{(t-1:t-p)}^T \beta_{xx} + z_{(t-1:t-p)}^T \beta_{xz} + \tilde{\epsilon}_{x,t} \quad (2.1b)$$

In these equations, x_t, y_t and z_t are column vectors containing the history dependence of their respective signals, *i.e.*, for a model of order p , $x_{(t-1:t-p)}^T = [x_{t-1} \dots x_{t-p}]$, where T indicates transpose, and $\beta_{xx}, \beta_{xy}, \beta_{xz}$ are column vectors containing the corresponding coefficients of the history dependence, *i.e.*, $\beta_{xx} = [\beta_{xx,1} \dots \beta_{xx,p}]^T$.

Choosing the best model order p is a challenging task as there are several statistically guided techniques, such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which may yield different recommendations. Furthermore, different time windows of data and different pairs of electrodes may also yield different recommendations. Because model order can significantly affect results, we follow the recommendation to fix p across all electrode pairs and time windows (Cohen, 2014). We also choose the same p for both full and nested models as is conventionally

done when estimating causality in neuroscience research (Barnett & Seth, 2014; Bressler & Seth, 2011; Ding et al., 2006; Granger, 1969; Karalis et al., 2016; Luo et al., 2012; Mitra et al., 2018; Place, Farovik, Brockmann, & Eichenbaum, 2016; Schmitt et al., 2017; Seth, 2010; Smith et al., 2015; Uddin, Supekar, Ryali, & Menon, 2011; Zaghera, Ge, & McCormick, 2015). We use the *fitglm* function in MATLAB to determine the model coefficients and compute AIC.

The *standard-Granger method* refers to the procedure for determining a functional connectivity network from multiple signals by fitting the full and nested *standard AR models* in (2.1a) for every signal pair. To determine if the proportion of variance explained by introducing the variables associated with y_t in (2.1a) is significant, we construct an F-test, comparing the residuals of the full and nested models (Barnett & Seth, 2014). The null hypothesis is that the regions are not functionally connected, or that, for our trivariate example, y_t does not significantly improve model residuals. The F-statistic is defined as follows:

$$F = \frac{(RSS_{restricted} - RSS_{full})/p}{RSS_{full}/(N - k * p)},$$

where $RSS_{restricted}$ and RSS_{full} are the residual sum of squared errors for the restricted (2.1a) and full models (2.1a), respectively. The F-statistic accounts for the number of free parameters in each model; p is the model order, k is the number of nodes in the network, and N is the number of observations used to fit the models. The F-statistic is compared to an F-distribution with parameters $d_1 = p$ and $d_2 = N - k * p$.

Both equations in (2.1a) are conditioned on the history of z_t to remove the possibility of a spurious result in the case that z_t acts to confound the relationship between x_t and y_t , often referred to as the common input problem (Bastos & Schoffelen, 2016). That is, suppose z_t has predictive power for both x_t and y_t , and that x_t and y_t are conditionally independent. In this scenario, if we did not include the history of z_t in either of the models, we would likely conclude that y_t drives x_t , or x_t drives y_t , due to the indirect effect of both signals being correlated with z_t . However, by conditioning each model on the rest of the network, here by including the history of z_t , we would correctly identify no direct interaction between x_t and y_t . This example illustrates the importance of conditioning autoregressive models on the network as a whole rather than computing only pairwise comparisons.

In what follows, we compute the F-statistic and its associated p-value for every pair of nodes in the network. When doing so, we use the False-Discovery-Rate (FDR) to control for multiple testing, with $q = 0.05$ (*i.e.*, an upper bound of 5% on the expected proportion of false positives among all declared edges in the inferred network) (Benjamini & Hochberg, 1995).

To estimate confidence bounds for the model coefficients (e.g., β_{xx} in (2.1a)) we employ a bootstrapping procedure. To do so, we use the observed coefficient estimates and their estimated covariance to generate 10,000 normally distributed samples of the coefficients. From the resulting distribution, we determine the 0.025 and 0.975 quantiles of estimated model coefficients. In this way, we use the surrogate distribution to define the 95% confidence interval for the history dependence estimates.

2.2.2 Implementation of the spline AR model and the spline-Granger method

The method we develop in this paper is a modified version of the standard-Granger method, which we refer to as the *spline-Granger method*. The difference between the two methods is that, in the spline-Granger method, we rewrite the standard AR model (2.1a), such that the coefficients of the lagged variables are written in a Cardinal spline basis (Hearn & Baker, 1996). This spline basis fits a smooth curve to the data via local interpolation. It does so by estimating third-degree polynomials between specified points, called knots, in the data such that connections at the knots are smooth. To estimate a third-degree polynomial between two knots, we require that the tangent line at each knot is determined by the slope of the surrounding points whose shape is controlled by a tension parameter, τ , which we set to 0.5. That is, if we are estimating a polynomial, $f(x)$, between the two discretely indexed knots at $x = u_2$ and $x = u_3$ from the set of knots u_1, u_2, \dots, u_ℓ , then we impose: $f'(u_2) = \tau(f(u_3) - f(u_1))$ and $f'(u_3) = \tau(f(u_4) - f(u_2))$. The resulting estimated curves are continuous and have continuous first derivatives. By using a spline basis to estimate the history dependence in the Granger model, we reduce the number of model parameters to estimate from p (the number of history terms in the standard-AR model (2.1a)) to the number of knots. This method allows us to estimate directed functional connectivity with the same duration of history dependence as the standard-Granger model while reducing the number of parameters in the system.

We express the model coefficients in (2.1a) in a spline basis via the transformation $\beta_{ij} = M\alpha_{ij}$, where M is a matrix of size p by ℓ , the number of knots, and whose columns

consist of the spline basis vectors. We will refer to this model as the *spline AR model* with equations:

$$x_t = x_{(t-1:t-p)}^T M \alpha_{xx} + y_{(t-1:t-p)}^T M \alpha_{xy} + z_{(t-1:t-p)}^T M \alpha_{xz} + \epsilon_{x,t} \quad (2.2a)$$

$$x_t = x_{(t-1:t-p)}^T M \alpha_{xx} + z_{(t-1:t-p)}^T M \alpha_{xz} + \tilde{\epsilon}_{x,t} \quad (2.2b)$$

We choose the number of knots in the system, ℓ , to be less than p in the standard-Granger model. We note that for $\ell = p$, the number of knots equals the order, and the history dependent model in the spline basis and standard basis are equivalent.

Many alternative approaches exist to reduce the number of parameters required in a multivariate autoregressive model of neural activity. Common approaches include downsampling the time series data (Cohen, 2014; Seth et al., 2015; Seth, Chorley, & Barnett, 2013) or using regularization techniques (Smith et al., 2011; Valdés-Sosa et al., 2005) which makes the assumption that network connectivity is sparse. Both downsampling and an assumption of sparse network connectivity are inappropriate for some types of neural data, for example seizure activity, during which high frequency rhythms (Frauscher et al., 2017) and dense functional network connectivity are common (Burns, Santaniello, Yaffe, Jouny, & Crone, 2014; Jiruska et al., 2013; Kramer & Cash, 2012; Schindler, Elger, & Lehnertz, 2007). In addition, interhemispheric interactions may occur with prolonged delays (de Pasquale et al., 2010; Ringo et al., 1994), and developing a model that omits these delayed interactions may result in inaccurate functional networks (Bressler & Seth, 2011).

In what follows, we develop an approach to estimate multivariate autoregressive models of neural activity with prolonged history dependence and many nodes yet maintain

a relatively small number of parameters. Many continuous systems can be well approximated with low-order AR models because the system exerts itself instantaneously. However, in neural systems, there exist delays between the field signals that are influenced by the neural mechanisms themselves, such as the neural tissue and myelination of the fibers. We suspect that higher-order AR models are needed to fully capture these delays. Therefore, we choose to use a spline basis to aid in capturing the effects at longer lags, as previously used in the point process framework, where splines have been shown to efficiently capture the effects of a spike on a point process through some delay (Deng et al., 2013, Eden et al., 2012). The spline basis helps us achieve the goal of maintaining the spatial scale, *i.e.*, the size of the network, while extending our temporal scale, *i.e.*, including longer lags.”

To fit the spline AR model equations, we must first choose the location of the knots. Although procedures exist to select knot locations, an optimal knot placement is difficult to determine (Dimatteo, Genovese, & Kass, 2001). We choose here to place knots uniformly spaced every five indices starting at zero. When the sampling rate of the data is 500 Hz, this corresponds to a knot every 10 ms. Because there is missing information at the first and last boundary knots, we include one (invisible) knot at -200 ms before the first time index, and we require that the first derivative of the spline function be zero at the last knot. That is, at the last knot, $x = u_\ell$, we fix $f'(u_\ell) = 0$. In doing so, we assume that the history dependence does not change dramatically at long delays. In Figure 2.1, we plot an example set of basis functions for an example history dependence spanning 60 ms

(sampling rate 500 Hz). We show the eight fitted basis functions (for knots at -200 ms, 0 ms, 10 ms, 20 ms, 30 ms, 40 ms, 50 ms, and 60 ms).

In what follows, we refer to the *spline-Granger method* as the procedure for determining a functional connectivity network from multiple signals by fitting the full and nested *spline AR models* (2.2a) for every signal pair and applying the same F-test as described for the standard-Granger method. We estimate the confidence bounds on the coefficients via the same bootstrapping procedure as described for the standard-Granger method.

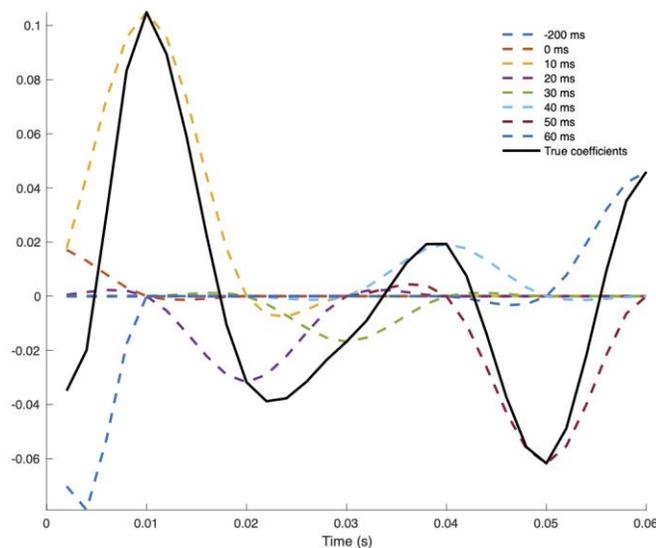


Figure 2.1: Example approximation of history dependence with Cardinal spline basis functions.

Example history dependence (black, solid) spanning 60 ms, and its approximation (gray, solid) using eight Cardinal spline basis functions (colored, dashed) with knots at -200 ms, 0 ms, 10 ms, 20 ms, 30 ms, 40 ms, 50 ms, and 60 ms. A multiplicative factor scales each spline basis function, so that sum of all eight basis functions approximates the history dependence.

2.2.3 Goodness-of-fit: Grenander and Rosenblatt test

As one measure of the model goodness-of-fit, we compare the spectra of the estimated model signals to the observed signal. To do so, we implement a modification of the Grenander and Rosenblatt test of the integrated spectrum, as described in (Priestley, 1981). This test compares the true cumulative spectrum of the signal, $H(\omega)$, with the cumulative spectrum of a new realization of the estimated model $\hat{H}(\omega)$. Because we are most interested in observations that consist of a single, noisy realization of a process, we replace the true cumulative smoothed spectrum with the spectrum estimated from the observed data. In this way, the test compares the observed spectrum with the spectrum simulated from the model. We estimate the spectrum using the multitaper method with a frequency resolution of 2 Hz. Doing so reduces the variability of the spectral estimates, compared to other approaches (Bokil, Andrews, Kulkarni, Mehta, & Mitra, 2010). The 95% confidence intervals for the cumulative spectra are computed using the following two equations:

$$\hat{H}(\omega) - a \sqrt{\frac{8\pi\hat{G}(\pi)}{N}} \leq H(\omega) \leq \hat{H}(\omega) + a \sqrt{\frac{8\pi\hat{G}(\pi)}{N}}$$

$$\hat{G}(\pi) = \frac{1}{4\pi} \sum_{s=-(N-1)}^{N-1} \hat{R}^2(s),$$

where N is the number of samples in the signal from which the spectrum is estimated, $\hat{R}(s)$ is the sample autocovariance function, and $a = 2.2414$ specifies the 95% confidence interval. Grenander and Rosenblatt define the statistic: $k_{GR} = \max_{\omega} \sqrt{N} |\hat{H}(\omega) - H(\omega)|$, which here is the weighted absolute difference between the observed cumulative spectrum and the estimated model spectrum. The model is considered a good fit with 95% confidence

if the value of the statistic k_{GR} is less than 2.2414 (Priestley, 1981). We determine the p-value associated with this statistic using the table in (Grenander & Rosenblatt, 1984), and correct for multiple comparisons using the False Discovery Rate with a significance level of 0.05 (Benjamini & Hochberg, 1995).

2.2.4 Goodness-of-fit: Durbin-Watson test

As a second measure of the model goodness-of-fit, we analyze the model residuals by computing the Durbin-Watson statistic. The Durbin-Watson test checks for serial correlation of the model residuals, as described in (Durbin & Watson, 1950). The Durbin-Watson statistic is defined as:

$$k_{dw} = \frac{\sum_{t=2}^N (e_t - e_{t-1})^2}{\sum_{t=1}^N e_t^2},$$

where e_t is the residual value at time t , and N is the number of observations. To compute a p-value for this statistic, we use the approximation method described in (Durbin & Watson, 1950) and commonly implemented, such as in the Multivariate Granger Causality Toolbox (Barnett & Seth, 2014), and correct for multiple comparisons using the False Discovery Rate with a significance level of 0.05 (Benjamini & Hochberg, 1995).

2.2.5 Generation of synthetic signals: single-node simulations

We generate a signal with a long history dependent structure to examine the performance of the standard AR and spline AR models. To generate the signal, we use an autoregressive model of order 20, AR(20):

$$x_t = x_{(t-1:t-20)}^T \beta_{xx} + \epsilon_{x,t}, \quad (2.3)$$

where we choose the coefficients β_{xx} to create a signal dominated by high frequency activity, $\beta_{xx} = [-0.023, 0.100, 0.050, -0.160, -0.170, -0.160, -0.123, -0.086, -0.008, 0.056, 0.083, 0.079, 0.056, 0.027, 0.005, 0.002, 0.003, 0.013, 0.021, 0.019]^T$. We note that these coefficients establish a relatively smooth history dependence (for example, see Figure 2.3c). The last term in the model, $\epsilon_{x,t}$, is a normal random variable (mean = 0, variance = 0.0625) and t corresponds to a sample every $1/f$ of a second, where f is the sampling frequency. In all simulations, we set the sampling frequency to 500 Hz. To fit the models, we simulate 8 s of data, and analyze the last 2 s (1000 samples) of data to avoid the effects of initial transients. In autoregressive modeling, it is required that the signals are weakly stationary, meaning that the mean and variance do not change over time (Cohen, 2014). To assess the stationarity of our simulated data, we apply to each simulation the KPSS test implemented in the Multivariate Granger Causality toolbox (Barnett & Seth, 2014; Seth et al., 2015). All simulations passed these tests at significance level $\alpha = 0.05$.”

To examine the impact of downsampling the signal as an alternative means of reducing the number of parameters, we first simulate a signal at 500 Hz as described above, and then downsample this signal using the MATLAB function `decimate`, which first low-pass filters the data to prevent aliasing, and then downsamples the signal. We downsample by a factor of 5 such that the new sampling rate is 100 Hz.

2.2.6 Generation of synthetic signals: nine-node simulations

To simulate activity in the nine-node networks, we implement a multivariate autoregressive (MVAR) model of order 30. The model for the nine signals is:

$$\mathbf{u}_t = \sum_{i=1}^{30} \mathbf{B}_i \mathbf{u}_{t-i} + \boldsymbol{\epsilon}_{u,t}, \quad (2.4)$$

where \mathbf{u}_t is a 9 x 1 vector representing the value for each signal at time t , \mathbf{B} is a multidimensional array with dimensions 9 x 9 x p , and $\boldsymbol{\epsilon}_{u,t}$ is a 9 x 1 vector of normal random variables (mean = 0, variance = 0.0625). B_i contains the model coefficients at time $t - i$, where the n^{th}, m^{th} entry, $B_i(n, m)$, is the coefficient of influence of signal m at time $t - i$ on signal n at time t . To generate $B_{1:30}(n, m)$ we first define two smooth functions (black lines in Figure 2.1) as references for the history dependence of nodes connecting to themselves, $n = m$ (self history; Figure 2.1a) and for nodes connected to other nodes, $n \neq m$ (cross history; Figure 2.1b). Then, for each B coefficient in the network (*i.e.*, for all n, m , and i), we add random noise (mean=0, variance=0.01) to each value to create different (noisy) functions for the history dependencies. The resulting values of $B_i(n, m)$ approximate the smooth functions, but are jagged (*i.e.*, not smooth). Example history dependencies for the self-history $B_{1:p}(n, n)$ and cross-history $B_{1:p}(n, m)$, with $n \neq m$ are plotted in Figure 2.2. In this way, we disadvantage the spline-Granger method because the true history dependence violates the spline AR model assumption that the history dependence is smooth.

We implement two simulations with this MVAR model. In the first, we fix the estimated model order $p = 30$, simulate 8 s of data, and analyze the last 2 s of data to avoid the effects of initial transients. In the second simulation, we fit two models with different model orders, $p = 5$ and $p = 30$, and analyze 2 s, 4 s and 8 s of data after simulating an initial 6 s of data to avoid initial transients. In all simulations, we set the sampling

frequency to 500 Hz. All simulations passed the KPSS test for stationarity (Barnett & Seth, 2014).

The single-node and nine-node simulations (include values of \mathbf{B}_i) are provided for reuse and further development at the repository: <https://github.com/erss/spline-granger-causality>.

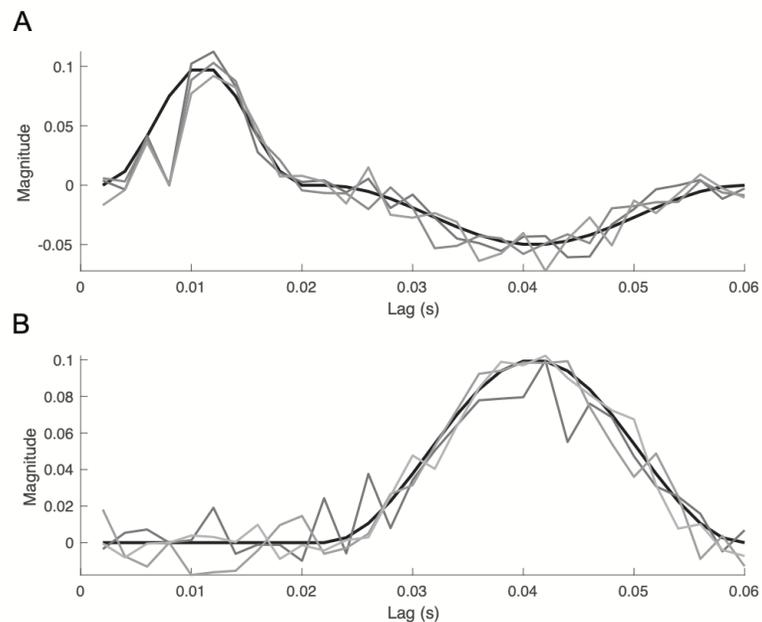


Figure 2.2: Illustration of model coefficients for network simulations.

Plots of example functions used for history dependence in the nine-node simulations. The network model coefficients (gray curves, three examples shown) were created by adding noise to the smooth functions (thick black curves) for the self-influence terms (a), and cross-influence terms (b).

2.2.7 Calculation of network accuracy

We represent the functional networks inferred from the signals as binary matrices in which 1 defines a Granger-caused connection (or edge) between two nodes and 0 defines a non-edge, *i.e.*, where there is no evidence for functional connectivity between two nodes.

We define the accuracy as the proportion of correctly identified edges, *i.e.*, true positives (TP), and non-edges, *i.e.*, true negatives (TN):

$$\frac{TP + TN}{N^2},$$

where N^2 is the total number of edges in the network. We note that here we allow self-connections (*i.e.*, an edge from a node to itself) which represent significant self-history dependence.

We note that computing the accuracy requires knowledge of where edges exist, and therefore only applies to simulated data.

2.2.8 *Calculation of computation time*

All simulations were run on a standard personal computer (64-bit Mac OS X 10.10.5, two Quad-Core Intel Xeon 2.27 GHz CPUs, 32 GB memory). Computation time was calculated using the *tic* and *toc* functions and averaged over multiple simulation trials (as described below) in MATLAB version 8.6.0.

2.2.9 *In vivo recordings from a human subject*

One patient (male age 45 years) with medically intractable focal epilepsy underwent clinically indicated intracranial electroencephalogram (ECoG) recordings for epilepsy monitoring. The recordings were performed using a standard clinical recording system (XLTEK, subsidiary of Natus Medical) with a 500 Hz sampling rate. A two-dimensional subdural electrode array grid (Adtech Medical) was placed on the pia to confirm the hypothesized seizure focus, and locate epileptogenic tissue in relation to essential cortex, thus directing surgical treatment. The reference electrode was a strip of

electrodes placed outside the dura and facing the skull at a region remote from the electrode array grid. Clinical electrode implantation, positioning, duration of recordings and medication schedules were based purely on clinical need as judged by an independent team of physicians without reference to this research. Analysis of these data was performed retrospectively under protocols monitored by the local Institutional Review Boards according to NIH guidelines.

In what follows, we analyze a 10 s interval of these data selected to occur before seizure onset. Before applying the Granger methods, we filter these data to remove 60 Hz line noise (Butterworth filter, order 2, with passband [59,61] Hz). All signals analyzed passed the KPSS test for stationarity (Barnett & Seth, 2014).

2.2.10 Simulation and model fitting code

The single-node and nine-node simulations (including values of β_{xx} in ((2.3) and \mathbf{B}_i in (2.4)) and code for implementing both the spline-Granger and standard-Granger network inference procedures are provided for reuse and further development at the repository: <https://github.com/erss/spline-granger-causality>.

2.3 Results

We begin with simulations to examine the performance of two approaches to history dependent modelling, one in which the model history is estimated in a standard basis, and another in which the history is estimated in a Cardinal spline basis. We show that, in both cases, the spline AR model accurately estimates the history dependence of the simulated signal, with more certainty than the standard AR model, and fits the data well. Next,

through simulation of a nine-node network, we show that the spline-Granger network inference procedure accurately infers the functional connectivity, and has more statistical power to detect true edges. Finally, we apply the model fitting and network inference procedures to an example multi-electrode voltage recording from human cortex. We show that both the spline-Granger and standard-Granger procedures infer similar functional connectivity networks, while the spline-Granger method can be implemented on shorter duration datasets and has more power to detect edges. These results illustrate the accuracy and efficiency of the proposed spline-Granger procedure.

2.3.1 *Simulation: Single Node*

We begin by examining the performance of the spline AR model on two seconds of activity simulated from a single node. To construct the simulated signal, we implement an autoregressive (AR) model of order 20, corresponding to 40 ms, (see Methods; Equation (2.3)) with activity dominated by frequencies in the 15-20 Hz range (Figure 2.3a,b). From this simulated signal, we then estimate history dependent autoregressive models, with history dependence expressed in the spline and standard bases, and include up to 60 ms of history. We choose the estimated history dependence to exceed the true dependence because, in practice, estimating model order from noisy neural data is not straightforward, and a common approach is to choose a model order between 5 and 100 ms (Barrett et al., 2012; de Pasquale et al., 2010; Smith et al., 2011). We estimate the parameters of the history dependent AR model in two ways. First, we estimate the coefficients β_{xx} of the standard AR model at each integer lag, in the univariate case where β_{xy} and β_{xz} are zero

vectors (Equation ((2.1a)a)) (Barnett & Seth, 2014). For 60 ms of history, this corresponds to 30 parameters to estimate where one coefficient is estimated for every lag. Second, we transform the history of the data into a spline basis and estimate the coefficients α_{xx} of the spline AR model, in the univariate case where α_{xy} and α_{xz} are zero vectors (Equation (2.2a)). For 60 ms of history, this corresponds to 8 parameters in the spline basis where coefficients are estimated for each of the 8 basis functions (see example of basis functions in Figure 2.1). Both the standard and the spline AR models estimate 60 ms of history dependence, but the spline AR model does so with fewer parameters.

We find that both models accurately estimate the true coefficients used to generate the simulated data (Figure 2.3c). For both methods, the coefficient estimates approximate the true model coefficients. However, those estimated in the standard AR model lack the smooth dependence that exists in the true model coefficients used to generate the signal, and have large confidence bounds that frequently include 0, meaning there is no evidence of influence at that lag. Visual inspection of Figure 2.3c reveals that the bounds of the estimate from the spline AR model are much tighter, and frequently exclude 0; by using fewer parameters and fitting on the same amount of data, we increase the amount of data used to estimate each parameter. Thus, we increase our effective degrees of freedom, gain statistical power to detect nonzero influences, and gain more certainty in the parameter estimates. We note that the spline AR model accurately captures the non-zero influence of history for broad intervals of time near 0.01 s and 0.03 s, while the standard AR model fails to do so (Figure 2.3c). At larger delays beyond the true model order, the estimated coefficients of both models fluctuate around zero.

To investigate further the performance of both models, we calculate two complementary goodness-of-fit measures that evaluate different aspects of the model's agreement with the data. First, we compare estimates of the integrated spectrum calculated from the simulated signal and estimated from the two history dependent models (Figure 2.3d, see Methods). Visual inspection of Figure 2.3d suggests that both the standard AR model and the spline AR model produce signals with similar spectra to the observed data spectrum. Consistent with this observation, we calculate the Grenander and Rosenblatt test statistic (k_{GR}) and find that both models pass the Grenander and Rosenblatt test in this realization of the process. This test provides no evidence of a significant difference between the spline AR model and the observed signal ($k_{GR} = 1.21, p = 0.45$), and no evidence of a significant difference between the standard AR model and the observed signal ($k_{GR} = 1.89, p = 0.12$). Second, we analyze the model residuals and compute the Durbin-Watson test statistic (k_{DW}) for autocorrelation of the residuals (Figure 2.3e) (Barnett & Seth, 2014). With the exception of the residual at 0 lag, we observe no evidence of significant autocorrelation in the residuals of the standard AR model ($k_{DW} = 2.00, p = 0.99$) or the spline AR model ($k_{DW} = 2.00, p = 0.99$). In this case, we find that both the spline AR model and standard AR model pass both goodness-of-fit tests.

Repeating this analysis for 1000 realizations of the simulated data and model estimates, we find consistent results. The mean and standard deviation of k_{GR} for the standard AR model is 2.28 and 0.79, respectively, and for the spline AR model is 2.21 and 0.81, respectively. A good model fit with 95% confidence is indicated when k_{GR} is below 2.2414 as computed in (Grenander & Rosenblatt, 1984). The standard AR model passed

the GR test in 55% of the realizations, and the spline model passed the GR test in 59% of the realizations (Figure 2.3f). The mean k_{DW} for the standard AR model is 1.998 (95% confidence interval [1.9978, 1.9983]) and for the spline AR model is 1.993 (95% confidence interval [1.992, 1.994]). Both the standard and spline AR models passed the DW test in 100% of the realizations (see Methods; Figure 2.3g) (Barnett & Seth, 2014; Seth, 2010). In addition, we find that the computation time required to estimate the standard AR model (mean 94.0 ms, 95% confidence interval [93.8, 94.2] ms) is significantly greater ($p < 10^{-15}$, t-test) than that required for the spline AR model (mean 61.0 ms, 95% confidence interval [60.6, 61.3] ms, Figure 2.3h). However, we note that the reduction in computation time (approximately 33 ms) is of little practical importance for the univariate data analyzed here. We conclude that both models are rapidly estimated for these univariate data and tend to pass the two goodness-of-fit tests in the same way.

However, we find that the standard AR and spline AR model results differ in two ways. First, for the 1000 realizations of the simulated data, we compute the AIC values for model orders up to 30 for both the spline AR model and the standard AR model (Figure 2.3i). For the standard AR model, we conclude that 16 parameters minimize the AIC, corresponding to a history dependence of 32 ms, while for the spline AR model, we conclude that 4 parameters minimize the AIC, corresponding to a history dependence of 20 ms (knots placed at -200 ms, 0 ms, 10 ms, 20 ms). We conclude that, for these simulated data, the model order that minimizes the AIC is smaller in the spline AR model than the standard AR model; the spline AR model requires fewer parameters to optimize model quality. Second, for the 1000 realizations of the simulated data, we determine the width of

the 95% confidence interval at 10 ms, and how often this interval correctly excludes the null value, for both models. We find that the widths of the confidence bands at 10 ms are tighter for the spline AR model (mean width 0.081, 95% confidence interval [0.075, 0.088]) than for the standard AR model (mean width 0.128, 95% confidence interval [0.123, 0.133]). The tighter confidence intervals of the spline AR model result in more accurate detection of the non-zero effect at this lag. In the spline AR model, the confidence intervals correctly exclude zero for 99.9% of the realizations, while for the standard model only 84.5% of the realizations exclude zero (Figure 2.3j). We conclude that the spline AR model has more power to detect small nonzero influences in the history dependence not detected by the standard AR model.

An alternative approach to reduce the number of parameters when estimating an AR model is to downsample the data (Seth et al., 2015). In some cases, downsampling may be advantageous by allowing longer history dependence with fewer model parameters. However, downsampling below the timescale of interactions between nodes hinders detection of causality, and downsampling typically involves filtering, which is generally ill-advised, as it has been shown to obfuscate underlying causal structure (Barnett & Seth, 2014; Seth et al., 2013). The spline-Granger method with uniformly spaced knots could be considered a type of downsampling. To verify that this is not the case, we examine the standard AR model fit to a downsampled and filtered signal (see Methods). We downsample the data so that the sampling frequency is reduced from 500 Hz to 100 Hz; in the resulting signal, each sample is separated by 10 ms, which matches the interval between knots in the spline AR model (Figure 2.4a). We then fit the standard AR model to the

downsampled data using the same history dependence of 60 ms, as in Figure 2.3. We find that the estimated coefficients for the downsampled signal provide a poor estimate of the true history of the signal (Figure 2.4b). We note that, for this example, the standard AR model fails the GR test (k_{GR} statistic is 7.84, $p < 10^{-15}$) and passes the DW test (k_{DW} statistic is 1.83 $p = 0.22$).

