Directed Spectral Measures Improve Latent Network Models Of Neural Populations

Neil M. Gallagher
Department of Neurobiology
Duke University
Durham, NC 27708
neil.gallagher@duke.edu

Kafui Dzirasa
Howard Hughes Medical Institute
Department of Psychiatry and Behavioral Sciences
Department of Neurobiology
Duke University
Durham, NC 27710
kafui.dzirasa@duke.edu

David Carlson
Department of Biostatistics and Bioinformatics
Department of Civil and Environmental Engineering
Duke University
Durham, NC 27708
david.carlson@duke.edu

Abstract

Systems neuroscience aims to understand how networks of neurons distributed throughout the brain mediate computational tasks. One popular approach to identify those networks is to first calculate measures of neural activity (e.g. power spectra) from multiple brain regions, and then apply a linear factor model to those measures. Critically, despite the established role of directed communication between brain regions in neural computation, measures of directed communication have been rarely utilized in network estimation because they are incompatible with the implicit assumptions of the linear factor model approach. Here, we develop a novel spectral measure of directed communication called the Directed Spectrum (DS). We prove that it is compatible with the implicit assumptions of linear factor models, and we provide a method to estimate the DS. We demonstrate that latent linear factor models of DS measures better capture underlying brain networks in both simulated and real neural recording data compared to available alternatives. Thus, linear factor models of the Directed Spectrum offer neuroscientists a simple and effective way to explicitly model directed communication in networks of neural populations.

1 Introduction

A major goal in neuroscience is to characterize how populations of neurons work together to carry out computational tasks [1]. A well-known example is the biological neural network that identifies low-level visual features such as edges. That network sends signals from the retina to the lateral geniculate nucleus, then from there to the primary visual cortex, then finally to higher cortical visual processing regions [2, 3]. While some biological neural networks are well known, we expect that the vast majority remain undiscovered due to the enormous variety of tasks the brain performs. Many methods have been developed to help discover latent networks of neural populations (i.e. brain networks) [4–6]. A key aspect of such methods is that they should be interpretable, meaning that a neuroscientist must be able to use the model to draw conclusions about brain function [7].

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

Unfortunately, many types of common models in the machine learning literature do not offer this type of interpretability [8].

One class of widely-used models that is considered interpretable for modeling brain networks is Linear Factor Models (LFMs) [9, 10]. This family includes popular methods like principal component analysis (PCA) and independent component analysis (ICA), and are used across almost all modalities in neuroscience to model brain networks. For example, ICA is regularly used in functional magnetic resonance imaging (fMRI) applications to identify latent brain networks such as the default mode network [11, 12]. A common approach in neuroscience is to apply an LFM to some measure or extracted feature from the recorded signal, rather than to the recorded signals themselves, in order to get a model of brain networks that are characterized in the desired measure. For example, power spectra are often calculated from raw electroencephalogram (EEG) recordings before an LFM is applied, yielding brain networks defined by power spectra [13, 14]. In multi-site local field potential (LFP) data, LFMs have been applied to identify latent brain networks that are defined by a cross-spectral covariance matrix [15].

One kind of measure that could be especially informative for defining latent brain networks is directed communication between brain regions. However, LFMs that directly incorporate measures of directed communication are lacking. We believe that this is because existing measures of directed communication, such as Granger causality [16], are incompatible with the implicit assumptions of using LFMs to identify brain networks. This incompatibility between standard measures of directed communication and LFMs is discussed further in Sections 2 and 4. In a previous attempt to capture directed communication within an interpretable brain network model, Gallagher et al. [15] modeled networks in terms of phase shifts in spectral content between neural populations, but it is unclear if such phase shifts are an appropriate proxy for directed transmission of signals, and those models are bottlenecked by significant computational time.

