

Imperial College London
Department of Electrical and Electronic Engineering

Decoding Causal Mechanisms in Brain-Muscle Activation Through AI-driven Signal Processing

Farwa Abbas

Supervised by: Dr. Wei Dai

Submitted in part fulfillment of the requirements for the degree of Doctor of Philosophy in
Electrical and Electronic Engineering of the Imperial College London and the Diploma of
Imperial College London, 2025

Declaration of Originality

I hereby declare that this thesis and the work reported herein was composed by and originated entirely from me. Information derived from the published and unpublished work of others has been acknowledged in the text and references are given in the list of sources.

Farwa Abbas (2025)

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.



Abstract

Movement disorders, particularly dystonia, present persistent challenges in diagnosis and treatment due to complex, non-linear interactions between central neural circuits and peripheral motor systems. These bidirectional interactions can become pathologically altered following neurophysiological disturbances. Traditional connectivity analyses often fail to capture the high-dimensional dynamics of such systems, motivating the development of advanced computational frameworks to uncover underlying causal mechanisms.

This thesis introduces several advancements that improve our understanding of causal mechanisms in neurophysiology. It presents a novel framework for modeling physiological signals using wavelet sparsity and Gaussian assumptions to distinguish physiological noise from measurement artifacts. The method also enforces mathematical stability constraints while allowing noise separation to identify feature signatures, enhancing causal analysis robustness.

The methodology begins with interpretable linear models before advancing to random forests and neural networks to better represent the neural-muscular interactions in movement disorders. A novel dynamic Bayesian network framework enables causal inference in temporal modeling, distinguishing direct and indirect causal effects via a two-layered analysis. This addresses key limitations of Granger causality, offering a clearer interpretation of relationships. The framework incorporates counterfactual analysis grounded in Bayesian principles to assess how observed changes impact the neural-muscular system. Unlike conventional methods, it embeds interventions into a dynamic causal context, inferring potential system responses.

These advancements contribute to the study of neuroplasticity, motor control, and various neurological conditions. By integrating mathematically rigorous causal inference with clinically interpretable models, this research enhances the role of computational neuroscience in clinical decision-making. The methods support data-driven diagnostics and intervention planning, paving the way for precision neurology.

At its core, this work presents a framework for dissecting neural-muscular communication, revealing causal pathways and breakdowns in sensorimotor signaling. The models uncover hidden dynamical structures, identify biomarkers, and inform interventions, bridging theory and practice in computational neurophysiology.

Acknowledgements

I would like to express my deepest gratitude to all those who have supported and guided me throughout the journey of completing this thesis.

First and foremost, I extend my sincere thanks to my supervisor, Dr. Wei Dai, for his unwavering guidance, support, and mentorship throughout my research. His expertise, patience, and encouragement have been invaluable to me. I am deeply grateful for his insightful feedback and for challenging me to think critically and approach problems from new perspectives. This work would not have been possible without his constant motivation and belief in my abilities.

I am also immensely thankful to my co-supervisors, Dr. Zoran and Dr. Verity, for their collaboration, stimulating discussions, and consistent support. It has been a privilege to work alongside such talented and dedicated individuals.

My sincere gratitude goes to all the participants in my study for their time, patience, and willingness to contribute. Their involvement was crucial to the success of this research.

I am deeply grateful to the Department of Electrical and Electronic Engineering at Imperial College London, as well as my sponsors from the Government of Pakistan, for providing the support necessary to make this research possible. Their assistance has been fundamental in enabling me to carry out my work and pursue my academic goals.

I would also like to express my heartfelt thanks to my parents and my sisters, Wafa and Alina, for their unwavering love, support, and encouragement throughout this journey. They have always believed in my dream of pursuing a PhD at Imperial, and despite the obstacles, their constant inspiration never wavered. Their faith in me has been my strength, especially during the toughest times.

I am deeply grateful to my friends, particularly Azka, whose support and quiet encouragement carried me through my moments of doubt, and for that, I am forever grateful.

Finally, I would like to thank my husband, Hussain, who has always encouraged me to live up to my true potential. His support was essential in making this work possible.

Thank you all for being a part of this journey.

To my family and teachers.

List of Abbreviations

VAR	Vector Autoregressive
EEG	Electroencephalography
sEMG	Surface Electromyography
PET	Positron Emission Tomography
fMRI	Functional Magnetic Resonance Imaging
MEG	Magnetoencephalography
DBS	Deep Brain Stimulation
GC	Granger Causality
CMC	Corticomuscular Coherence
ERP	Event-Related Potential
ICA	Independent Component Analysis
RMS	Root Mean Square
MAV	Mean Absolute Value
FT	Fourier Transform
PSD	Power Spectral Density
WT	Wavelet Transform
STFT	Short-Time Fourier Transform
EMD	Empirical Mode Decomposition
DTF	Directed Transfer Function
MI	Mutual Information
TE	Transfer Entropy
PDC	Partial Directed Coherence
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
HAMPD	Healthy Adults Motor Perturbation Dataset
FDI	First Dorsal Interosseous
FPB	Flexor Pollicis Brevis
PMPD	Pediatric Motor Perturbation Dataset
BFMDRS-m	Burke-Fahn-Marsden Dystonia Rating Scale (modified)
ADMM	Alternating Direction Method of Multipliers
CCM	Convergent Cross Mapping
DBN	Dynamic Bayesian Network
MVC	Maximum Voluntary Contraction
EEGD	Electroencephalography channel for dominant motor cortex
EEGND	Electroencephalography channel for non-dominant motor cortex
ARMA	Autoregressive Moving Average
SS-ADMM	Stationary and Sparse Alternating Direction Method of Multipliers
LOG	Latent Overlapping Group

OMP	Orthogonal Matching Pursuit
BP	Basis Pursuit
PDOM	Proximal Dogleg Opportunistic Majorization
ARFIMA	Autoregressive Fractionally Integrated Moving Average
HS-ADMM	Hierarchical Sparsity Alternating Direction Method of Multipliers
NMSE	Normalized Mean Squared Error
SSD-ADMM	Stationary and Sparse Denoising Alternating Direction Method of Multipliers
PG	Proximal Gradient
DLGC	Dictionary Learning-based Granger Causality
LSTM	Long Short-Term Memory
MINE	Mutual Information Neural Estimator
INFR-GC	INterpretable Feature Representations for Granger Causality
MLE	Maximum Likelihood Estimation
OLS	Ordinary Least Squares
MCMC	Markov Chain Monte Carlo
SGD	Stochastic Gradient Descent
GRU	Gated Recurrent Unit
RNN	Recurrent Neural Network
RF	Random Forest
EM	Expectation-Maximization
CNN	Convolutional Neural Network
RC	Reservoir Computing
CI	Confidence Interval
WSR	Wilcoxon Signed-Rank Test
SHAP	SHapley Additive exPlanations
SCM	Structural Causal Model
ATE	Average Treatment Effect
CATE	Conditional Average Treatment Effect
SUTVA	Stable Unit Treatment Value Assumption
ICP	Invariant Causal Prediction
DAG	Directed Acyclic Graph
KDE	Kernel Density Estimation
SVM	Support Vector Machine
PCA	Principal Component Analysis
LASSO	Least Absolute Shrinkage and Selection Operator
EHR	Electronic Health Record
MS	Multiple Sclerosis

List of Tables

2.1	EEG Frequency Bands Decomposition	17
2.2	Assumptions for Granger causal Analysis	34
4.1	Comparative Analysis of Time Series Causal Methods	83
4.2	Performance Comparison of LSTM, RC, and INFR-GC Models on Lorenz96 Data	100
5.1	Performance Comparison of LSTM, RC, INFR-GC, and DBN-Informed CCM Models on Lorenz96 Data	125
5.2	Comparison of Standard CCM and DBN-informed CCM results . . .	136

List of Figures

2.1	Overview of cortico-muscular coupling estimation	33
2.2	Data collection during controlled motor task	40
3.1	Proposed method overview: with (SSD-ADMM) vs. without (SS-ADMM) denoising.	61
3.2	Block diagram for our proposed method DLGC	63
3.3	Model order identified by SS-ADMM Vs BIC [104].	68
3.4	Eigenvalues of Companion matrix visualized	69
3.5	Log of normalized error of HS-ADMM [64] vs SS-ADMM.	70
3.6	Signal decomposition for simulated signal	71
3.7	NMSE Vs $\text{var}(\Delta Y)$	72
3.8	First row: Traditional GC [70], Second row: VAR (BIC+Wiener filtering [111]), Third row: SS-ADMM, Fourth row: SSD-ADMM . . .	73
3.9	EEG signal components for subject D	75
3.10	First col: Traditional GC [70], Second col: SS-ADMM, Third col: DLGC.	76
3.11	Comparative performance analysis of PDOM and PG	77
4.1	Overview of Linear and Non-linear methods	82
4.2	Overview of INFR-GC model	92
4.3	<i>Granger</i> -causality for healthy subjects and dystonia patients with critical value (pink) and stimulus onset (green)	101
4.4	Comparison of GC distribution for first 20 trials.	101
4.5	Feature Importance for difference frequency bands for randomly selected 4 dystonia patients (top) and 4 healthy subjects (bottom) . . .	102
4.6	SHAP values for different frequencies in healthy subjects (top) and patients (bottom).	103
5.1	Overview of the proposed method DBN-Informed CCM	123
5.2	Dynamic Bayesian Network learnt from Lorenz96 data	124
5.3	Effect of intervention applied on variable 3 and its neighbor variable . . .	12126
5.4	CCM values before and after intervention on variable 3	127
5.5	DBN for each frequency band for healthy subject 13	129
5.6	DBN for each frequency band for dystonia subject 10	130
5.7	Connection strength for different frequency bands	131
5.8	DBN for each frequency band in healthy subjects (left) and individuals with dystonia (right), showing causality from EEG to EMG (odd rows) and from EMG to EEG (even rows).	132
5.9	Average CCM values for brain to muscle causality	133

5.10	Average CCM values for muscle to brain causality	133
5.11	Average connection strength for bidirectional causality	134
5.12	Standard CCM approach vs DBN-Informed CCM	137
5.13	Noise intervention in EEG signal for healthy subject 23	138
5.14	Noise intervention in EMG signal for healthy subject 23	139
5.15	Pre and Post Intervention CCM values for all subjects	140
6.1	Time series causality method selection framework	146
6.2	Proposed Clinical Implementation Framework	159

Contents

Abstract	3
Acknowledgements	4
List of Abbreviations	7
1 Introduction	1
1.1 Significance of Neural-Muscular Coupling in Movement Disorders . . .	3
1.2 Challenges in Modeling Neural-Muscular Coupling	5
1.3 Biomarkers for Dystonia	6
1.4 Research Questions and Objectives	8
1.5 Methodological Approach Overview	10
1.6 Thesis Contributions and Organization	11
2 Fundamentals of Cortico-Muscular Coupling Analysis	16
2.1 Physiological Signal Processing	17
2.1.1 EEG Signal Characteristics	18
2.1.2 EMG Signal Characteristics	19
2.1.3 Time-Frequency Analysis Methods	20
2.2 Coupling Analysis Metrics	21
2.2.1 Linear Metrics	22
2.2.1.1 Coherence Analysis	23
2.2.1.2 Directed Transfer Function	23
2.2.2 Information Theoretic Metrics	25
2.2.2.1 Mutual Information	26
2.2.2.2 Transfer Entropy	26
2.2.3 Advanced Coupling Metrics	27
2.2.3.1 Partial Directed Coherence	27
2.2.3.2 Convergent Cross Mapping	29
2.3 Granger Causality Framework	32
2.3.1 Key Assumptions	34
2.3.2 Non-linear Granger Causality	36
2.4 Experimental Paradigm	38
2.4.1 Healthy Adults Motor Perturbation Dataset (HAMPD)	39
2.4.2 Pediatric Motor Perturbation Dataset (PMPD)	39
2.4.3 Lorenz96 Dynamical System Implementation	41

3	Linear Autoregressive Models: Enhancing Model Integrity	43
3.1	Model Order Identification by Hierarchical Sparsity	46
3.2	Proximal Operator for Stationarity Condition	48
3.3	Globally Convergent Method for Model Order and Stationarity	50
3.4	Non-convex ADMM for Noise Robust Estimation	53
3.5	Data-Driven Learned Dictionary for VAR Modeling	61
3.6	Experimental Results	66
3.6.1	Evaluation for SS-ADMM	67
3.6.2	Evaluation for SSD-ADMM	70
3.6.3	Evaluation for DLGC	74
3.7	Insights and Implications	78
4	Towards Interpretable Non-linear Approaches	80
4.1	Advanced Non-linear Approaches	84
4.1.1	Long Short-Term Memory networks	85
4.1.2	Reservoir Computing for Non-linear Dynamics	87
4.2	Interpretable Random Forest-Based Coupling Analysis	89
4.2.1	Mathematical Formulation	91
4.2.2	Feature Selection Framework	94
4.2.3	Insights on Model Interpretability	96
4.2.4	Comparative Analysis on Lorenz96 Data	99
4.2.5	Experimental Evaluation on Physiological Data	99
4.3	Insights and Implications	104
5	Beyond Predictive Ability: Counterfactual Perspective on Causality	106
5.1	Dynamic Bayesian Networks	109
5.1.1	Structure Learning	110
5.1.2	Parameter Learning	111
5.2	DBN-Informed Convergent Cross Mapping	115
5.2.1	Theoretical Foundation	116
5.2.2	Intervention Implementation	119
5.2.3	Experimental Evaluation on Lorenz96 Data	123
5.2.4	Experimental Evaluation on Physiological Data	128
5.2.5	Case Study: Comparative Analysis of Standard and DBN-informed CCM	135
5.2.6	Case Study: Analysis of Counterfactual Interventions	137
5.3	Insights and Implications	142
6	Conclusion	145
6.1	Principled Approach to Causal Inference in Time Series	145
6.2	Summary of Thesis Achievements	147
6.3	Methodological Constraints and Assumptions	149
6.3.1	Stationarity Considerations	149
6.3.2	Scalability Considerations	150
6.3.3	Model Order Selection	151
6.3.4	Regularization Parameters	152
6.3.5	Numerical Stability	153
6.3.6	Validation Strategies	153

6.4	Navigating Implementation Barriers in Clinical Analytics	154
6.5	Real-time Implementation and Clinical Translation	157
6.5.1	Potential Clinical Applications in Dystonia Management . . .	157
6.5.2	Proposed Clinical Implementation Framework	158
6.5.3	Theoretical Case Study Design	160
6.5.3.1	Initial Assessment Phase	160
6.5.3.2	Longitudinal Monitoring Phase	160
6.5.3.3	Outcome Evaluation	161
6.6	Future Work	161
6.6.1	Extensions to Other Neurological Conditions	165
6.7	Conclusion	167

Chapter 1

Introduction

Motor control in humans embodies a complex, multimodal system of neural computation and biomechanical execution that demonstrates causal relationships across multiple biological scales. The proper functioning of the nervous system requires seamless coordination of distributed neural networks, peripheral nerve fibers, and contractile tissues to deliver precisely timed signals to skeletal muscles, ensuring both temporal and spatial coordination. This intricate interplay between neural and muscular systems is foundational to our ability to move, react, and adapt to dynamic environments.

Despite significant advances in understanding motor control, a critical gap exists in the accurate diagnosis and mechanistic understanding of movement disorders, particularly dystonia. This condition affects tens of thousands of people worldwide yet continues to present substantial diagnostic challenges, with patients often experiencing prolonged suffering due to misdiagnosis and delayed access to effective therapies. The fundamental challenge lies in quantitatively characterizing the disrupted causal pathways between neural activity and muscular response that underlie dystonic movements. Current diagnostic approaches lack the sensitivity and specificity needed for early detection, while existing analytical methods struggle to capture the complex, nonlinear dynamics of cortico-muscular interactions in pathological states.

This thesis addresses these challenges by developing novel mathematical frame-

works for analyzing electroencephalography (EEG) and surface electromyography (sEMG) signals that can reveal causal disruptions in neural-muscular coupling. The work is guided by the central research question:

Can robust, interpretable causal analysis of cortico-muscular interactions reveal distinctive patterns that reliably differentiate dystonia from healthy motor control and inform targeted therapeutic interventions?

In recent years, advancements in computational neuroscience and biomechanics have provided new methods to study neural-muscular processes in greater detail. These innovations have enabled unprecedented investigation into the causal mechanisms underlying the nerve-muscle interactions that drive motor control, offering deeper insights into how these systems work together to produce coordinated movement. High-temporal resolution techniques, including simultaneous electromyography (EMG) and motor unit recordings, have revealed intricate patterns of neural recruitment and muscle fiber activation [125]. These findings illuminate the hierarchical nature of motor control, where descending neural commands are dynamically modulated through both feedforward and feedback mechanisms.

The integration of mechanistic modeling with empirical observations has been especially valuable in uncovering the computational principles that govern neuromuscular junction signaling, providing a clearer understanding of how these processes coordinate to enable muscle activation. Studies using optogenetic manipulation combined with high-precision force measurements have revealed that motor neurons exhibit exceptional specificity in their timing patterns, which directly impacts the contractile properties of muscle fibers [92, 121]. This deeper mechanistic understanding of how neuromuscular control systems function holds significant implications for both theoretical neuroscience and clinical applications.

Moreover, advanced imaging techniques, in conjunction with machine learning approaches, have allowed researchers to build detailed computational models of the neuromuscular system [130]. These models capture the full complexity of the system,

from cellular calcium dynamics to the mechanical properties of tissues, offering a more integrated framework for understanding how neural signals are converted into coordinated movement patterns [17]. The ability to precisely identify the causal relationships between neural activity and muscular response not only enhances our basic understanding of motor control but also guides the development of more effective treatments for movement disorders.

Our theoretical framework builds upon these advances by integrating causality theory with neurophysiological modeling. We conceptualize neural-muscular coupling as a dynamic causal system where temporal relationships between cortical activity and muscle activation can be quantified through advanced signal processing techniques. This framework extends beyond traditional correlation-based analyses by incorporating both predictive causality (Granger causality) and interventional perspectives (counterfactual analysis) to better characterize the complex, bidirectional information flow between brain and muscle.

1.1 Significance of Neural-Muscular Coupling in Movement Disorders

Cortico-muscular interactions play a central role in motor control, forming the basis for the seamless execution of voluntary movements. These interactions involve the communication between the motor cortex inside the brain, which plans and initiates movement, and the peripheral musculature, which executes it. Neural signals generated in the motor cortex travel through descending pathways, including the corticospinal tract, to activate spinal motor neurons, which then innervate muscles, leading to contraction and movement. Sensory feedback from muscles and joints is integrated by the central nervous system to refine motor output in real time. This bidirectional communication ensures precision, adaptability, and coordination in motor tasks.

Movement disorders such as Parkinson’s disease, dystonia, and ataxia illustrate

the devastating consequences of disruptions in these cortico-muscular pathways. For instance, Parkinson’s disease is characterized by the degeneration of dopaminergic neurons in the substantia nigra, leading to impaired motor planning and execution [65]. Symptoms include tremors, rigidity, bradykinesia, and postural instability, all of which stem from altered neural connectivity and impaired cortico-muscular coupling. Dystonia, on the other hand, involves involuntary muscle contractions resulting in abnormal postures or repetitive movements, often linked to dysfunctional basal ganglia circuits and their interactions with cortical motor areas [78]. Ataxia is another movement disorder that reflects cerebellar dysfunction, leading to a lack of coordination, balance problems, and erratic movement patterns [3], [76]. These disorders highlight how complex interactions between brain regions and the motor system maintain functional movement. Therefore, studying cortico-muscular interactions is vital for understanding and treating movement disorders [19, 53].

By investigating the mechanisms underlying normal motor control, researchers can identify how disruptions in neural pathways contribute to pathological motor symptoms. Advanced techniques such as Electroencephalography (EEG), Electromyography (EMG), and functional neuroimaging (PET, fMRI, MEG, etc.) allow for the detailed analysis of cortico-muscular interactions. Insights from these studies inform the development of targeted therapies, such as Deep Brain Stimulation (DBS) that modulates abnormal neural activity in specific brain regions. It also helps in refining rehabilitation strategies, neuroprosthetics, and pharmacological interventions aimed at restoring motor function. Understanding cortico-muscular dynamics advances both fundamental neuroscience and clinical applications, from personalized DBS protocols to adaptive rehabilitation technologies [44].

1.2 Challenges in Modeling Neural-Muscular Coupling

Modeling the neural-muscular coupling presents significant methodological challenges, which this thesis aims to address. We identify four critical limitations in current approaches:

- The neural control of motor systems, particularly in the presence of pathological conditions like dystonia, is highly nonlinear. For instance, the process of motor unit recruitment (the activation of individual muscle fibers through motor neurons) is nonlinear and state-dependent. Similarly, the neural signal processing in motor cortex areas involves complex nonlinear dynamics that are difficult to model using linear systems alone. Understanding the precise way these nonlinearities manifest in conditions like dystonia, where neural feedback loops may be distorted, requires advanced models that can account for these complexities.
- One of the most significant challenges in any neurophysiological research, especially involving EEG and EMG, is the presence of physiological and measurement noise. In clinical settings, signal quality can vary greatly due to factors such as electrode placement, motion artifacts, and individual differences in skin conductivity. Additionally, extraneous signals from other sources (e.g., eye movements, cardiac activity, or even external environmental noise) can severely compromise the accuracy of causal inference.
- Dystonia, like other movement disorders, can manifest in a variety of ways, including focal dystonias, where one part of the body is affected, or generalized dystonias, where multiple muscle groups are involved. This heterogeneity poses a major challenge for identifying reliable biomarkers or causal signatures that can accurately reflect the disease state across different individuals. Additionally,

variability in disease progression, medication use, and external factors makes it difficult to pinpoint stable, reproducible markers of pathology.

- Despite advances in neuroscience, we still lack a clear, mechanistic understanding of many movement disorders. In particular, conditions like dystonia are poorly understood, with multiple potential causes including basal ganglia dysfunction, abnormal sensory processing, and impaired central nervous system plasticity. While we can observe and quantify changes in neural and muscular activity, understanding the causal mechanisms that underlie these changes is a significant challenge. For example, how do certain changes in cortical oscillations lead to the abnormal muscle contractions observed in dystonia? Without a reliable method for inferring causal relationships, treatment strategies remain largely symptom-focused, rather than targeted at the root causes of the disease.

These limitations have direct clinical implications that have impeded the progress in developing reliable diagnostic biomarkers. This underscores the importance of advancing analytical tools and deepening our understanding of the condition’s underlying mechanisms.

1.3 Biomarkers for Dystonia

At its core, dystonia disrupts the intricate communication between the brain and muscles, leading to involuntary and often debilitating movements. However, the precise mechanisms driving these disruptions remain poorly understood. This gap in knowledge reflects broader challenges in connecting brain activity with physical manifestations in movement disorders. The complexity of these neural-muscular interactions necessitates advanced analytical approaches to identify reliable diagnostic markers.

Recent advances in neuroimaging and signal processing have identified promising biomarkers for dystonia diagnosis and characterization. These biomarkers can be categorized into three types:

- *Diagnostic biomarkers:* Indicators that confirm the presence of dystonia and differentiate it from other movement disorders. For instance, specific genetic mutations in the TOR1A gene (DYT1) provide definitive diagnostic evidence for early-onset generalized torsion dystonia [13].
- *Prognostic biomarkers:* Measurements that predict disease progression or severity. The pattern of cerebellar connectivity abnormalities detected through advanced neuroimaging techniques correlates strongly with long-term clinical outcomes in cervical dystonia patients [59].
- *Predictive biomarkers:* Signals that forecast treatment response, such as to Deep Brain Stimulation (DBS) therapy. Baseline sensorimotor plasticity measured via paired associative stimulation protocols accurately predicts individual patient responsiveness to botulinum toxin treatment in focal hand dystonia [96].

The development of machine learning-driven frameworks has enabled unprecedented insight into structural and functional alterations associated with dystonic conditions. The DystoniaNet framework [119] has revealed significant structural alterations in specific brain regions, including the corpus callosum, thalamus, and parts of the frontal and temporal lobes. These regions demonstrate crucial roles in interhemispheric information transfer and heteromodal sensorimotor processing, with the DystoniaNet-based biomarker achieving remarkable diagnostic accuracy of 98.8%. The identification of these structural biomarkers represents a significant advancement in our understanding of dystonia’s neuroanatomical basis.

Building upon these structural insights, comprehensive analysis of coherence patterns across multiple neural substrates has provided additional insights into dystonia pathophysiology [57]. This analytical approach extends beyond static structural alterations to examine dynamic functional relationships within the neural network. A review encompassing 41 studies and 395 adult dystonia patients revealed distinctive patterns across corticocortical, corticopallidal, corticomuscular, pallidopallidal,

pallidomuscular, and intermuscular coherence measures. The consistent observation of increased 4-12 Hz coherence across various dystonia subtypes presents a promising direction for biomarker development. In a study in [55], Hjort Mobility is identified as a biomarker that correlates with the long-term effects of Deep Brain Stimulation (DBS) [57]. This indicates its potential utility as a predictive biomarker for patient response to DBS treatment. These findings establish a crucial link between neural synchronization patterns and the clinical manifestations of dystonia, as discussed in [57].

The emergence of advanced computational methodologies has further enhanced our capacity to characterize neural-muscular interactions in dystonia. Recent methodological advances in causal inference have introduced more advanced frameworks for analyzing temporal relationships in physiological signals. Developments in Granger Causal structure learning from heterogeneous interventional data [137] and self-explaining neural networks for multivariate Granger causality [72] have enabled more nuanced analysis of temporal dependencies and causal relationships. These approaches transcend traditional correlation-based analyses by enabling the extraction of diverse causal information across different delay ranges, integrated through learned attention weights to capture comprehensive relationships between neural time series [31].

Despite these advances, significant gaps remain in translating these findings into clinically useful tools. The integration of these methodological advances with clinical observations provides the foundation for our research approach, which aims to develop more robust, interpretable, and clinically relevant biomarkers for dystonia.

1.4 Research Questions and Objectives

Based on the identified research gaps and challenges, this thesis addresses four critical challenges in neurophysiological signal analysis, each representing the intersection of methodological innovation and clinical need in movement disorder research.

1. *How can we enforce mathematical constraints to ensure stability and reliability in time series modeling of neurophysiological signals?* Dystonia exhibits highly variable symptom patterns, which makes consistent mathematical characterization challenging. Our primary objective is to develop analytical methods that can accommodate this clinical heterogeneity while preserving physiological plausibility. To that end, we applied stationarity-enforcing transformations to EEG and EMG time-series data prior to modeling their interactions, thereby ensuring that our models capture genuine neural dynamics and motor responses rather than artifacts arising from non-stationary statistics.
2. *How can we develop more robust methods for estimating directional relationships in EEG-EMG signals that account for physiological and measurement noise?* The analysis of neural-muscular coupling is consistently undermined by the inherent noise in clinical recordings from motion artifacts to electrical interference. Building on the identified challenge of noise contamination in neurophysiological signals, we aim to develop analytical approaches specifically designed to maintain accuracy even under suboptimal recording conditions.
3. *What analytical frameworks can capture complex neural-muscular interactions while maintaining interpretability for clinical applications?* The recognized nonlinearity in motor system function, particularly evident in motor unit recruitment and cortical processing in dystonia, demands models that transcend simple linear approximations. Our objective is to develop methods that capture these complex nonlinear dynamics while producing results that clinicians can confidently interpret and apply in practice.
4. *How can intervention-based analytical approaches enhance our understanding of underlying mechanisms in dystonia beyond traditional correlational frameworks?* Our incomplete mechanistic understanding of movement disorders like dystonia limits treatment to management of symptoms rather than addressing root causes. Our final objective develops frameworks for simulating system responses

to potential interventions, creating a foundation for mechanistic insights that could guide more targeted therapeutic approaches.

These questions form a progressive research agenda that builds from fundamental signal processing challenges toward clinical application, with each advance laying essential groundwork for subsequent innovations. Together, they address the critical need for more reliable, interpretable, and clinically relevant neurophysiological analysis in movement disorder research.

1.5 Methodological Approach Overview

This thesis develops and evaluates a unified analytical pipeline for inferring causal relationships in physiological signals, with particular application to movement disorders such as dystonia. Emphasizing both methodological rigor and clinical interpretability, the pipeline proceeds through four integrated phases. First, simultaneous EEG–EMG recordings are obtained from dystonia patients and matched healthy controls under a standardized protocol; the raw data are then subjected to automated artifact detection and rejection, band-pass filtering, and precise temporal alignment to ensure high signal quality. In the second phase, these preprocessed time series are mapped into enriched mathematical representations via time–frequency decompositions (for example, wavelet or multitaper spectral analysis) and state-space embedding; from these representations we extract physiologically relevant features such as spectral power, phase dynamics, and state-space coordinates.

Building on these representations, the third phase implements a multi-level causal analysis framework. At the time-domain level, we fit sparse Vector AutoRegressive (VAR) models using the Alternating Direction Method of Multipliers (ADMM) to recover linear interactions. In parallel, a component-based analysis employs dictionary learning and structural decomposition to isolate interacting signal sub-components. Finally, a nonlinear feature-based analysis applies machine-learning classifiers, augmented by expert-guided interpretation, to embeddings and derived

features, capturing complex dependencies beyond linear assumptions.

To validate our approach, the fourth phase combines synthetic-data experiments, where ground-truth causal graphs are known, with k-fold cross-validation on clinical EEG-EMG datasets annotated by established pathophysiological markers. We benchmark our results against standard causality metrics and perform statistical tests to quantify the discriminative power between dystonia and control cohorts. Each analytical technique is evaluated on four key performance criteria: accuracy of causal-structure recovery, robustness to measurement noise, computational efficiency, and clinical discriminative power. By integrating these phases into a coherent workflow, our methodology achieves both high analytical precision and the interpretability required for translational applications in clinical neurophysiology.

1.6 Thesis Contributions and Organization

This thesis makes four principal contributions to the fields of signal processing, causal inference, and clinical neurophysiology:

1. **Stationary-Sparse ADMM Framework:** We introduce a globally convergent optimization framework using ADMM that simultaneously determines model order through hierarchical sparsity constraints and ensures stationarity in VAR models via a novel proximal operator. This contribution addresses the fundamental challenge of obtaining reliable signal representations from noisy physiological recordings, enabling more accurate downstream causal analysis.
2. **Dictionary Learning for Noise-Robust Causality:** We develop a structural decomposition method that distinguishes between physiological and measurement noise, coupled with a Dictionary Learning approach that identifies characteristic signal components while determining causal relationships. This novel integration overcomes a key limitation in current approaches by explicitly modeling different noise sources, significantly improving causal discovery in noisy real-world signals.

3. **Interpretable Feature-Based Causality Framework:** We create an interpretable nonlinear causality framework combining Random Forest modeling with neural network-guided feature selection, enabling robust characterization of complex physiological interactions while maintaining analytical transparency. This contribution bridges the gap between model complexity and clinical interpretability, a critical requirement for translational neurophysiology.
4. **Counterfactual Analysis for Physiological Signals:** We design a systematic approach to counterfactual analysis in dynamic Bayesian networks, providing novel insights into intervention effects and their implications for cross-convergent mapping in physiological signal analysis. This methodological innovation extends causality analysis beyond traditional correlation or prediction to examine potential effects of clinical interventions.

These contributions have significant translational potential. By improving the ability to detect and characterize causal disruptions in cortico-muscular interactions, our methods can enhance diagnostic precision for dystonia, potentially reducing time to diagnosis and improving treatment selection. The interpretable nature of our frameworks facilitates adoption in clinical settings by providing actionable insights into the mechanisms underlying movement disorders.

This thesis is structured into six chapters, each building upon the previous to address the overarching research objectives.

Chapter 1 serves as an introduction, outlining the problem statement, the research objectives, and the methodological approach. This foundational chapter establishes the context for the research, ensuring clarity in the objectives and the analytical framework taken throughout the thesis.

Building upon this introduction, Chapter 2 lays the theoretical groundwork by reviewing corticomuscular coupling fundamentals and relevant metrics. It explores contemporary signal processing and information-theoretic methods for detecting corticomuscular interactions, delving into their theoretical underpinnings, extensions,

and applications. This chapter critically evaluates the assumptions and limitations of existing methods, highlighting the gaps that motivate the novel approaches developed in subsequent chapters.

In response to these limitations, Chapter 3 introduces innovative methodologies aimed at enhancing model integrity. By incorporating structural and modeling constraints, this chapter emphasizes improved noise handling in physiological signals as a crucial step for reliable neurophysiological analysis. These advancements set the stage for the exploration of more advanced models.

Chapter 4 continues this progression by focusing on the development of non-linear yet interpretable models for feature extraction and causality estimation. Comparative analyses of these approaches provide valuable insights into their performance and applicability, addressing the complexities inherent in neural-muscular interaction analysis.

To further advance the analytical framework, Chapter 5 investigates counterfactual intervention analysis for improved causality estimation. This chapter bridges theoretical innovation with clinical relevance, demonstrating how counterfactual approaches can yield more robust and actionable insights into causality in physiological systems.

Finally, Chapter 6 synthesizes the contributions of the thesis, presenting a comprehensive summary of findings and their implications. This concluding chapter also outlines potential directions for future research, emphasizing the broader impact of the methodologies developed in this work on neurological investigations and movement disorder analysis.

The organization reflects a progression from theoretical foundations to applied innovations, providing a structured road map for advancing our understanding of dystonia and related movement disorders. The relevant research outputs are as follows:

1. F. Abbas, V. McClelland, W. Dai , Z. Cvetkovic, “ INFR-GC: Interpretable Feature Representations For Granger Causality In Cortico-Muscular Interac-

- tions ”, *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2025. doi: 10.1109/ICASSP49660.2025.10887822.
2. F. Abbas, V. McClelland, Z. Cvetkovic, W. Dai, “Stationary and Sparse Denoising Approach for Corticomuscular Causality Estimation”, *IEEE Transactions on Biomedical Engineering (TBME)*, 2025. doi: 10.1109/TBME.2024.3518602.
 3. F. Abbas, V. McClelland, Z. Cvetkovic, W. Dai, “DLGC: Dictionary Learning based Granger Causal Discovery for Cortico-muscular Coupling”, *European Conference on Signal Processing (EUSIPCO)*, 2024. doi: 10.23919/EUSIPCO63174.2024.10715346.
 4. F. Abbas, V. McClelland, Z. Cvetkovic, W. Dai, “SS-ADMM: Stationary and Sparse Granger Causal Discovery for Cortico-muscular Coupling”, *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2023. doi: 10.1109/ICASSP49357.2023.10095111.
 5. F. Abbas, V. McClelland, Z. Cvetkovic, W. Dai, “DBN-Informed Cross Convergent Mapping: An Interventional Framework for Causal Discovery”, *To be submitted in Nature Biomedical Engineering*.

Our methodological advances have been rigorously validated on both synthetic datasets, where ground-truth causal structures are known, and on clinical EEG–EMG recordings from dystonia patients and matched healthy controls. We assess performance in terms of causality detection between patient and control signals, robustness to additive and physiological noise relative to benchmark methods, and clinical interpretability of the extracted features.

Nevertheless, several limitations warrant consideration. First, the present models were trained and tested primarily on select dystonia subtypes, so further work is required to ensure generalizability across the full clinical spectrum. Second, variability in EEG–EMG acquisition protocols across centers may affect reproducibility and will need harmonization strategies. Third, surface-recorded signals inherently

underrepresent deep-brain dynamics, suggesting the value of integrating invasive recordings or complementary imaging modalities.

Future research should therefore pursue multi-center studies with standardized acquisition pipelines and explore hybrid multimodal frameworks, combining EEG–EMG with, for example, local field potentials or functional neuroimaging, to overcome these limitations and extend the translational impact of our causal-analysis pipeline.

Chapter 2

Fundamentals of Cortico-Muscular Coupling Analysis

The analysis of brain-muscle interactions through simultaneous recording of cortical and muscular electrical activity has emerged as a powerful tool for understanding motor control mechanisms. EEG and EMG signals recorded during controlled movement tasks serve as key indicators of corticomuscular coupling, quantifying the functional connectivity between motor cortex and muscles. This coupling reflects the intricate communication pathways essential for precise movement execution and control. This allows researchers to characterize both normal motor control patterns and their disruption in various movement disorders, offering a window into the underlying pathophysiological mechanisms. For instance, the temporal evolution of corticomuscular coupling provides crucial information about motor control dynamics. Phase relationships between cortical and muscular signals can elucidate the temporal ordering of neural events, helping to distinguish between direct cortical commands and feedback-modulated responses. This temporal precision in signal analysis enables researchers to map the sequential activation patterns that underlie coordinated movement execution.

Furthermore, the spatial distribution of corticomuscular coupling across the cortical surface offers insights into the functional organization of motor control networks.

Table 2.1: EEG Frequency Bands Decomposition

Frequency Band	Description
Gamma (30-70 Hz)	Reflects higher cognitive functions such as memory, attention, and sensory processing.
Beta (14-30 Hz)	Associated with active thinking, attention, and motor activity.
Mu (10-13 Hz)	Linked to sensorimotor processes and often associated with motor control.
Alpha (8-13 Hz)	Prominent during relaxed wakefulness and reduced during active mental engagement.
Theta (4-8 Hz)	Linked to drowsiness, meditation, or early stages of sleep.
Delta (0.5-4 Hz)	Dominates during deep sleep or brain injury.

Source localization techniques [69] applied to EEG data can reveal the specific cortical regions most strongly coupled with particular muscle groups, advancing our understanding of the somatotopic organization of motor control systems. This spatial mapping capability proves particularly valuable in investigating pathological conditions where normal motor control patterns are disrupted by analyzing the recorded physiological signals. However, effectively leveraging these insights requires addressing the unique challenges associated with recording and processing EEG and EMG signals, as discussed in the following section.

2.1 Physiological Signal Processing

The accurate recording and analysis of physiological signals present unique challenges due to their complex nature, susceptibility to noise, and the need for advanced processing techniques. Each signal carries distinct properties and requires specific approaches for acquisition, preprocessing, and analysis. Signal processing methods must account for various artifacts, frequency content, and temporal dynamics while preserving the underlying physiological information. Moreover, the interpretation of these signals demands careful consideration of both their physiological origins

and the technical limitations of recording methods. The following subsections detail the specific characteristics and processing considerations for EEG and EMG signals, forming the foundation for understanding corticomuscular coupling analyses. These signals, when properly processed and analyzed together, reveal crucial information about the neural mechanisms of motor control and their potential alterations in pathological conditions.

2.1.1 EEG Signal Characteristics

Electroencephalography (EEG) signals represent the electrical activity of large populations of neurons in the cortex, providing crucial insights into brain function and neural processing. These signals exhibit characteristic oscillations within specific frequency bands, each associated with distinct brain states or cognitive functions [118]. The EEG spectrum can be decomposed into several key frequency bands as outlined in Table 2.1. The human brain exhibits distinct oscillatory patterns across various frequency bands, with each band reflecting specific cognitive states, behavioral conditions, and neurophysiological processes. This frequency-specific organization of neural activity provides valuable insights into brain function and dysfunction across different behavioral and cognitive contexts.

In the time domain, EEG signals reveal instantaneous electrical activity patterns, where specific waveform shapes and amplitudes indicate various normal and pathological brain functions. Of particular interest are Event-Related Potentials (ERPs) [139], which represent time-locked changes in neural activity in response to specific stimuli or cognitive events. For instance, the N400 component is crucial for investigating semantic processing deficits in conditions like schizophrenia [37], while the P300 component, prominently observed in oddball paradigms, reflects attention and cognitive processing of rare stimuli [97].

The spatial distribution of EEG activity across the scalp also provides valuable diagnostic information. Different brain regions exhibit characteristic patterns of activity, and the spatial organization of these signals can help localize neurological

abnormalities. Additionally, artifacts in EEG signals, including eye movements, muscle activity, and cardiac signals, require careful consideration during analysis. Common preprocessing steps include filtering, artifact removal techniques such as Independent Component Analysis (ICA), and reference selection, which significantly impact signal quality and interpretation [48].

2.1.2 EMG Signal Characteristics

Surface electromyography (sEMG) signals capture the electrical activity generated by skeletal muscles during contraction and relaxation. In the time domain, these signals provide direct insights into muscle activation patterns, including contraction onset, duration, and intensity. Key quantitative measures such as root mean square (RMS) and mean absolute value (MAV) help assess muscle activation strength and timing. The baseline EMG signal during rest should exhibit minimal activity, with deviations potentially indicating neuromuscular disorders. In the frequency domain, EMG signals reflect the recruitment patterns of different muscle fiber types, with low frequencies corresponding to slow-twitch fibers and higher frequencies indicating fast-twitch fiber activation [81]. Notably, muscle fatigue manifests as a characteristic decrease in the signal's median frequency, reflecting changes in muscle fiber conduction velocity and recruitment patterns [133].

The complexity of EMG signals is further influenced by anatomical factors such as subcutaneous tissue thickness, electrode placement, and cross-talk between adjacent muscles. Modern recording techniques employ various electrode configurations, including monopolar, bipolar, and array recordings, each offering distinct advantages for specific applications [11]. Additionally, the relationship between EMG amplitude and force production is non-linear and can be affected by factors such as muscle length, contraction velocity, and fatigue state, necessitating careful interpretation in both clinical and research settings.

Given these challenges in interpreting EEG and EMG amplitude, it becomes essential to apply advanced analytical techniques that can effectively capture the

dynamic characteristics of physiological signals, which is addressed through time-frequency analysis methods in the following section.

2.1.3 Time-Frequency Analysis Methods

The analysis of EEG and EMG signals employs various time-frequency techniques to capture their dynamic characteristics. The Fourier Transform (FT) serves as a fundamental tool for decomposing signals into their constituent frequency components, enabling the calculation of Power Spectral Density (PSD) to quantify frequency-specific signal power. However, the FT's assumption of signal stationarity limits its application to dynamic physiological signals. To address this limitation, advanced techniques such as the Wavelet Transform (WT) provide simultaneous temporal and spectral analysis. The WT offers superior resolution for analyzing non-stationary signals through multi-resolution decomposition, making it particularly valuable for detecting transient events like epileptic spikes in EEG or fatigue-related changes in EMG. The Short-Time Fourier Transform (STFT) offers an alternative approach, using a sliding window to capture temporal changes in frequency content. These time-frequency analysis methods have crucial applications in both clinical diagnosis and research. In EEG analysis, they help identify pathological conditions through specific frequency band alterations and detect transient neurological events. For EMG, these techniques enable the assessment of muscle fatigue, characterization of movement disorders like Parkinson's disease through tremor frequency analysis, and evaluation of muscle fiber recruitment patterns. This comprehensive analysis framework supports various applications, from neurological diagnosis to the development of brain-computer interface systems.

Recent advances in signal processing have introduced more analysis techniques that extend beyond traditional methods. Empirical Mode Decomposition (EMD) [116] adaptively decomposes signals into intrinsic mode functions, providing powerful tools for analyzing non-linear and non-stationary processes in physiological signals. Higher-order spectral analysis has emerged as a valuable technique for revealing

phase coupling between different frequency components, offering deeper insights into signal interactions. Additionally, time-varying coherence measures [47] have become essential for studying dynamic relationships between EEG and EMG signals during movement, capturing the temporal evolution of corticomuscular coupling. These advanced methods have significantly enhanced our ability to detect subtle changes in neural and muscular activity patterns, proving particularly valuable in early disease detection and monitoring treatment responses. The integration of these techniques has opened new avenues for understanding complex neurophysiological processes and their alterations in pathological conditions.

By employing these advanced analytical techniques, more meaningful and precise patterns can be extracted from EEG and EMG data, facilitating a deeper understanding of both physiological processes and their pathological deviations. Enhanced representations, such as time-domain waveforms, frequency-domain power spectra, and wavelet scalograms, further enrich this interpretation of complex signal behaviors.

2.2 Coupling Analysis Metrics

The quantification of cortico-muscular coupling represents a fundamental challenge in neuroscience that necessitates rigorous methodological consideration. The selection and application of appropriate coupling metrics serve as critical determinants in explaining the complex interactions between cortical and muscular systems. These metrics must not only capture the underlying physiological phenomena but also address the inherent complexities of signal acquisition, processing, and interpretation.

The theoretical foundations of coupling analysis encompass a diverse array of mathematical frameworks, each offering distinct advantages in characterizing specific aspects of neural-muscular interactions. Linear coupling metrics, while computationally tractable and interpretable, may fail to capture the full complexity of physiological signal relationships. Conversely, nonlinear approaches offer enhanced sensitivity to complex coupling patterns but often present challenges in computational

efficiency and result interpretation. This fundamental trade-off between analytical power and practical implementation underlies many methodological decisions in coupling analysis.

Contemporary approaches to coupling quantification have evolved to address these challenges through increasingly complex mathematical frameworks. The following sections present an examination of several coupling metrics that have widely been employed. Each metric offers unique insights into different aspects of cortico-muscular coupling, from linear frequency-domain relationships to complex nonlinear interactions and causal dependencies.

The progressive development of these metrics reflects broader advances in signal processing and information theory, enabling more nuanced understanding of neural-muscular communication pathways. This comprehensive examination of coupling metrics establishes a foundation for subsequent analysis of cortico-muscular interactions in both physiological and pathological conditions. The sections that follow detail each metric’s mathematical basis, implementation considerations, and specific advantages in characterizing cortico-muscular coupling. This analysis facilitates informed selection of appropriate methodologies for specific research objectives while acknowledging the inherent complexities of physiological signal analysis.

2.2.1 Linear Metrics

Linear coupling metrics provide foundational tools for quantifying relationships between time series in the frequency domain, offering insights into both direct and indirect interactions. These metrics have gained widespread adoption due to their mathematical tractability and interpretability, particularly in scenarios where the underlying system dynamics can be reasonably approximated by linear models. The analytical framework of linear coupling encompasses both temporal correlations and spectral relationships, enabling a comprehensive understanding of signals’ interdependencies. Some of the most widely used linear metrics to model corticomuscular coupling are as follows:

2.2.1.1 Coherence Analysis

Corticomuscular Coherence (CMC) represents a foundational technique in understanding neural-muscular interactions through spectral analysis. This method emerges from the critical need to quantify the linear relationships between cortical neural signals and muscular activity, typically measured through simultaneous electroencephalographic (EEG) and electromyographic (EMG) recordings. Traditionally, CMC focuses on the beta-band frequency range (13-30 Hz), which has been consistently associated with motor control processes [78, 27]. Mathematically, CMC is defined as the coherence function between cortical neural signals $x(t)$ and muscular signals $y(t)$:

$$\text{CMC}_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)} \quad (2.1)$$

where $P_{xy}(f)$ represents the cross-spectral density. $P_{xx}(f)$ and $P_{yy}(f)$ are the power spectral densities of $x(t)$ and $y(t)$ respectively.

The coherence value ranges between 0 and 1, with values closer to 1 indicating stronger linear relationships between neural and muscular signals. The metric primarily captures linear, zero-phase-lag interactions, which may sometimes oversimplify the complex neural dynamics. Caution must be exercised in interpretation, particularly concerning potential confounding factors such as volume conduction and reference electrode placement. Despite these challenges, CMC remains a powerful tool in neurophysiological research, offering valuable insights into motor control mechanisms.

2.2.1.2 Directed Transfer Function

The Directed Transfer Function (DTF) is a frequency-domain measure used to quantify the directional influence or causal interaction between two signals, typically in the context of a multivariate time-series analysis [52, 134]. It is commonly used in fields such as neuroscience, economics, and physics to assess the directionality of causality between time series, especially in systems where interactions are complex and involve multiple variables. The DTF is typically estimated using a VAR model,

which models the time series data as a linear function of past values. Given an n -variate time series $\mathbf{Y}(t) = \{Y_1(t), Y_2(t), \dots, Y_n(t)\}$, a VAR model of order m can be written as:

$$\mathbf{Y}(t) = \sum_{i=1}^m \mathbf{A}_i \mathbf{Y}(t-i) + \boldsymbol{\epsilon}(t),$$

where:

- $\mathbf{Y}(t)$ is the vector of observed variables at time t ,
- $\mathbf{A}_i$ is the coefficient matrix of the i -th lag in the VAR model,
- $\boldsymbol{\epsilon}(t)$ is the error term (or residual) at time t .