Repeating this analysis of downsampled data for 1000 realizations, we find consistent results. The mean k_{GR} for the standard AR model is 8.25 (95% confidence interval, [8.20, 8.30]), passing the GR test for none of the realizations. The mean k_{DW} for the standard AR model is 1.86 (95% confidence interval [1.86, 1.87]), passing the DW test in 97% of the realizations (Barnett & Seth, 2014; Seth, 2010). Thus, we conclude that the standard AR model is a poor fit to the downsampled and filtered signal, and not equivalent to the spline AR model, which passes both model fitting tests, and more accurately captures the true history dependence of the data. We note that the spline AR model acts to downsample the history dependence, not directly the signal itself.

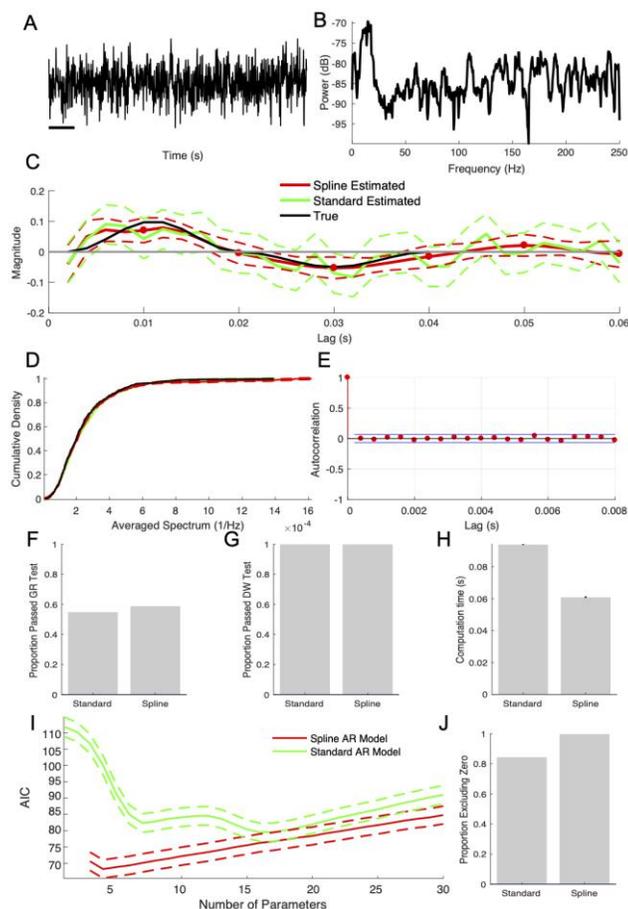


Figure 2.3: Both modeling approaches perform well for a single node with a dominant high frequency spectral peak and long, smooth history dependence.

(a,b) Example trace (a) and spectrum (b) of the simulated signal. Rhythms near 15 Hz dominate the activity. Scale bar in (a) indicates 200 ms.

(c) The true autoregressive coefficients (black curve) estimated using the spline AR model (solid red curve, 95% confidence interval in dashed red), and using the standard AR model (green curve, 95% confidence interval in dashed green).

(d) Illustration of the Grenander and Rosenblatt goodness-of-fit test of the integrated spectrum for the data in (a). The cumulative distribution of the observed signal power (black) and both estimated averaged signals' power (standard AR in green, spline

AR in red) overlap. Dashed lines indicate 95% confidence intervals.

(e) Illustration of the Durbin-Watson goodness-of-fit test for the spline AR model for the data in (a). The autocorrelation of residuals remains small for all nonzero lags.

(f-h) Population results (1000 instances) of the simulation for the (f) Grenander and Rosenblatt statistic, (g) Durbin-Watson statistic, and (h) computation time. The computation time (h) is significantly smaller for the spline AR model; error bars indicate two standard errors of the mean. For both models, approximately 60% of the simulations pass the Grenander and Rosenblatt test (f), and nearly 100% of the simulations pass the Durbin-Watson test (g).

(i) Averaged AIC values for 1000 instances of the simulation for a range of parameters for both the spline AR (red) and standard AR (green) models; solid curves indicate the mean, and dashed curves represent two standard errors of the mean.

(j) Population results (1000 instances) of the simulation for the exclusion of the null value at 10 ms delay. The confidence bounds at 10 ms correctly exclude the null value in approximately 80% of the simulations for the standard AR model, and nearly all simulations for the spline AR model.

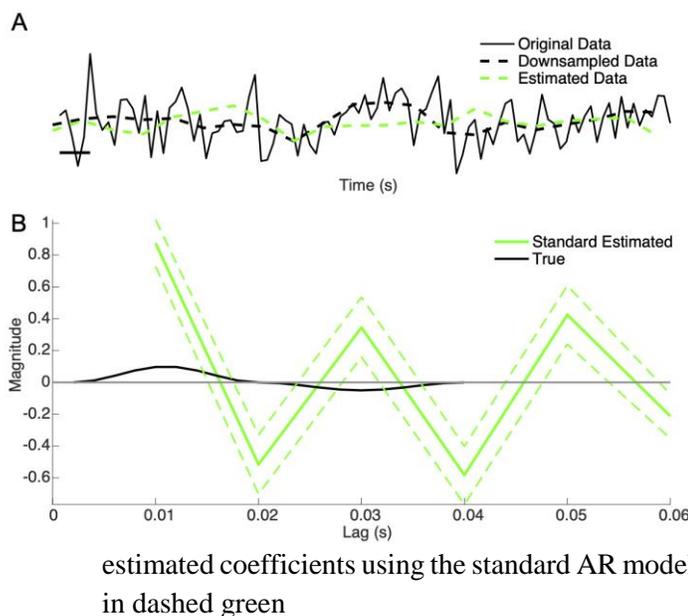


Figure 2.4: The spline-Granger method is not equivalent to filtering and downsampling the signal.

(a) Example trace of a simulated signal (black), filtered and downsampled signal (dashed black) and signal estimated from the downsampled signal using the standard AR model (green). Scale bar indicates 10 ms.

(b) The true autoregressive coefficients (black curve), and

estimated coefficients using the standard AR model (green curve, 95% confidence interval in dashed green)

2.3.2 Simulation: Nine-Node Network

In the previous section, we showed that, in a single node simulation, estimation of an AR model in two ways (at each integer lag, and with a spline interpolation of history) performed similarly; both models adequately reconstructed the data (*i.e.*, tended to pass two goodness-of-fit tests), and accurately captured the history dependence. In this section, we examine the performance of the spline-Granger and standard-Granger methods with a more complicated network of signals. To do so, we simulate a nine-node network using a multivariate autoregressive (MVAR) model with history dependent effects extending to 60 ms (see Methods). In these simulated data, the activity (Figure 2.5a,b) at each node depends on its own history (up to 60 ms), and may depend on the history of activity at other nodes (again, up to 60 ms). We show a representation of the true network connectivity, defined as nonzero influence in the MVAR, in Figure 2.5c. In this figure, we indicate the maximal

influence (across all lags) between nodes; darker colors indicate stronger connectivity between nodes. We expect an accurate network inference to mimic this connectivity pattern.

We begin by estimating the functional connectivity given two seconds of data simulated from the nine-node network. We determine the functional connectivity in the traditional way by implementing Granger causality to assess the predictive power of each node on all nodes (including self-influences) (Barnett & Seth, 2014); see Methods. We do this in two ways: the standard-Granger causality method, and the modified spline-Granger causality method. In the standard-Granger approach, we fit AR models and estimate the history dependence at every time delay for each node, which requires estimation of a large number of parameters; each node in the nine-node network of order 30 (*i.e.*, 60 ms history dependence) requires estimation of 270 parameters (9 nodes * 30 parameters) for the full model fit (see Methods, Equation (2.1a)). In the spline-Granger approach, we fit AR models whose coefficients are written in a lower dimensional spline basis. For this example, we use eight spline basis functions that span the specified 60 ms (see Methods, Equation (2.2a)). Doing so reduces the number of parameters to estimate for each node to 72 (9 nodes * 8 knots).

We fit both the standard AR model (Equation (2.1a)) and the spline AR model (Equation (2.2a)) to the simulated data. For the example data in Figure 2.5a, we find that the inferred spline-Granger networks accurately capture the true network connectivity (Figure 2.5d). In this example for one realization of the network, the accuracy, or the proportion of correctly identified edges and non-edges (see Methods), is 100% for the

spline-Granger network (Figure 2.5d) and is 96.3% for the standard-Granger method (*i.e.*, the inferred functional networks using the standard-Granger method correctly identify 78 of the 81 edges and non-edges in the true network). Repeating this MVAR simulation 1000 times with different noise instantiations, we find that nodes more strongly connected in the true network (Figure 2.5c) appear more frequently in the inferred networks with less variability (mean and standard deviation of networks inferred across the 1000 realizations using the spline-Granger method in Figure 2.5e and f, respectively). Computing the accuracy between the standard- and spline-Granger networks, we find 96.33 +/- 2.00 %. This indicates that functional networks inferred using standard-Granger and spline-Granger are similar.

To examine the impact of model order on network inference, we repeat the MVAR simulation and estimation of the spline and standard AR models with history dependence up to: (1) 10 ms, corresponding to 5 parameters in the standard AR model and 3 parameters in the spline AR model, and (2) 60 ms, corresponding to 30 parameters in the standard AR model and 8 parameters in the spline AR model. We chose these two history dependencies to test the hypothesis that, if the true history dependence of a system is long, then models estimated using a longer history dependence infer more accurate networks. We compare models fit using 5 integer lags in the standard AR, a typical model order chosen for standard AR models in the literature (Barnett & Seth, 2014; Barrett et al., 2012) with models fit using 30 integer lags in the standard AR model. The latter corresponds to the true model order, but is traditionally not computationally feasible for larger-scale biological networks. The goal of these simulations is to compare how network inference is affected by the

amount of history dependence included in the model. In both cases, we infer functional networks from simulated data of duration 2 s, 4 s, and 8 s. For each combination of model order and data duration, we simulate 1000 instances of the MVAR process, and compare the accuracy of the inferred networks and computation time of both the spline-Granger and standard-Granger methods.

We find that both the spline-Granger and standard-Granger methods infer more accurate functional networks when the model order is higher and closer to the true model order for all durations of data analyzed, as expected. For both models, the accuracy is higher when estimating 60 ms of history compared to 10 ms of history (compare pink and gray bars, Figure 2.5g,h). We also observe that, for the longer history dependence of 60 ms, the accuracy increases as the duration of data analyzed increases (compare gray bars in Figure 2.5g,h); additional data provides more evidence to estimate the additional parameters in the model with 60 ms of history. We conclude that the most accurate estimation of the functional network occurs when the model order is large and consistent with the simulated system, and the duration of data analyzed is large.

For the network fitting scenarios with less observed data and more parameters to estimate (*i.e.*, 2 s of data, and a model history dependence of 60 ms), the spline-Granger method is significantly more accurate ($p < 10^{-15}$, t-test; Figure 2.5g,h), capturing 98.69% (mean for 1000 simulations, 95% confidence interval [0.9861, 0.9878]) of the true network, while the standard-Granger captures 96.96% (mean for 1000 simulations, 95% confidence interval intervals [0.9686, 0.9706]). We note that the spline-Granger method produces more accurate functional networks, despite the fact that the true simulated history

dependence was not smooth. This is because the spline AR model has more statistical power to correctly infer edges, *i.e.*, correctly reject the null hypothesis that there is no correlation between two nodes. The difference is more pronounced when there is less available data to estimate the model parameters.

To illustrate further the increased statistical power of the spline AR model to detect edges, we compare the estimated coefficients from the standard AR and the spline AR models for an example node pair at which the two methods produced different results; the standard-Granger method failed to detect an edge, while the spline-Granger method correctly detected the edge (Figure 2.5i). In this case, consistent with the single node simulation (Figure 2.3c), the confidence bounds on the coefficients of the standard AR model fluctuate around zero and are much larger than those of the spline AR model. Thus, there is not enough evidence to reject the null hypothesis in the standard-Granger method, while in the spline-Granger method the decreased uncertainty in the model coefficients correctly identifies the nonzero effects, resulting in an edge between the two nodes.

For both methods, the computation time increases as the model order and duration of data analyzed increase (Figure 2.5j,k). However, the computation time for the spline-Granger method is less than that required for the standard-Granger method for all combinations of model order and data duration considered ($p < 10^{-15}$ in all cases). We note that this difference is most pronounced when the model order and data duration are large, *i.e.*, in the scenario of the most accurate network inference. Specifically, for fitting models with 60 ms of history using 8 s of data, the standard-Granger method takes 72% longer to compute on average compared to the spline-Granger method.

This example highlights the utility of the spline-Granger approach; the method permits more accurate network inference because (1) we can fit models with longer history dependencies – which are more reflective of neural systems – with fewer parameters, and (2) we gain more statistical power to correctly identify edges.

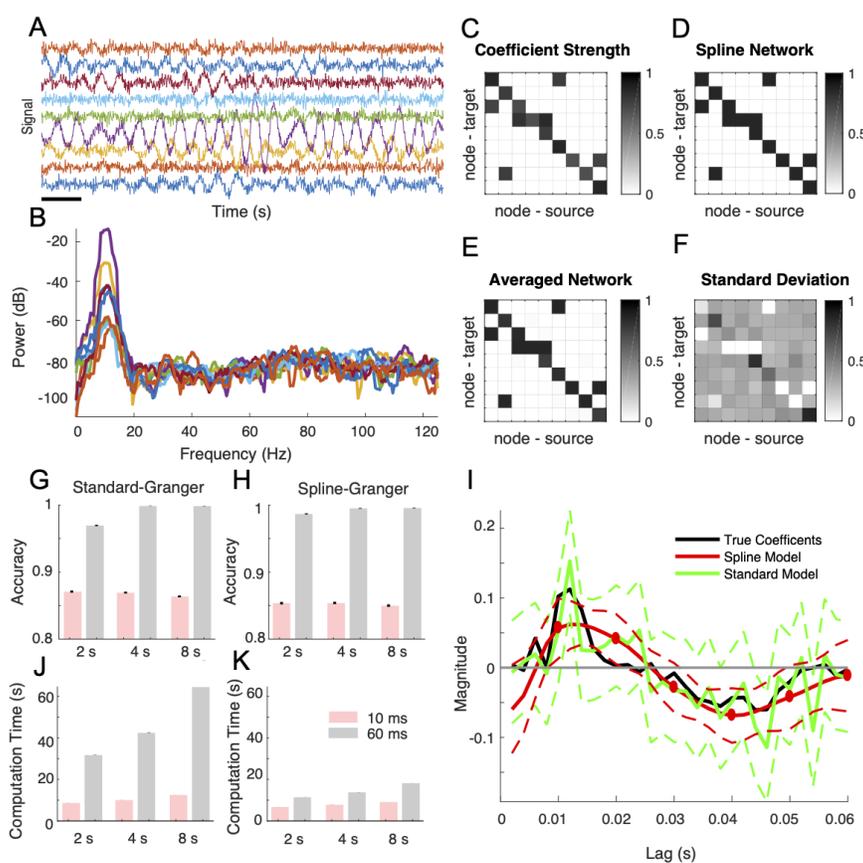


Figure 2.5: Increases in estimated model order produce accurate functional networks in minimal computation time using the spline-Granger method.

(a,b,c) Example simulated signals (a) and spectra (b) of the nine-node network with the connectivity strength between nodes shown in (c). Scale bar in (a) indicates 200 ms.

(d,e,f) Corresponding networks inferred using the spline-Granger method (d), the average network estimated from the spline-Granger method

with history dependence 60 ms (e), and the variability of those estimates (f) for 1000 realizations of the simulated network data. In each figure, darker shades indicate larger values.

(g,h) Accuracy for 1000 realizations of the nine-node network for the standard-Granger (g) and spline-Granger (h) models with history dependence extending to 10 ms (red) and 60 ms (gray), and data of duration 2 s, 4 s, and 8 s.

(i) An example of the estimated history dependence for which the standard-Granger method fails to detect an edge, while the spline-Granger method detects an edge. Inferred coefficients for standard AR model (green) and spline AR model (red) and true model

coefficients (black); solid curves indicate the mean, and dashed curves indicate the 95% confidence intervals.

- (j,k) Computation time for 1000 realizations of the nine-node network for the standard-Granger (j) and spline-Granger (k) models with history dependence extending to 10 ms (red) and 60 ms (gray), and data of duration 2 s, 4 s, and 8 s.

2.3.3 *In vivo Data: Single Node*

To examine the performance of the spline AR model on physiological data, we first consider *in vivo* brain voltage activity recorded from a single electrode. We show in Figure 2.6a an example voltage trace from human cortex at a time preceding seizure onset. The power spectrum of these data exhibits a $1/f$ structure, common in recordings of brain voltage activity (Figure 2.6b) (He et al., 2010). Although the true history dependence is unknown for these *in vivo* data, we may still compare the performance of the spline AR model with the standard AR model. First, we compute the AIC values for both models (Figure 2.6c). We find for model orders up to 30, no absolute minimum occurs, as expected for $1/f$ signals which theoretically require infinite model order (Tang et al., 2017). Because AIC is not an informative way to select model order for brain signals of this type, and other model selection procedures tend to produce variable results (Cohen, 2014), we choose to fit the univariate spline AR and standard AR models with history dependence up to 40 ms, which is within the range of history dependencies reported in brain signals (Barrett et al., 2012; de Pasquale et al., 2010; Smith et al., 2011). In the standard AR model this corresponds to estimating 20 parameters at each integer lag, and in the spline AR model this corresponds to estimating 6 parameters for the coefficients of 6 spline basis functions (with knots at -200 ms, 0 ms, 10 ms, 20 ms, 30 ms, and 40 ms). The coefficients inferred for both models are similar (Figure 2.6d), but the coefficients for the spline AR model are

smoother, as expected. Consistent with the single node simulations, we find tighter confidence bounds around the estimated coefficients for the spline AR model, and significant effects in the history dependence. The estimated coefficients of the standard AR model fluctuate around zero starting at and after 7 ms, whereas in the spline AR model, we find weak but significant nonzero effects up to 20 ms. The ability to detect these weak effects results from the gain in statistical power in the spline-Granger model, which uses the same amount of data to estimate fewer parameters. The k_{DW} statistics for the standard AR and the spline AR models are 2.00 ($p=0.97$) and 2.00 ($p=0.96$), respectively, indicating no evidence for autocorrelation of the model residuals. We find that both the standard AR and the spline AR models fail the GR test ($p < 10^{-15}$) in both cases the models fail to capture all of the observed signal power at low frequencies. Comparing the structure of the model coefficients and the model residuals, we conclude that both methods similarly reconstruct the data, although the spline-Granger method provides more statistical power to detect weak – but significant – effects.

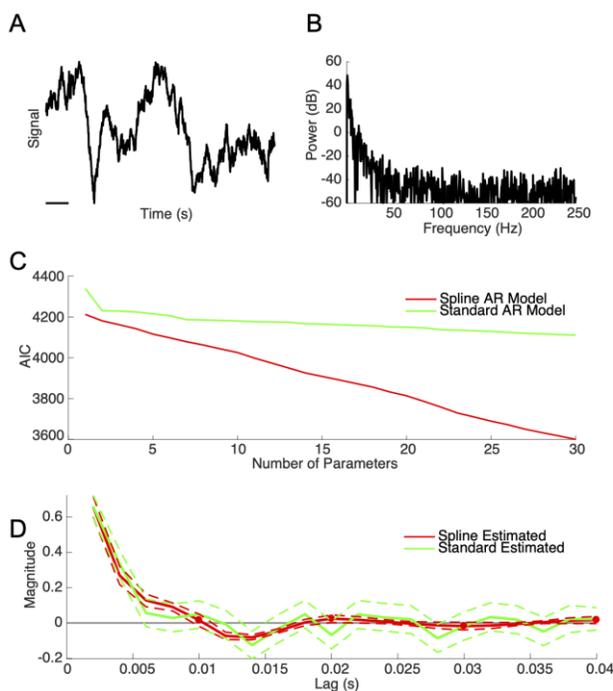


Figure 2.6: The spline AR and standard AR models produce similar results for in vivo data recorded from a single node.

(a,b) Example signal (a) and spectrum (b) recorded from human cortex preceding a seizure. Black bar in (a) indicates 200 ms.

(c) AIC computed for both the spline AR and standard AR models up to order 30.

(d) Fits to the history dependence using the spline AR model (red) and standard AR model (green); solid curves indicate the mean, and dashed curves indicate the 95% confidence intervals.

2.3.4 In vivo Data: Multiple Nodes

Finally, we apply both network inference methods to *in vivo* brain data recorded from a 26-electrode array. We chose to use only 26 signals of an 8 x 8 electrocortigraphy grid because while still a sizable number of signals, it is small enough to fit, and thus compare, both methods. Typically, computing networks of this scale using the standard-Granger approach is infeasible due to the limited amount of stationary data available relative to the number of parameters required for model inference. To demonstrate the utility of the spline-Granger approach, we fit models using 2 s and 10 s of data, and a model history dependence up to 40 ms. For 2 s of data, estimation of the standard AR model is not appropriate; the number of parameters to estimate for each node ($26 \text{ nodes} * 20 \text{ lags} = 520$) is nearly half the number of observations per node ($2 \text{ s} * 500 \text{ Hz} - 20 \text{ lags} = 980$ observations). In general, a good model fit requires a ratio of the number of observations

to the number of parameters much greater than 1 (Harrell, 2001). For the standard AR model, this ratio is 1.9, and too small to support confident results. However, because the spline AR model requires estimation of only 156 parameters for each node (26 nodes * 6 knots), this ratio is 6.3, and large enough to perform the model fit. We note that estimating the functional network using the spline-Granger method requires 2.90 minutes of computation time (Figure 2.7a).

While, ideally, we would not fit AR models on 10 s of data due to the nonstationary nature of brain signals, we do so here as a means of comparison for both network inference procedures; despite expecting the underlying brain system generating the observed activity to lack stationarity over a 10 s interval, we note that all signals analyzed passed the KPSS test for stationarity. For 10 s of data, the number of observations is sufficient to infer network structure with both methods. We find that, while the standard-Granger method requires 25.4 minutes to compute (Figure 2.7b), the spline-Granger method requires 5.93 minutes (Figure 2.7c). Excluding the self-influence terms, both methods infer similar functional networks; the overlap in edges and non-edges is 93.1%. Additionally, we note that the spline-Granger method infers more edges than the standard-Granger method (compare Figure 2.7b and c). As we illustrated previously, we compare the estimated coefficients for an instance in which the standard-Granger method fails to detect an edge and the spline-Granger method detects an edge (Figure 2.7d). Although both methods detect a similar mean effect, the 95% confidence bounds on the estimated coefficients for the spline AR model are tighter than those for the standard AR model, consistent with the simulation results from the nine-node network (Figure 2.5i). Thus, the influence between

the two signals is nonzero, and we infer that the two nodes are functionally connected. We conclude that the reduction in parameters when using spline-Granger method permits larger functional networks to be fit using shorter durations of data, often required when analyzing non-stationary data, and that it has more statistical power to detect edges.

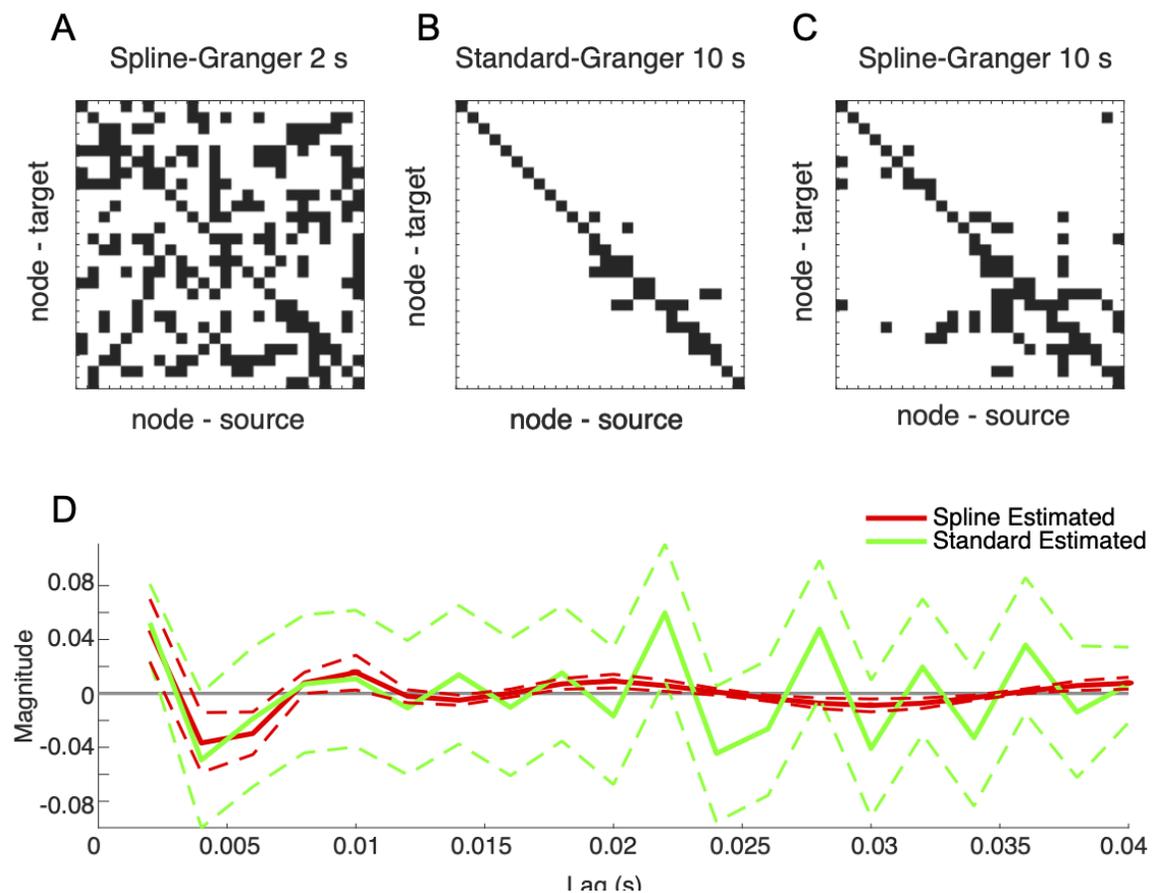


Figure 2.7: The spline-Granger method makes inference of a 26-node network computationally tractable and more sensitive.

- (a) Network inferred applying the spline-Granger method to 26 voltage recordings of duration two seconds. The standard-Granger method requires estimation of too many parameters given the amount of data observed.
- (b,c) Networks inferred using standard-Granger (b) and spline-Granger (c) methods applied to voltage recordings from 26 electrodes of duration ten seconds.
- (d) Estimated history dependence for a signal pair for which a connection was detected using the spline-Granger method (solid red curve, 95% confidence interval in dashed red), but

not using the standard-Granger method (solid green curve, 95% confidence interval in dashed green).

2.4 Discussion

Granger causality is a powerful approach to infer functional networks from multi-sensor recordings of brain activity. However, the standard-Granger approach requires estimation of many model parameters, limiting its applicability to small or otherwise restricted networks. To address this limitation, we developed a modification of the standard-Granger approach by assuming the model coefficients of the time lagged history dependent terms are smooth functions. This assumption reduces the number of model parameters to estimate and makes inference of functional networks from stationary intervals of multi-sensor recordings computationally tractable.

In univariate and multivariate simulations, we showed that the spline AR model accurately reconstructs the signals and closely estimates the true model parameters, and that the spline-Granger method accurately infers functional network structure. We also showed that when the true history dependence of a system is long (e.g., 60 ms, consistent with brain signals), accuracy of functional network inference improves when models approximate the true history dependence of the system. We applied the proposed method to *in vivo* brain data and found that a univariate brain signal can be modeled by history dependencies that extend to 40 ms, and that multivariate functional networks can be inferred using the spline-Granger method on shorter intervals of data. Additionally, in the simulations and *in vivo* brain data, we found that using the same amount of data to estimate fewer parameters in the spline-Granger approach produced more precise parameter

estimates, *i.e.*, smaller confidence bounds around the parameter estimates, and ultimately detection of weak – but present – connections in the functional networks.

Two challenges limit the applicability of the standard-Granger approach to large brain networks. First, the computation time required for the standard-Granger model rapidly increases with network size. Second, brain signals are highly nonstationary which limits the amount of data and leads to statistically underpowered models. The problem of computation time may be addressed through brute force approaches given powerful computers and efficient algorithms. The problem of statistically underpowered models may be addressed through appropriate choice of model. Combining too many model parameters with too few observations leads to Type II errors, meaning many true functional connections between nodes will be missed (Kelley & Maxwell, 2003). By assuming that the history dependence of the model is smooth, we express the MVAR models in a lower dimensional basis, thus reducing the number of parameters to estimate. Compared to the standard AR model, the spline AR model proposed here reduces the number of parameters by a factor of approximately 5, allowing reliable models estimates from shorter data intervals, in the scenario that our model assumptions are appropriate, and reducing computation time.

Because Granger methods are model-based, there are many ways to assess goodness-of-fit and confidence in the model results. Since different goodness-of-fit tests assess different aspects of model performance, we chose to implement two tests here. The Durbin-Watson statistic analyzes the model fit in the time domain by checking for serial correlations in the model residuals of the reconstructed signal. The Grenander-Rosenblatt

statistic analyzes the model fit in the frequency domain. This test is derived from stochastic process theory and compares a known spectrum to a model estimated spectrum. Ultimately, this test assesses the model's ability to generate signals with appropriate spectral properties. We chose these two methods because the Durbin-Watson test is commonly used in MVAR modeling (Barnett & Seth, 2014) and because the Grenander-Rosenblatt test assess the spectrum - an important characteristic of brain data (Buzsáki & Draguhn, 2004).