In response to this methodological gap, we introduce a novel measure of directed communication that we will refer to as the Directed Spectrum (DS). The Directed Spectrum estimates directed communication between time series in the frequency domain. We prove that these measures are a linear function of latent brain networks under reasonable assumptions, making them compatible with LFMs in this application. We then demonstrate that an LFM of DS measures recovers latent networks in a simulated dataset where the ground truth networks are known. We compare the performance to several competing directed communication measures, including Spectral Granger Causality [17], and show that using the Directed Spectrum results in significantly higher quality reconstruction of the true brain networks. Finally, we show that latent brain networks identified from real neural data via the Directed Spectrum can decode behaviorally relevant information with much higher fidelity.

2 Linear factor models (LFMs) for identifying brain networks

The term *linear factor model* [18] describes any model that seeks to approximate a data vector $\boldsymbol{x}_n$ as,

$$\boldsymbol{x}_n = \sum_{j=1}^{J} Z_n^{(j)} \boldsymbol{x}^{(j)} + \boldsymbol{\epsilon}_n. \tag{1}$$

The vectors $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(J)}$ are the J different latent factor loadings, $\boldsymbol{\epsilon}_n$ is a noise term, and $Z_n^{(j)}$ is the activation score of the jth factor in the nth sample. $\boldsymbol{x}_n$ represents a single sample from a larger dataset, $[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$. Many well-known models, such as PCA [19], ICA [20], and nonnegative matrix factorizaton [21], are LFMs. LFMs are frequently used for modeling and discovering brain networks [10]. In that context, it is straightforward to interpret the factor scores, $Z_n^{(j)}$, as the activation levels for a set of brain networks. Likewise, we can interpret the factor loading $\boldsymbol{x}^{(j)}$ as the observable signature of the jth brain network. LFMs are quite flexible; a wide variety of assumptions can be made regarding the structure of the factors, the scores, and the noise distribution [6, 9, 22]. The fact that LFMs are interpretable and adaptable to most data types has made them a standard model class for studying latent brain networks.

Note: For simplicity, the subscript n will be dropped from all variables for the remainder of this document. In cases where it is important that a variable is constant over all samples, such as the factor loadings $\boldsymbol{x}^{(j)}$, we will explicitly state it.

2.1 Compatibility between linear factor models and measures of neural activity

It is common to model measures of neural activity recordings rather than modeling the recording directly. In this work, we assume that the vector $\boldsymbol{x}$ contains measures calculated from multi-channel time-series recordings, $V = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K]^\intercal$, where K is the number of channels and $\boldsymbol{v}_c \in \mathbb{R}^T$ is the row of V corresponding to a recording from a single channel c. The frequency domain representations of these quantities will be marked by a tilde (e.g. $\tilde{v}_c(\omega)$ is the frequency domain representation of $\boldsymbol{v}_c$). A typical choice for $\boldsymbol{x}$ is to convert the observed data into a set of power spectra (S_{cc}) for each channel [13, 23],

$$S_{cc}(\omega) \equiv E[|\tilde{v}_c(\omega)|^2]. \tag{2}$$

Modeling these measures in an LFM leads to the latent factor loadings $\boldsymbol{x}^{(j)}$ being described in terms of power spectra associated with each channel. The chosen measures dictate the representation of the discovered networks, so we want to use relevant and easy-to-interpret measures.

An implicit assumption in this application of LFMs is that the measures contained in $\boldsymbol{x}$ are *linear* functions of the latent brain networks. This is seen in (1) where the observed measure is a linear function of the latent factor score $Z^{(j)}$ within some noise tolerance. We let $\mathcal{X}(\cdot)$ represent a function, such as (2), that produces the set of measures from our observed neural data V.

Definition 1. *We call $\mathcal{X}(\cdot)$ a <u>linear function</u> of latent brain networks if*

$$\mathcal{X}(V) = \mathcal{X}(\textstyle\sum_{j=1}^{J} Z^{(j)}\Omega^{(j)}) = \sum_{j=1}^{J} Z^{(j)}\mathcal{X}(\Omega^{(j)}), \tag{3}$$

where $\Omega^{(j)}$ represents the component of the neural data (V) that is due to the jth network, normalized to represent a score of 1. In this way, $\mathcal{X}(\Omega^{(j)})$ represents the expected value of the measure if the jth network was active with a score of 1 in isolation.

