An Approach to Learning the Hierarchical Organization of the Frontal Lobe

Kurt Butler,[†] Duncan Cleveland,[◇] Charles B. Mikell,[*] Sima Mofakham,[†*]
Yuri B. Saalmann,[◇] and Petar M. Djurić[†]

[†]*Department of Electrical and Computer Engineering*, Stony Brook University, Stony Brook, NY 11794, USA
[*]*Department of Neurosurgery*, Stony Brook University, Stony Brook, NY 11794, USA
[◇]*Department of Psychology*, University of Wisconsin-Madison, Madison, WI, USA

***Abstract*—In neuroscience, hierarchical models of brain connectivity, particularly in the prefrontal cortex (PFC), are used to understand how the brain can process sensory information, make decisions and perform other high level tasks. Despite extensive research, understanding the structure of the PFC remains a crucial challenge. To this end, we propose a data-driven approach to studying brain signals based on Gaussian processes and causal strengths. For discovering causations, we propose a metric referred to as double-averaged differential causal effect. The differential causal effect has been proposed recently, and it can be used to quantify causal strengths in a principled way. We studied real multivariate time series data that represent local field potentials from the frontal lobe. The interest was in finding the causal relationship between the medial and lateral PFC areas of the brain. Our results suggest that the medial PFC causally influences the lateral PFC.**

***Index Terms*—brain, causal strength, Gaussian processes, medical signal processing, time series**

I. Introduction

Cognitive control is our ability to flexibly adapt behavior according to goals and context. Cognitive control is considered hierarchical [5] in the sense that when we plan and perform actions, we often start with an overarching/abstract goal like "make coffee," which can be broken down into more concrete subgoals, e.g., "grind beans" or "get a cup." The frontal lobe of the brain, particularly the prefrontal cortex (PFC), is vital for cognitive control. The PFC is comprised of 47 areas [13] differentially contributing to cognitive control, but how information flows between areas of PFC and its hierarchical causal network structure overall, is unclear. This question takes on added importance because perturbed cognitive control characterizes several neuropsychiatric disorders, including schizophrenia and depression.

Neuroscientists have studied the PFC extensively to discover how it is connected and how information flows from one part to another. However, determining functional connectivity and the direction of information flow within the large-scale organization of brain networks, and PFC in particular due to its dense recurrent bidirectional connectivity, is a critical challenge in neuroscience. At present, functional connectivity metrics have only been coarsely applied to PFC. One can divide current methods of functional connectivity based on whether they are directed (vs. non-directed) and model-based (vs. model-free) [3]. Directed models try to identify the cause and effect, while non-directed models only show statistical interdependence. Cross-correlation and Granger's causality are examples of a directed model where cause precedes the effect [10], [14]. Model-based approaches assume a relationship between two signals (Pearson correlation, for example).

In contrast, model-free approaches (such as mutual information) speak only to signal transmission between regions but can specify the direction of signal transmission (i.e., with transfer entropy [17], [23]). The PFC is astonishingly complex, with innumerable bidirectional connections with the rest of the brain, leading to difficulty in creating tractable models. Nonetheless, some large-scale organizational principles have been proposed with the assistance of functional connectivity-based models. Koechlin and colleagues proposed an anterior-to-posterior gradient wherein higher-order control is imposed by the frontal pole. Low-level sensory feedback is implemented at more posterior regions, such as the premotor cortex [16]. The authors of [16] used a linear model of blood flow patterns to support this view, which parcellated the prefrontal cortex into relatively large regions and assumed linear relationships between activity in these regions. Badre and Nee extrapolated this and other data into a general model of antero-posterior gradient organization of PFC, based on studies that largely rely on similar approaches [2]. A related proposal by Shenhav and colleagues suggests that the medial parts of the frontal lobe compute the expected value of control signals needed to optimize behavior and transmit these signals to the lateral parts of the frontal lobe for implementation [24]. This model, however, also depends on studies that only coarsely identify causal influences in the prefrontal cortex.