The transfer function $H_{ij}(\omega)$ from variable $Y_j(t)$ to $Y_i(t)$ is given by the ratio of the output-to-input in the frequency domain. In a VAR model, the transfer function $H_{ij}(\omega)$ at frequency ω is given by:

$$H_{ij}(\omega) = \frac{S_{ij}(\omega)}{S_{jj}(\omega)},$$

where:

- $S_{ij}(\omega)$ is the cross-spectral density between $Y_i(t)$ and $Y_j(t)$,
- $S_{jj}(\omega)$ is the auto-spectral density of $Y_j(t)$.

The DTF quantifies the directed influence from variable $Y_j(t)$ to variable $Y_i(t)$ at frequency ω , reflecting how much of the variation in $Y_i(t)$ can be attributed to $Y_j(t)$ in the frequency domain. Mathematically, the Directed Transfer Function (DTF) is then defined as:

$$\text{DTF}_{ij}(\omega) = \frac{|H_{ij}(\omega)|^2}{\sum_{k=1}^n |H_{ik}(\omega)|^2},$$

where:

- $|H_{ij}(\omega)|^2$ is the squared magnitude of the transfer function from $Y_j(t)$ to $Y_i(t)$,

- The denominator $\sum_{k=1}^n |H_{ik}(\omega)|^2$ represents the total power that influences $Y_i(t)$ from all sources.

The DTF provides a normalized measure of directional influence, where the sum of influences from all variables to a specific variable $Y_i(t)$ equals 1 at each frequency ω . This normalization enables the DTF to quantify the relative contribution of each variable $Y_j(t)$ to $Y_i(t)$ within the multivariate system. Operating across the frequency spectrum, DTF values range from 0 to 1, where higher values indicate stronger directional influence at that particular frequency. A modification of the original method has been proposed in [56] to handle non-stationary signals.

2.2.2 Information Theoretic Metrics

Information theoretic approaches to coupling analysis offer a powerful framework for quantifying dependencies between system components without assuming underlying linear relationships or requiring specific model structures. These metrics, derived from Shannon’s information theory, capture both linear and nonlinear associations through the fundamental concept of uncertainty reduction. By measuring how knowledge of one variable reduces uncertainty about another, these approaches provide a more general characterization of coupling than their linear counterparts. The theoretical foundation of these metrics rests upon the estimation of probability distributions and their relationships, presenting both opportunities and challenges in practical applications. Their model-free nature makes them particularly valuable in scenarios where the underlying system dynamics are poorly understood or highly nonlinear. However, this generality comes with increased computational complexity and demands careful consideration of estimation methods, particularly in the presence of finite, noisy data. Some of the most commonly used information theoretic measures are described in following sections.

2.2.2.1 Mutual Information

Mutual Information (MI) represents a more comprehensive approach to quantifying neural-muscular interactions [51, 63]. Unlike traditional coherence-based methods, MI provides a robust framework for detecting both linear and non-linear relationships between cortical and muscular signals. By leveraging advanced statistical methods such as kernel density estimation and k-nearest neighbor approaches, MI can capture intricate dependencies that linear methods might overlook. This makes it particularly valuable in complex neurophysiological systems where interactions are rarely purely linear. The mathematical formulation of MI is rooted in information theory:

$$\text{MI}(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right), \quad (2.2)$$

where X and Y are random variables representing cortical and muscular signals, $p(x, y)$ is the joint probability distribution, and $p(x)$ and $p(y)$ are marginal probability distributions. Alternatively, MI can be expressed in terms of entropy as:

$$\text{MI}(X; Y) = H(X) + H(Y) - H(X, Y), \quad (2.3)$$

where $H(X)$ and $H(Y)$ are marginal entropies and $H(X, Y)$ is the joint entropy of X and Y . The versatility of MI becomes evident in its application to non-stationary signals and diverse signal types. However, this complexity comes with computational challenges, including increased computational complexity and sensitivity to estimation parameters.

2.2.2.2 Transfer Entropy

Transfer Entropy (TE) emerges as an information-theoretic measure that provides unprecedented insights into directional information flow between neural and muscular signals [38, 20]. TE goes beyond simple correlation measures to quantify the directed transfer of information between complex physiological systems.

Unlike traditional measures, TE can distinguish between TE is able to assess

asymmetric information transfer. This directional insight is crucial in understanding the nuanced feedback mechanisms underlying motor control. Insights into the causal mechanisms of neural-muscular interactions are gained by constructing conditional probability distributions and computing information transfer. The mathematical formulation is defined as:

$$\text{TE}_{X \rightarrow Y} = \sum_{x_{t+1}, x_t, y_t} p(x_{t+1}, x_t, y_t) \log \left(\frac{p(x_{t+1}|x_t, y_t)}{p(x_{t+1}|x_t)} \right), \quad (2.4)$$

where X and Y are time series representing cortical and muscular signals respectively. x_{t+1} is the future state of X whereas x_t and y_t are past states of X and Y respectively. $p(x_{t+1}|x_t, y_t)$ is the conditional probability of x_{t+1} given the past states.

Implementation of TE requires careful consideration of methodological nuances. The technique demands sufficient data length and is sensitive to embedding parameters. Moreover, the computational intensity of TE necessitates robust computational resources and advanced analytical approaches.

2.2.3 Advanced Coupling Metrics

Advanced coupling metrics extend beyond traditional linear and information theoretic approaches, offering methods for detecting and quantifying complex relationships in multivariate systems. These methods are particularly valuable when analyzing systems where conventional approaches may fail to capture subtle interactions or when the underlying relationships are nonlinear and dynamic in nature.

2.2.3.1 Partial Directed Coherence

Partial Directed Coherence (PDC) represents an extension of traditional spectral analysis methods for investigating directed influences in multivariate time series data. This frequency-domain approach, grounded in multivariate autoregressive modeling, offers distinct advantages in characterizing causal relationships within complex neurophysiological systems [125]. The mathematical formulation of PDC

incorporates both direct and indirect causal pathways, enabling more nuanced interpretation of corticomuscular interactions.

The theoretical foundation of PDC emerges from the parametric representation of multivariate processes through VAR modeling. For an n -variate time series $X(t)$, the VAR model of order m is expressed as:

$$X(t) = \sum_{i=1}^m A(i)X(t-i) + E(t)$$

where $A(i)$ represents the model coefficients and $E(t)$ denotes the innovation process. The frequency-domain representation of PDC is then derived through Fourier transformation of these autoregressive coefficients:

$$\pi_{ij}(f) = \frac{\bar{A}_{ij}(f)}{\sqrt{\sum_{k=1}^n |\bar{A}_{kj}(f)|^2}}$$

where $\bar{A}_{ij}(f)$ represents the frequency-domain coefficient matrix, and the denominator ensures proper normalization across frequencies. This formulation enables quantification of direct causal influences from channel j to channel i at frequency f , while accounting for the potential influences of all other channels in the system.

The methodological advantages of PDC in corticomuscular coupling analysis manifest through several key properties. First, its frequency-specific characterization allows identification of causal influences within physiologically relevant frequency bands, particularly crucial for distinguishing between different modes of motor control. Second, the normalization inherent in the PDC formulation facilitates comparison of causal strengths across different channel pairs and frequency ranges. Third, its foundation in VAR modeling provides a robust statistical framework for hypothesis testing and confidence interval estimation.

In practical applications, PDC computation requires careful consideration of model order selection, typically addressed through information criteria such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). The method's

sensitivity to common preprocessing choices, including filtering and reference selection, necessitates rigorous validation procedures. Additionally, the assumption of stationarity underlying VAR modeling must be carefully evaluated in the context of dynamic motor control processes.

Recent methodological extensions have enhanced the utility of PDC for corticomuscular coupling analysis [122]. These include generalized PDC (gPDC) [9] accounting for different scales across channels, and adaptive implementations capable of tracking time-varying causal relationships. Such advances have expanded its applicability to more complex experimental paradigms while maintaining its interpretability within established frameworks of spectral analysis.

2.2.3.2 Convergent Cross Mapping

Convergent Cross Mapping (CCM) is a method used for testing causal relationships between time series data [62, 128]. It is based on the principle of attractor reconstruction from time-delay embeddings, a concept rooted in dynamical systems theory. CCM is particularly useful in the context of nonlinear and deterministic systems. The key idea behind CCM is to determine whether the state of one variable can be used to predict the state of another, thereby implying a causal relationship.

Given a time series $Y(t) = \{Y(1), Y(2), \dots, Y(N)\}$, the method begins by reconstructing the state-space of the system using time-delay embeddings. The embedding process constructs a state vector $M_X(t), M_Y(t)$ at each time point t , which captures the state of the system at that time. The embedding vector $M_Y(t), M_X(t)$ is constructed as follows:

$$M_X(t) = (X(t), X(t - \tau), X(t - 2\tau), \dots, X(t - (E - 1)\tau)),$$

$$M_Y(t) = (Y(t), Y(t - \tau), Y(t - 2\tau), \dots, Y(t - (E - 1)\tau)).$$

where:

- $X(t), Y(t)$ are the observed values of the time series at time t ,

- τ is the time delay (a chosen integer),
- E is the embedding dimension (the number of delayed points to be included).

The goal of state-space reconstruction is to capture the underlying dynamical system that generates the observed time series. Once the state-space of the system is reconstructed, CCM tests for causal relationships by comparing the state-space of one variable with that of another. The method works by checking if the state of one time series, say $X(t)$, can be predicted using the reconstructed attractor of another time series $Y(t)$.

For each time point t , the reconstructed attractor M_Y of the time series Y is used to estimate the state of $X(t)$. This is done by cross-mapping, which is the process of predicting the state of $X(t)$ using the reconstructed manifold of Y .

The cross-mapping prediction $\hat{X}(M_Y)$ of the state of X at time t is based on the nearest neighbors in the reconstructed state-space of Y . Specifically, $\hat{X}(M_Y(t)), \hat{Y}(M_X(t))$ are calculated as follows:

$$\hat{X}(M_Y(t)) = \frac{1}{k} \sum_{i=1}^k X(t_i), \quad \hat{Y}(M_X(t)) = \frac{1}{k} \sum_{i=1}^k Y(t_i),$$

where t_i are the times corresponding to the nearest neighbors of $M_X(t)$ and $X(t_i)$ is the state of X at the i -th nearest neighbor of $M_Y(t)$, and k is the number of nearest neighbors chosen for the prediction. Similarly for $\hat{Y}(M_X(t))$.

CCM tests for causality by comparing the correlation between the predicted state $\hat{X}(M_Y(t)), \hat{Y}(M_X(t))$ and the actual state $X(t), Y(t)$. The strength of the cross-mapping is quantified by the Pearson correlation coefficient ρ , which measures how well the predicted state $\hat{X}(M_Y(t))$ matches the observed state $X(t)$:

$$\rho(M_Y) = \frac{\text{cov}(X(t), \hat{X}(M_Y(t)))}{\sigma_X \sigma_{\hat{X}}}, \quad \rho(M_X) = \frac{\text{cov}(Y(t), \hat{Y}(M_X(t)))}{\sigma_Y \sigma_{\hat{Y}}}.$$

A high correlation indicates that the reconstructed state of Y can effectively predict the state of X , suggesting a causal influence of Y on X . The process is

symmetrical, meaning that M_X can also be used to predict $Y(t)$, and the results from both directions should converge if there is a bidirectional causal relationship.

A significant correlation $\rho(M_Y)$ suggests that Y influences X , and a significant $\rho(M_X)$ suggests the reverse. Hence, CCM proves to be a powerful method for identifying causal relationships in nonlinear dynamical systems. By using state-space reconstruction and cross-mapping techniques, CCM allows us to assess the influence of one time series on another without the need for explicit parametric modeling, making it particularly useful for complex, chaotic systems. The method's ability to handle deterministic, nonlinear systems makes it an invaluable tool in causal inference when standard linear methods may not be applicable.

The methodological approaches examined thus far illuminate distinct yet complementary aspects of neural signal interdependencies through their respective mathematical frameworks. Spectral coherence provides insights into frequency-domain synchronization patterns through complex-valued cross-spectral density estimation, while mutual information quantifies statistical dependencies via information-theoretic principles without assuming linear relationships or specific probability distributions. Transfer entropy extends this information-theoretic foundation to capture asymmetric information flow between processes, offering a model-free characterization of directed interactions. Building upon these foundational approaches, the subsequent section introduces Granger causality, a statistical framework that operationalizes temporal precedence through vector autoregressive modeling and tests of predictive improvement in the presence of hypothesized causal variables. This methodology specifically examines whether past values of one time series contain information that helps predict another time series beyond its own history, providing a distinct perspective from the spectral, information-theoretic, and dynamical systems approaches discussed above.

2.3 Granger Causality Framework

Granger Causality considers the temporal precedence principle through predictive modeling frameworks, addressing a fundamental question in time series analysis: whether incorporating the historical dynamics of a causal signal enhances the prediction accuracy of target signal beyond what is achievable through the past trajectories of target itself. This formalization of causality through prediction has generated extensive theoretical developments in statistical inference, from its foundational linear autoregressive implementations to contemporary non-linear extensions incorporating kernel methods, transfer entropy equivalences, and state-space reconstructions.

In neurophysiological applications, particularly in cortico-muscular coupling analysis where signals exhibit complex non-stationarities and may operate at multiple temporal scales, rigorous examination of GC's theoretical assumptions and practical limitations becomes crucial. The ability of Granger Causality to quantify directional influences among multiple interacting variables has established it as a cornerstone in biological time series analysis. This framework naturally incorporates conditional dependencies, making it particularly suitable for complex physiological systems. However, the unique characteristics of biological data, including non-stationarity and noisy interferences, have driven continuous methodological refinements to the original framework. Linear Granger causality is mathematically formulated through autoregressive models. Consider two time series X and Y and consider $y(t)$ is the EEG signal and $x(t)$ is the corresponding sEMG signal at time t , assuming the observed data does not involve measurement noise. They can be represented as:

$$y(t) = \sum_{j=1}^m a_{yy}^{(j)} y(t-j) + \sum_{j=1}^m a_{yx}^{(j)} x(t-j) + \epsilon_y(t), \quad (2.5)$$

$$x(t) = \sum_{j=1}^m a_{xy}^{(j)} y(t-j) + \sum_{j=1}^m a_{xx}^{(j)} x(t-j) + \epsilon_x(t), \quad (2.6)$$

where $\epsilon_y(t)$ and $\epsilon_x(t)$ are time-dependent innovation processes and m indicates the dependency on past instants or order of the autoregressive model as outlined in 2.1.

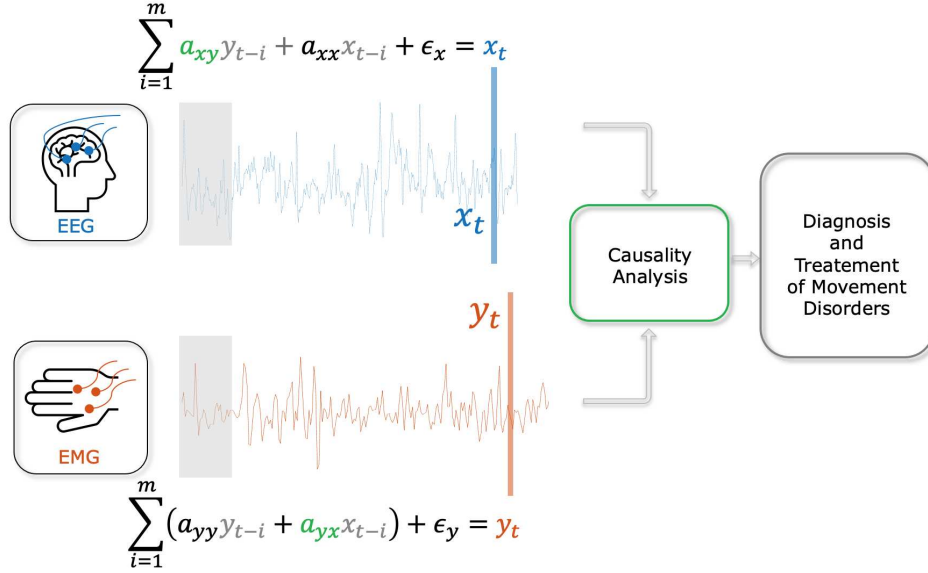


Figure 2.1: Overview of cortico-muscular coupling estimation

The objective is to fit an autoregressive model to estimate parameters $a_{yy}^{(j)}, a_{yx}^{(j)}, a_{xy}^{(j)}, a_{xx}^{(j)} \forall j = 1, \dots, m$ where m is predetermined. After estimating the parameters, Granger causality can be estimated using an F-test [110] comparing the *restricted model* (without cross-coupling i.e. $a_{yx} = a_{xy} = 0$) and *unrestricted model* (with cross-coupling i.e. $a_{yx} \neq 0, a_{xy} \neq 0$) as:

$$GC_{x \rightarrow y} = \frac{RSS_{y_{\text{res}}} - RSS_{y_{\text{unr}}}/(p - p')}{RSS_{y_{\text{unr}}}/(T - p)}, \quad (2.7)$$

$$GC_{y \rightarrow x} = \frac{RSS_{x_{\text{res}}} - RSS_{x_{\text{unr}}}/(p - p')}{RSS_{x_{\text{unr}}}/(T - p)}, \quad (2.8)$$

where RSS means *residual sum of squares*, T is the length of the process, p and p' are the number of parameters in unrestricted and restricted models, respectively. This test shows that if the observed statistic F exceeds a certain quantile, say 95th, of an F -distribution with $p - p'$ and $T - p$ degrees of freedom then Granger causality exists.

It is worth noting that under the Gaussian assumption, transfer entropy is equivalent to Granger causality. The linear GC method's strength lies in its computational efficiency and clear statistical interpretation. However, the linear approach inherently limits the complexity of neural-muscular interactions.

Table 2.2: Assumptions for Granger causal Analysis

Assumption	Description
Known lag	The dependency on lagged observations is assumed to have a known and uniform order across series.
Stationarity	The statistics of the process are assumed to be time-invariant
Perfectly observed	The variables need to be observed without measurement errors.
Linearity	The data-generating process and causal effects are assumed to be linear
Complete system	All relevant variables are assumed to be included, with no unmeasured confounders affecting the analysis.

2.3.1 Key Assumptions

The practical applicability of Granger causality is bounded by several fundamental assumptions, as outline in [110] and summarized in Table 2.2. These assumptions warrant careful consideration in physiological applications:

In the context of autoregressive modeling, the linearity assumption implies that the future values of a time series can be expressed as a linear combination of its own past values and those of potential causal variables. While this assumption simplifies the mathematical framework and enables efficient estimation, it may not fully capture the complex, non-linear dynamics often present in physiological systems. This linear framework is intrinsically linked to the known lag assumption, which requires a priori specification of the temporal order of dependencies. This presupposes that causal influences operate within a fixed time window and that the temporal structure of these influences remains consistent across the analyzed period. In physiological systems, where interactions may occur across multiple time scales or with variable delays, this assumption can be particularly restrictive.

Beyond these structural constraints, the framework relies heavily on stationarity, which demands that the statistical properties of the time series remain constant over time. Specifically, this requires that the mean, variance, and auto-covariance

structure do not change throughout the observation period. Physiological signals, however, often exhibit non-stationary behavior due to adaptation, fatigue, or changing experimental conditions.

The practical implementation of these theoretical requirements further depends on the perfect observation assumption, which stipulates that variables are measured without error or noise contamination. This is particularly challenging in biological recordings where measurement noise is inevitable and may be substantial. The presence of measurement error can lead to biased estimates of causal relationships and potentially spurious conclusions.

Perhaps most crucially from a systems perspective, the complete system assumption requires that all relevant variables are included in the analysis. Unobserved confounding variables that influence multiple observed variables can lead to spurious causal relationships or mask true causal effects. In complex physiological systems, where numerous interacting components may be unmeasured or unmeasurable, this assumption is often violated.

These assumptions constitute the foundational framework of classical Granger causality analysis, where violations can significantly impact the validity of causal inferences. The severity of assumption violations in physiological data necessitates methodological adaptations for meaningful causal analysis. While various extensions to the basic framework have been proposed in literature, they often address these limitations in isolation.

This thesis presents an integrated approach to address multiple GC assumptions simultaneously in the context of physiological signal analysis. We develop novel methodologies for robust model order selection in non-stationary conditions, accounting for measurement noise, and handling non-linear dependencies. Our framework particularly focuses on the challenges of cortico-muscular coupling, where traditional GC assumptions are frequently violated due to inherent system complexity. The proposed extensions relax these restrictive assumptions while maintaining the theoretical rigor of the original framework, enabling reliable causal inference in real-world

physiological applications.

2.3.2 Non-linear Granger Causality

Recognizing the limitations of linear methods, non-linear Granger Causality emerges as an extension of traditional linear approaches [73, 115]. The transition from linear to non-linear frameworks represents a crucial methodological advancement in capturing the inherent complexity of neurophysiological interactions. The classical linear framework relies on several critical assumptions that may not hold in complex neurophysiological systems. First, the assumption of linearity constrains the model to detect only linear relationships between variables, potentially missing important non-linear interactions that are prevalent in neural dynamics. Second, the statistical inference in linear Granger causality typically employs F-tests or likelihood ratio tests, which rely on assumptions of normally distributed errors and homoscedasticity that may be violated in practice.

To address these limitations, more flexible frameworks have been developed in recent years that can accommodate non-linear relationships while maintaining the fundamental concept of predictive causality. This approach employs advanced techniques such as kernel-based methods [74], mutual nearest neighbor approaches [129], and nonparametric regression models [84] to capture more complex interactions. A general formulation extends the classical linear framework by introducing non-linear functional relationships:

$$\begin{aligned} Y(t) &= f(Y(t-1), \dots, Y(t-m)) + \varepsilon_1(t) \\ Y(t) &= g(Y(t-1), \dots, Y(t-m), X(t-1), \dots, X(t-m)) + \varepsilon_2(t) \end{aligned} \tag{2.9}$$

where f and g are unknown non-linear functions, m is the model order, and $\varepsilon_1(t)$ and $\varepsilon_2(t)$ represent the prediction errors for the restricted and full models respectively.

While this theoretical formulation provides a foundation for non-linear analysis, implementing such models presents substantial computational and methodological

challenges. A more practical implementation utilizes kernel methods, which implicitly map the input space to a high-dimensional feature space where complex non-linear relationships can be captured through linear operations. This approach, known as the kernel trick, leads to the following representation using a commonly employed Gaussian kernel:

$$Y(t) = \sum_{i=1}^n \alpha_i \exp \left(-\frac{\|\mathbf{Z}_i - \mathbf{Z}_j\|^2}{2\sigma^2} \right) + \varepsilon(t) \quad (2.10)$$

where σ is the kernel bandwidth parameter controlling the degree of non-linearity, $\mathbf{Z}_i$ represents the delay embedding vectors (constructed as $[Y(t-1), \dots, Y(t-m)]$ for the restricted model and $[Y(t-1), \dots, Y(t-m), X(t-1), \dots, X(t-m)]$ for the full model), α_i are the weights to be estimated, and n is the number of observations.

The transition to non-linear methods necessitates a fundamental reconsideration of statistical inference procedures. The assessment of causality in this non-linear framework requires alternative statistical approaches, as traditional F-tests are no longer applicable due to the non-linear transformations and potential non-Gaussian error distributions. Researchers typically employ permutation tests or bootstrapping procedures to evaluate the significance of causal relationships, comparing prediction errors between restricted and full models in the kernel-induced feature space. These non-parametric approaches offer greater flexibility in capturing complex dependencies while maintaining statistical rigor.

The evolution of non-linear Granger causality continues with recent methodological innovations in machine learning. Recent advances have further extended the framework through neural network architectures [115], which can automatically learn appropriate non-linear transformations from data. These approaches leverage deep learning's capacity for representation learning while maintaining interpretability through carefully designed architecture constraints and regularization schemes. The integration of modern machine learning techniques with traditional causal inference frameworks represents a promising direction for capturing increasingly complex causal relationships in neurophysiological data.

Despite these methodological advances, significant challenges remain in the prac-

tical application of non-linear Granger causality. These include selecting appropriate model complexity to balance flexibility with overfitting, determining optimal kernel parameters for kernel-based methods, and ensuring computational tractability for high-dimensional time series data. Moreover, interpreting non-linear causal relationships demands careful attention, as these relationships are often more complex and less intuitive than their linear counterparts. As a result, recent efforts have concentrated on developing more robust model selection criteria, enhancing computational efficiency, and establishing clearer frameworks for understanding non-linear causal relationships within neurophysiological contexts.

Considering these advancements, the validation of our proposed methodological framework encompasses both synthetic and empirical physiological datasets. This would enable a comprehensive evaluation of causal inference methodologies in various experimental settings. In the next section, we will introduce the experimental paradigm for this thesis and describe various types of datasets used for experimental validation.

2.4 Experimental Paradigm

The validation framework of our proposed methods employs both synthetic and physiological datasets to rigorously evaluate the proposed causal inference methodology. The synthetic validation utilizes the Lorenz96 dynamical system, which provides analytically tractable ground truth causality while incorporating key characteristics relevant to physiological systems - non-linear interactions and multi-scale temporal dependencies. This controlled environment enables quantitative assessment of causal structure recovery under varying noise conditions.

The empirical validation examines cortico-muscular datasets that capture the inherent complexity of biological systems, including non-stationary dynamics and measurement noise. These data present practical challenges in disambiguating direct and indirect causal pathways within densely connected physiological networks. The

combination of synthetic and empirical validation enables the evaluation of both methodological robustness and clinical utility, establishing reliability bounds for real-world applications in physiological signal analysis.

2.4.1 Healthy Adults Motor Perturbation Dataset (HAMPD)

The investigation recruited fourteen neurotypical subjects (nine female, five male; age range: 24-62 years), all exhibiting right-hand dominance through self-reporting. The experimental protocol comprised three distinct paradigms, following established methodologies in motor control research [78].

The electrical stimulus investigation (n=5) utilized cutaneous afferent stimulation of the digital nerve of the index finger, with pulse width maintained at 200 μ s and intensity set to 1.5 times motor threshold. The interstimulus interval was pseudorandomized at 7 ± 1.4 s, with 200 epochs recorded per subject. Mechanical perturbation analysis (n=12) implemented precise displacement parameters, achieving 1 mm lateral movement at 0.2 m/s velocity, with 5 ms rise time and 20 ms duration. The non-dominant hand electrical stimulus evaluation (n=5) replicated the primary electrical stimulus protocol with contralateral application for hemispheric comparison.

Electrophysiological data acquisition employed a comprehensive recording paradigm. High-density EEG recordings were obtained using a 64-channel system (sampling rate: 2048 Hz, bandwidth: DC-500 Hz) with linked mastoid reference. Surface EMG recordings targeted the first dorsal interosseous (FDI) and flexor pollicis brevis (FPB) muscles using a belly-tendon montage configuration [80] as in Figure 2.2. The EMG signal chain maintained a bandwidth of 20-500 Hz with a sampling rate of 2048 Hz and common mode rejection ratio exceeding 100 dB.

2.4.2 Pediatric Motor Perturbation Dataset (PMPD)

The pediatric investigation encompassed sixteen children presenting with dystonia (eleven acquired, five idiopathic/genetic; nine female; age range: 12-18 years) recruited through the Complex Motor Disorders Service at Evelina London Children’s Hospital,

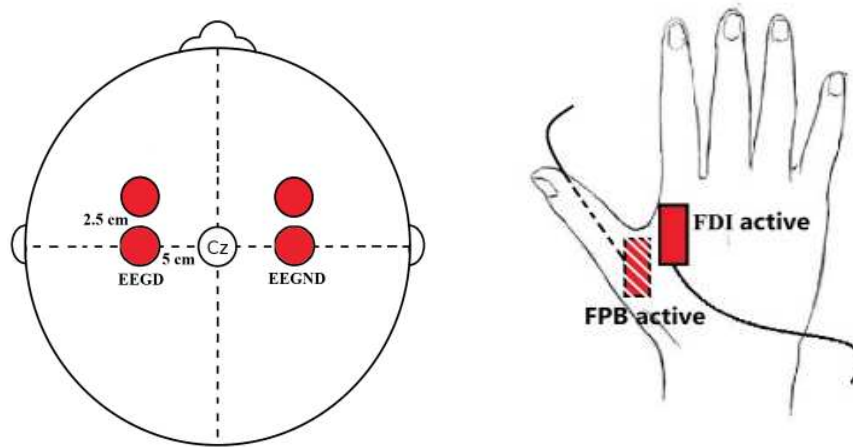


Figure 2.2: Data collection during controlled motor task

alongside thirteen neurotypical controls (seven female). Dystonia diagnosis and classification underwent verification by a consultant pediatric neurologist specializing in movement disorders, implementing standardized assessment protocols.

The Burke-Fahn-Marsden Dystonia Rating Scale (BFMDRS-m) provided quantitative assessment of dystonia severity, administered by specialized pediatric therapists blind to neurophysiological data. Motor task execution replicated the adult protocol, with grip force standardized to 5-15% of maximal voluntary contraction (MVC) and continuous position monitoring via three-dimensional motion capture.

Mechanical perturbations were delivered through an electromechanical tapper system, generating a Gaussian-approximated force profile with 2 N peak force over 1 cm² contact area. The experimental structure implemented blocks of 10-25 trials with 120-second inter-block intervals, accumulating up to 200 epochs per subject as permitted by individual capability.

During this task, the muscle pair consisting of the first dorsal interosseous (FDI) and flexor pollicis brevis (FPB) was predominantly activated. EMG signals were recorded using self-adhesive electrodes in a belly-tendon configuration over the FDI and FPB of the dominant hand (Fig. 3.1a). Additionally, two bipolar EEG recordings were obtained: one from the dominant motor cortex (EEGD) and the other from the non-dominant motor cortex (EEGND). These EEG signals were recorded from the

scalp overlying the bilateral motor cortex, with two electrodes placed 5 cm lateral to the vertex along the interaural line and another two electrodes positioned 2.5 cm anterior to the first pair (Fig. 3.1b). Data were collected at a sampling frequency of 1024 Hz, with EMG and EEG signals amplified and band-pass filtered (0.5-100 Hz for EEG; 5-500 Hz for EMG). Raw data were reviewed offline through visual inspection, and epochs containing movement or blink artifacts were excluded [78]. In this thesis, the primary focus was on measuring cortico-muscular coupling between the EEG from the dominant motor cortex (EEGD) and the EMG from the FDI, unless stated otherwise. In the original neurophysiology study [78], coherence was relatively strong in some subjects but very weak in others.

2.4.3 Lorenz96 Dynamical System Implementation

The Lorenz96 model, originally conceptualized for atmospheric predictability studies [66], serves as our synthetic validation framework due to its ability to generate complex, non-linear interactions with known ground truth causal relationships. This dynamical system implements $n = 40$ coupled variables evolving according to the differential equation:

$$\frac{dX_i}{dt} = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F$$

where X_i is the value of the i -th variable (representing, e.g., atmospheric quantities at grid point i), F is an external forcing parameter (e.g., representing constant energy input) set to 8 to ensure chaotic behavior that mimics the complexity observed in physiological systems. The system is usually defined with periodic boundary conditions i.e. $X_{i+n} = X_i$, where n is the total number of variables. The distinctive feature of model lies in its ring topology, where each variable exhibits direct causal influences on specific neighboring variables through multiplicative coupling terms, creating a well-defined network of causality.

The numerical implementation employs a fourth-order Runge-Kutta integration

scheme with a step size of 0.01 over 1000 time units, ensuring stable and accurate trajectory computation. This fine temporal resolution captures the system’s intricate dynamics while providing sufficient data density for robust causal inference. The resulting trajectories exhibit key characteristics relevant to physiological time series, including non-linear state transitions, multi-scale temporal dependencies, and complex feedback mechanisms.

To enhance the physiological relevance of our synthetic validation, we incorporate measurement noise and perturbations into the generated time series. This modification allows us to evaluate robustness of our causal inference methodology under conditions that more closely approximate real-world physiological recordings. The known causal structure of the Lorenz96 system provides an unambiguous ground truth for quantifying the accuracy and reliability of our proposed methods across varying noise levels and coupling strengths.

These three complementary datasets form a comprehensive validation framework for evaluating our causal inference methodology across increasing levels of physiological complexity. The Lorenz96 system provides analytical ground truth for initial validation. The cortico-muscular datasets test our methods’ ability to distinguish neural-muscular causal pathways, while the motor control recordings extend validation to clinically relevant scenarios, where voluntary movement and feedback mechanisms create complex causal networks. This hierarchical approach assesses both the robustness and clinical applicability of our methodology, addressing the challenges of transitioning from synthetic to physiological data. Cross-comparisons across these paradigms establish reliability metrics and define the operational boundaries of our framework, providing a foundation for its application in physiological signal analysis and clinical assessment.

Chapter 3

Linear Autoregressive Models: Enhancing Model Integrity

This chapter advances existing frameworks for addressing fundamental limitations in classical Granger causal inference, with particular emphasis on the restrictive assumptions underlying Vector Autoregressive (VAR) models. While linear VAR models offer analytical tractability and interpretability in capturing temporal dependencies, their underlying assumptions often constrain their applicability in complex physiological systems. Nevertheless, their proven efficacy in approximating many real-world relationships and their transparent interpretation make them an invaluable foundation for algorithmic development.

We examine and develop solutions for three critical constraints: the requirement for known and fixed lag structures, the assumption of strict stationarity in time series processes, and the premise of perfect observability in measurement systems. Through careful theoretical development and empirical validation, we establish more flexible modeling approaches that relax these assumptions while maintaining analytical rigor.

The progression begins with foundational extensions to conventional VAR frameworks, introducing adaptive techniques for lag structure identification, robust handling of non-stationary dynamics, and explicit incorporation of measurement uncertainty. Building upon these fundamental advances, we then expand our analytical

framework to encompass more complex model architectures, including non-linear dynamics and confounding variable analysis.

The chapter concludes with comprehensive experimental validation of our proposed methods, employing both synthetic and physiological datasets to assess performance improvements over classical approaches, providing quantitative evidence for the practical utility of our methodological advances in real-world applications.

To frame these advancements within a broader context, it is essential to consider the foundational models and concepts that have shaped causal inference methodologies. Autoregressive models have commonly been employed to capture the relationships between different processes and discern directional causality [46],[70],[34]. The notion of causality, as applied to autoregressive models, was originally introduced by Granger [36] where the fundamental concept is to quantify the degree to which knowledge about one process reduces uncertainty associated with another.

In autoregressive modeling, the sEMG signal can be modelled as a linear combination of lagged values of relevant EEG control activity, previous muscle activity, and excitation noise from a random innovation process. The EEG signal can also be expressed in a similar way with sensory feedback proportionate to sEMG activity sent back to the cortex.

Let $y(t)$ be the EEG signal and $x(t)$ be the corresponding sEMG signal at time t . They can be represented as:

$$y(t) = \sum_{k=1}^m a_{yy}^{(k)} y(t-k) + \sum_{k=1}^m a_{yx}^{(k)} x(t-k) + \epsilon_y(t), \quad (3.1)$$

$$x(t) = \sum_{k=1}^m a_{xy}^{(k)} y(t-k) + \sum_{k=1}^m a_{xx}^{(k)} x(t-k) + \epsilon_x(t), \quad (3.2)$$

where $\epsilon_y(t)$ and $\epsilon_x(t)$ are time-dependent excitation noise signals and m indicates the dependency on past instants or order of the autoregressive model. The objective is to fit an autoregressive model to estimate parameters $a_{yy}^{(k)}, a_{yx}^{(k)}, a_{xy}^{(k)}, a_{xx}^{(k)} \forall k = 1, \dots, m$ where m is known a priori. However, in practice, determining the order of the autoregressive model is a non-trivial problem.

The system of equations in (3.1-3.2) can be compactly represented as follows:

$$\begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} a_{yy}^{(1)} \cdots a_{yy}^{(m)} & a_{yx}^{(1)} \cdots a_{yx}^{(m)} \\ a_{xy}^{(1)} \cdots a_{xy}^{(m)} & a_{xx}^{(1)} \cdots a_{xx}^{(m)} \end{bmatrix} \begin{bmatrix} y(t-1) \\ \vdots \\ y(t-m) \\ x(t-1) \\ \vdots \\ x(t-m) \end{bmatrix} + \begin{bmatrix} \epsilon_y(t) \\ \epsilon_x(t) \end{bmatrix}, \quad (3.3)$$

$$\mathbf{Y} = \mathbf{A}\mathbf{H} + \mathbf{E}, \quad (3.4)$$

where the matrices in Equation (3.3) are assigned to corresponding variables in Equation (3.4) for $t = m + 1, \dots, m + T$ where T is the length of the process. Hence, $\mathbf{Y} \in \mathbb{R}^{2 \times T}$, $\mathbf{A} \in \mathbb{R}^{2 \times 2m}$, $\mathbf{H} \in \mathbb{R}^{2m \times T}$ and $\mathbf{E} \in \mathbb{R}^{2 \times T}$. The model in Equation (3.1) is termed as *unrestricted* model. On the other hand, a *restricted* model would be the one that does not involve any cross-coupling between $x(t)$ and $y(t)$ i.e. $a_{xy}^{(i)} = a_{yx}^{(i)} = 0$, $i = 1, \dots, m$.

Mathematically, the *restricted* model amounts to the one below:

$$\begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} a_{yy}^{(1)'} \cdots a_{yy}^{(m)'} & 0 \cdots \cdots \cdots 0 \\ 0 \cdots \cdots \cdots 0 & a_{xx}^{(1)'} \cdots a_{xx}^{(m)'} \end{bmatrix} \begin{bmatrix} y(t-1) \\ \vdots \\ y(t-m) \\ x(t-1) \\ \vdots \\ x(t-m) \end{bmatrix} + \begin{bmatrix} \epsilon'_y(t) \\ \epsilon'_x(t) \end{bmatrix}, \quad (3.5)$$

$$\mathbf{Y} = \mathbf{A}'\mathbf{H} + \mathbf{E}', \quad (3.6)$$

where the matrices in Equation (3.5) are assigned to corresponding variables in Equation (3.6) and $\mathbf{A}' \in \mathbb{R}^{2 \times 2m}$ and $\mathbf{E}' \in \mathbb{R}^{2 \times T}$ are the coefficients and noise matrices for restricted model, respectively. Granger causality can be tested using an

F-test [110] comparing the two models as:

$$GC_{x \rightarrow y} = \frac{RSS_{y_{res}} - RSS_{y_{unr}} / (p - p')}{RSS_{y_{unr}} / (T - p)}, \quad (3.7)$$

$$GC_{y \rightarrow x} = \frac{RSS_{x_{res}} - RSS_{x_{unr}} / (p - p')}{RSS_{x_{unr}} / (T - p)}, \quad (3.8)$$

where RSS means *residual sum of squares*, p and p' are the number of parameters in unrestricted and restricted models, respectively. This metric, used to determine Granger causality, quantifies the impact of cross-coupling terms in prediction of a variable.

3.1 Model Order Identification by Hierarchical Sparsity

Model order selection represents a fundamental challenge in causal inference, particularly within the framework of Granger causality analysis. An appropriate model order, that quantifies the relevant process history for prediction, is assumed to be known a priori for Granger causality analysis [110]. This assumption, while theoretically convenient, often proves problematic in practical applications where the temporal dependencies are complex and potentially unknown.

Numerous techniques have been put forward for determining the appropriate order of the autoregressive model, reflecting the critical nature of this selection process in causal inference. The optimal model order should be sufficiently small to guarantee that predictions depend primarily on direct observations rather than interim estimations, yet large enough to encompass the inherent trend within the underlying process. This fundamental tension between model parsimony and explanatory power presents a significant methodological challenge. When Granger causality is estimated using a small model order, it tends to limit the frequency resolution of physiological signals, making it challenging to distinguish effectively between different frequency ranges [32].

Traditionally, model identification is performed through visual assessment of the autocorrelation function or by minimizing loss functions like negative log-likelihood or least squares loss. Popular methods like the AIC and BIC have followed this approach. However, these methods may struggle to accurately infer the true data distribution [109] and can potentially overfit in practical scenarios [23], particularly when dealing with high-dimensional time series data where the underlying causal structure is complex.

Recent methodological advances have sought to address these limitations. More recently, a method for simultaneous model order identification and parameter estimation for ARMA (Auto-Regressive Moving Average) models is proposed [64]. This approach employs a sparsity-based regularization technique to estimate the order and leverages a Block Coordinate Descent based method for parameter estimation. This regularization has proven effective in determining the model order by progressively curtailing the irrelevant history of the process in a hierarchical manner, however, the method does not guarantee convergence to a global optimal point, highlighting the need for further methodological refinement.

Building upon these developments, we introduce a novel way to augment stationarity constraint with model identification by imposing sparsity on autoregressive coefficients in a hierarchical manner. This approach synthesizes the benefits of sparse estimation with the theoretical guarantees provided by stationarity constraints. We assume prior knowledge of the maximum allowable model order denoted as $\bar{m}$ i.e. $m < \bar{m}$. The hierarchical sparsity pattern is created by incorporating the Latent Overlapping Group (LOG) lasso penalty, as detailed in [64, 50], into the optimization objective function. The LOG penalty function is defined as:

$$\Omega_{\text{LOG}}(\mathbf{c}) = \min_{\mathbf{l}^{(g)}, g \in \mathcal{G}} \left\{ \sum_{g \in \mathcal{G}} w_g \|\mathbf{l}^{(g)}\|_2 \mid \sum_{g \in \mathcal{G}} \mathbf{l}^{(g)} = \mathbf{c}, \mathbf{l}_{g^c}^{(g)} = 0 \right\},$$

where $\mathcal{G} = \{\{1\}, \dots, \{1, \dots, \bar{m}\}, \{\bar{m}+1\}, \dots, \{\bar{m}+1, \dots, 2\bar{m}\}, \{2\bar{m}+1\}, \dots, \{2\bar{m}+1, \dots, 3\bar{m}\}, \{3\bar{m}+1\}, \dots, \{3\bar{m}+1, \dots, 4\bar{m}\}\}$ is the set of all groups, $\mathbf{l}^{(g)} \in \mathbb{R}^{4\bar{m} \times 1}$ is a latent vector indexed

by g , and w_g is the weight for set g . The key concept is that the sparsity-inducing penalty maintains the hierarchical structure of non-zero parameters. For instance, it avoids setting the first AR parameter to zero while the second AR parameter remains non-zero. As a result of solving $\Omega_{\text{LOG}}(\mathbf{c})$ where $\mathbf{c}$ represents the parameter vector $\mathbf{a}$, the model orders for $a_{yy}, a_{yx}, a_{xy}, a_{xx}$ will be chosen adaptively rather than a fixed order. We propose to couple this hierarchical-sparsity based penalty with the stationarity requirement to ensure stability of the model.

3.2 Proximal Operator for Stationarity Condition

A critical consideration in vector autoregressive modeling is ensuring the stability of the estimated system through appropriate stationarity constraints. Note that the bivariate VAR model in Equation (3.3) can also be written as follows:

$$\boldsymbol{\xi}_t = \mathbf{A}_1 \boldsymbol{\xi}_{t-1} + \mathbf{A}_2 \boldsymbol{\xi}_{t-2} + \cdots + \mathbf{A}_{\bar{m}} \boldsymbol{\xi}_{t-\bar{m}} + \mathbf{n}_t,$$

where $\boldsymbol{\xi}_t := \begin{bmatrix} y(t) \\ x(t) \end{bmatrix}$, $\mathbf{A}_i := \begin{bmatrix} a_{yy}^{(i)} & a_{yx}^{(i)} \\ a_{xy}^{(i)} & a_{xx}^{(i)} \end{bmatrix}$ and $\mathbf{n}_t := \begin{bmatrix} n_y(t) \\ n_x(t) \end{bmatrix}$. The $\bar{m}$ -th order process can be written as a first order VAR process VAR(1) by stacking the variables as:

$$\begin{bmatrix} \xi_t \\ \xi_{t-1} \\ \xi_{t-2} \\ \vdots \\ \xi_{t-\bar{m}+1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{\bar{m}-1} & \mathbf{A}_{\bar{m}} \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \xi_{t-1} \\ \xi_{t-2} \\ \xi_{t-3} \\ \vdots \\ \xi_{t-\bar{m}} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_t \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix},$$

$$\boldsymbol{\xi}_t = \boldsymbol{\Gamma}(\mathbf{A}) \boldsymbol{\xi}_{t-1} + \mathbf{v}_t,$$

where $\mathbf{\Gamma}(\cdot) : \mathbb{R}^{2 \times 2\bar{m}} \rightarrow \mathbb{R}^{2\bar{m} \times 2\bar{m}}$ is the transformation to extract the companion matrix corresponding to $\mathbf{A}$. For the VAR model to be stationary, the eigenvalues of $\mathbf{\Gamma}(\mathbf{A})$ must lie inside the unit circle [110]. Our proposed approach incorporates a spectral norm-based regularization technique aimed at constraining the eigenvalues of the companion matrix linked to the estimated parameters to reside within the unit circle. This methodology provides a mathematically rigorous framework for ensuring model stability while maintaining computational tractability.

The spectral norm of a matrix, which serves as our primary mathematical tool for enforcing stationarity, is defined as its largest singular value:

$$\|\mathbf{\Gamma}(\mathbf{A})\|_2 := \max_i |\sigma_i|.$$

This definition naturally leads to the fundamental condition for stationarity: $\|\mathbf{\Gamma}(\mathbf{A})\|_2 < 1$. To operationalize this constraint within our optimization framework, we introduce the proximal operator $\Psi_{\text{SP}}(\mathbf{A})$ for the spectral norm, which can be formulated as:

$$\Psi_{\text{SP}}(\mathbf{A}) := \arg \min_{\mathbf{A}} \gamma \|\mathbf{\Gamma}(\mathbf{A})\|_2 + \frac{\rho}{2} \|\mathbf{X} - \mathbf{\Gamma}(\mathbf{A})\|_F^2.$$

This optimization problem, at its core, represents the computation of the proximal operator for the ℓ_∞ norm. Through singular value decomposition, we can express the solution in a more computationally amenable form. Specifically, if we denote $\mathbf{X} = \mathbf{U} \text{diag}(\boldsymbol{\sigma}) \mathbf{V}^T$, then $\Psi_{\text{SP}}(\mathbf{A})$ can be estimated as:

$$\Psi_{\text{SP}}(\mathbf{A}) \approx \mathbf{\Gamma}^{-1}(\mathbf{U} \text{diag}(\mathbf{v}^*) \mathbf{V}^T),$$

where $\mathbf{\Gamma}^{-1}(\cdot) : \mathbb{R}^{2\bar{m} \times 2\bar{m}} \rightarrow \mathbb{R}^{2 \times 2\bar{m}}$ represents the inverse mapping of $\mathbf{\Gamma}$. The optimal solution vector $\mathbf{v}^*$ is obtained through:

$$\begin{aligned} \mathbf{v}^* &= \arg \min_{\mathbf{v}} \gamma/\rho \|\mathbf{v}\|_\infty + \frac{1}{2} \|\mathbf{v} - \boldsymbol{\sigma}\|_2^2, \\ \Pi_{\gamma/\rho}(\mathbf{X}) &:= \mathbf{\Gamma}^{-1}(\mathbf{U} \text{diag}(\mathbf{v}^*) \mathbf{V}^T), \end{aligned}$$

For notational convenience, these operations are consolidated within the operator $\Pi(\cdot)$. The projection operator $\mathcal{P}_{\|\cdot\|_1 \leq 1}$, which ensures projection onto the unit ℓ_1 norm ball, can be efficiently computed using non-iterative techniques as detailed in [127]. This computational efficiency is particularly valuable when dealing with high-dimensional time series data, where iterative approaches might prove prohibitively expensive.

The elegance of this approach lies in its ability to simultaneously enforce stationarity constraints while maintaining computational feasibility, thereby providing a practical solution to the fundamental challenge of ensuring model stability in vector autoregressive systems.