It can be shown that power spectra are theoretically consistent with this requirement (see Supplemental Section A), but there are many measures that cannot obey this assumption. For example, the Pearson correlation coefficient and Granger causality are two measures that cannot obey this assumption (see Section 4). Including such measures in LFMs will lead to suboptimal representations of latent brain networks.

3 Modeling directed communication in brain networks with LFMs

Measures capturing directed communication between neural populations are rarely included in LFMs despite their frequent use in neuroscience [24–26]. We believe this is because such measures, including Granger causality [16, 17], are not linear functions of brain networks and therefore are poorly modeled by LFMs (see Section 4). Nonetheless, measures of directed communication such as Granger causality are vital for the study of brain networks [27, 28].

Below, we introduce a novel measure of directed communication within a network, which we refer to as the Directed Spectrum (DS). In order to derive the Directed Spectrum, we first build a general time-series model of latent brain networks in Section 3.1, and then define how latent networks combine to produce the observed signal in Section 3.2. The Directed Spectrum is formally defined in Section 3.3. It captures communication in a similar manner to Granger causality while also being compatible with a linear model of brain networks.

3.1 Modeling brain networks as independent VAR processes

We outline a model where the output for each latent network is defined by a vector autoregressive (VAR) process. For additional background on VAR models, see Supplemental Section B. The model outlined below represents a general framework for understanding what the DS measures capture, even though inferring it directly would be quite challenging. We begin by assuming that we have J latent brain networks that each generate vector timeseries outputs. Rather than slicing by channel as in Section 2, we define the output series as $V^{(j)} = [\boldsymbol{v}_1^{(j)}, \ldots, \boldsymbol{v}_T^{(j)}]$, where $\boldsymbol{v}_t^{(j)} \in \mathbb{R}^K$ is the column of $V^{(j)}$ representing signal for all channels at time t. The output series is associated with a VAR process,

$$\boldsymbol{v}_t^{(j)} = \textstyle\sum_{\tau=1}^{p_j} A_\tau^{(j)} \boldsymbol{v}_{(t-\tau)}^{(j)} + \boldsymbol{\sigma}_t^{(j)}, \qquad \boldsymbol{\sigma}_t^{(j)} \sim \mathcal{N}(0, Z^{(j)}\Sigma^{(j)}). \tag{4}$$

We refer to $\boldsymbol{\sigma}_t^{(j)}$ as the *innovations* in the jth network at time t, and assume that they are drawn *iid* from a zero mean Gaussian distribution. Each network has a single covariance structure over all samples ($\Sigma^{(j)}$) that is scaled by an activation score ($Z^{(j)}$) to define the innovation distribution of the current sample. The set of autocovariance matrices for the jth network ($A_\tau^{(j)}$) are also assumed to be constant over all samples. The innovation terms represent new signal introduced into the network that cannot be explained by the past network outputs $\{\boldsymbol{v}_{(t-1)}^{(j)}, \dots, \boldsymbol{v}_{(t-p_j)}^{(j)}\}$. We can represent this network in the frequency domain given standard assumptions regarding the stability of the VAR model [29],

$$\tilde{\boldsymbol{v}}^{(j)}(\omega) = H^{(j)}(\omega)\tilde{\boldsymbol{\sigma}}^{(j)}(\omega), \qquad H^{(j)}(\omega) \equiv \left(I - \textstyle\sum_{\tau=1}^{p_j} A_\tau^{(j)} e^{-i\tau\omega}\right)^{-1}, \qquad 0 \le \omega \le 2\pi, \tag{5}$$

where $\tilde{\boldsymbol{v}}^{(j)}(\omega) \in \mathbb{R}^K$ and $\tilde{\boldsymbol{\sigma}}^{(j)}(\omega) \in \mathbb{R}^K$ are the frequency domain representations of the series $\boldsymbol{v}_1^{(j)}, \dots, \boldsymbol{v}_T^{(j)}$ and $\boldsymbol{\sigma}_1^{(j)}, \dots, \boldsymbol{\sigma}_T^{(j)}$, respectively. The network transfer matrix ($H^{(j)}(\omega)$) is assumed to be constant over all samples.