Our contributions in this work are that we introduce novel analyses to study the functional connectivity and dynamics of the frontal lobe in a data-driven manner. Our proposed method is based on the concept of causal strength, and it is directed, non-parametric, and principled. We apply this method to the analysis of local field potentials (LFP), which are strongly correlated to local neuronal activity, and thus are closer to the "ground truth" of brain network function [15]. However, the proposed method is also applicable to time series data in general.

II. Problem Formulation

A. Hierarchical organization of the frontal lobe

Information about the external world is transmitted from our sensory organs to the occipital, temporal and parietal lobes of the brain; and there is a well-studied hierarchical organization of sensory processing across these lobes. In contrast, the hierarchical organization of the frontal lobe is unclear. One major theory suggests that PFC, along its anterior-posterior dimension, is broadly organized according to the level of abstraction of behavioral goals. Here, anterior PFC, at the top of the hierarchy, represents abstract goals/rules, and more concrete goals/rules are represented in posterior PFC [1], [16]. Another major theory suggests that PFC, along its medial-lateral dimension, is broadly organized according to its role in mediating exploration-exploitation trade-offs which are central to our daily decision-making. Here, medial PFC enables exploration of new options, whereas lateral PFC enables exploitation of the current option [11]. These theories derive from functional MRI and brain lesion data, and higher spatiotemporal resolution of intracranial neural recordings and methods to measure causal influences are required to resolve PFC organization, in conjunction with cognitive control tasks that activate PFC.

B. Description of the data

In this investigation, our neural time series data are multi-channel, local field potentials (LFPs sampled at 2kHz) derived from intracranial EEG recordings in 10 patients evaluated for surgical treatment of epilepsy. Recording sites, based on clinical requirements, include extensive coverage of the frontal lobe. We filtered data from 4-200Hz, removed artifacts/ictal activity, and analyzed bipolar derivations of LFPs (difference between two adjacent channels, to remove shared noise across channels). An important way to implement cognitive control is to apply rules which map cues to actions according to context. During recordings, patients performed a hierarchical rule task requiring switching and application of abstract and concrete rules. The abstract rule cue specified the relevant dimension (shape or orientation) of the subsequent concrete rule cue, which specified the relevant feature to report (rectangle/oval/bowtie or N/SW/SE). Preliminary data herein focus on medial (anterior cingulate cortex, ACC) and lateral (dorsolateral PFC, DLPFC) frontal lobe during the working memory period after the abstract rule cue.

III. The methodology

A. Modeling the hierarchy

We propose to learn the structure of the frontal lobe by analyzing the strength of causation between brain regions as represented by the LFP signals. To illustrate this process, consider two time series x_t and y_t, where we suppose that y_t is 'downstream' in the network hierarchy. We use a nonlinear additive noise model to represent the interaction of x_t and y_t, i.e.

$$y_t = F(y_{t-1}, ..., y_{t-Q}, x_{t-1}, ..., x_{t-Q}) + w_t \qquad (1)$$

where w_t represents noise due to unobserved influences and other background brain activity, and F is an unknown nonlinear function. The model order parameter Q controls how much of the signals' history is used for prediction, and it is assumed to be fixed, but later we consider the behavior as the parameter is varied.

While several different time lags of the signal x_t are included in the model, not all of these features contribute equally to produce the next value of y_t. For example, if the time required to send a signal from the recording site of x_t to the recording site of y_t is τ, then the partial derivative of F with respect to x_{t-q} should be zero for $q < \tau$. Since the internal state of the brain varies from moment to moment, the propagation delay τ also varies from trial to trial [9]. As a result, the function F may change from trial to trial, and so we will usually estimate the function F on a trial-by-trial basis. The primary innovation of this work concerns how to continue our analyses despite the trial-to-trial variability of the neural data.

Additionally, we note that the model in (1) considers only the pairwise relationship of x_t and y_t. As with other causal methods such as Granger causality, pairwise analysis is susceptible to bias due to latent confounding variables that influence both x_t and y_t [6], [19]. While this situation is perhaps unavoidable in neuroscience [20], we note the influence of latent confounders are more likely to result in a false-positive connection (due to the presence of information that is useful for prediction) rather than a false-negative (which would require that the unobserved influences perfectly cancel out the influence of x_t, which may be a non-generic situation [4]). In this way, we interpret the pairwise analysis as being optimistic when we attempt to detect causal influences.[1] A general way to improve the approach is to starting with pairwise analysis, and to consider ternary and higher order relationships for pairs that passed the pairwise test. For the current work, we find that pairwise analysis is a useful starting point, and we approach our results with an awareness of the need for future work.