3.3 Globally Convergent Method for Model Order and Stationarity

Building upon our framework for hierarchical sparsity and stationarity constraints, we now formulate a comprehensive convex optimization program that synthesizes these methodological elements. In addition to reconstruction error, the optimization program will include the Ω_{LOG} and Ψ_{SP} terms to enforce sparsity and stationarity respectively, yielding a mathematically rigorous framework for stable parameter estimation:

$$\begin{aligned} & \min \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{H}\|_F^2 + \lambda \Omega_{\text{LOG}}(\mathbf{c}) + \gamma \|\mathbf{\Gamma}(\mathbf{Z})\|_2 \\ & + \frac{1}{2} \|\mathbf{Y}' - \mathbf{A}'\mathbf{H}'\|_F^2 + \lambda' \Omega_{\text{LOG}}(\mathbf{c}') + \gamma' \|\mathbf{\Gamma}(\mathbf{Z}')\|_2 \\ & \text{s.t. } \mathbf{c} = \text{vec}(\mathbf{A}^T), \quad \mathbf{c}' = \text{vec}(\mathbf{A}'^T), \quad \mathbf{A} = \mathbf{Z}, \quad \mathbf{A}' = \mathbf{Z}'. \end{aligned} \quad (3.9)$$

$$\begin{aligned} & \min \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{H}\|_F^2 + \lambda \Omega_{\text{LOG}}(\mathbf{c}) + \gamma \Psi_{\text{SP}}(\mathbf{Z}) \\ & \text{s.t. } \mathbf{c} = \text{vec}(\mathbf{A}^T), \quad \mathbf{A} = \mathbf{Z}. \end{aligned} \quad (3.10)$$

In contrast to the method proposed in [64], which employs an iterative minimization approach to achieve a Euclidean projection within a feasible set, we introduce specific constraints on the companion matrix of the model parameters to promote stable estimates by penalizing unstable estimates in the optimization program (3.9). By expanding $\Omega_{\text{LOG}}(\mathbf{c})$ in (3.9) we get:

$$\begin{aligned}
 & \min \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{H}\|_F^2 + \frac{1}{2} \|\mathbf{Y}' - \mathbf{A}'\mathbf{H}'\|_F^2 + \gamma \|\mathbf{\Gamma}(\mathbf{Z})\|_2 \\
 & + \gamma' \|\mathbf{\Gamma}(\mathbf{Z}')\|_2 + \lambda \sum_{g \in \mathcal{G}} w_g \|\mathbf{P}_{.g}\|_2 + \frac{1}{2} \left\| \sum_{g \in \mathcal{G}} \mathbf{Q}_{.g} - \mathbf{c} \right\|_2^2 \\
 & + \lambda' \sum_{g \in \mathcal{G}} w_g \|\mathbf{P}'_{.g}\|_2 + \frac{1}{2} \left\| \sum_{g \in \mathcal{G}} \mathbf{Q}'_{.g} - \mathbf{c}' \right\|_2^2 \\
 & \text{s.t.} \\
 & \mathbf{c} = \text{vec}(\mathbf{A}^T), \mathbf{P} = \mathbf{Q}, \mathbf{c}' = \text{vec}(\mathbf{A}'^T), \mathbf{P}' = \mathbf{Q}', \mathbf{A} = \mathbf{Z}, \\
 & \mathbf{A}' = \mathbf{Z}', (\mathbf{P}_{.g})_{g^c} = 0, a_{yx}^{(j)'} = a_{xy}^{(j)'} = 0 \quad \forall j = 1, \dots, \bar{m}.
 \end{aligned}$$

The update for parameter matrix $\mathbf{A}$ can be computed by finding the gradient of objective with respect to $\mathbf{A}$ as follows:

$$\mathbf{A} = [\rho_3 \tilde{\mathbf{Z}} + \rho_1 \text{vec}^{-1}(\tilde{\mathbf{c}})^T] + \mathbf{Y}\mathbf{H}^T (\mathbf{H}\mathbf{H}^T + (\rho_1 + \rho_3)\mathbf{I})^{-1},$$

where $\tilde{\mathbf{Z}} := \mathbf{Z} - \mathbf{U}_3$, $\tilde{\mathbf{c}} := \mathbf{c} - \mathbf{u}_1$, $\rho := \rho_1 + \rho_3$. To estimate $\mathbf{A}'$ for restricted model, there is an additional constraint $a_{xy}^{(i)'} = a_{yx}^{(i)'} = 0 \quad \forall i = 1, \dots, \bar{m}$. To address this assumption of known support, two approaches can be employed: i) first, by increasing the hierarchical sparsity parameter λ' , especially on cross-coupling terms, to force them to become zero, ii) second, by discarding the cross-coupling terms in each iteration after solving the ordinary least squares problem. To avoid difficulties associated with both approaches, we use the idea to shift the sparsity from $\mathbf{A}'$ to data matrices $\mathbf{Y}$ and $\mathbf{H}$ such that cross-coupling terms do not play any role in optimization for *restricted* model. By redefining matrices in Equation (3.3) we can

write:

$$\mathbf{Y}' = \mathbf{A}'\mathbf{H}' + \mathbf{E}',$$

where $\mathbf{Y}' := \begin{bmatrix} y(t) & \mathbf{0} \\ \mathbf{0} & x(t) \end{bmatrix} \in \mathbb{R}^{2 \times 2T'}$, $\mathbf{H}' := \begin{bmatrix} y(t-1) & \mathbf{0} \\ \vdots & \mathbf{0} \\ y(t-\bar{m}) & x(t-1) \\ \mathbf{0} & \vdots \\ & x(t-\bar{m}) \end{bmatrix} \in \mathbb{R}^{2\bar{m} \times 2T'}$, $\mathbf{E}' := \begin{bmatrix} n_y(t)' & \mathbf{0} \\ \mathbf{0} & n_x(t)' \end{bmatrix} \in \mathbb{R}^{2 \times 2T'}$, and $\mathbf{A}' := \begin{bmatrix} a_{yy}^{(1)'} \cdots a_{yy}^{(\bar{m})'} & a_{yx}^{(1)'} \cdots a_{yx}^{(\bar{m})'} \\ a_{xx}^{(1)'} \cdots a_{xx}^{(\bar{m})'} & a_{xy}^{(1)'} \cdots a_{xy}^{(\bar{m})'} \end{bmatrix} \in \mathbb{R}^{2 \times 2\bar{m}}$. Now we can update $\mathbf{A}'$ according to:

$$\mathbf{A}' = [\rho'_3 \tilde{\mathbf{Z}}' + \rho_1 \text{vec}^{-1}(\tilde{\mathbf{c}}')^T] + \mathbf{Y}'\mathbf{H}'^T(\mathbf{H}'\mathbf{H}'^T + \rho'\mathbf{I})^{-1},$$

where $\rho' := \rho'_1 + \rho'_3$. After computing $\mathbf{A}'$ we can safely discard cross-coupling terms without any loss of information. Closed-form solutions for hierarchical sparsity terms $\mathbf{P}, \mathbf{Q}$ and $\mathbf{P}', \mathbf{Q}'$ are obtained as discussed in [64]. For details, refer to Algorithm 1. To find the update for $\mathbf{Z}$, we leverage the following two properties of $\Gamma(\cdot)$ operator.

Property 1

$$\|\Gamma(\mathbf{X})\|_F^2 = \|\mathbf{X}\|_F^2 + 2(\bar{m} - 1) \text{ for any } \mathbf{X} \in \mathbb{R}^{2 \times 2\bar{m}}.$$

Property 2

$$\Gamma(\mathbf{X} - \mathbf{Y}) = 2\Gamma(\frac{1}{2}\mathbf{X}) - \Gamma(\mathbf{Y}) \text{ for any } \mathbf{X}, \mathbf{Y} \in \mathbb{R}^{2 \times 2\bar{m}}.$$

These two properties allow us to find the closed-form solution for the update for $\mathbf{Z}$

as:

$$\begin{aligned}\mathbf{Z} &= \arg \min_{\mathbf{Z}} \gamma \|\mathbf{\Gamma}(\mathbf{Z})\|_2 + \frac{\rho_3}{2} \|\mathbf{A} + \mathbf{U}_3 - \mathbf{Z}\|_F^2, \\ &= \arg \min_{\mathbf{Z}} \gamma \|\mathbf{\Gamma}(\mathbf{Z})\|_2 + \frac{\rho_3}{2} \|2\mathbf{\Gamma}(\frac{1}{2}(\mathbf{A} + \mathbf{U}_3)) - \mathbf{\Gamma}(\mathbf{Z})\|_F^2, \\ \mathbf{Z} &= \Pi_{\gamma/\rho_3}(2\mathbf{\Gamma}(\frac{1}{2}(\mathbf{A} + \mathbf{U}_3))).\end{aligned}$$

The complete algorithm is shown in Algorithm 1.

Algorithm 1 SS-ADMM for GC Estimation

Input: $\mathbf{Y}, \mathbf{H}, \bar{m}, \lambda, \gamma, \rho_1, \rho_2, \rho_3$

```

while stopping criterion not met do
   $\mathbf{B} := \rho(\mathbf{Z} - \mathbf{U}_3 + \text{vec}^{-1}(\mathbf{c} - \mathbf{u}_1)^T)$ 
   $\mathbf{A} = (\mathbf{B} + \mathbf{Y}\mathbf{H}^T)(\mathbf{H}\mathbf{H}^T + \rho\mathbf{I})^{-1}$ 
  for  $g = 1, \dots, 4\bar{m}$  do
     $\mathbf{P}_{gg} = \text{prox}_{\lambda w_g \|\cdot\|}(\mathbf{P}_{gg} + \mathbf{q}_g - \mathbf{u}_{2g} - \mathbf{p}_g)$ 
     $\mathbf{P}_{g^c g} = 0$ 
  end for
   $\mathbf{p} = 1/4\bar{m} \sum_{g \in \mathcal{G}} \mathbf{P}$ 
   $\mathbf{q} = 1/(\rho_2 + 4\bar{m})(\mathbf{c} + \rho_2(\mathbf{u}_2 + \mathbf{p}))$ 
   $\mathbf{c} = \text{vec}(\mathbf{A}^T) + \sum_{g \in \mathcal{G}} \mathbf{P} + \text{vec}(\mathbf{U}_1)$ 
   $\mathbf{Z} = \Pi_{\gamma/\rho_3}(2\mathbf{\Gamma}(\frac{1}{2}(\mathbf{A} + \mathbf{U}_3)))$ 
   $\mathbf{u}_1 = \mathbf{u}_1 + \text{vec}(\mathbf{A}^T) - \mathbf{c}$ 
   $\mathbf{u}_2 = \mathbf{u}_2 + \mathbf{p} - \mathbf{q}$ 
   $\mathbf{U}_3 = \mathbf{U}_3 + \mathbf{A} - \mathbf{Z}$ 
   $m_i = \text{card}(\mathbf{P}_{(i-1)\bar{m}+1:i\bar{m}} \neq 0) \ \forall i = 1, 2, 3, 4$ 
end while

```

Output: $\mathbf{A}, m_1, m_2, m_3, m_4$

3.4 Non-convex ADMM for Noise Robust Estimation

The development of robust methodologies for causal inference in the presence of measurement noise represents a critical frontier in contemporary time series analysis. In the above-mentioned model the excitation noise term represents the process noise associated with inherent dynamics of the system being modeled. In the context of Granger causality estimation based on an autoregressive model, one of the primary

assumptions is that the variables need to be observed without measurement errors [110]. For simplicity, most of the analysis methods assume that the measurement noise is negligible and the recorded data represent the true underlying process dynamics. However, this assumption is technically incorrect and, therefore, leads to significantly biased estimate of the model and false inference of causation when the underlying model is affected by considerable noise [82]. Hence, handling measurement noise in the model is crucial to obtain reliable Granger causality estimates.

Estimating Granger causality in the presence of measurement errors has scarcely been investigated in the literature. Existing approaches mostly rely on Kalman filtering and Expectation Maximization [82], [18], [85] techniques using a state space representation of the model. However, Kalman filtering assumes the knowledge of an exact mathematical model of the system and Gaussian assumption for both the state and measurement noise. Deviations from these assumptions can lead to suboptimal performance. Expectation Maximization, on the other hand, is sensitive to the choice of initial parameters and involves iterative optimization steps resulting in high computational complexity, especially for large and complex systems.

To overcome the limitations associated with these approaches, leveraging empirical knowledge about the recorded signals can result in a realistic model for incorporating measurement noise. In well-controlled experimental settings with high-quality EEG equipment, where the noise characteristics are relatively stable over time, it is reasonable to assume the measurement noise to be stationary Gaussian noise. The physiological noise, on the other hand, can be represented in the form of non-Gaussian trends such as skewness or sudden bursts of very large amplitudes at random times.

In the context of EEG and sEMG signals, observing such non-Gaussian trends and measurement noise is unavoidable. Therefore, to account for this, instead of assuming Gaussianity on the excitation noise sequence as a whole, we propose to assume Gaussianity on the recorded measurement noise and the excitation noise term containing non-Gaussian trends is assumed to be sparse in a wavelet basis. This assumption offers a practical approach to modeling real signals, as supported

by previous research in [8] and [7]. As a result, it is expected to obtain improved causality estimates.

Revisiting the bivariate VAR model in Equation (3.3), now we consider the presence of measurement noise in the observed signals. The observed signals $\hat{\mathbf{Y}}$ and their past values $\hat{\mathbf{H}}$ can, therefore, be represented as a sum of measurement noise-free signal and an additive Gaussian measurement noise term, $\hat{\mathbf{Y}} = \mathbf{Y} + \Delta\mathbf{Y}$ and $\hat{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}$, respectively. From the original autoregressive model in (3.3) it follows then that the observed signals are related according to:

$$\hat{\mathbf{Y}} - \Delta\mathbf{Y} = \mathbf{A}(\hat{\mathbf{H}} - \Delta\mathbf{H}) + \mathbf{E}, \quad (3.11)$$

$$\hat{\mathbf{Y}} = \mathbf{A}\mathbf{H} + \mathbf{E} + \Delta\mathbf{Y}. \quad (3.12)$$

The model in Equation (3.11) is ill-posed in a conventional setting, since both the excitation noise $\mathbf{E}$ and the measurement noise $\Delta\mathbf{Y}$ and $\Delta\mathbf{H}$ are, traditionally, assumed to be Gaussian. However, considering a realistic signal model we propose to enforce a wavelet sparsity assumption on the excitation noise $\mathbf{E}$, whereas the measurement noise terms $\Delta\mathbf{Y}$ and $\Delta\mathbf{H}$ are assumed to be Gaussian and relatively weak compared to the physiological signal. By using these constraints, we formulate a non-convex optimization problem based on the model proposed in Equation (3.12).

We formulate the objective in terms of $\mathbf{H}$ that will eventually be computed by subtracting the noise values. The overall objective function can be written as follows:

$$\begin{aligned} \min \quad & \kappa \|\mathcal{W}(\mathbf{E})\|_1 + \frac{\alpha}{2} \|\hat{\mathbf{Y}} - \mathbf{A}\mathbf{H} - \mathbf{E}\|_F^2 + \lambda \Omega_{\text{LOG}}(\mathbf{c}) + \gamma \Psi_{\text{SP}}(\mathbf{Z}) \\ \text{s.t.} \quad & \mathbf{c} = \text{vec}(\mathbf{A}^T), \quad \mathbf{A} = \mathbf{Z}, \end{aligned}$$

where $\mathcal{W}(\cdot)$ indicates an orthogonal wavelet transform. In the above constrained optimization problem, the last two terms enforce hierarchical sparsity and stationarity, as in (3.10).

The above program is non-convex due to the presence of bilinear term $\mathbf{A}\mathbf{H}$.

However, the matrix $\mathbf{H}$ has a special Toeplitz structure that can be leveraged to efficiently handle the non-convex bilinear term. The augmented Lagrangian of the above program in scaled form, can be written as follows:

$$\begin{aligned} \min \kappa & \|\mathcal{W}(\mathbf{E})\|_1 + \frac{\alpha}{2} \|\hat{\mathbf{Y}} - \mathbf{A}\mathbf{H} - \mathbf{E}\|_F^2 + \lambda \Omega_{\text{LOG}}(\mathbf{c}) + \gamma \Psi_{\text{SP}}(\mathbf{Z}) \\ & + \frac{\rho_1}{2} \|\mathbf{A} - \text{vec}^{-1}(\mathbf{c} - \mathbf{u}_1)\|_F^2 + \frac{\rho_3}{2} \|\mathbf{A} - \mathbf{Z} + \mathbf{U}_3\|_F^2, \end{aligned}$$

which can be solved using non-convex ADMM. The closed form solutions for all the matrices can be obtained by fixing the other variables and optimizing with respect to one of them. The closed form solutions to enforce hierarchical sparsity and stationarity conditions are estimated by solving (3.9). By keeping all other matrices fixed, the update for parameter matrix $\mathbf{A}$ can be computed as follows:

$$\begin{aligned} \mathbf{A} &= \arg \min_{\mathbf{A}} \frac{\alpha}{2} \|\hat{\mathbf{Y}} - \mathbf{A}\mathbf{H} - \mathbf{E}\|_F^2 + \frac{\rho_3}{2} \|\mathbf{A} - \mathbf{Z} + \mathbf{U}_3\|_F^2 \\ &+ \frac{\rho_1}{2} \|\mathbf{A} - \text{vec}^{-1}(\mathbf{c} - \mathbf{u}_1)\|_F^2, \\ &= [\rho_3(\tilde{\mathbf{Z}}) + \rho_1(\text{vec}^{-1}(\tilde{\mathbf{c}})^T) + \alpha \tilde{\mathbf{Y}}\mathbf{H}^T][\alpha \mathbf{H}\mathbf{H}^T + (\rho_3 + \rho_1)\mathbf{I}]^{-1}, \end{aligned} \tag{3.13}$$

where $\tilde{\mathbf{Z}} := \mathbf{Z} - \mathbf{U}_3$, $\tilde{\mathbf{c}} := \mathbf{c} - \mathbf{u}_1$, and $\tilde{\mathbf{Y}} := \hat{\mathbf{Y}} - \mathbf{E}$. The update for excitation noise matrix $\mathbf{E}$ can be simplified as follows:

$$\begin{aligned} \mathbf{E} &= \arg \min_{\mathbf{E}} \kappa \|\mathcal{W}(\mathbf{E})\|_1 + \frac{\alpha}{2} \|\hat{\mathbf{Y}} - \mathbf{A}\mathbf{H} - \mathbf{E}\|_F^2, \\ &= \mathcal{W}^{-1}(\mathcal{S}_{\kappa/\alpha}(\mathcal{W}(\hat{\mathbf{Y}} - \mathbf{A}\mathbf{H}))), \end{aligned}$$

where $\mathcal{W}^{-1}(\cdot)$ is the inverse of the orthogonal wavelet transform and $\mathcal{S}(\cdot)$ is the soft-thresholding operator. The update for the measurement noise matrix $\Delta \mathbf{Y}$ turns out to be:

$$\Delta \mathbf{Y} = \hat{\mathbf{Y}} - \mathbf{A}\mathbf{H} - \mathbf{E} .$$

In order to find the update for $\mathbf{H}$, we first estimate the noise matrix $\Delta\mathbf{H}$. After updating all the matrices, $\Delta\mathbf{H}$ can be updated as a function of these matrices by leveraging the inherent Toeplitz structure as follows:

$$\Delta\mathbf{H} = \begin{bmatrix} n_y(\bar{m}) & \cdots & n_y(2\bar{m}-1) & \cdots & n_y(\bar{m}+T'-1) \\ n_y(\bar{m}-1) & \cdots & n_y(2\bar{m}-2) & \cdots & n_y(\bar{m}+T'-2) \\ & & \vdots & & \\ n_y(1) & \cdots & n_y(\bar{m}) & \cdots & n_y(T') \\ n_x(\bar{m}) & \cdots & n_x(2\bar{m}-1) & \cdots & n_x(\bar{m}+T'-1) \\ n_x(\bar{m}-1) & \cdots & n_x(2\bar{m}-2) & \cdots & n_x(\bar{m}+T'-2) \\ & & \vdots & & \\ n_x(1) & \cdots & n_x(\bar{m}) & \cdots & n_x(T') \end{bmatrix}, \quad (3.14)$$

where $\Delta\mathbf{H} = \begin{bmatrix} \mathbf{T}_y \\ \mathbf{T}_x \end{bmatrix}$, and $\mathbf{T}_y \in \mathbb{R}^{\bar{m} \times T'}$ and $\mathbf{T}_x \in \mathbb{R}^{\bar{m} \times T'}$ are rectangular Toeplitz matrices containing noise terms in $y(t)$ and $x(t)$, respectively.

Due to the Toeplitz structure both $\mathbf{T}_y$ and $\mathbf{T}_x$ can be fully determined by their first row and first column. To determine the first row of $\mathbf{T}_y$ and $\mathbf{T}_x$, by comparing Equation(3.11) and Equation(3.14), we observe

$$\mathbf{T}_{y1,2:} = \Delta\mathbf{Y}_{1,:}, \quad (3.15)$$

$$\mathbf{T}_{x1,2:} = \Delta\mathbf{Y}_{2,:}. \quad (3.16)$$

The first column of $\Delta \mathbf{H} = \begin{bmatrix} \mathbf{T}_y \\ \mathbf{T}_x \end{bmatrix}$ can be determined efficiently by solving a simple least square problem. Consider the Equation (3.11) for $t = \bar{m} + 1$:

$$\begin{aligned} & \begin{bmatrix} \hat{y}(\bar{m} + 1) \\ \hat{x}(\bar{m} + 1) \end{bmatrix} - \begin{bmatrix} n_y(\bar{m} + 1) \\ n_x(\bar{m} + 1) \end{bmatrix} \\ &= \begin{bmatrix} a_{yy}^{(1)} \cdots a_{yy}^{(\bar{m})} & a_{yx}^{(1)} \cdots a_{yx}^{(\bar{m})} \\ a_{xy}^{(1)} \cdots a_{xy}^{(\bar{m})} & a_{xx}^{(1)} \cdots a_{xx}^{(\bar{m})} \end{bmatrix} \left(\begin{bmatrix} \hat{y}(\bar{m}) \\ \hat{y}(\bar{m} - 1) \\ \vdots \\ \hat{y}(1) \\ \hat{x}(\bar{m}) \\ \hat{x}(\bar{m} - 1) \\ \vdots \\ \hat{x}(1) \end{bmatrix} - \begin{bmatrix} n_y(\bar{m}) \\ n_y(\bar{m} - 1) \\ \vdots \\ n_y(1) \\ n_x(\bar{m}) \\ n_x(\bar{m} - 1) \\ \vdots \\ n_x(1) \end{bmatrix} \right) \\ &+ \begin{bmatrix} \epsilon_y(\bar{m} + 1) \\ \epsilon_x(\bar{m} + 1) \end{bmatrix} . \end{aligned}$$

Assigning the matrices below to their corresponding values in the equation above gives:

$$\hat{\mathbf{y}} - \Delta \mathbf{y} = \mathbf{A}(\mathbf{h} - \Delta \mathbf{h}) + \mathbf{e},$$

$$\mathbf{A} \Delta \mathbf{h} = \mathbf{A} \mathbf{h} - \hat{\mathbf{y}} + \mathbf{e} + \Delta \mathbf{y},$$

$$\Delta \mathbf{h} = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} (\mathbf{A} \mathbf{h} - \hat{\mathbf{y}} + \mathbf{e} + \Delta \mathbf{y}) .$$

Hence we can write

$$\Delta \mathbf{H}_{:,1} = \begin{bmatrix} \mathbf{T}_{y:,1} \\ \mathbf{T}_{x:,1} \end{bmatrix} = \Delta \mathbf{h} . \quad (3.17)$$

The measurement noise matrix $\Delta \mathbf{H}$ can then be obtained from equations (3.15-3.16) and equation(3.17) and by exploiting its Toeplitz structure. The uncontaminated matrix $\mathbf{H}$ can then be computed by subtracting estimated noise from the noisy observations as:

$$\mathbf{H} = \hat{\mathbf{H}} - \Delta \mathbf{H}.$$

Finally, the Lagrange multipliers are computed and all the variables are updated iteratively. The extension of SS-ADMM to incorporate measurement noise in recorded signals is termed as SSD-ADMM: Stationary and Sparse Denoising ADMM as outlined in Algorithm 2.

Figure 3.1 provides a comparison of our two methodological frameworks: the Stable Sparse Denoising ADMM (SSD-ADMM) and the Stable Sparse ADMM (SS-ADMM), highlighting their distinct approaches to noise handling in causal inference. The flowchart illustrates how both methodologies initiate with multivariate time series data, but diverge fundamentally in their treatment of noise components. The SS-ADMM framework proceeds with a conventional assumption of negligible measurement noise, focusing primarily on the optimization of stable sparse parameters within the autoregressive framework. In contrast, the SSD-ADMM framework implements a more nuanced approach to noise characterization, explicitly differentiating between measurement and excitation noise through distinct structural constraints. This methodology leverages the empirically-justified assumption of Gaussian distributions for measurement noise, a characteristic typically observed in well-calibrated data acquisition systems, while simultaneously accounting for the non-Gaussian nature

Algorithm 2 SSD-ADMM for GC Estimation

Input: $\hat{\mathbf{Y}}, \hat{\mathbf{H}}, \bar{m}, \lambda, \gamma, \alpha, \kappa, \rho_1, \rho_2, \rho_3$

while stopping criterion not met **do**

$$\mathbf{A} = [\rho_3(\mathbf{Z} - \mathbf{U}_3) + \rho_1(\text{vec}^{-1}(\mathbf{c} - \mathbf{u}_1)^T) + \alpha(\hat{\mathbf{Y}} - \mathbf{E})(\mathbf{H})^T][\alpha\mathbf{H}\mathbf{H}^T + (\rho_1 + \rho_3)\mathbf{I}]^{-1}$$

for $g = 1, \dots, 4\bar{m}$ **do**

$$\mathbf{P}_{gg} = \text{prox}_{\lambda w_g \|\cdot\|}(\mathbf{P}_{gg} + \mathbf{q}_g - \mathbf{u}_{2g} - \mathbf{p}_g)$$

$$\mathbf{P}_{g^c g} = 0$$

end for

$$\mathbf{p} = 1/4\bar{m} \sum_{g \in \mathcal{G}} \mathbf{P}$$

$$\mathbf{q} = 1/(\rho_2 + 4\bar{m})(\mathbf{c} + \rho_2(\mathbf{u}_2 + \mathbf{p}))$$

$$\mathbf{c} = \text{vec}(\mathbf{A}^T) + \sum_{g \in \mathcal{G}} \mathbf{P} + \text{vec}(\mathbf{U}_1)$$

$$\mathbf{Z} = \Pi_{\gamma/\rho_3}(2\Gamma(\frac{1}{2}(\mathbf{A} + \mathbf{U}_3)))$$

$$\mathbf{E} = \mathcal{W}^{-1}(\mathcal{S}_{\kappa/\alpha}(\mathcal{W}(\hat{\mathbf{Y}} - \mathbf{A}\mathbf{H})))$$

$$\Delta \mathbf{Y} = \hat{\mathbf{Y}} - \mathbf{A}\mathbf{H} - \mathbf{E}$$

$$\Delta \mathbf{H}_{\{1, \bar{m}+1\},:} = \begin{bmatrix} \Delta \mathbf{Y}_{1,:} \\ \Delta \mathbf{Y}_{2,:} \end{bmatrix}$$

$$\Delta \mathbf{H}_{:,1} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{h} - \hat{\mathbf{y}} + \mathbf{e} + \Delta \mathbf{y})$$

$$\mathbf{H} = \hat{\mathbf{H}} - \Delta \mathbf{H}$$

$$\mathbf{u}_1 = \mathbf{u}_1 + \text{vec}(\mathbf{A}^T) - \mathbf{c}$$

$$\mathbf{u}_2 = \mathbf{u}_2 + \mathbf{p} - \mathbf{q}$$

$$\mathbf{U}_3 = \mathbf{U}_3 + \mathbf{A} - \mathbf{Z}$$

$$m_i = \text{card}(\mathbf{P}_{(i-1)\bar{m}+1:i\bar{m}} \neq 0) \quad \forall i = 1, 2, 3, 4$$

end while

Output: $\mathbf{A}, \mathbf{Y}, \mathbf{H}, \mathbf{E}, m_1, m_2, m_3, m_4$

of excitation noise, which often manifests as asymmetric distributions or sporadic high-amplitude perturbations in physiological signals.

Both frameworks converge in their utilization of ADMM-based optimization strategies, but the SSD-ADMM's noise handling mechanism enables more robust parameter estimation under realistic measurement conditions. This distinction is particularly crucial in contexts where measurement noise cannot be dismissed as negligible, such as in complex physiological recording scenarios. The parallel representation of these methodologies effectively illustrates how explicit consideration of noise characteristics can enhance the reliability of causal inference while maintaining computational tractability.

These methodological considerations, while addressing fundamental noise characteristics, reveal a deeper challenge in causal inference: the need to discover intrinsic

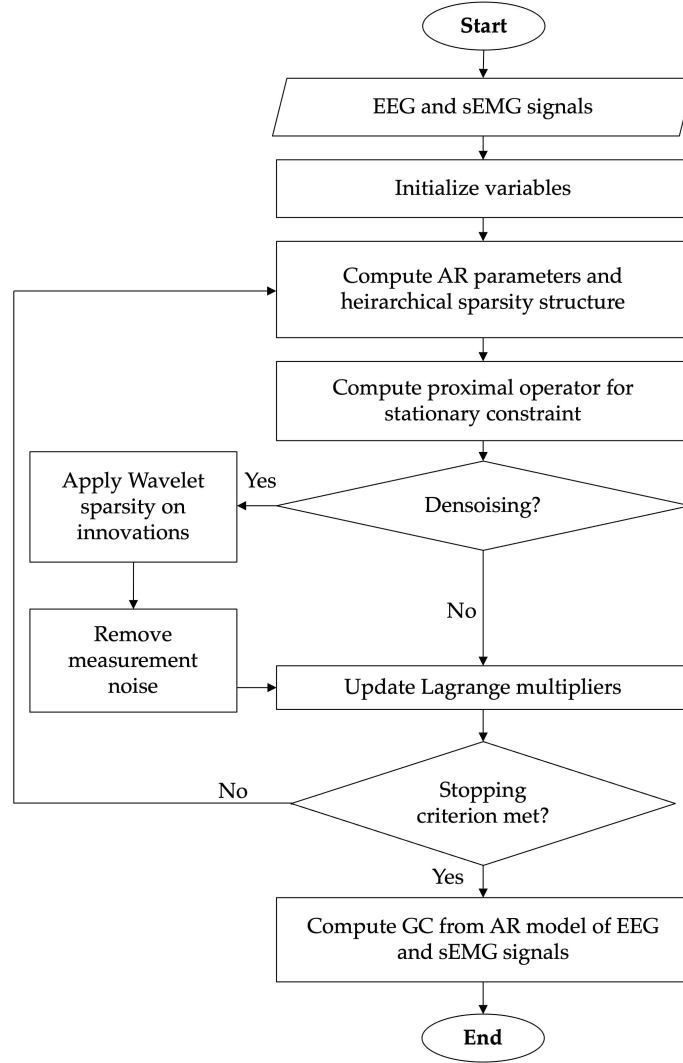


Figure 3.1: Proposed method overview: with (SSD-ADMM) vs. without (SS-ADMM) denoising.

signal representations that can effectively disambiguate meaningful temporal dependencies from measurement artifacts. This recognition motivates our exploration of other approaches to feature representation within the autoregressive framework.

3.5 Data-Driven Learned Dictionary for VAR Modeling

The limitations of conventional autoregressive models in capturing complex signal structures stem partly from their reliance on fixed parameter matrices that implicitly assume direct temporal relationships in the observed data. We propose a fundamental

reconceptualization of autoregressive parameter representation through dictionary learning, whereby the coefficient matrix encoding temporal dependencies is expressed as a linear combination of learned atomic elements.

Traditional VAR models employ an identity matrix as their implicit dictionary, essentially constraining the model to operate directly in the observed signal space. This representation proves inadequate for capturing the rich, hierarchical structure inherent in physiological signals. While previous work in EEG and EMG signal processing [27] has demonstrated the utility of shared dictionaries for feature extraction, the application of dictionary learning to autoregressive parameter estimation in causal discovery remains unexplored.

Our proposed method reformulates the autoregressive coefficient matrix as a composition of learned dictionary elements, enabling a more expressive representation of temporal dependencies. Each coefficient in the autoregressive model is represented as a linear combination of dictionary atoms, providing a flexible framework for capturing complex relationships while maintaining interpretability. This decomposition allows the model to adapt its representational basis to the underlying structure of the temporal dependencies, rather than imposing a fixed parametric form. By learning an optimal basis for parameter representation, the model can better distinguish between genuine temporal dependencies and spurious relationships induced by measurement noise. This approach naturally aligns with the nature of physiological processes, where complex interactions often manifest through combinations of fundamental patterns. Traditional approaches to dictionary learning and sparse coding, such as Orthogonal Matching Pursuit (OMP) or Basis Pursuit (BP), often exhibit sensitivity to noise due to their reliance on purely signal-based decision criteria [135]. We introduce a more robust optimization strategy based on the Proximal Dogleg Opportunistic Majorization (PDOM) algorithm [140]. This approach enables comprehensive exploration of the solution space through dynamic step size and direction adjustments based on local curvature information. The incorporation of local surrogate functions provides accurate approximations of the

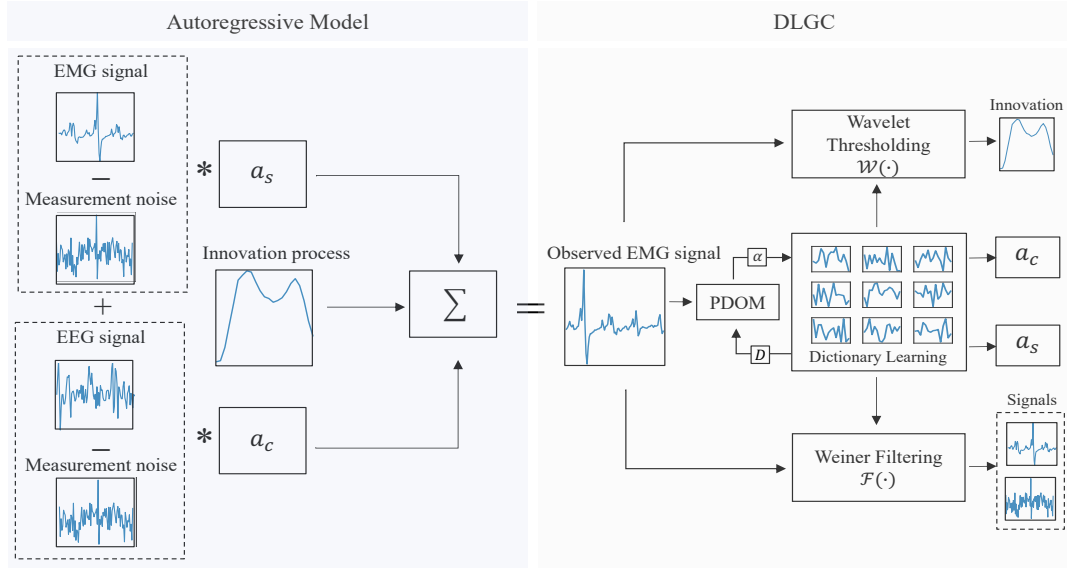


Figure 3.2: Block diagram for our proposed method DLGC

optimization landscape, enhancing the robustness of sparse signal recovery in noisy environments [58].

Our proposed algorithm, Dictionary Learning based Granger Causality (DLGC), implements an alternating minimization strategy to jointly learn feature representations and separate noise components. The algorithm combines a gradient-descent based dictionary learning for atomic element updates, PDOM-based sparse coefficient estimation, and Wiener filtering for optimal noise reduction in the mean squared error sense [25]. The overview of the optimization process using DLGC is shown in Figure 3.2.

To formalize mathematically, consider $\mathbf{y} \in \mathbb{R}^T$ is the clean EEG signal and $\mathbf{x} \in \mathbb{R}^T$ is the corresponding clean sEMG signal, then the model in Equations (3.1), (3.2) can be rewritten in matrix form as follows:

$$\mathbf{y} = \mathbf{Y}\mathbf{a}_{yy} + \mathbf{X}\mathbf{a}_{yx} + \boldsymbol{\epsilon}_y, \quad (3.18)$$

$$\mathbf{x} = \mathbf{Y}\mathbf{a}_{xy} + \mathbf{X}\mathbf{a}_{xx} + \boldsymbol{\epsilon}_x, \quad (3.19)$$

where $\boldsymbol{\epsilon}_y, \boldsymbol{\epsilon}_x \in \mathbb{R}^{T \times 1}$, $\mathbf{a}_i \in \mathbb{R}^{m \times 1}$, $\forall i \in \{yy, yx, xy, xx\}$, $\mathbf{Y}, \mathbf{X} \in \mathbb{R}^{T \times m}$ represent the Toeplitz matrices associated with past values of $\mathbf{y}, \mathbf{x}$, respectively. To generalize the

two equations above, we define the future values of the signal as $\mathbf{z} \in \mathbb{R}^T$ that can either be $\mathbf{y}$ or $\mathbf{x}$ i.e. $\mathbf{z} \in \{\mathbf{y}, \mathbf{x}\}$, and the above equations can be written, generically, as:

$$\mathbf{z} = \mathbf{Z}_s \mathbf{a}_s + \mathbf{Z}_c \mathbf{a}_c + \boldsymbol{\epsilon}_z, \quad (3.20)$$

where we refer to self-coupling coefficients $\mathbf{a}_{yy}, \mathbf{a}_{xx}$ as $\mathbf{a}_s$, cross-coupling coefficients $\mathbf{a}_{yx}, \mathbf{a}_{xy}$ as $\mathbf{a}_c$, and their corresponding Toeplitz matrices as $\mathbf{Z}_s, \mathbf{Z}_c$, respectively.

Now, we consider the measurement noise $\mathbf{n}_z \in \mathbb{R}^T$ in the observed noisy signal $\tilde{\mathbf{z}} \in \mathbb{R}^T$ as follows:

$$\mathbf{n}_z = \tilde{\mathbf{z}} - \mathbf{z} = \tilde{\mathbf{z}} - (\mathbf{Z}_s \mathbf{a}_s + \mathbf{Z}_c \mathbf{a}_c + \boldsymbol{\epsilon}_z),$$

where $\mathbf{Z}_s, \mathbf{Z}_c$ are constructed from past values of clean signals $\mathbf{z}_s, \mathbf{z}_c$ (self and cross signals based on the value of $\mathbf{z} \in \{\mathbf{y}, \mathbf{x}\}$), transformed by Toeplitz operator $\mathcal{T}(\cdot)$. Since we only have access to noisy observations, to get an estimate of the past values of the clean signals we employ Wiener filter $\mathcal{F}(\cdot)$ as follows:

$$\mathbf{Z}_s = \mathcal{T}(\mathbf{z}_s) \approx \mathcal{T}(\mathcal{F}(\tilde{\mathbf{z}}_s)), \quad (3.21)$$

$$\mathbf{Z}_c = \mathcal{T}(\mathbf{z}_c) \approx \mathcal{T}(\mathcal{F}(\tilde{\mathbf{z}}_c)). \quad (3.22)$$

Consider $\mathbf{z} \in \mathbb{R}^T$ is the EEG/sEMG signal, then we propose to minimize the non-convex, non-smooth optimization:

$$\begin{aligned} & \min \frac{1}{2} \|\tilde{\mathbf{z}} - (\mathcal{T}(\mathcal{F}(\tilde{\mathbf{z}}_s)) \mathbf{a}_s + \mathcal{T}(\mathcal{F}(\tilde{\mathbf{z}}_c)) \mathbf{a}_c + \boldsymbol{\epsilon}_z)\|_2^2 + \lambda \|\boldsymbol{\alpha}_c\|_0 \\ & + \lambda \|\boldsymbol{\alpha}_s\|_0 + \gamma \|\mathcal{W}(\boldsymbol{\epsilon}_z)\|_1 \\ & \text{s.t. } \mathbf{a}_s = \mathbf{D} \boldsymbol{\alpha}_s, \quad \mathbf{a}_c = \mathbf{D} \boldsymbol{\alpha}_c. \end{aligned} \quad (3.23)$$

where $\mathcal{W}(\cdot)$ indicates Wavelet transform operator to enforce Wavelet sparsity assumption on innovation process, $\mathbf{D}$ is the Dictionary used to represent the autoregressive

process whereas $\boldsymbol{\alpha}_{s,(c)}$ denotes the corresponding sparse coefficients vector. The first term in the objective function is data fidelity term to enforce the autoregressive structure while simultaneously ensuring that the measurement noise in data is small. The optimization problem leverages ℓ_0 norm minimization for feature selection and model interpretability by identifying a small subset of significant components in $\boldsymbol{\alpha}_s$ and $\boldsymbol{\alpha}_c$, while ℓ_1 norm minimization controls noise and prevents overfitting by shrinking small values to zero and penalizing large deviations. In unconstrained form using quadratic penalty terms, the objective function can be written as follows:

$$\begin{aligned} \min & \frac{1}{2} \|\tilde{\mathbf{z}} - (\mathcal{T}(\mathcal{F}(\tilde{\mathbf{z}}_s))\mathbf{a}_s + \mathcal{T}(\mathcal{F}(\tilde{\mathbf{z}}_c))\mathbf{a}_c + \boldsymbol{\epsilon}_z)\|_2^2 + \lambda \|\boldsymbol{\alpha}_c\|_0 \\ & + \lambda \|\boldsymbol{\alpha}_s\|_0 + \frac{\rho}{2} \|\mathbf{a}_s - \mathbf{D}\boldsymbol{\alpha}_s\|_2^2 + \frac{\rho}{2} \|\mathbf{a}_c - \mathbf{D}\boldsymbol{\alpha}_c\|_2^2 + \gamma \|\mathcal{W}(\boldsymbol{\epsilon}_z)\|_1. \end{aligned}$$

To compute the update for $\mathbf{a}_s, \mathbf{a}_c$, the least square solution can be employed by using $\mathbf{Z}_s, \mathbf{Z}_c$ from equations (3.21) and (3.22).

$$\mathbf{a}_s = (\mathbf{Z}_s^T \mathbf{Z}_s + \rho \mathbf{I})^{-1} [\mathbf{Z}_s^T (\tilde{\mathbf{z}} - \mathbf{Z}_c \mathbf{a}_c - \boldsymbol{\epsilon}_z) + \rho \mathbf{D} \boldsymbol{\alpha}_s], \quad (3.24)$$

$$\mathbf{a}_c = (\mathbf{Z}_c^T \mathbf{Z}_c + \rho \mathbf{I})^{-1} [\mathbf{Z}_c^T (\tilde{\mathbf{z}} - \mathbf{Z}_s \mathbf{a}_s - \boldsymbol{\epsilon}_z) + \rho \mathbf{D} \boldsymbol{\alpha}_c]. \quad (3.25)$$

For the sake of brevity, we use a generic notation to represent either $\mathbf{a}_s$ or $\mathbf{a}_c$ i.e. $\mathbf{a}_{s,(c)} \in \{\mathbf{a}_s, \mathbf{a}_c\}$. The complete DLGC algorithm is shown in Algorithm 1. After estimating the parameters, the next step is to find the sparse coefficients for the dictionary $\mathbf{D}$ using PDOM. In Algorithm 1, L_q refers to the Lipschitz constant of $\nabla q(\boldsymbol{\alpha}_{s,(c)})$ and is lower bounded by the maximum eigenvalue of the Hessian of $q(\boldsymbol{\alpha}_{s,(c)})$. The proximal operator for ℓ_0 norm is employed to efficiently handle the non-convex part of the objective function.

$$\begin{aligned} \boldsymbol{\alpha}^k &= \text{prox}_{\tau h}(\boldsymbol{\alpha}^{k-1} - \tau \nabla q(\boldsymbol{\alpha}^{k-1})), \\ &:= \arg \min_{\boldsymbol{\alpha}} h(\boldsymbol{\alpha}) + \frac{1}{2\tau} \|\boldsymbol{\alpha} - (\boldsymbol{\alpha}^{k-1} - \tau \nabla q(\boldsymbol{\alpha}^{k-1}))\|_2^2. \end{aligned}$$

The modified dogleg path $\mathbf{p}(\cdot)$ defined in [140] efficiently explores the optimization

landscape by dynamically adjusting the step size τ_β and direction based on the local curvature and gradient information provided by the surrogate function $m(\cdot, \cdot)$ as in [140]. The surrogate function using second order information, in turn, enables the optimization algorithm to make informed decisions about the direction and magnitude of each step. The estimate obtained by this search path is then compared against the Proximal Gradient (PG) [131] estimate to select the better one. The technical details regarding the complete PDOM algorithm and modified dogleg path can be found in [140]. The dictionary is updated next, by backpropagating the loss gradients by taking a gradient descent step to minimize the loss $f(\boldsymbol{\alpha}_{s,(c)}^k)$. Following that, the estimate for the innovation process is then computed by employing Wavelet thresholding operator. Eventually, the measurement noise is removed from the signal. All the above steps are repeated iteratively until convergence.

3.6 Experimental Results

In this section, we perform a comparative study to investigate the performance of proposed methods Stationary and Sparse ADMM (SS-ADMM) and its extension Stationary and Sparse Denoising ADMM (SSD-ADMM) against existing methods on both synthetic and real datasets. All the experiments have been performed in MATLAB 2022b with Core i7 CPU (2.90 GHz), 8 GB RAM, and Windows 11 operating system. The proposed method is tested on physiological data (HAMPD), introduced in Chapter 1, which is collected from healthy subjects in which subjects performed a controlled motor task, grasping a ruler between thumb and index finger [78]. An electromechanical tapper provided mechanical perturbations of lateral displacement to the ruler at pseudorandom intervals of 5.6 - 8.4 s (mean 7 s). The experiment comprised 8 blocks of 25 trials. sEMG was recorded from first dorsal interosseous and bipolar scalp EEG was recorded over left sensorimotor cortex. Both signals were sampled at 1024 Hz, amplified and band-pass filtered (0.5-100 Hz for EEG; 5-500 Hz for (sEMG)). Offline, data were divided into 5 s epochs (1.1 s pre- and

Algorithm 3 DLGC: Dictionary Learning for GC Esimation

Input: $\lambda, \gamma, \eta, \beta \in (1, 2], \tau \in (0, 1/L_q), \mathbf{z}, \mathbf{Z}_c, \mathbf{Z}_s$

Initialize $\mathbf{a}_c, \mathbf{a}_s, \boldsymbol{\alpha}_c, \boldsymbol{\alpha}_s, \mathbf{D}$

while stopping criterion not met **do**

 Compute least square update for $\mathbf{a}_s^k, \mathbf{a}_c^k$ by (3.24), (3.25)

 Compute updates for $\boldsymbol{\alpha}_s^k, \boldsymbol{\alpha}_c^k$ using PDOM as follows:

 Set $h(\boldsymbol{\alpha}_{s,(c)}^k) = \lambda \|\boldsymbol{\alpha}_{s,(c)}^k\|_0$

 Set $q(\boldsymbol{\alpha}_{s,(c)}^k) = \frac{\rho}{2} \|\mathbf{a}_{s,(c)}^k - \mathbf{D}^k \boldsymbol{\alpha}_{s,(c)}^k\|_2^2$

 Set $f(\boldsymbol{\alpha}_{s,(c)}^k) = q(\boldsymbol{\alpha}_{s,(c)}^k) + h(\boldsymbol{\alpha}_{s,(c)}^k)$

 Set $\tau_\beta = -\|\mathbf{p}(\beta)\|^2 / \langle \nabla q(\boldsymbol{\alpha}_{s,(c)}^k), \mathbf{p}(\beta) \rangle$

 Compute $\boldsymbol{\alpha}_{s,(c)}^k = \text{prox}_{\tau_\beta h}(\boldsymbol{\alpha}_{s,(c)}^{k-1} + \mathbf{p}(\beta^{k-1}))$ for β_{max}

 such that $m_\beta(\boldsymbol{\alpha}_{s,(c)}^k, \boldsymbol{\alpha}_{s,(c)}^{k-1}) \geq q(\boldsymbol{\alpha}_{s,(c)}^k)$

 Compute $\boldsymbol{\alpha}_s^{\text{PG}} = \text{prox}_{\tau h}(\boldsymbol{\alpha}_{s,(c)}^{k-1} - \tau \nabla q(\boldsymbol{\alpha}_{s,(c)}^{k-1}))$

if $f(\boldsymbol{\alpha}_{s,(c)}^k) \geq f(\boldsymbol{\alpha}_{s,(c)}^{\text{PG}})$ **then**

 Set $\boldsymbol{\alpha}_{s,(c)}^k = \boldsymbol{\alpha}_{s,(c)}^{\text{PG}}$

end if

 Backpropagate the loss gradients $\Delta \mathbf{d}^k = -\frac{\delta f(\boldsymbol{\alpha}_{s,(c)}^k)}{\delta \mathbf{d}^k}$

 Update dictionary atoms in $\mathbf{D}^k$ by $\mathbf{d}^k \leftarrow \mathbf{d}^{k-1} + \eta \Delta \mathbf{d}^k$

 Compute $\mathbf{Z}_s = \mathcal{T}(\mathcal{F}(\tilde{\mathbf{z}}_s)), \mathbf{Z}_c = \mathcal{T}(\mathcal{F}(\tilde{\mathbf{z}}_c))$

 Compute $\boldsymbol{\epsilon}_z^k = \mathcal{W}^{-1}(\mathcal{S}_\gamma(\mathcal{W}(\tilde{\mathbf{z}} - \mathbf{Z}_s \mathbf{a}_s^k - \mathbf{Z}_c \mathbf{a}_c^k)))$

 Compute denoised signal $\mathbf{z}^k = \mathbf{Z}_s \mathbf{a}_s^k + \mathbf{Z}_c \mathbf{a}_c^k + \boldsymbol{\epsilon}_z^k$

 Compute noise $\mathbf{n}_z^k = \tilde{\mathbf{z}} - \mathbf{z}^k$

end while

Output: $\mathbf{D}, \mathbf{a}_c, \mathbf{a}_s, \mathbf{n}_z, \boldsymbol{\epsilon}_z$

3.9s post-stimulus). Epochs containing movement or blink artefacts were eliminated.