3.2 Modeling observed signal as a superposition of latent transmitted signals

In order to complete our model of latent brain networks, we must relate it to the observed signal. We model each observed sample by a VAR in the frequency domain, analogous to the way we modeled latent network outputs,

$$\tilde{\boldsymbol{v}}(\omega) = H(\omega)\tilde{\boldsymbol{\sigma}}(\omega). \tag{6}$$

We assume that the K channels in our data are partitioned into a set of non-overlapping groups, $\mathcal{G} = \{\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}, \dots\}$. For example, each group can be a single channel to model all inter-channel relationships, or could be all channels within a given brain region. We partition the observed data and the transfer matrix $H(\omega)$ and innovations $\tilde{\boldsymbol{\sigma}}(\omega)$ into these groups,

$$\begin{aligned}
\tilde{\boldsymbol{v}}(\omega) &= \begin{bmatrix} \tilde{\boldsymbol{v}}_{\boldsymbol{b}}(\omega) \\ \tilde{\boldsymbol{v}}_{\boldsymbol{c}}(\omega) \\ \tilde{\boldsymbol{v}}_{\boldsymbol{d}}(\omega) \\ \vdots \end{bmatrix}, \quad H(\omega) = \begin{bmatrix} H_{\boldsymbol{bb}}(\omega) & H_{\boldsymbol{bc}}(\omega) & H_{\boldsymbol{bd}}(\omega) & \\ H_{\boldsymbol{cb}}(\omega) & H_{\boldsymbol{cc}}(\omega) & H_{\boldsymbol{cd}}(\omega) & \cdots \\ H_{\boldsymbol{db}}(\omega) & H_{\boldsymbol{dc}}(\omega) & H_{\boldsymbol{dd}}(\omega) & \\ & \vdots & & \end{bmatrix}, \\
\tilde{\boldsymbol{\sigma}}(\omega) &= \begin{bmatrix} \tilde{\boldsymbol{\sigma}}_{\boldsymbol{b}}(\omega) \\ \tilde{\boldsymbol{\sigma}}_{\boldsymbol{c}}(\omega) \\ \tilde{\boldsymbol{\sigma}}_{\boldsymbol{d}}(\omega) \\ \vdots \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{\boldsymbol{bb}} & \Sigma_{\boldsymbol{bc}} & \Sigma_{\boldsymbol{bd}} & \\ \Sigma_{\boldsymbol{cb}} & \Sigma_{\boldsymbol{cc}} & \Sigma_{\boldsymbol{cd}} & \cdots \\ \Sigma_{\boldsymbol{db}} & \Sigma_{\boldsymbol{dc}} & \Sigma_{\boldsymbol{dd}} & \\ & \vdots & & \end{bmatrix}.
\end{aligned} \tag{7}$$

We assume that the networks outputs and corresponding VAR parameters are partitioned in the same way. The ordering of these groups is arbitrary. Without loss of generality, we will focus on modeling communication between groups $\boldsymbol{b}$ and $\boldsymbol{c}$.

We examine how our VAR models represent communication between groups by noting that

$$\tilde{\boldsymbol{v}}_{\boldsymbol{c}}(\omega) = \textstyle\sum_{\boldsymbol{g}\in\mathcal{G}} H_{\boldsymbol{cg}}(\omega)\tilde{\boldsymbol{\sigma}}_{\boldsymbol{g}}(\omega). \tag{8}$$

This shows that $\tilde{\boldsymbol{v}}_{\boldsymbol{c}}(\omega)$ is a sum of contributions from each group $\boldsymbol{g}$, including the self-contribution from $\boldsymbol{c}$. If the innovations in $\boldsymbol{b}$ and $\boldsymbol{c}$ are independent, then $H_{\boldsymbol{cb}}(\omega)\tilde{\boldsymbol{\sigma}}_{\boldsymbol{b}}(\omega)$ would unambiguously represent the signal in $\boldsymbol{c}$ that can be attributed to innovations $\boldsymbol{b}$. If the innovations in $\boldsymbol{b}$ and $\boldsymbol{c}$ are not independent, then there is ambiguity regarding how to assign 'responsibility' for the observed signal in $\boldsymbol{c}$. Specifically, the component of the innovations in $\boldsymbol{b}$ that is correlated with innovations in $\boldsymbol{c}$ is not necessary to explain the observed signal in $\boldsymbol{c}$, since the innovations in $\boldsymbol{c}$ are sufficient to explain this part of the observed signal. We will assume that only the uncorrelated component of the innovations in $\boldsymbol{b}$ contribute to the observed signal in $\boldsymbol{c}$.