B. Gaussian process regression

To learn the function F in (1), we need a nonlinear regression tool. To this end, Gaussian process regression (GPR) is a principled, Bayesian and non-parametric tool that has been proven useful to model time series in several ways [8], [12], [18], [26]. Suppose we want to estimate a functional relationship of the form $f : \mathbb{R}^D \to \mathbb{R}$ such that $y_i \approx f(\mathbf{x}_i)$ for some observed data set $\mathcal{D} = \{(\mathbf{x}_i, y_i); i = 1, ..., N\}$. In GPR, we specify a Gaussian process (GP) prior for the function f,

[1] In a full neuroscientific study, such pairwise analyses should be supplemented with additional analyses that further scrutinize the detected relationships, by controlling for other observed variables and carefully outlining one's assumptions (which are technically required to make the inference problem identifiable [22]).

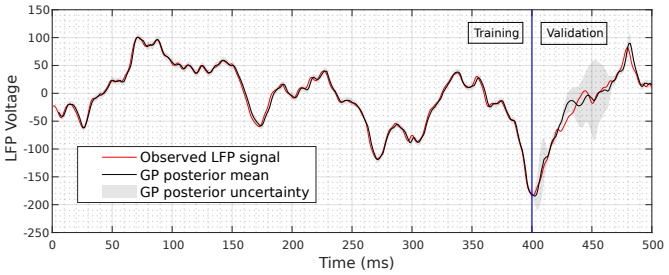

Fig. 1. Demonstration of GPR being used to model a portion of an LFP signal. For $Q = 8$, we use the GP posterior to predict the evolution of LFP signal. A GP model of the form (1) was fitted using the first 400 ms of data, and the learned model was validated on the remaining 100 ms by examining the prediction quality. We observe that the pairwise model fits confidently within sample, and can also make reasonable predictions out-of-sample.

denoted as $f \sim \mathcal{GP}(m, k)$, which means that

$$(f(\mathbf{x}_1), ..., f(\mathbf{x}_N)) \sim \mathcal{N}(\mathbf{m}, \mathbf{K}) \tag{2}$$

$$\mathbf{m}_i = \mathbb{E}(f(\mathbf{x}_i)) = m(\mathbf{x}_i) \tag{3}$$

$$\mathbf{K}_{ij} = \text{Cov}(f(\mathbf{x}_i), f(\mathbf{x}_j)) = k(\mathbf{x}_i, \mathbf{x}_j). \tag{4}$$

The functions $m(\cdot)$ and $k(\cdot, \cdot)$ are called the mean and covariance functions respectively, and are chosen when training the model. To estimate $f(\mathbf{x})$ at a new point $\mathbf{x}$, we can obtain a posterior distribution for $f(\mathbf{x})$ by conditioning on the known data. The resulting posterior distribution happens to also be a Gaussian process:

$$f(\mathbf{x})|\mathcal{D} \sim \mathcal{GP}(m_p, k_p) \tag{5}$$

$$\begin{aligned} m_p(\mathbf{x}) &= \mathbf{k}_*(\mathbf{x})^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} \\ &= \sum_{n=1}^{N} k(\mathbf{x}, \mathbf{x}_n)\, \alpha_n, \end{aligned} \tag{6}$$

$$k_p(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}_*(\mathbf{x})^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{k}_*(\mathbf{x}) \tag{7}$$

where the vectors $\mathbf{k}_*(\mathbf{x})$, $\boldsymbol{\alpha}$ and $\mathbf{y}$ are common notation used to express these formulas [21], and they are defined $(\mathbf{k}_*(\mathbf{x}))_i = k(\mathbf{x}, \mathbf{x}_i)$, $\boldsymbol{\alpha} = (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y}$ and $\mathbf{y} = (y_1, ..., y_N)$.

In Figure 1, we show an example of GPR being used to fit the model in (1) for an LFP signal.