In the case of real data, there is no ground truth to assess which model performs more accurately. Therefore, we test our models on simulated datasets that are multivariate and include an extended timescale of interactions between nodes, consistent with *in vivo* brain activity. Motivated by these simulated results, and our analysis of the ECoG data, we speculate that the spline-Granger method supports more accurate and representative functional networks. By allowing estimation simultaneously from more brain regions, the method permits inference of larger functional networks, *i.e.*, networks with more brain regions or nodes, and therefore accounts for more indirect interactions. And, by permitting estimation with longer history dependence, the method captures longer scale brain dynamics. The fundamental modification in this model compared to the standard-Granger approach is the smooth history dependent structure. The benefit of the spline AR model is that spline functions are sufficiently flexible to reflect variations in the history dependence while using fewer parameters than the standard AR model. We emphasize that neither method is more correct, but that the spline-Granger method can operate on more brain regions with smaller intervals of data than the standard-Granger method. In the case that the model assumptions are met, the effective degrees of freedom in the spline AR model

are greater, giving the model more statistical power to detect weak interactions. However, we note that this assumption is poor when the history dependence changes quickly, such that the history dependence is not smooth, and that this assumption is not particularly useful when the true history dependence is short.

While we cannot know the true history dependence that drives observed neural signals, nor necessarily interpret the Granger coefficients as a representation of the underlying mechanisms (Barrett & Barnett, 2013), there exist two observations that motivate an assumption of smoothness for the model coefficients. First, comparing the standard AR to the spline AR model coefficients fit to in vivo data (Figure 2.6 and Figure 2.7), the standard AR coefficients fluctuate around the spline AR coefficients. While these fluctuations may contain meaningful information, we hypothesize that – instead – the rapid fluctuations of the standard AR coefficients represent non-informative noise. By smoothing these rapid fluctuations, the spline AR model both reduces the impact of this noise and requires less parameters to estimate. Second, in the nine node simulations, we designed a system such that critical information appeared at longer lags, and the history dependencies between signals were not smooth. Yet, despite an incorrect assumption of smoothness, the spline-Granger model accurately captured the network structure. In this case, the omission of rapid fluctuations in the history dependence did not impact network inference. Importantly, because the spline-Granger model required fewer parameters, this approach also permits more accurate network inference on a smaller duration of data compared to the standard-Granger model. Ultimately, our analysis suggests that the spline basis acts as a reasonable approximation for the system.

In our model selection procedure, there remain three challenges: determining the model order, the choice of knot placement, and the number of knots. Common approaches to determining model order, such as the Akaike information criterion (AIC), Bayesian information criterion (BIC), and partial autocorrelation functions, aim to select the most parsimonious model. However, parsimony may not be desirable; in brain signals, lags up to 100 ms can drive the network (Barrett et al., 2012; Smith et al., 2011; Tang et al., 2017) and models that omit these influences may miss important features. We also note that AIC in particular may be a poor choice for brain data because neural signals are dominated by pink noise, which requires an infinite model order (Tang et al., 2017). Determining the optimal model order for a multivariate system, rather than a univariate signal, is an even more challenging task. For example, the optimal model order may differ for each signal (Cohen, 2014). Because of the smoothness induced by the spline basis, small fluctuations in the history dependence near zero are damped to zero, consistent with a weak neural influence. We speculate that the spline bases could be used *a posteriori* to determine model order as the last non-zero term in the history dependence.

The transformation into a spline basis also depends heavily on the choice of knot placement and the number of basis functions. We chose to place the knots uniformly every 10 ms to standardize our models across our analysis and to limit how much prior information we impose on the models. Other work has explored principled approaches, including Bayesian techniques, to determine the optimal number and location of splines (Dimatteo et al., 2001), which may potentially yield more accurate models.

Continuing research to reduce the number of parameters, and choose appropriate models that reflect the system, in these approaches remains an important goal to accurately infer functional networks. Ideally, studies are designed to detect features with enough power to obtain confidence intervals - of any size - that correctly exclude the null value. However, when the feature values are nonzero, but small, model estimates need to be more precise, *i.e.*, the confidence intervals need to be small, to correctly exclude the null value. Thus, not only is it important to design studies that have enough power to determine the significance of parameters, but also to apply methods that support more precise parameter estimates. By obtaining tighter confidence bounds, we achieve more accurate parameter estimates and thus obtain more power to correctly identify weak, but present, functional connections between nodes (Kelley & Maxwell, 2003). Approaches exist to design studies that achieve both a desired statistical power level and a desired confidence band width, such as Accuracy in Parameter Estimation analysis (Kelley & Maxwell, 2003). Knowing the optimal number of parameters to achieve desired power and precision might provide insight when choosing the number of nodes and knots to include in spline-Granger network models. For instance, if the optimal number of parameters can be predetermined, these parameters could then be distributed uniformly across all node pairs in the network.

In conclusion, the spline-Granger method provides a flexible and useful tool for network inference of large models. Because accurate network inference is limited by the stationarity of the data available, we develop an approach that reduces the number of parameters to estimate. The number of parameters in the model is a function of the number of brain regions included and the amount of history dependence estimated between these

brain regions. To account for the possible confounding influences of indirect interactions between brain regions, we develop an approach that does not reduce the number of brain areas, but instead simplifies the estimated history dependence. We showed that we gain statistical power and precision in our parameters estimates at the cost of making assumptions on how the influence of past values changes over time. By fitting more statistically powerful models with more nodes and longer history dependence, the method can produce more precise and more informative functional networks.

2.5 Acknowledgements

This work was supported by the National Science Foundation DMS #1451384, and the National Institutes of Health NINDS R01NS095369.

3 CHAPTER THREE

Source estimates reveal regional spindle deficits and predict cognitive symptoms in Rolandic epilepsy³

ABSTRACT

Rolandic epilepsy is the most common form of epileptic encephalopathy, characterized by seizures and cognitive deficits in school-age children that spontaneously resolve by adolescence. The pathophysiology of this disease is unknown, but growing evidence suggests that transient disruption of thalamocortical circuits could cause both the stereotyped, focal, sleep-activated epileptiform spikes observed in the inferior Rolandic cortex, and a paucity of sleep spindles, physiological rhythms associated with sleep-dependent learning. We utilized electrical source imaging to study the distribution of spindle deficits in this disease and the relationship with cognitive symptoms. In specific, we hypothesized that: 1) spindle rate would be decreased in the inferior Rolandic cortex in active epilepsy, 2) spindle deficits would extend beyond the epileptic cortex in active epilepsy, and 3) regional spindle deficits would better predict cognitive dysfunction than focal estimates. We obtained high-resolution MRI, high-density EEG recordings, and focused neuropsychological assessments in children with Rolandic epilepsy during active

³ Chapter 3, in full, is a reprint of the following article as it appears under review: Spencer, E. R., Chinappen, D., Emerton, B., Morgan, A. K., Emerton, B., Manoach, D. S., Eden, U. T., Kramer, M. A., & Chu C. (2021). *Source estimates reveal regional spindle deficits and predict cognitive symptoms in Rolandic epilepsy*, under review.

($n=8$, age 9-14.7 years, 3F) and resolved ($n=10$, age 10.3-16.7 years, 1F) stages of disease and age-matched controls ($n=8$, age 8.9-14.5 years, 5F). We computed spindle rates using a validated spindle detector applied to the source activity in the inferior Rolandic cortex as well as each label in the Desikan-Killiany cortical atlas during non-rapid eye movement sleep. The results were compared across groups and to measures of fine motor dexterity, phonological processing, attention, and global intelligence (IQ) using generalized linear models and a bootstrap analysis of goodness-of-fit. Among detected spindles, we also compared spindle features (power, duration, coherence, bilateral synchrony) between groups. We found that spindle rate was reduced in the inferior Rolandic cortices in active Rolandic epilepsy (active $P=0.007$; resolved $P=0.2$) compared to controls. We found statistically significant decreased bilateral synchrony between spindles in the active group ($P=0.005$) but not resolved ($P=0.1$) compared to controls and no difference in the other tested spindle features. Compared to controls, spindle rate in the active group was also found to be reduced in the prefrontal, insular, superior temporal, and posterior parietal regions ($P<0.009$ for all). Regional spindle rates positively correlated with cognitive function ($P<1e-4$) and provided improved estimates of fine motor dexterity compared to focal estimates from the inferior Rolandic cortex alone ($P=0.001$). These results provide evidence for a regional disruption to the thalamocortical circuit in Rolandic epilepsy and a potential mechanistic explanation for the broad cognitive deficits observed in children with this epileptic encephalopathy.

3.1 Introduction

Rolandic epilepsy, previously known as benign epilepsy with centrotemporal spikes, is the most common form of childhood epileptic encephalopathy, characterized by epileptic spikes and seizures arising from the inferior Rolandic cortex during non-rapid eye movement (NREM) sleep and neurocognitive deficits ranging from subtle to severe in school-age children (Berg et al., 2010; Besenyei et al., 2012; Carvill et al., 2013; Tovia et al., 2011). Common cognitive deficits grossly localize to frontoparietal circuits and include sensorimotor dysfunction, attention-regulation difficulties, and phonological processing difficulties (Callenbach et al., 2010; Katewa & Parakh, 2015; Vannest, Tenney, Gelineau-Morel, Maloney, & Glauser, 2015; Wickens et al., 2017). This disease is self-limited, in that by adolescence, seizures spontaneously resolve. Additionally, the neurocognitive deficits are also transient, where formal neuropsychological testing identifies symptoms in most children tested within approximately five years of onset (Wickens et al., 2017) but cannot detect deficits when children are tested nine years after diagnosis (Ross, Stoyell, Kramer, Berg, & Chu, 2020). Importantly, cognitive deficits are observed in children independent of anticonvulsant treatment status (Wickens et al., 2017).

While epileptic spikes provide a robust biomarker of seizure risk, the neurophysiological basis for cognitive deficits in this epileptic encephalopathy remains largely unknown. The activation of epileptic spikes during NREM sleep characteristic of Rolandic epilepsy (Carvill et al., 2013; Katewa & Parakh, 2015) suggest involvement of the thalamus, a prominent brain nucleus involved in synchronizing and regulating sleep rhythms (De Gennaro & Ferrara, 2003; Gent, Bassetti, & Adamantidis, 2018). This

hypothesis has been further supported by recent work identifying abnormal thalamocortical white matter connectivity to the Rolandic cortex (Thorn et al., 2020) and a paucity of sleep spindles, characteristic 10-15 Hz rhythms produced during NREM sleep, in central regions on scalp EEG (Kramer et al., 2021). Sleep spindles are generated and amplified within the thalamocortical circuits and have been associated with sleep-dependent memory consolidation, and general cognitive functioning (Beenhakker & Huguenard, 2009; De Gennaro & Ferrara, 2003). Epileptiform spikes are anticorrelated with spindles in Rolandic epilepsy (Kramer et al., 2021), suggesting a competitive relationship, whereby spikes may commandeer spindle thalamocortical circuitry (Beenhakker & Huguenard, 2009). Because the thalamus is comprised of nuclei with distinct cortical projections (Bastuji, Lamouroux, Villalba, Magnin, & Garcia-Larrea, 2020; Behrens et al., 2003; Fama & Sullivan, 2015), thalamocortical circuit dysfunction may lead to abnormal physiology widespread across cortex. Although in Rolandic epilepsy the epileptiform spikes have been well-localized to the inferior Rolandic cortices (Mirandola et al., 2013; Ross et al., 2020), the spatial extent of the spindle disruption and the relationship to cognitive function is not known.

Prior work evaluating sleep spindles in Rolandic epilepsy utilized standard scalp EEG data, which is limited in spatial resolution due to skull blurring and inconsistent electrode placement across subjects (Kramer et al., 2021). To better evaluate the spatial extent of the spindle deficit in Rolandic epilepsy, we utilized co-registered high-density EEG, high-resolution MRI, digitized electrode coordinates, and a validated biophysical electrical source imaging (ESI) model to estimate cortical signals (Hamalainen & Sarvas, 1987, 1989). To relate these findings to cognitive function, we evaluated performance on

tasks targeting sensorimotor, attention, phonological processing, and global intellectual (IQ) skills, the canonical challenges reported in Rolandic epilepsy (Scheffer et al., 2017; Vannest et al., 2015; Wickens et al., 2017). We hypothesized that: 1) spindle rate would be decreased in the inferior Rolandic cortex, 2) spindle deficits would extend beyond the epileptic cortex, and 3) regional spindle deficits would better predict cognitive dysfunction than focal estimates. Identification of regional spindle deficits in Rolandic epilepsy may provide an improved biomarker and mechanistic explanation for the variable cognitive deficits observed in children with this epileptic encephalopathy and evidence for a regional disruption to the thalamocortical circuit in this disease.

3.2 Materials and methods

3.2.1 Subject data

Children with a documented EEG showing sleep-activated centrotemporal spikes and a clinical diagnosis of Rolandic epilepsy by a child neurologist ($n=18$, ages 9-16.7, 4F) and control subjects ($n=8$, age 8.9-14.5 years, 5F) were recruited to participate. The epilepsy subjects were further divided into two groups based on seizure risk (Ross et al., 2020): active disease ($n=8$, age 9-14.7 years, 3F), and resolved disease ($n=10$, age 10.3-16.7 years, 1F).

This research was approved by the Massachusetts General Hospital and Boston University institutional review boards, and assent and informed consent were obtained from each subject and guardian.

3.2.2 *Electrical source imaging and minimum norm estimation*

Each subject underwent an EEG recording session, MRI recording session, and assessment of neuropsychological function. MRI and neuropsychological evaluations were separated from the EEG by 3.8 days (range 0-36) and 28.6 days (range 0-142), respectively. EEG data (70 channel cap based on the 10-10 electrode placement system with additional electrodes at T1 and T2, 2035 Hz sampling rate (Easycap, Vectorview, Elekta-Neuromag, Helsinki, Finland)) collected during stages 2 and 3 NREM sleep epochs, when spindles are present and epileptiform spikes are activated, were selected for analysis (mean duration 811.9 min, minimum duration 63.7 min, maximum duration 2644.2 min).

MRI data acquisition included T1-weighted multi-echo magnetization-prepared rapid acquisition gradient-echo (MEMPRAGE) images that were collected on a 3 T MAGNETOM Prisma Scanner (Siemens, Germany) with the following parameters: TR = 2,530 ms, TE = (1.69, 3.55, 5.41, 7.27 ms), voxel size 1x1x1 mm, flip angle =7 degrees.

Source analysis of EEG data was performed using the MNE-C software package (Gramfort et al., 2014; Hamalainen & Sarvas, 1989). Briefly, MNE provides a distributed source estimate of cortical currents incorporating constraints from the patients' MRI, transforming the data to brain space without requiring heuristic choices or strong assumptions about the sources (Chu et al., 2015). Continuous EEG data were acquired at a sampling rate of 2035 Hz after bandpass filtering (low pass cutoff frequency of 671.55 Hz, filter order 16384). Subsequently, channels with no signal or high noise and periods of artifact were identified through visual analysis by an experienced electroencephalographer and manually removed.

EEG electrode positions were digitized prior to recording using a 3D digitizer (Fastrak, Polhemus Inc., Colchester, VA). Anatomical cortical surfaces of the brain were reconstructed using FreeSurfer from the MEMPRAGE data (Fischl, 2012). Digitized electrode coordinates were aligned to the MEMPRAGE data using the nasion and auricular points as fiducial markers (Figure 3.1A).

For the forward model, a three-layer boundary element model consisting of the inner skull, outer skull and outer skin surfaces was generated using the watershed algorithm in FreeSurfer (Figure 3.1B). The digitized EEG electrode coordinates were co-registered to the reconstructed surface using the nasion and auricular points. Cortical surfaces were parcellated using FreeSurfer to identify the regions of interest within each subject. To generate the inferior Rolandic cortex label within each subject, a sphere centered on the most inferior vertex in the pre- or post-central gyrus with a radius equal to half of the distance between the most inferior and most superior vertices in the pre- or post-central gyrus was generated. The overlap between this sphere and the pre- and post-central gyrus labels was the inferior Rolandic cortex ROI (Song et al., 2019) (Figure 3.1C).

For each subject, 10,242 source space points per hemisphere were computed using the topology of a recursively subdivided icosahedron fitted to the cortical surface inflated to a sphere. A similar procedure was followed to produce a lower-density model with 162 source space points per hemisphere. The inverse operator was computed from the forward solution with a loose orientation constraint of 0.6 to eliminate implausible sources and 2 microvolts as the estimate of EEG noise. The normal component of dipoles at each source space point were used for source data estimates.

For each subject, we calculated the activity at each of the 162 source space points as follows. First, a circle of approximately 1 cm diameter on the cortical surface was drawn around the source space point using the full-width half-max smoothing kernel (Fischl, 2012). Then, the mean activity of the high-density source space points within this circle was computed; this mean activity defined the average source space solution for the low-density source space point. Example source data estimates during a Rolandic epileptic spike are shown in Figure 3.1D.

The source activities were down-sampled to 407 Hz using MATLAB's function *decimate*. We restricted our initial analysis to sources in the inferior Rolandic cortices, which are the cortical origins of the epileptic spikes in Rolandic epilepsy subjects. For regional analysis, we evaluated all cortical labels produced using the Desikan-Killiany atlas (Desikan et al., 2006).

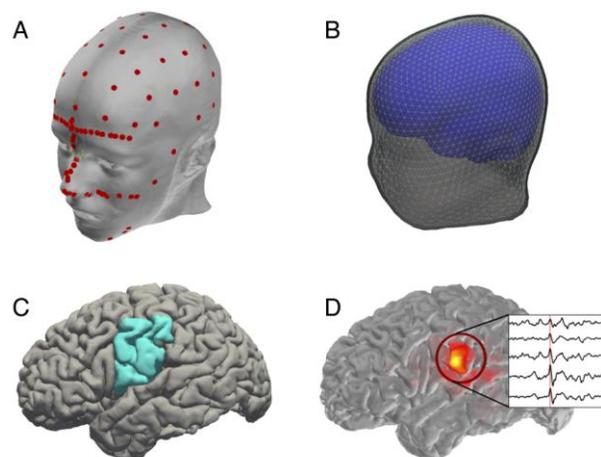


Figure 3.1: Example of electrical source imaging procedure.

(A) Digitized EEG electrode placement and anatomical landmarks (red circles).

(B) Reconstructed anatomical surfaces: inner skull (blue), outer skull (light gray), and outer skin surfaces (dark gray).

(C) Example cortical surface reconstruction with lateral inferior Rolandic cortex indicated (blue).

(D) Example source activity during an interictal spike in the inferior Rolandic cortex (inset). Red heat-map indicates amplitude of interictal spike (corresponding to time indicated by red line in inset) averaged over time for each source.

3.2.3 *Artifact and epileptic spike removal procedures*

To minimize the impact of muscle movements, we adapted the artifact removal procedure developed in Chu et al., (2014). First, for a 1 s interval of data we computed the power spectrum (Hanning taper). Then, we computed a linear fit to the logarithm of power versus logarithm of frequency for frequencies between 30 - 95 Hz. Given the typical $1/f^\alpha$ property of EEG activity (He et al., 2010), if the slope of the linear fit was not sufficiently negative, then the interval was marked as an artifact. We chose a threshold of $\alpha=1.5$, which exceeds the values of α typically observed in human brain activity (He et al., 2010).

Large amplitude interictal spikes common in subjects with Rolandic epilepsy produce broadband spectral content and may impact detection and characterization of spindles. Although our spindle detector is robust to the impact of spikes (*e.g.*, see section *Automated spindle detection* and Kramer et al., (2021)), to remove any potential impact of interictal spikes on our source estimates and subsequent analysis, we applied an automated spike detection method - the Persyst 13 algorithm (Scheuer, Bagic, & Wilson, 2017) - to each patient's scalp EEG data, and identified all spikes at the standard 10–20 EEG channels. We then removed 200 ms around spikes detected on the central, temporal, and frontal electrodes in all subsequent analysis.

3.2.4 *Neuropsychological assessment*

Each subject completed a focused neuropsychological assessment performed by clinical neuropsychologists (BCE, AKM) including assessments of fine motor

performance, attention, global intellectual function, and speech sound processing. To test fine motor performance, subjects completed the Grooved Pegboard task, where the time required to correctly place grooved pegs into irregular holes at a specific orientation is recorded, thereby providing an assessment of hand-eye coordination, motor speed and sensorimotor control and integration (Merker, Podell, & Wingate, 2018). Subjects (active epilepsy $n=6$; resolved epilepsy $n=9$; control $n=8$) completed the grooved pegboard task once using their dominant hand and once using their nondominant hand (Figure 3.6A). Attentional performance was assessed using the processing speed index from the Wechsler Intelligence Scale for Children, 5th ed (WISC-V), which examines the state of preparedness to respond to stimuli, incorporating both sensory registration and timing of motor response (Jacobson et al., 2011). The processing speed index is derived from subtests that require children to attend to visual material and sort or classify targets and symbols in a time-limited setting (active epilepsy $n=7$; resolved epilepsy $n=9$; control $n=8$). General intellectual functioning was assessed using the WISC-V to quantify full-scale IQ, which is derived from subtests of verbal comprehension, perceptual reasoning, working memory, and processing speed (Wechsler, 2014) (active epilepsy $n=6$; resolved epilepsy $n=8$; control $n=8$). Speech-sound processing was assessed using the Comprehensive Test of Phonological Processing, 2nd ed. to measure phonological awareness. The test is comprised of three subscales: the ability to divide a spoken word into its individual phonological components, the ability to blend individual phonemes presented auditorily and articulate them into spoken words, and the ability to break a real word into phonemic pieces, remove

one, and combine the remaining pieces together to produce a real word (Wagner, Torgesen, & Rashotte, 1991) (active epilepsy $n=5$; resolved epilepsy $n=7$; control $n=8$).

For all tests, z-scores representing each individuals' deviation from standardized score distributions for his or her age and sex were evaluated.

3.2.5 Spectral analysis

All spectral analyses were computed using the multitaper method as implemented in the Chronux toolbox (Bokil et al., 2010), unless otherwise noted. For each source, we computed the power spectrum on non-overlapping one second windows (frequency resolution 1 Hz; single Hanning taper) for the entire duration of the session, and then averaged these spectra. Then, for each region of interest, in each hemisphere, we averaged the power spectra of all sources within the label to create one power spectrum per label. We normalized this spectrum by the total power between 0-50 Hz to compute a relative power spectrum. We then computed two measures of sigma band activity. First, we computed sigma power as the average of relative power between frequencies 10-15 Hz for each label in each hemisphere. Second, we computed the sigma bump (Donoghue et al., 2020; Ouyang et al., 2020), which we define as the sigma power (10-15 Hz) with background activity subtracted. Specifically, we first fit a line between the power at 10 Hz and 15 Hz to approximate the $1/f^\alpha$ spectral background. Then, we subtracted the fit line from the power spectrum, and summed over the positive values between 10-15 Hz to approximate the contribution of sigma band activity above the spectral background. We computed this statistic for the inferior Rolandic cortex in both hemispheres.

3.2.6 Automated spindle detection

Spindles - sigma-band (10-15 Hz) activity of duration 0.5-2 s - are characteristic rhythms present in stages 2 and 3 NREM sleep (examples in Figure 3.2A). To identify spindles in subjects with Rolandic epilepsy, we applied a spindle detection method developed to accurately measure sleep spindles in subjects with epilepsy to source activity in the left and right inferior Rolandic cortices (Figure 3.2 B) of subjects with Rolandic epilepsy and control subjects (Kramer et al., 2021). Briefly, the method estimates the probability of the spindle state given three features calculated from the source activity: (i) theta power (4-8 Hz), (ii) sigma power (9-15 Hz), and (iii) a measure of the consistency of time intervals between subsequent peaks and subsequent troughs in the signal. This detector was trained and validated using scalp EEG data from subjects with Rolandic epilepsy and control subjects; for details see Kramer et al. (2021). Here, we applied this detector to the activity of each source within a chosen cortical label (*e.g.*, within the left and right inferior Rolandic cortices). The method returns the time interval of each spindle detection, with spindle durations restricted to be at least 0.5 s. We computed the number of spindles over time to define the spindle rate (spindles/minute). To detect spindles from scalp data directly, we applied the spindle detector to central and temporal electrode recordings (C3, C5, T3, C4, C6, T4) referenced to an average reference, as described in (Kramer et al, 2021).

3.2.7 Assessment of spindle features

We computed five features to characterize the identified spindles: duration, sigma-band power, intra-hemispheric and inter-hemispheric coupling, and bilateral spindle synchrony. We define each feature here.

Duration: We measured spindle duration as the time between onset and offset of each spindle detection. As part of the spindle detection procedure, durations were restricted to exceed 0.5 s.

Sigma-band power: To compute the sigma-band power of a spindle, we first applied a Hanning window to the source activity during the spindle detection. For spindle durations less than 4 s, we then zero padded the signal to 4 s, and evaluated the mean power between 10.25 Hz and 14.75 Hz to span the [10, 15] Hz range. We averaged the sigma band power over all spindles from sources within the left inferior Rolandic cortex and the right inferior Rolandic cortex for each patient.

Intra-hemispheric and inter-hemispheric sigma-band coherence: To assess sigma-band coherence, we first identified time intervals of spindle activity. To do so, we defined spindle indicator vectors for the set of sources within the left and right inferior Rolandic cortices. The spindle indicator vector is a time series containing ones when at least one source in a region of interest exhibits a spindle (example indicator vectors computed for sources from the left and right inferior Rolandic gyrus shown in Figure 3.4C). We note that, within a spindle indicator vector, multiple sources may be involved, and not all sources may be involved for the entire duration of the vector. We selected ± 1 s around the

center of each spindle indicator vector to create a 2 s spindle epoch. In these spindle epochs, only sources exhibiting spindles were included to compute the coherence.

Within each spindle epoch, we investigated the coherence of sources within and between the left and right inferior Rolandic cortices. To assess intra-hemispheric coupling, we computed the coherence between sources in the left (or right) inferior Rolandic cortex during each spindle run detected in the left (or right) inferior Rolandic cortex, yielding two measurements per subject (see ‘intra’ example in Figure 3.4C). To assess inter-hemispheric coupling, we computed coherence between the left inferior Rolandic cortex and the right inferior Rolandic cortex during bilateral and simultaneous spindle epochs detected in both cortices, yielding one measurement per patient (see ‘inter’ example in Figure 3.4C).

We computed pairwise coherence between sources with a 2.5 Hz frequency resolution and 9 tapers. We evaluated coherence at 12.5 Hz covering the 12.5 ± 2.5 Hz to estimate the sigma coherence.

Bilateral synchrony: Because Rolandic spikes occur independently in the left and right hemisphere in Rolandic epilepsy, and sleep spindles often occur synchronously between hemispheres in the Rolandic regions after ~12 months of age (Gruber & Wise, 2016), we evaluated the interhemispheric synchrony between Rolandic spindles. To do so, we computed the dot product between the spindle indicator vectors for each hemisphere, as defined in the previous section (example indicator vectors in Figure 3.4C). The result estimates how often at least one Rolandic cortical source in each hemisphere produce spindles that temporally overlap. To account for potential biases due to differences in spindle rate, we divided the product by the sum of the bilateral spindle run (*i.e.*, a vector

that indicates when spindles occurred in either the left or the right hemisphere). For example, in Figure 3.4C, we compute the dot product of the left and right indicator vectors (resulting in a value of 9) and divide by the sum of the joint bilateral spindle run (value of 11).

3.2.8 Statistical analysis

We tested three *a priori* hypotheses: 1) spindle rate would be decreased in the inferior Rolandic cortex in active epilepsy; 2) spindle deficits would extend beyond the epileptic cortex in active epilepsy; and 3) regional spindle deficits would better predict cognitive dysfunction than focal estimates.

To test hypothesis (1), we implemented a mixed effects model with spindle rate as the dependent variable and group as the predictor (indicator vectors for the active group and for the resolved group; both zero if in the control group) and controlling for age. In addition to a direct measure of spindle rate, we also tested two related measures: sigma power and sigma bump (see *Spectral analysis*). We used a linear model fit using maximum likelihood for sigma power ($P=0.79$, Lilliefors test, no evidence of violation of normality) and a quasi-Poisson model fit using pseudo likelihood for spindle rate and sigma bump. We chose a quasi-Poisson model for the spindle rate data for three reasons. First, visual inspection of the data suggested a concentration of spindle rate values near zero (Figure 3.4A). Second, spindle rate is directly related to the (discrete and nonnegative) spindle count, consistent with this discrete probability distribution. Third, we find a near violation of normality ($P=0.076$, Lilliefors test). We note that, assuming a normal distribution for

the spindle rate and repeating all analyses, we found consistent results. We chose a quasi-Poisson model for the sigma bump due to the violation of normality for these data ($P=0.01$, Lilliefors test) and to maintain consistency with the model of spindle rate. We included a random intercept term to account for repeat measurements (*e.g.*, from the left and right inferior Rolandic cortices) taken from the same subject. Significant differences between the active or resolved epilepsy subjects and control subjects were identified if the p-value of the corresponding variable was less than 0.05.

In addition to spindle rate, we tested whether focal spindle properties (*i.e.*, sigma power of spindles, duration, intra-hemispheric coupling, inter-hemispheric coupling, or bilateral synchrony) in the inferior Rolandic cortex differed by group. To do so, we implemented a likelihood ratio test comparing a null and a full model. The null model included age, and the full model additionally included the group variable (active and resolved epilepsy groups). We chose a linear model because we found no violations of normality for any measures ($P>0.1$, Lilliefors test). For sigma power, duration, and intra-hemispheric coherence, we included a random intercept term because there were two measurements per subject (*e.g.*, from the left and right inferior Rolandic cortices). For inter-hemispheric coherence and bilateral synchrony, we only include the fixed effects because there was only one measurement per subject. We tested for significant differences ($P<0.05$, chi-squared distribution) between the models using the MATLAB functions, *compare* and *lratiotest*, for the mixed and fixed effects models, respectively. In the full model, we identified significant differences between the active or resolved epilepsy subjects and

control subjects if the p-value of the corresponding variable was less than 0.05 in the full model.

To test hypothesis (2), we implemented the same quasi-Poisson mixed effects model for spindle counts used for the inferior Rolandic cortex for each of the 31 Desikan-Killiany Atlas labels. We use false discovery rate (FDR) (Benjamini & Hochberg, 1995) with $q=0.05$ to correct for multiple comparisons.

To test hypothesis (3), we first (a) determined whether focal or regional spindle rate predicted cognitive performance. We then (b) tested which of these spindle estimates performed better. The focal spindle rate is defined as the average over sources in the inferior Rolandic cortices, and the regional spindle rate is defined as the average across all cortical regions identified to have a significantly lower spindle rate in subjects with active epilepsy in hypothesis (2) (see Figure 3.5).