3.6.1 Evaluation for SS-ADMM

Synthetic data has been generated using an AutoRegressive Fractionally Integrated Moving Average (ARFIMA) model with long memory as outlined in [15]. The bivariate process has 942 samples where the model orders for a_{yy}, a_{yx}, a_{xy} and, a_{xx} are set to be 17, 21, 20 and, 18 respectively. We assume that the upper limit on the model orders $\bar{m}$ is 30. The measurement noise variance is set to be 0.01. System parameters have been generated randomly based on an underlying mathematical model that governs variables $\mathbf{Y}$, $\mathbf{A}$, and $\mathbf{H}$. The results, obtained through multiple experimental

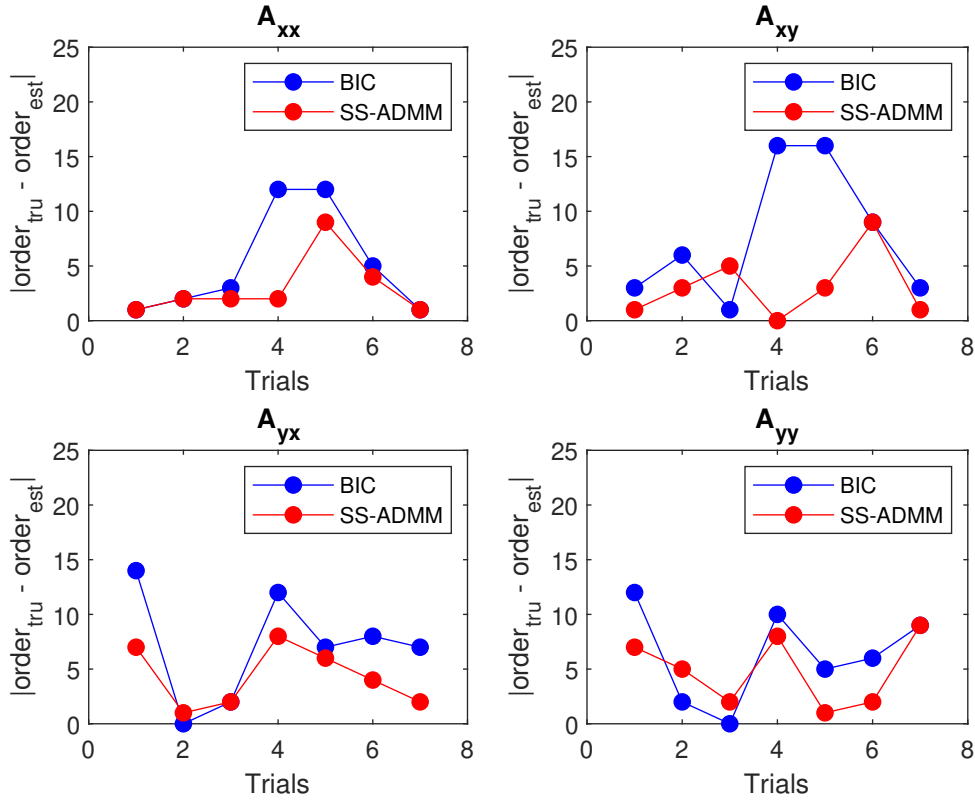


Figure 3.3: Model order identified by SS-ADMM Vs BIC [104].

repetitions, are compared to evaluate performance in terms of stationarity, sparsity and denoising. The performance of model identification is depicted in Figure 3.3, comparing it with BIC [104], which demonstrates that the proposed method achieves more accurate system order estimation. We observed also (not shown in the figure) that SS-ADMM tends to estimate the model order as either equal to or higher than the true order, while BIC typically underestimates the model order, thereby losing significant information about the underlying process. It is worth highlighting that SS-ADMM offers flexibility by allowing for fine-tuning of the hyperparameter λ to precisely estimate the model order.

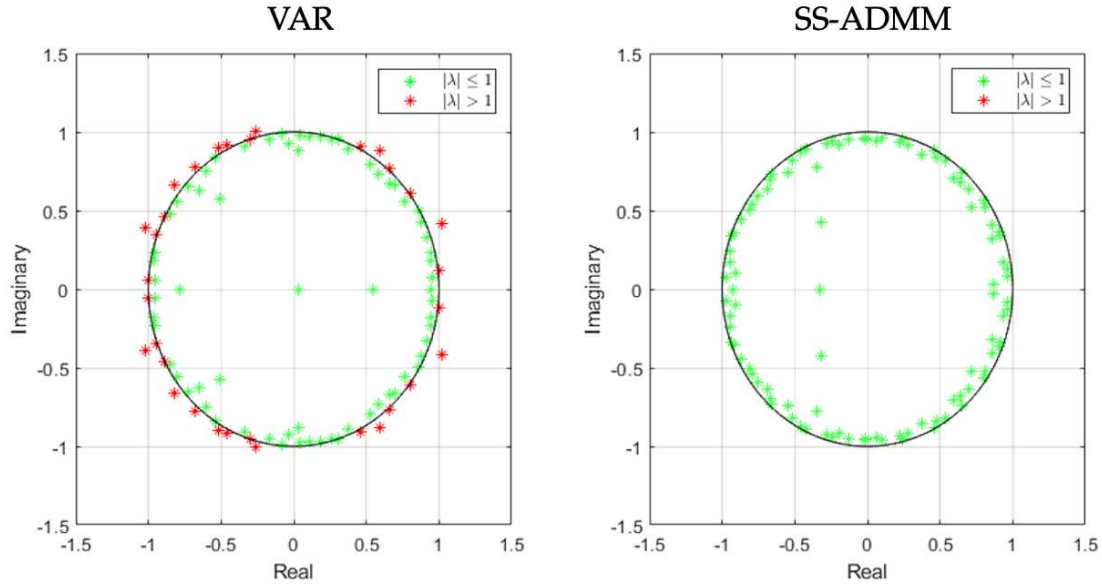


Figure 3.4: Eigenvalues of Companion matrix visualized

We performed grid search method [108] to find the optimal set of hyperparameter values where the domain of the hyperparameters is divided into a discrete grid and for every combination of values in the grid the performance metrics are computed using cross-validation. The combination of values that maximizes the average value in cross-validation is eventually selected. The proposed method also distinguishes between four distinct model orders for self-coupling and cross-coupling in a bivariate model, as opposed to just two in the case of BIC.

Figure 3.4 illustrates the effect of the stationarity constraint enforced by SS-ADMM. After applying the constraint, all eigenvalues are confined within the unit circle, unlike the standard VAR model, which is prone to instability.

Next we evaluate our method in comparison with Hierarchical Sparsity ADMM (HS-ADMM) [64] in terms of computational efficiency. The results obtained through the enforcement of stationarity using Euclidean projections, as demonstrated in [64], are contrasted with our proposed regularization based method, Ψ_{SP} , in Figure 3.5. Notably, the projection-based approach exhibits a longer execution time. Conversely, our proposed SS-ADMM approach achieves a speedup of $2\times$ or more per iteration, and the model-fitting error reduces more steadily when executed for the same number of iterations, as illustrated in Figure 3.5. The total execution time for HS-ADMM

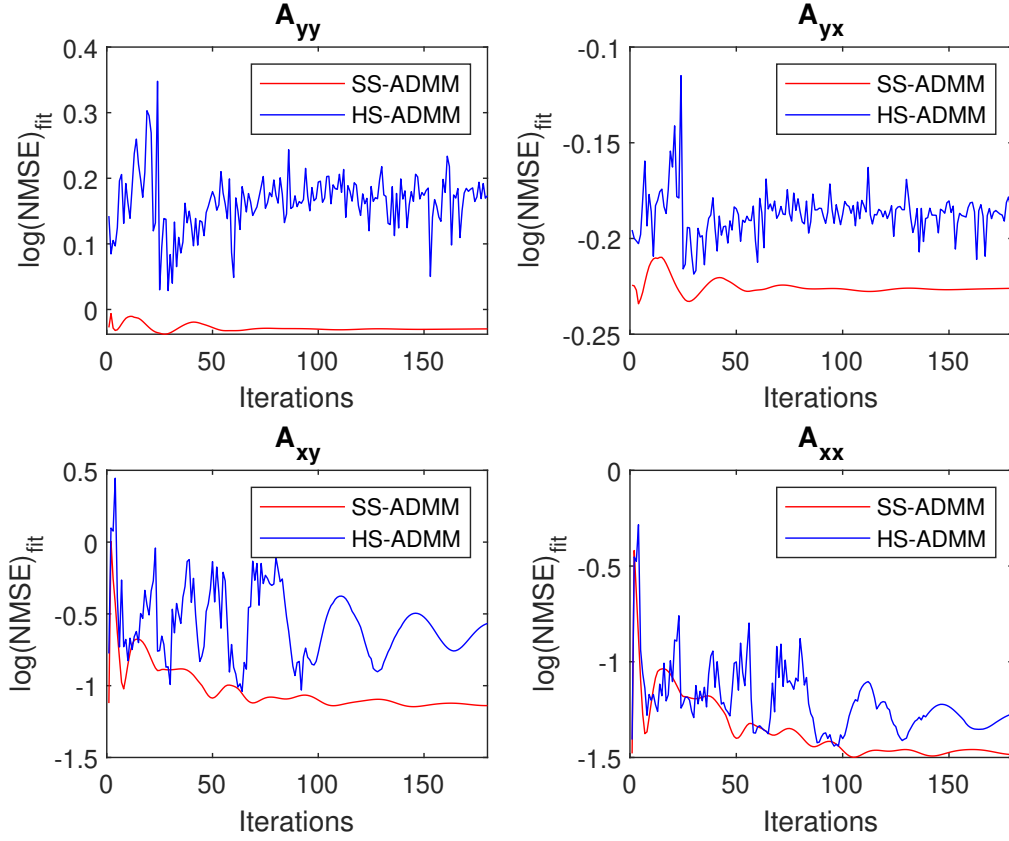


Figure 3.5: Log of normalized error of HS-ADMM [64] vs SS-ADMM.

was 1.9 minutes whereas SS-ADMM only took 0.39 seconds.

3.6.2 Evaluation for SSD-ADMM

To observe the denoising performance of SSD-ADMM, a measurement noise signal sampled from a multivariate normal distribution has been added to the simulated EEG and EMG signals. The noise variance $\sigma_{\Delta Y}$ is varied from 0 to 1.41. A sensing dictionary has been generated using a discrete Daubechies wavelet transform.

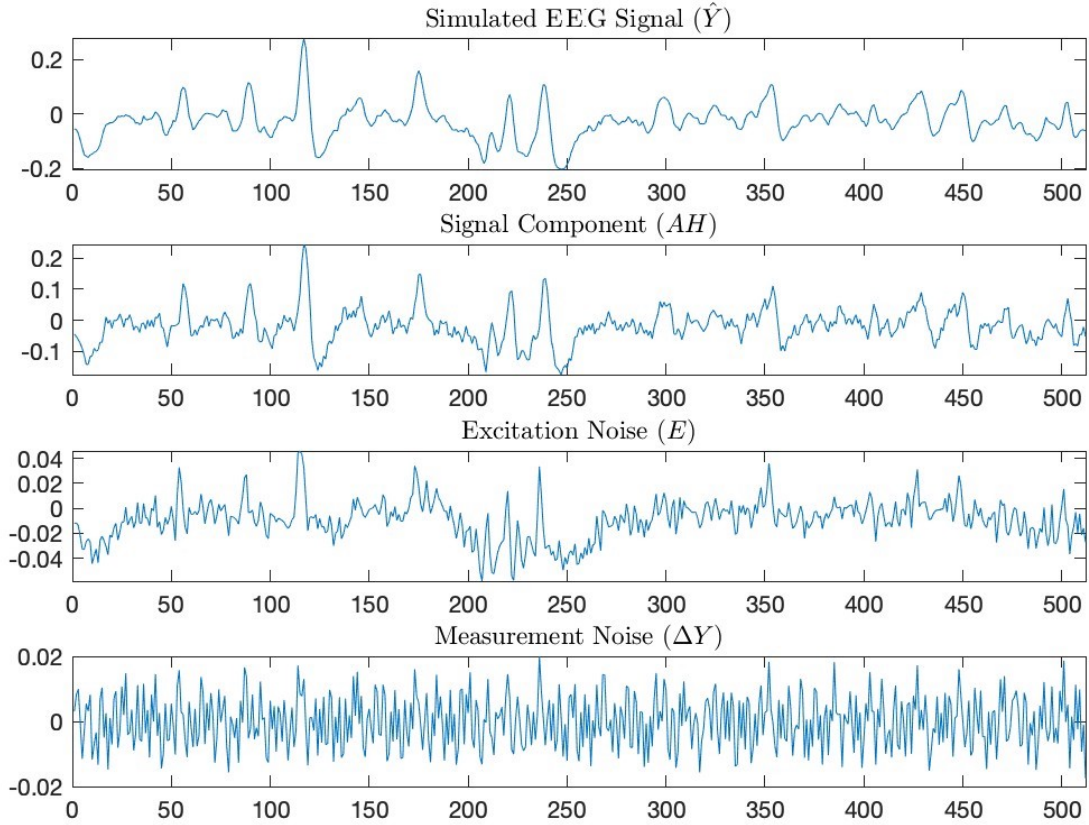


Figure 3.6: Signal decomposition for simulated signal

A randomly generated sparse vector, representing predominantly zero coefficients in the wavelet basis, is then subject to transformation with a sub-matrix derived from the dictionary, resulting in a wavelet sparse signal for modeling the excitation noise. The sparsity level is set to 70%. According to the model discussed in Equation (3.12), the observed noisy signal $\hat{\mathbf{Y}}$ can be decomposed as a combination of an autoregressive signal component $\mathbf{AH}$, an excitation noise $\mathbf{E}$ and a measurement noise term $\Delta\mathbf{Y}$. The estimates obtained for fitting the proposed model for simulated data as shown in Figure 3.6. In Figure 3.7, it can be observed that if signals are denoised first using a standard method such as Total Variation denoising [123], for instance, and then modeled using SS-ADMM, it experiences a degradation in the quality of fit quantified by Normalised Mean Squared Error (NMSE) because of its inability to handle measurement noise in data. On the other hand, SSD-ADMM effectively accounts for the measurement noise to estimate the clean signals and, thus, yields a more robust model. The proposed method is tested on physiological data collected

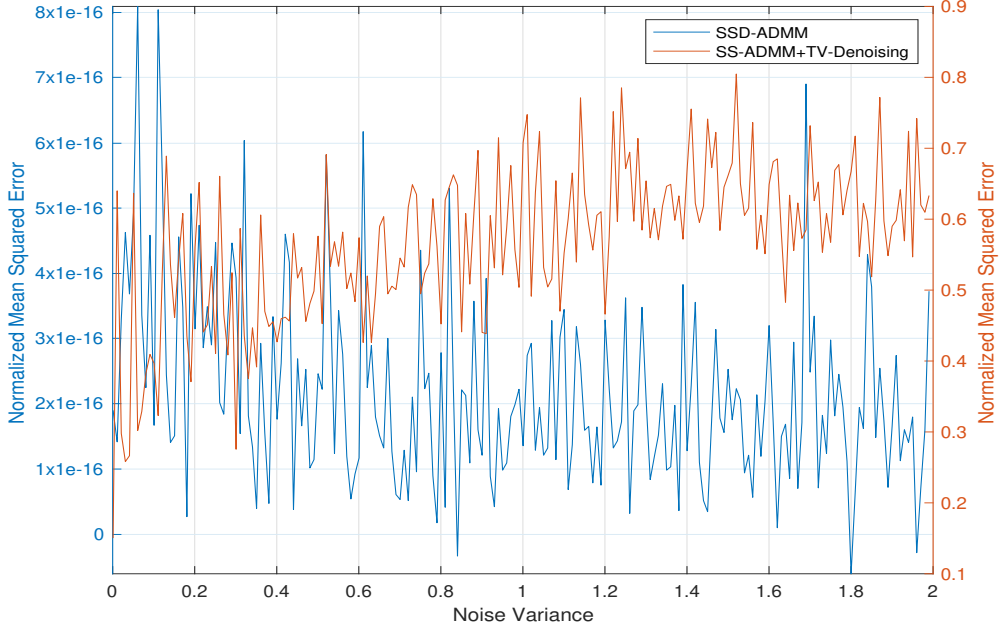


Figure 3.7: NMSE Vs $\text{var}(\Delta Y)$

from eight healthy subjects in a previously published study [78]. In order to evaluate how cortex-muscle interactions might vary over the 5 second data epoch in relation to the stimulus, we employed a half-overlapping sliding window of 500ms duration to estimate causality over successive time intervals, as is typically performed in motor neuroscience studies [70]. The same approach was applied with each method being evaluated as indicated in [40], [39]. Figure 3.8 shows the resulting Granger causality over the 5 second data epoch, as detected using our proposed algorithms SS-ADMM and SSD-ADMM in comparison with the traditional Granger causality [70].

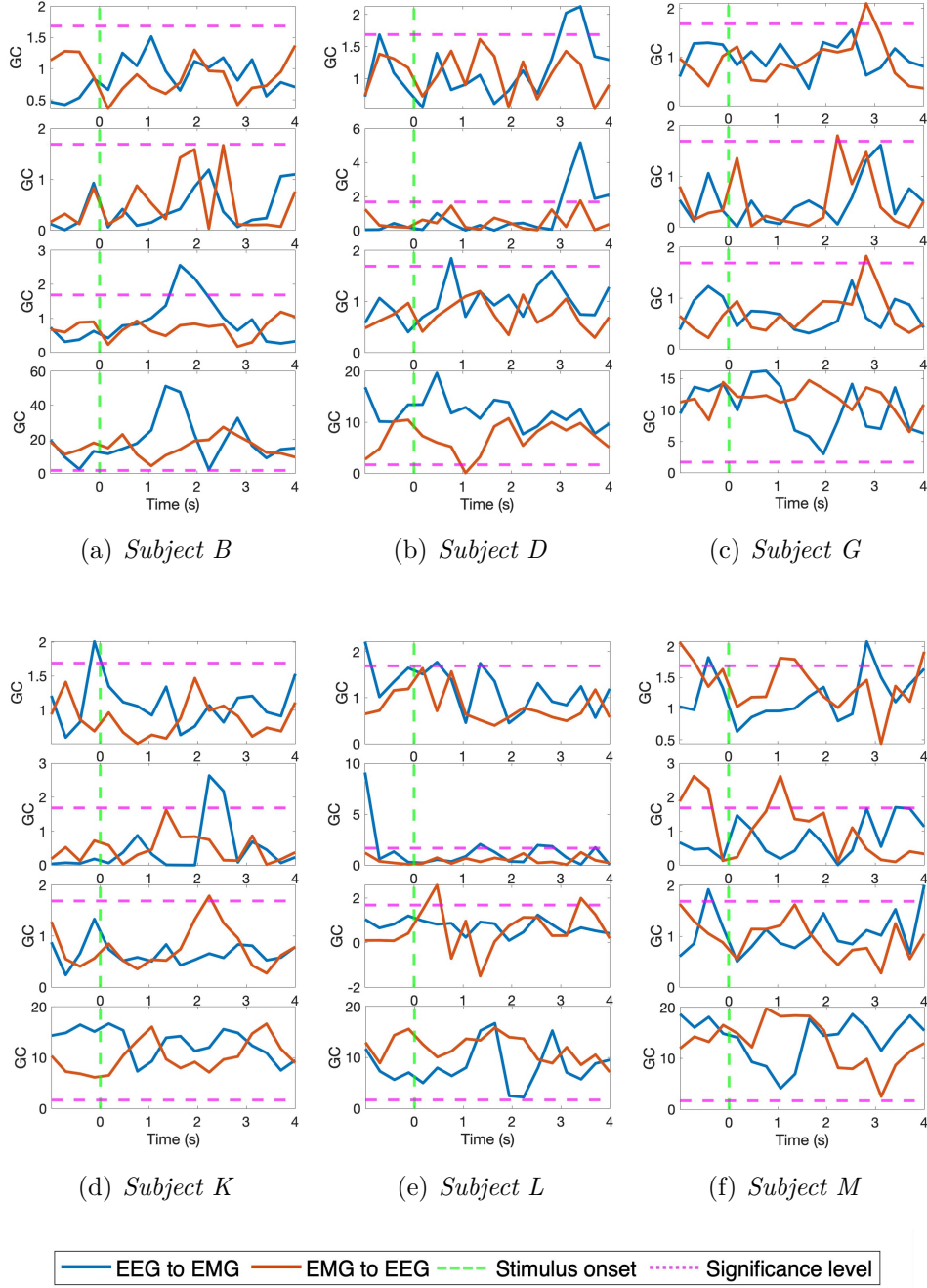


Figure 3.8: First row: Traditional GC [70], Second row: VAR (BIC+Wiener filtering [111]), Third row: SS-ADMM, Fourth row: SSD-ADMM

As another comparative approach, we utilized a Wiener filter [111] as a pre-processing step to effectively denoise the signals, after which we fitted a VAR model and employed the Bayesian Information Criterion (BIC) for model order identification. These results for HAMPD are compared in Figure 3.8 .

Granger causality computed from all four algorithms across the 5 second epoch is

shown for eight subjects. An F-test applied with 95% significance level tests the null hypothesis that the first time series does not Granger-cause the second. In Figure 3.8, the vertical and horizontal lines represent stimulus onset and critical values of the F-test, respectively. In this active task we expect causality between brain and muscles, so the observed pre-stimulus causality in some subjects in Figure 3.8, is not unanticipated. The results in Figure 3.8 demonstrate that Granger causality detection has been significantly improved by modelling the measurement noise. The causality from sEMG to EEG signals has been discovered in many subjects using the proposed method SSD-ADMM that has not been discovered earlier using the other approaches.

3.6.3 Evaluation for DLGC

In this section, we perform a comparative study to investigate the performance of our proposed method DLGC against existing methods on real data. All experiments were performed in PyTorch 2.1.2 with CUDA version 11.2, GPU Tesla T4, 15 GB RAM, and Windows 11 operating system. The proposed method is tested on physiological data HAMPD collected from healthy subjects in a previously published study [78]. In our experimental settings, a Daubechies wavelet dictionary ‘db4’ is used to extract the innovation process, and γ is set to 70%. The level of wavelet decomposition is set at 3. We set the model order $m = 15$ and $T = 495$. The learning rate for dictionary D is $\eta = 10^{-3}$ and sparsity level $\lambda = 0.1$.

It is important to acknowledge that our method comprises several parameters, and the adjustment of hyperparameters can be accomplished through cross-validation or alternative techniques to achieve optimal estimates. Figure 3.9 shows that our method successfully disentangles the observed signal into autoregressive component, the innovation process, and the measurement noise.

In Figure 3.10, we present a detailed comparison of the performance of our proposed DLGC algorithm with two alternative methods: the traditional Blockwise Granger Causality approach [70] and the recently introduced SS-ADMM method

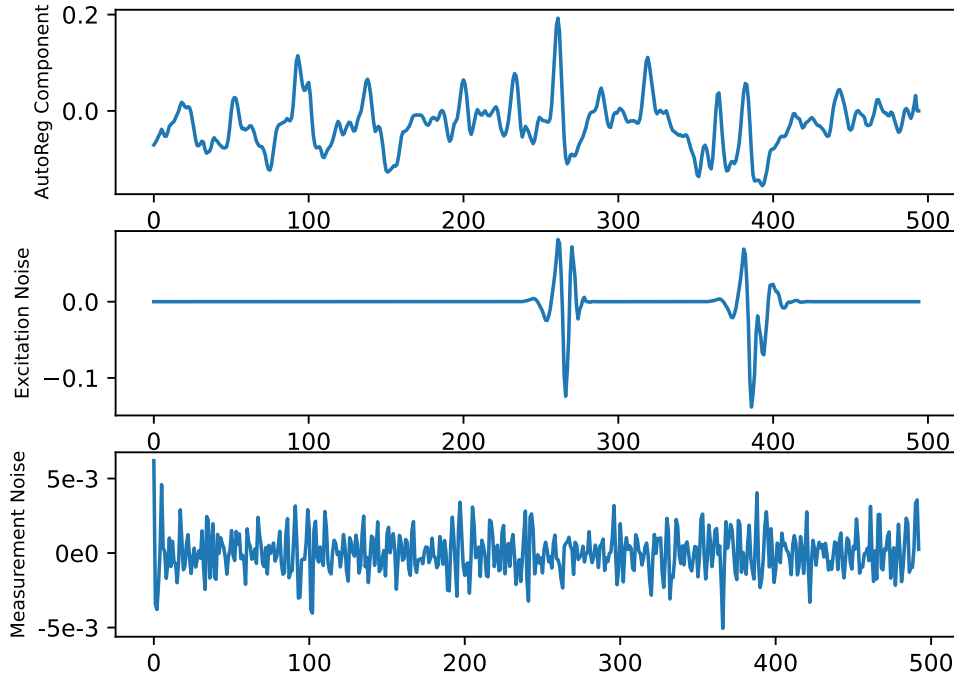


Figure 3.9: EEG signal components for subject D

[1]. The comparison involves Granger causality estimates computed by all three algorithms across a 5-second epoch for seven healthy subjects. An F-test with a 95% significance level is applied to test the null hypothesis that the first time series does not jointly Granger-cause the second. The results consistently show that the DLGC algorithm detects higher levels of significant causality compared to the other methods, highlighting its improved sensitivity in identifying causal relationships.

Additionally, in Figure 3.11, we compare the performance of the sparse coding optimization achieved using PDOM with the traditional Proximal Gradient (PG) method. Using the same initialization and hyperparameter values for both methods, PDOM demonstrates superior performance by achieving a lower objective function value and producing a sparser solution.

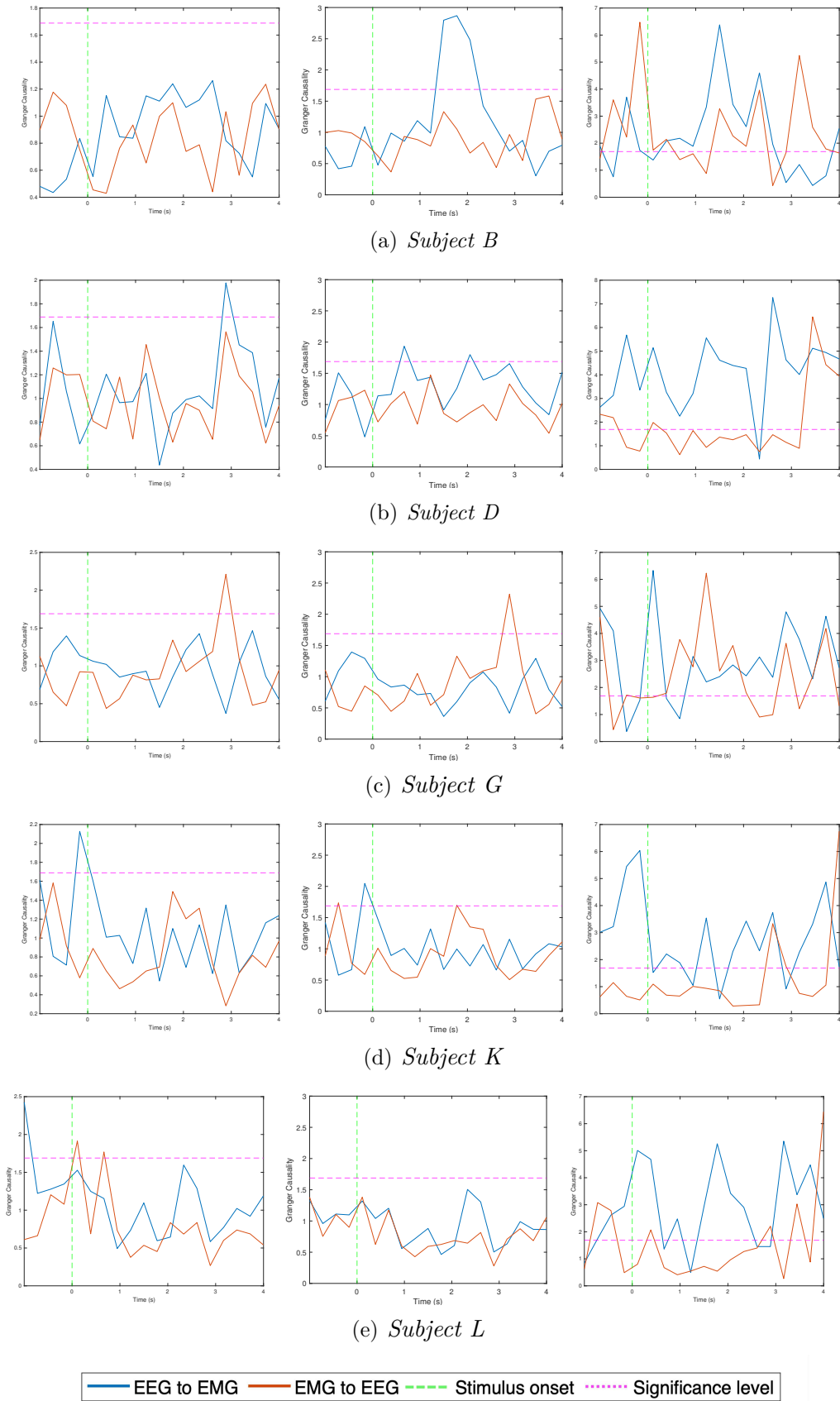


Figure 3.10: First col: Traditional GC [70], Second col: SS-ADMM, Third col: DLGC.

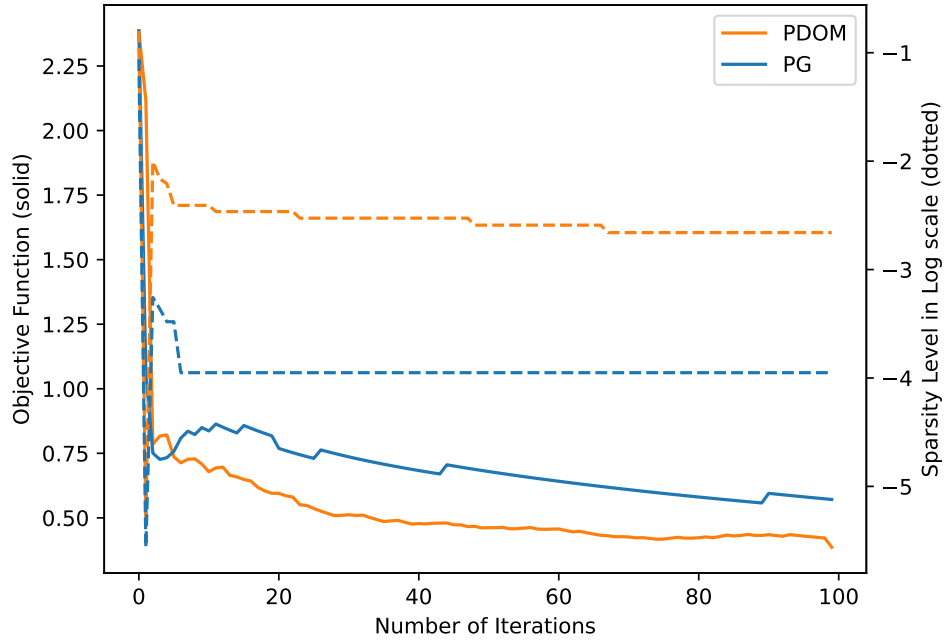


Figure 3.11: Comparative performance analysis of PDOM and PG

This suggests that PDOM identifies a potentially better optimal point, showcasing its efficiency and effectiveness in sparse coding tasks.

The figure provides a detailed comparison over 100 iterations, illustrating two key metrics: the objective function value (solid lines) and sparsity level (log scale, dotted lines). The objective function value, which represents the optimization goal, decreases more rapidly for PDOM than for PG. PDOM converges to a significantly lower value, indicating a more effective minimization of the objective function. In contrast, PG shows slower convergence and stabilizes at a higher objective function value, suggesting suboptimal performance relative to PDOM.

In terms of sparsity, PDOM maintains a consistently higher sparsity level throughout the iterations, as evidenced by the lower values on the logarithmic scale of the dotted orange line. This reflects PDOM's stronger ability to enforce sparse solutions, which are critical in high-dimensional settings for achieving interpretable and generalizable results. Conversely, the sparsity level achieved by PG is notably lower, underscoring its relative limitations in promoting sparsity.

Hence, the results in Figure 3.11 demonstrate the clear advantages of PDOM over PG in sparse coding tasks. PDOM not only achieves faster convergence and a

better optimization outcome but also enforces stronger sparsity, making it a more effective and reliable choice for problems requiring sparse representations.

3.7 Insights and Implications

In this chapter, we addressed a fundamental challenge in neural signal processing: the detection of cortico-muscular interactions amid noise contamination. Two distinct noise sources affect these recordings: measurement noise from environmental interference and equipment limitations, and physiological noise inherent to neural signals. While measurement noise can be partially mitigated through shielding and filtering during data acquisition, physiological noise remains an intrinsic component of the signal.

Our proposed signal processing method directly accounts for both noise types, revealing previously masked cortico-muscular interactions in healthy subjects. The key finding emerges from the comparison between traditional and enhanced approaches: while standard Granger Causality (GC) failed to detect significant causal relationships, our noise-aware method successfully identified bidirectional causal interactions both pre- and post-stimulus, a pattern expected in healthy subjects but often missed by conventional analyses.

These results, particularly evident in Figure 3.10, demonstrate the DLGC algorithm’s superior performance in detecting Granger causality above critical values. During active tasks, we expect to observe causality between brain and muscles in healthy subjects throughout the experiment. However, established methods like blockwise GC and SS-ADMM often yielded sub-threshold or transiently significant results. In contrast, DLGC consistently detected bidirectional causality above critical values across both temporal phases, directly attributable to its explicit noise handling.

While two important limitations warrant consideration, their impact on our findings is constrained by several methodological safeguards. First, regarding the absence of ground truth Granger causality in EEG and EMG recordings, we mitigate

this limitation through convergent validation across multiple subjects and tasks. The consistent emergence of expected physiological patterns, specifically the bidirectional causality in healthy subjects during active tasks, aligns with established neurophysiological principles and provides strong circumstantial support for our method’s validity.

Second, while classical Granger causality’s reliance on finite-order autoregressive models introduces potential biases [112, 29, 6, 5], our analysis focuses on short-term causal relationships within well-defined temporal windows, where these models demonstrate robust performance. The temporal specificity of our investigation, combined with the clear physiological basis for expected causal patterns, suggests that any model-induced biases would be insufficient to invalidate our core findings. Moreover, the comparative improvement in causality detection across subjects provides internal validation independent of absolute model accuracy.

Additionally, our method’s enhanced noise handling directly addresses a primary source of uncertainty in causal estimation. By explicitly modeling both measurement and physiological noise, we reduce the confounding effects that typically challenge causal inference in biological systems. This noise-aware approach strengthens our confidence in the detected causal relationships, even in the absence of absolute ground truth validation.

This work implies that previous studies likely underestimated cortico-muscular interactions by inadequately addressing noise effects. This finding necessitates a reexamination of established results in the field and emphasizes the importance of comprehensive noise handling in neural signal analysis. Our method provides a concrete step toward more accurate characterization of neural interactions, though further validation through diverse experimental paradigms remains essential.

Chapter 4

Towards Interpretable Non-linear Approaches

In this chapter, we transition from linear models to more expressive non-linear models, broadening the scope for better understanding complex neural systems. While linear approaches have formed the backbone of neuroscientific analysis, they face significant limitations in modeling the intricate interactions between cortical and muscular activity. These limitations become particularly clear when attempting to model the dynamic and non-stationary nature of neural signals. To better navigate these challenges, we explore a diverse range of methods that extend beyond the traditional linear models.

Figure 4.1 presents a taxonomy that categorizes causal estimation methods across five critical dimensions: VARs, State Space Models, Random Forests, Neural Networks, and Bayesian Networks. This framework illustrates the methodological progression from deterministic, linear models toward more flexible, non-linear approaches that better capture the complexities of neural communication. This transition reflects a clear shift in how we approach the estimation of causal relationships, from rigid parametric assumptions to more adaptive, data-driven methods.

Vector Autoregressive Models and State Space Models, positioned within the domain of model-driven, deterministic, and linear frameworks, are heavily reliant on

strict parametric assumptions. While they have been foundational in the analysis of temporal dependencies, they often fall short in capturing the full range of neural dynamics. This is especially problematic when working with non-stationary signals or complex temporal dependencies inherent in cortico-muscular communication. These models, by nature, are limited in their ability to account for intricate, time-varying relationships that are commonly observed in real-world neural data.

To overcome these limitations, we turn to non-linear approaches that are more adept at handling such complexities. Among these, Long Short-Term Memory (LSTM) networks and the Reservoir Computing framework stand out due to their ability to process temporal sequences effectively. LSTMs, with their gated architecture, excel at capturing long-range dependencies in neural signals, making them well-suited for modeling the intricate communication between the brain and muscles. Similarly, Reservoir Computing leverages recurrent dynamics in a fixed architecture to efficiently process sequences without the need for extensive training. Both methods demonstrate a powerful capability to model non-linear interactions, yet they are often criticized for their lack of interpretability, which presents a significant challenge in neuroscience research, where understanding the underlying mechanisms is crucial.

To address the interpretability challenge of these non-linear models, we propose a novel hybrid approach that combines the power of Random Forests with feature selection via Mutual Information Neural Estimation (MINE) [49]. This approach, situated at the intersection of deterministic and non-linear methods as illustrated in Figure 4.1, leverages the algorithmic transparency of Random Forests, while incorporating MINE-based feature selection to enhance the detection of relevant features in complex, non-linear relationships. MINE, by quantifying statistical dependencies in data through neural estimation, enables the identification of crucial features that contribute to understanding the underlying neural processes, providing a bridge between computational power and interpretability. This hybrid methodology addresses a critical gap in non-linear modeling, where increased model complexity often comes at the expense of explanatory clarity. By integrating transparent feature

	Vector Autoregressive Models	State Space Models	Random Forrest	Neural Networks	Bayesian Networks
Model-driven	✓	✓			✓
Data-driven			✓	✓	
Deterministic	✓	✓	✓	✓	
Probabilistic					✓
Linear	✓	✓			
Non-linear			✓	✓	✓

Figure 4.1: Overview of Linear and Non-linear methods

selection techniques into a non-linear framework, we provide a more interpretable approach to causal estimation, advancing beyond the traditional linear-nonlinear dichotomy. This allows researchers to explore neural motor control mechanisms with greater confidence in the model’s explanatory power, making it a valuable tool for clinical applications.

In the following sections, we provide a detailed exploration of these methods, highlighting both their strengths and the challenges they present. A comparative analysis of existing causality frameworks is provided in Table 4.1, offering a comprehensive overview of key causal inference techniques in time-series analysis. These frameworks are evaluated across five technical dimensions: parameter estimation, temporal dependency, causal criterion, uncertainty quantification, and model identifiability.

This table highlights how each framework handles the core elements of causal time-series modeling:

First, parameter estimation ranges from closed-form solutions in VARs to sampling-based inference in Bayesian Networks, with neural nets and state-space models using iterative numerical optimization. Second, temporal dependencies are explicit and linear in VARs, implicit in engineered features for Random Forests, embedded in latent states for state-space methods, and captured via memory cells in recurrent

Table 4.1: Comparative Analysis of Time Series Causal Methods

Property	VARs	State Space	Random Forest	Neural Networks	Bayesian Networks
Parameter Estimation	MLE/OLS with order p	EM Algorithm/ Kalman	CART with bootstrapping	SGD with backprop	MCMC/VI inference
Temporal Structure	Fixed p -lag dependency	Hidden Markov Process	Ensemble time windows	LSTM/GRU cells	Dynamic BN evolution
Causal Framework	Granger/ Transfer Entropy	State transitions	Feature importance	Gradient attribution	D-separation
Uncertainty Metrics	Asymptotic variance	Innovation covariance	Out-of-bag error	Dropout variance	Posterior credible intervals
Identifiability	Order selection criterion	Observability matrix rank	Feature correlation threshold	Layer-wise relevance	Markov equivalence class

networks. Third, causality is formalized via statistical tests (Granger), information-theoretic measures (transfer entropy), attribution methods (gradients, permutation), and graphical criteria (d-separation). Fourth, uncertainty quantification moves from analytic standard errors to bootstrap intervals, particle-filter credible bands, Monte Carlo dropout, and full Bayesian posterior intervals. Finally, identifiability considerations vary: classical order-selection in VARs, linear-systems tests in state-space, non-uniqueness under collinearity in ensembles, weight-symmetry issues in deep nets, and equivalence classes in graphical models.

These trade-offs, between statistical tractability, modeling flexibility, and inferential rigor, must guide the choice of method for any given causal-inference task in neural and cortico-muscular time-series data. Subsequent sections will explore these trade-offs in concrete applications.

4.1 Advanced Non-linear Approaches

Recent methodological innovations in computational frameworks have fundamentally transformed our capacity to interrogate neural mechanisms. The emergence of deep learning architectures represents a paradigmatic shift in modeling physiological systems, particularly in capturing non-linear dynamics and complex temporal dependencies inherent in neural processes.

Recurrent neural architectures, from classical RNNs to advanced LSTM and GRU variants, have enabled robust modeling of temporal dependencies in high-dimensional neural data [132]. These frameworks transcend traditional time-series analysis by incorporating adaptive memory mechanisms that selectively preserve or attenuate information flows, crucial for capturing the varying temporal scales characteristic of neural communication. The gated architecture of LSTMs provides a theoretically grounded approach to managing the vanishing gradient problem while maintaining temporal coherence in long-range dependencies [136].

Complementing these temporal modeling capabilities, convolutional architectures have demonstrated remarkable efficacy in extracting hierarchical spatial features from neural signals. The integration of CNNs with recurrent architectures has given rise to hybrid frameworks that simultaneously capture spatiotemporal dynamics [132]. These hybrid models leverage the locality and translation invariance properties of convolution operations while preserving the temporal dependency structures essential for understanding neural sequence dynamics.

The probabilistic paradigm offers a fundamentally different but complementary perspective through its rigorous treatment of uncertainty quantification [114]. Bayesian networks and state-space models provide a theoretical framework for decomposing complex cortico-muscular coupling into interpretable components while explicitly accounting for measurement noise and system variability. These approaches enable formal hypothesis testing about causal relationships in neural circuits through their ability to represent conditional independence structures.

The methodological implications extend beyond statistical accuracy to fundamental questions in neuroscience, enabling novel insights into neural plasticity mechanisms [12], motor adaptation processes, and pathological dysfunction patterns. However, their application necessitates careful consideration of critical methodological challenges: model selection requires balancing computational tractability against interpretability, while validation strategies must account for both the high dimensionality and complex temporal structure inherent in physiological data [87].

The combination of these advanced approaches has become instrumental in modern neuroscience, offering frameworks for decoding neural system complexity while maintaining the methodological rigor necessary for scientific inference. This convergence marks a significant advance in our ability to understand and characterize the intricate dynamics of neural systems, while simultaneously raising important questions about model interpretability and validation in high-dimensional biological systems.

4.1.1 Long Short-Term Memory networks

Long Short-Term Memory (LSTM) networks have emerged as powerful tools for causal inference in time series data, offering multiple frameworks for detecting and quantifying causal relationships [2], [31], [137]. Their ability to capture long-term dependencies and non-linear interactions makes them particularly suitable for complex causal analysis.

To understand how LSTMs facilitate causal inference, it's important to examine their fundamental implementation. The most straightforward application of LSTMs in causality estimation involves comparing two predictive models i.e. restricted and full model. Within this framework, the LSTM architecture processes the input sequences through its gates (input, forget, and output) to generate predictions. The difference in predictive performance between these models, typically measured through metrics like mean squared error or likelihood ratios, indicates potential causal relationships.

Building upon this basic structure, advanced LSTM frameworks sometimes incorporate latent confounding variables to address hidden common causes. These models extend the basic framework to:

$$P(Y_{t+1}|Y_{1:t}, X_{1:t}, Z_{1:t}), \quad (\text{full model}), \quad P(Y_{t+1}|Y_{1:t}, Z_{1:t}), \quad (\text{restricted model}), \quad (4.1)$$

where $Z_{1:t}$ represents latent confounders. To accommodate these additional variables, the LSTM architecture is augmented with:

$$h_t = \text{LSTM}([y_t, x_t, z_t], h_{t-1}), \quad (4.2)$$

where h_t represents the hidden state incorporating all variables.

Beyond static causal relationships, LSTMs can also model time-varying causal relationships through adaptive architectures:

$$\alpha_t = g(h_t^{(X)}, h_t^{(Y)}), \quad (4.3)$$

where α_t represents a time-dependent causal strength parameter, and $g(\cdot)$ is a neural network mapping LSTM hidden states to causal coefficients. This enables detection of dynamic causal relationships that change over time.

To ensure the reliability of these causal inferences, robust statistical testing methods are essential. Statistical testing in LSTM-based causality frameworks often employs likelihood ratio tests between the full and restricted models [71]. These tests can be augmented with bootstrap procedures to obtain confidence intervals on causal effect estimates. Permutation tests [86], where the temporal ordering of potential causal variables is randomized, provide an additional tool for assessing the significance of discovered causal relationships.

The implementation of these frameworks requires careful attention to validation procedures. Model validation requires careful consideration of temporal dependencies

in the data. Standard cross-validation must be modified to respect the temporal ordering of observations, typically through rolling window schemes or blocked cross-validation approaches. The choice of architecture and hyperparameters can significantly impact causal inference results, necessitating robust validation procedures that account for model stability and reliability. Thus, this framework can handle multivariate time series, non-stationary processes, and complex temporal dependencies commonly encountered in real-world applications [2]. Building upon the strengths of LSTM networks in capturing temporal dependencies, we now shift our focus to another non-linear approach, Reservoir Computing, which offers distinct advantages in handling complex dynamical systems. While LSTMs are well-regarded for their ability to manage long-range dependencies, Reservoir Computing presents a complementary strategy for modeling non-linear interactions in time series,

4.1.2 Reservoir Computing for Non-linear Dynamics

Reservoir computing (RC) provides a powerful framework for modeling nonlinear interactions in time series and is particularly suited for Granger causality analysis [138], [30], [28]. A time series $y(t)$ can be modeled using a nonlinear autoregressive process as follows:

$$y(t) = f(y(t-1), y(t-2), \dots, y(t-m)) + \varepsilon_y(t), \quad (4.4)$$

where f is an unknown nonlinear function that captures the dependence of $y(t)$ on its m -lagged historical values, and $\varepsilon_y(t)$ represents the residual noise. The reservoir serves as a dynamic kernel that projects the input time series into a high-dimensional state space through its recurrent connections. This projection is analogous to the feature space mapping in kernel methods, but with RC, the mapping is performed explicitly through the reservoir dynamics rather than implicitly through a kernel function [102]. The reservoir's nonlinear activation functions and recurrent connections create a rich representation of the temporal dependencies in the input data, making it easier to

detect causal relationships that might be difficult to identify in the original input space.