C. Differential causal effect

We now introduce our measure of causal strength. Consider two signals x_t, y_t and a model of the form (1). Given that y_t depends on x_i for $i = t-1, ..., t-Q$, we wish to quantify how strongly each x_i influences y_t. When the function F is differentiable, the partial derivatives of F measure the sensitivity of y_t to perturbations in each input. Thus, the *differential causal effect* (DCE) of x_i on y is defined to be the partial derivative of F with respect to x_i [7], i.e.

$$\text{DCE}_{x_i \to y_t}(\mathbf{y}_{t-1}, \mathbf{x}_{t-1}) \triangleq \frac{\partial F(\mathbf{y}_{t-1}, \mathbf{x}_{t-1})}{\partial x_i}, \tag{8}$$

where $\mathbf{x}_{t-1} = (x_{t-1}, ..., x_{t-Q})$ and $\mathbf{y}_{t-1} = (y_{t-1}, ..., y_{t-Q})$. Since $\text{DCE}_{x_i \to y}$ is potentially a non-constant function of $(\mathbf{y}_t, \mathbf{x}_t)$, we may wish to summarize the DCE in a principled way. A natural approach is to average the DCE over the probability distribution of its inputs, $p(\mathbf{y}_t, \mathbf{x}_t)$. We then define the *averaged magnitude* of the DCE,

$$\langle \text{DCE}_{x_i \to y} \rangle \triangleq \iint |\text{DCE}_{x_i \to y}(\mathbf{y}_t, \mathbf{x}_t)| p(\mathbf{y}_t, \mathbf{x}_t) d\mathbf{y}_t d\mathbf{x}_t. \tag{9}$$

To estimate the DCE from data, the GPR approach also yields an elegant solution. If a function F is distributed according to a GP with mean function m and covariance function k, i.e. $F \sim \mathcal{GP}(m, k)$, then the partial derivatives of F are also GP-distributed [25]:

$$\frac{\partial F}{\partial x_i} \sim \mathcal{GP}(m_i, k_i) \tag{10}$$

$$m_i(\mathbf{y}_t, \mathbf{x}_t) = \frac{\partial m(\mathbf{y}_t, \mathbf{x}_t)}{\partial x_i} \tag{11}$$

$$k_i(\mathbf{y}_t, \mathbf{x}_t, \mathbf{y}'_t, \mathbf{x}'_t) = \frac{\partial k(\mathbf{y}_t, \mathbf{x}_t, \mathbf{y}'_t, \mathbf{x}'_t)}{\partial x_i \partial x'_i} \tag{12}$$

In particular, this result also applies to the posterior estimate of the function F as in (5), and also a posterior estimate of the partial derivatives. Combining (5) and (10) yields the following estimator [7]:

$$\begin{aligned} \widehat{\text{DCE}}_{x_i \to y}(\mathbf{y}, \mathbf{x}) &= \mathbb{E}\left(\frac{\partial \hat{F}(\mathbf{y}, \mathbf{x})}{\partial x_i} \middle| \mathcal{D}, \mathbf{x}, \mathbf{y} \right) \\ &= \sum_{n=1}^{N} \frac{\partial k(\mathbf{y}, \mathbf{x}, \mathbf{y}_n, \mathbf{x}_n)}{\partial x_i} \alpha_n, \end{aligned} \tag{13}$$

where the coefficients α_n are obtained from (5) and n is ranging over the set of training vectors $\mathbf{x}_n, \mathbf{y}_n$ that we produce from observed signals. Often times, the kernel function k can be differentiated easily, and there are simple expressions available for common kernels [7], [25]. The estimator in (13) is again a function of the test point $(\mathbf{y}, \mathbf{x})$, and it can be averaged over the input space, just as we did for $\langle \text{DCE}_{x_i \to y} \rangle$. Since the probability distribution $p(\mathbf{y}_t, \mathbf{x}_t)$ is not usually available, we can resort to a bootstrap estimate by substituting the observed values of the signal. Thus, we have an estimator for the averaged magnitude of the DCE,

$$\langle \widehat{\text{DCE}}_{x_i \to y} \rangle = \sum_{m=1}^{N} \left| \sum_{n=1}^{N} \frac{\partial k(\mathbf{y}_m, \mathbf{x}_m, \mathbf{y}_n, \mathbf{x}_n)}{\partial x_i} \alpha_n \right|. \tag{14}$$

If desired, the posterior uncertainty in (14) can also be derived from the GP posterior.