For (a) and (b), to estimate a model of combined cognitive function, we fit individual models for each of the four neuropsychological tasks as a function of focal or regional spindle rate (individual models described below). We summed the deviance across each of the individual models, assuming independence between cognitive functions after conditioning on spindle rate. The summed deviance provides a measure of goodness-of-fit for a model of combined cognitive performance using either the focal or regional estimates of spindle rate, *i.e.*, the focal deviance or the regional deviance respectively.

For (a), to test for a relationship between spindle rate and combined cognitive function, we additionally computed the null deviance of a model of combined cognitive function

excluding spindle rate. We used a chi-square test (4 degrees of freedom) to test if the focal or regional deviance is significantly different than the null deviance. If spindle rate was found to predict performance on the combined model, we then analyzed the relationship between spindle rate and performance of each individual neuropsychological. For each task, a significant relationship was determined if $P < 0.05$. For (b), we used a bootstrap approach to test if the focal deviance significantly differed from the regional deviance (for more details see *Supplementary Material*). To determine whether focal or regional estimates of spindle rate better predict individual cognitive functions, we applied the same analysis to each individual neuropsychological task model. Finally, we compared the results of (a) and (b) to models using spindle rate estimated from centroparietal electrodes in the scalp EEG.

For the individual neuropsychological tasks: (i) fine motor dexterity, (ii) processing speed, (iii) full-scale IQ, and (iv) phonological awareness, we built the following models. For (i), we modeled motor dexterity as a linear function of spindle rate. We paired performance by hand with spindle rate in the contralateral hemisphere (see Figure 3.6A). We fit a linear mixed effects model with a random intercept to account for repeat measures from the same subject (*i.e.*, left and right hand performance), which accounts for differences in baseline fine motor skill for each subject. For (ii) and (iii), we compared task performance with the average spindle rate over the left and right hemispheres. For (iv), we compared task performance with the spindle rate in the left hemisphere, which is typically dominant in language. For (ii) to (iv), we fit a linear model estimating performance as a function of spindle rate. For all models, we applied the inverse hyperbolic sine (IHS)

transform to the spindle rate to reduce the influence of extreme observations (Kramer et al., 2021). We tested age as a covariate in each model and included age as an independent variable in the model if $P < 0.1$. Doing so, we only found evidence to include age in (i); we therefore modeled motor dexterity as a function of spindle rate, controlling for age. The systematic part of the model for (i) fine motor dexterity is:

$$\beta_0 + \beta_1 \text{IHS}(\text{spindle rate}) + \beta_2 \text{age},$$

where β_0 , β_1 , and β_2 are unknown parameters to estimate, and the model includes a random-effects term indexed by subject.

For (ii) – (iv), the systematic part of the model is:

$$\beta_0 + \beta_1 \text{IHS}(\text{spindle rate}).$$

3.2.9 Data availability

Raw data were generated at Massachusetts General Hospital and the Athinoula A. Martinos Center for Biomedical Imaging. Derived data supporting the findings of this study are available from the corresponding author on request. Software for the detection of spindle events is available at <https://github.com/Mark-Kramer/Spindle-Detector-Method>.

3.3 Results

3.3.1 Spindle rate in the inferior Rolandic cortex is reduced in Rolandic epilepsy compared to control subjects.

We found a significant spindle deficit in the inferior Rolandic cortices of active subjects compared to control subjects (70.9% decrease, $P = 0.007$, quasi-Poisson model),

and no significant difference in spindle rate between resolved and control subjects ($P=0.2$). Progressing from the most to least affected group, the average spindle rate increased (Figure 3.2C): active subjects mean 0.65 spindles/min, standard deviation 0.78 spindles/min; resolved subjects mean 1.09 spindles/min, standard deviation 0.65 spindles/min; control subjects mean 1.84 spindles/min, standard deviation 1.04 spindles/min. We conclude that spindle rate is transiently decreased in the inferior Rolandic cortex in Rolandic epilepsy during the active period of disease.

We note that alternative spectral measures have been used to estimate spindle activity. In particular, sigma power is frequently used as a surrogate measure of spindle activity (Beelke et al., 2000; Nobili et al., 2001, 1999; Tucker & Fishbein, 2009), although the effects are weakened by background EEG activity (De Gennaro & Ferrara, 2003; Wamsley et al., 2012). Visual inspection of the average spectrum of source activity in the Rolandic cortices suggests lower sigma-band power in subjects with Rolandic epilepsy compared to control subjects (Figure 3.3A). To examine this surrogate measure of spindle activity, we analyzed both sigma power and sigma bump, in which the background sigma activity is removed (see *Methods*). As expected (Purcell et al., 2017), both measures positively correlated with spindle rate (sigma power, $r=0.60$, $P<1e-5$; sigma bump, $r=0.82$, $P<1e-13$; Figure 3.3B,C). However, we found no difference in these spectral measures between the active epilepsy and control groups (sigma power, $P=0.119$; sigma bump, $P=0.08$). We conclude that spindle rate is reduced in the inferior Rolandic cortices and is a more sensitive measure of the difference in spindle activity between subject groups than spectral measures alone.

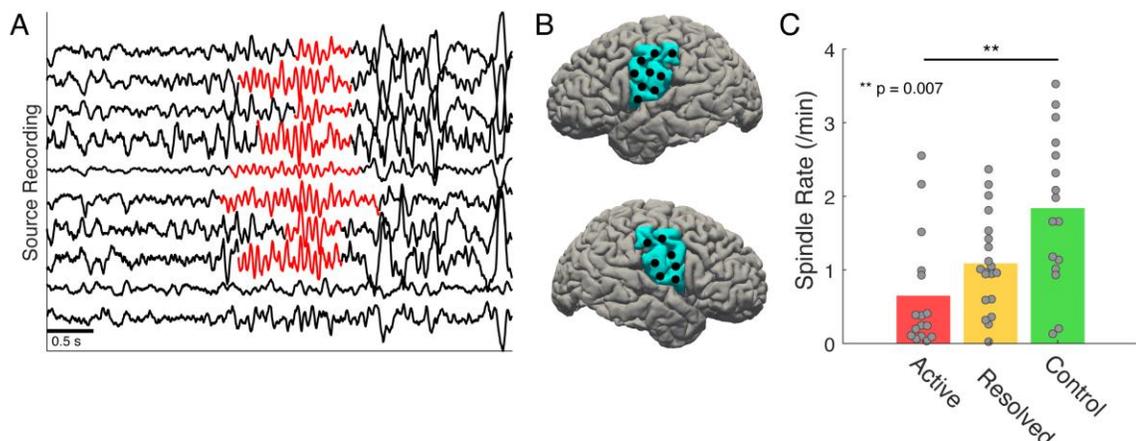


Figure 3.2: Spindle rate is lower in subjects with Rolandic epilepsy.

- (A) Example spindle activity (10-15 Hz rhythms characteristic of stage 2 sleep) from 10 sources in the left inferior Rolandic cortex with detected spindles (orange).
- (B) Left and right lateral views of the brain with the inferior Rolandic region indicated (blue). Black circles indicate example source locations used to compute spindle rate.
- (C) Spindle rate in the inferior Rolandic cortices for active (red), resolved (yellow), and control (green) subjects. Bar heights indicate population mean, and circles indicate the spindle rate for each hemisphere of each patient.

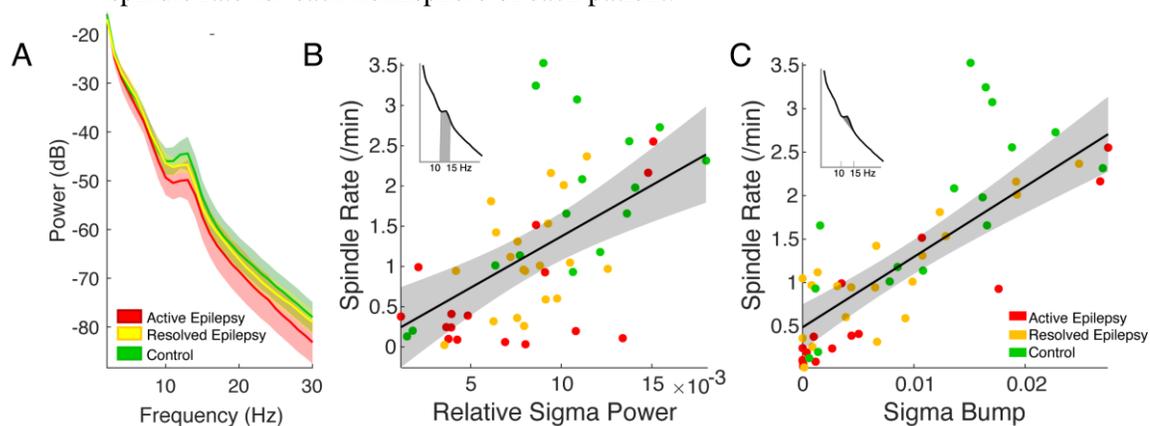


Figure 3.3: Spindle rate is a more sensitive measure than other spectral estimates.

- (A) Averaged power spectra for each patient group, active (red), resolved (yellow), and control subjects (green). Solid lines indicate the mean, and shading indicates 95% confidence intervals. (B, C) Spindle rate in the inferior Rolandic cortex correlates with sigma power (B) and sigma bump (C). Shaded regions in the power spectra insets in the upper left of (B) and (C) represent areas used to compute sigma power and sigma bump, respectively. Black line indicates the linear fit, shading the 95% confidence intervals, and circles the values for each subject (see legend).

3.3.2 *Spindle features are typical in subjects with Rolandic epilepsy but are less bilaterally synchronous.*

To test whether spindle properties in the inferior Rolandic cortices differed between subject groups, we analyzed five features computed during identified spindles occurring from sources within the inferior Rolandic cortices (Figure 3.4A-C). We found no difference between groups in four of the features: sigma power ($P=0.25$, likelihood ratio test, see *Methods*; Figure 3.4D), duration ($P=0.21$, likelihood ratio test; Figure 3.4E), or the intra-hemispheric or inter-hemispheric sigma-band coherence ($P=0.64$, likelihood ratio test; Figure 3.4F; $P=0.11$, Figure 3.4G, respectively). The co-occurrence of bilateral spindles was lower in active ($P=0.005$, linear model), but not resolved ($P=0.13$, linear model) epilepsy subjects compared to control subjects (Figure 3.4H). We conclude that – although spindle rate is reduced in active Rolandic epilepsy – when spindles occur, spindle features are similar in Rolandic epilepsy and control subjects. However, spindles are more bilaterally independent in subjects with active epilepsy compared to control subjects, consistent with the bilaterally independent nature of the epileptiform spike activity in Rolandic epilepsy subjects (Callenbach et al., 2010; Carvill et al., 2013).

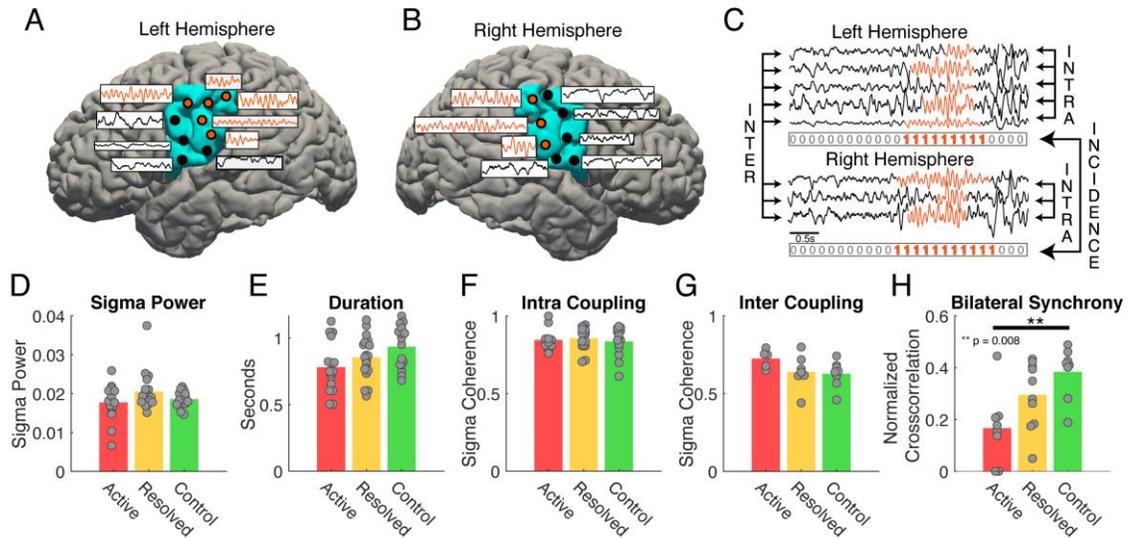


Figure 3.4: Rolandic epilepsy subjects produce fewer, but healthy, spindles.

(A) Left and (B) right hemispheres of a subject with sources (circles) in the inferior Rolandic cortices (blue). The subset of sources in the inferior Rolandic cortices with detected spindles are colored orange, otherwise black. (C) Example recordings from source in the left hemisphere (top) and the right hemisphere (bottom) with detected spindles in orange. Arrows between sources within each cortex indicate intra-hemispheric coherence, and arrows between sources from the left to the right cortices indicate inter-hemispheric coherence. Below the recordings from each hemisphere is the corresponding spindle indicator function that contains ones if at least one source is exhibiting a spindle at that moment in time and is used to compute the bilateral synchrony of spindles. (D-H) Spindle characteristics sigma power (D), duration (E), intra-hemispheric sigma band coherence (F), inter-hemispheric sigma band coherence (G), and bilateral synchrony (H). We find evidence of a difference between active and control subjects only for the bilateral synchrony (asterisks).

3.3.3 Spindle deficit extends beyond inferior Rolandic cortices.

Next, to test the hypothesis that the spindle deficit extends beyond the epileptic cortex, we analyzed spindle rates measured from each Desikan-Killiany atlas label in each subject (see *Methods*). We found significantly lower spindle rates in active epilepsy compared to control subjects (quasi-Poisson mixed effects model) in frontal cortical regions (superior frontal; rostral middle frontal; triangularis; lateral orbitofrontal; medial

orbitofrontal), insula, temporal cortical regions (superior temporal, entorhinal), as well as in the superior parietal region ($P < 0.009$ for all regions, controlling for multiple comparisons using FDR; Figure 3.5). The most affected region is the inferior Rolandic cortex which, as reported above, results in a mean 70.9% decrease in the baseline spindle rate for the active group. For the remaining affected regions, we find mean decreases between 49.6-60.8% in the baseline spindle rate for the active group (Figure 3.5). We conclude that spindle deficits in Rolandic epilepsy involve broader cortical regions beyond the inferior Rolandic cortex.

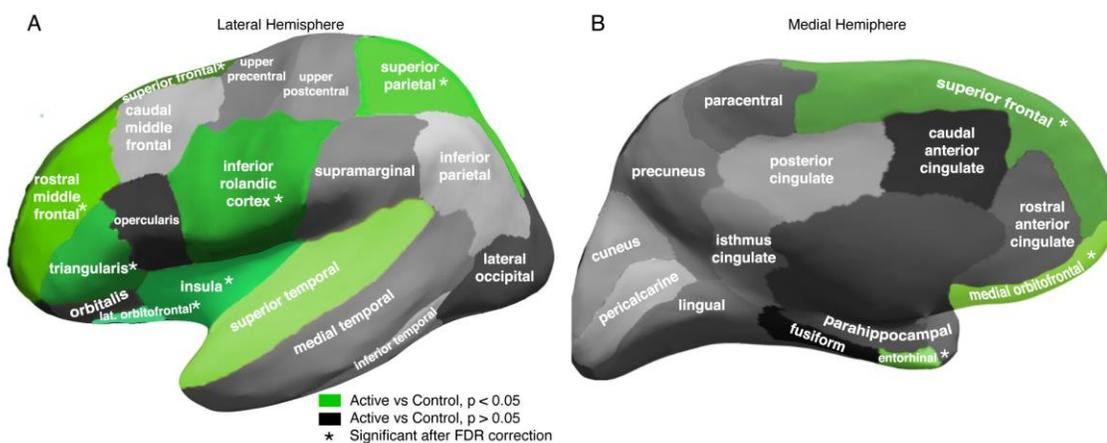


Figure 3.5: Spindle rate deficit extends beyond inferior Rolandic cortices.

Parcellation of the cortex into 31 regions per hemisphere. Lateral (A) and medial (B) regions with a significant reduction in spindle rate in active versus control subjects indicated in green. Asterisk (*) indicates significant differences after correction for multiple comparisons.

| Brain region | Percent (%) | P-value |
|-------------------------|-------------|---------|
| Lateral Orbito. Frontal | 60.8 | 0.001* |
| Rostral Middle Frontal | 55 | 0.002* |
| Superior Frontal | 50.4 | 0.006* |
| Medial Orbito. Frontal | 56.7 | 0.006* |
| Triangularis | 56.2 | 0.006* |

| | | |
|--|------|--------|
| Inferior Rolandic | 70.9 | 0.007* |
| Superior Parietal | 49.6 | 0.007* |
| Insula | 51.7 | 0.008* |
| Entorhinal | 58.2 | 0.008* |
| Superior Temporal | 54.1 | 0.038 |
| Cuneus | 46.6 | 0.051 |
| Pericalcarine | 42 | 0.053 |
| Opercularis | 46.2 | 0.056 |
| Caudal middle frontal | 57.4 | 0.063 |
| Rostral Ant. Cingulate | 61.5 | 0.074 |
| Inferior Parietal | 58.5 | 0.075 |
| Fusiform | 49.5 | 0.103 |
| Caudal Ant. Cingulate | 47.8 | 0.103 |
| Lateral Occipital | 52.9 | 0.123 |
| Inferior Temporal | 33.5 | 0.125 |
| Lingual | 31.3 | 0.135 |
| Orbitalis | 43.9 | 0.14 |
| Medial Temporal | 48.2 | 0.157 |
| Precentral gyrus (excluding inferior Rolandic cortex) | 42.6 | 0.202 |
| Precuneus | 43.8 | 0.21 |
| Posterior Cingulate | 39.4 | 0.225 |
| Supramarginal | 41 | 0.249 |
| Isthmus Cingulate | 18 | 0.68 |
| Parahippocampal | 8.6 | 0.783 |
| Paracentral | 4.7 | 0.877 |
| Postcentral gyrus (excluding inferior Rolandic cortex) | -3.2 | 0.92 |

Table 1: Percent reduction in spindle rate by being in the active group relative to the control group.

Asterisk (*) indicates significant differences after correcting for multiple comparisons using the false discovery rate.

3.3.4 *Regional estimates of spindle rate improve predictions of cognitive function better than focal estimates*

We found that focal source estimates of spindle rate from the inferior Rolandic cortex only and regional estimates of spindle rate from all affected cortical regions both predicted cognitive performance in our combined model across all neuropsychological tasks (focal $P=0.001$, regional $P<1e-4$, focal deviance =192.4, regional deviance = 185.1, null deviance = 210.5, chi-squared test, 4 degrees of freedom). The performance of the model using regional estimates of spindle rate was significantly better than that using focal estimates of spindle rate ($P=0.016$).

Across individual neuropsychological tasks, we found positive relationships between both focal and regional source estimates of spindle rate and performance in each domain tested: fine motor skills, processing speed, full-scale IQ, and phonological awareness (Figure 3.6). Using the focal spindle estimates, we found strong positive relationships between spindle rate and fine motor performance ($\beta_1 = 0.9$, see Methods, 95% CI [0.15,1.65], $P=0.02$, $R^2=0.88$), and processing speed ($\beta_1 = 0.84$, 95% CI [0.11, 1.57], $P=0.03$, $R^2=0.19$), and weaker positive relationships with full-scale IQ ($\beta_1 = 0.76$, 95% CI [0.04, 1.5], $P=0.052$), and phonological awareness ($\beta_1 = 0.79$, 95% CI [0.005, 1.58], $P=0.06$). Using the regional spindle estimates (Figure 3.6), we found strong positive relationships between spindle rate and motor performance ($\beta_1 = 1.36$, 95% CI [0.63, 2.1], $P<1e-3$, $R^2=0.9$), processing speed ($\beta_1 = 0.87$, 95% CI [0.19, 1.54], $P=0.02$, $R^2=0.22$), and IQ ($\beta_1 = 0.76$, 95% CI [0.07, 1.44], $P=0.04$, $R^2=0.19$), and no relationship with phonological awareness ($P=0.1$); see Table 2.

Across individual tasks, we found regional estimates of spindle rate significantly improved performance in the model of fine motor skills ($P=0.001$, focal deviance=130.2, regional deviance=123.5, bootstrap test) compared to focal estimates. We found no significant difference between models using focal or regional spindle estimates to predict processing speed, full-scale IQ, and phonological awareness ($P>0.3$, bootstrap test).

Repeating this analysis using centrottemporal spindle estimates from the scalp EEG, we found no significant difference compared to regional source estimates in modeling individual task performance ($P>0.2$ for all tasks, bootstrap test), and improved performance compared to focal ESI estimates only for the fine motor task ($P=0.035$, bootstrap test; $P>0.27$ for all other tasks).

We note that repeating these analyses without removing interictal spikes from the data (see *Artifact and epileptic spike removal procedures*) yielded qualitatively consistent results. We also note that the model residuals of IQ are correlated with the model residuals of motor dexterity, processing speed, and phonological awareness (Pearson's correlation coefficient $r>0.53$, $P<0.012$ for all models). To account for this, we repeated our analysis assuming conditional dependence on IQ (see *Supplementary Material*) which removed all correlation between the models ($r<0.18$, $P>0.12$) and found qualitatively similar results.

We conclude that both focal and regional ESI estimates of spindle rate predict cognitive function. However, regional estimates of spindle rate from all affected cortical regions significantly improves prediction of combined cognitive function and motor performance compared to spindle estimates limited to the inferior Rolandic cortex.

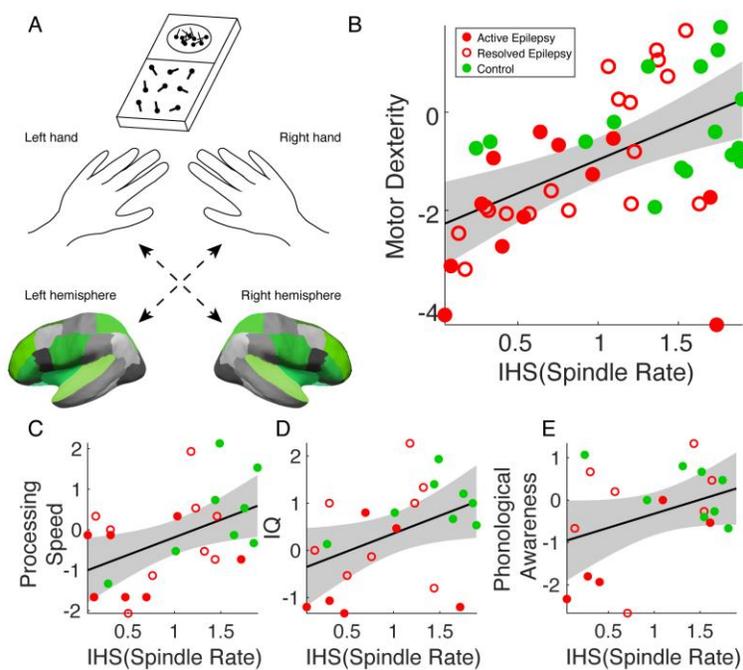


Figure 3.6: Spindle rate correlates with neuropsychological assessments.

Schematic of the grooved pegboard experiment. Subjects perform a grooved pegboard task with their left and right hand. Performance is paired with spindle rate in the contralateral hemisphere (blue). **(B-D)** As regional measure of spindle rate increases, motor performance **(B)**, processing speed **(C)**, and IQ **(D)** significantly increase. **(E)** Phonological awareness shows an increasing trend. Circles represent three disease groups: active (red filled), resolved (red unfilled), and control subjects (green). The solid line indicates the model fit, and shaded regions indicate 95% confidence intervals.

| | Fine Motor Skills | Processing Speed | Full-Scale IQ | Phonological Awareness |
|--|-------------------|------------------|---------------|------------------------|
| | | | | |

| | | | | | |
|-----------------------|-----------------------------------|-------------------|--------------------|-------------------|---------------------|
| Focal Spindle Rate | Beta coefficient [95% confidence] | 0.9; [0.15,1.65] | 0.84; [0.11, 1.57] | 0.76; [0.04, 1.5] | 0.79; [0.005, 1.58] |
| | P-value | 0.02 | 0.03 | 0.052 | 0.06 |
| | Deviance | 130.2 | 23.3 | 18.9 | 19.9 |
| | R ² | 0.88 | 0.19 | 0.18 | 0.18 |
| Regional Spindle Rate | Beta coefficient [95% confidence] | 1.36; [0.63, 2.1] | 0.87; [0.19, 1.54] | 0.76; [0.07,1.44] | 0.66; [-0.08, 1.41] |
| | P-value | <1e-3 | 0.02 | 0.04 | 0.2 |
| | Deviance | 123.5 | 22.3 | 18.6 | 20.7 |
| | R ² | 0.9 | 0.22 | 0.19 | 0.14 |

Table 2: Model fits for each neuropsychological assessment using focal and regional spindle rate.

3.4 Discussion

While neurocognitive deficits commonly occur in Rolandic epilepsy, it is unknown how the pathology of epilepsy disrupts cognition in this disease and related epileptic encephalopathies. Here, using ESI, we investigated the spatial extent of the sleep spindle deficit and the relationships between cortical sleep spindle deficits and performance on neurocognitive tasks. We found that children with active Rolandic epilepsy have regional spindle deficits that extend beyond the epileptic Rolandic cortices, involving parts of the pre-frontal, insula, temporal, and parietal cortices. We also found that spindle rates estimated from these broadly affected regions better predicted cognitive performance on a range of tasks compared to spindle rate estimated from the inferior Rolandic cortex alone. These results suggest that the cognitive symptoms in Rolandic epilepsy might be due to

involvement of broader regional networks beyond the Rolandic cortex and contributes to growing evidence of thalamocortical circuit dysfunction in Rolandic epilepsy.

We found no significant differences in spindle duration or sigma power of spindles produced in the active epilepsy subjects versus the control subjects. We additionally analyzed alternative measures of spindle activity, sigma power and sigma bump. Despite correlation with spindle rate, spindle amplitude and spindle duration (Beelke et al., 2000; De Gennaro & Ferrara, 2003; Nobili et al., 2001) were not as sensitive as spindle rate alone to detect a difference between groups. We also found no significant difference in the fine temporal coupling (*i.e.*, sigma coherence) between spindles; however, the likelihood of spindles to coincidentally occur between hemispheres was reduced. This finding indicates that spindle production is more bilaterally independent in active Rolandic epilepsy, similar to the bilaterally independent epileptic spiking activity (Galicchio et al., 2021). Altogether, these results suggest that the spindle pathology and the associated neural plastic changes that contribute to cognitive deficits in Rolandic epilepsy are restricted to the spindle rate, and not characteristics or coupling properties of the spindles themselves. Further, these data suggest that the process resulting in a reduction of spindles occurs independently in the left and right thalamocortical circuits.

In analyzing neurocognitive task performance, we found a strong relationship between the regional spindle rate and motor dexterity. Source estimates of regional spindle rate improved prediction of contralateral fine motor performance compared to focal estimates of spindle rate from the inferior Rolandic estimates, where regional estimates explained 90% of the variance in fine motor performance and increased the mean

coefficient estimate (β_1) by approximately 50% compared to the focal spindle rate. Although the inferior Rolandic cortex is involved in primary sensorimotor processing, several regions identified to have a spindle deficit, including the posterior parietal cortex and pre-frontal cortex, are involved in the planning, initiation and execution of motor movements (Andersen, Aflalo, & Kellis, 2019; Purves et al., 2004). Additionally, the insula has strong functional connectivity with the sensorimotor cortex (Fink, Frackowiak, Pietrzyk, & Passingham, 1997). Thus, the fine motor impairments measured in this task may reflect dyscoordination in these distributed motor networks that extend beyond the primary sensorimotor (*e.g.*, Rolandic) cortex. Although we did not find that the regional model had improved performance to predict processing speed and full-scale IQ compared to the focal model using our bootstrap approach, we note increases in mean coefficient estimate (β_1) in models using regional spindle rate, suggesting nominal improvements. Thus, distributed cortical networks may also contribute to these cognitive deficits observed. Models using regional spindle estimates performed similarly to models using spindle estimates from centrotemporal scalp EEG channels, which also reflects regional cortical activity due to the limited spatial resolution of scalp EEG. To limit risk of false detection due to multiple comparisons, we did not test spindle estimates from mixed combinations of cortical sources here. Given the regional spindle deficits observed here, future work could investigate whether spindle estimates from different combinations of cortical regions provide the best models for specific cognitive functions.

The identification of a regional spindle deficits suggests two potential sources of malfunctioning in the thalamocortical circuitry. First, the thalamus is comprised of many

nuclei that have broad and discrete thalamocortical connectivity (Bastuji et al., 2020; Behrens et al., 2003; Fama & Sullivan, 2015). Spindles are generated in the thalamic reticular nucleus (TRN) and propagated throughout cortex via other cortical and thalamic circuits (Beenhakker & Huguenard, 2009; De Gennaro & Ferrara, 2003). The TRN encases all the other thalamic nuclei (De Gennaro & Ferrara, 2003) and is comprised of GABAergic cells (Clemente-Perez et al., 2017). Both the thalamic nuclei (Behrens et al., 2003) and GABAergic subpopulations (Clemente-Perez et al., 2017; Li et al., 2020) have discrete thalamocortical circuitry potentially leading to the spatially discrete regional disturbances in cortical spindle activity (Bastuji et al., 2020). For example, a focal reduction of spindle activity in the inferior Rolandic cortices could implicate ventroanterior, ventrolateral, and ventroposterior thalamic nuclei (Andersen, Andersson, & Lomo, 1967). However, regional spindle deficits involving the prefrontal, superior parietal, insular, and temporal regions could implicate more thalamic nuclei, including the anterior pulvinar, mediodorsal nucleus and parts of the anterior nucleus (Behrens et al., 2003). Second, it has been shown in mouse models that parvalbumin (PV) and somatostatin (SOM) cells in the TRN are part of distinct functional circuits. For example, PV cells have strong inputs to thalamic relay nuclei (*e.g.*, ventromedial, ventrolateral, ventroposteriormedial, ventroposteriorlateral) whereas SOM cells have strong inputs to intralaminar nuclei (Clemente-Perez et al., 2017). Although sleep spindles enable identification of cortical networks affected, future work is required to understand the regional thalamocortical circuitry leading to the distributed cortical abnormalities observed.