To understand how RC facilitates causal inference, it is important to first examine the fundamental concept of nonlinear Granger causality. This can be defined using two predictive models:

$$y(t) = f(y_{t-m:t-1}) + \varepsilon_y(t), \quad (4.5)$$

$$y(t) = g(y_{t-m:t-1}, x_{t-m:t-1}) + \varepsilon_{y|x}(t). \quad (4.6)$$

If the variance of the residual noise $\text{Var}(\varepsilon_y(t))$ is significantly larger than $\text{Var}(\varepsilon_{y|x}(t))$, then $x(t)$ is said to Granger-cause $y(t)$. Building upon this foundation, reservoir computing extends this framework by utilizing a dynamical reservoir to capture the memory and non-linear relationships in the time series. The key advantage of RC lies in its ability to implicitly model complex temporal dependencies through its reservoir states. In RC, the output $y(t)$ is predicted using the input $x(t)$ and its historical information mixed within the reservoir states:

$$y(t) = f(x(t-1), x(t-2), \dots, x(t-m)) + \varepsilon(t), \quad (4.7)$$

where the order m is implicitly determined by the memory capacity of the reservoir.

A key metric in RC-based causality analysis is the causality index (CI), which quantifies the strength of potential causal relationships [126]. The CI is typically computed by comparing prediction errors with and without the potential causal variable:

$$\text{CI} = 1 - \frac{\text{RMSE}(x, y)}{\text{RMSE}(y)}. \quad (4.8)$$

A positive CI value indicates that including the potential causal variable improves prediction accuracy, suggesting a causal relationship. The flexibility of reservoir computing (RC) architectures enables diverse approaches to causality estimation, ranging from single-reservoir frameworks to advanced multi-reservoir configurations

[67]. These architectures can be adapted to characterize multifaceted causal relationships, encompassing direct effects, indirect pathways, and higher-order interactions between variables [35, 120]. The reservoir’s rich dynamics serve as a non-linear expansion of the input space, potentially capturing complex temporal dependencies that might elude traditional time series methods.

Statistical validation constitutes a critical component in RC-based causality detection frameworks. Contemporary approaches typically employ rigorous surrogate data testing methodologies to assess the statistical significance of detected causal relationships. These validation procedures generate null distributions through perturbation of the original time series, enabling the computation of empirical significance thresholds while preserving essential temporal characteristics of the underlying processes. This statistical framework provides a robust foundation for distinguishing genuine causal relationships from spurious correlations that might arise from the high-dimensional nature of reservoir states [103].

The integration of RC architectures with established causal inference frameworks has led to notable methodological advances. Recent work has demonstrated the efficacy of combining reservoir computing with existing methods, extending classical approaches to handle non-linear and non-stationary dynamics characteristic of complex systems. These hybrid approaches leverage the computational efficiency of reservoir computing while maintaining the theoretical rigor of established causal inference methodologies.

In the next section, we introduce interpretable methodologies that can effectively capture non-linear causal relationships while maintaining analytical transparency.

4.2 Interpretable Random Forest-Based Coupling Analysis

Recent advances in machine learning have demonstrated the exceptional capability of non-linear methods, particularly neural networks and tree-based regressors, in

discovering and modeling complex patterns within data [119], [106], [88]. In the context of autoregressive modeling, contemporary approaches to investigating non-linear *Granger*-causal interactions frequently employ sparsity-inducing regularization within neural network frameworks [115], [110]. This methodology promotes weight group sparsity, facilitating the identification of genuinely relevant lagged inputs and potentially revealing underlying *Granger*-causal structures.

However, the inherent complexity that enables such non-linear models to achieve superior performance simultaneously presents a significant challenge: their limited interpretability. This limitation is particularly problematic in scientific and medical contexts, where it is commonly referred to as the *issue of clinical interpretability* [79]. For instance, in dystonia research, merely demonstrating that a neural network can accurately model brain-muscle interactions is insufficient; understanding the specific interactions driving the model’s predictions is essential for advancing therapeutic interventions. The necessity of aligning model outputs with established medical knowledge underscores the importance of addressing this interpretability challenge.

Recent efforts to overcome these limitations have shown promise. For example, the authors in [72] have integrated self-explaining neural networks with sparsity-inducing penalties to infer *Granger*-causality, advancing the field of interpretable causal learning. Nevertheless, neural networks, particularly Recurrent Neural Networks (RNNs), present significant practical challenges, including training complexity and substantial computational resource requirements. Furthermore, these methods often exhibit high sensitivity to tuning parameters, potentially leading to overfitting issues [54].

To address these methodological constraints, we propose a novel approach that synthesizes the representational power of non-linear models with the interpretability advantages of linear frameworks. Our method introduces an interpretable model for estimating *Granger*-causal interactions from simultaneously recorded EEG and EMG signals, leveraging gradient-boosted random forests to prioritize the most informative features during regression analysis. We employ mutual information as a preliminary

analytical tool to detect non-linear dependencies that may indicate potential Granger causality, while incorporating both time-lagged and frequency-specific features to enhance the regression process.

The resulting algorithm, which we term INterpretable Feature Representations for Granger Causality (INFR-GC), effectively captures the complex dynamics inherent in brain-muscle interactions through comprehensive feature integration. Random Forests, comprising ensembles of decision trees, offer enhanced interpretability compared to neural networks while maintaining computational efficiency. For mutual information estimation, we utilize a neural mechanism previously proposed in [49], which informs our regression model’s feature selection process.

To validate our approach, we employ rigorous statistical methodology, comparing results obtained with and without coupled time series data using the Wilcoxon signed-rank (WSR) test as detailed in [33]. Furthermore, we utilize Shapley Additive Explanations (SHAP) values [68] to identify and quantify the most influential features driving model outputs and to understand feature interactions, thereby facilitating more informed interpretations of the results.

While our method incorporates a compact two-layer neural network for mutual information estimation, it is specifically designed to minimize computational demands compared to larger neural architectures, achieving an optimal balance between model complexity and computational efficiency. This approach represents a significant advancement in interpretable causal analysis, particularly for neurophysiological applications where both accuracy and interpretability are crucial considerations.

4.2.1 Mathematical Formulation

Consider $\mathbf{x} \in \mathbb{R}^{L \times 1}$ is the predictor signal and $\mathbf{y} \in \mathbb{R}^{L \times 1}$ is the response signal i.e. $\mathbf{x} \neq \mathbf{y}$ and $\mathbf{x}, \mathbf{y} \in \{\text{EEG signals, sEMG signals}\}$, then the response $\mathbf{y}$ can be approximated by unrestricted and restricted models:

$$\hat{\mathbf{y}}_{unr}^{(t)} \approx \mathcal{M}([\mathbf{y}^{(t-1)} \mathbf{x}^{(t-1)}]; \tau), \quad \hat{\mathbf{y}}_{res}^{(t)} \approx \mathcal{M}(\mathbf{y}^{(t-1)}; \tau),$$

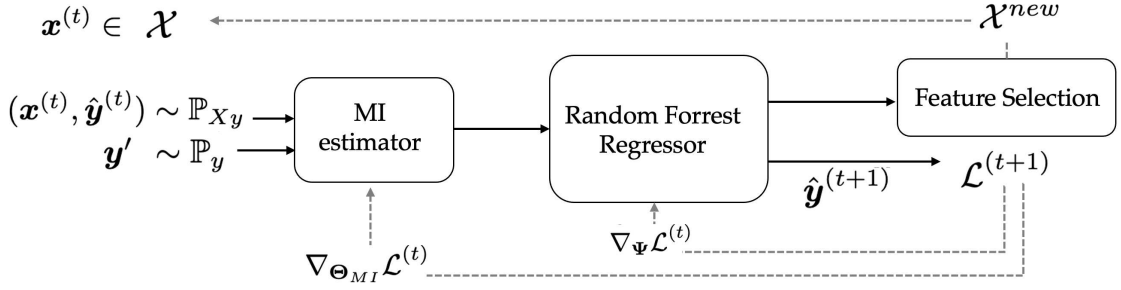


Figure 4.2: Overview of INFR-GC model

where $\mathcal{M}(\cdot; \tau)$ is a non-linear function parameterized by τ . We decompose both EEG and sEMG signals into six frequency bands i.e. Gamma (30-70 Hz), Beta (14-30 Hz), Mu (10-13 Hz), Alpha (8-13 Hz), Theta (4-8 Hz), and Delta (0.5-4 Hz). We then construct the feature matrices $\mathbf{X}^{(t)}, \mathbf{Y}^{(t)} \in \mathbb{R}^{L \times m+6}$ from time-lagged and frequency domain features as $\mathbf{X}^{(t)} = [\mathbf{x}^{(t-1)}, \dots, \mathbf{x}^{(t-m)}, \mathbf{x}_\delta, \mathbf{x}_\theta, \mathbf{x}_\alpha, \mathbf{x}_\mu, \mathbf{x}_\beta, \mathbf{x}_\gamma]$, $\mathbf{Y}^{(t)} = [\mathbf{y}^{(t-1)}, \dots, \mathbf{y}^{(t-m)}, \mathbf{y}_\delta, \mathbf{y}_\theta, \mathbf{y}_\alpha, \mathbf{y}_\mu, \mathbf{y}_\beta, \mathbf{y}_\gamma]$ where m is the relevant predictive history of the response signal. The unrestricted and restricted models use feature sets $\mathcal{X}_{unr} \in \{\mathbf{X}, \mathbf{Y}\}$ and $\mathcal{X}_{res} \in \{\mathbf{Y}\}$, respectively. The objective is to find the future values of response $\mathbf{y}^{(t)}$ along with the set of features that are most relevant for the target series. Restricted and unrestricted models are then compared using Wilcoxon signed-rank test to compute *Granger*-causality from $\mathbf{x}$ to $\mathbf{y}$.

In order to estimate mutual information (MI) in a reliable way we employ a two-layer neural network $f_{\Theta_{MI}}(\mathbf{g}, \mathbf{z})$ using the approach proposed [49], to leverage Donsker-Varadhan representation of the MI estimate [26] as follows:

$$I = \frac{1}{N} \sum_{n=1}^N f_{\Theta_{MI}}(\mathbf{g}, \hat{\mathbf{z}}) - \log\left(\frac{1}{N} \sum_{n=1}^N e^{f_{\Theta_{MI}}(\mathbf{g}, \mathbf{z}')} \right), \quad (4.9)$$

where the first term is the average of the MI Estimator's output on the joint distribution $(\mathbf{u}, \hat{\mathbf{v}}) \sim \mathbb{P}_{Xy}$ and the second term is the log of the average of the exponentiated outputs on the marginal distribution $\mathbf{y}' \sim \mathbb{P}_y$. To maximize the mutual information, the loss function has the negative of I term. In the above equation, the marginal distribution is constructed by dropping $\mathbf{u}$ from the joint distribution i.e. $(\mathbf{u}, \mathbf{v}') \sim \mathbb{P}_y$. The MI estimator then guides the gradient-boosted

random forest model to minimize the overall objective function as follows:

$$\begin{aligned} \mathcal{L} = & \frac{1}{N} \sum_{n=1}^N (z_n - \hat{z}_n)^2 - \lambda \left(\frac{1}{N} \sum_{n=1}^N f_{\Theta_{MI}}(\mathbf{g}_n, \hat{z}_n) \right. \\ & \left. + \log \left(\frac{1}{N} \sum_{n=1}^N e^{f_{\Theta_{MI}}(\mathbf{g}_n, z'_n)} \right) \right). \end{aligned} \quad (4.10)$$

The gradient of the loss is computed and backpropagated to update the parameters Θ of MI estimator and Ψ of the random forest regressor. The complete flow of the method is detailed in Figure 4.2. The random forest regressor model employs gradient-boosting to estimate parameters of j -th tree Ψ_j based on the residuals from previous trees and the bootstrap samples.

Algorithm 4 MI-Augmented Autoregressive Forecasting

```

function  $\mathcal{M}(\mathbf{z}, \mathcal{X})$ 
  Initialization: Set  $T, \eta, \lambda$ , and  $\hat{\mathbf{z}}^{(0)} = \text{mean}(\mathbf{z})$ ,  $f_{\Theta_{MI}}(\mathbf{g}, \mathbf{z}) = \text{ReLU}(\mathbf{g}\Theta_1)\Theta_2$ 
  where  $\mathbf{g} \in \mathcal{X}$ .
  while stopping criterion is not met do
    for all  $t = 1, \dots, T$  do
      Step 1: Find  $I = \frac{1}{N} \sum_{n=1}^N f_{\Theta_{MI}}(\mathbf{g}, \hat{\mathbf{z}}^{(t-1)})$ 
       $-\log(\frac{1}{N} \sum_{n=1}^N e^{f_{\Theta_{MI}}(\mathbf{g}, z'_n)})$ 
      Step 2: Compute residuals  $\mathbf{a}^{(t)} = \hat{\mathbf{y}}^{(t-1)} - \mathbf{y}$ 
       $+ \lambda \frac{\partial I}{\partial \hat{\mathbf{z}}^{(t-1)}}$ 
      Step 3: Update  $j$ -th tree parameters by  $\Psi_j^{(t)} \leftarrow$ 
       $\Psi_j^{(t-1)} - \eta \nabla_{\Psi_j} \sum_{i \in \text{Bootstrap}(\mathcal{X})} (\mathbf{a}_i^{(t)} - h(\mathbf{g}_i; \Psi_j))^2$ 
      Step 4 : Update parameters  $\Theta_{MI}$  by adam
      optimizer  $\Theta_{MI}^{(t)} \leftarrow \text{Adam}(\Theta_{MI}^{(t-1)}, \nabla_{\Theta_{MI}} \mathcal{L})$ 
      Step 5: Compute random forest update with  $J$ 
      trees  $\hat{\mathbf{z}}^{(t)} = \hat{\mathbf{z}}^{(t-1)} + \eta \sum_{j=1}^J \alpha_j h(\mathbf{g}; \Psi_j^{(t)})$ 
    end for
    Step 6: Get importance score  $\iota_k$  and SHAP values
     $\phi_k^{(l)}$  for  $k$ -th feature and  $l$ -th sample  $\forall k, l$  and
    update feature set  $\mathcal{X}$ .
  end while
return  $\hat{\mathbf{z}}^{(T)}$ 
end function
    
```

Algorithm 5 INFR-GC : Interpretable Feature Representations for *Granger-Causality*

Input: Training data $\{(\mathbf{x}_n, \mathbf{y}_n)_{n=1}^N\}$

Step 1: Estimate time-lagged and frequency-domain features $\mathcal{X}_{unr} = \{\mathbf{X}, \mathbf{Y}\}$, $\mathcal{X}_{res_y} = \{\mathbf{Y}\}$, and $\mathcal{X}_{res_x} = \{\mathbf{X}\}$

Step 2: Learn restricted models $\hat{\mathbf{x}}_{res} = \mathcal{M}(\mathbf{x}, \mathcal{X}_{res_x})$ and $\hat{\mathbf{y}}_{res} = \mathcal{M}(\mathbf{y}, \mathcal{X}_{res_y})$

Step 3: Learn unrestricted models $\hat{\mathbf{x}}_{unr} = \mathcal{M}(\mathbf{x}, \mathcal{X}_{unr})$ and $\hat{\mathbf{y}}_{unr} = \mathcal{M}(\mathbf{y}, \mathcal{X}_{unr})$

Step 4: Estimate $GC_{x \rightarrow y} = \text{WSR-test}(\hat{\mathbf{y}}_{res}, \hat{\mathbf{y}}_{unr})$ and $GC_{y \rightarrow x} = \text{WSR-test}(\hat{\mathbf{x}}_{res}, \hat{\mathbf{x}}_{unr})$

Step 5: Evaluate feature importance ι_k and SHAP values ϕ_k for all features

Output: $GC_{x \rightarrow y}, GC_{y \rightarrow x}, \iota_k, \phi_k \forall k$

4.2.2 Feature Selection Framework

We propose a comprehensive feature selection framework that integrates multiple complementary approaches to identify the most informative predictors for Granger-causal inference. This framework synthesizes mutual information estimation, random forest-based importance metrics, and SHAP values through an iterative optimization process.

The feature selection process begins with the construction of a candidate feature set $\mathcal{X}$ comprising both temporal and spectral components. Temporal features include time-lagged values up to order p , while spectral features encompass power in conventional frequency bands extracted through wavelet decomposition. This rich feature space captures both linear and non-linear dependencies across multiple temporal and spectral scales. The selection mechanism operates through a three-fold evaluation criterion:

1. **Mutual Information Estimation:** We employ a neural estimator based on MINE to quantify the statistical dependence between each feature and the target variable. This approach captures non-linear relationships through the following optimization:

$$I(X; Y) = \sup_{\Theta_{MI}} \mathbb{E}_{\mathbb{P}_{XY}}[f_{\Theta_{MI}}(\mathbf{g}, \mathbf{z})] - \log(\mathbb{E}_{\mathbb{P}_X \otimes \mathbb{P}_Y}[e^{f_{\Theta_{MI}}(\mathbf{g}, \mathbf{z})}])$$

2. **Random Forest Importance:** The algorithm computes feature importance

scores through the gradient-boosted random forest framework, quantifying the reduction in prediction error attributable to each feature:

$$\text{Importance}(x_i) = \sum_{t=1}^T \sum_{n \in N_t} \Delta \text{Impurity}(x_i, n)$$

where N_t represents the nodes in tree t where feature x_i is used for splitting and T is the total number of trees.

3. SHAP Value Analysis: We incorporate SHAP values to provide a game-theoretic perspective on feature contributions:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)]$$

where N is the set of all feature indices, $S \subseteq N \setminus \{i\}$ is a subset of features not including i , and $v(S)$ is the model prediction when only features in S are known (others are marginalized or imputed). The term $\frac{|S|!(|N| - |S| - 1)!}{|N|!}$ is the Shapley weighting term, which ensures fairness by averaging over all possible feature orderings.

The feature selection algorithm iteratively updates the feature set $\mathcal{X}$ by retaining variables that demonstrate strong predictive power across all three metrics. This is formalized through a composite score:

$$\text{Score}(x_i) = w_1 I(x_i; Y) + w_2 \text{Importance}(x_i) + w_3 \phi_i$$

where w_1 , w_2 , and w_3 are weights determined through cross-validation.

Features are selected if their composite score exceeds a threshold τ :

$$\mathcal{X}_{\text{selected}} = \{x_i : \text{Score}(x_i) > \tau\}$$

This iterative process continues until convergence, defined as stability in the selected feature set across consecutive iterations. The resulting framework ensures robust feature selection by incorporating multiple complementary perspectives on feature

relevance, while maintaining computational efficiency through the streamlined neural architecture of MINE and the inherent parallelizability of random forests.

The selected features then serve as inputs for the subsequent Granger causality analysis, providing a refined and interpretable basis for investigating temporal dependencies in neurophysiological signals. This approach significantly reduces the dimensionality of the feature space while preserving the most informative predictors for causal inference.

4.2.3 Insights on Model Interpretability

To transcend the conventional “black box” limitations of non-linear models, we implement a comprehensive interpretability framework centered on SHAP values. This approach facilitates granular insights into the model’s decision-making process, particularly within the context of neurophysiological signal analysis and *Granger*-causal inference.

The SHAP methodology, grounded in cooperative game theory, provides a unified framework for interpreting model predictions through feature importance attribution. For each feature i , SHAP values $\phi_i(x)$ are computed as:

$$\phi_i(x) = \sum_{S \subseteq \mathcal{F} \setminus \{i\}} \frac{|S|!(|\mathcal{F}| - |S| - 1)!}{|\mathcal{F}|!} [f_x(S \cup \{i\}) - f_x(S)] \quad (4.11)$$

where $\mathcal{F}$ represents the complete feature set, S denotes feature subsets, and f_x represents the model prediction function. This formulation satisfies the fundamental properties of local accuracy, missingness, and consistency, ensuring reliable feature attribution.

Our analysis reveals several key interpretability insights that include temporal dependency patterns, frequency-band contributions, and feature interaction effects. The SHAP value distributions across time-lagged features demonstrate distinctive patterns of temporal influence. We observe that specific time lags consistently exhibit higher attribution values, suggesting temporal windows where brain-muscle coupling

is most pronounced. These patterns are quantified through the Temporal Attribution Profile (TAP):

$$T(\tau) = \mathbb{E}_{x \sim \mathbf{X}}[|\phi_\tau(x)|] \quad (4.12)$$

where τ represents the time lag and the expectation $\mathbb{E}_x$ is taken over the entire dataset $\mathbf{X}$. SHAP analysis of spectral features reveals differential contributions across frequency bands, providing insights into the spectral characteristics of *Granger*-causal relationships. The relative importance of each frequency band β is computed as:

$$R(\beta) = \frac{\sum_{x \in \mathcal{X}_\beta} |\phi(x)|}{\sum_{x \in \mathcal{X}} |\phi(x)|} \quad (4.13)$$

where $\mathcal{X}_\beta$ represents features from frequency band β . The SHAP interaction values Φ_{ij} capture pairwise feature dependencies, revealing complex relationships between temporal and spectral features:

$$\Phi_{ij} = \sum_{S \subseteq \mathcal{F} \setminus \{i,j\}} \frac{|S|!(|\mathcal{F}| - |S| - 2)!}{2(|\mathcal{F}| - 1)!} \Delta_{ij} f_x(S) \quad (4.14)$$

where $\Delta_{ij} f_x(S)$ represents the interaction effect between features i and j . The analysis of local SHAP values enable subject-specific analysis, revealing individualized patterns of brain-muscle coupling that may inform personalized therapeutic approaches.

Through this interpretability framework, we bridge the gap between complex non-linear modeling and clinical utility, providing transparent insights into the mechanisms underlying brain-muscle interactions. The combination of local and global interpretability measures ensures that our model's predictions can be effectively translated into actionable clinical insights while maintaining rigorous statistical foundations.

This approach represents a significant advancement in interpretable machine learning for neurophysiological analysis, offering a framework for understanding complex causal relationships while maintaining the transparency necessary for clinical applications. The integration of SHAP values with domain-specific knowledge enables

both theoretical validation and practical application of our modeling framework.

To evaluate the effectiveness of different causality detection approaches, we compare the performance of two non-linear models: Long Short-Term Memory networks (LSTM) and Reservoir Computing (RC). Both models are particularly well-suited for capturing temporal dependencies in non-linear dynamical systems. The statistical significance of causal relationships is assessed using the Wilcoxon signed-rank test [21]. This non-parametric test is chosen for its robustness and flexibility in handling non-linear relationships. Two key assumptions must be satisfied:

1) **Independence:** To ensure this, we partition the time series into training and test sets, maintaining a gap of m samples between them to prevent data leakage and temporal correlation.

2) **Symmetry:** The differences between paired observations should follow a symmetric distribution. However, this assumption can be relaxed for one-sided tests by employing the concept of *stochastic dominance* [95], which allows us to compare the relative frequencies of negative and positive ranks without requiring strict symmetry.

The testing procedure involves comparing restricted and unrestricted models. For each potential causal relationship, we compute the vector of differences $\mathbf{d}$ between the absolute prediction errors of these models:

$$\mathbf{d} = |\mathbf{y} - \hat{\mathbf{y}}_{res}| - |\mathbf{y} - \hat{\mathbf{y}}_{unr}|,$$

where $\hat{\mathbf{y}}_{res}$ and $\hat{\mathbf{y}}_{unr}$ represent predictions from the restricted and unrestricted models, respectively. The signed ranks $\mathcal{R}$ are then computed by:

$$\mathcal{R}(\mathbf{d}) = \text{Rank}(|\mathbf{d}|) \circ \text{sgn}(\mathbf{d}),$$

where $\circ$ denotes element-wise multiplication. The test statistic W is calculated as:

$$W = \min \left(\sum_{i:d_i>0} \mathcal{R}(\mathbf{d}_i), \left| \sum_{i:d_i<0} \mathcal{R}(\mathbf{d}_i) \right| \right).$$

This approach offers several advantages over traditional Granger causality tests. Unlike the F-test used in linear autoregressive models, the Wilcoxon signed-rank test remains valid for non-normally distributed data, ordinal measurements, and scenarios with heteroscedasticity. A predominance of positive values in $\mathbf{d}$ indicates that the unrestricted model consistently outperforms the restricted model, suggesting a causal relationship.

4.2.4 Comparative Analysis on Lorenz96 Data

The Lorenz96 model is a classical benchmark dataset in dynamical systems analysis, originally introduced by Edward Lorenz to study atmospheric predictability. In our experiments, we use $N = 40$ variables, creating a complex system with multiple non-linear interactions. This dataset serves as an ideal testbed for causality analysis due to its known ground truth relationships and challenging non-linear, chaotic dynamics.

These results are estimated on $N = 40$ variables for Lorenz96 dataset and summarized in Table 4.2. The results indicate that both LSTM and RC models achieve comparable performance, with RC showing slightly better recall and F1 scores. Moreover, our proposed INFR-GC method outperforms both LSTM and RC methods in terms of accuracy and recall. The perfect precision scores suggest that all approaches are highly reliable in identifying true causal relationships, though they may occasionally miss some existing connections as indicated by the recall values.

4.2.5 Experimental Evaluation on Physiological Data

In this section, we present the results of our method applied to real physiological data PMPD, introduced in Chapter 1, that is recorded from healthy subjects and

Table 4.2: Performance Comparison of LSTM, RC, and INFR-GC Models on Lorenz96 Data

Metric	LSTM	RC	INFR-GC
Accuracy	0.7125	0.7125	0.8751
Precision	1.0000	1.0000	1.0000
Recall	0.6406	0.6618	0.8330
F1 Score	0.7810	0.7965	0.9092

dystonia patients. Experiments were conducted using PyTorch 2.1.2 with CUDA 11.2, GPU Tesla T4, 15GB RAM, and Windows 11. The study involved 15 children with dystonia (10 acquired, 5 idiopathic/genetic, including 9 females) and 13 typically developing children (7 females), all aged 12 to 18 years. Participants performed a motor task holding a 15 cm plastic ruler, with mechanical perturbations delivered using an electromechanical tappe as described in [77]. Each 5-second trial began with the stimulus 1.1 seconds after data collection started. After removing noise and artifacts, features were extracted and used to train the model. Each trial, comprising 5118 samples at 1024 Hz, was divided into 300-sample blocks for Granger-causality estimation, with cumulative causality recorded by aggregating errors across blocks. The experimental results reveal distinctive patterns in both feature importance and Granger causality across healthy subjects and dystonia patients.

Figure 4.3 presents the *Granger*-causality estimates obtained from our proposed method. The results show that healthy subjects generally exhibit significant causality values exceeding the critical threshold for the Wilcoxon signed-rank (WSR) test. In contrast, individuals with dystonia display less consistent causality, suggesting altered or diminished predictive relationships in this group.

Figure 4.4 compares the distribution of Granger causality (GC) measures between healthy subjects and patients across three different methodologies: SS-ADMM, Neural GC [115], and INFR-GC. The bidirectional analysis (EEG to sEMG and sEMG to EEG) reveals important asymmetries in causal influence. In the EEG to sEMG direction, healthy subjects consistently show higher GC values across

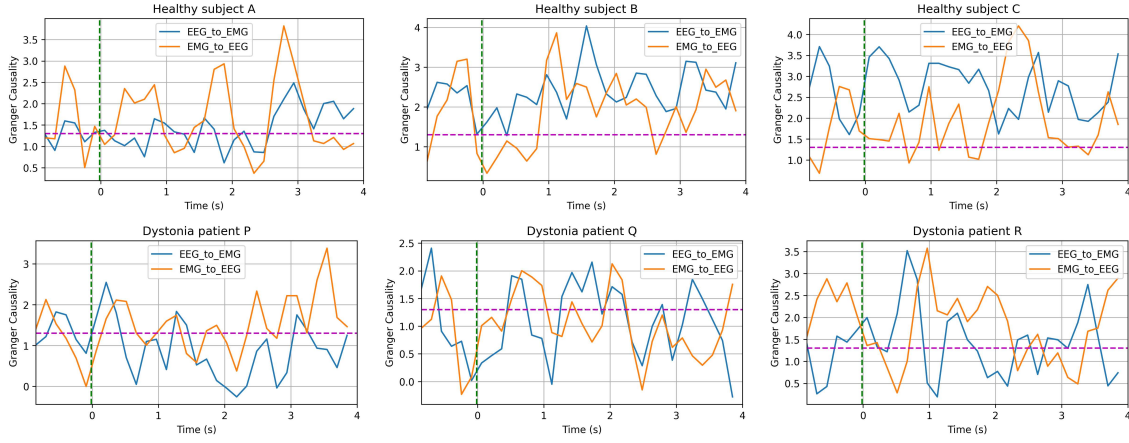


Figure 4.3: *Granger-causality* for healthy subjects and dystonia patients with critical value (pink) and stimulus onset (green)

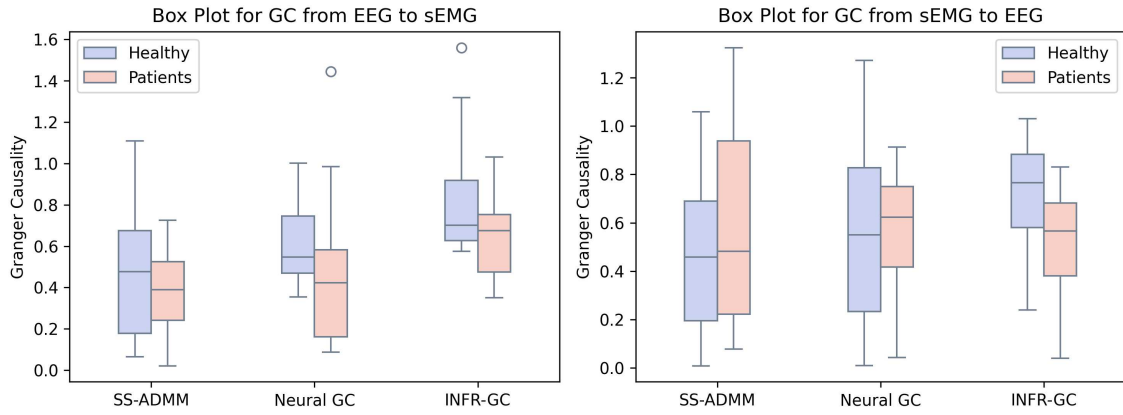


Figure 4.4: Comparison of GC distribution for first 20 trials.

all methods, with INFR-GC demonstrating the most pronounced differences. The reverse direction (sEMG to EEG) shows more variable results but maintains consistent group differences. Notably, INFR-GC produces more tightly clustered distributions compared to other methods, suggesting improved reliability in causal estimation.

Figure 4.5 provides a detailed heatmap visualization of feature importance across different frequency bands for randomly selected subjects (4 dystonia patients and 4 healthy controls). The temporal evolution of feature importance, displayed across 20 time lags, reveals distinct patterns between groups. Healthy subjects demonstrate more consistent and structured patterns of feature importance, particularly in the alpha and beta frequency ranges, indicated by the more organized distribution of higher importance values (shown in yellow and light green). In contrast, dystonia patients exhibit more scattered and irregular patterns of feature importance, suggest-

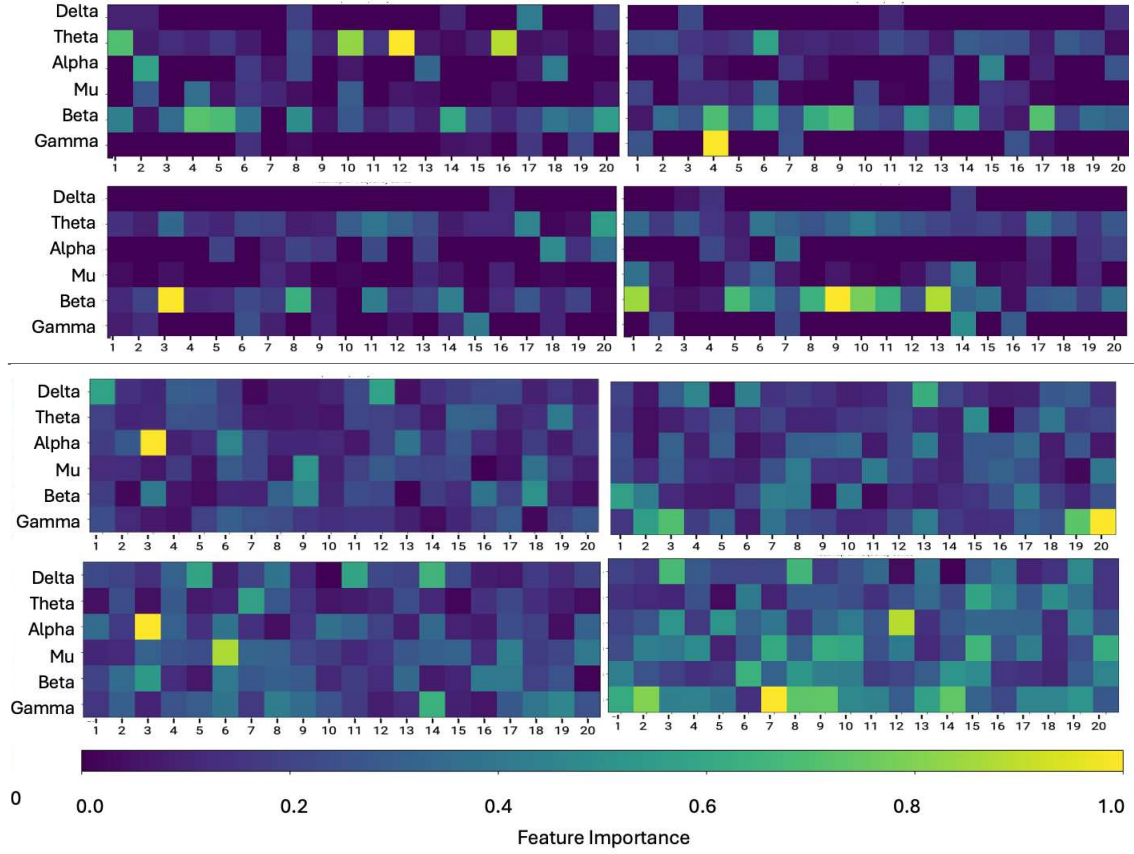


Figure 4.5: Feature Importance for difference frequency bands for randomly selected 4 dystonia patients (top) and 4 healthy subjects (bottom)

ing disrupted temporal organization of cortico-muscular coupling. This disruption is particularly evident in the mu and beta frequency ranges, which are known to be crucial for motor control. Figure 4.6 presents SHAP value distributions across different frequency bands, demonstrating marked differences between healthy subjects and patients. In healthy subjects, alpha band activity shows the highest magnitude of SHAP values, followed by beta and gamma bands, suggesting these frequencies play crucial roles in normal cortico-muscular coupling. Notably, the magnitude of influence is substantially higher in healthy subjects compared to patients, with SHAP values reaching up to 0.040 versus 0.00030 respectively. In contrast, for dystonia patients, the theta band plays a prominent role, potentially reflecting abnormal network connectivity and pathological brain rhythms. Similar to healthy individuals, beta rhythms remain crucial in dystonia for sensory feedback, underscoring their their

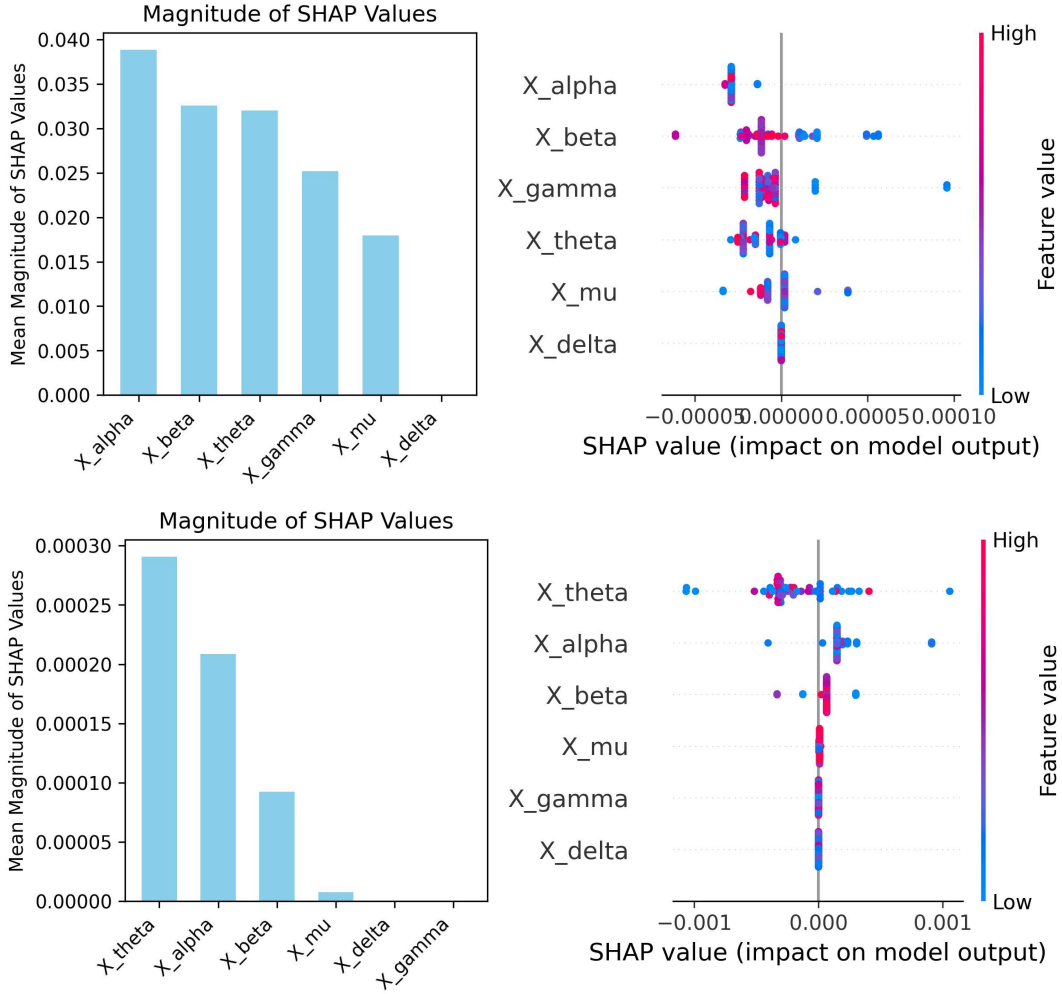


Figure 4.6: SHAP values for different frequencies in healthy subjects (top) and patients (bottom).

relevance for sensory feedback in both cases. The dramatic reduction in SHAP values among patients indicates a significant disruption in the normal frequency-specific coupling mechanisms in dystonia.

Our approach demonstrates comparable performance to these advanced methods while maintaining computational efficiency. Moreover, INFR-GC offers clear interpretability of the model, providing valuable insights into the underlying causal relationships in both healthy and affected subjects. The trade off between accuracy, efficiency, and interpretability highlights the practical benefits of our proposed method.

These findings collectively indicate a fundamental alteration in cortico-muscular coupling mechanisms in dystonia, characterized by reduced frequency-specific influ-

ences and disrupted temporal organization of causal interactions. The INFR-GC framework successfully captures these pathological changes while providing more stable and interpretable measures compared to traditional approaches. The pronounced differences in SHAP values and feature importance patterns between groups validate the clinical utility of our approach in quantifying pathological changes in motor control circuits. The consistent pattern of reduced GC values in patients, particularly in the feedforward (EEG to sEMG) direction, aligns with clinical observations of impaired motor control in dystonia and suggests a specific disruption in the efferent control pathways. This asymmetric disruption provides new insights into the pathophysiological mechanisms underlying dystonia and may inform more targeted therapeutic approaches.

4.3 Insights and Implications

The INFR-GC framework demonstrates significant potential in explaining Granger-causal relationships within neurophysiological systems, yet several methodological constraints warrant careful consideration. The current implementation, while effective for standard neurophysiological signals, faces fundamental challenges in capturing the intricate structures inherent in complex physiological systems. Most notably, the neural estimation approach for mutual information, though providing robust results, encounters theoretical and practical limitations in its estimation accuracy. Moreover, the underlying estimation framework remains sensitive to sample size requirements, network architecture specifications, and computational constraints in high-dimensional feature spaces.

To address these limitations, future developments must explore the integration of probabilistic frameworks for uncertainty estimation, dealing with the critical need for confidence quantification in causal relationships. Moreover, the issue of unobserved variables in cortico-muscular studies presents a critical challenge to causal identification. While GC effectively identifies predictive relationships between

observed variables, it cannot definitively disambiguate between direct causal effects and indirect influences propagated through unmeasured pathways. This limitation becomes particularly acute in neurophysiological investigations where subcortical structures, though not directly recorded, serve as crucial mediators of cortical-muscular interactions.

These methodological constraints motivate the exploration of counterfactual frameworks, which offer a more rigorous theoretical foundation for causal inference. Unlike GC's prediction-based approach, counterfactual models explicitly formalize interventional reasoning and potential outcomes, enabling systematic modeling of time-varying causal effects, mediation pathways, and multi-scale interactions. This transition represents more than a mathematical refinement and constitutes a fundamental reconceptualization of causality in physiological systems, providing a more nuanced framework for understanding cortico-muscular interactions while acknowledging the challenges in empirical implementation and validation.

The integration of these advanced methodological frameworks with advanced signal processing techniques and uncertainty quantification methods presents promising avenues for advancing physiological causal inference, while maintaining the interpretability and computational feasibility necessary for practical applications in clinical settings.

Chapter 5

Beyond Predictive Ability: Counterfactual Perspective on Causality

In this chapter, we move beyond models focused solely on prediction to explore approaches capable of capturing and reasoning about causal relationships. This progression marks a critical step toward developing models that are not only robust but also interpretable. While traditional machine learning methods excel in recognizing patterns and making predictions, they often fall short in uncovering the underlying causal mechanisms [94]. Historically, causal learning algorithms have relied on predictive power as a proxy for causal insight. However, while predictive models can identify correlational patterns with impressive accuracy, understanding true causal relationships demands a fundamentally distinct framework that goes beyond correlation to uncover the directional and mechanistic links within data. Consider a medical diagnosis system trained on historical patient data. While it might achieve high accuracy in predicting disease outcomes, it cannot answer crucial questions about interventions: “What would happen if we administered treatment A instead of B?” Such questions are inherently counterfactual and require causal reasoning.

For this purpose, the counterfactual perspective on causality, pioneered by [89], provides a rigorous mathematical foundation for reasoning about cause and effect, marking a significant advancement in our ability to understand complex systems. This chapter deals with methodologies for transcending predictive capabilities to design models that explicitly encode and analyze underlying causal mechanisms.

Despite its theoretical elegance, a fundamental challenge in counterfactual reasoning arises from unmeasured confounding variables that simultaneously influence both treatment assignment and outcomes, violating the backdoor criterion [89]. When such confounders are present, causal effects become non-identifiable from observational data alone, necessitating either experimental intervention or additional identifying assumptions such as instrumental variables [43] or proxy controls [24]. Even when measured confounders are available, high-dimensional settings introduce practical challenges for conventional adjustment methods, requiring careful balance between bias and variance in estimation [75]. These identification challenges are particularly acute in observational studies where the confounding structure may be complex and partially unknown.

While the Neyman-Rubin potential outcomes framework [99] established foundational principles for causal inference through randomized experiments, Pearl’s structural causal models (SCMs) [89] introduced a complementary algebraic framework that explicitly encodes causal relationships through directed acyclic graphs (DAGs) and structural equations. These approaches differ significantly in their treatment of confounding variables and identification strategies. In particular, while SCMs provide a rich graphical language for representing and reasoning about causal relationships in complex systems, the underlying assumptions and methodological emphases of these frameworks reflect their distinct historical origins and intended applications.

The potential outcomes framework primarily addresses causal effects through the lens of treatment assignment, focusing on average treatment effects (ATEs) and conditional average treatment effects (CATEs) [10]. This approach has proven

particularly effective in domains where randomized controlled trials are feasible, such as clinical trials and econometric studies [45]. However, its reliance on the stable unit treatment value assumption (SUTVA) and complete randomization can limit its applicability in complex systems with interference between units or partial controllability [100].

In contrast, SCMs provide a more flexible framework for representing and reasoning about causal relationships through the do-calculus [91]. This approach enables the identification of causal effects from observational data under certain conditions, even when randomization is impossible or unethical. The do-operator formalizes interventional queries, allowing researchers to derive identifiability criteria and estimation strategies for causal effects. Recent work by [93] has extended these methods to handle high-dimensional settings through invariant causal prediction (ICP), which exploits stability across different environmental conditions to identify causal parents.

The integration of machine learning with causal inference has led to novel methodological developments. [107] introduced counterfactual regression using representation learning, addressing the fundamental problem of causal inference through balanced representations. This approach minimizes selection bias while maintaining expressiveness, achieving superior performance in scenarios with strong confounding. Similarly, [14] developed counterfactual reasoning techniques for policy evaluation in reinforcement learning, enabling more efficient exploration strategies through off-policy evaluation.

The theoretical advances in causal inference have profound implications for contemporary challenges in artificial intelligence and healthcare. Recent work by [90] demonstrates how counterfactual reasoning frameworks enhance the interpretability of deep learning systems by enabling attribution of effects to specific architectural components and training decisions. This capability proves crucial for debugging complex models and ensuring their reliability in high-stakes applications. In precision medicine, counterfactual methods have enabled more detailed approaches to treatment effect heterogeneity, moving beyond average effects to estimate patient-specific

responses through methods such as causal forests and deep structural models [124]. These developments represent a crucial bridge between theoretical advances in causal methodology and pressing practical challenges in modern applications, particularly where decisions have significant societal impact.

5.1 Dynamic Bayesian Networks

Bayesian networks have emerged as a foundational framework for probabilistic reasoning and causal inference, offering a mathematical apparatus for modeling complex dependencies among variables through directed acyclic graphs (DAGs). They are designed to encode conditional independence relationships through their structure, while maintaining computational tractability through the chain rule of probability. Mathematically, the joint probability distribution $P(\mathbf{X})$ over a set of variables $\mathbf{X} = \{X_1, \dots, X_n\}$ can be factorized according to the network structure:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid \text{Pa}(X_i)) \quad (5.1)$$

where $\text{Pa}(X_i)$ denotes the parents of variable X_i in the graph. While static Bayesian networks effectively model time-invariant relationships, where the underlying probability distribution remains constant across time, many real-world phenomena exhibit temporal dynamics that necessitate more expressive modeling approaches.

Dynamic Bayesian Networks (DBNs) extend the traditional Bayesian network framework to incorporate temporal dependencies, enabling the representation of stochastic processes that evolve over time [42, 4]. This approach enables DBNs to model both contemporaneous relationships within a single time slice and temporal relationships across adjacent time points. This temporal extension retains the key benefits of Bayesian networks, including their compact representation of complex probability distributions through local independence assumptions, while introducing the capability to model dynamic systems. For a sequence of observations $\mathbf{X}_{1:T} =$

$\{\mathbf{X}_1, \dots, \mathbf{X}_T\}$, the joint distribution factorizes as:

$$P(\mathbf{X}_{1:T}) = P(\mathbf{X}_1) \prod_{t=2}^T P(\mathbf{X}_t \mid \mathbf{X}_{t-1}) \quad (5.2)$$

where $P(\mathbf{X}_t \mid \mathbf{X}_{t-1})$ represents the transition model between consecutive time slices.

5.1.1 Structure Learning

The challenge of structure learning in Bayesian networks, whether static or dynamic, represents a fundamental problem in causal discovery. Structure learning algorithms attempt to recover the underlying graph structure from observed data, effectively identifying the dependencies that give rise to the observed patterns of association. This process confronts several theoretical and computational challenges. In the context of static Bayesian networks, structure learning algorithms typically fall into two broad categories: constraint-based methods [101, 16, 60], which leverage conditional independence tests to identify graph structure, and score-based methods [83, 22, 61], which search through the space of possible structures to optimize a scoring criterion.

The challenge of structure learning in Bayesian networks represents a fundamental problem in causal discovery, where the objective is to recover the underlying graph structure $\mathcal{G}$ from observed data $\mathcal{D}$. In static networks, this process typically employs score-based or constraint-based approaches. Score-based methods optimize a scoring function $S(\mathcal{G}, \mathcal{D})$ over possible structures:

$$\mathcal{G}^* = \arg \max_{\mathcal{G}} S(\mathcal{G}, \mathcal{D}) - \lambda |\mathcal{G}| \quad (5.3)$$

where $|\mathcal{G}|$ denotes the complexity of the graph and λ serves as a regularization parameter. Common scoring functions include the BIC and the Bayesian Dirichlet equivalent uniform (BDeu) score [105]:

$$\text{BIC}(\mathcal{G}, \mathcal{D}) = \ell(\boldsymbol{\theta}^* \mid \mathcal{D}) - \frac{\log n}{2} |\mathcal{G}| \quad (5.4)$$

where $\ell(\boldsymbol{\theta}^*|\mathcal{D})$ represents the maximized log-likelihood and n is the sample size.