D. Double-averaged DCE

As mentioned in Section III-A, the precise input features x_i that strongly influence y_t may vary from trial to trial. To account for this, we require an approach that can average across time lags, but still retains the ability for us to ascertain what time lags are important. To this end, we consider several different values for the model order parameter Q. For each $Q = 1, ..., Q_{max}$, where Q_{max} is specified beforehand, we fit a model of the form (1) and we estimate $\langle \text{DCE}_{x_i \to y} \rangle$ using the GP posterior. We then define the *double-averaged DCE* to

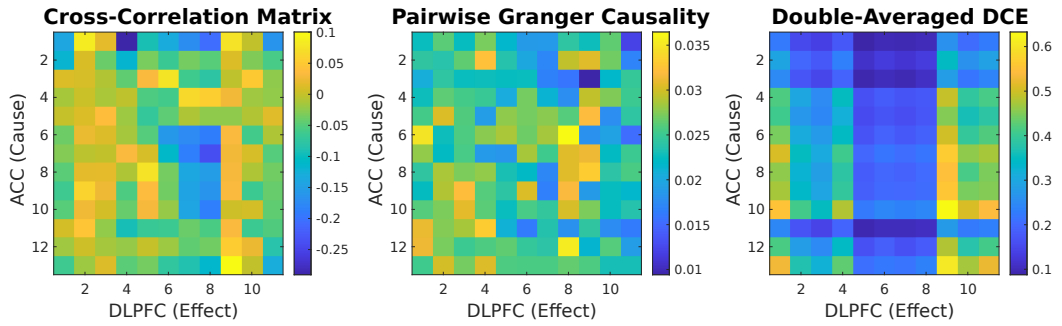

Fig. 2. Comparison of the cross-correlation, pairwise Granger causality [6], and the double-averaged DCE across pairs of channels from the ACC and DLPFC. cross-correlation and Granger causality analyses are suggestive of a relationship between ACC and DLPFC, but the resulting plots don't show recognizable spatial structure across channels. The double-averaged DCE detected causal strength which was more spatially organized than the two linear measures. In all three plots, the matrices were computed using 212 trials of the hierarchical rule task, where each channel recorded 1,000 samples of data per trial.

be the average of the averaged magnitude of the DCE over each x_i within a model with fixed order Q, i.e.

$$\langle\langle \widehat{\mathrm{DCE}}_{x\to y}\rangle\rangle(Q) \triangleq \frac{1}{Q}\sum_{i=t-1}^{t-Q}\langle \widehat{\mathrm{DCE}}_{x_i\to y}\rangle. \tag{15}$$

The measure $\langle\langle \widehat{\mathrm{DCE}}_{x\to y}\rangle\rangle$ compensates for the fact that the propagation delay from cause to effect varies from trial-to-trial by averaging over the observed lags. By considering $\langle\langle \widehat{\mathrm{DCE}}_{x\to y}\rangle\rangle$ as a function of Q, one can study how changes to the time window change the causal strength measurement, without concern that the resulting measurements are distributed across several lags in time.

IV. Results

In this section, we perform several analyses of the LFP data using the double-averaged DCE. Before performing DCE analysis, we first applied linear cross-correlation analyses to explore the data set. Peaks in the cross-correlation generally occurred well within 15ms, and thus $Q_{max} = 30$ was selected after considering the 2kHz sampling rate. To emphasize the relationship of model order to the time window of the model, we will often specify the model order in terms of milliseconds.

In Figure 2, we visualize how the causal strength between ACC and DLPFC varies across sensor locations. In the proposed method using double-averaged DCE, we observe patterns in which some groups of adjacent channels detect more causal strength than other regions. In contrast, we also computed the corresponding cross-correlations and pairwise Granger causalities, to observe how other pairwise analyses fared. The Granger causality method use linear models with the same model order as the double-averaged DCE approach.