Here, we have provided evidence that spindle rate is a sensitive biomarker that tracks with disease state and extends beyond the region of focal spiking activity implicating regional thalamocortical circuit dysfunction. Although Rolandic epilepsy is considered a focal epilepsy, we found the regional model of dysfunction better predicts cognitive function, providing a potential mechanistic explanation for the range of cognitive deficits observed in children with this epileptic encephalopathy. Alongside treating seizures, future therapeutic trials in Rolandic epilepsy could target increased spindle production with the goal of improving cognitive symptoms in this common disease (Mednick et al., 2013; Ngo et al., 2015).

3.5 Acknowledgements

The authors thank Sally M. Stoyell, Dan Song, Emily L. Thorn, and Lauren M. Ostrowski for assistance with data collection for this project.

3.6 Funding

ERS is supported by National Science Foundation GRFP. CJC was supported by NIH NINDS K23NS092923 during data collection for this work. CJC, MAK, and DM are supported by NIH NINDS R01NS115868.

3.7 Competing interests

CJC and MAK consult for Biogen Inc.

3.8 Supplementary material

To determine whether focal or regional estimates of spindle rate better predict combined cognitive function, we employed a bootstrap procedure. Under the null hypothesis of no difference between focal and regional spindle rates in a model of cognitive

performance, we randomly selected focal or regional estimates within each subject to model cognitive performance and computed the difference in deviance. Repeating this procedure 1000 times, we generated a distribution of deviance differences under the null hypothesis, and used this distribution to compute the probability of the observed difference in deviances.

We repeated our analysis assuming conditional dependence on IQ by including full-scale IQ as an additional predictor in each of the models for the remaining three tasks. For

(i) motor dexterity, the systematic part of the model is updated to:

$$\beta_0 + \beta_1 \text{IHS}(\text{spindle rate}) + \beta_2 \text{age} + \beta_3 \text{IQ}.$$

For (ii) processing speed and (iv) phonological awareness, the models are updated to:

$$\beta_0 + \beta_1 \text{IHS}(\text{spindle rate}) + \beta_2 \text{IQ}.$$

The model for IQ is the same. Then, we summed the deviance across each of these individual models as a measure of goodness-of-fit for a model of combined cognitive performance conditioned on IQ and repeat all subsequent analyses.

4 CHAPTER FOUR

Deviations from the natural history of delta power in Angelman syndrome reflect treatment effect size and correlate with UBE3A expression⁴

ABSTRACT

Background: Angelman syndrome (AS) is a neurodevelopmental disorder caused by deficiency of the maternally inherited *Ube3a* gene in neurons. Antisense oligonucleotide (ASO) therapies are under development to reinstate UBE3A protein production. Non-invasive biomarkers to detect target engagement and treatment response are needed to support clinical trials. Delta power measured in the scalp electroencephalogram (EEG) is a reliable biomarker for AS, but varies widely across individuals and throughout development, making detection of a treatment effect using single measurements challenging.

Methods: We utilized a longitudinal dataset of 204 EEG recordings from 56 patients with AS to develop a natural history model of delta (2-4 Hz) power, with predictors of age, elapsed time, and delta power at initial recording. Using this model, we computed the sample and effect sizes needed to detect a treatment effect in a human clinical trial with

⁴ Chapter 4, in full, is a reprint of the following article as it appears under review: Spencer, E. R., Shi, W., Komorowski, R. W., Gilbert, J. P., Ostrowski, L. M., Bird, L. M., Thibert, R., Bao, C., Molloy, F., Calhoun, M., Koirala, S., Jafar-nejad, P., Rigo, F., Kramer, M. A., & Chu, C. J. (2021). Deviations from the natural history of delta power in Angelman syndrome reflect treatment effect size and correlate with UBE3A expression, *eLife*, under review.

80% power. We applied the model to a mouse model of AS (n=41) to detect ASO-mediated treatment effects on delta activity and *Ube3a* expression.

Results: In humans, delta power at a second timepoint can be reliably predicted using the natural history model. In mice, a treatment effect can be detected after *Ube3a-ATS* ASO-treatment through at least 8 weeks post-treatment ($p < 1e-15$). Deviations in delta power from the expected natural history correlated with *Ube3a* expression in the mouse model ($p < 0.001$).

Conclusions: Deviations in delta power from a human natural history model in AS can detect ASO-mediated improvement in *Ube3a* expression in AS mice and may be relevant for human clinical trials.

4.1 Introduction

Angelman syndrome (AS) is a rare neurodevelopmental disorder (Kyllerman, 1995; Mertz et al., 2013; Petersen, Brøndum-Nielsen, Hansen, & Wulff, 1995b) characterized by severe developmental delay and epilepsy, along with impairments in speech and motor skills (Thibert et al., 2013; Williams et al., 2006). AS is caused by a deficit of UBE3A protein due to genetic abnormalities resulting in loss of *Ube3a* expression from the maternal allele (Bird, 2014; Sonzogni et al., 2019). Promising disease modifying therapies to reinstate production of UBE3A are under development (Bi et al., 2016). In particular, antisense oligonucleotides (ASOs) have been developed to target the endogenous *Ube3a* antisense transcript (*Ube3a-ATS*), which normally silences the paternal *Ube3a* allele in neurons (Meng et al., 2015). In a mouse model of AS, ASO treatment unsilences the paternal *Ube3a* allele and increases the production of UBE3A protein (Meng et al., 2015).

Such potentially transformative disease modifying treatments give rise to the need for an accurate, non-invasive approach to detect target engagement and treatment effect in clinical trials.

Many studies indicate that abnormal delta power (2-4 Hz) measured in the scalp electroencephalogram (EEG) is a reliable and sensitive biomarker for AS. Delta power is highly elevated in AS patients compared to typically developing individuals (Frohlich et al., 2019; Martinez et al., 2020; Sidorov et al., 2017), and in mouse models of AS compared to wild-type mice (Sidorov et al., 2017). Additionally, delta power correlates with genotype (Frohlich et al., 2019) and cognitive function (Ostrowski et al., 2021), where increased delta power correlates with more severely affected phenotypes. Therefore, delta power may present a useful biomarker for severity of disease and may provide a simple, non-invasive metric to track improvement in clinical trials. However, although delta power is reliably increased in AS compared to healthy control subjects, in cross-sectional studies, delta power measurements vary widely between individuals and throughout development. Longitudinal measurements of delta power, accounting for each subject's age and elapsed time between measurements, are required to develop a more sensitive measure of target engagement in clinical trials.

Here, we first analyze a small database of prolonged continuous EEG recordings in AS subjects to show that delta power estimates remain stable over the course of a 24-hour sample. Then, utilizing a large longitudinal dataset of EEG recordings from patients with AS, we develop a natural history model to predict delta power at a second visit from delta power at an initial visit, age, and elapsed time between visits. We utilize this model to

compute the sample and effect sizes needed to detect a treatment effect in a human clinical trial with 80% power. We then fit the model on a longitudinal AS mouse dataset and measure for a treatment effect after ASO-treatment. Finally, we compare deviations from the natural history model with *Ube3a* expression in *Ube3a-ATS* ASO-treated and control ASO-treated mice. This work provides a non-invasive method to detect potential treatment effects with confidence in AS and validates that increased *Ube3a* expression corresponds to deviations from the natural history of delta power in this disorder.

4.2 Materials and Methods

4.2.1 Human subject data collection

Human subject data were obtained from (1) a database of EEG recordings from patients with AS seen at Massachusetts General Hospital (MGH); and (2) a database of EEG recordings from the multicenter AS Natural History Study (NHS; ClinicalTrials.gov identifier: NCT00296764) conducted as part of the Rare Diseases Clinical Research Network, Angelman, Rett and Prader-Willi syndrome consortium.

In (1), all AS patients with longitudinal EEG recordings obtained between 2005 and 2019 were included. To prevent a disproportionate impact of subjects with multiple visits, and to represent EEG recordings obtained across varying intervals of elapsed time between visits (i.e., inter-visit intervals or IVIs), pairings of longitudinal EEG recordings separated by hourly, daily, weekly, monthly, and yearly IVIs were included with an approximately equal distribution, as available for each subject. In total 116 EEG recordings from 26 subjects (age 0.89 - 32.5 years, 8F:18M, 2-13 visits per subject) were included.

Recordings in this dataset were separated by a median of 8.2 months (range: 2 days - 4.4 years).

In (2), subjects were recruited at six sites between 2006 and 2017 and EEG recordings were collected from the sites at Rady Children's Hospital/University of California San Diego and Boston Children's Hospital. Consent was obtained according to the Declaration of Helsinki and was approved by the institutional review boards of the participating sites. In total, 88 EEG recordings from 30 subjects (age 1.3-21 years, 9F:21M, 2-6 visits per subject) were included. Recordings in this dataset were separated by a median of 1.1 years (range: 8.4 months - 7 years).

The final combined dataset included 56 subjects with longitudinal EEG data from 204 total visits.

All EEG recordings were collected using the international 10-20 montage on either BioLogic or Xltek systems (Klem, Lüders, Jasper, & Elger, 1999). For the MGH data, five recordings lasting approximately 24 hours were available. For the remainder of the data, recordings were of approximately 50 minutes duration (mean 50 min, range: 3 min - 6.5 hrs). For the NHS data, 30 minutes of wake recording and 30 minutes of sleep recording were attempted at each session (mean 28 min, range 1 min – 2.8 hrs). Some of the NHS subjects contributed EEG recordings that were obtained for clinical purposes.

All EEG data were manually staged for wake and sleep states by an experienced clinical neurophysiologist (CJC), and wake data selected for model development. We note that delta rhythms are abnormal during wake and sleep in AS (Ostrowski et al., 2021; Sidorov et al., 2017). We focused on wake data because it was more widely available and

more reliably identified. The final dataset included wake recordings of mean duration 37.4 minutes (range 1 min - 6.5 hrs).

4.2.2 *Human EEG data analysis*

Power spectra were calculated using the Chronux toolbox (Bokil et al., 2010). Following the procedure in Ostrowski et al., (2021), to reduce the impact of muscle artifacts, we analyzed only occipital and parietal electrodes, O1, O2, P3, Pz and P4 referenced to a group average. For each channel, we computed the power spectrum on non-overlapping one second intervals (1 Hz frequency resolution; single Hanning taper). Within each interval, we estimated the relative delta power as the average 2-4 Hz summed power over channels divided by the 1-50 Hz total power over channels. We then averaged the delta power over all intervals to yield a single relative delta power value for each subject.

To examine the stability of delta power estimates across full day recordings, we analyzed EEG recordings from five subjects, each with at least 24 hours of continuous recording and different genotypes (n=2 deletion, n=2 *UBE3A* mutation, n=1 uniparental disomy). For this, we implemented a resampling procedure to estimate the delta power from 1 to 3,600 randomly sampled (without replacement) 1-second epochs from the 24-hour dataset, reflecting data durations ranging from 1 second to 60 minutes. For each sample size (1 to 3,600 seconds), and for each subject, we repeated this procedure 1000 times to compute the standard error of the mean (SEM) delta power estimate. For a given estimate of relative delta power, δ , we empirically estimate the 95% confidence interval.

For our natural history model, we fit a linear mixed effect model with dependent

variable delta power at a second time point, and predictors of delta power at a previous time point, age, IVI, and subject as a random effect. We declared a variable significant when the p-value of the F-test was less than 0.05. We also tested the significance of including genotype in the model by evaluating the Akaike information criterion.

To simulate treatment and control groups for power calculations, we sampled (with replacement) 25, 50, 100, or 150 patient EEG recording pairs for each group from the longitudinal dataset. Then, from each subject in the simulated treatment group, we subtracted a fixed offset from the subject's observed second time point delta power to represent a treatment effect under the hypothesis that treatment reduces delta power. We considered delta power offsets ranging from 0 to 0.1. For each fixed delta power offset, we repeated this simulation 2000 times and computed the proportion of times a difference between simulated treatment and simulated control groups was detected using a one-sided t-test with p-value < 0.05.

4.2.3 *Animals*

All experiments were conducted in compliance with the rules set forth by the Biogen Institutional Animal Use and Care Committee in accordance with the guidelines established in the National Institutes of Health Guide for the Care and Use of Laboratory Animals. Mice were group housed on a 12 h light/dark cycle with ad libitum access to food and water. Both male and female offspring were used for experiments. Breeding was performed internally at Biogen by crossing female *Ube3a*^{m+/p-} (JAX Stock No: 016590) × male *Ube3a*^{m+/p+} (JAX Stock No: 000664) breeders to generate offspring, including

experimental *Ube3a*^{m⁻/p⁺} (AS) mice and littermate Wild-Type (WT) *Ube3a*^{m⁺/p⁺} controls. Wild-type and AS littermates were housed in the same cage whenever possible.

4.2.4 Oligonucleotides

Synthesis and purification of all chemically modified oligonucleotides was performed as previously described (Swayze et al., 2007). The 2'-MOE gapmer ASOs are 20 nucleotides in length, wherein the central gap segment comprising ten 2'-deoxynucleotides is flanked on the 5' and 3' wings by five 2'-MOE modified nucleotides. The sequences of the ASOs are as follows: control ASO, 5'-CTATAGGACTATCCAGGAA-3' and *Ube3a-ATS* ASO, 5'-CCAGCCTTGTTGGATATCAT-3'.

4.2.5 ASO in vivo administration

Lyophilized ASOs were dissolved in sterile PBS without calcium or magnesium and quantified by ultraviolet spectrometry. The ASOs were then diluted to the desired concentration required for dosing mice and sterilized through a 0.2 µm filter. Surgeries were performed +/- 3 days of postnatal day 35 (P35). Mice were anaesthetized with 2% isoflurane and placed in a stereotaxic frame (David Kopf Instruments). After exposing the skull, a needle (Hamilton, 1701 RN 10 µl micro syringe, needle 26 s/2''/2) was used to penetrate the skull at 0.3 mm posterior and 1.0 mm lateral to the bregma and lowered to a depth of 2.25-3.0 mm (based on weight), to deliver a non-targeting control ASO or *Ube3a-ATS* ASO (500 µg) at a rate of approximately 1 µl per 30 s into the cerebral ventricle. The needle was left in place for 5 min, slowly withdrawn and the incision was sutured.

4.2.6 SDS-PAGE Western Blotting

Mice were euthanized 8 weeks after intracerebroventricular (ICV) injection and the brains were rapidly harvested and cortical pinches of gray matter (200-300mg) were flash-frozen. The tissues were homogenized and lysed in Pierce RIPA lysis and extraction buffer (Thermo Fisher Scientific, Waltham, MA) supplemented with 1% Halt protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific, Waltham, MA) and centrifuged at 14,000×g for 20 min at 4 °C to clear the lysate. Protein concentrations were determined using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA) and 10 µg of each sample was denatured in 6X SDS sample buffer (Boston Bioproducts, Ashland, MA) by boiling for 8 min at 90 °C. Proteins were loaded into a Criterion 7.5% tris-glycine gel (Bio-Rad, Hercules, CA) and separated by SDS-PAGE at 120 V for 120 min. Gels were transferred to an IBlot2 nitrocellulose membrane (Invitrogen, Carlsbad, CA), blocked with TBST blocking buffer (Li-cor Biosciences, Lincoln, NE) for 1 h, and washed three times with TBST. The membrane was probed with primary antibodies (1:1000 dilution) in antibody dilution buffer (1:1 TBST blocking buffer and 1X TBST) overnight at 4 °C. The following antibodies were used for immunoblotting: UBE3A (1:1000; E8655, Sigma-Aldrich) and GAPDH (1:3000; 5174S, Cell Signaling Technology). After primary antibody staining, the blot was washed in triplicate with TBST and incubated for 1 h with secondary antibody (1:10,000 dilution of IR Dye 800 anti-mouse IgG and IR Dye 680 anti-rabbit IgG, Li-cor Biosciences, Lincoln, NE) in antibody dilution buffer. After a final triplicate wash with TBST, the blot was visualized using the Odyssey CLx imaging system (Li-cor Biosciences, Lincoln, NE).

4.2.7 *Quantitative Reverse Transcription Polymerase Chain Reaction*

Samples for PCR were lysed in RLT buffer (Qiagen) + 0.1% β -mercaptoethanol and total RNA was extracted using the RNeasy mini kit (Qiagen) and RNase-free DNase set (Qiagen) following the manufacturers protocol. Total RNA concentration was determined using the Nanodrop 8000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and reverse transcription was performed using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems) using up to 2 μ g of RNA as a template. 50 ng of the resulting cDNA product was subjected to duplex PCR reactions using Gene Expression Master Mix (Applied Biosystems, Waltham, MA) containing Taqman primers for *Ube3a* (Cat# Mm00839910_m1), *Ube3a-ATS* (Cat# Mm02580988_m1) and housekeeping gene GAPDH (Cat# Mm99999915), all primers are from Applied Biosystems). Real-time PCR reactions were performed on the Via7 Real-time PCR system (Thermo Fisher Scientific, Waltham, MA) using the following thermocycling conditions: 2 min at 50 °C, 10 min at 95 °C, followed by 50 cycles of 15 s at 95 °C and 1 min at 60 °C. Relative gene expression levels of *Ube3a* and *Ube3a-ATS* were calculated using the ddCt algorithm.

4.2.8 *Surgeries and Local Field Potential (LFP) Recordings*

Mice were surgically implanted with depth electrodes targeting layer IV of primary visual cortex 4-7 days after ASO administration (~P42). Primary visual cortex was chosen to replicate and extend the delta phenotype observed previously in Sidorov et al. (2017). Mice were anaesthetized with 2% isoflurane and placed in a stereotaxic frame (David Kopf

Instruments). A steel headpost was affixed to the skull anterior to bregma using cyanoacrylate glue. Burr holes (< 0.5 mm) were then drilled in the skull over binocular V1 (3.2 mm lateral of lambda). Tungsten electrodes (FHC, Bowdoinham, ME, US), 75 μ m in diameter at their widest point, were implanted in each hemisphere, 450 μ m below cortical surface. Reference electrodes consisted of 000-120 CS screws inserted into the skull touching dura located over prefrontal cortex. Wires extending from the electrodes were connected to female gold pins and inserted into a plastic pedestal connector (MS363 PlasticsOne, Roanoke, Va). Finally, dental cement (C&B Metabond, Parkell Inc., Bentwood, NY) was applied to form a stable, protective head-cap. Recordings were performed using a Tucker-Davis Technologies (TDT) R25D system at ~ 3 K sampling rate. Mice were head-fixed for all recording sessions viewing a full-field gray screen in an enclosed dark, quiet environment. They were habituated to this environment for 2 consecutive days, at 15 minutes per session, after-which recordings were acquired during three consecutive daily 15-minute sessions for at least 2 weeks post ICV infusion and then, for a portion of mice, recorded weekly thereafter.

4.2.9 *Mouse LFP data analysis*

LFP analyses were performed with the experimenter blinded to treatment and genotype. Power spectra were calculated for each hemisphere using the Chronux software package (Bokil et al., 2010) in MATLAB (Mathworks) for the last 10 minutes of the ~ 15 minute recording session and averaged across hemispheres (5 tapers, time bandwidth product of 3, using 5 second windows with a 1 second overlap). A custom algorithm was

used to remove recordings from channels with noisy LFPs, in some cases due to damage to the electrode. Delta activity was measured as the peak activity between 1 and 5Hz.

To examine model performance in the AS mouse model, we applied the natural history model developed with the human data to the longitudinal mouse LFP recordings. To estimate the model parameters for the mouse data, for each control ASO-treated mouse, pairs of longitudinal recordings were defined using all possible combinations of subsequent recordings, resulting in 6 to 21 longitudinal pairs per mouse, and 546 total longitudinal pairs.

We then applied the natural history model to estimate a treatment effect in *Ube3a-ATS* ASO-treated versus control ASO-treated mice. To do so we compared the distribution of model residuals for the *Ube3a-ATS* ASO-treated and control ASO-treated mice. We note that because electrode implantation interferes with subsequent ICV injections (due to the required use of dental cement on the skull), we did not have pre-treatment delta power values in the *Ube3a-ATS* ASO-treated mice. Therefore, to compute the model residual for an ASO-treated mouse, we selected (randomly) the delta power value from a mouse in the youngest control ASO-treated group and used the ASO-treated mouse's subject-specific variance to predict the delta power value of the ASO-treated mouse at all subsequent ages. We repeated this procedure to compute the model residuals for each *Ube3a-ATS* ASO-treated mouse at all subsequent visits. We applied the same procedure to compute the model residuals for the control ASO-treated mice, while excluding pairing of the same mouse at both visits. To compute a treatment effect size for each postnatal week, we computed the difference in the median residuals of the *Ube3a-ATS* ASO-treated mice and

the control ASO-treated mice. We repeated this entire procedure 10,000 times to generate a distribution of treatment effect sizes for each postnatal week.

To test the null hypothesis of no difference in median residuals between *Ube3a-ATS* ASO-treated and control ASO-treated mice, we implemented a resampling procedure. Under the null hypothesis of no difference between groups, we first created a combined group including both the control ASO-treated (n=26) mice and *Ube3a-ATS* ASO-treated (n=15) mice at week 2 (approximately postnatal day 49). Then, from this list, we selected randomly (without replacement) two pseudo-groups of mice to represent pseudo-control ASO-treated (n=26) and pseudo-ASO-treated (n=15). We repeated this entire procedure 10,000 times to create distributions of treatment effect size for each week, assuming no difference between groups. To test our hypothesis that *Ube3a-ATS* ASO-treated mice would have larger positive residuals compared to control ASO-treated mice, we compared the observed and resampled distributions of treatment effect size at each week using a one-sided t-test with significance at $p < 0.05$.

4.2.10 Statistical analysis for Ube3a mRNA upregulation and model residuals

To assess the correlation between the relative *Ube3a* mRNA upregulation and the model residual, we utilized the natural history model to compute residuals for mice at Week 7 and compare them to measured *Ube3a* mRNA levels obtained in the same mice at Week 8, in both control ASO-treated (n=8) and *Ube3a-ATS* ASO-treated (n=4) mice. To compute model residuals, for each Week 7 mouse, we compared the predicted and observed delta values using a randomly selected control ASO-treated mouse at Week 2. We then matched

this residual with the *Ube3a* mRNA measurements from the same mouse. We repeated this procedure for all 12 mice with *Ube3a* mRNA measurements at Week 8. We then performed linear regression to estimate the slope from a dataset of 12 samples. We repeated this entire procedure 1000 times (each using randomly selected control ASO-treated mouse at Week 2) to create a distribution of slope estimates.

4.3 Results

4.3.1 *Delta power estimates are stable over a 24-hour period within an individual*

Spectral estimates from scalp EEG are highly variable within individuals over the course of years (Frohlich et al., 2019; Ostrowski et al., 2021). To assess the stability of delta power estimates over shorter time intervals, we compared delta power estimates using increasing data sample sizes, randomly selected from one-second intervals in 5 AS patients with 24-hour continuous recordings (see Methods). As the amount of data increases, the standard error decreases at a rate of approximately $1/\sqrt{x}$ (Figure 4.1). When the sample size exceeds 8 minutes, the SEM plateaus and remains less than 0.009 (95% CI [0.008, 0.01]). Therefore, we conclude that delta power estimates are stable over the course of a 24-hour period and can be reliably estimated from just 8 minutes (i.e., 480 one-second samples) of EEG data.

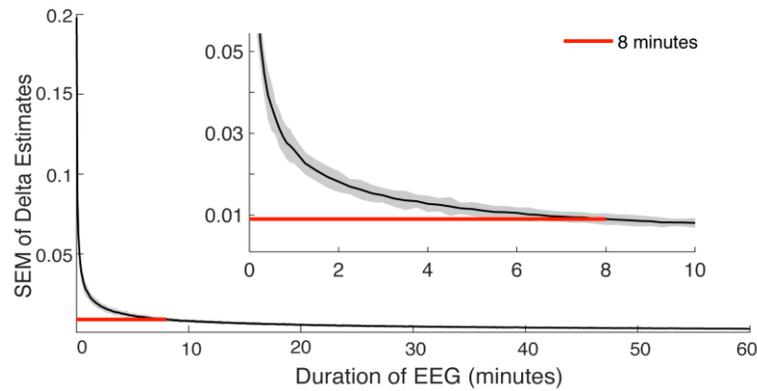


Figure 4.1: Only a few minutes of EEG data are needed to estimate delta power with high precision.

The black line represents the standard error of the mean (SEM) in the estimate of relative delta power calculated on a given amount of data and shaded bars represent 95% confidence. Red line indicates 8 minutes of data.

4.3.2 *Delta power at a future visit can be reliably predicted from a longitudinal natural history model*

We use the human longitudinal dataset to estimate parameters in a natural history model of delta power in AS to predict delta power at a future visit. To do so, we constructed a linear regression model with three predictors: delta power at the initial visit, $\log_{10}(\text{age})$ at the first visit (Chu et al., 2014; Sidorov et al., 2017), and elapsed time between visits (the IVI). Because the data consist of multiple longitudinal observations from repeat subjects, we include a random intercept to allow inter-subject variability in baseline delta power, for example due to genotype (Frohlich et al., 2019). The final natural history model is (Figure 4.2A):

$$\Delta_{\text{Visit } 2} \sim \Delta_{\text{Visit } 1} + \log_{10}(\text{Age}_{\text{Visit } 1}) + \text{IVI} + (1|\text{Subject}).$$

Fitting this model, we find that each predictor is significant: ($\log_{10}(\text{age})$, effect size -0.096,

95% CI [-0.215, -0.116], $p=0.005$; IVI, effect size -0.016, 95% CI [-0.029, -0.0016], $p=0.029$; Table 3). Delta power at the first time point (visit) is the strongest predictor; for every unit increase in delta power at the first time point, there is a 0.33 unit increase in delta power at the second time point (95% CI [0.17, 0.49], $p<1e-4$). We note that including genotype in the model did not improve model fit.

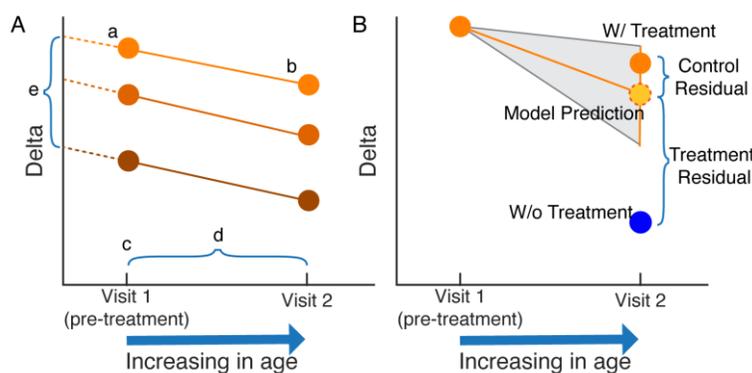


Figure 4.2: Natural history model overview.

(A) Schematic of variables included in the natural history model: a. Delta power at visit 1; b. Delta power at visit 2; c. Age at visit 1; d. Inter-visit interval (IVI); e. Random intercept.

(B) Schematic of procedure

to detect treatment effect. Using the natural history model, generate an estimate of delta power at visit 2 (model prediction) and compute the difference between predicted and observed delta power at visit 2 with (orange) and without (blue) treatment to compute the treatment and control residuals, respectively.

| | Parameter estimate | p-value |
|--|--------------------|----------|
| $\log_{10}(\text{Age}_{\text{Visit 1}})$ | -0.096 | 0.005 |
| IVI | -0.016 | 0.029 |
| $\text{Delta}_{\text{Visit 1}}$ | 0.33 | 6.40e-05 |

Table 3: Natural history model parameter estimates and p-values.

Delta power at a return visit depends on age, inter-visit interval (IVI), and delta power at prior visit.

4.3.3 *Power to detect a deviation from the natural history model following treatment*

An accurate natural history model can be used to identify a deviation from the expected natural history due to treatment, *i.e.*, a treatment effect. Delta power is abnormally increased in AS (Frohlich et al., 2019; Martinez et al., 2020; Sidorov et al., 2017) and higher delta power correlates with more severe disease (Frohlich et al., 2019; Ostrowski et al., 2021); therefore, we expect effective treatment to reduce delta power (illustration in Figure 4.2B). To determine if delta power is significantly reduced beyond the expected natural variability for an individual subject, the model-predicted delta power can be compared to the observed delta power at a future visit. To test the hypothesis that treatment impacts delta power beyond natural variability, we compare the residuals, or deviations from the model, of the simulated control and treatment groups. We expect a treatment effect that reduces delta power would result in significantly larger model residuals in the simulated treatment group compared to the simulated control group (Figure 4.2B).

To determine the sensitivity of the natural history model to detect a difference between groups, we simulate power calculations based on varying sample sizes and treatment effect sizes. To illustrate this simulation procedure, we consider a theoretical control group (n=50 subjects) and a simulated treatment group (n=50 subjects, see Methods). For the treatment group, we simulate varying treatment effects and sample sizes. For example, if we assume that treatment results in a 0.1 decrease in the observed delta power, using the longitudinal natural history model, we find larger residuals in the treatment group compared to the control group in this example, as expected (effect size 0.11, 95% CI [0.08, Inf], $p < 1e-8$, Figure 4.3A,B). Repeating this simulation for different

effect and sample sizes (Figure 4.3C), we conclude we can detect with 80% power a treatment effect size of 0.064 relative delta power in a sample of 25 patients per group, 0.046 in a sample of 50 patients per group, 0.033 for 100 patients per group, and 0.027 for 150 patients per group.