In the dynamic setting, for a two-slice temporal Bayes net (2TBN) representation, the structure learning problem can be formulated as:

$$\mathcal{G}^* = \arg \max_{\mathcal{G}} \left[S(\mathcal{G}_1, \mathcal{D}_1) + \sum_{t=2}^T S(\mathcal{G}_2, \mathcal{D}_{t,t-1}) \right] \quad (5.5)$$

where $\mathcal{G}_1$ represents the initial network structure and $\mathcal{G}_2$ captures the temporal transitions. The structure learning problem becomes substantially more complex in the dynamic setting [117]. Dynamic Bayesian Networks introduce additional constraints and challenges due to their temporal nature. The search space expands significantly, as the algorithm must identify both intra-slice dependencies (relationships within a time slice) and inter-slice dependencies (relationships between variables across time). Traditional approaches often decompose the structure learning problem into separate phases for learning intra-slice and inter-slice dependencies [41], though more recent methods have explored unified approaches that leverage the temporal structure more effectively.

5.1.2 Parameter Learning

Parameter learning, following the establishment of a network structure, involves estimating the conditional probability distributions that govern the relationships between nodes within the network. In the Bayesian network context, these parameters quantify the strength and nature of the relationships encoded by the graph structure. The parameter learning process must account for various challenges, including incomplete data, hidden variables, and the potential for complex interactions among variables.

For static Bayesian networks, parameter learning typically employs either maximum likelihood estimation or Bayesian approaches, depending on the availability of prior knowledge and the specific requirements of the application. The maximum likelihood framework provides point estimates of parameters, while Bayesian approaches

yield full posterior distributions over parameters, incorporating prior knowledge and uncertainty in a principled manner. These methods often leverage the local independence properties of Bayesian networks, allowing parameters to be estimated independently for each family of variables (a node and its parents).

In the dynamic setting, parameter learning must address additional complexities introduced by temporal dependencies. The parameters must capture not only the instantaneous relationships within time slices but also the evolution of variables across time. This temporal aspect often necessitates specialized learning algorithms that can effectively leverage the repeated structure across time slices while accounting for potential non-stationarity in the underlying process. Contemporary approaches frequently employ variational methods [98] or sampling-based techniques [113] to handle the increased computational complexity introduced by temporal dependencies, particularly in the presence of hidden states or missing data. In the presence of complete data, the maximum likelihood estimates for discrete variables can be obtained through sufficient statistics:

$$\theta_{ijk} = P(X_i = k \mid \text{Pa}(X_i) = j) = \frac{N_{ijk}}{N_{ij}} \quad (5.6)$$

where N_{ijk} represents the count of instances where $X_i = k$ and $\text{Pa}(X_i) = j$. For continuous variables, kernel density estimation methods can be employed to approximate the conditional densities:

$$P(X_i \mid \text{Pa}(X_i)) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (5.7)$$

where K represents a kernel function and h is the bandwidth parameter.

In DBNs, the transition parameters can be estimated through a modified likelihood function that incorporates temporal structure:

$$\ell(\boldsymbol{\theta} \mid \mathcal{D}) = \log P(\mathbf{X}_1; \boldsymbol{\theta}_1) + \sum_{t=2}^T \log P(\mathbf{X}_t \mid \mathbf{X}_{t-1}; \boldsymbol{\theta}_2) \quad (5.8)$$

where $\boldsymbol{\theta}_1$ represents the initial network parameters and $\boldsymbol{\theta}_2$ captures the transition

parameters. For non-stationary processes, time-varying parameters can be estimated through state-space models:

$$P(\mathbf{X}_t | \mathbf{X}_{t-1}, \boldsymbol{\theta}_t) = \int P(\mathbf{X}_t | \mathbf{X}_{t-1}, \boldsymbol{\theta}_t) P(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) d\boldsymbol{\theta}_{t-1} \quad (5.9)$$

This formulation allows for adaptive parameter estimation that captures evolving relationships in the dynamic system. Through this framework, dynamic Bayesian networks provide a methodological framework for causality estimation by integrating temporal dynamics, probabilistic modeling, and counterfactual reasoning to uncover causal relationships in complex systems. However, despite their widespread application in temporal modeling, they exhibit certain limitations. Firstly, their graphical structure enforces strict directionality in relationships through acyclic graph constraints, making it difficult to properly capture systems where variables influence each other simultaneously or through complex feedback mechanisms. Additionally, the parametric nature of DBNs often requires assumptions about linear relationships or specific probability distributions, typically Gaussian or discrete categorical, that may not hold in real-world systems. These constraints become particularly problematic when analyzing physiological signals or neural dynamics, where nonlinear interactions and feedback loops are prevalent.

These limitations can be partially mitigated by alternative methods such as Convergent Cross Mapping (CCM), which is specifically designed to detect bidirectional causality and nonlinear dynamics. Unlike the rigid graphical constraints of DBNs, CCM excels at capturing bidirectional causality and intricate feedback mechanisms by employing state-space reconstruction. Leveraging Takens' embedding theorem, CCM reconstructs the attractor of the underlying dynamical system, enabling the identification of nonlinear causal relationships without relying on restrictive parametric assumptions. This flexibility makes CCM particularly well-suited for analyzing complex systems where traditional linear methods or probabilistic models fail to fully capture the intricacies of causal interactions.

However, CCM introduces its own set of methodological challenges. The method

lacks a formal framework for handling measurement uncertainty and temporal evolution of probability distributions, which are inherently well-handled by DBNs’ probabilistic foundation. CCM’s reliance on precise state-space reconstruction makes it sensitive to noise. In addition to that, the absence of an explicit uncertainty quantification framework makes it difficult to assess the confidence in detected causal relationships, particularly in scenarios with substantial measurement noise or system stochasticity.

To augment the capabilities of both methodologies and address their complementary limitations, we propose an integrated approach that leverages their respective strengths, creating a more robust framework for causality analysis. The integration addresses these limitations through three key mechanisms. First, the DBN structure guides the shadow manifold construction by incorporating known temporal dependencies, improving the robustness of state-space reconstruction through probabilistic constraints. Second, the kernel weights combine geometric distances with conditional probabilities, creating a hybrid similarity metric that accounts for both deterministic dynamics and uncertainty. Third, the normalized combined weights enable adaptive blending of geometric and probabilistic information, leading to more reliable causal inference in complex systems. This adaptive weighting scheme ensures that the method can automatically adjust to varying levels of uncertainty and nonlinearity in the data.

This unified approach yields several practical benefits that extend beyond the capabilities of either method alone: it can detect bidirectional causality while maintaining probabilistic rigor, handle measurement noise through principled uncertainty quantification, and capture nonlinear relationships without sacrificing the interpretability of the model structure. The resulting correlation metric provides a quantitative measure of causality that incorporates both geometric and probabilistic aspects of the system dynamics, offering a more comprehensive assessment of causal relationships. This is particularly valuable for analyzing real-world systems where complex interactions, feedback loops, and uncertainty coexist, such as in neural

signal processing, climate systems, or financial markets. The method’s ability to handle both nonlinear dynamics and uncertainty makes it especially suitable for applications where traditional linear methods or purely deterministic approaches may fail to capture the full complexity of the system. In the following section, we will formally present our proposed methodology, detailing its framework, underlying principles, and potential applications.

5.2 DBN-Informed Convergent Cross Mapping

This section introduces a novel methodological framework that synthesizes Dynamic Bayesian Networks (DBNs) with Convergent Cross Mapping (CCM) through kernel-based estimation, leveraging the geometric properties of shadow manifolds while incorporating probabilistic temporal dependencies. Our proposed approach, DBN-Informed CCM, addresses key computational challenges in traditional kernel-based methods while preserving essential causal information encoded in the manifold structure.

Traditional Kernel Density Estimation (KDE) approaches require computing kernel functions centered at each point in the dataset, leading to substantial computational overhead when analyzing temporal sequences. In contrast, DBN-Informed CCM employs a targeted estimation strategy that exploits the local geometry of shadow manifolds. Specifically, for each point in the embedded space, we compute Gaussian kernel weights: $u_i = \exp(\frac{-d_i^2}{2\sigma^2})$ where d_i represents the Euclidean distance in the embedded space, and σ^2 denotes the kernel bandwidth parameter. Crucially, these computations are restricted to the $E + 1$ nearest neighbors in the embedded space, where E denotes the embedding dimension determined through delay embedding theorems. This selective computation strategy reduces the computational complexity from $O(N^2)$ in traditional KDE to $O(NE)$, where N represents the time series length and typically $E \ll N$.

The methodological innovation extends beyond computational efficiency. By

focusing on local geometric relationships within the shadow manifold, the kernel weights capture fine-grained dynamic patterns that may be obscured in global analyses. These local geometric relationships are then integrated with global structural information through the incorporation of conditional probabilities p_i derived from the DBN framework. Specifically, the kernel weights are modulated by these conditional probabilities, creating a dual representation that encodes both the local manifold geometry through kernel-based similarity measures and the global temporal dependencies through DBN-derived conditional probabilities.

This integration provides a principled framework for combining geometric and probabilistic approaches to temporal causal discovery. The resulting methodology maintains computational tractability while preserving the capacity to detect complex causal relationships in nonlinear dynamical systems. The key theoretical advantage of this approach lies in its ability to leverage complementary forms of causal information: the geometric properties of shadow manifolds that capture nonlinear dynamic relationships, and the probabilistic dependencies encoded in DBN structures. This enables more robust causal discovery in systems characterized by both nonlinear dynamics and complex temporal dependencies.

5.2.1 Theoretical Foundation

The theoretical underpinning of our framework emerges from the intersection of dynamical systems theory, differential geometry, and probabilistic inference. We develop a comprehensive mathematical framework that synthesizes the geometric properties of state space reconstruction with probabilistic causal inference. This synthesis enables robust detection of causal relationships in complex, nonlinear dynamical systems while maintaining computational tractability.

Consider two time series $X = \{x_1, \dots, x_N\}$ and $Y = \{y_1, \dots, y_N\}$ representing observations from potentially coupled dynamical systems. Our approach leverages Takens' embedding theorem, which establishes that under suitable smoothness conditions and generic observables, the reconstructed manifold preserves the essential

topological and dynamical properties of the original system's attractor. Specifically, each point in the shadow manifold $M_X[i]$ is constructed through delay embedding:

$$M_X[i] = [x_i, x_{i-\tau}, \dots, x_{i-(E-1)\tau}] \quad (5.10)$$

The fidelity of this reconstruction critically depends on two parameters: the embedding dimension E and time delay τ . Their selection follows from both theoretical requirements and practical considerations.

In the reconstructed state space, we identify the optimal set of $E + 1$ nearest neighbors for each point through:

$$\mathcal{I} = \arg \min_{j \in J, |J|=E+1} \sqrt{\sum_{k=1}^E (M_X[i]_k - M_X[j]_k)^2} \quad (5.11)$$

The selection of exactly $E + 1$ neighbors is theoretically motivated: it represents the minimal set required for unique point determination in E -dimensional space plus one additional point for robust estimation, thereby achieving an optimal balance between estimation stability and computational efficiency.

Our novel hybrid weighting scheme integrates geometric proximity with probabilistic causality through two fundamental components:

$$u_i = \exp\left(-\frac{d_i^2}{2\sigma^2}\right), \quad i \in \mathcal{I} \quad (5.12)$$

$$p_i = P(Y[i]|\text{Pa}(Y[i])), \quad i \in \mathcal{I} \quad (5.13)$$

The exponential kernel weight u_i is specifically chosen for its natural distance-decay properties aligned with manifold diffusion processes and its optimality for local density estimation in the presence of noise. The DBN conditional probabilities p_i incorporate known causal structure, providing a probabilistic framework for temporal

Algorithm 6 DBN-Informed CCM

Input: Time series $X = \{x_1, \dots, x_N\}$, $Y = \{y_1, \dots, y_N\}$, embedding dimension E , time delay τ , DBN structure $\mathcal{G}$
 $M_X[i] = [x_i, x_{i-\tau}, \dots, x_{i-(E-1)\tau}]$ $\triangleright$ Construct shadow manifold M_X with each point $M_X[i]$
for $t \leftarrow 1$ to N **do**
 $\mathcal{I} = \arg \min_{j \in J, |J|=E+1} \sqrt{\sum_{k=1}^E (M_X[t]_k - M_X[j]_k)^2}$ $\triangleright$ Find $E + 1$ nearest neighbor indices
 $u_i = \exp(-\frac{d_{ti}^2}{2\sigma^2})$, $i \in \mathcal{I}$ $\triangleright$ Compute kernel weights for neighbors
 $p_i = P(Y[i] | \text{Pa}(Y[i]))$, $i \in \mathcal{I}$ $\triangleright$ Get DBN conditional probabilities
 $w_i = \frac{u_i p_i}{\sum_{j \in \mathcal{I}} u_j p_j}$, $i \in \mathcal{I}$ $\triangleright$ Compute normalized combined weights
 $Y_{pred}[t] = \sum_{i \in \mathcal{I}} w_i Y[i]$ $\triangleright$ Predict Y value as weighted sum
end for
 $\rho = \frac{\sum_{t=1}^N (Y[t] - \bar{Y})(Y_{pred}[t] - \bar{Y}_{pred})}{\sqrt{\sum_{t=1}^N (Y[t] - \bar{Y})^2} \sqrt{\sum_{t=1}^N (Y_{pred}[t] - \bar{Y}_{pred})^2}}$ $\triangleright$ Compute correlation between Y and Y_{pred}
Output: Correlation $\rho_{X \rightarrow Y}$

dependency modeling. These components are integrated through normalized weights:

$$w_i = \frac{u_i p_i}{\sum_{i \in \mathcal{I}} u_i p_i} \quad (5.14)$$

This normalization ensures proper probabilistic interpretation while preserving the relative influence of both geometric and causal factors. The prediction framework employs these weights in a locally linear approximation scheme:

$$Y_{pred}[t] = \sum_{i \in \mathcal{I}} w_i Y[i] \quad (5.15)$$

The strength of causal relationships is quantified through correlation analysis:

$$\rho = \frac{\sum_{i=1}^N (Y[i] - \bar{Y})(Y_{pred}[i] - \bar{Y}_{pred})}{\sqrt{\sum_{i=1}^N (Y[i] - \bar{Y})^2} \sqrt{\sum_{i=1}^N (Y_{pred}[i] - \bar{Y}_{pred})^2}} \quad (5.16)$$

This choice of correlation measure provides natural normalization and invariance to linear scaling, facilitating comparative analysis across different variable pairs and systems. The complete algorithmic implementation is detailed in Algorithm 6.

5.2.2 Intervention Implementation

Since dynamic Bayesian networks provide a formal way to implement counterfactual interventions analysis, we can extend our proposed method to incorporate counterfactual interventions within a unified theoretical framework. The implementation of interventions in our framework is designed to capture both local and global changes in the system’s dynamics, allowing for comprehensive analysis of causal effects across multiple temporal scales. This dual-scale approach enables robust identification of causal relationships while accounting for the inherent complexity of temporal dependencies. The intervention analysis proceeds along two fundamental dimensions that capture complementary aspects of the system’s causal structure:

First, the geometric changes in the shadow manifold M_X reflect local perturbations in the system’s state space structure. This choice is motivated by the understanding that causal influences should manifest as detectable changes in the system’s dynamical structure, particularly in the geometric relationships between temporally adjacent states. Post-intervention, the points in $M_X[i]$ reflect modified dynamics, leading to reorganized nearest neighbor relationships. These geometric modifications provide crucial information about the system’s response to interventions, as genuine causal relationships necessarily alter the underlying manifold structure.

Second, we track changes in DBN conditional probabilities $P(Y[i]|\text{Pa}(Y[i]))$. This probabilistic analysis is crucial because interventions should modify the probabilistic relationships between variables if true causation exists. These changes propagate through our hybrid weights w_i , affecting the prediction accuracy in ways that reveal the underlying causal structure. The integration of geometric and probabilistic approaches provides a more complete characterization of causal effects than either method alone.

Our approach to causal effect estimation is based on the principle that true causal relationships should manifest as measurable changes in predictability under intervention. This fundamental insight leads us to quantify causal effects through pre- and post-intervention correlations:

$$\rho_{pre} = \text{corr}(Y, Y_{pred} | M_X^{pre}) \quad (5.17)$$

$$\rho_{post} = \text{corr}(Y, Y_{pred} | M_X^{post}) \quad (5.18)$$

The difference in these correlations provides a natural measure of causal effect, capturing the intervention-induced change in predictability:

$$\Delta\rho = \rho_{post} - \rho_{pre} \quad (5.19)$$

To ensure robust structural identification, we apply a sparsity-induced structural learning technique within a dynamic Bayesian framework, combining ℓ_1 -regularized and score-based approaches to model temporal dependencies. This methodology integrates proximal gradient optimization with classical DBN learning principles, offering an approach to discovering temporal structures through sparsity constraints as outlined in Algorithm 7. The integration of sparsity constraints helps prevent overfitting while capturing the most salient temporal dependencies.

Our proposed approach integrates structure learning, parameter estimation, and causal inference through convergent cross mapping in a framework designed to quantify intervention effects in complex temporal systems. This unified framework allows for consistent treatment of uncertainty across all analysis stages while maintaining computational tractability.

- 1) **Structural Learning:** We begin by learning the temporal dependency structure of the system through dynamic Bayesian network analysis. This involves identifying the optimal network structure $\mathcal{G}$ that captures the temporal relationships between variables X and Y , utilizing ℓ_1 regularized algorithms adapted for temporal data 7.
- 2) **Parameter Learning:** Given the learned structure $\mathcal{G}$, we estimate the conditional probability distributions $P(X_t | Pa(X_t))$ and $P(Y_t | Pa(Y_t))$ that char-

Algorithm 7 DBN Structure Learning

Input: $X = \{x_1, \dots, x_T\}$, $Y = \{y_1, \dots, y_T\}$, λ , Maximum lag L , threshold θ
Initialize $W_{X \rightarrow Y} \leftarrow 0_L$, $W_{Y \rightarrow X} \leftarrow 0_L$
while iter $\leq$ max_iter **do** ▷ Build lagged matrices
 for $l = 1$ to L **do**
 $X^l \leftarrow [x_{t-l}]_{t=l+1}^T$
 $Y^l \leftarrow [y_{t-l}]_{t=l+1}^T$
 end for
 $\nabla L_{X \rightarrow Y} = \frac{1}{T-L} \sum_{t=L+1}^T (y_t - \sum_{l=1}^L w_l^{X \rightarrow Y} x_{t-l}) x_{t-l}$ ▷ Update weight matrices
 $W_{X \rightarrow Y} \leftarrow \text{ProxL1}(W_{X \rightarrow Y} - \eta \nabla L_{X \rightarrow Y}, \lambda)$
 $\nabla L_{Y \rightarrow X} = \frac{1}{T-L} \sum_{t=L+1}^T (x_t - \sum_{l=1}^L w_l^{Y \rightarrow X} y_{t-l}) y_{t-l}$
 $W_{Y \rightarrow X} \leftarrow \text{ProxL1}(W_{Y \rightarrow X} - \eta \nabla L_{Y \rightarrow X}, \lambda)$
 if $\|W^{\text{new}} - W^{\text{old}}\| < \epsilon$ **then**
 break
 end if
end while
for $t = 1$ to $T - 1$ **do** ▷ Construct DBN graph structure
 for $l = 1$ to L **do**
 if $|W_{X \rightarrow Y}[l]| > \theta$ **then**
 Add edge (X^l, Y^{t+l}) to $\mathcal{G}$
 end if
 if $|W_{Y \rightarrow X}[l]| > \theta$ **then**
 Add edge (Y^l, X^{t+l}) to $\mathcal{G}$
 end if
 end for
end for
Output: $\mathcal{G}$

function PROXL1(w, λ)
 return $\text{sign}(w) \odot \max(|w| - \lambda, 0)$
end function

acterize the temporal dependencies. This step employs maximum likelihood estimation or Bayesian inference to obtain robust parameter estimates θ that quantify the strength of temporal relationships as in 8.

3) Initial CCM Estimation: We implement DBN-informed CCM to establish baseline causal relationships. This involves constructing shadow manifolds M_X and M_Y , computing kernel weights u_i and normalized combined weights w_i incorporating DBN probabilities, and estimating the pre-intervention correlation measures $\rho_{X \rightarrow Y}$ and $\rho_{Y \rightarrow X}$ as detailed in 6.

4) Intervention Implementation: We introduce targeted interventions on

Algorithm 8 DBN Parameter Learning

Input: $X = \{x_1, \dots, x_T\}$, $Y = \{y_1, \dots, y_T\}$, Graph structure $\mathcal{G}$, Maximum lag L
 $P_{X \rightarrow Y} \leftarrow 0_L$, $P_{Y \rightarrow X} \leftarrow 0_L$ ▷ Initialize parameter matrices
for each edge $(u, v) \in \mathcal{G}$ **do** ▷ Compute parameters for each edge
 if u is type X and v is type Y **then**
 $P(v|u) \leftarrow \text{EstimateConditional}(v|u, \text{lag}(u, v))$
 $P_{X \rightarrow Y}[\text{lag}(u, v)] \leftarrow P(v|u)$
 else
 $P(v|u) \leftarrow \text{EstimateConditional}(v|u, \text{lag}(u, v))$
 $P_{Y \rightarrow X}[\text{lag}(u, v)] \leftarrow P(v|u)$
 end if
end for
for each node $v \in \mathcal{G}$ **do** ▷ Normalize probabilities
 Normalize $P(v|\text{Pa}(v))$ such that $\sum_v P(v|\text{Pa}(v)) = 1$
end for
Output: $\{P_{X \rightarrow Y}, P_{Y \rightarrow X}\}$

function ESTIMATECONDITIONAL($(v|u, \text{lag})$)
 $N_{u,v} \leftarrow$ count of joint occurrences of (u, v) with specified lag
 $N_u \leftarrow$ count of u occurrences in total
 return $N_{u,v}/N_u$
end function

selected variables, modifying either X or Y while maintaining observation of the system’s temporal evolution. These interventions are designed to perturb the system while preserving its fundamental dynamic structure.

- 5) **Post-intervention Parameter Estimation:** Following the intervention, we re-estimate the conditional probability distributions to capture potential changes in the system’s temporal dynamics. This yields updated parameter estimates θ' that reflect intervention-induced modifications to the temporal relationships.
- 6) **Post-intervention CCM Estimation:** We recompute DBN-informed CCM metrics using post-intervention data, following the same procedure as Step 3 but with the updated parameter estimates. This generates post-intervention correlation measures $\rho'_{X \rightarrow Y}$ and $\rho'_{Y \rightarrow X}$.
- 7) **Comparative Analysis:** We quantify the intervention’s causal impact by analyzing changes in CCM metrics across both causal directions and analyze

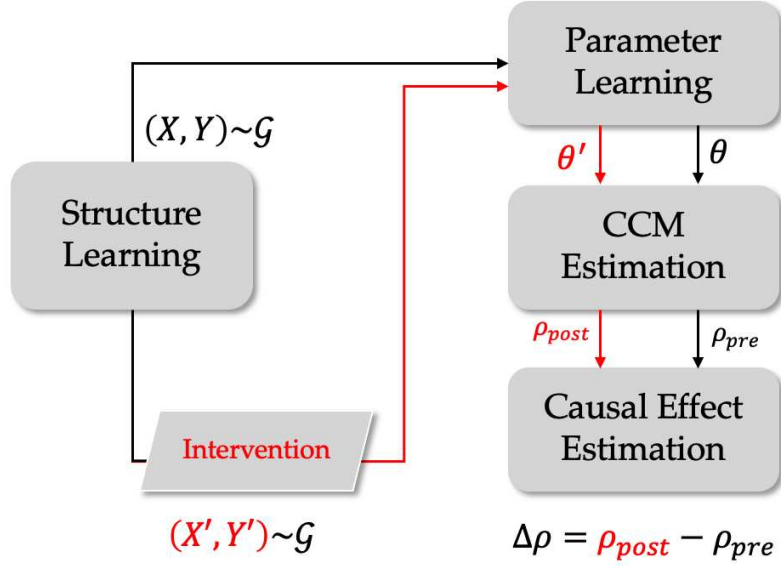


Figure 5.1: Overview of the proposed method DBN-Informed CCM

the relative magnitudes and signs of these changes to assess directional causal effects.

By explicitly bridging the geometric and probabilistic dimensions of causal inference, our approach provides a robust framework for identifying causal effects in complex temporal systems. In the next section, we demonstrate the effectiveness of our proposed method, DBN-Informed CCM, through evaluations on both synthetic and real-world datasets.

5.2.3 Experimental Evaluation on Lorenz96 Data

We conducted a series of experiments on the Lorenz96 dataset to validate the proposed methodology. The network structure was initially inferred using the first 20 variables, as depicted in Figure 5.2. This step provides critical insights into the underlying dependencies within the dataset. By visualizing the learned dependency structure, we can identify the interactions and relationships between different variables, including their neighbors within the network. Such visualization not only facilitates a better understanding of the dataset’s inherent dynamics but also aids in interpreting how information propagates across variables, which is essential for downstream analyses

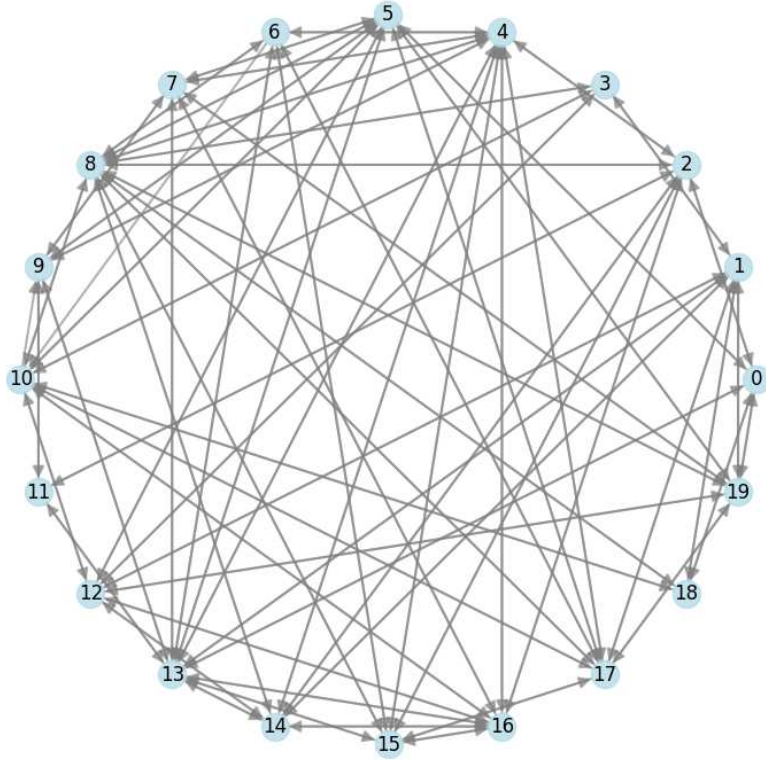


Figure 5.2: Dynamic Bayesian Network learnt from Lorenz96 data

and applications. The experimental evaluation on the Lorenz96 dataset provides a comparative analysis of various previously discussed methods and causality detection architectures, highlighting their relative effectiveness in capturing and identifying complex temporal dependencies. The performance metrics reveal a clear progression in detection capabilities across the model spectrum, from traditional sequence models to causal inference frameworks.

The DBN-Informed CCM architecture demonstrates superior performance across all evaluation metrics as outlined in Table 5.1, achieving an accuracy of 0.9380 and an F1 score of 0.9570. All methods demonstrate the ability to identify some true causal connections, and none incorrectly flag non-causal connections as causal. However, only a subset of the actual causal connections is discovered by these methods. Notably, DBN-Informed CCM outperforms the others by detecting nearly all causal connections with high accuracy, achieving the highest recall value among the evaluated approaches. This represents a substantial improvement over the baseline LSTM and RC models, which achieved accuracies of 0.7125 and 0.7125 respectively.

The INFR-GC model occupies an intermediate position, with an accuracy of 0.8751. Notably, while all models achieved perfect precision (1.0000), their recall values varied significantly, with DBN-Informed CCM achieving 0.9170 compared to the markedly lower 0.6406 and 0.6618 for LSTM and RC respectively.

The dynamic Bayesian network structure learned from the data, as visualized in the second figure, reveals a complex web of temporal dependencies among 20 variables (labeled 0-19). The circular arrangement of nodes with dense interconnections suggests a highly coupled dynamical system, characteristic of the Lorenz96 model’s inherent chaos and nonlinearity. The network structure exhibits significant symmetry in its connectivity patterns, with each node maintaining multiple incoming and outgoing edges to both neighboring and distant nodes. This learned structure effectively captures the spatiotemporal relationships inherent in the Lorenz96 system, where each variable’s evolution is influenced by both local and non-local interactions.

After performing the convergent cross-mapping analysis, we extended the learned DBN structure to include counterfactual intervention analysis. This is shown in Figure 5.3, which highlights the complex interactions and causal relationships within the Lorenz96 system. The figure compares two visualizations that illustrate both the direct and indirect effects of the interventions.

In this particular intervention, we begin with a baseline system operating under normal conditions, where each point in the system evolves according to the standard Lorenz-96 dynamics with a forcing parameter $F=8$. The intervention introduces

Table 5.1: Performance Comparison of LSTM, RC, INFR-GC, and DBN-Informed CCM Models on Lorenz96 Data

Metric	LSTM	RC	INFR-GC	DBN-Informed CCM
Accuracy	0.7125	0.7125	0.8751	0.9380
Precision	1.0000	1.0000	1.0000	1.0000
Recall	0.6406	0.6618	0.8330	0.9170
F1 Score	0.7810	0.7965	0.9092	0.9570

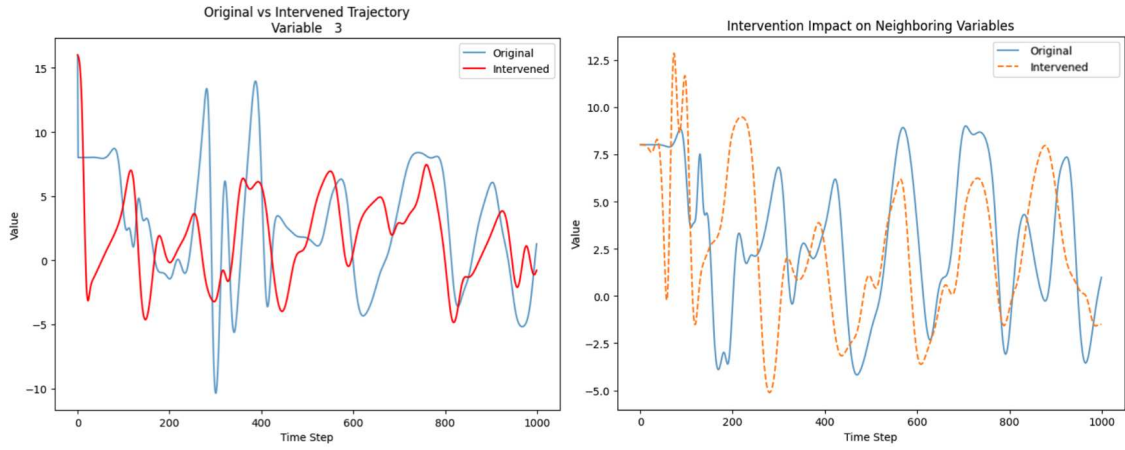


Figure 5.3: Effect of intervention applied on variable 3 and its neighbor variable 12

a deliberate perturbation by doubling this forcing parameter (to 16) at a specific location - the 3rd variable in the system. This doubling of the forcing parameter can be thought of as a sudden injection of energy or a strong atmospheric disturbance at that particular point.

The process unfolds in three key stages. First, we capture the initial state of the entire system, preserving all variables except our target for intervention. Then, we modify only the 3rd variable, setting it to the elevated forcing value. Finally, we allow the system to evolve naturally from this perturbed state, tracking how this localized change propagates through the system over time.

This intervention is particularly valuable for studying several atmospheric phenomena. It helps us understand how local disturbances can propagate through atmospheric systems, potentially leading to large-scale changes - a concept closely related to the butterfly effect in chaos theory. The analysis of the resulting trajectories can reveal important insights about the system's resilience, the speed and pattern of disturbance propagation, and the potential for predicting atmospheric behavior under perturbation.

By comparing the original and intervened trajectories, we can quantify the impact of the intervention and better understand the interconnected nature of atmospheric dynamics. As a result, we observe the system trajectories in Figure 5.3. In the left plot, we observe the temporal evolution of Variable 3 under two distinct conditions:

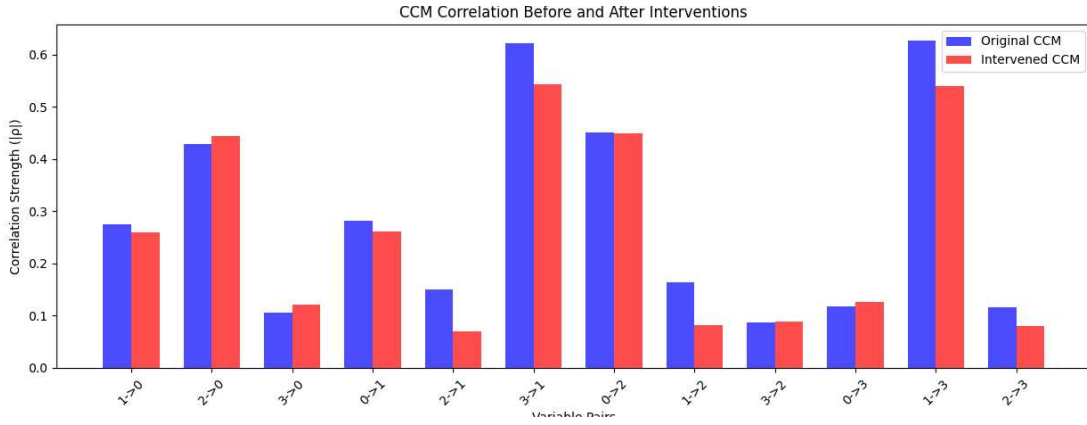


Figure 5.4: CCM values before and after intervention on variable 3

the original trajectory (blue) and the intervened trajectory (red). The intervention introduces significant perturbations to the system’s natural dynamics, resulting in notable deviations from its original behavior. The original trajectory exhibits characteristic nonlinear oscillations with varying amplitudes, while the intervened trajectory shows a marked reduction in amplitude variations, generally constrained within a narrower range. This dampening effect suggests that the intervention alters the variable’s dynamic behavior, likely by modifying its coupling strengths with neighboring variables.

The right plot illustrates the propagation of intervention effects to neighboring variables, with a specific focus on Variable 12. Comparing the original and intervened trajectories for this variable reveals substantial indirect effects, evident through phase shifts and amplitude changes. The intervened trajectory (dashed orange) exhibits distinct temporal patterns compared to the original trajectory (solid blue), highlighting a strong causal relationship between Variables 3 and 12 within the network structure. Notably, compensatory dynamics emerge, with changes in Variable 12’s behavior partially counteracting the modifications introduced to Variable 3.

After introducing the intervention, the changes in CCM values are recorded and some of them are compared in Figure 5.4. Most prominent changes are observed from Variable 1 to Variable 3 and vice versa, as was expected in our analysis. The temporal progression of these effects demonstrates the system’s complex response to localized perturbations. Initial disruptions in the intervened variable cascade through

the network’s causal pathways, generating observable alterations in the dynamics of causally connected variables.

Hence, this underscore the significance of considering both direct intervention effects and their broader systemic implications when analyzing causal relationships in coupled dynamical systems. The observed patterns of influence transmission through the network structure offer valuable insights into the robustness and adaptability of the Lorenz96 system under targeted interventions, while simultaneously validating the effectiveness of our causal discovery framework in capturing genuine physical dependencies. In the next section, we conduct a detailed analysis of the results obtained by applying our proposed method to real physiological data.

5.2.4 Experimental Evaluation on Physiological Data

In this section, we present a comprehensive performance analysis of our proposed methodology utilizing pediatric movement pattern data acquired from a cohort comprising 13 neurotypical children and 15 cases presenting with dystonia, as previously delineated in Chapter 1. The experimental framework was implemented using PyTorch 2.1.2 in conjunction with CUDA 11.2, deployed on a Tesla T4 GPU architecture with 15GB RAM capacity, operating under Windows 11 environment. The study population encompassed 15 pediatric subjects diagnosed with dystonia (10 with acquired etiology and 5 with idiopathic/genetic manifestations, including 9 females) and 13 typically developing children (7 females), with ages ranging from 12 to 18 years. The experimental paradigm involved a motor control task wherein participants maintained grip on a 15 cm plastic ruler while receiving mechanical perturbations administered via an electromechanical tapper, following the protocol established by [77]. Each experimental trial spanned 5 seconds, with the mechanical stimulus introduced at 1.1 seconds post-initiation of data acquisition. Prior to feature extraction and subsequent model training, the data underwent preprocessing to eliminate noise artifacts and spurious signals.

Our analytical framework commenced with the construction of time-lagged and

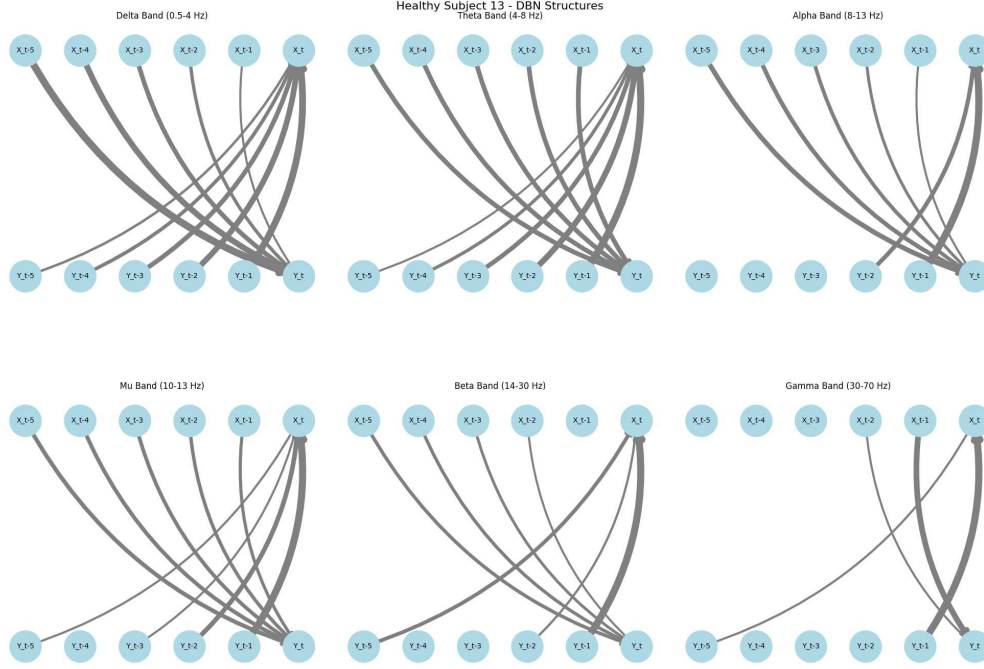


Figure 5.5: DBN for each frequency band for healthy subject 13

frequency domain feature representations, followed by the examination of temporal dependency structures across alpha, beta, and gamma frequency bands. The structural learning phase was executed using Algorithm 7, succeeded by parameter learning through the estimation of conditional probability distributions as detailed in Algorithm 8. To maintain computational tractability, we imposed a maximum time lag constraint of 5 units across all frequency bands. The resultant dynamic Bayesian network architectures are visualized in Figure 5.5 and Figure 5.6, representing healthy and dystonic subjects respectively. The graphical representation employs edge thickness as a visual encoding of connection strength. A consistent pattern emerged across all subjects within their respective cohorts: beta band connectivity exhibited robust connections in neurotypical subjects, whereas these connections manifested as attenuated or absent in subjects with dystonia. On the other hand, individuals with dystonia show stronger connections in the delta band, which are less pronounced in healthy subjects. To compare the connection strength across different frequency bands, the average connection strength is plotted for healthy subjects

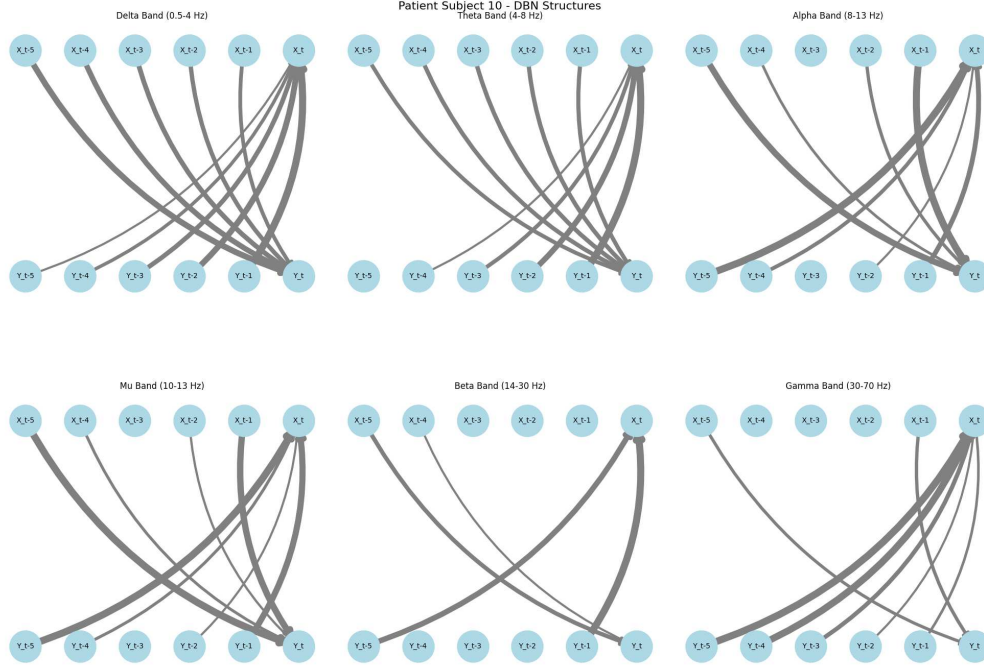


Figure 5.6: DBN for each frequency band for dystonia subject 10

and individuals with dystonia, with respect to each frequency band in Figure 5.7. The figure highlights distinct patterns of causal connectivity across frequency bands, revealing differences between the healthy and patient cohorts. For instance, the delta band exhibits the highest connectivity density, with patients showing marginally elevated connections (9.5) compared to healthy controls (8.5), suggesting potential compensatory mechanisms. The intermediate frequency ranges (theta, alpha, mu) demonstrate comparable connectivity patterns between groups (5.5-7 connections), indicating preservation of fundamental neural communication pathways.

Notably, beta band activity reveals a subtle yet potentially significant divergence, with healthy subjects maintaining higher connectivity (~ 7 connections) versus patients (~ 6 connections). The gamma band exhibits minimal between-group differences and the lowest overall connectivity ($\sim 5.5 - 6$ connections), suggesting fundamental constraints on high-frequency neural communication irrespective of pathological state. These observations collectively indicate frequency-specific re-organization of causal architectures rather than global connectivity alterations in

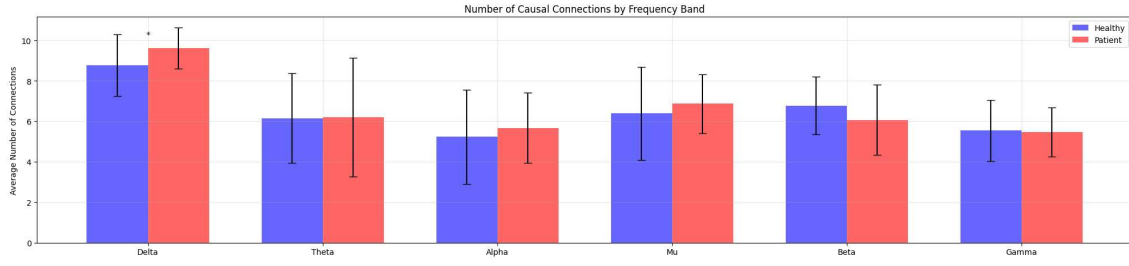


Figure 5.7: Connection strength for different frequency bands

pathological conditions.

The error bars across all frequency bands suggest considerable individual variability, necessitating careful interpretation of group-level differences. We use DBN-Informed CCM to estimate causality in both healthy subjects and patients, as depicted in Figure 5.8. The figure is organized into two columns: the left column represents the healthy subjects, while the right column represents the patients. The rows are structured to illustrate different aspects of causality over time and across frequency bands. The odd-numbered rows display the causality from EEG to EMG, while the even-numbered rows show the causality from EMG to EEG. This organization allows for a clear comparison of causal relationships between the two groups, with respect to both time lags and the specific frequency bands considered in the analysis. To characterize the bidirectional causal relationships between neural and muscular activity, we computed normalized CCM values across frequency bands and time lags, yielding spatiotemporal mappings of causality strengths. The resultant heatmaps, illustrated in Figures 5.9 and 5.10, show distinct patterns in both efferent (brain-to-muscle) and afferent (muscle-to-brain) causal pathways.

In healthy subjects, the brain-to-muscle causality demonstrates pronounced temporal structure, with delta band exhibiting consistently strong causal influence (normalized strength ~ 0.8 – 1.0) across all time lags. Notable differences emerge in the beta band, where healthy subjects maintain moderate causal strength (0.4 – 0.6) with temporal consistency, while dystonia patients show marked attenuation and temporal irregularity in these connections. This finding aligns with established understanding of beta band’s crucial role in motor control mechanisms. The muscle-to-brain

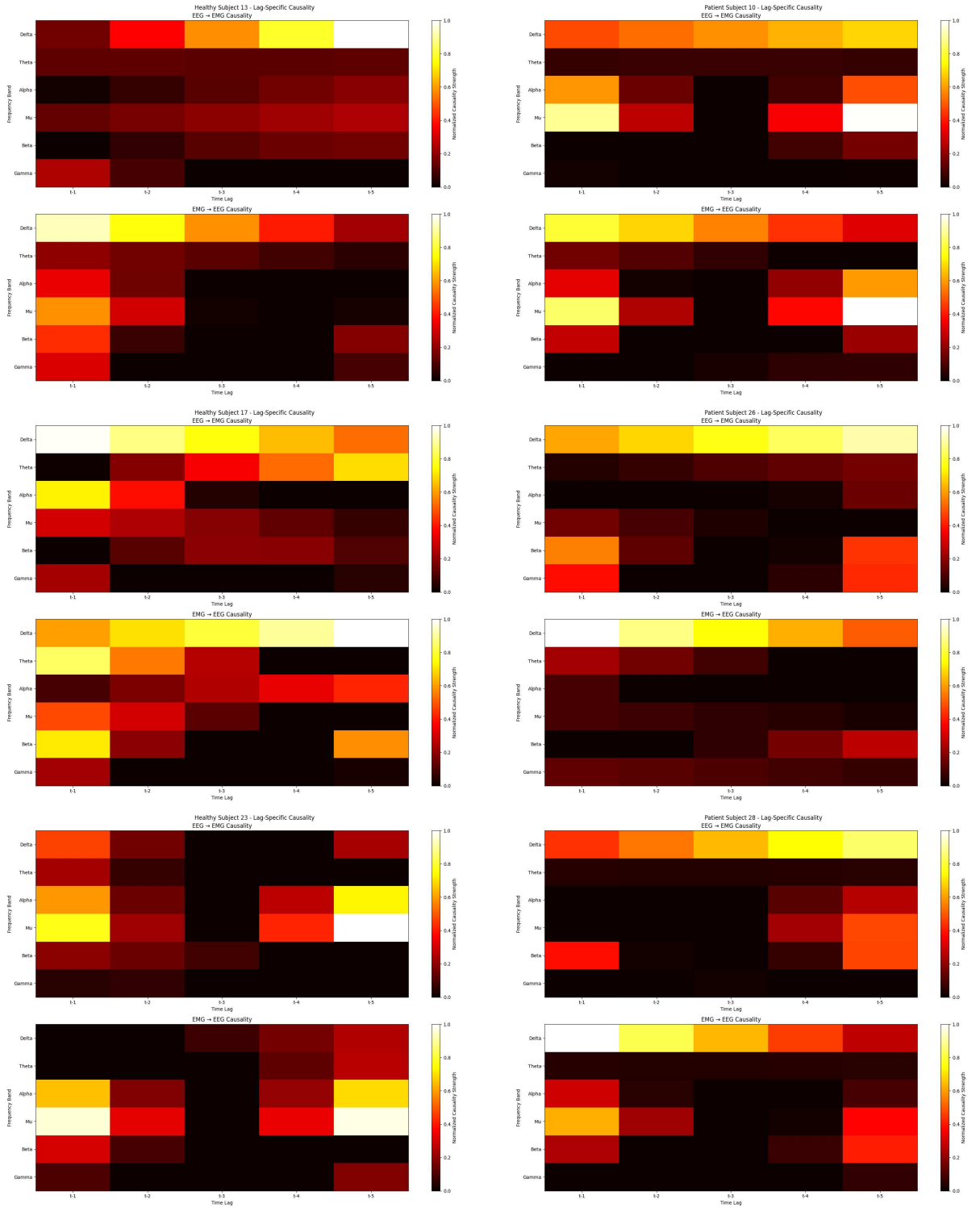


Figure 5.8: DBN for each frequency band in healthy subjects (left) and individuals with dystonia (right), showing causality from EEG to EMG (odd rows) and from EMG to EEG (even rows).