Definition 2. *The* <u>*transmitted signal*</u> *represents the contribution of the innovations in* $\boldsymbol{b}$ *to the observed signal in* $\boldsymbol{c}$

$$\mathcal{TS}_{\boldsymbol{b}\to\boldsymbol{c}}(\omega) = \begin{cases} H_{\boldsymbol{cb}}(\omega)\left(\tilde{\boldsymbol{\sigma}}_{\boldsymbol{b}}(\omega) - \Sigma_{\boldsymbol{bc}}\Sigma_{\boldsymbol{cc}}^{-1}\tilde{\boldsymbol{\sigma}}_{\boldsymbol{c}}(\omega)\right), & \boldsymbol{b} \neq \boldsymbol{c} \\ \left(H_{\boldsymbol{cc}}(\omega) + \sum_{k=1}^{K} H_{\boldsymbol{c}k}(\omega)\Sigma_{k\boldsymbol{c}}\Sigma_{\boldsymbol{cc}}^{-1}\right)\tilde{\boldsymbol{\sigma}}_{\boldsymbol{c}}(\omega), & \boldsymbol{b} = \boldsymbol{c}. \end{cases} \tag{9}$$

By restricting $\mathcal{TS}_{\boldsymbol{b}\to\boldsymbol{c}}(\omega)$ to only convey uncorrelated innovations, we construct a more conservative estimate of the effect of $\boldsymbol{b}$ on $\boldsymbol{c}$. This is not the only choice that could be made; for example, we could use the full innovations, but we believe that using the uncorrelated innovations provides the cleanest interpretation.

We complete our model of the observed signal $\tilde{\boldsymbol{v}}(\omega)$ with the following two assumptions:

Assumption 1. *The innovations associated with any network are independent of the innovations in all other networks.*

Assumption 2. *The transmitted signals of the observed data are the sum of the corresponding transmitted signals for the latent networks,*

$$\mathcal{TS}_{k\to\boldsymbol{c}}(\omega) = \textstyle\sum_{j=1}^{J} \mathcal{TS}^{(j)}_{k\to\boldsymbol{c}}(\omega), \tag{10}$$

where the transmitted signal is defined for the latent networks as

$$\mathcal{TS}^{(j)}_{\boldsymbol{b}\to\boldsymbol{c}}(\omega) = \begin{cases} H^{(j)}_{\boldsymbol{cb}}(\omega)\left(\tilde{\boldsymbol{\sigma}}^{(j)}_{\boldsymbol{b}}(\omega) - \Sigma^{(j)}_{\boldsymbol{bc}}{\Sigma^{(j)}_{\boldsymbol{cc}}}^{-1}\tilde{\boldsymbol{\sigma}}^{(j)}_{\boldsymbol{c},n}(\omega)\right), & \boldsymbol{b}\neq\boldsymbol{c} \\ \left(H^{(j)}_{\boldsymbol{cc}}(\omega) + \sum_{k=1}^{K} H^{(j)}_{\boldsymbol{c}k}(\omega)\Sigma^{(j)}_{k\boldsymbol{c}}{\Sigma^{(j)}_{\boldsymbol{cc}}}^{-1}\right)\tilde{\boldsymbol{\sigma}}^{(j)}_{\boldsymbol{c}}(\omega), & \boldsymbol{b}=\boldsymbol{c} \end{cases}. \tag{11}$$