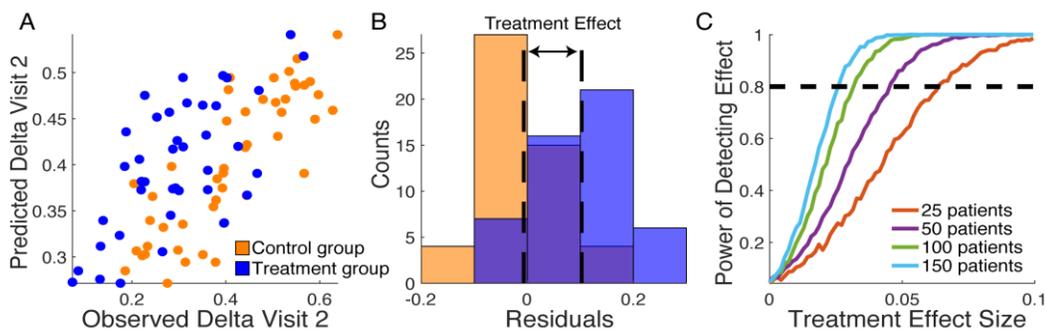


Figure 4.3: Simulation of model implementation.

Example simulation (A-B) where the effect size of the treatment group is a 0.1 reduction in delta power. (A) Predicted values from the model versus the observed values of delta power at visit 2 for the control (orange) and treatment simulated data (blue). (B) Histograms of the corresponding control and treatment residuals. Black dashed lines indicate mean residuals of each group. Black arrow indicates the difference between mean residuals of each group, i.e., the treatment effect. (C) Power to detect a treatment effect versus treatment effect size when sampling 25, 50, 100 and 150 patients per group. Black dashed line indicates 80% power.

4.3.4 The natural history model of delta activity detects deviations in delta power in

ASO-treated AS mice

To determine whether the *Ube3a-ATS* ASO was able to correct abnormal delta-like (1-5Hz) oscillations, AS mice were injected with a single ICV infusion of either the non-targeting control ASO or the *Ube3a-ATS* ASO (500 ug) at postnatal day 35 (P35). WT mice only were injected with the control ASO. LFP recordings in the visual cortex were obtained weekly from 2 weeks through 8 weeks post ICV infusion (WT Control ASO (n=28), AS Control ASO (n=26), AS *Ube3a-ATS* ASO (n=15)). On visual inspection, delta

power in AS control-treated ASO mouse models decrease with age, consistent with prior reports (Sidorov et al., 2017) (Figure 4.4 orange). Direct comparisons of delta power did not detect a significant difference between AS *Ube3a-ATS* ASO mice and AS control ASO mice for any post treatment week (one sided t-test, $p > 0.05$) except at week 7 ($p < 1e-2$, Figure 4.4).

To assess the ability of the natural history model to detect treatment effects in the mouse data, we first estimated model parameters on all available longitudinal data from the AS mice treated with control ASO (see Methods). Doing so, we found trends in all three predictors consistent with analysis of the human data; relative delta power at recording 2 increased with delta power at recording 1 (effect size 0.06, 95% CI [-0.02, 0.14], $p = 0.14$), and decreased with $\log_{10}(\text{age})$ (effect size -255.1, 95% CI [-572.2, 62.1], $p = 0.11$) and IVI (effect size -40.44, 95% CI [-75.73, -5.15], $p = 0.02$).

To assess the impact of treatment, we compared model residuals for control ASO-treated and *Ube3a-ATS* ASO-treated mice using a resampling procedure (see Methods; Figure 4.5A). We computed the treatment effect (*i.e.*, median residual difference; Figure 4.5B) between the control ASO-treated and *Ube3a-ATS* ASO-treated mice for each age following treatment (example in Figure 4.5C) and compared to a null distribution (Figure 4.5D). We found a significant treatment effect for all ages lasting through the 8 weeks observed, with the strongest treatment effect present at age week 7, corresponding to 7 weeks after treatment. We conclude that the natural history model can detect a significant and long-lasting treatment effect in *Ube3a-ATS* ASO-treated mice compared to control ASO-treated mice.

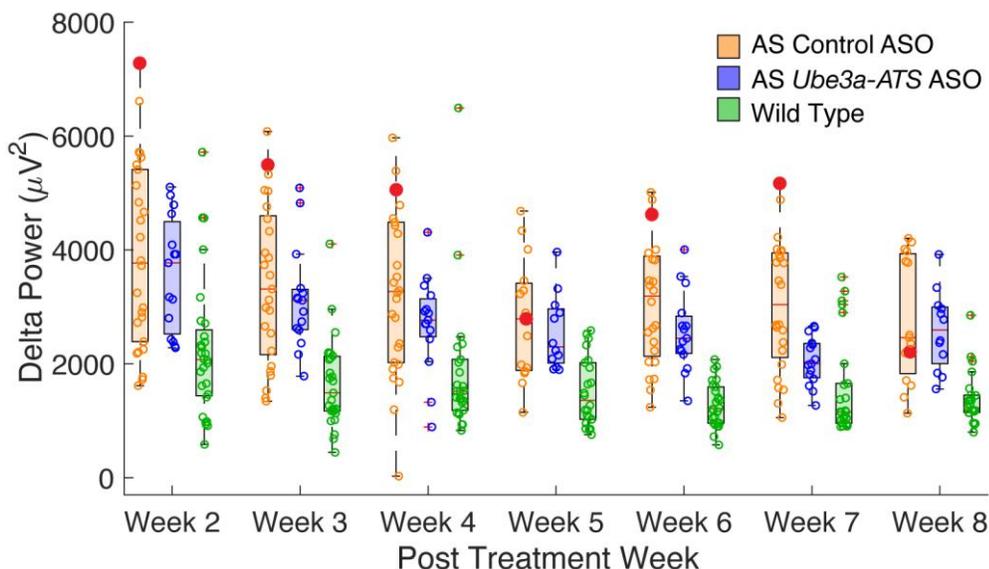


Figure 4.4: Longitudinal mouse LFP data.

Boxplots of delta power across age per group at each time point: mice with AS that did not receive control ASO treatment (orange), mice with AS that received *Ube3a-ATS* ASO treatment (blue), and wild-type mice (green). The filled marker (red) represents the progression of an example control ASO-treated mouse across weeks. The median of each group is indicated by a horizontal line. Using cross-sectional measures, no difference in delta power can be detected between control treated AS mice and *Ube3a-ATS* ASO treated AS mice at any week ($p > 0.05$) except week 7 ($p = 0.0016$). Control treated animals have higher delta values than wild type at each week ($p < 0.01$). ASO-treated animals have higher delta values than wildtype at each week ($p < 0.01$), except week 4 ($p = 0.2$).

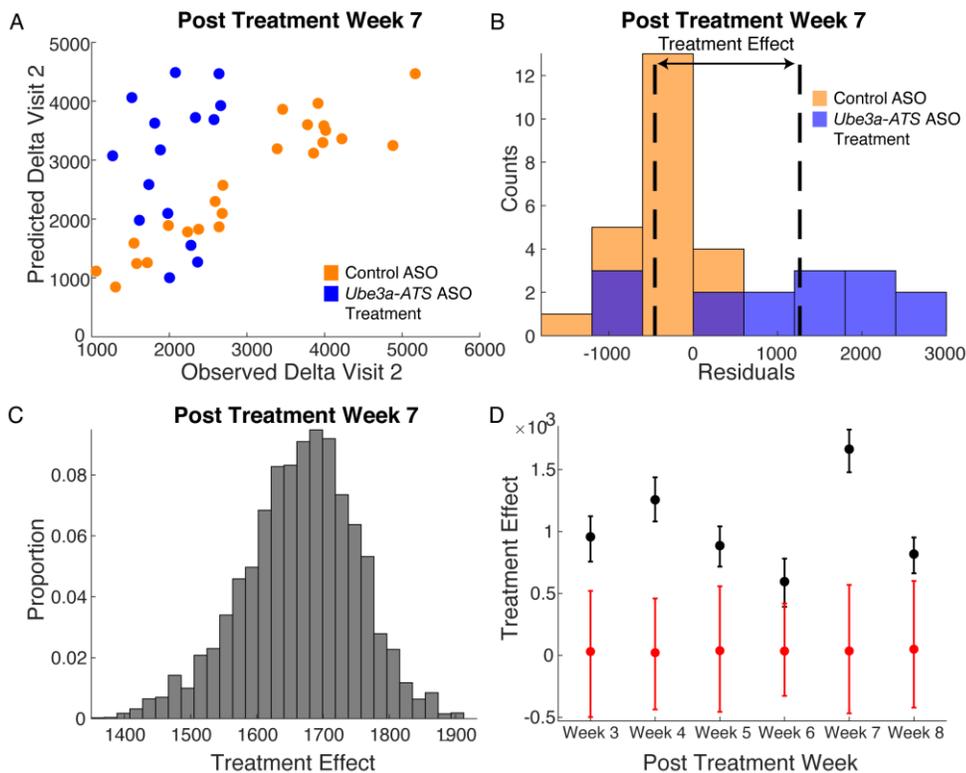


Figure 4.5: Application of natural history model to mouse model of AS.

- (A) Example of predicted versus observed delta power values at visit 2 from one prediction iteration of the resampling procedure corresponding to Week 7 post-treatment after *Ube3a-ATS* ASO (blue) or control ASO (orange) treatment.
- (B) Corresponding example histogram of *Ube3a-ATS* ASO treatment and control ASO residuals. Dashed lines indicate the median residual of each group. The distance between the dashed lines is the treatment effect ($\mu V^2/Hz$).
- (C) Example histogram of estimated treatment effect at Week 7 post-treatment from 10,000 iterations.
- (D) Treatment effect size versus age. Black circles (error bars) are mean (95% confidence bounds) of treatment effect at each age. Red circles (error bars) are the treatment effect between two groups under null hypothesis.

4.3.5 *Unsilencing of the Ube3a paternal allele with a Ube3a-ATS ASO in AS mice*

To assess *Ube3a* unsilencing *in vivo*, we compared the levels of *Ube3a-ATS* and *Ube3a* mRNA by qPCR in cortical tissue 8-weeks after treatment with control ASO or *Ube3a-ATS* ASO (500 ug) in WT (n=4 control ASO) or AS mice (n=10 control ASO; n=4

Ube3a-ATS ASO). We observed a ~70% reduction in *Ube3a-ATS* with the *Ube3a-ATS* ASO compared to control ASO treated mice at 8-weeks post-ASO administration. In a subset of mice, this level of *Ube3a-ATS* knockdown corresponded to ~2-fold increase in *Ube3a* mRNA levels compared to AS control ASO treated mice and levels that were ~50% of WT control ASO treated mice (Figure 4.6A). To understand the correlation between *Ube3a* mRNA and protein levels, we next quantified (by western blot) UBE3A protein levels in cortex in a subset of mice used for the RNA quantification at 8-weeks after ASO administration. After *Ube3a-ATS* ASO administration, UBE3A protein levels were 44% of WT control ASO-treated mice (~2-fold increase compared to the AS control-ASO group) 8-weeks post-ASO (Figure 4.6B). We conclude that *Ube3a-ATS* ASO ICV infusion successfully increases cortical UBE3A protein levels in AS mice.

4.3.6 *Increased model residuals correlate with increased Ube3a mRNA production*

To assess correlation between the relative *Ube3a* levels and the model residuals, we utilized the direct measures of *Ube3a* mRNA available for *Ube3a-ATS* ASO treated (n=4) and control ASO treated mice (n=8). We then implemented a resampling procedure (controlling for the small sample size, see Methods) to estimate the slope relating the model residuals and the relative *Ube3a* mRNA levels. We found a positive relationship between relative *Ube3a* mRNA and the model residuals (mean(std) slope = 29.17(0.73), standard error = 0.02), all fitted slopes are positive with $p < 0.001$, Figure 4.6C,D). We conclude that larger deviations from the natural history of delta power correlate with increased relative *Ube3a* expression.

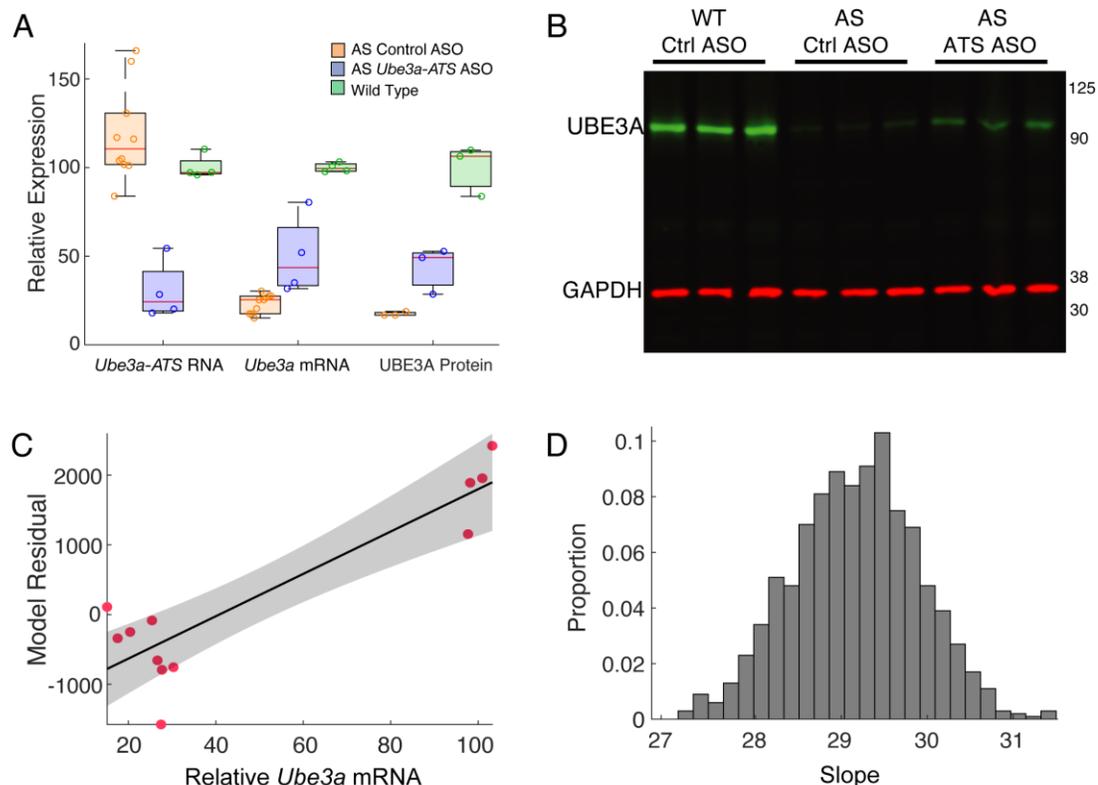


Figure 4.6: Unsilencing of the *Ube3a* paternal allele with a *Ube3a-ATS* ASO in AS mice leads to increase *Ube3a* mRNA and UBE3A protein and correlation to model residuals.

(A) *Ube3a-ATS* RNA levels (normalized to Wild-Type (WT) control) after treatment with a non-targeting control ASO (orange) or a *Ube3a-ATS* ASO in Wild-type or Angelman syndrome (AS) mice. Mice were ICV dosed at P35 and cortical tissue was collected at 8-weeks post-ASO treatment. (*Ube3a-ATS*: WT Control ASO = $100 \pm 3\%$, $n = 4$; AS Control ASO = $113 \pm 8\%$, $n = 10$; AS *Ube3a-ATS* ASO = $30 \pm 8\%$, $n = 4$) (left). *Ube3a* mRNA levels (normalized to WT control) after control or *Ube3a-ATS* ASO treatment (WT Control ASO = $100 \pm 1\%$, $n = 4$; AS Control ASO = $22 \pm 2\%$, $n = 10$; AS *Ube3a-ATS* ASO = $50 \pm 11\%$, $n = 4$) (middle). UBE3A protein levels. UBE3A signal intensity was quantified relative to GAPDH (*Ube3a*: WT Control ASO = $100 \pm 8\%$, $n = 3$; AS Control ASO = $17 \pm 1\%$, $n = 3$; AS *Ube3a-ATS* ASO = $44 \pm 8\%$, $n = 3$) (right). Each group (*Ube3a* mRNA, *Ube3a-ATS* RNA and UBE3A protein) was normalized to its own respective WT control group.

(B) Western blot from WT and AS mouse cortical tissue.

(C) Example model residuals (red dots) and *Ube3a* mRNA expression (normalized to WT control) and the linear fit to these data (mean, black line; gray shaded region, 95% confidence intervals).

(D) Histogram of estimated slopes from all resamples.

4.4 Discussion

As potentially disease-modifying therapies are on the horizon for AS patients, reliable methods to measure and detect treatment response are needed. Here, we utilized a large database of longitudinal EEG recordings from AS patients and to develop a natural history model of delta power in this disorder. As demonstrated in simulation, the model allows estimation of the populations required to detect treatment effects of various sizes on delta power for use in clinical trial planning. We also validated the therapeutic utility of this model in showing that it can detect deviations in delta power in a mouse model of AS following *Ube3a-ATS* ASO treatment compared with mice treated with a control ASO. These results support utilizing non-invasive measures of delta power to demonstrate target engagement and potential treatment effect in human clinical trials in AS.

Abnormally increased delta power is a consistent electrophysiologic phenotype of AS across species that shows promise to provide a meaningful biomarker for treatment efficacy. One challenge in employing this biomarker in clinical trials is that delta power varies dramatically across subjects and with age (Frohlich et al., 2019; Ostrowski et al., 2021). Here, we showed that stable estimates of delta power can be obtained from less than 10 minutes of EEG data. We then used a large longitudinal human dataset to develop a natural history model of delta power and showed that an individual's future delta power values could be predicted using readily available non-invasive clinical data (*e.g.*, delta power at a previous visit, age, and IVI). By including the prior delta power estimates and a random intercept, the longitudinal model controls for specific variations in baseline delta power due to genotype and disease severity and reduces noise due to intersubject

variability. Noise is further reduced in our longitudinal model by accounting for the known impact of IVI and interaction of age:genotype on delta power. This model then can be used to identify deviations in delta power outside of the expected natural history with confidence and thereby identify a significant treatment effect.

Using the natural history model, we performed power calculations to characterize the effect sizes and sample sizes needed to identify a treatment effect. We showed by simulation that in a sample of approximately 50 treated and 50 untreated AS patients, a decrease in delta power of only .046 could be detected with 80% confidence. This reduction in delta power corresponds to an increase in the raw score of the cognitive domain of the Bayley Scales of Infant and Toddler Development, 3rd edition (Ostrowski et al., 2021) of only 0.6 points. We conclude that small deviations in delta power, potentially signaling a treatment effect, can be detected by the model, even when below the threshold of detection on a performance-based test.

Several groups have reported promising approaches to reinstate *Ube3a* expression by unsilencing the paternal allele using an ASO treatment in mouse models of AS (Meng et al., 2015). Similar to these reports, we found that *Ube3a-ATS* ASO treatment correlated with increased *Ube3a* expression. Using our natural history model of delta power, we were able to both detect a persistent treatment effect following *Ube3a-ATS* ASO treatment and found that the changes in delta power correlated with *Ube3a* expression. These differences were not reliably detected using direct comparisons, but by controlling for variations expected with age, IVI, and between subjects, the longitudinal natural history model had the power to detect a significant treatment effect between the treated and control groups,

beyond changes expected with the natural history. The mice data analyzed here were limited by requiring us to implement model predictions between different individual mice, as in a cross-sectional dataset. If longitudinal data were available, we would expect the model to be even more sensitive to detect a treatment effect between groups. Additionally, *Ube3a-ATS* ASO treatment has been shown to result in a sustained increase in *Ube3a* expression up to four months after treatment, corresponding to improved synaptic plasticity and cognitive functions (Meng et al., 2015). The longitudinal model developed herein enables testing for and detection of a treatment effect at any IVI post treatment, thus enabling detection of peak effects and duration. Future work evaluating the natural history of delta power across the entire lifespan in AS mouse models, including early development, may be helpful to elucidate the impact and duration of *Ube3a-ATS* ASO treatment at different ages.

With many potentially disease-modifying treatments for AS in development, we introduce a natural history model and statistical procedure to utilize deviations from expected measurements of delta power as a sensitive indicator of target engagement and possible treatment efficacy. Measures of UBE3A expression in mouse models can be done by directly assaying neuronal tissue. Estimates of UBE3A neuronal levels in humans may be performed indirectly through cerebrospinal fluid (CSF) sampling using lumbar puncture. A reliable EEG biomarker to complement CSF data would reduce the need for invasive procedures and would enable multiple repeated measures for longitudinal observations. We find that delta power provides a simple, non-invasive alternative to invasive UBE3A measurements. Future work to validate the relationship between delta

power and UBE3A expression after effective treatment in humans with AS would secure delta power as a mechanistic biomarker to gauge both target engagement and therapeutic response in clinical trials.

4.4.1 Competing interests

CJC and MAK consult for Biogen. RWK, JPG, CB, FM, MC and SK are employees and shareholders of Biogen, Inc. The work presented here was funded in part by Biogen. PJ and FR are employees and shareholders of Ionis Pharmaceuticals. LMB receives funding from Biogen to conduct clinical trials.

5 CHAPTER FIVE

Conclusion

In this dissertation, we addressed various computational challenges when analyzing brain data in three different translational neuroscience applications. Using disease inspired methodology and developing interpretable models, we tested hypotheses underlying the mechanisms of brain dysfunction. Here we summarize the contributions of this dissertation to biomarker discovery and statistical modeling in elucidating the neural mechanisms of disease and providing potential avenues for treatment.

5.1 Innovation and Impact

5.1.1 Disease inspired methodology

In studying Rolandic epilepsy and Angelman syndrome, we encountered a common problem in neural data analysis: limited signal in the presence of highly variable data. Data are limited due to the nonstationarity of brain activity and the fact that spontaneous activity restricts analysis to one trial (Cohen, 2014). Brain rhythms are also highly dynamic and variable within and between subjects to support the diversity of brain functioning (Buzsáki, 2006; Chu et al., 2014). These facts in tandem lead to statistically underpowered models. To address these challenges, we (i) reduced the dimensionality of the models by incorporating biological assumptions, (ii) quantified the variability in spectral estimation by resampling the data, and (iii) leveraged knowledge of inter-subject variance by implementing mixed effects models.

For (i), we incorporated knowledge of biological processes to increase statistical power in functional network inference in Chapters 2 and 3. Inferring functional network models are limited in statistical power because there are many tests performed (to infer a network of size N nodes, there can be N^2 tests performed), and uncertainty exists in the network measurements themselves (Kramer et al., 2009), thus *a priori* knowledge is critical when employing such techniques. In Chapter 2, we extended a common method of functional inference, Granger causality. Granger causal analysis was originally developed for application to economic time series but is prevalently used in neuroscience due to its utility in determining conditional dependencies between brain regions (Bastos & Schoffelen, 2016; Granger, 1969). To increase statistical power in the Granger causality model, we adopted methodology from the spike train literature in which spike trains are modeled as point processes estimating the probability of spiking as a function of history. In Frank et al., (2002) and Eden et al., (2012), the authors used a lower dimensional spline basis to estimate the coefficients of the lagged history dependent terms thus reducing the number of parameters and permitting longer time lags to be included in the model. This is important because in spike trains longer lags account for the refractory period of neurons. Similarly, in electrophysiological data, which integrates activity over many neurons, longer lags are needed to account for the multiple delays and timescales influenced by neural mechanisms, such as, the myelination of connections and distance between regions. Thus, by incorporating knowledge of neural signaling, we developed a spline-Granger model that more accurately represents neuronal signaling dynamics. In Chapter 3, we developed functional network models to understand how Rolandic epilepsy disrupted cognitive

function. There are myriad ways to measure functional connectivity (Bastos & Schoffelen, 2016). However, we refined our analysis to assess coupling in a specific frequency band, sigma (10-15 Hz), *i.e.*, the frequency of sleep spindle oscillations, because spindles are disrupted in Rolandic epilepsy and support overnight learning and memory (Kramer et al., 2021). Additionally, in this dataset, we had 324 source-localized electrical recordings, meaning potentially 324^2 tests per subject. However, we restricted our analysis to a specific brain region, the inferior Rolandic cortex, *i.e.*, the affected region of brain in Rolandic epilepsy because the disease is focal, greatly reducing the number of tests. In both cases of functional connectivity in Chapters 2 and 3, we included biologically motivated methodology to increase confidence in our scientific inferences.

For (ii), in Chapters 3 and 4, we showed the utility of rhythmic biomarkers, sigma and delta power, for Rolandic epilepsy and Angelman syndrome, respectively; however, spectral estimates are highly variable across time within and between subjects. To increase precision of our sigma and delta power estimates, we used multitaper spectral analysis to gain independent samples from the data and more accurately estimate the spectral content (Bokil et al., 2010). Additionally, in Chapter 4, we quantified the variability in delta power estimates. We used 24-hour recordings from five subjects and resampled different durations of activity across the 24-hour period to estimate delta power precision as a function of duration of recording. We identified eight minutes as the smallest amount of data with sufficiently high precision. By reducing variability in estimates of sigma and delta power, we increased confidence when utilizing them as dependent variables in our models.

For (iii), we utilized random intercept models in Chapters 3 and 4 to address the high variability of behavioral and neural estimates. Instead of removing the correlation between data points (via averaging or only using one point per subject, *i.e.*, removing data), we modeled the correlation as a random intercept term to leverage the complete dataset. In Chapter 3, we used four measures of cognitive function to test our hypothesis that spindle rate correlates with cognitive function. One of the tasks was a fine motor task in which we had two measurements per patient for each hand's performance, which we modeled using a random intercept to account for baseline differences in subjects. For the other three tasks, we only had one measurement per patient, meaning roughly half as many data points as the fine motor task. We were still able to detect significant effects with cognitive function in these three tasks ($p < 0.1$ for all); however, the strongest effect was in the motor performance task (explained 90% of the variance). Additionally, in preliminary analysis, we averaged the two data points and only used dominant hand and implemented a linear regression model; however, the random intercept model was the most powerful approach⁵. In Chapter 4, we also used a random intercept in a longitudinal natural history model fit on patients with multiple EEG recording sessions. Delta power is highly variable in subjects with Angelman syndrome (Frohlich et al., 2019) thus making tracking changes in delta power within individual subjects challenging. However, on average, we were able to detect a

⁵ Approximately only 30% of the variance was explained in simple linear regression modeling only dominant hand motor performance or averaging motor performance as a function of spindle rate.

decrease in delta, as expected by age in Angelman syndrome (Frohlich et al., 2019) and consistent with what is observed in healthy subjects (Chu et al., 2014). The model disentangled reductions in delta power due to age and due to efficacious treatment in a simulated population of treated subjects and in a mouse model of AS that received an ASO treatment. To investigate changes in individual patients, we developed a bootstrap procedure to propagate delta measurement error and model error as required when computing individual predictions (Kass et al., 2014). However, because of the high variability in delta trajectories, detecting a treatment in an individual required a large treatment effect, *i.e.*, a large decrease in delta power, and was ultimately unnecessary for a clinical trial setting in which a mean effect across a population is sufficient.

5.1.2 Pathways to understanding disease

The primary goal in studying neurophysiological biomarkers is to increase understanding of disease and ultimately improve patient outcomes. In this thesis, we utilized interpretable models and hypothesis driven questions to understand the mechanisms of disease. First, in Chapter 2, we proposed a method that increased statistical power and interpretability in building functional network models. Functional network models are a useful lens into how disease alters brain connectivity (Bassett & Sporns, 2017). Second, in Chapters 3 and 4, we studied physiological biomarkers in two specific disease applications, Rolandic epilepsy and Angelman syndrome. We built interpretable models to make conclusions about how their respective biomarkers are impacted by and evolve with disease, identifying these biomarkers as potential therapeutic targets.

In Chapter 2, we developed a statistical modeling approach, spline-Granger causality, to help improve understanding of how neuronal signaling is impacted by disease. Because the coefficients of the spline-Granger model are in smooth spline basis, the magnitude of the coefficients versus time lag are interpretable: large coefficients indicate periods during which a signal exerts an influence upon itself, or upon another signal. Therefore, via this approach, we can not only infer the strength of coupling, but also the temporal relationship between signals.

In Chapter 3, we utilized source-localized EEG to characterize the spatial extent of the spindle deficits in Rolandic epilepsy. By identifying the specific cortical regions implicated by the disease, we can identify specific thalamic circuitry, either thalamic nuclei or cellular subtypes, that may be malfunctioning and leading to the cognitive deficits and seizures (Behrens et al., 2003; Clemente-Perez et al., 2017). Additionally, we developed models that indicated that (1) there is a spindle deficit in active, but not resolved epilepsy as compared to controls, and (2) as spindles increase, cognitive performance increases. This shows a temporal correlation between when the abnormal electrophysiological activity (both spindle and spiking activity) and cognitive deficits occur (Kramer et al., 2021). Additionally, Rolandic epilepsy subjects exhibit reduced white matter connectivity in resolution relative to controls (Thorn et al., 2020). In tandem, these results suggest there is some evolving corticothalamic circuitry that may be driving recovery. Clinically, this is important because it points to therapeutics that can restore healthy spindle activity and potentially hasten resolution of the cognitive comorbidities.

In Chapter 4, we developed a procedure to detect the effect of a therapeutic on delta activity in Angelman syndrome. Because the model can distinguish changes in delta power due to age and due to treatment, we can directly interpret the impact of the drug on the neural activity, *i.e.*, the treatment effect. For example, because we know delta power and cognition are inversely related as measured by the cognitive domain of the Bayley Scales of Infant and Toddler Development (Ostrowski et al., 2021), we can interpret the treatment effect in terms of clinically relevant outcomes such as cognitive score. For illustration, if the treatment reduces delta power by 0.05, the smallest treatment effect that can be detected by our proposed model in a sample of 50 patients with 80% confidence, this indicates a one point improvement in Bayley cognitive score. Additionally, in order to see a cognitive improvement relative to the difference between a deletion and non-deletion subject for a ten-year-old, a much larger reduction of delta power, 0.68, is needed. Finally, we applied this procedure to a mouse model of Angelman syndrome who received *Ube3a-ATS* ASO treatment and found that larger deviations from our model, *i.e.*, treatment effect, correlated with increased UBE3A expression, suggesting some potential mechanistic relationship between UBE3A and delta activity.