In Figure 3, we consider how varying the model order Q allows the double-averaged DCE approach to discover the time delays of interest for the ACC to DLPFC mechanism. We observe that by varying the model order Q, the largest values of the double-averaged DCE typically corresponded to a time window of 6-8 ms. The rising causal strength as Q increases, while $Q \leq 6$, may correspond to the arrival of increasing relevant information as the model important time lags are

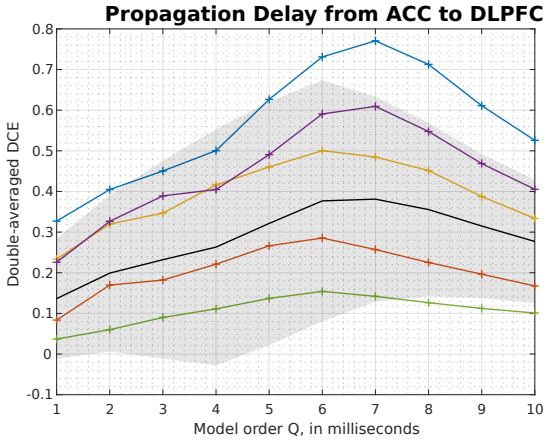

Fig. 3. Double-averaged DCE vs. the model order. In black, we show the average across all pairs of channels and trials, with one standard deviation in the shaded region. The colored curves correspond to the results for fixed pairs of channels, but we still average over trials to produce each curve. The most strong influence appears when the model order (interpreted as a window length) is around 6 to 8 ms.

added to the model. Before the lags, are available, e.g., $Q = 2$ ms, the causal strength is low but not zero, which is expected because even if a more delayed copy of the ACC signal was omitted from the model, the autocorrelation or smoothness of the signal ensures that nearby lags are still statistically meaningful. When $Q \geq 8$, increasing Q appears to decrease the average causal strength, which may indicate that additional useful information is not being added to the model when the time window exceeds the physically meaningful range.

In Figure 4, we examine the trial-to-type variation of the DCE measures. We see that individual trials vary significantly in their behavior, and thus to get a generalized assessment of the strength of causation between the ACC and DLPFC, one must either aggregate over trials, or find a way to explain the variability in terms of a measurable variable. Despite the trial-to-trial variability, the measured causal strength responds comparably to Q; for small Q, the causal strength is low, and for Q exceeding some trial-specific threshold the causal

strength is larger, which agrees with the trends seen in Figure 3.

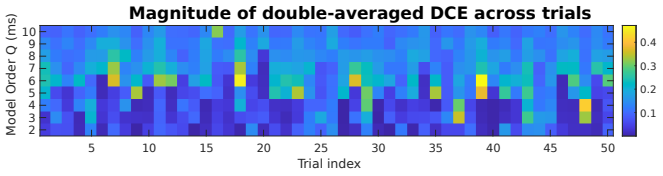

Fig. 4. Heatmap illustrating how the double-averaged DCE varies from trial to trial, for a fixed pair of channels and a subset of the total number of trials. There is noticeably variability from trial to trial, but in general all trials show the behavior that the causal strength is small for small Q, and the measured causal strength of the ACC on the DLPFC increases noticeably when Q passes some trial-dependent threshold.

V. Discussion

Our DCE analyses suggest that medial PFC (ACC) causally influenced lateral PFC (DLPFC) during the cognitive control task. This influence was maximal at a time lag of around 7ms, indicating the signal transmission time from ACC to DLPFC. Although the time lag varied from trial-to-trial of the task, it was consistently between 6-10ms, which is within the expected physiological range. The medial PFC influence occurred during processing of the abstract rule cue, which specifies the subset of subsequent relevant concrete rules to complete the task. Hence, the medial PFC influence on lateral PFC is consistent with a shift from exploration to exploitation of the relevant task rules. This provides support for the theory of medial-to-lateral PFC interactions mediating exploration-exploitation trade-offs [11].

VI. Conclusion

In this paper, we addressed the problem of learning the hierarchical organization of the frontal lobe from LFP signal data. More specifically, the interest was on determining if the medial PFC causally influenced the lateral PFC. Our methodology for exploring causations was based on a recently introduced concept, DCE. In this paper, we introduced the notion of double-averaged DCE as a way to address for large trial-to-trial variability while attempting to measure causality strength from the LFP signals. The current results of the work show evidence that medial PFC causally influences lateral PFC.