Assumption 1 can be thought of as enforcing the independence of inputs to each network. Assumption 2 is equivalent to the following two statements: 1) the propagation of signal within each network is independent of propagation in each other network, and 2) the observed outputs occur in some medium where simultaneously occurring phenomena obey the laws of superposition, which is a property of many neural activity measurement modalities (e.g., electrical potentials, fluorescence). By combining these two assumptions with (8), we get the result that the observed data are a superposition of the network outputs,

$$\tilde{\boldsymbol{v}}_{\boldsymbol{c}}(\omega) = \textstyle\sum_{\boldsymbol{g}\in\mathcal{G}} H_{\boldsymbol{cg}}(\omega)\tilde{\boldsymbol{\sigma}}_{\boldsymbol{g}}(\omega) \qquad = \sum_{\boldsymbol{g}\in\mathcal{G}} \mathcal{TS}_{\boldsymbol{g}\to\boldsymbol{c}}(\omega) \tag{12}$$

$$= \textstyle\sum_{\boldsymbol{g}\in\mathcal{G}}\sum_{j=1}^{J} \mathcal{TS}^{(j)}_{\boldsymbol{g}\to\boldsymbol{c}}(\omega) \quad = \sum_{\boldsymbol{g}\in\mathcal{G}}\sum_{j=1}^{J} H^{(j)}_{\boldsymbol{cg}}(\omega)\tilde{\boldsymbol{\sigma}}^{(j)}_{\boldsymbol{g}}(\omega) = \sum_{j=1}^{J}\tilde{\boldsymbol{v}}^{(j)}_{\boldsymbol{c}}(\omega). \tag{13}$$

Note in (12) and (13), summing the *transmitted signals* to $\boldsymbol{c}$ over all source groups, including $\boldsymbol{c}$, causes the conditioning terms in our definitions of the *transmitted signal* to cancel out. Because the relationship above does not contain a noise term, non-physiological sources of signal would simply be represented as additional networks in the model. This completes our model of the relationship between latent brain networks and observed signals.

3.3 The Directed Spectrum

The model defined above provides a framework for determining whether a measure of neural recordings can be considered a linear function of latent brain networks.

Definition 3. *We define our measure, <u>the Directed Spectrum (DS)</u>, as the second moment of the transmitted signal,*

$$\mathcal{DS}_{\boldsymbol{b}\to\boldsymbol{c}}(\omega) \equiv \mathbb{E}\left[\widetilde{\mathcal{TS}}_{\boldsymbol{b}\to\boldsymbol{c}}(\omega)\widetilde{\mathcal{TS}}^{*}_{\boldsymbol{b}\to\boldsymbol{c}}(\omega)\right] = H_{\boldsymbol{cb}}(\omega)\Sigma_{\boldsymbol{b}|\boldsymbol{c}}H^{*}_{\boldsymbol{cb}}(\omega), \tag{14}$$

$$\Sigma_{\boldsymbol{b}|\boldsymbol{c}} = \Sigma_{\boldsymbol{bb}} - \Sigma_{\boldsymbol{bc}}\Sigma^{-1}_{\boldsymbol{cc}}\Sigma^{*}_{\boldsymbol{bc}},$$

where $\Sigma_{\boldsymbol{b}|\boldsymbol{c}}$ represents the innovation variance for $\boldsymbol{b}$ conditioned on the innovations in $\boldsymbol{c}$.

The Directed Spectrum corresponds to the portion of the power spectrum for $\boldsymbol{c}$ that is explained by signal that originated in $\boldsymbol{b}$ (see Supplemental Section C for more details). It is a linear function of the latent networks when data is appropriately modeled as described above (for proof, see Supplemental Section D).

The Directed Spectrum can be estimated efficiently by fitting a VAR model to the observed data or via factorization of the cross-spectral matrix associated with the observed data [24, 30, 31]. Additional algorithmic details can be found in Supplemental Section F.1. These DS measures can be calculated according to (14) for data partitioned into any number of groups greater than one. In some situations, it is desirable to calculate the DS values separately for each pair of groups, by estimating $H(\omega)$ and Σ using reduced models that contains only two groups per model. We refer to this method as the *Pairwise Directed Spectrum* (PDS). One benefit of the PDS is that it simplifies the handling of missing channels from the data, since the PDS for the non-missing groups of channels can still be calculated without any adjustments.