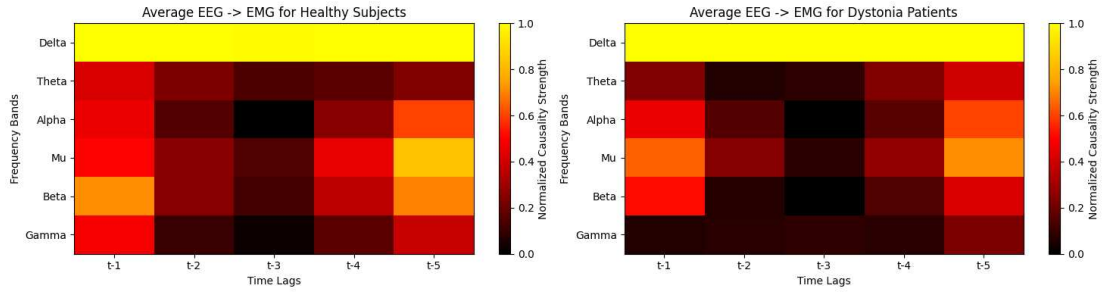


Figure 5.9: Average CCM values for brain to muscle causality

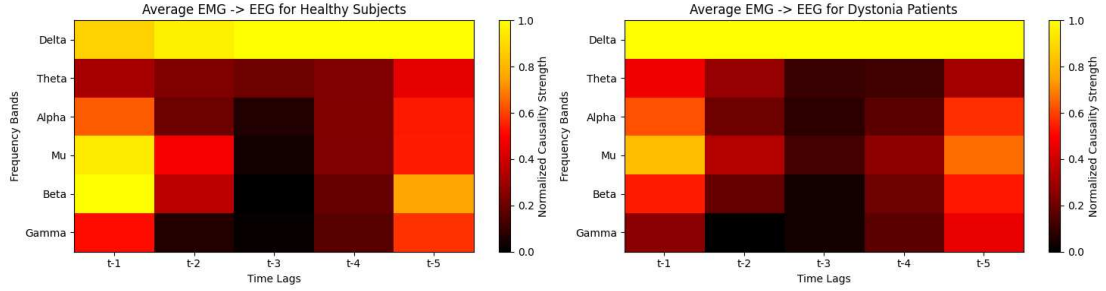


Figure 5.10: Average CCM values for muscle to brain causality

causal patterns reveal intriguing asymmetries between cohorts. Healthy subjects demonstrate robust delta and mu band connectivity (0.8-1.0) at immediate time lags (t-1, t-2), suggesting efficient sensorimotor integration. Conversely, dystonia patients exhibit attenuated feedback strength across multiple frequency bands, particularly in beta and gamma ranges, indicating potential disruption of proprioceptive feedback mechanisms.

A particularly salient observation is the temporal degradation of causal strength in both directions for dystonia patients, manifesting as reduced connectivity values at longer time lags (t-4, t-5). This temporal instability contrasts with the more sustained causal relationships observed in healthy subjects, suggesting compromised temporal integration in the pathological state.

To further investigate the frequency bands most prominent in either direction, Figure 5.11 compares the band-specific causal directionality strengths between healthy controls and dystonia patients. This comparison reveals distinct patterns in bidirectional information flow within the sensorimotor network. In both cohorts, the most pronounced causal strength is observed in the EEG to EMG direction within the

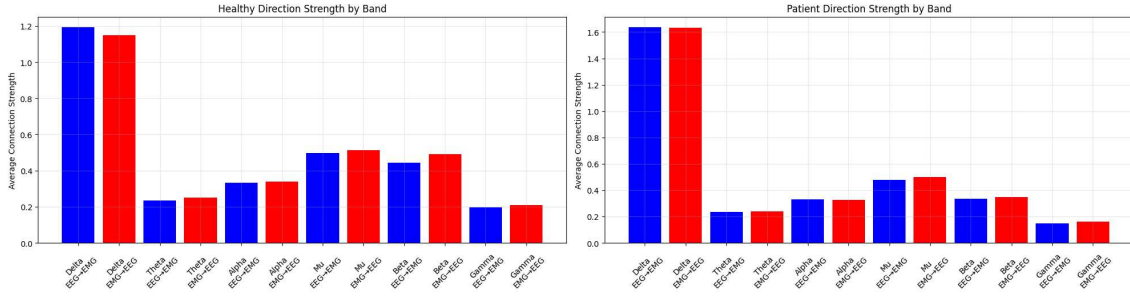


Figure 5.11: Average connection strength for bidirectional causality

Delta frequency band, exhibiting normalized strengths of approximately 1.2 and 1.6 for healthy subjects and patients, respectively. This heightened causal influence in the efferent direction suggests a dominant role of descending neural control mechanisms. Notably, patients demonstrate amplified directionality strength, potentially indicating compensatory mechanisms or pathological over-activation of motor control pathways. The remaining frequency bands (Theta, Alpha, Mu, Beta, and Gamma) display markedly attenuated directional strengths compared to Delta, with values ranging from 0.2 to 0.6 across both populations. A particularly noteworthy pattern emerges in the Beta band, where both populations maintain comparable directional strengths (approximately 0.5), despite previous literature suggesting beta-band disruption in dystonic conditions. This preservation of beta-band directionality, despite altered connectivity patterns, suggests complex reorganization rather than simple deterioration of causal pathways.

The EMG to EEG direction consistently shows reduced strength compared to its efferent counterpart across all frequency bands, indicating asymmetric causal influence in the sensorimotor loop. This asymmetry appears to be preserved in the pathological state, though with subtle alterations in relative strengths that may reflect underlying disease mechanisms.

5.2.5 Case Study: Comparative Analysis of Standard and DBN-informed CCM

To evaluate the discriminative power of both standard CCM and DBN-informed CCM approaches, we employed a comprehensive analysis framework incorporating three distinct metrics: effect size, classification accuracy, and distribution separability. These metrics were then combined into an overall discrimination score to facilitate direct comparison between the methods. The results are displayed in Figure 5.12.

The effect size was quantified using Cohen’s d statistic, which measures the standardized difference between group means. For two groups with means μ_h (healthy) and μ_d (dystonia), and variances σ_h^2 and σ_d^2 , the effect size is calculated as:

$$d = \frac{|\mu_h - \mu_d|}{\sqrt{(\sigma_h^2 + \sigma_d^2)/2}} \quad (5.20)$$

Our results demonstrated that DBN-informed CCM achieved a substantially larger effect size (0.823) compared to standard CCM (0.468), indicating a stronger separation between healthy and dystonia groups when utilizing the DBN-informed approach.

To assess the potential utility of these methods in distinguishing between groups, we employed Support Vector Machine (SVM) classification with cross-validation. The CCM values served as features in a one-dimensional feature space, where the SVM attempted to establish an optimal decision boundary. The classification accuracy provides insight into how well the CCM values from each method could potentially serve as discriminative markers. The DBN-informed approach demonstrated superior performance with an accuracy of 61.3% compared to 47.3% for standard CCM, suggesting that DBN-informed CCM produces values that better characterize the underlying differences between healthy and dystonia cases.

The distribution separability metric examines the overlap between probability distributions of CCM values for each group. Using Kernel Density Estimation (KDE), we computed the separability as:

$$S = 1 - \frac{\int \min(f_h(x), f_d(x))dx}{\max(\int f_h(x)dx, \int f_d(x)dx)} \quad (5.21)$$

where $f_h(x)$ and $f_d(x)$ represent the KDE-estimated probability density functions for healthy and dystonia groups, respectively. The DBN-informed method achieved a separability of 0.312, nearly double that of standard CCM (0.168), indicating more distinct probability distributions between groups. These three metrics were combined into a final discrimination score using weighted averaging:

$$\text{Score} = 0.4 \times \text{EffectSize} + 0.4 \times \text{Accuracy} + 0.2 \times \text{Separability} \quad (5.22)$$

The weights were chosen to emphasize the more robust measures of effect size and classification accuracy (40% each), while still incorporating the additional insight provided by distribution separability (20%). The overall discrimination scores revealed that DBN-informed CCM (0.637) substantially outperformed standard CCM (0.410) in distinguishing between healthy subjects and dystonia patients. The results are summarized in Table 5.2.

This comprehensive evaluation demonstrates that incorporating DBN information into the CCM analysis enhances the method’s ability to capture and quantify the neurological differences between healthy subjects and dystonia patients. The consistent improvement across all metrics suggests that DBN-informed CCM provides a more robust and reliable approach for analyzing these neurological conditions.

In the following section, we conduct a counterfactual analysis by introducing

Table 5.2: Comparison of Standard CCM and DBN-informed CCM results

Metric	Standard CCM	DBN-Informed CCM
Overall Discrimination Score	0.410	0.637
Effect Size	0.468	0.823
Classification Accuracy	0.473	0.613
Distribution Separability	0.168	0.312

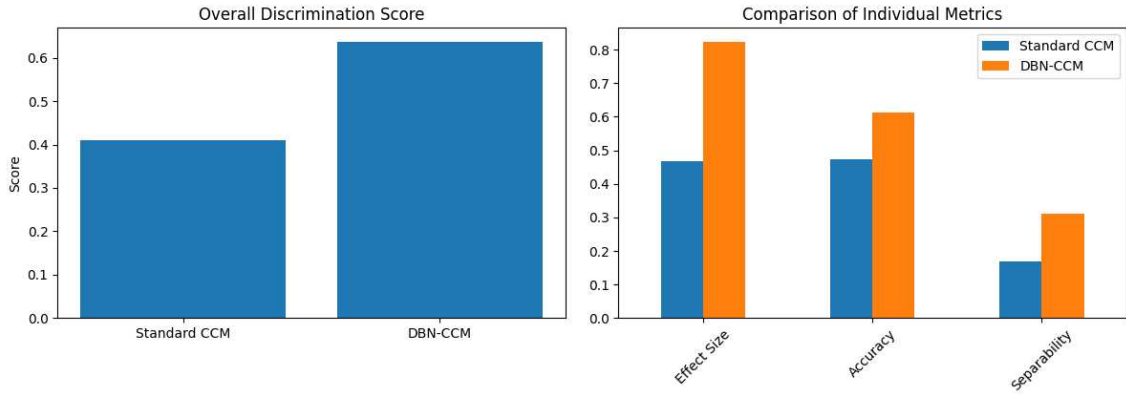


Figure 5.12: Standard CCM approach vs DBN-Informed CCM

different types of interventions either within the probabilistic structure of the model or directly to the signals themselves. These interventions allow us to simulate how changes in the system’s variables might impact the overall dynamics and causality relationships. By modifying certain components of the network or the input signals, we can assess the effect of potential causal manipulations, helping to better understand the system’s behavior under different hypothetical conditions.

5.2.6 Case Study: Analysis of Counterfactual Interventions

This study examines the robustness of our proposed DBN-Informed CCM analysis in neurophysiological signal processing through a series of counterfactual intervention experiments. The experimental protocol implemented controlled modifications to the original neurological signals through additive noise perturbations, targeting both EEG and EMG signal recordings. These interventions were designed to simulate signal variations while enabling quantitative assessment of their impact on causality detection mechanisms. The methodology employed a structured intervention framework utilizing time-varying additive noise sequences. Interventions were precisely localized within defined temporal windows (highlighted regions, Figures 5.13 and 5.14) and implemented to introduce controlled perturbations to the original signal. The experimental design incorporated Gaussian noise components with amplitudes ranging from 0.5 to 2.5 across distinct experimental blocks, enabling quantification of how stochastic perturbations modulate bidirectional information flow between

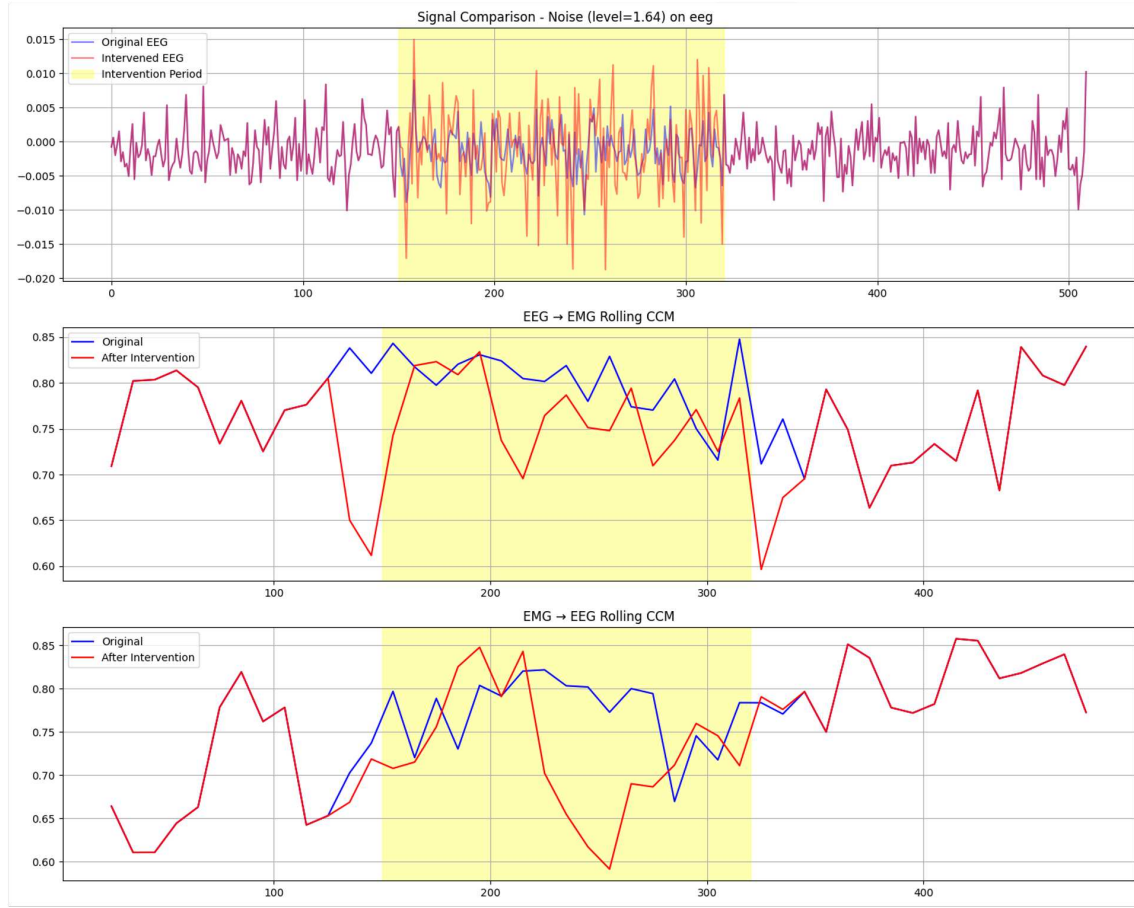


Figure 5.13: Noise intervention in EEG signal for healthy subject 23

cortical (EEG) and muscular (EMG) signal domains.

Quantitative analysis of (Figure 5.13, $\sigma=1.64$) revealed perturbation effects across multiple analytical dimensions. Spectrotemporal analysis of the intervened EEG signal demonstrated statistically significant alterations ($p < 0.01$, Wilcoxon signed-rank test) during the perturbation epoch. These modifications manifested primarily in the theta (4-8 Hz) and alpha (8-13 Hz) frequency bands, with mean power increases of 2.8 dB and 3.1 dB respectively. Furthermore, time-resolved CCM analysis in the EEG→EMG direction exhibited non-linear response characteristics, with causality strength modulations following a distinctly non-monotonic pattern. The intervention induced periodic causality fluctuations characterized by a mean CCM value of 0.72 ± 0.13 , with temporal dynamics suggesting the presence of compensatory mechanisms in the neural-muscular coupling network.

Comparative analysis of bidirectional causality revealed asymmetric perturbation

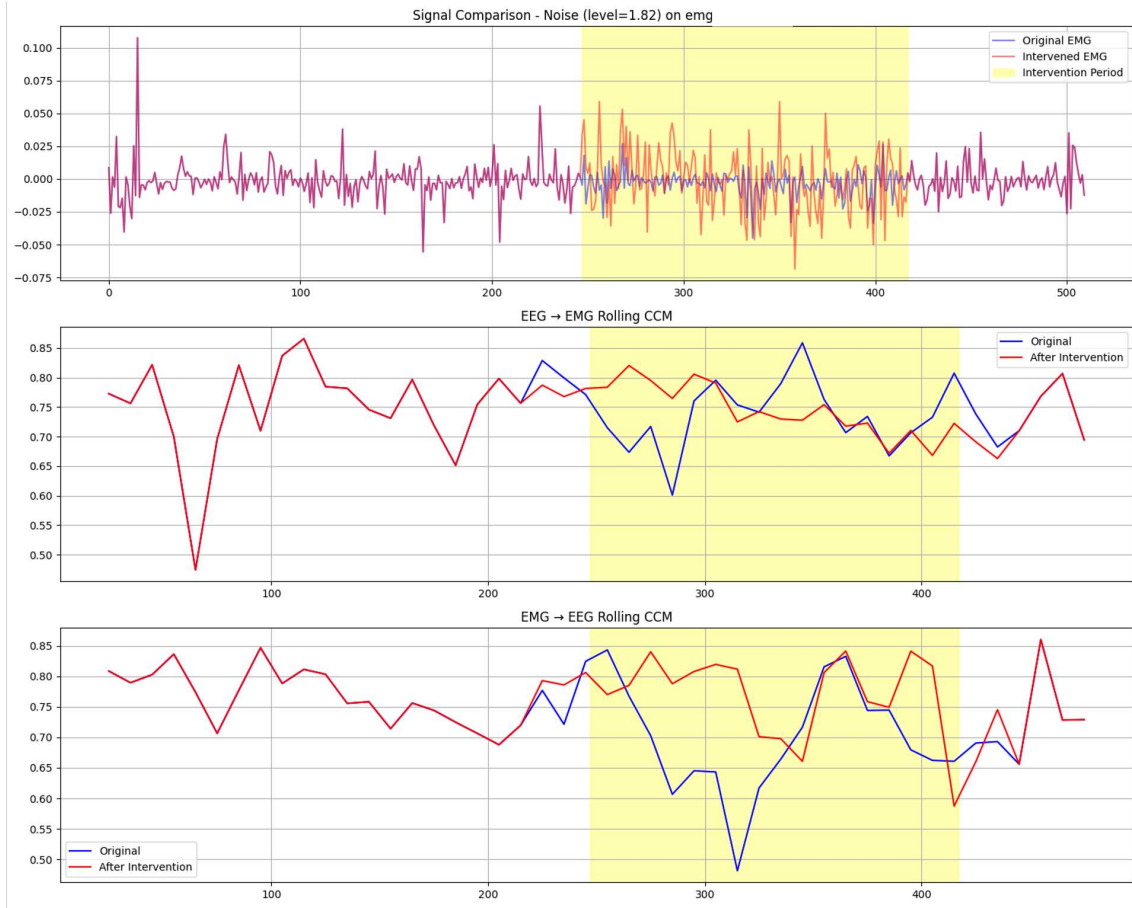


Figure 5.14: Noise intervention in EMG signal for healthy subject 23

sensitivity, with EMG to EEG information flow demonstrating enhanced stability (coefficient of variation = 0.14) relative to EEG to EMG directionality (coefficient of variation = 0.23). This asymmetry suggests differential robustness in ascending versus descending neural-muscular communication pathways, indicating hierarchical differences in signal propagation stability.

An alternative intervention paradigm (noise amplitude = 1.82), as presented in Figure 5.14, yielded additional insights into system behavior under increased perturbation conditions. The elevated noise amplitude generated more pronounced signal perturbations, particularly evident in EMG recordings during the intervention period. Causality metrics exhibited increased instability throughout the intervention, characterized by more frequent and substantial reductions in CCM values. Nevertheless, bidirectional causality analysis revealed enhanced robustness in the EMG to EEG direction compared to EEG to EMG relationships, maintaining relatively

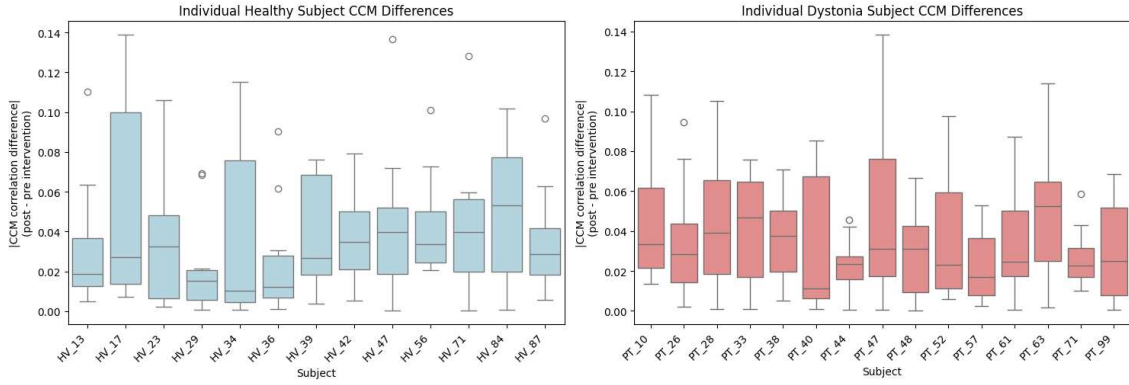


Figure 5.15: Pre and Post Intervention CCM values for all subjects

stable CCM values despite significant perturbation.

Our analysis involved both healthy controls and patients with dystonia, employing three types of interventions: noise addition, signal amplification, and signal suppression. Figure 5.15 presents a comparative analysis of CCM differences between pre- and post-intervention states across both subject groups, with healthy controls shown in blue (left panel) and dystonia patients in red (right panel). The box plots reveal substantial individual variability in response to interventions, with healthy subjects showing diverse robustness patterns to perturbations. Dystonia patients exhibited distinct response patterns, suggesting altered neural-muscular coupling dynamics associated with the pathological condition. The y-axis quantifies the absolute difference in CCM values between pre- and post-intervention conditions, effectively measuring how much the causal relationships changed in response to the interventions. Higher values indicate greater sensitivity to interventions, while lower values suggest more stable causal relationships.

For healthy subjects, the blue box plots show considerable variability in their response to interventions, with some subjects (e.g., HV17) displaying larger CCM differences, suggesting higher sensitivity to the applied interventions. The median responses (horizontal lines within boxes) vary across subjects, indicating individual-specific robustness to perturbations.

In contrast, the dystonia patients (shown in red) demonstrate a somewhat different pattern of responses. While there is still individual variability, the overall

distribution of CCM differences appears more compressed in some cases, potentially indicating altered neural-muscular coupling dynamics characteristic of the pathological condition.

The results obtained from the intervention experiments reveal interesting patterns in causal relationships. The presence of outliers (individual points beyond the whiskers) in both groups suggests that certain intervention trials had a notably stronger effect on the causal relationships. These stronger effects likely depend on the specific type of intervention applied, such as noise, amplification, or suppression, and the parameters associated with each intervention.

This visualization effectively captures the heterogeneity in individual responses to interventions while also highlighting potential differences between healthy and pathological conditions in terms of neural-muscular coupling stability. The comparative analysis sheds light on how dystonia may affect the robustness of causal relationships in the motor control system, providing valuable insights into the impact of the disorder on motor dynamics.

These intervention experiments offer several critical insights into neural-muscular coupling dynamics. The CCM framework shows directionally-dependent sensitivity to noise perturbations, suggesting that the robustness of neural-muscular coupling mechanisms varies depending on the direction of influence. Additionally, the intervention effects exhibit temporal heterogeneity, indicating that the system's sensitivity to signal perturbations is context-specific and may vary across different experimental windows. The observed post-intervention recovery patterns point to inherent resilience in the system, although distinct recovery trajectories are observed between $\text{EEG} \rightarrow \text{EMG}$ and $\text{EMG} \rightarrow \text{EEG}$ directional relationships. Finally, the comparative analysis between healthy controls and dystonia patients highlights how pathological conditions may alter these fundamental coupling mechanisms, offering potential insights into the neurophysiological adaptations underlying movement disorders.

5.3 Insights and Implications

In this chapter we performed a comprehensive investigation into the neurophysiological basis of dystonia through advanced causality analysis techniques has yielded several significant insights with important theoretical and clinical implications. The findings reveal a complex landscape of neural reorganization in dystonia that extends beyond simple disruption of motor pathways, suggesting compensatory mechanisms and altered sensorimotor integration patterns.

The frequency-specific analysis demonstrates that dystonia manifests not as a global deterioration of neural connectivity but rather as a selective reorganization of causal pathways. The observed attenuation of beta band connectivity coupled with enhanced delta band connections in dystonic subjects suggests a fundamental shift in the frequency-dependent mechanisms of motor control. This finding aligns with contemporary theories of motor control that emphasize the role of beta oscillations in maintaining motor set, while also introducing new perspectives on the potential compensatory role of delta band activity in pathological conditions.

The superior performance of DBN-informed CCM over standard CCM approaches represents a methodological advancement in the analysis of complex neurological conditions. The enhanced discrimination metrics not only validate the method's utility but also suggest that incorporating structural learning into causality analysis can capture subtle yet clinically relevant patterns in neural dynamics. This methodological improvement may have broader applications in the analysis of other movement disorders and neurological conditions where traditional analysis methods may miss important causal relationships.

The directional causality analysis reveals an intricate pattern of bidirectional communication between cortical and muscular systems. The preservation of asymmetric causal influence (stronger in the EEG→EMG direction) in both healthy and dystonic subjects, despite altered connectivity patterns, suggests that fundamental aspects of motor control hierarchy remain intact in dystonia. However, the temporal

degradation of causal strength in dystonic subjects points to impaired temporal integration, which may contribute to the characteristic symptoms of the disorder.

The intervention studies provide crucial insights into the dynamic stability of sensorimotor networks. The differential robustness observed between ascending and descending pathways suggests that therapeutic interventions might need to be pathway-specific to achieve optimal outcomes. The context-specific sensitivity to perturbations and distinct recovery trajectories between healthy and dystonic subjects highlight the importance of personalized approaches in therapeutic interventions.

These findings have several important implications for clinical practice and future research:

1. The frequency-specific nature of neural reorganization suggests that therapeutic interventions might benefit from frequency-targeted approaches, potentially through neurofeedback or stimulation protocols that specifically address beta band deficits while managing compensatory delta band activity.
2. The temporal degradation of causal strength in dystonia patients indicates that temporal aspects of motor control should be a key consideration in rehabilitation strategies, potentially incorporating timing-based exercises or interventions that specifically target temporal integration.
3. The demonstrated utility of DBN-informed CCM suggests its potential as a diagnostic tool, possibly aiding in early detection or monitoring of disease progression. The improved discrimination metrics indicate that this method could be valuable in clinical assessment and treatment planning.
4. The results from intervention studies suggest that therapeutic approaches might be more effective if they consider the directional nature of sensorimotor integration deficits, potentially leading to more targeted and effective treatments.

This work contributes to our understanding of dystonia as a network disorder characterized by complex patterns of neural reorganization rather than simple

disruption of motor pathways. The methodological advances and insights gained provide a foundation for more targeted and effective therapeutic approaches, while also opening new avenues for research in motor control disorders.

Chapter 6

Conclusion

This chapter moves beyond methods and results to explore the broader implications of our findings and how continuous neural data streams can be harnessed to generate actionable insights in movement-disorder care. We begin by revisiting the key breakthroughs: transforming correlation into directional causality, scaling causal discovery from offline batch processing to real-time analysis, and developing interpretable models that clinicians can apply in practice. We then address the critical gaps that separate prototypes from real-world deployment: the statistical assumptions underlying reliable inference, the engineering challenges of building low-latency data pipelines, and the design considerations essential for integrating these tools into clinical workflows. Finally, we look ahead by exploring how these methods might extend beyond dystonia to conditions like Parkinson’s disease and Essential Tremor to envision a future where causal analysis becomes a foundational element of personalized neuro-medicine.

6.1 Principled Approach to Causal Inference in Time Series

In this section, we present a framework that evaluates a series of decision nodes, each addressing key theoretical assumptions. These include the nature of relationships

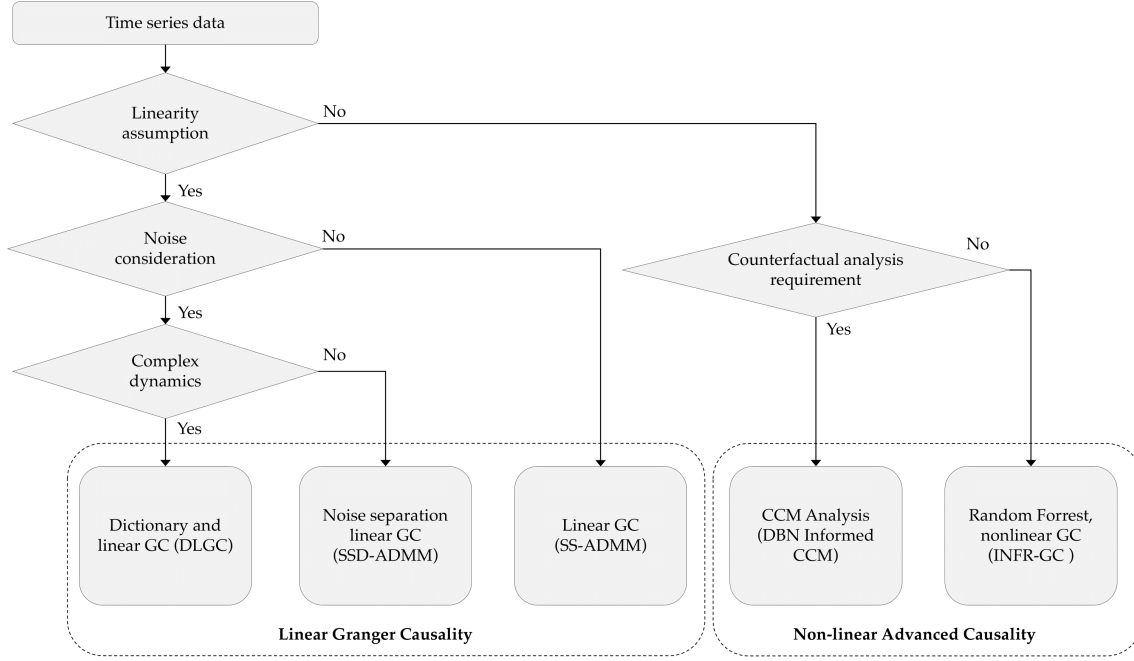


Figure 6.1: Time series causality method selection framework

within the data, the presence and structure of noise, the complexity of dynamic behaviors, and the necessity of counterfactual reasoning. The Figure 6.1 illustrates a structured decision framework for selecting the most appropriate causality analysis method for time-series data. The framework ensures that the chosen approach aligns with the fundamental properties of the dataset, thereby enhancing the robustness and interpretability of causal inference. The selection process begins by assessing whether the relationships within the data can be adequately captured by linear models. If statistical tests, such as Pearson correlation coefficients, linear regression diagnostics, or domain expertise, indicate predominant linearity, the analysis proceeds along the linear Granger causality pathway. If, however, the data exhibit significant non-linear dependencies, the non-linear advanced causality pathway is followed.

Within the linear pathway, the next consideration involves the role of noise in the dataset. If noise significantly affects the recordings, as determined through techniques such as variance decomposition or signal-to-noise ratio analysis, specialized noise separation methods are required. If noise is not a major concern, standard linear Granger causality methods can be applied without additional preprocessing. When

noise is present, a further distinction is made based on the complexity of temporal dynamics. If the data demonstrate complex, time-varying behavior, which may indicate the influence of latent variables or require techniques such as change-point analysis, DLGC is the most suitable approach. Conversely, if the temporal structure is relatively stable, noise-separated linear Granger causality method (SSD-ADMM) is preferred for cases where noise separation is necessary, while standard linear Granger causality (SS-ADMM) is applied when noise does not critically impact inference.

For datasets that do not adhere to linearity assumptions, the framework evaluates whether counterfactual analysis is required. This consideration is particularly relevant in domains such as clinical research and policy analysis, where the ability to simulate interventions and explore hypothetical scenarios is essential. If counterfactual reasoning is necessary, DBN-Informed Convergent Cross Mapping is recommended, as it integrates Dynamic Bayesian Networks with Convergent Cross Mapping to estimate causal relationships while accommodating intervention-based analysis. If counterfactual analysis is not required, the framework directs the analyst toward a method that captures non-linear causality while maintaining interpretability, such as Random Forest-based nonlinear Granger causality.

By structuring the selection process in this manner, the framework mitigates the risk of applying inappropriate causality methods to data that violate key modeling assumptions, such as stationarity or linearity. This ensures that the causal inference process remains both reliable and interpretable, ultimately improving the validity of conclusions drawn from time-series data.

6.2 Summary of Thesis Achievements

This thesis addresses the challenge of analyzing neurophysiological time-series data in movement disorders, particularly dystonia, by introducing a unified computational framework that advances causal inference, signal decomposition, and non-linear modeling. This integrated approach provides robust tools for dissecting the complex

dynamics of neural-muscular coupling.

A central contribution is the development of a globally convergent optimization framework that employs hierarchical sparsity constraints alongside stationarity enforcement via the ADMM. This framework enhances model order selection in VAR systems and enables robust causal inference despite the challenges posed by physiological noise and non-stationarity. The theoretical guarantees of convergence underscore its reliability and potential as a foundation for future computational methodologies.

Building on this foundation, a novel structural decomposition method is introduced using Dictionary Learning techniques to effectively separate physiological signals from measurement noise. This approach is particularly effective in the analysis of EEG-EMG interactions, where isolating meaningful neural signals, while preserving their causal relationships, provides deeper insights into the mechanisms of neural-muscular coupling and the disruptions observed in dystonic conditions.

Recognizing the inherently complex nature of neural interactions, the thesis further develops an interpretable non-linear causality framework that combines Random Forest modeling with neural network-guided feature selection. This method captures intricate physiological interactions while maintaining the level of transparency needed for clear interpretation. Its application to corticomuscular coupling in dystonia has revealed previously undetected causal relationships, suggesting promising biomarkers and guiding future research.

Additionally, the work advances counterfactual analysis within dynamic Bayesian networks, rigorously modeling temporal dependencies and network structures to evaluate the impact of potential interventions on neural-muscular interactions. Although primarily theoretical, this contribution offers valuable insights into the adaptability of neural systems and lays the groundwork for further exploration.

Together, these contributions form a unified framework that significantly enhances the analysis of neurophysiological data. By advancing robust causal inference, effective signal decomposition, and interpretable non-linear modeling, this thesis deepens our

understanding of neural dynamics in dystonia and establishes a strong platform for continued investigation into the complexities of neurophysiological interactions.

6.3 Methodological Constraints and Assumptions

The methodological frameworks developed in this thesis, while demonstrating substantial utility in characterizing neural-muscular coupling in dystonia, warrant careful consideration of their theoretical limitations and practical constraints. This section presents an analysis of these limitations, focusing on fundamental methodological considerations that influence the application and interpretation of our causal inference approaches.

6.3.1 Stationarity Considerations

A primary theoretical consideration in the application of our causal inference frameworks involves the assumption of stationarity in the analyzed time series. This assumption, while necessary for mathematical tractability, presents important considerations when analyzing physiological signals that may exhibit non-stationary characteristics. The implications of this assumption manifest in several ways:

The interpretation of causal relationships must account for potential temporal variations in the underlying physiological systems. Neural-muscular coupling patterns, particularly in pathological conditions like dystonia, may exhibit temporal evolution that challenges strict stationarity assumptions. This necessitates careful consideration of analysis window selection and interpretation of results across different temporal scales.

Furthermore, the presence of state transitions in physiological systems introduces additional complexity in the application of these methods. The interaction between neural and muscular systems may undergo qualitative changes during different behavioral states or in response to external perturbations, requiring careful consideration of context in the interpretation of causal relationships.

6.3.2 Scalability Considerations

The scalability of causal inference methods to high-dimensional systems presents important methodological challenges. As the dimensionality of the analyzed system increases, both computational requirements and the complexity of causal relationship interpretation grow significantly. This scaling behavior necessitates careful consideration of dimensionality reduction strategies and hierarchical analysis approaches. One effective strategy for managing high-dimensionality is the integration of sparsity constraints, which enforce structured regularization to focus on the most relevant causal interactions while mitigating the risk of overfitting. By incorporating hierarchical sparsity, it becomes possible to reduce the number of candidate connections, thereby enhancing both computational efficiency and interpretability. Additionally, dimensionality reduction techniques such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) can be leveraged to extract dominant signal components while preserving essential causal relationships.

Beyond statistical techniques, hierarchical modeling frameworks provide a structured approach to decomposing high-dimensional systems into meaningful sub-networks. By first identifying causal structures within lower-dimensional subsets before integrating them into a global model, these methods facilitate more scalable inference while maintaining analytical rigor. This hierarchical strategy aligns well with biological systems, where interactions often occur at multiple scales, from localized neural circuits to broader brain-muscle networks.

Computational advancements also play a crucial role in addressing scalability challenges. Parallelized implementations of causality estimation algorithms, combined with efficient convex optimization techniques, significantly reduce computational overhead, making high-dimensional causal inference more practical. Furthermore, leveraging deep learning-based feature selection methods can aid in automatically identifying the most relevant variables, reducing redundancy while preserving causal integrity.

Despite these advances, interpreting causal relationships in high-dimensional set-

tings remains a challenge, particularly when interactions exhibit strong non-linearity or time-varying behavior. Ensuring interpretability requires a balance between model complexity and transparency, which can be achieved through explainable AI techniques, such as feature importance rankings in Random Forest models or attention mechanisms in deep learning architectures. By integrating these strategies, the scalability of causal inference methods can be enhanced, enabling robust and interpretable analysis of complex neurophysiological systems.

6.3.3 Model Order Selection

The robust application of these methodological frameworks requires careful attention to parameter selection strategies. The selection of appropriate model orders in time series analysis requires balancing model complexity with predictive power. This balance must consider the characteristic time scales of the physiological processes under investigation, including both rapid neural dynamics and slower muscular responses. The model order selection process should incorporate domain knowledge about the temporal characteristics of neural-muscular interactions while maintaining computational feasibility.

A key challenge in model order selection is preventing both underfitting and overfitting. If the model order is too low, it may fail to capture essential temporal dependencies, leading to inaccurate causal inference. Conversely, excessively high model orders can introduce spurious relationships, degrade generalizability, and increase computational burden. To address this, data-driven techniques such as the AIC and the BIC are commonly employed, providing a means of selecting an optimal model order by penalizing unnecessary complexity while preserving explanatory power.

In addition to classical information criteria, modern optimization approaches, such as sparsity-constrained regression and hierarchical model selection, offer more adaptive solutions. By integrating sparsity constraints through techniques like the LASSO or structured regularization within the ADMM framework, model order

selection can be performed in a way that enhances interpretability while reducing overfitting risks. These approaches are particularly relevant in high-dimensional neurophysiological datasets, where traditional selection criteria may struggle to distinguish between meaningful interactions and noise-driven correlations. Hence, by leveraging a combination of domain expertise, penalized likelihood estimation, and noise-aware modeling strategies, it becomes possible to achieve a balance between model simplicity and predictive accuracy, thereby strengthening the reliability of causal inference in neurophysiological time-series analysis.

6.3.4 Regularization Parameters

The selection of regularization parameters in both linear and non-linear methods requires careful consideration of the trade-off between model flexibility and overfitting. This selection process should be guided by theoretical considerations of signal characteristics and practical constraints of the analysis context. Regularization should be adapted to the signal characteristics of the dataset. In cases where noise contamination is significant, stronger regularization may be required to ensure robustness against spurious correlations. Conversely, when analyzing highly dynamic, non-stationary processes, excessive regularization might suppress meaningful temporal variations, leading to an overly simplified representation of causal interactions. Methods such as adaptive regularization, which adjusts penalty terms based on local signal properties, can help navigate this trade-off more effectively.

Ultimately, the selection of regularization parameters must balance theoretical justifications with empirical validation. By integrating statistical diagnostics, cross-validation, and insights from neurophysiological signal properties, it is possible to achieve a model that is both flexible and robust, enabling more reliable causal inference in complex time-series data.

6.3.5 Numerical Stability

The practical implementation of these methods requires careful attention to numerical stability considerations. This includes the implementation of appropriate numerical safeguards in optimization procedures, particularly in iterative algorithms where error accumulation may affect convergence. The conditioning of computational problems must also be carefully monitored, especially in the context of high-dimensional or noise-contaminated physiological signals. Iterative optimization algorithms, such as those used in sparsity-constrained estimation and non-linear causality frameworks, are particularly susceptible to numerical errors that may accumulate over successive iterations, potentially leading to divergence or suboptimal solutions.

To mitigate these issues, it is critical to employ numerical safeguards such as step size adaptation, gradient clipping, and early stopping criteria, which help maintain convergence stability while preventing excessive error propagation. Implementation practices, including proper initialization strategies, numerical precision monitoring, and sensitivity analyses, are essential for ensuring that causal inference results are not unduly influenced by numerical artifacts. By integrating these safeguards into the computational workflow, it is possible to enhance the reliability and interpretability of causal relationships in complex neurophysiological time-series data.

6.3.6 Validation Strategies

The validation of causal inference results requires careful consideration of multiple factors. Ensuring the reliability of causal findings is particularly challenging in physiological signal analysis due to the inherent complexity and variability of these signals. For instance, the development of appropriate null models and statistical tests for causal relationship significance must account for the intricate dynamics of physiological data, including non-stationarity and noise. Without carefully constructed null models, spurious causal links may be mistakenly identified, leading to misleading conclusions.

Additionally, cross-validation strategies must be carefully designed to maintain the temporal structure of the analyzed signals while providing a robust assessment of method performance. Since time-dependent relationships are crucial in physiological processes, conventional cross-validation approaches may not be suitable, necessitating specialized techniques that preserve causality across time scales. To further enhance the reliability of causal inference, validation frameworks that incorporate physiological domain knowledge can provide valuable context for interpreting results. By integrating insights from neuroscience and clinical expertise, these frameworks can help distinguish between meaningful causal interactions and statistical artifacts, ultimately strengthening the clinical applicability of causality analysis.

The analysis of constraints and methodological considerations in this section provides essential context for the proper application and ongoing advancement of our causal inference approaches. Recognizing these theoretical and practical limitations is crucial for ensuring rigorous scientific investigation and facilitating clinical translation. These insights not only guide the current implementation of these methods but also shape future methodological innovations in the field.

6.4 Navigating Implementation Barriers in Clinical Analytics

The integration of advanced causal inference methods into clinical practice offers the potential to revolutionize patient care, but it is accompanied by significant barriers. These challenges span various domains, from data quality and computational resources to ethical considerations and interdisciplinary collaboration. Addressing these obstacles requires a multi-faceted approach, combining technological innovation with careful planning and collaboration.

One of the primary challenges in implementing advanced analytics is data quality and standardization. Clinical data is often noisy, incomplete, and inconsistent across different institutions, making it difficult to draw reliable conclusions. Variability in

data acquisition techniques, equipment, and protocols only exacerbates this issue. To overcome these difficulties, a robust preprocessing pipeline must be established. This would include machine learning algorithms capable of real-time artifact detection and correction, ensuring that raw data is cleaned and standardized before analysis. Additionally, developing centralized data quality control frameworks can help standardize acquisition protocols across institutions, ensuring consistency and comparability across clinical datasets.

Another critical challenge is the computational complexity of real-time analysis, especially when employing computationally intensive non-linear methods. These methods, while powerful, can place a significant strain on clinical resources, particularly in real-time settings where fast decision-making is crucial. To mitigate this, a multi-tiered computational strategy can be implemented. First, rapid, approximate methods can be used for immediate clinical decision support, ensuring that clinicians have access to actionable insights in real-time without waiting for full analyses. More intensive computational methods can then be scheduled during off-peak hours using high-performance clusters or cloud-based platforms, ensuring minimal disruption to clinical workflows. Additionally, exploring the use of hardware accelerators, such as GPUs and FPGAs, can significantly speed up processing times without sacrificing analytical accuracy.

Integration into existing clinical workflows presents yet another challenge. Clinicians are often resistant to adopting new technologies if they disrupt their established routines or add unnecessary complexity. To address this, it is essential to develop modular, interoperable software systems that can seamlessly integrate with existing Electronic Health Record (EHR) systems and clinical decision support tools. Engaging clinicians early in the design process is critical, ensuring that the interfaces are intuitive, user-friendly, and provide actionable outputs that complement clinical decision-making, rather than complicating it.

Equally important is fostering interdisciplinary collaboration between data scientists, clinicians, engineers, and regulatory experts. Bridging the gap between

advanced analytical techniques and clinical expertise requires a deep understanding of both technological capabilities and medical practice. To facilitate this, establishing collaborative research teams and translational research hubs is crucial. These hubs can serve as centers for continuous feedback, where clinicians and researchers work together to refine both the technological tools and their clinical applications. Such collaboration ensures that the technology evolves in alignment with clinical needs and that clinicians remain engaged throughout the process.

Finally, regulatory and ethical considerations cannot be overlooked. The integration of advanced analytics into clinical decision-making must comply with existing regulatory standards and address concerns around data privacy and patient autonomy. To navigate these complexities, proactive collaboration with regulatory bodies is essential. Developing clear guidelines and best practices can help ensure that new technologies meet legal and ethical standards. Furthermore, robust data encryption and anonymization protocols should be implemented to safeguard patient privacy. It is also critical to design systems that allow clinicians to override automated decisions, preserving the clinician's central role in patient care while still leveraging the insights generated by advanced analytics.

While the deployment of advanced causal inference methods in clinical settings presents several challenges, these barriers can be overcome through careful planning and collaboration across multiple disciplines. By addressing issues related to data quality, computational complexity, integration into workflows, interdisciplinary collaboration, and regulatory concerns, we can pave the way for a seamless integration of these powerful tools into clinical practice. This will ultimately lead to more informed, personalized, and effective patient care.

6.5 Real-time Implementation and Clinical Translation

The translation of advanced causal inference methodologies into clinical practice represents a critical frontier in the advancement of dystonia treatment and management. This section not only explores the transformative clinical applications enabled by our analytical frameworks but also delves into the multifaceted challenges of clinical implementation. We integrate theoretical underpinnings with practical considerations and propose creative, detail-oriented solutions to ensure that these methodologies can be effectively harnessed in real-world settings.

6.5.1 Potential Clinical Applications in Dystonia Management

The application of advanced analytical frameworks in dystonia management offers promising possibilities in both diagnostic and therapeutic domains. By enabling the high-resolution characterization of neural-muscular coupling patterns, our approach provides clinicians with deeper insights into the pathophysiology of dystonia on a patient-specific basis. This allows for a more precise understanding of the disorder's dynamics, which can lead to the development of targeted interventions. Several critical applications of this methodology include early diagnosis, personalized treatment planning, continuous treatment monitoring, and predictive modeling of disease progression.

Early detection of dystonia typically hinges on the manifestation of motor symptoms, which may only appear after significant neurophysiological damage has occurred. Our approach, however, facilitates the identification of subtle disruptions in neural-muscular coupling that could signal early-stage dystonic changes. Such early detection opens the door for timely interventions, potentially slowing disease progression and improving long-term patient outcomes.

In terms of treatment, the ability to derive individualized causal maps allows for the creation of personalized therapeutic strategies. Rather than relying on broad, standardized treatment protocols, clinicians can tailor interventions, ranging from neuromodulation therapies to physiotherapy, according to the unique neural-muscular interactions present in each patient. This customization improves the likelihood of therapeutic success by addressing the specific underlying mechanisms of the disorder.