5.2 Future Directions

Neurophysiological biomarkers are enormously useful in that they reflect the dynamic aspects of the brain, measuring disease severity and progression. Here in this thesis, we have demonstrated how the EEG biomarkers, spindle rate and delta power, can be used to measure cognitive symptoms and disease state in Rolandic epilepsy, and to track disease course in Angelman syndrome. Neurophysiological biomarkers are dynamic by

definition and while we have addressed the evolution of both to some extent, more formalized approaches could be used to characterize spatiotemporal dynamics. In Rolandic epilepsy, we found that spindle deficits resolved with seizure resolution and correlated with cognitive deficits in a cross-sectional cohort of subjects, suggesting compensatory changes in the brain are driving resolution of spindle and cognitive deficits. Future longitudinal studies could help determine the exact time course of spindle rate resolution in relation to resolution of cognitive deficits within subjects to confirm our findings. Additionally, while we characterized to some extent the temporal dynamics of spindle rate in the inferior Rolandic cortex by estimating the co-occurrence of spindles, we observed a broader distribution of spindle deficits across cortex. Future work building a spatiotemporal model to characterize these dynamics across cortex would be useful in understanding how spindle production evolves spatially during sleep. Similarly, in Angelman syndrome, we studied how delta power changed across sessions on the timescale of days, months, weeks, and years, but we did not study how delta power evolved within an EEG recording session. Delta activity in Angelman syndrome has been reported to have unique temporal dynamics, exhibiting either “bursty” behavior or long runs of moderate to high amplitude slow waves. Furthermore, bursts of “notched” delta, when delta activity coincides with epileptic spikes or sharp waves, are a common characteristic of Angelman syndrome. Throughout childhood, delta activity becomes more regular and synchronous as well as having an increased incidence of notched delta activity (Korff, Kelley, & Nordli, 2005; Valente et al., 2003). Characterizing the evolution of delta activity in relation to cognitive symptoms may help identify potential mechanisms underlying poor cognition in Angelman syndrome.

Finally, we have demonstrated two examples of links between the lack of healthy brain rhythms and cognitive decline, leading to the fundamentally important question: is it sufficient to restore healthy brain rhythms to resolve cognitive comorbidities in disease? Auditory stimulation during sleep has been shown to increase production of sleep spindles (Antony & Paller, 2017) and suggests a potential therapeutic avenue for the cognitive symptoms of Rolandic epilepsy. In Angelman syndrome, ASO treatments have been shown to restore healthy levels of UBE3A expression and delta activity in mouse models. If these treatments improve brain rhythms, which in turn improve cognitive function in humans, then this suggests potential therapeutic avenues addressing the restoration of brain activity directly. Future work would help characterize the temporal relationship between restoration of healthy neural activity and cognition in Angelman syndrome and Rolandic epilepsy.

BIBLIOGRAPHY

- Andersen, P., Andersson, S. A., & Lomo, T. (1967). Nature of thalamo-cortical relations during spontaneous barbiturate spindle activity. *The Journal of Physiology*, *192*(2), 283–307. <https://doi.org/10.1113/jphysiol.1967.sp008300>
- Andersen, R. A., Aflalo, T., & Kellis, S. (2019). From thought to action: The brain-machine interface in posterior parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(52), 26274–26279. <https://doi.org/10.1073/pnas.1902276116>
- Angelman, H. (1965). ‘Puppet’ Children A Report on Three Cases. *Developmental Medicine & Child Neurology*, *7*(6), 681–688. <https://doi.org/10.1111/j.1469-8749.1965.tb07844.x>
- Anscombe, F. J. (1973). Graphs in Statistical Analysis. *The American Statistician*, *27*(1), 17–21. https://doi.org/10.1007/978-3-540-71915-1_35
- Antony, J. W., & Paller, K. A. (2017). Using oscillating sounds to manipulate sleep spindles. *Sleep*, *40*(3). <https://doi.org/10.1093/sleep/zsw068>
- Astradsson, A., Olafsson, E., Ludvigsson, P., Björgvinsson, H., & Hauser, W. A. (1998). Rolandic epilepsy: An incidence study in Iceland. *Epilepsia*, *39*(8), 884–886. <https://doi.org/10.1111/j.1528-1157.1998.tb01185.x>
- Barnett, L., & Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*, *223*, 50–68. <https://doi.org/10.1016/j.jneumeth.2013.10.018>
- Barrett, A. B., & Barnett, L. (2013). Granger causality is designed to measure effect, not

- mechanism. *Frontiers in Neuroinformatics*, 7(April), 6–7.
<https://doi.org/10.3389/fninf.2013.00006>
- Barrett, A. B., Murphy, M., Bruno, M.-A., Noirhomme, Q., Boly, M., Laureys, S., & Seth, A. K. (2012). Granger Causality Analysis of Steady-State Electroencephalographic Signals during Propofol-Induced Anaesthesia. *Public Library of Science ONE*, 7(1), e29072. <https://doi.org/10.1371/journal.pone.0029072>
- Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, 20(3), 353–364. <https://doi.org/10.1038/nn.4502>
- Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M., & Grafton, S. T. (2011). Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*, 108(18), 7641–7646.
<https://doi.org/10.1073/pnas.1018985108>
- Bassett, D. S., Yang, M., Wymbs, N. F., & Grafton, S. T. (2015). Learning-induced autonomy of sensorimotor systems. *Nature Neuroscience*, 18(5), 744–751.
<https://doi.org/10.1038/nn.3993>
- Bastos, A. M., & Schoffelen, J.-M. (2016). A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. *Frontiers in Systems Neuroscience*, 9(January), 1–23. <https://doi.org/10.3389/fnsys.2015.00175>
- Bastuji, H., Lamouroux, P., Villalba, M., Magnin, M., & Garcia-Larrea, L. (2020). Local sleep spindles in the human thalamus. *Journal of Physiology*, 598(11), 2109–2124.
<https://doi.org/10.1113/JP279045>
- Beelke, M., Nobili, L., Baglietto, M. ., De Carli, F., Robert, A., De Negri, E., & Ferrillo,

- F. (2000). Relationship of sigma activity to sleep interictal epileptic discharges: a study in children affected by benign epilepsy with occipital paroxysms. *Epilepsy Research*, 40(2–3), 179–186. [https://doi.org/10.1016/S0920-1211\(00\)00131-5](https://doi.org/10.1016/S0920-1211(00)00131-5)
- Beenhakker, M. P., & Huguenard, J. R. (2009). Neurons that Fire Together Also Conspire Together: Is Normal Sleep Circuitry Hijacked to Generate Epilepsy? *Neuron*, 62(5), 612–632. <https://doi.org/10.1016/j.neuron.2009.05.015>
- Behrens, T. E. J., Johansen-Berg, H., Woolrich, M. W., Smith, S. M., Wheeler-Kingshott, C. A. M., Boulby, P. A., ... Matthews, P. M. (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience*, 6(7), 750–757. <https://doi.org/10.1038/nn1075>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*. <https://doi.org/10.2307/2346101>
- Berg, A. T., Berkovic, S. F., Brodie, M. J., Buchhalter, J., Cross, J. H., Van Emde Boas, W., ... Scheffer, I. E. (2010). Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on Classification and Terminology, 2005-2009. *Epilepsia*, 51(4), 676–685. <https://doi.org/10.1111/j.1528-1167.2010.02522.x>
- Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv Für Psychiatrie Und Nervenkrankheiten*, 87(1), 527–570. <https://doi.org/10.1007/BF01797193>
- Besenyi, M., Varga, E., Fekete, I., Puskás, S., Hollódy, K., Fogarasi, A., ... Clemens, B.

- (2012). EEG background activity is abnormal in the temporal and inferior parietal cortex in benign rolandic epilepsy of childhood: A LORETA study. *Epilepsy Research*, 98(1), 44–49. <https://doi.org/10.1016/j.eplepsyres.2011.08.013>
- Besseling, R. M. H., Jansen, J. F. A., Overvliet, G. M., Van Der Kruijs, S. J. M., Vles, J. S. H., Ebus, S. C. M., ... Backes, W. H. (2013). Reduced functional integration of the sensorimotor and language network in rolandic epilepsy. *NeuroImage: Clinical*, 2(1), 239–246. <https://doi.org/10.1016/j.nicl.2013.01.004>
- Bi, X., Sun, J., Ji, A. X., & Baudry, M. (2016). Potential therapeutic approaches for Angelman syndrome. *Expert Opinion on Therapeutic Targets*, 20(5), 601–613. <https://doi.org/10.1517/14728222.2016.1115837>
- Bird, L. M. (2014). Angelman syndrome: review of clinical and molecular aspects. *The Application of Clinical Genetics*, 7, 93–104. <https://doi.org/10.2147/TACG.S57386>
- Birnbaum, M. H. (1973). The devil rides again: Correlation as an index of fit. *Psychological Bulletin*, 79(4), 239–242. <https://doi.org/10.1037/h0033853>
- Blumenfeld, H., & Taylor, J. (2003). Why do seizures cause loss of consciousness? *Neuroscientist*, 9(5), 301–310. <https://doi.org/10.1177/1073858403255624>
- Bokil, H., Andrews, P., Kulkarni, J. E., Mehta, S., & Mitra, P. P. (2010). Chronux: A platform for analyzing neural signals. *Journal of Neuroscience Methods*, 192(1), 146–151. <https://doi.org/10.1016/j.jneumeth.2010.06.020>
- Bonnefond, M., Kastner, S., & Jensen, O. (2017). Communication between Brain Areas Based on Nested Oscillations. *ENeuro*, 4(2), ENEURO.0153-16.2017. <https://doi.org/10.1523/ENeuro.0153-16.2017>

- Brankač, J., Scheffzük, C., Kukushka, V. I., Vyssotski, A. L., Tort, A. B. L., & Draguhn, A. (2012). Distinct features of fast oscillations in phasic and tonic rapid eye movement sleep. *Journal of Sleep Research, 21*(6), 630–633.
<https://doi.org/10.1111/j.1365-2869.2012.01037.x>
- Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., ... Bassett, D. S. (2015). Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences, 112*(37), 11678–11683. <https://doi.org/10.1073/pnas.1422487112>
- Bressler, S. L., & Seth, A. K. (2011). Wiener-Granger Causality: A well established methodology. *NeuroImage, 58*(2), 323–329.
<https://doi.org/10.1016/j.neuroimage.2010.02.059>
- Brown, E. N., & Kass, R. E. (2009). What Is Statistics? *The American Statistician, 63*(2), 105–110. <https://doi.org/10.1198/tast.2009.0019>
- Brown, E. N., Lydic, R., & Schiff, N. D. (2010). General Anesthesia, Sleep, and Coma. *New England Journal of Medicine, 363*(27), 2638–2650. Retrieved from <http://www.nejm.org/doi/10.1056/NEJMra0808281>
- Buiting, K., Williams, C., & Horsthemke, B. (2016). Angelman syndrome-insights into a rare neurogenetic disorder. *Nature Reviews Neurology, 12*(10), 584–593.
<https://doi.org/10.1038/nrneurol.2016.133>
- Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience, 10*(3), 186–198.
<https://doi.org/10.1038/nrn2575>

- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research, 33*(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Burns, S. P., Santaniello, S., Yaffe, R. B., Jouny, C. C., & Crone, N. E. (2014). Network dynamics of the brain and influence of the epileptic seizure onset zone. *Proceedings of the National Academy of Sciences of the United States of America, 111*(49), E5321–5330. <https://doi.org/10.1073/pnas.1401752111>
- Buzsáki, G. (2006). *Rhythms of the Brain. Rhythms of the Brain*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195301069.001.0001>
- Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science, 304*(5679), 1926–1929. <https://doi.org/10.1126/science.1099745>
- Buzsáki, G., & Watson, B. O. (2012). Brain rhythms and neural syntax: Implications for efficient coding of cognitive content and neuropsychiatric disease. *Dialogues in Clinical Neuroscience, 14*(4), 345–367. <https://doi.org/10.31887/dcns.2012.14.4/gbuzsaki>
- Calin-Jageman, R., & Cumming, G. (2019). Estimation for Better Inference in Neuroscience. *ENeuro, 6*(August). <https://doi.org/10.31234/osf.io/rzk5m>
- Callenbach, P. M. C., Bouma, P. A. D., Geerts, A. T., Arts, W. F. M., Stroink, H., Peeters, E. A. J., ... Brouwer, O. F. (2010). Long term outcome of benign childhood epilepsy with centrotemporal spikes: Dutch Study of Epilepsy in Childhood. *Seizure, 19*(8), 501–506. <https://doi.org/10.1016/j.seizure.2010.07.007>
- Camfield, C. S., & Camfield, P. R. (2014). Rolandic epilepsy has little effect on adult life

30 years later. *Neurology*, 82(13), 1162 LP – 1166.

<https://doi.org/10.1212/WNL.0000000000000267>

Carvill, G. L., Regan, B. M., Yendle, S. C., O’Roak, B. J., Lozovaya, N., Bruneau, N., ...

Mefford, H. C. (2013). GRIN2A mutations cause epilepsy-aphasia spectrum

disorders. *Nature Genetics*, 45(9), 1073–1076. <https://doi.org/10.1038/ng.2727>

Chu, C. J., Leahy, J., Pathmanathan, J., Kramer, M. A., & Cash, S. S. (2014). The

maturation of cortical sleep rhythms and networks over early development. *Clinical*

Neurophysiology, 125(7), 1360–1370. <https://doi.org/10.1016/j.clinph.2013.11.028>

Chu, C. J., Tanaka, N., Diaz, J., Edlow, B. L., Wu, O., Hämäläinen, M., ... Kramer, M.

A. (2015). EEG functional connectivity is partially predicted by underlying white matter connectivity. *NeuroImage*, 108, 23–33.

<https://doi.org/10.1016/j.neuroimage.2014.12.033>

Clemens, B., Puskás, S., Spisák, T., Lajtos, I., Opposits, G., Besenyi, M., ... Emri, M.

(2016). Increased resting-state EEG functional connectivity in benign childhood epilepsy with centro-temporal spikes. *Seizure*, 35, 50–55.

<https://doi.org/10.1016/j.seizure.2016.01.001>

Clemente-Perez, A., Makinson, S. R., Higashikubo, B., Brovarney, S., Cho, F. S., Urry,

A., ... Paz, J. T. (2017). Distinct Thalamic Reticular Cell Types Differentially

Modulate Normal and Pathological Cortical Rhythms. *Cell Reports*, 19(10), 2130–

2142. <https://doi.org/10.1016/j.celrep.2017.05.044>

Cohen, M. X. (2014). Granger Prediction. In *Analyzing Neural Time Series Data: Theory and Practice* (pp. 371–388). MIT Press.

- De Gennaro, L., & Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Medicine Reviews*, 7(5), 422–440. [https://doi.org/10.1016/S1087-0792\(02\)00116-8](https://doi.org/10.1016/S1087-0792(02)00116-8)
- de Pasquale, F., Della Penna, S., Snyder, A. Z., Lewis, C., Mantini, D., Marzetti, L., ... Corbetta, M. (2010). Temporal dynamics of spontaneous MEG activity in brain networks. *Proceedings of the National Academy of Sciences*, 107(13), 6040–6045. <https://doi.org/10.1073/pnas.0913863107>
- De Vico Fallani, F., Richiardi, J., Chavez, M., & Achard, S. (2014). Graph analysis of functional brain networks: practical issues in translational neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1653), 20130521–20130521. <https://doi.org/10.1098/rstb.2013.0521>
- Desai, M., Stockbridge, N., & Temple, R. (2006). Blood pressure as an example of a biomarker that functions as a surrogate. *The Journal of the American Association of Pharmaceutical Scientists*, 8(1), E146–E152. <https://doi.org/10.1208/aapsj080117>
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3), 968–980. <https://doi.org/10.1016/j.neuroimage.2006.01.021>
- Dimatteo, I., Genovese, C. R., & Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika*, 88(4), 1055–1071. <https://doi.org/10.1093/biomet/88.4.1055>
- Ding, M., Chen, Y., & Bressler, S. L. (2006). Granger Causality: Basic Theory and Application to Neuroscience. In *Handbook of Time Series Analysis* (pp. 437–460).

Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.

<https://doi.org/10.1002/9783527609970.ch17>

Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., ... Voytek,

B. (2020). Parameterizing neural power spectra into periodic and aperiodic components. *Nature Neuroscience*, *23*(12), 1655–1665.

<https://doi.org/10.1038/s41593-020-00744-x>

Durbin, J., & Watson, G. S. (1950). Trust Testing for Serial Correlation in Least Squares

Regression I. *Biometrika Trust*, *37*(1), 159–177.

Eden, U. T., Gale, J. T., Amirnovin, R., & Eskandar, E. (2012). Characterizing the

spiking dynamics of subthalamic nucleus neurons in Parkinson's disease using generalized linear models. *Frontiers in Integrative Neuroscience*, *6*(June), 28.

<https://doi.org/10.3389/fnint.2012.00028>

Engel, A. K., & Fries, P. (2010). Beta-band oscillations-signalling the status quo?

Current Opinion in Neurobiology, *20*(2), 156–165.

<https://doi.org/10.1016/j.conb.2010.02.015>

Engel Jr, J. (2011). Biomarkers in epilepsy: introduction. *Biomarkers in Medicine*,

11(62), 537–544. <https://doi.org/10.2217/BMM>

Englot, D. J., Konrad, P. E., & Morgan, V. L. (2016). Regional and global connectivity

disturbances in focal epilepsy, related neurocognitive sequelae, and potential mechanistic underpinnings. *Epilepsia*, *57*(10), 1546–1557.

<https://doi.org/10.1111/epi.13510>

Fama, R., & Sullivan, E. V. (2015). Thalamic structures and associated cognitive

- functions: Relations with age and aging. *Neuroscience and Biobehavioral Reviews*, 54, 29–37. <https://doi.org/10.1016/j.neubiorev.2015.03.008>
- Fejerman, N. (2009). Atypical rolandic epilepsy. *Epilepsia*, 50(SUPPL. 7), 9–12. <https://doi.org/10.1111/j.1528-1167.2009.02210.x>
- Fejerman, N., Caraballo, R., & Tenenbaum, S. N. (2000). Atypical evolutions of benign localization-related epilepsies in children: Are they predictable? *Epilepsia*, 41(4), 380–390. <https://doi.org/10.1111/j.1528-1157.2000.tb00177.x>
- Fernandez, I. S., Loddenkemper, T., Galanopoulou, A., & Moshe, S. L. (2015). Should epileptiform discharges be treated? *Epilepsia*, 56(10), 1492–1504. <https://doi.org/10.1111/epi.13108>.
- Fink, G. R., Frackowiak, R. S. J., Pietrzyk, U., & Passingham, R. E. (1997). Multiple nonprimary motor areas in the human cortex. *Journal of Neurophysiology*, 77(4), 2164–2174. <https://doi.org/10.1152/jn.1997.77.4.2164>
- Fischl, B. (2012). FreeSurfer. *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2012.01.021>
- Frank, L. M., Eden, U. T., Solo, V., Wilson, M. A., & Brown, E. N. (2002). Contrasting Patterns of Receptive Field Plasticity in the Hippocampus and the Entorhinal Cortex: An Adaptive Filtering Approach. *Journal of Neuroscience*, 22(9), 3817–3830. <https://doi.org/10.1523/jneurosci.22-09-03817.2002>
- Frauscher, B., Bartolomei, F., Kobayashi, K., Cimbalnik, J., van 't Klooster, M. A., Rampp, S., ... Gotman, J. (2017). High-frequency oscillations: The state of clinical research. *Epilepsia*. <https://doi.org/10.1111/epi.13829>

- Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, 88(1), 220–235. <https://doi.org/10.1016/j.neuron.2015.09.034>.Rhythms
- Frohlich, J., Bird, L. M., Dell’Italia, J., Johnson, M. A., Hipp, J. F., & Monti, M. M. (2020). High-voltage, diffuse delta rhythms coincide with wakeful consciousness and complexity in Angelman syndrome. *Neuroscience of Consciousness*, 6(1), 1–15. <https://doi.org/10.1093/nc/niaa005>
- Frohlich, J., Miller, M. T., Bird, L. M., Garces, P., Purtell, H., Hoener, M. C., ... Hipp, J. F. (2019). Electrophysiological Phenotype in Angelman Syndrome Differs Between Genotypes. *Biological Psychiatry*, 85(9), 752–759. <https://doi.org/10.1016/j.biopsych.2019.01.008>
- Frohlich, J., Toker, D., & Monti, M. M. (2021). Consciousness among delta waves: a paradox? *Brain*. <https://doi.org/10.1093/brain/awab095>
- Gałecki, A., & Burzykowski, T. (2013). Linear Mixed-Effects Model. In *Linear Mixed-Effects Models Using R: A Step-by-Step Approach* (pp. 245–273). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-3900-4_13
- Galicchio, S., Espeche, A., Cersosimo, R., Chacon, S., Gamboni, B., Adi, J., ... Caraballo, R. H. (2021). Self-limited epilepsy with centro-temporal spikes: A study of 46 patients with unusual clinical manifestations. *Epilepsy Research*, 169(October 2020), 106507. <https://doi.org/10.1016/j.eplepsyres.2020.106507>
- Gelman, A. (2014). *Bayesian data analysis* (Third edit). Boca Raton: Boca Raton : CRC Press.
- Gent, T. C., Bassetti, C. LA, & Adamantidis, A. R. (2018). Sleep-wake control and the

thalamus. *Current Opinion in Neurobiology*, 52, 188–197.

<https://doi.org/10.1016/j.conb.2018.08.002>

Gentile, J. K., Tan, W. H., Horowitz, L. T., Bacino, C. A., Skinner, S. A., Barbieri-Welge, R., ... Peters, S. U. (2010). A neurodevelopmental survey of Angelman syndrome with genotype-phenotype correlations. *Journal of Developmental and Behavioral Pediatrics*, 31(7), 592–601.

<https://doi.org/10.1097/DBP.0b013e3181ee408e>

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data.

NeuroImage. <https://doi.org/10.1016/j.neuroimage.2013.10.027>

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3), 424. <https://doi.org/10.2307/1912791>

Greenblatt, R. E., Pflieger, M. E., & Ossadtchi, A. E. (2012). Connectivity measures applied to human brain electrophysiological data. *Journal of Neuroscience Methods*, 207(1), 1–16. <https://doi.org/10.1016/j.jneumeth.2012.02.025>

Grenander, U., & Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time Series* (2nd ed.). Chelsea Publishing Company.

Gruber, R., & Wise, M. S. (2016). Sleep Spindle Characteristics in Children with Neurodevelopmental Disorders and Their Relation to Cognition. *Neural Plasticity*, 2016. <https://doi.org/10.1155/2016/4724792>

Hamalainen, M. S., & Sarvas, J. (1987). Feasibility of the homogeneous head model in the interpretation of neuromagnetic fields. *Physics in Medicine and Biology*, 32(1),

91–97. <https://doi.org/10.1088/0031-9155/32/1/014>

Hamalainen, M. S., & Sarvas, J. (1989). Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. *Institute of Electrical and Electronics Engineers Transactions on Biomedical Engineering*, 36(2), 165–171. <https://doi.org/10.1109/10.16463>

Hauser, W. A. (1994). The Prevalence and Incidence of Convulsive Disorders in Children. *Epilepsia*, 35(s2), S1–S6. <https://doi.org/10.1111/j.1528-1157.1994.tb05932.x>

He, B. J., Zempel, J. M., Snyder, A. Z., & Raichle, M. E. (2010). The temporal structures and functional significance of scale-free brain activity. *Neuron*, 66(3), 353–369. <https://doi.org/10.1086/498510.Parasitic>

Hearn, D. D., & Baker, M. P. (1996). *Computer Graphics: C Version* (2nd ed.). Pearson Education.

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures*. (A. C. Tamhane, Ed.). New York: New York : Wiley.

Hughes, J. R. (2010). Benign epilepsy of childhood with centrotemporal spikes (BECTS): To treat or not to treat, that is the question. *Epilepsy and Behavior*, 19(3), 197–203. <https://doi.org/10.1016/j.yebeh.2010.07.018>

Jäckel, D., Bakkum, D. J., Russell, T. L., Müller, J., Radivojevic, M., Frey, U., ...

Hierlemann, A. (2017). Combination of High-density Microelectrode Array and Patch Clamp Recordings to Enable Studies of Multisynaptic Integration. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-00981-4>

- Jacobson, L. A., Ryan, M., Martin, R. B., Ewen, J., Mostofsky, S. H., Denckla, M. B., & Mark Mahone, E. (2011). Working memory influences processing speed and reading fluency in ADHD. *Child Neuropsychology*.
<https://doi.org/10.1080/09297049.2010.532204>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. (D. Witten, T. Hastie, & R. Tibshirani, Eds.) (1st ed. 20). New York, NY: New York, NY : Springer New York : Imprint: Springer.
- Jiruska, P., de Curtis, M., Jefferys, J. G. R., Schevon, C. A., Schiff, S. J., & Schindler, K. (2013). Synchronization and desynchronization in epilepsy: Controversies and hypotheses. *Journal of Physiology*. <https://doi.org/10.1113/jphysiol.2012.239590>
- Judson, M. C. C., Wallace, M. L. L., Sidorov, M. S. S., Burette, A. C. C., Gu, B., van Woerden, G. M. M., ... Philpot, B. D. D. (2016). GABAergic Neuron-Specific Loss of Ube3a Causes Angelman Syndrome-Like EEG Abnormalities and Enhances Seizure Susceptibility. *Neuron*, *90*(1), 56–69.
<https://doi.org/10.1016/j.neuron.2016.02.040>
- Kabbara, A., El Falou, W., Khalil, M., Wendling, F., & Hassan, M. (2017). The dynamic functional core network of the human brain at rest. *Scientific Reports*, *7*(1), 2936.
<https://doi.org/10.1038/s41598-017-03420-6>
- Karalis, N., Dejean, C., Chaudun, F., Khoder, S., R Rozeske, R., Wurtz, H., ... Herry, C. (2016). 4-Hz oscillations synchronize prefrontal-amygdala circuits during fear behavior. *Nature Neuroscience*, *19*(4), 605–612. <https://doi.org/10.1038/nn.4251>
- Kass, R. E., Amari, S. I., Arai, K., Brown, E. N., Diekman, C. O., Diesmann, M., ...

- Kramer, M. A. (2018). Computational Neuroscience: Mathematical and Statistical Perspectives. *Annual Review of Statistics and Its Application*, 5, 183–214.
<https://doi.org/10.1146/annurev-statistics-041715-033733>
- Kass, R. E., Eden, U. T., & Brown, E. N. (2014). *Analysis of Neural Data*. Springer Series in Statistics. https://doi.org/10.1007/978-1-4614-9602-1_1
- Katewa, V., & Parakh, M. (2015). A Review of the Not So Benign- Benign Childhood Epilepsy with Centrottemporal Spikes. *Journal of Neurology & Neurophysiology*, 06(04), 4–7. <https://doi.org/10.4172/2155-9562.1000314>
- Kelley, K., & Maxwell, S. E. (2003). Sample Size for Multiple Regression: Obtaining Regression Coefficients That Are Accurate, Not Simply Significant. *Psychological Methods*, 8(3), 305–321. <https://doi.org/10.1037/1082-989X.8.3.305>
- Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology*, 52, 3–6.
- Korff, C. M., Kelley, K. R., & Nordli, D. R. J. (2005). Notched delta, phenotype, and Angelman syndrome. *Journal of Clinical Neurophysiology : Official Publication of the American Electroencephalographic Society*, 22(4), 238–243.
<https://doi.org/10.1097/01.wnp.0000167930.90824.0f>
- Kramer, M. A., & Cash, S. S. (2012). Epilepsy as a Disorder of Cortical Network Organization. *The Neuroscientist*, 18(4), 360–372.
<https://doi.org/10.1177/1073858411422754>
- Kramer, M. A., & Eden, U. T. (2016). *Case Studies in Neural Data Analysis: A Guide for*

the Practicing Neuroscientist. MIT Press (1st ed.).

- Kramer, M. A., Eden, U. T., Cash, S. S., & Kolaczyk, E. D. (2009). Network inference with confidence from multivariate time series. *Physical Review E*, 79(6), 061916. <https://doi.org/10.1103/PhysRevE.79.061916>
- Kramer, M. A., Ostrowski, L. M., Song, D. Y., Thorn, E. L., Stoyell, S. M., Parnes, M., ... Chu, C. J. (2019). Scalp recorded spike ripples predict seizure risk in childhood epilepsy better than spikes. *Brain*, 142(5), 1296–1309. <https://doi.org/10.1093/brain/awz059>
- Kramer, M. A., Stoyell, S. M., Chinappen, D., Ostrowski, L. M., Spencer, E. R., Morgan, A. K., ... Chu, C. J. (2021). Focal sleep spindle deficits reveal focal thalamocortical dysfunction and predict cognitive deficits in childhood epilepsy with centrotemporal spikes. *The Journal of Neuroscience*, 00(00), JN-RM-2009-20. <https://doi.org/10.1523/jneurosci.2009-20.2020>
- Kramer, M. A., Tort, A. B. L., & Kopell, N. J. (2008). Sharp edge artifacts and spurious coupling in EEG frequency comodulation measures. *Journal of Neuroscience Methods*, 170(2), 352–357. <https://doi.org/10.1016/j.jneumeth.2008.01.020>
- Kyllerman, M. (1995). On the prevalence of Angelman syndrome. *American Journal of Medical Genetics*. <https://doi.org/10.1002/ajmg.1320590331>
- Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4), 194–208. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)8:4<194::AID-HBM4>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C)
- Laureys, S., Gosseries, O., & Tononi, G. (2015). *The Neurology of Consciousness* :

Cognitive Neuroscience and Neuropathology. Elsevier Science & Technology.

Retrieved from

<http://ebookcentral.proquest.com/lib/bu/detail.action?docID=2166960>

- Lemke, J. R., Lal, D., Reinthaler, E. M., Steiner, I., Nothnagel, M., Alber, M., ... Von Spiczak, S. (2013). Mutations in GRIN2A cause idiopathic focal epilepsy with rolandic spikes. *Nature Genetics*, *45*(9), 1067–1072. <https://doi.org/10.1038/ng.2728>
- Li, Y., Lopez-Huerta, V. G., Adiconis, X., Levandowski, K., Choi, S., Simmons, S. K., ... Feng, G. (2020). Distinct subnetworks of the thalamic reticular nucleus. *Nature*, *583*(7818), 819–824. <https://doi.org/10.1038/s41586-020-2504-5>
- Luo, C., Guo, Z. wei, Lai, Y. xiu, Liao, W., Liu, Q., Kendrick, K. M., ... Li, H. (2012). Musical training induces functional plasticity in perceptual and motor networks: Insights from resting-state fMRI. *Public Library of Science ONE*, *7*(5). <https://doi.org/10.1371/journal.pone.0036568>
- Lynall, M.-E., Bassett, D. S., Kerwin, R., McKenna, P. J., Kitzbichler, M., Muller, U., & Bullmore, E. (2010). Functional Connectivity and Brain Networks in Schizophrenia. *Journal of Neuroscience*, *30*(28), 9477–9487. <https://doi.org/10.1523/JNEUROSCI.0333-10.2010>
- Maccotta, L., He, B. J., Snyder, A. Z., Eisenman, L. N., Benzinger, T. L., Ances, B. M., ... Hogan, R. E. (2013). Impaired and facilitated functional networks in temporal lobe epilepsy. *NeuroImage: Clinical*, *2*(1), 862–872. <https://doi.org/10.1016/j.nicl.2013.06.011>
- Martinez, L. A., Born, H. A., Harris, S., Regnier-Golanov, A., Grieco, J. C., Weeber, E.