References

[1] D. Badre and M. D'esposito. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*, 10(9):659–669, 2009.

[2] D. Badre and D. E. Nee. Frontal cortex and the hierarchical control of behavior. *Trends in cognitive sciences*, 22(2):170–188, 2018.

[3] A. M. Bastos and J.-M. Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, 9:175, 2016.

[4] M. Besserve, N. Shajarisales, B. Schölkopf, and D. Janzing. Group invariance principles for causal generative models. In *International Conference on Artificial Intelligence and Statistics*, pages 557–565. PMLR, 2018.

[5] M. M. Botvinick. Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, 12(5):201–208, 2008.

[6] S. L. Bressler and A. K. Seth. Wiener–Granger Causality: A well established methodology. *Neuroimage*, 58(2):323–329, 2011.

[7] K. Butler, G. Feng, and P. M. Djurić. A Differential Measure of the Strength of Causation. *IEEE Signal Processing Letters*, 29:2208–2212, 2022.

[8] K. Butler, G. Feng, C. B. Mikell, S. Mofakham, and P. M. Djurić. Predicting Latent States of Dynamical Systems With State-Space Reconstruction and Gaussian Processes. In *2022 30th European Signal Processing Conference (EUSIPCO)*, pages 2216–2220. IEEE, 2022.

[9] G. Deco, Y. Sanz Perl, H. Bocaccio, E. Tagliazucchi, and M. L. Kringelbach. The INSIDEOUT framework provides precise signatures of the balance of intrinsic and extrinsic dynamics in brain states. *Communications Biology*, 5(1):572, 2022.

[10] M. Ding, Y. Chen, and S. L. Bressler. Granger causality: basic theory and application to neuroscience. *Handbook of time series analysis: recent theoretical developments and applications*, pages 437–460, 2006.

[11] P. Domenech and E. Koechlin. Executive control and decision-making in the prefrontal cortex. *Current Opinion in Behavioral Sciences*, 1:101–106, 2015.

[12] D. Foreman-Mackey, E. Agol, S. Ambikasaran, and R. Angus. Fast and scalable Gaussian process modeling with applications to astronomical time series. *The Astronomical Journal*, 154(6):220, 2017.

[13] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016.

[14] C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.

[15] Y. Kajikawa and C. E. Schroeder. How local is the local field potential? *Neuron*, 72(5):847–858, 2011.

[16] E. Koechlin, C. Ody, and F. Kouneiher. The architecture of cognitive control in the human prefrontal cortex. *Science*, 302(5648):1181–1185, 2003.

[17] M. Lindner, R. Vicente, V. Priesemann, and M. Wibral. Trentool: A matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC neuroscience*, 12:1–22, 2011.

[18] Y. Liu, C. Cui, M. Ajirak, and P. M. Djurić. Estimation of time-varying graph topologies from graph signals. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[19] J. Pearl. *Causality*. Cambridge University Press, 2009.

[20] R. Ramb, M. Eichler, A. Ing, M. Thiel, C. Weiller, C. Grebogi, C. Schwarzbauer, J. Timmer, and B. Schelter. The impact of latent confounders in directed network analysis in neuroscience. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1997):20110612, 2013.

[21] C. E. Rasmussen and C. K. Williams. *Gaussian Processes for Machine Learning*, volume 2. MIT Press, 2006.

[22] B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio. Toward Causal Representation Learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

[23] T. Schreiber. Measuring information transfer. *Physical review letters*, 85(2):461, 2000.

[24] A. Shenhav, M. M. Botvinick, and J. D. Cohen. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, 79(2):217–240, 2013.

[25] E. Solak, R. Murray-Smith, W. Leithead, D. Leith, and C. Rasmussen. Derivative observations in Gaussian process models of dynamic systems. *Advances in Neural Information Processing Systems*, 15, 2002.

[26] J. Wang, A. Hertzmann, and D. J. Fleet. Gaussian Process Dynamical Models. *Advances in Neural Information Processing Systems*, 18, 2005.