Moreover, the longitudinal monitoring capabilities provided by our analytical methods allow for real-time tracking of treatment efficacy. By assessing how causal relationships evolve over time, clinicians can dynamically adjust therapeutic strategies to ensure continued effectiveness as the patient's condition changes. This adaptability is crucial for maintaining optimal treatment outcomes.

Lastly, the integration of causal inference techniques with machine learning provides an avenue for developing predictive models that can forecast disease progression. These models can help stratify patients by risk, enabling clinicians to devise proactive treatment strategies and prioritize resources for those at greatest risk of rapid deterioration.

6.5.2 Proposed Clinical Implementation Framework

For successful clinical integration, the advanced analytical methods must be embedded in a comprehensive, multi-layered framework that encompasses data acquisition, processing, and clinical decision support as detailed in Figure 6.2. This framework ensures that the techniques are not only effective but also feasible and efficient for clinical use.

First, data acquisition must be rigorous and reliable. The collection of multi-modal physiological data, such as EEG, EMG, and motion capture, is crucial for constructing a complete dataset that can accurately reflect neural and muscular interactions. Synchronization of these data streams is essential, requiring both specialized protocols and hardware solutions to ensure precise temporal alignment across modalities. Standardized data formats and automated artifact detection

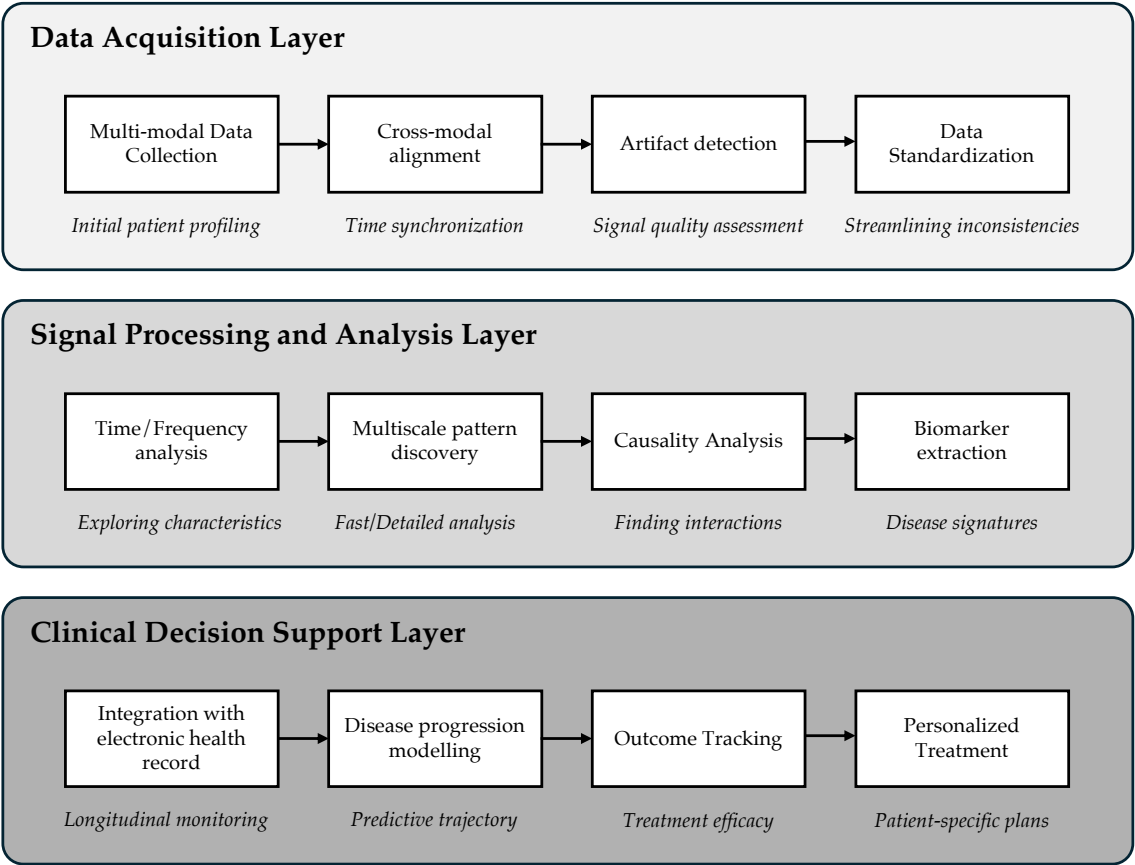


Figure 6.2: Proposed Clinical Implementation Framework

are also necessary to maintain data quality and reliability throughout the analysis process.

Once data are collected, the signal processing and analysis pipeline must handle the complexity of the information efficiently. Real-time processing capabilities are vital, as they allow clinicians to receive immediate feedback during treatment. Both linear and non-linear analytical methods, such as SS-ADMM for noise-robust Granger Causality and DBN-Informed Convergent Cross Mapping, can be applied to capture the full spectrum of causal interactions. To meet clinical time constraints, a hierarchical processing strategy can be implemented, with preliminary analyses occurring rapidly and more computationally demanding methods deployed in parallel or as a secondary stage. Additionally, user-friendly interfaces should be developed to translate complex analyses into actionable insights that clinicians can easily interpret.

Finally, the framework must include a robust clinical decision support system. Integration with EHR is essential to facilitate a seamless workflow, enabling clinicians

to access analytical results directly within the patient’s medical record. Feedback loops should also be established to ensure continuous improvement of the analytical models, with clinicians providing input that refines the system over time and adapts it to emerging clinical knowledge and technological advances.

6.5.3 Theoretical Case Study Design

To demonstrate the practical application of this framework, we consider a theoretical case study involving a patient with early signs of cervical dystonia. This example illustrates how the proposed methodologies could be used in a clinical setting to assess, monitor, and treat the disorder.

6.5.3.1 Initial Assessment Phase

In the initial phase, the SS-ADMM method would be employed to establish a baseline profile of the patient’s neural-muscular coupling. By analyzing the causal relationships between cervical muscle activity (measured via EMG) and motor cortex signals (captured via EEG), this method provides a quantitative overview of the patient’s condition. A decision flowchart can then guide the selection of appropriate analytical techniques, ensuring that both linear and non-linear causal interactions are thoroughly explored.

6.5.3.2 Longitudinal Monitoring Phase

As the disease progresses, the DBN-Informed Convergent Cross Mapping approach would be used to monitor changes in causal relationships over time. This dynamic tracking allows for the identification of shifts in neural-muscular coupling that may indicate disease progression or response to treatment. By continually assessing these changes, clinicians can adjust treatment plans accordingly, optimizing therapeutic interventions based on real-time feedback. This adaptive approach ensures that treatment remains aligned with the evolving needs of the patient.

6.5.3.3 Outcome Evaluation

The final phase involves evaluating the efficacy of the treatment over time. Quantitative biomarkers derived from the causal analysis can be used to assess the effectiveness of the interventions. Additionally, these biomarkers can be incorporated into predictive models that estimate future disease trajectories, helping clinicians plan ahead and adjust treatment strategies proactively to optimize long-term outcomes.

By utilizing this comprehensive approach, clinicians can make more informed decisions, improving the management of dystonia through a combination of early detection, personalized treatments, and continuous monitoring.

6.6 Future Work

This thesis has developed novel computational frameworks for causal inference in neural-muscular systems, progressing from linear to nonlinear interpretable models. While these advances contribute significantly to understanding movement disorders, several promising research directions remain unexplored. This section outlines key avenues for future investigation that build upon the methodological foundations established in this work.

- **Real-Time Causal Analysis for Clinical Applications**

A critical next step in this research trajectory is the adaptation of these computational frameworks for real-time clinical use. The current implementation primarily operates on pre-recorded data sets, requiring substantial computational resources and time for processing. This limitation restricts the immediate clinical utility of these causal inference techniques during interventional procedures or therapeutic sessions. Future work should focus on algorithmic optimizations that maintain analytical integrity while significantly reducing computational demands.

The development of streamlined processing pipelines could enable intraoperative

monitoring during neuromodulation procedures, where real-time causal information could guide electrode placement or stimulation parameter adjustments. Similarly, rehabilitation sessions could benefit from immediate feedback on neural-muscular causal dynamics, allowing therapists to modify interventions based on patient-specific responses. This would represent a paradigm shift from the current model of retrospective analysis to proactive, adaptive clinical decision-making informed by causal insights.

The technical challenges associated with this transition are substantial. They include the development of feature extraction methods that operate efficiently on streaming data, implementation of incremental learning algorithms that update causal models as new data becomes available, and creation of intuitive visualization interfaces that convey complex causal information to clinicians without overwhelming them during time-sensitive procedures. Addressing these challenges will require close collaboration between computational scientists and clinical practitioners to ensure that technical innovations align with practical clinical workflows.

- **Uncertainty Quantification in Causal Inference**

One limitation of current causal inference approaches in neurophysiological systems is the limited attention given to uncertainty quantification. Biological systems inherently contain stochasticity, and measurements invariably include noise. Future work should develop rigorous frameworks for quantifying uncertainty in causal relationships, allowing clinicians to make informed decisions with appropriate confidence.

Bayesian extensions of both linear and nonlinear causal models offer one promising direction. These approaches would not only identify potential causal relationships but also provide probabilistic assessments of their strength and reliability. This would be particularly valuable in clinical settings, where treatment decisions often require balancing multiple factors and understanding

the confidence associated with various causal hypotheses could inform risk-benefit analyses.

Beyond Bayesian approaches, ensemble methods that combine multiple causal models could provide complementary insights into uncertainty. Different models might agree on certain causal relationships while disagreeing on others, providing a natural measure of confidence based on consensus across methodological approaches. Developing formal frameworks for integrating these diverse perspectives would enhance the robustness of causal inferences in complex neural systems.

- **Closed-Loop Validation Systems**

A fundamental challenge in causal inference for neurophysiological systems is validation, as true causal relationships are rarely known with certainty. Future work should develop closed-loop experimental systems specifically designed to validate causal models under controlled conditions. These systems would allow researchers to introduce controlled perturbations to neural circuits while measuring responses across multiple modalities, creating ground-truth data sets for evaluating causal inference methods.

Such validation frameworks might include in vitro neural cultures with optical stimulation and multielectrode recording, animal models with optogenetic interventions, or non-invasive stimulation protocols in human subjects. By comparing model predictions against observed outcomes following controlled interventions, researchers could rigorously assess the accuracy and limitations of different causal inference approaches.

These validation systems would also provide valuable insights into the conditions under which causal inference methods succeed or fail. Understanding these boundaries would guide the appropriate application of these techniques in clinical settings and inform future methodological refinements. Furthermore, validated models would increase clinician confidence in adopting these

approaches for treatment planning.

- **Personalized Intervention Optimization**

Perhaps the most promising direction for future research is leveraging causal models for personalized intervention planning. The framework developed in this thesis provides a foundation for understanding individual differences in neural-muscular dynamics, which could inform tailored therapeutic approaches. Future work should explore how these individual causal profiles might guide optimization of interventions ranging from medication selection to neuromodulation parameters.

This research direction would involve developing computational methods for simulating intervention outcomes based on patient-specific causal models. These simulations could predict how perturbations to different components of the neural-muscular system might propagate through causal networks, allowing clinicians to anticipate treatment responses before implementation. Such an approach would represent a significant advance over current trial-and-error methods, potentially reducing adverse effects while improving therapeutic outcomes.

The complexity of this endeavor should not be underestimated. It requires not only accurate causal models but also detailed pharmacological or biophysical models of interventions and their neural effects. Integrating these diverse modeling approaches presents substantial computational and theoretical challenges. However, the potential clinical impact, moving from standardized protocols to mechanistically informed precision interventions, justifies this ambitious research agenda.

- **Integration with Emerging Sensing Technologies**

Advances in wearable and implantable sensing technologies offer unprecedented opportunities to monitor neural and muscular activity in naturalistic settings. Future work should investigate how data from these emerging technologies

can be integrated into causal models, extending analysis beyond controlled laboratory environments to real-world conditions where movement disorders most impact patients' lives.

This integration presents both opportunities and challenges. On one hand, continuous monitoring could provide much richer data sets, capturing rare events and environmental interactions that might reveal new causal mechanisms. On the other hand, data collected outside controlled settings typically contains more noise and confounding factors, requiring robust analytical methods. Developing techniques that handle sparse, heterogeneous data from multiple sources while maintaining causal interpretability would significantly advance the field.

Longitudinal monitoring through these technologies could also enable the study of how causal relationships evolve over time during disease progression, in response to treatment, or through natural aging. Such temporal dynamics remain poorly understood in many movement disorders but could provide crucial insights into when and how to intervene for optimal outcomes.

The interdisciplinary nature of this future work cannot be overstated. Progress will require collaboration among computational neuroscientists developing mathematical frameworks, engineers creating sensing technologies, and clinicians interpreting results in patient contexts. By pursuing these research directions, we can build upon the foundation established in this thesis to develop more powerful tools for understanding and treating movement disorders, ultimately improving patient outcomes through mechanistically-informed, personalized interventions.

6.6.1 Extensions to Other Neurological Conditions

While this thesis primarily focused on dystonia, the methodologies developed here have the potential for broader application across a wide range of neurological con-

ditions. Future work should explore transfer learning approaches that leverage knowledge gained from well-studied conditions to accelerate understanding of rarer or more complex disorders. This approach recognizes that while movement disorders may present differently, they often share underlying neural mechanisms that can be captured through similar causal modeling approaches.

Transfer learning in this context would involve identifying invariant causal structures across disorders while allowing for disorder-specific parameters or connections. This would require careful consideration of how to represent prior knowledge from one condition in a form that informs but does not overly constrain analyses of another. Techniques such as meta-learning, which aims to extract general principles that apply across tasks, could prove particularly valuable in this context.

The clinical impact of successful transfer learning would be substantial. Rare movement disorders often suffer from limited research due to small patient populations, making it difficult to collect sufficient data for comprehensive analysis. By transferring insights from more common conditions, researchers could develop more robust models even with limited data, accelerating the development of therapeutic approaches for underserved patient populations. Some of the key areas for future research include Parkinson’s disease, essential tremor, cerebellar ataxia, and multiple sclerosis.

Parkinson’s disease presents an interesting opportunity for applying causal inference methods. By examining the causal relationships between the basal ganglia and cortical motor regions, researchers could gain a deeper understanding of the mechanisms behind motor fluctuations and tremors. This could lead to more targeted therapeutic interventions that address the root causes of these motor symptoms rather than simply alleviating their surface manifestations.

In the case of essential tremor, non-linear causal inference methods could be applied to explore the complex interactions between central oscillatory activity and peripheral muscle control. Understanding how these systems influence one another in the context of tremor could help develop more precise, individualized treatments, potentially improving outcomes for patients who currently do not respond well to

conventional therapies.

Cerebellar ataxia, a condition characterized by coordination deficits and balance issues, could benefit from a similar approach. By investigating the directional influences within cerebellar-motor circuits, researchers could better understand the underlying causes of these coordination difficulties. Insights from such studies would not only improve the understanding of cerebellar function but also inform the development of tailored rehabilitation strategies aimed at restoring motor control and balance.

Lastly, Multiple Sclerosis (MS) represents another promising area for research. The disease's progressive nature, marked by demyelination and subsequent changes in neural circuitry, could be explored using causal inference methods to understand how these alterations impact the brain's ability to communicate effectively. By analyzing how compensatory mechanisms evolve over time in response to demyelination, clinicians could improve strategies for tracking disease progression and evaluating the efficacy of interventions, particularly those aimed at slowing or reversing the effects of demyelination.

6.7 Conclusion

This thesis has advanced our understanding of causal interactions within neurophysiological systems, establishing a robust foundation for both theoretical insights and clinical innovations in movement disorders. By integrating advanced computational frameworks with advanced causal inference techniques, the work effectively addresses the complex interplay between structural connectivity and dynamic neural processes. These methodological advancements pave the way for improved diagnostic precision and more targeted therapeutic interventions.

Looking ahead, unifying diverse causal inference methods into a cohesive framework remains a key objective. Combining linear and non-linear approaches can overcome challenges such as high-dimensionality and intricate feedback loops, ulti-

mately leading to more reliable models. Establishing clear identifiability conditions within these networks will further enhance the credibility and clinical utility of causal inferences drawn from complex physiological data.

Equally critical is the development of transparent and explainable AI systems that empower clinicians to trust and act upon model-driven insights. By advancing counterfactual analysis and other interpretability techniques, future research can demystify the causal relationships underlying neural function, supporting personalized treatment strategies and ensuring that AI-driven recommendations are both comprehensible and actionable.

Translating these innovations into practical clinical tools poses a significant challenge that demands real-time data processing, adaptive algorithm design, and seamless integration with existing clinical infrastructures such as electronic health records and decision support systems. Future work must focus on creating user-friendly interfaces that bridge the gap between complex computational methods and everyday clinical practice.

Expanding these methodologies to encompass a broader spectrum of neurological conditions beyond dystonia, such as Parkinson's disease, Essential Tremor, and Multiple Sclerosis, will not only validate the approaches but also enhance their generalizability. This cross-condition exploration is essential for developing personalized treatment protocols that cater to the unique neurophysiological profiles of individual patients.

Ultimately, the future of causal analysis in neurophysiology lies at the intersection of computational neuroscience, clinical research, and biomedical engineering. Interdisciplinary collaboration will be vital in refining these analytical tools and transforming them into clinical innovations that drive early diagnosis, precise interventions, and improved long-term patient outcomes. Embracing these challenges promises to usher in a new era of personalized medicine, where the integration of theory, technology, and clinical practice leads to transformative advances in patient care and quality of life.

Bibliography

- [1] Farwa Abbas et al. “SS-ADMM: Stationary and Sparse Granger Causal Discovery for Cortico-Muscular Coupling”. In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023, pp. 1–5.
- [2] Saima Absar, Yongkai Wu, and Lu Zhang. “Neural time-invariant causal discovery from time series data”. In: *2023 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2023, pp. 1–8.
- [3] Alberto Albanese, M Di Giovanni, and S Lalli. “Dystonia: diagnosis and management”. In: *European journal of neurology* 26.1 (2019), pp. 5–17.
- [4] Stefano V Albrecht and Subramanian Ramamoorthy. “Exploiting causality for selective belief filtering in dynamic Bayesian networks”. In: *Journal of Artificial Intelligence Research* 55 (2016), pp. 1135–1178.
- [5] Yuri Antonacci et al. “Estimation of Granger causality through Artificial Neural Networks: applications to physiological systems and chaotic electronic oscillators”. In: *PeerJ Computer Science* 7 (2021), e429.
- [6] Yuri Antonacci et al. “Information transfer in linear multivariate processes assessed through penalized regression techniques: validation and application to physiological networks”. In: *Entropy* 22.7 (2020), p. 732.
- [7] Anestis Antoniadis et al. “Wavelet thresholding for some classes of non-Gaussian noise”. In: *Statistica neerlandica* 56.4 (2002), pp. 434–453.

- [8] Roland Averkamp and Christian Houdré. “Wavelet thresholding for non-necessarily Gaussian noise: idealism”. In: *The Annals of Statistics* 31.1 (2003), pp. 110–151.
- [9] Shi-Chun Bao, Wing-Cheong Leung, Vincent C K. Cheung, Ping Zhou, and Kai-Yu Tong. “Pathway specific modulatory effects of neuromuscular electrical stimulation during pedaling in chronic stroke survivors”. In: *Journal of neuroengineering and rehabilitation* 16 (2019), pp. 1–15.
- [10] Falco J Bargagli-Stoffi, Riccardo Cadei, Kwonsang Lee, and Francesca Dominici. “Causal rule ensemble: Interpretable discovery and inference of heterogeneous treatment effects”. In: *arXiv preprint arXiv:2009.09036* (2020).
- [11] Ernesto H Bedoy, Efrain A Guirola Diaz, Ashley N Dalrymple, Isaiah Levy, Thomas Hyatt, Darcy M Griffin, George F Wittenberg, and Douglas J Weber. “Improving localization and measurements of M-waves using high-density surface electromyography”. In: *Journal of Neurophysiology* (2024).
- [12] Paolo Belardinelli, Leonard Laer, Erick Ortiz, Christoph Braun, and Alireza Gharabaghi. “Plasticity of premotor cortico-muscular coherence in severely impaired stroke patients with hand paralysis”. In: *NeuroImage: Clinical* 14 (2017), pp. 726–733.
- [13] Félix Bermejo-Pareja. “Essential tremor—a neurodegenerative disorder associated with cognitive defects?” In: *Nature Reviews Neurology* 7.5 (2011), pp. 273–282.
- [14] Léon Bottou, Jonas Peters, Joaquin Quiñonero-Candela, Denis X Charles, D Max Chickering, Elon Portugaly, Dipankar Ray, Patrice Simard, and Ed Snelson. “Counterfactual reasoning and learning systems: The example of computational advertising”. In: *The Journal of Machine Learning Research* 14.1 (2013), pp. 3207–3260.
- [15] George EP Box et al. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

- [16] Alessandro Bregoli, Marco Scutari, and Fabio Stella. “A constraint-based algorithm for the structural learning of continuous-time Bayesian networks”. In: *International Journal of Approximate Reasoning* 138 (2021), pp. 105–122.
- [17] Marco Capogrosso, Josep-Maria Balaguer, Genis Prat-Ortega, Nikhil Verma, Prakarsh Yadav, Erynn Sorensen, Roberto de Freitas, Scott Ensel, Luigi Borda, Serena Donadio, et al. “Supraspinal control of motoneurons after paralysis enabled by spinal cord stimulation”. In: *Research Square* (2024).
- [18] Ömer Çayır and Çağatay Candan. “Maximum likelihood autoregressive model parameter estimation with noise corrupted independent snapshots”. In: *Signal Processing* 186 (2021), p. 108118.
- [19] Xiaoling Chen, Jinyuan Zhang, Yingying Hao, Ying Liu, Yinan Yang, and Ping Xie. “Change in EEG-EMG synchronization reflecting abnormal functional corticomuscular coupling following stroke: A pilot study”. In: *Medicine in Novel Technology and Devices* 24 (2024), p. 100335.
- [20] Xiaoling Chen, Yuanyuan Zhang, Shengcui Cheng, and Ping Xie. “Transfer spectral entropy and application to functional corticomuscular coupling”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.5 (2019), pp. 1092–1102.
- [21] William Jay Conover. *Practical nonparametric statistics*. Vol. 350. John Wiley & Sons, 1999.
- [22] Alvaro Henrique Chaim Correia, James Cussens, and Cassio de Campos. “On pruning for score-based Bayesian network structure learning”. In: *International conference on artificial intelligence and statistics*. PMLR, 2020, pp. 2709–2718.
- [23] Nicandro Cruz-Ramírez, Héctor-Gabriel Acosta-Mesa, Rocío-Erandi Barrientos-Martínez, and Luis-Alonso Nava-Fernández. “How good are the Bayesian information criterion and the minimum description length principle for model selection? A Bayesian network analysis”. In: *Mexican International Conference on Artificial Intelligence*. Springer, 2006, pp. 494–504.

- [24] Ben Deaner. “Proxy controls and panel data”. In: *arXiv preprint arXiv:1810.00283* (2018).
- [25] Laura-Maria Dogariu, Jacob Benesty, Constantin Paleologu, and Silviu Ciochină. “An insightful overview of the Wiener filter for system identification”. In: *Applied Sciences* 11.17 (2021), p. 7774.
- [26] MD Donsker and SRS Varadhan. “Large deviations for Markov processes and the asymptotic evaluation of certain Markov process expectations for large times”. In: *Probabilistic Methods in Differential Equations: Proceedings of the Conference Held at the University of Victoria, August 19–20, 1974*. Springer. 2006, pp. 82–88.
- [27] Shengjia Du et al. “Dictionary Learning Strategies for Cortico-Muscular Coherence Detection and Estimation”. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2021, pp. 240–244.
- [28] Andrea Duggento, Maria Guerrisi, and Nicola Toschi. “Echo state network models for nonlinear granger causality”. In: *Philosophical Transactions of the Royal Society A* 379.2212 (2021), p. 20200256.
- [29] Luca Faes et al. “On the interpretability and computational reliability of frequency-domain Granger causality”. In: *F1000Research* 6 (2017).
- [30] Joan Falcó-Roget, Adrian I Onicas, Felix Akwasi-Sarpong, and Alessandro Crimi. “Directed networks and resting-state effective brain connectivity with state-space reconstruction using reservoir computing causality”. In: *bioRxiv* (2023), pp. 2023–06.
- [31] Chenchen Fan, Yixin Wang, Yahong Zhang, and Wenli Ouyang. “Interpretable multi-scale neural network for granger causality discovery”. In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023, pp. 1–5.

- [32] Esther Florin, Johannes Pfeifer, Veerle Visser-Vandewalle, Alfons Schnitzler, and Lars Timmermann. “Parkinson subtype-specific Granger-causal coupling and coherence frequency in the subthalamic area”. In: *Neuroscience* 332 (2016), pp. 170–180.
- [33] Wojciech Fulmyk et al. “Nonlinear Granger Causality using Kernel Ridge Regression”. In: *arXiv preprint arXiv:2309.05107* (2023).
- [34] Rainer Goebel et al. “Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping”. In: *Magnetic resonance imaging* 21.10 (2003), pp. 1251–1261.
- [35] Alireza Goudarzi, Matthew R Lakin, and Darko Stefanovic. “DNA reservoir computing: a novel molecular computing approach”. In: *International Workshop on DNA-Based Computers*. Springer. 2013, pp. 76–89.
- [36] CWJ Granger. “Investigating causal relations by econometric models and cross-spectral methods”. In: *Essays in econometrics: collected papers of Clive WJ Granger*. 2001, pp. 31–47.
- [37] Nianchun Guo, Jing Wang, and Xin YU. “A review of early identification and conversion studies of event-related potentials in individuals with high risk of schizophrenia”. In: *Chinese Mental Health Journal* (2024), pp. 200–205.
- [38] Zhenghao Guo et al. “Multiscale wavelet transfer entropy with application to corticomuscular coupling analysis”. In: *IEEE Transactions on Biomedical Engineering* 69.2 (2021), pp. 771–782.
- [39] Zhenghao Guo et al. “Subband independent component analysis for coherence enhancement”. In: *IEEE Transactions on Biomedical Engineering* (2024).
- [40] Zhenghao Guo et al. “Unravelling Causal Relationships Between Cortex and Muscle with Errors-In-Variables Models”. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2021, pp. 967–970.

- [41] Dan Halbersberg and Boaz Lerner. “Local to global learning of a latent dynamic Bayesian network”. In: *ECAI 2020*. IOS Press, 2020, pp. 2600–2607.
- [42] Dylan Harries and Terence John O’Kane. “Dynamic Bayesian networks for evaluation of Granger causal relationships in climate reanalyses”. In: *Authorea Preprints* (2022).
- [43] Jason Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. “Counterfactual prediction with deep instrumental variables networks”. In: *arXiv preprint arXiv:1612.09596* (2016).
- [44] Shenghong He, Timothy O West, Fernando R Plazas, Laura Wehmeyer, Alek Pogosyan, Alceste Deli, Christoph Wiest, Damian M Herz, Thomas Simpson, Pablo Andrade, et al. “Cortico-thalamic tremor circuits and their associations with deep brain stimulation effects in essential tremor”. In: *Brain* (2024), awae387.
- [45] Paul W Holland. “Statistics and Causal Inference”. In: *Journal of the American Statistical Association* 81.396 (1986), pp. 945–960.
- [46] Cheng Hsiao. “Autoregressive modeling and causal ordering of economic variables”. In: *Journal of Economic Dynamics and Control* 4 (1982), pp. 243–259.
- [47] Guiting Hu, Wenjuan Yang, Xiaoling Chen, Wenjing Qi, Xinxin Li, Yihao Du, and Ping Xie. “Estimation of time-varying coherence amongst synergistic muscles during wrist movements”. In: *Frontiers in Neuroscience* 12 (2018), p. 537.
- [48] Scott Huberty, James Desjardins, Tyler Collins, Mayada Elsabbagh, and Christian O’Reilly. “PyLossless: A non-destructive EEG processing pipeline”. In: *bioRxiv* (2024), pp. 2024–01.
- [49] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeswar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and R Devon Hjelm. “MINE: mutual information neural estimation”. In: *arXiv e-prints* (2018), arXiv–1801.

- [50] Laurent Jacob et al. “Group lasso with overlap and graph lasso”. In: *Proceedings of the 26th annual international conference on machine learning*. 2009, pp. 433–440.
- [51] Seung-Hyun Jin et al. “Linear and nonlinear information flow based on time-delayed mutual information method and its application to corticomuscular interaction”. In: *Clinical Neurophysiology* 121.3 (2010), pp. 392–401.
- [52] Marcin Jan Kaminski and Katarzyna J Blinowska. “A new method of the description of the information flow in the brain structures”. In: *Biological cybernetics* 65.3 (1991), pp. 203–210.
- [53] Weijie Ke and Zhizeng Luo. “Analysis of cortico-muscular coupling and functional brain network under different standing balance paradigms”. In: *Brain Sciences* 14.1 (2024), p. 81.
- [54] Kangil Kim. “Time-sensitive adaptation of regularization strength of recurrent neural networks for accurate learning”. In: *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE. 2017, pp. 194–198.
- [55] Ondřej Klempř, Radim Krupička, Václav Čejka, and Robert Jech. “Microelectrode neuronal activity biomarker of the internal globus pallidus in dystonia correlates with long-term neuromodulation effects”. In: *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE. 2018, pp. 1–6.
- [56] Anna Korzeniewska, Malgorzata Manczak, Maciej Kaminski, Katarzyna J Blinowska, and Stefan Kasicki. “Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method”. In: *Journal of neuroscience methods* 125.1-2 (2003), pp. 195–207.
- [57] SAJEA Lagerweij, M Smit, LM Centen, JMC van Dijk, ME van Egmond, JW Elting, and MAJ Tijssen. “Connecting the dots—A systematic review on coherence analysis in dystonia”. In: *Neurobiology of Disease* (2024), p. 106616.

- [58] Jason D Lee, Yuekai Sun, and Michael A Saunders. “Proximal Newton-type methods for minimizing composite functions”. In: *SIAM Journal on Optimization* 24.3 (2014), pp. 1420–1443.
- [59] Stéphane Lehéricy, Marina AJ Tijssen, Marie Vidailhet, Ryuji Kaji, and Sabine Meunier. “The anatomical basis of dystonia: current view using neuroimaging”. In: *Movement Disorders* 28.7 (2013), pp. 944–957.
- [60] Honghao Li, Vincent Cabeli, Nadir Sella, and Hervé Isambert. “Constraint-based causal structure learning with consistent separating sets”. In: *Advances in neural information processing systems* 32 (2019).
- [61] Hongjian Li, Kam-Heung Sze, Gang Lu, and Pedro J Ballester. “Machine-learning scoring functions for structure-based drug lead optimization”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 10.5 (2020), e1465.
- [62] Long Li, Jing Guo, Yanlong Zhang, Hongjian Wu, Lina Li, Tian Liu, and Jue Wang. “Pattern reorganization of corticomuscular connection with the tactile stimulation”. In: *Annals of Biomedical Engineering* 48 (2020), pp. 834–847.
- [63] Tie Liang et al. “Time-frequency maximal information coefficient method and its application to functional corticomuscular coupling”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28.11 (2020), pp. 2515–2524.
- [64] Yin Liu and Sam Davanloo Tajbakhsh. “Fitting arma time series models without identification: A proximal approach”. In: *arXiv preprint arXiv:2002.06777* (2020).
- [65] C Lo Bianco, JL Ridet, BL Schneider, N Deglon, and P Aebischer. “ α -Synucleinopathy and selective dopaminergic neuron loss in a rat lentiviral-based model of Parkinson’s disease”. In: *Proceedings of the National Academy of Sciences* 99.16 (2002), pp. 10813–10818.

- [66] Edward N Lorenz. “Predictability: A problem partly solved”. In: *Proc. Seminar on predictability*. Vol. 1. 1. Reading. 1996.
- [67] Mantas Lukoševičius. “A practical guide to applying echo state networks”. In: *Neural Networks: Tricks of the Trade: Second Edition*. Springer, 2012, pp. 659–686.
- [68] Scott Lundberg. “A unified approach to interpreting model predictions”. In: *arXiv preprint arXiv:1705.07874* (2017).
- [69] Gianvittorio Luria, Alessandro Viani, Annalisa Pascarella, Harald Bornfleth, Sara Sommariva, and Alberto Sorrentino. “The SESAMEEG package: a probabilistic tool for source localization and uncertainty quantification in M/EEG”. In: *Frontiers in Human Neuroscience* 18 (2024), p. 1359753.
- [70] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2007.
- [71] Sepehr Maleki, Sasan Maleki, and Nicholas R Jennings. “Unsupervised anomaly detection with LSTM autoencoders using statistical data-filtering”. In: *Applied Soft Computing* 108 (2021), p. 107443.
- [72] Ričards Marcinkevičs and Julia E Vogt. “Interpretable models for granger causality using self-explaining neural networks”. In: *arXiv preprint arXiv:2101.07600* (2021).
- [73] Daniele Marinazzo, Wei Liao, Huaifu Chen, and Sebastiano Stramaglia. “Non-linear connectivity by Granger causality”. In: *Neuroimage* 58.2 (2011), pp. 330–338.
- [74] Daniele Marinazzo, Mario Pellicoro, and Sebastiano Stramaglia. “Kernel method for nonlinear Granger causality”. In: *Physical review letters* 100.14 (2008), p. 144103.

- [75] Maya B Mathur and Tyler J VanderWeele. “Methods to address confounding and other biases in meta-analyses: review and recommendations”. In: *Annual review of public health* 43.1 (2022), pp. 19–35.
- [76] Verity M McClelland and Jean-Pierre Lin. “Dystonia in Childhood: How Insights from Paediatric Research Enrich the Network Theory of Dystonia”. In: *Basic and Translational Applications of the Network Theory for Dystonia*. Springer, 2023, pp. 1–22.
- [77] Verity M McClelland et al. “Abnormal patterns of corticomuscular and intermuscular coherence in childhood dystonia”. In: *Clinical Neurophysiology* 131.4 (2020), pp. 967–977.
- [78] Verity M McClelland et al. “Modulation of corticomuscular coherence by peripheral stimuli”. In: *Experimental brain research* 219 (2012), pp. 275–292.
- [79] Jiri Mekyska, Zdenek Smekal, Zoltan Galaz, Zdenek Mzourek, Irena Rektorova, Marcos Faundez-Zanuy, and Karnele López-de Ipiña. “Perceptual features as markers of Parkinson’s Disease: the issue of clinical interpretability”. In: *Recent advances in nonlinear speech processing* (2016), pp. 83–91.
- [80] Wolf Muellbacher, Ulf Ziemann, Babak Boroojerdi, and Mark Hallett. “Effects of low-frequency transcranial magnetic stimulation on motor excitability and basic motor behavior”. In: *Clinical neurophysiology* 111.6 (2000), pp. 1002–1007.
- [81] Remya R Nair, G Venugopal, and Ramakrishnan Swaminathan. “Analysis of Muscle Fiber Type Proportions in Surface Electromyography Signals of Athletes Using Reassigned Morlet Scalogram”. In: *IEEE Transactions on Instrumentation and Measurement* (2024).
- [82] Hariharan Nalatore et al. “Mitigating the effects of measurement noise on Granger causality”. In: *Physical Review E* 75.3 (2007), p. 031123.

- [83] Preetam Nandy, Alain Hauser, and Marloes H Maathuis. “High-dimensional consistency in score-based and hybrid structure learning”. In: *The Annals of Statistics* 46.6A (2018), pp. 3151–3183.
- [84] Nicoletta Nicolaou and Timothy G Constandinou. “A nonlinear causality estimator based on non-parametric multiplicative regression”. In: *Frontiers in neuroinformatics* 10 (2016), p. 19.
- [85] Sebin Park et al. “Measurement noise recommendation for efficient Kalman filtering over a large amount of sensor data”. In: *Sensors* 19.5 (2019), p. 1168.
- [86] Magdalini Paschali, Qingyu Zhao, Ehsan Adeli, and Kilian M Pohl. “Bridging the Gap Between Deep Learning and Hypothesis-Driven Analysis via Permutation Testing”. In: *International Workshop on PRedictive Intelligence In MEDicine*. Springer. 2022, pp. 13–23.
- [87] Abhidnya Patharkar, Fulin Cai, Firas Al-Hindawi, and Teresa Wu. “Predictive modeling of biomedical temporal data in healthcare applications: review and future directions”. In: *Frontiers in Physiology* 15 (2024), p. 1386760.
- [88] Robert Peach, Maximilian Friedrich, Lara Fronemann, Muthuraman Muthuraman, Sebastian R Schreglmann, Daniel Zeller, Christoph Schrader, Joachim K Krauss, Alfons Schnitzler, Matthias Wittstock, et al. “Head movement dynamics in dystonia: a multi-centre retrospective study using visual perceptive deep learning”. In: *npj Digital Medicine* 7.1 (2024), p. 160.
- [89] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.
- [90] Judea Pearl. *The Seven Tools of Causal Inference with Reflections on Machine Learning*. Vol. 62. 3. 2019, pp. 54–60.
- [91] Judea Pearl and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, 2018.

- [92] Jiazheng Peng, Talifu Zikereya, Zhongshu Shao, and Kaixuan Shi. “The neuromechanical of Beta-band corticomuscular coupling within the human motor system”. In: *Frontiers in Neuroscience* 18 (2024), p. 1441002.
- [93] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.
- [94] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, 2017.
- [95] John Winsor Pratt and Jean Dickinson Gibbons. *Concepts of nonparametric theory*. Springer Science & Business Media, 2012.
- [96] Angelo Quartarone, Vincenzo Rizzo, Carmen Terranova, Francesca Morgante, Sybille Schneider, Nasrien Ibrahim, Paolo Girlanda, Kailash P Bhatia, and John C Rothwell. “Abnormal sensorimotor plasticity in organic but not in psychogenic dystonia”. In: *Brain* 132.10 (2009), pp. 2871–2877.
- [97] Alberto Raggi, Alessandro Serretti, and Raffaele Ferri. “The P300 component of the auditory event-related potential in adult psychiatric and neurologic disorders: a narrative review of clinical and experimental evidence”. In: *International Clinical Psychopharmacology* (2024), pp. 10–1097.
- [98] Joshua W Robinson, Alexander J Hartemink, and Zoubin Ghahramani. “Learning Non-Stationary Dynamic Bayesian Networks.” In: *Journal of Machine Learning Research* 11.12 (2010).
- [99] Donald B Rubin. “Causal inference using potential outcomes: Design, modeling, decisions”. In: *Journal of the American Statistical Association* 100.469 (2005), pp. 322–331.
- [100] Donald B Rubin. “Randomization analysis of experimental data: The Fisher randomization test comment”. In: *Journal of the American statistical association* 75.371 (1980), pp. 591–593.

- [101] Kayvan Sadeghi and Terry Soo. “Conditions and assumptions for constraint-based causal structure learning”. In: *Journal of Machine Learning Research* 23.109 (2022), pp. 1–34.
- [102] Benjamin Schrauwen, David Verstraeten, and Jan Van Campenhout. “An overview of reservoir computing: theory, applications and implementations”. In: *Proceedings of the 15th european symposium on artificial neural networks. p. 471-482 2007*. 2007, pp. 471–482.
- [103] Johannes Schumacher, Hazem Toutounji, and Gordon Pipa. “An introduction to delay-coupled reservoir computing”. In: *Artificial Neural Networks: Methods and Applications in Bio-/Neuroinformatics*. Springer. 2015, pp. 63–90.
- [104] Gideon Schwarz. “Estimating the dimension of a model”. In: *The annals of statistics* (1978), pp. 461–464.
- [105] Marco Scutari. “Dirichlet Bayesian network scores and the maximum relative entropy principle”. In: *Behaviormetrika* 45 (2018), pp. 337–362.
- [106] Syed Ahmar Shah, Peter Brown, Hortensia Gimeno, Jean-Pierre Lin, and Verity M McClelland. “Application of machine learning using decision trees for prognosis of deep brain stimulation of globus pallidus internus for children with dystonia”. In: *Frontiers in neurology* 11 (2020), p. 825.
- [107] Uri Shalit, Fredrik D Johansson, and David Sontag. “Estimating individual treatment effect: generalization bounds and algorithms”. In: *International Conference on Machine Learning*. 2017, pp. 3076–3085.
- [108] BH Shekar and Guesh Dagnew. “Grid search-based hyperparameter tuning and classification of microarray cancer data”. In: *2019 second international conference on advanced computational and communication paradigms (ICACCP)*. IEEE. 2019, pp. 1–8.
- [109] Galit Shmueli. “To explain or to predict?” In: (2010).

- [110] Ali Shojaie and Emily B Fox. “Granger causality: A review and recent advances”. In: *Annual Review of Statistics and Its Application* 9.1 (2022), pp. 289–319.
- [111] Ben Somers et al. “A generic EEG artifact removal algorithm based on the multi-channel Wiener filter”. In: *Journal of neural engineering* 15.3 (2018), p. 036007.
- [112] Patrick A Stokes and Patrick L Purdon. “A study of problems encountered in Granger causality analysis from a neuroscience perspective”. In: *Proceedings of the national academy of sciences* 114.34 (2017), E7063–E7072.
- [113] Polina Suter, Jack Kuipers, Giusi Moffa, and Niko Beerenwinkel. “Bayesian structure learning and sampling of Bayesian networks with the R package BiDAG”. In: *arXiv preprint arXiv:2105.00488* (2021).
- [114] Jiyu Tan, Yurong Li, Qiurong Xie, and Xiaoling Wang. “Modeling and causality analysis of human sensorimotor control system based on NVAR method”. In: *International Journal of Network Dynamics and Intelligence* (2023), pp. 100014–100014.
- [115] Alex Tank, Ian Covert, Nicholas Foti, Ali Shojaie, and Emily B Fox. “Neural granger causality”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.8 (2021), pp. 4267–4279.
- [116] Chaolin Teng, Yanyan Zhang, and Gang Wang. “The removal of EMG artifact from EEG signals by the multivariate empirical mode decomposition”. In: *2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. IEEE. 2014, pp. 873–876.
- [117] Ghada Trabelsi, Philippe Leray, Mounir Ben Ayed, and Adel M Alimi. “Benchmarking dynamic Bayesian network structure learning algorithms”. In: *2013 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO)*. IEEE. 2013, pp. 1–6.

- [118] Nelson J Trujillo-Barreto, David Araya, Aland Astudillo, and Wael El-Deredy. “Explicit modeling of brain state duration using hidden semi Markov models in EEG data”. In: *IEEE Access* (2024).
- [119] Davide Valeriani and Kristina Simonyan. “A microstructural neural network biomarker for dystonia diagnosis identified by a DystoniaNet deep learning platform”. In: *Proceedings of the National Academy of Sciences* 117.42 (2020), pp. 26398–26405.
- [120] David Verstraeten, Benjamin Schrauwen, Michiel d’Haene, and Dirk Stroobandt. “An experimental unification of reservoir computing methods”. In: *Neural networks* 20.3 (2007), pp. 391–403.
- [121] Livia Vignozzi, Francesca Macchi, Elena Montagni, Maria Pasquini, Alessandra Martello, Antea Minetti, Elea Coulomb, Anna Letizia Allegra Mascaro, Silvestro Micera, Matteo Caleo, et al. “Combining Gamma Neuromodulation and Robotic Rehabilitation Restores Parvalbimin-mediated Gamma Function and Boosts Motor Recovery in Stroke Mice”. In: *bioRxiv* (2024), pp. 2024–08.
- [122] Elisa Visani, Ferruccio Panzica, Silvana Franceschetti, Nico Golfrè Andreasi, Roberto Cilia, Sara Rinaldo, Davide Rossi Sebastiano, Paola Lanteri, and Roberto Eleopra. “Early cortico-muscular coherence and cortical network changes in Parkinson’s patients treated with MRgFUS”. In: *Frontiers in Neurology* 15 (2024), p. 1362712.
- [123] Curtis R Vogel and Mary E Oman. “Iterative methods for total variation denoising”. In: *SIAM Journal on Scientific Computing* 17.1 (1996), pp. 227–238.
- [124] Stefan Wager and Susan Athey. “Estimation and inference of heterogeneous treatment effects using random forests”. In: *Journal of the American Statistical Association* 113.523 (2018), pp. 1228–1242.

- [125] Guozheng Wang, Yi Yang, Kangli Dong, Anke Hua, Jian Wang, and Jun Liu. “Multisensory Conflict Impairs CorticoMuscular Network Connectivity and Postural Stability, Insights from Partial Directed Coherence Analysis”. In: *Neuroscience Bulletin* 40.1 (2024), pp. 79–89.
- [126] Mingzhao Wang and Zuntao Fu. “A new method of nonlinear causality detection: Reservoir computing Granger causality”. In: *Chaos, Solitons & Fractals* 154 (2022), p. 111675.
- [127] Weiran Wang and Miguel A Carreira-Perpinán. “Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application”. In: *arXiv preprint arXiv:1309.1541* (2013).
- [128] Yalin Wang, Wei Chen, Minghui Liu, and Yan Xu. “Time-shift permutation cross-mapping: a robust data-driven causality detection framework for complex physiology system”. In: *Available at SSRN 4267027* ().
- [129] Axel Wismüller, Adora M Dsouza, M Ali Vosoughi, and Anas Abidin. “Large-scale nonlinear Granger causality for inferring directed dependence from short multivariate time-series data”. In: *Scientific reports* 11.1 (2021), p. 7817.
- [130] Bowen Xiao, Limeng Liu, Lin Chen, Xing Wang, Xin Zhang, Xiaoyu Liu, Wensheng Hou, and Xiaoying Wu. “Neuro-muscular Responses Adaptation to Dynamic Changes in Grip Strength”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2024).
- [131] Lin Xiao and Tong Zhang. “A proximal-gradient homotopy method for the sparse least-squares problem”. In: *SIAM Journal on Optimization* 23.2 (2013), pp. 1062–1091.
- [132] Fei Ye, Ziyang Sun, Donghui Yang, Huijiao Wang, and Xugang Xi. “Cortico-muscular coupling analysis based on improved LSTM and transfer entropy”. In: *Neuroscience Letters* 760 (2021), p. 136012.

- [133] Jinxu Yu, Lijie Zhang, Yihao Du, Xiaoran Wang, Jianhua Yan, Jie Chen, and Ping Xie. “Exploration and Application of a Muscle Fatigue Assessment Model Based on NMF for Multi-Muscle Synergistic Movements”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2024).
- [134] Xiaolin Yu, Chong Zhang, Longfei Su, Jianbao Zhang, and Nini Rao. “Estimation of the cortico-cortical and brain-heart functional coupling with directed transfer function and corrected conditional entropy”. In: *Biomedical Signal Processing and Control* 43 (2018), pp. 110–116.
- [135] Hadi Zayyani, Massoud Babaie-Zadeh, and Christian Jutten. “An iterative Bayesian algorithm for sparse component analysis in presence of noise”. In: *IEEE Transactions on Signal Processing* 57.11 (2009), pp. 4378–4390.
- [136] Chaozhu Zhang, Hongxing Chu, and Mingyuan Ma. “Decoding Algorithm of Motor Imagery Electroencephalogram Signal Based on CLRNet Network Model”. In: *Sensors* 23.18 (2023), p. 7694.
- [137] Ziyi Zhang, Shaogang Ren, Xiaoning Qian, and Nick Duffield. “Learning Flexible Time-windowed Granger Causality Integrating Heterogeneous Interventional Time Series Data”. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024, pp. 4408–4418.
- [138] Jintong Zhao, Zhongxue Gan, Ruixi Huang, Chun Guan, Jifan Shi, and Siyang Leng. “Detecting dynamical causality via intervened reservoir computing”. In: *Communications Physics* 7.1 (2024), p. 232.
- [139] Shuzhi Zhao, Guangyan Dai, Jingting Li, Xiaoxia Zhu, Xiyan Huang, Yongxue Li, Mingdan Tan, Lan Wang, Peng Fang, Xi Chen, et al. “An interpretable model based on graph learning for diagnosis of Parkinson’s disease with voice-related EEG”. In: *NPJ Digital Medicine* 7.1 (2024), p. 3.
- [140] Yiming Zhou and Wei Dai. “Proximal Dogleg Opportunistic Majorization for Nonconvex and Nonsmooth Optimization”. In: *arXiv preprint arXiv:2402.19176* (2024).