- J., & Anderson, A. E. (2020). Quantitative EEG Analysis in Angelman Syndrome: Candidate Method for Assessing Therapeutics. *Clinical EEG and Neuroscience*, 1–10. <https://doi.org/10.1177/1550059420973095>
- Matlis, S., Boric, K., Chu, C. J., & Kramer, M. A. (2015). Robust disruptions in electroencephalogram cortical oscillations and large-scale functional networks in autism. *BMC Neurology*, 15(1), 1–17. <https://doi.org/10.1186/s12883-015-0355-8>
- McCullagh, P., & Nelder, J. (1989). *Generalized Linear Models*. (J. A. Nelder, Ed.) (2nd ed.). London: London.
- Mednick, S. C., McDevitt, E. A., Walsh, J. K., Wamsley, E., Paulus, M., Kanady, J. C., & Drummond, S. P. A. (2013). The critical role of sleep spindles in hippocampal-dependent memory: A pharmacology study. *Journal of Neuroscience*, 33(10), 4494–4504. <https://doi.org/10.1523/JNEUROSCI.3127-12.2013>
- Mellish, L. C., Dunkley, C., Ferrie, C. D., & Pal, D. K. (2015). Antiepileptic drug treatment of rolandic epilepsy and Panayiotopoulos syndrome: Clinical practice survey and clinical trial feasibility. *Archives of Disease in Childhood*, 100(1), 62–67. <https://doi.org/10.1136/archdischild-2013-304211>
- Meng, L., Ward, A. J., Chun, S., Bennett, C. F., Beaudet, A. L., & Rigo, F. (2015). Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature*, 518(7539), 409–412. <https://doi.org/10.1038/nature13975>
- Merker, B., Podell, K., & Wingate, M. (2018). Grooved Pegboard Test. In *Encyclopedia of Clinical Neuropsychology*. https://doi.org/10.1007/978-3-319-57111-9_187
- Mertz, L. G. B., Christensen, R., Vogel, I., Hertz, J. M., Nielsen, K. B., Grønskov, K., &

- Ostergaard, J. R. (2013). Angelman syndrome in Denmark. Birth incidence, genetic findings, and age at diagnosis. *American Journal of Medical Genetics, Part A*, *161*(9), 2197–2203. <https://doi.org/10.1002/ajmg.a.36058>
- Mirandola, L., Cantalupo, G., Vaudano, A. E., Avanzini, P., Ruggieri, A., Pisani, F., ... Meletti, S. (2013). Centrotemporal spikes during NREM sleep: The promoting action of thalamus revealed by simultaneous EEG and fMRI coregistration. *Epilepsy and Behavior Case Reports*, *1*(1), 106–109. <https://doi.org/10.1016/j.ebcr.2013.06.005>
- Mitra, A., Kraft, A., Wright, P., Acland, B., Snyder, A. Z., Rosenthal, Z., ... Raichle, M. E. (2018). Spontaneous Infra-slow Brain Activity Has Unique Spatiotemporal Dynamics and Laminar Structure. *Neuron*, *98*(2), 297-305.e6. <https://doi.org/10.1016/j.neuron.2018.03.015>
- Ngo, H.-V. V., Miedema, A., Faude, I., Martinetz, T., Molle, M., & Born, J. (2015). Driving Sleep Slow Oscillations by Auditory Closed-Loop Stimulation--A Self-Limiting Process. *Journal of Neuroscience*, *35*(17), 6630–6638. <https://doi.org/10.1523/JNEUROSCI.3133-14.2015>
- Nobili, L., Baglietto, M. G., Beelke, M., De Carli, F., De Negri, E., Gaggero, R., ... Ferrillo, F. (2001). Distribution of epileptiform discharges during nREM sleep in the CSWSS syndrome: relationship with sigma and delta activities. *Epilepsy Research*, *44*(2–3), 119–128. [https://doi.org/10.1016/S0920-1211\(01\)00191-7](https://doi.org/10.1016/S0920-1211(01)00191-7)
- Nobili, L., Ferrillo, F., Baglietto, M. ., Beelke, M., De Carli, F., De Negri, E., ... De Negri, M. (1999). Relationship of sleep interictal epileptiform discharges to sigma

- activity (12–16 Hz) in benign epilepsy of childhood with rolandic spikes. *Clinical Neurophysiology*, *110*(1), 39–46. [https://doi.org/10.1016/S0168-5597\(98\)00041-0](https://doi.org/10.1016/S0168-5597(98)00041-0)
- Nunez, P. L., & Srinivasan, R. (2006). *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford: Oxford University Press.
- <https://doi.org/10.1093/acprof:oso/9780195050387.001.0001>
- Ostrowski, L. M., Spencer, E. R., Bird, L. M., Thibert, R., Komorowski, R. W., Kramer, M. A., & Chu, C. J. (2021). Delta power robustly predicts cognitive function in Angelman syndrome. *Annals of Clinical and Translational Neurology*, *acn3.51385*. <https://doi.org/10.1002/acn3.51385>
- Ouyang, G., Hildebrandt, A., Schmitz, F., & Herrmann, C. S. (2020). Decomposing alpha and 1/f brain activities reveals their differential associations with cognitive processing speed. *NeuroImage*, *205*(September 2019), 116304. <https://doi.org/10.1016/j.neuroimage.2019.116304>
- Panayiotopoulos, C. P., Michael, M., Sanders, S., Valeta, T., & Koutroumanidis, M. (2008). Benign childhood focal epilepsies: Assessment of established and newly recognized syndromes. *Brain*, *131*(9), 2264–2286. <https://doi.org/10.1093/brain/awn162>
- Park, H.-J., & Friston, K. (2013). Structural and Functional Brain Networks: From Connections to Cognition. *Science*, *342*(6158), 1238411–1238411. <https://doi.org/10.1126/science.1238411>
- Peters, S. U., Goddard-Finegold, J., Beaudet, A. L., Madduri, N., Turcich, M., & Bacino, C. A. (2004). Cognitive and adaptive behavior profiles of children with Angelman

syndrome. *American Journal of Medical Genetics. Part A*, 128A(2), 110–113.

<https://doi.org/10.1002/ajmg.a.30065>

Petersen, M. B., Brøndum-Nielsen, K., Hansen, L. K., & Wulff, K. (1995a). Clinical, cytogenetic, and molecular diagnosis of Angelman syndrome: estimated prevalence rate in a Danish county. *American Journal of Medical Genetics*.

<https://doi.org/10.1002/ajmg.1320600317>

Petersen, M. B., Brøndum-Nielsen, K., Hansen, L. K., & Wulff, K. (1995b). Clinical, cytogenetic, and molecular diagnosis of Angelman syndrome: estimated prevalence rate in a Danish county. *American Journal of Medical Genetics*, 60(3), 261–262.

<https://doi.org/10.1002/ajmg.1320600317>

Petersen, S. E., & Sporns, O. (2015). Brain Networks and Cognitive Architectures.

Neuron, 88(1), 207–219. <https://doi.org/10.1016/j.neuron.2015.09.027>

Pfurtscheller, G., Stancák, A., & Neuper, C. (1996). Post-movement beta synchronization. A correlate of an idling motor area? *Electroencephalography and Clinical Neurophysiology*, 98(4), 281–293. [https://doi.org/10.1016/0013-4694\(95\)00258-8](https://doi.org/10.1016/0013-4694(95)00258-8)

Place, R., Farovik, A., Brockmann, M., & Eichenbaum, H. (2016). Bidirectional prefrontal-hippocampal interactions support context-guided memory. *Nature Neuroscience*, 19(8), 992–994. <https://doi.org/10.1038/nn.4327>

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

Purcell, S. M., Manoach, D. S., Demanuele, C., Cade, B. E., Mariani, S., Cox, R., ...

Stickgold, R. (2017). Characterizing sleep spindles in 11,630 individuals from the

- National Sleep Research Resource. *Nature Communications*, 8(May), 1–16.
<https://doi.org/10.1038/ncomms15930>
- Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., LaMantia, A.-S., & White, L. (2004). *Neuroscience* (Third Edit). Sinauer Associates, Inc.
- Ringo, J. L., Don, R. W., Demeter, S., & Simard, P. Y. (1994). Time Is of the Essence: A Conjecture that Hemispheric Specialization Arises from Interhemispheric Conduction Delay. *Cerebral Cortex*, (September).
- Ross, E. E., Stoyell, S. M., Kramer, M. A., Berg, A. T., & Chu, C. J. (2020). The natural history of seizures and neuropsychiatric symptoms in childhood epilepsy with centrottemporal spikes (CECTS). *Epilepsy and Behavior*.
<https://doi.org/10.1016/j.yebeh.2019.07.038>
- Scheffer, I. E., Berkovic, S., Capovilla, G., Connolly, M. B., French, J., Guilhoto, L., ... Zuberi, S. M. (2017). ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. *Epilepsia*.
<https://doi.org/10.1111/epi.13709>
- Scheltens-De Boer, M. (2009). Guidelines for EEG in encephalopathy related to ESES/CSWS in children. *Epilepsia*, 50(SUPPL. 7), 13–17.
<https://doi.org/10.1111/j.1528-1167.2009.02211.x>
- Scheuer, M. L., Bagic, A., & Wilson, S. B. (2017). Spike detection: Inter-reader agreement and a statistical Turing test on a large data set. *Clinical Neurophysiology*.
<https://doi.org/10.1016/j.clinph.2016.11.005>
- Schindler, K., Elger, C. E., & Lehnertz, K. (2007). Increasing synchronization may

- promote seizure termination: Evidence from status epilepticus. *Clinical Neurophysiology*, *118*(9), 1955–1968. <https://doi.org/10.1016/j.clinph.2007.06.006>
- Schmitt, L. I., Wimmer, R. D., Nakajima, M., Happ, M., Mofakham, S., & Halassa, M. M. (2017). Thalamic amplification of cortical connectivity sustains attentional control. *Nature*, *545*(7653), 219–223. <https://doi.org/10.1038/nature22073>
- Seth, A. K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods*, *186*(2), 262–273. <https://doi.org/10.1016/j.jneumeth.2009.11.020>
- 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>
- Seth, A. K., Chorley, P., & Barnett, L. C. (2013). Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. *NeuroImage*, *65*, 540–555. <https://doi.org/10.1016/j.neuroimage.2012.09.049>
- Shamshiri, E. A., Sheybani, L., & Vulliemoz, S. (2019). The Role of EEG-fMRI in Studying Cognitive Network Alterations in Epilepsy. *Frontiers in Neurology*, *10*, 1033. <https://doi.org/10.3389/fneur.2019.01033>
- Sidorov, M. S., Deck, G. M., Dolatshahi, M., Thibert, R. L., Bird, L. M., Chu, C. J., & Philpot, B. D. (2017). Delta rhythmicity is a reliable EEG biomarker in Angelman syndrome: A parallel mouse and human analysis. *Journal of Neurodevelopmental Disorders*, *9*(1), 1–14. <https://doi.org/10.1186/s11689-017-9195-8>
- Silva-Santos, S., Van Woerden, G. M., Bruinsma, C. F., Mientjes, E., Jolfaei, M. A.,

- Distel, B., ... Elgersma, Y. (2015). Ube3a reinstatement identifies distinct developmental windows in a murine Angelman syndrome model. *Journal of Clinical Investigation*, 125(5), 2069–2076. <https://doi.org/10.1172/JCI80554>
- Singer, W. (1993). Synchronization of Cortical Activity and Its Putative Role in and Learning. *Annual Review of Physiology*, 55, 349–374. <https://doi.org/10.1146/annurev.ph.55.030193.002025>
- Smith, E. H., Banks, G. P., Mikell, C. B., Cash, S. S., Patel, S. R., Eskandar, E. N., & Sheth, S. A. (2015). Frequency-Dependent Representation of Reinforcement-Related Information in the Human Medial and Lateral Prefrontal Cortex. *Journal of Neuroscience*, 35(48), 15827–15836. <https://doi.org/10.1523/JNEUROSCI.1864-15.2015>
- Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., ... Woolrich, M. W. (2011). Network modelling methods for FMRI. *NeuroImage*, 54(2), 875–891. <https://doi.org/10.1016/j.neuroimage.2010.08.063>
- Song, D. Y., Stoyell, S. M., Ross, E. E., Ostrowski, L. M., Thorn, E. L., Stufflebeam, S. M., ... Chu, C. J. (2019). Beta oscillations in the sensorimotor cortex correlate with disease and remission in benign epilepsy with centrotemporal spikes. *Brain and Behavior*, 9(3), 1–12. <https://doi.org/10.1002/brb3.1237>
- Sonzogni, M., Hakonen, J., Bernabé Kleijn, M., Silva-Santos, S., Judson, M. C., Philpot, B. D., ... Elgersma, Y. (2019). Delayed loss of UBE3A reduces the expression of Angelman syndrome-associated phenotypes. *Molecular Autism*, 10(1), 1–9. <https://doi.org/10.1186/s13229-019-0277-1>

- Strimbu, K., & Tavel, J. A. (2010). What are biomarkers? *Current Opinion in HIV and AIDS*, 5(6), 463–466. <https://doi.org/10.1097/COH.0b013e32833ed177>
- Swayze, E. E., Siwkowski, A. M., Wancewicz, E. V, Migawa, M. T., Wyrzykiewicz, T. K., Hung, G., ... Bennett, C. F. (2007). Antisense oligonucleotides containing locked nucleic acid improve potency but cause significant hepatotoxicity in animals. *Nucleic Acids Research*, 35(2), 687–700. <https://doi.org/10.1093/nar/gkl1071>
- Symms, M. R., Allen, P. J., Woermann, F. G., Polizzi, G., Krakow, K., Barker, G. J., ... Duncan, J. S. (1999). Reproducible localization of interictal epileptiform discharges using EEG-triggered fMRI. *Physics in Medicine and Biology*, 44(7), N161-8. <https://doi.org/10.1088/0031-9155/44/7/403>
- Tang, W., Liu, H., Douw, L., Kramer, M. A., Eden, U. T., Hämäläinen, M. S., & Stufflebeam, S. M. (2017). Dynamic connectivity modulates local activity in the core regions of the default-mode network. *Proceedings of the National Academy of Sciences*, 114(36), 201702027. <https://doi.org/10.1073/pnas.1702027114>
- Telesford, Q. K., Lynall, M.-E., Vettel, J., Miller, M. B., Grafton, S. T., & Bassett, D. S. (2016). Detection of functional brain network reconfiguration during task-driven cognitive states. *NeuroImage*, 142, 198–210. <https://doi.org/10.1016/j.neuroimage.2016.05.078>
- Thibert, R. L., Larson, A. M., Hsieh, D. T., Raby, A. R., & Thiele, E. A. (2013). Neurologic manifestations of Angelman syndrome. *Pediatric Neurology*, 48(4), 271–279. <https://doi.org/10.1016/j.pediatrneurol.2012.09.015>

- Thorn, E. L., Ostrowski, L. M., Chinappen, D. M., Jing, J., Westover, M. B., Stufflebeam, S. M., ... Chu, C. J. (2020). Persistent abnormalities in Rolandic thalamocortical white matter circuits in childhood epilepsy with centrottemporal spikes. *Epilepsia*, *61*(11), 2500–2508. <https://doi.org/10.1111/epi.16681>
- Tovia, E., Goldberg-Stern, H., Ben Zeev, B., Heyman, E., Watemberg, N., Fattal-Valevski, A., & Kramer, U. (2011). The prevalence of atypical presentations and comorbidities of benign childhood epilepsy with centrottemporal spikes. *Epilepsia*, *52*(8), 1483–1488. <https://doi.org/10.1111/j.1528-1167.2011.03136.x>
- Tucker, M. A., & Fishbein, W. (2009). The impact of sleep duration and subject intelligence on declarative and motor memory performance: How much is enough? *Journal of Sleep Research*, *18*(3), 304–312. <https://doi.org/10.1111/j.1365-2869.2009.00740.x>
- Uddin, L. Q., Supekar, K. S., Ryali, S., & Menon, V. (2011). Dynamic Reconfiguration of Structural and Functional Connectivity Across Core Neurocognitive Brain Networks with Development. *Journal of Neuroscience*, *31*(50), 18578–18589. <https://doi.org/10.1523/JNEUROSCI.4465-11.2011>
- Uhlhaas, P. J., Roux, F., Rodriguez, E., Rotarska-Jagiela, A., & Singer, W. (2010). Neural synchrony and the development of cortical networks. *Trends in Cognitive Sciences*, *14*(2), 72–80. <https://doi.org/10.1016/j.tics.2009.12.002>
- Valdés-Sosa, P. A., Sánchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., & Canales-Rodríguez, E. (2005). Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, 360(1457), 969–981.

<https://doi.org/10.1098/rstb.2005.1654>

Valente, K. D., Andrade, J. Q., Grossmann, R. M., Kok, F., Fridman, C., Koiffmann, C. P., & Marques-Dias, M. J. (2003). Angelman Syndrome: Difficulties in EEG Pattern Recognition and Possible Misinterpretations. *Epilepsia*, 44(8), 1051–1063.

<https://doi.org/10.1046/j.1528-1157.2003.66502.x>

van Ginkel, J. R. (2019). Significance Tests and Estimates for R² for Multiple Regression in Multiply Imputed Datasets: A Cautionary Note on Earlier Findings, and Alternative Solutions. *Multivariate Behavioral Research*, 54(4), 514–529.

<https://doi.org/10.1080/00273171.2018.1540967>

Vannest, J., Tenney, J. R., Gelineau-Morel, R., Maloney, T., & Glauser, T. A. (2015). Cognitive and behavioral outcomes in benign childhood epilepsy with centrotemporal spikes. *Epilepsy and Behavior*.

<https://doi.org/10.1016/j.yebeh.2015.01.041>

Viventi, J., Kim, D. H., Vigeland, L., Frechette, E. S., Blanco, J. A., Kim, Y. S., ... Litt, B. (2011). Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity in vivo. *Nature Neuroscience*, 14(12), 1599–1605.

<https://doi.org/10.1038/nn.2973>

Wagner, R., Torgesen, J., & Rashotte, C. (1991). *Comprehensive Test of Phonological Processing: Examiner's Manual*. *Journal of Experimental Child Psychology*. Austin, TX: Sage.

Wamsley, E. J., Tucker, M. A., Shinn, A. K., Ono, K. E., McKinley, S. K., Ely, A. V., ...

- Manoach, D. S. (2012). Reduced Sleep Spindles and Spindle Coherence in Schizophrenia: Mechanisms of Impaired Memory Consolidation? *Biological Psychiatry*, *71*(2), 154–161. <https://doi.org/10.1016/j.biopsych.2011.08.008>
- Wasserman, L. (2010). *All of statistics : a concise course in statistical inference. Concise course in statistical inference (Corrected)*. New York: New York : Springer.
- Wechsler, D. (2014). *Wechsler intelligence scale for children (5th ed.)*. San Antonio, TX: Psychological Corporation.
- Wedderburn, R. W. M. (1974). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*, *61*(3), 439–447.
- Wickens, S., Bowden, S. C., & D'Souza, W. (2017). Cognitive functioning in children with self-limited epilepsy with centrotemporal spikes: A systematic review and meta-analysis. *Epilepsia*. <https://doi.org/10.1111/epi.13865>
- Williams, C. A., Beaudet, A. L., Clayton-Smith, J., Knoll, J. H., Kyllerman, M., Laan, L. A., ... Wagstaff, J. (2006). Angelman syndrome 2005: Updated consensus for diagnostic criteria. *American Journal of Medical Genetics*. <https://doi.org/10.1002/ajmg.a.31074>
- Xiao, F., Lei, D., An, D., Li, L., Chen, S., Chen, F., ... Zhou, D. (2015). Functional brain connectome and sensorimotor networks in rolandic epilepsy. *Epilepsy Research*, *113*, 113–125. <https://doi.org/10.1016/j.eplepsyres.2015.03.015>
- Xie, W., Ross, E. E., Kramer, M. A., Eden, U. T., & Chu, C. J. (2018). Timing matters: Impact of anticonvulsant drug treatment and spikes on seizure risk in benign epilepsy with centrotemporal spikes. *Epilepsia Open*, *3*(3), 409–417.

<https://doi.org/10.1002/epi4.12248>

Zagha, E., Ge, X., & McCormick, D. A. (2015). Competing Neural Ensembles in Motor Cortex Gate Goal-Directed Motor Output. *Neuron*, 88(3), 565–577.

<https://doi.org/10.1016/j.neuron.2015.09.044>

Zaveri, H. P., Pincus, S. M., Goncharova, I. I., Duckrow, R. B., Spencer, D. D., & Spencer, S. S. (2009). Localization-related epilepsy exhibits significant connectivity away from the seizure-onset area. *NeuroReport*, 20(9), 891–895.

<https://doi.org/10.1097/WNR.0b013e32832c78e0>

CURRICULUM VITAE

ELIZABETH SPENCER

Phone: (443) 798-1830

liz.r.spencer1@gmail.com

GitHub: <https://github.com/erss>

Website: <https://erss.github.io/>

EDUCATION

Boston University 2015 - Present
 Graduate Program for Neuroscience (computational specialty)
 Advisors: Mark Kramer, Ph.D., Uri Eden, Ph.D., Catherine Chu, MD.

University of Maryland, College Park (Honors Program) 2011 - 2015
BS: Mathematics (GPA 3.74) & **BA:** French Language and Literature (GPA 4.0)
 Minor in Physics

RESEARCH GRANTS

National Science Foundation Graduate Research Fellow (NSF GRFP) Apr 2017
 Field of Study: Mathematics - Mathematical Biology
 Proposal to analyze how dynamic functional networks evolve along their anatomical white matter pathways, "A Dynamic Approach to Resolving the Brain."

RESEARCH EXPERIENCE

Dissertation, Boston University (BU), Boston, MA 2015 - Present
 Advisors: Mark Kramer, Ph.D. & Uri Eden, Ph.D. - Neural Data Analysis Lab (BU)
 Catherine Chu, M.D. - Chu Lab at Massachusetts General Hospital (MGH)
 "Biomarker discovery and statistical modeling with applications in childhood epilepsy and Angelman syndrome."

Industry Research Internship Summer 2020
 Biogen Inc & Massachusetts General Hospital, Boston MA

- Developed approach to use a clinical biomarker, delta power, to measure treatment response in Angelman syndrome.

Research in Industrial Projects for Students Summer 2014
 Institute for Pure and Applied Mathematics, University of California, Los Angeles, CA
 The Aerospace Corporation, El Segundo, CA

- Studied methods for computing the channel capacity of satellite communication systems.

Undergraduate research assistant

Fall 2013

Center for Harmonic Analysis, University of Maryland (UMD), College Park, MD

Advisor: Wojciech Czaja, Ph.D.

- Studied wavelet transforms with applications in medical imaging.

Summer Undergraduate Applied Mathematics Institute

Summer 2013

Center for Nonlinear Analysis, Carnegie Mellon University, Pittsburgh, PA

- Studied statistical regression techniques, including general additive models, linear models, and smoothing splines, to build models to predict galaxy redshifts.

PEER-REVIEWED PUBLICATION

SPENCER, E. R., SHI, W., OSTROWSKI, L., KOMOROWSKI, R., GILBERT, J. P., OSTROWSKI, L. M., BIRD L.M., THIBERT, R., BAO, C. MOLLOY, F., CALHOUN, M., KOIRALA, S., JAFAR-NEJAD, P., RIGO, F., KRAMER, M. A., & CHU, C. J. (2021). *Deviations from the natural history of delta power in Angelman syndrome reflect treatment effect size and correlate with UBE3A expression*, eLife, under review.

SPENCER, E. R., CHINAPPEN, D., EMERTON, B., MORGAN, A. K., EMERTON, B., MANOACH, D. S., EDEN, U. T., KRAMER, M. A., & CHU C. (2021). *Source estimates reveal regional spindle deficits and predict cognitive symptoms in Rolandic epilepsy*, under review.

OSTROWSKI, L., **SPENCER, E. R.**, BIRD, L., THIBERT, R., KOMOROWSKI, R., KRAMER, M., CHU, C. (2021). *Delta power robustly predicts cognitive function in Angelman syndrome*. Ann Clin Transl Neurol:acn3.51385.

KRAMER, M. A., STOYELL, S., CHINAPPEN, D., OSTROWSKI, L., **SPENCER, E. R.**, MORGAN, A., EMERTON, B., JING, J., WESTOVER, M., EDEN, U., STICKGOLD, R., MANOACH, D., & CHU C. J. (2021). *Focal sleep spindle deficits reveal focal thalamocortical dysfunction and predict cognitive deficits in childhood epilepsy with centrottemporal spikes*. The Journal of Neuroscience, JN-RM-2009-20.

MARTINET, L.-E., KRAMER, M. A., VILES, W., PERKINS, L. N., **SPENCER, E.**, CHU, C. J., CASH, S. S., & KOLACZYK, E. D. (2020). *Robust dynamic community detection with applications to human brain functional networks*. Nature Communications, 11(1), 2785.

SPENCER, E., MARTINET, L. E., ESKANDAR, E. N., CHU, C. J., KOLACZYK, E. D., CASH, S. S., EDEN, U. T., & KRAMER, M. A. (2018). *A procedure to increase the power of Granger-*

causal analysis through temporal smoothing. Journal of Neuroscience Methods, 308(July), 48–61.

PRESENTATIONS AND POSTERS

Presentations

Brain Networks in Epilepsy

Math Department Community Seminar, BU, Boston, MA

Apr 2019

Procedure to increase the power of Granger-causal analysis through temporal smoothing

Cognitive Rhythms Collaborative, MIT, Cambridge, MA

Apr 2018

NSF Graduate Research Fellowship Session

American Mathematical Society Professional Development Series, Boston, MA

Sept 2017

Posters

Characterizing the relationship between sleep spindles and neurocognitive deficits in epileptic encephalopathy

Society for Neuroscience Annual Meeting, Chicago, IL

Oct 2019

Characterizing the relationship between functional connectivity and neurocognitive deficits in benign epilepsy with centrottemporal spikes

Statistical Analysis of Neuronal Data, Pittsburgh, PA

May 2019

Society for Neuroscience Annual Meeting, San Diego, CA

Nov 2018

Network inference for dynamic modeling of epileptic seizures

Society for Neuroscience Annual Meeting, Washington, D.C.

Nov 2017

Calculating Channel Capacity of Satellite Communications

Joint Mathematics Meetings, San Antonio, TX

Nov 2017

Developing Regression Models to Predict Galaxy Redshifts

Joint Mathematics Meetings, Baltimore, MD

Jan 2015

Applied Harmonic Analysis for Retinal Imaging – Dimensionality Reduction and Classification

Biosciences Research and Technology Review, UMD, College Park, MD

Nov 2013

PROFESSIONAL TRAINING

Summer Workshop on the Dynamic Brain

Aug - Sept 2019

Allen Institute for Brain Science & University of Washington, Friday Harbor, WA

Two-week course featuring lectures on computational modeling applied to visual system neuroscience, Python bootcamp and group project using the Allen Institute's open datasets (project: *Characterization of Oscillations in Visual Cortex*).

The MIT IMPACT program

Spring 2019

Massachusetts Institute of Technology, Cambridge, MA

Semester long mentorship program to refine research focus to optimize scientific impact, develop communication skills and engage in professional development with faculty from institutions and companies in the Boston area.

Summer School in Computational Sensory-Motor Neuroscience

Aug - Sept 2016

University of Minnesota, Minneapolis, MN

Two-week course featuring lectures on modeling applied to sensory and movement science and a group project (project: *Modeling the Relationship Between Reward and Sensory Feedback in Sensorimotor Adaptation*).

Course in Mining and Modeling of Neuroscience Data

July - Aug 2016

Redwood Center for Theoretical Neuroscience, University of California, Berkeley, CA

Two-week course featuring computational and statistical modeling techniques for analyzing big data.

HONORS AND AWARDS

First Place Computational Neuroscience Poster

Feb 2018, Mar 2020

Poster session for all students in the BU Graduate Program for Neuroscience (faculty judged).

Third Place Henry I. Russek Student Achievement Award

Apr 2019

Departmental award for excellence in the BU Graduate Medical Sciences.

PROFESSIONAL ACTIVITIES AND SERVICE

NeuroArts Forum: Co-Organizer

Oct 2019

Lecture series bringing neuroscientists and artists from different communities to facilitate crosstalk between the arts and sciences.

Directed Reading Program: Mentor

Fall 2018 & Spring 2019

Semester-long mentoring of a BU undergraduate student in a private reading course, and completion of a related research project and presentation. Topics:

- Dr. Kolaczyk's *Statistical Analysis of Network Data: Methods and Models*.
- Drs. Kolaczyk and Csardi's *Statistical Analysis of Network Data with R*.

Women in Science and Engineering Undergraduate Mentorship Program Fall 2018

Semester long mentorship program of BU undergraduate interested in pursuing graduate school.

Neuroscience Graduate Student Organization Outreach Committee 2016 – 2019

- Prepared and organized activities for BU booth at Boston’s Museum of Science annual Brain Awareness Week (Winter 2016 – 19).
- Volunteer for SET (Science, Engineering, Technology) in the city – Day of Career Exploration for High School Girls (April 2018).
- Helped design four-part after school neuroscience club series at William H. Lincoln Elementary School (Fall 2017).

West End House Elementary Girls Science Club, volunteer Fall 2017 – Spring 2018
Computational Neuroscience Student Organization, Treasurer 2016

SKILLS

Computer Proficiencies: MATLAB (expert); R (advanced); and Python (intermediate).
Languages: English and French.