4 Related measures of directed communication

The Directed Spectrum is a necessary development because similar measures typically used in neuroscience are not linear functions of latent brain networks. We first examine Granger causality, which measures the degree to which one signal or group of signals helps predict future values of another signal or group [16]. Spectral Granger causality is a modification that separates Granger causality into components associated with individual frequencies [17]. It is a leading spectral measure of directed communication in neuroscience [24, 25], and we view it as conceptually the closest measure to the Directed Spectrum. The (unconditional) spectral Granger causality from $\boldsymbol{b}$ to $\boldsymbol{c}$ can be defined in terms of the Directed Spectrum,

$$\boldsymbol{f}_{\boldsymbol{b}\to\boldsymbol{c}}(\omega) = \ln\left(\frac{|S_{\boldsymbol{cc}}(\omega)|}{|S_{\boldsymbol{cc}}(\omega) - \mathcal{DS}_{b\to c}(\omega)|}\right). \tag{15}$$

It is also possible to calculate a conditional spectral Granger causality that accounts for all other groups in the recording before assigning influence from $\boldsymbol{b}$ to $\boldsymbol{c}$ [32]. Neither form of spectral Granger causality is a linear function of the network models described in Section 3. To see this, consider the case where only the jth network is present without any other signals being added to it. Note that changing $Z^{(j)}$ would then have no impact on $\boldsymbol{f}_{\boldsymbol{b}\to\boldsymbol{c}}(\omega)$ because both the numerator and denominator in (15) would scale equally with $Z^{(j)}$. Standard (non-spectral) Granger causality is equivalent to a scaled integral over all frequencies of the spectral Granger causality [17], meaning that it also would remain constant as $Z^{(j)}$ changes.

There are several other spectral measures of directed communication used in neuroscience, including phase slope index [33], partial directed coherence [34], and the directed transfer function [35]. Each of these can not be a linear function of latent brain networks under reasonable assumptions. We offer the following brief explanations here, with more details in Supplemental Section E. The phase slope index cannot capture bidirectional transmission of signal (all signal is considered unidirectional) [36]. Partial directed coherence and the directed transfer function are both based in VAR models like the Directed Spectrum, but they rely solely on the transfer matrix properties and do not use the innovations. As such, they are scale-invariant and are not linear functions of the brain networks.

There are a number of other non-spectral measures of directed communication used in neuroscience [25, 26]. The cross-correlation function between activity in different neural populations has been used to identify directed communication between neuronal populations [37]. In these applications it has been used to identify the lag associated with unidirectional transmission between populations, and is considered unreliable in situations involving bidirectional communication [25]. Transfer entropy provides an information theoretic measure of directed communication. In our application transfer entropy suffers from the same problem as Granger causality, where scaling the activation of a single network results in the same value [38, 39]. We discuss methods for measuring directed communication between neural populations further in Supplemental Section E.

One final method worth mentioning here is dynamic causal modeling (DCM) [40]. DCM produces a generative model to explain observed neural activity based on latent connectivity. Directed connectivity is represented by parameters of the model, which contrasts with our application where measures of directed connectivity are calculated first before being used as features in an LFM. DCM has even been extended in order to represent spectral content in neural signals [41]. One substantial difference between DCM and the approach we suggest here is that DCM is not designed to segregate the directed connectivity out into subnetworks in an unsupervised manner. Instead, DCM is typically used to identify differences in connectivity between two known conditions [42].

5 Directed Spectrum improves identification of latent networks

To test the effectiveness of the Directed Spectrum and LFMs, we generated a dataset of simulated latent networks. The dataset contains 10,000 independent recordings, each containing 5 channels sampled for five seconds at 500 Hz. The "observed" recordings were generated as a sum of contributions from three independent networks. Each network was associated with a vector autoregressive (VAR) model that generated network contributions as a different random draw for each recording with parameters that were fixed over all recordings. Each VAR model was designed to generate a particular pattern of oscillations and delayed directional transmission of signals. See Figure 1 for the network properties and Figure 2 for a visualization of the superposition. See Supplemental Section F.2 for additional details on the design of those VAR models.

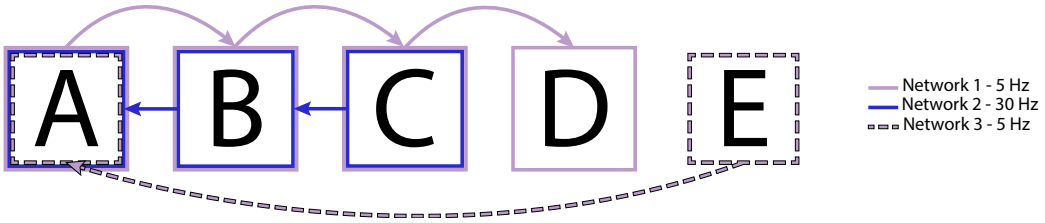

Figure 1: Graphical representation of the simulated latent networks. The five channels are represented by the letters A, B, C, D, and E. A box around a channel indicates the corresponding network induces oscillations in that channel. Arrows indicate propagation of signal at some delay. Networks 1 and 3 have a predominant frequency of 5 Hz (both in purple), and Network 2 has a predominant frequency of 30 Hz (in blue).

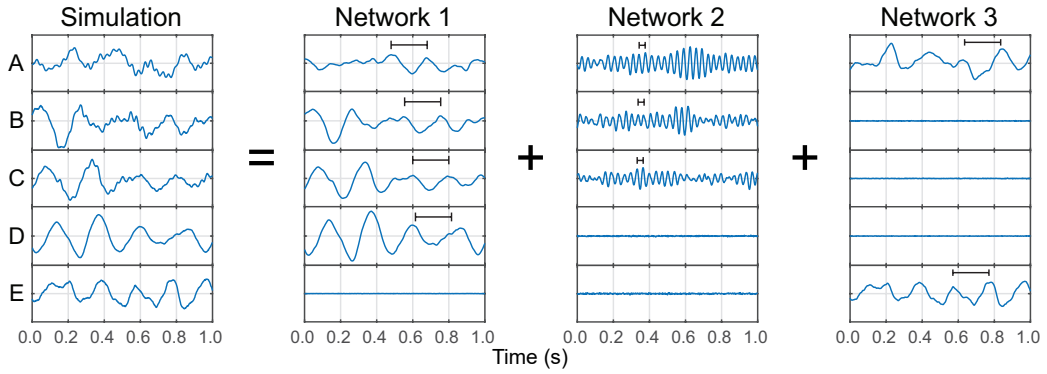

Figure 2: Simulated recordings are a sum of latent network contributions. The "observed" recording (left) is a sum of signals generated by each of the three networks. Each panel contains 5 signals associated with the 5 regions over the same one second period. Black scale bars indicate the period for oscillations associated with each network, and are placed to show that the distance between peaks in the signal is approximately one period.

We generated a model of the latent networks in the simulated dataset by applying an LFM to DS measures. We calculated the Directed Spectrum for frequencies between 1 Hz and 50 Hz at 1 Hz intervals, between all directed pairs of channels. A non-negative matrix factorization model (i.e. LFM) was trained using the Itakura-Saito (IS) divergence loss, with an L1 penalty on the factor loadings and scores, and multiplicative update steps for optimization [22]. The IS divergence loss was chosen because it is related to the gamma distribution and is a more appropriate loss for power than mean squared error loss. After training, the non-negative matrix factorization (NMF) factors were re-ordered to correspond to the matching network in the simulation, based on maximizing the average Spearman's correlation between the estimated and true network activation scores. For comparison, we repeated this process with the Pairwise Directed Spectrum (PDS), unconditional spectral Granger causality (GC) [17], conditional spectral Granger causality (cGC) [32], phase-slope index (PSI) [33], directed transfer function (DTF) [35], and partial directed coherence (PDC) [34] features substituted for the DS features. We also use the difference between (unconditional) Granger causality values in either direction as suggested by Roebroeck et al. [43] as an additional comparison method. Because both this difference and PSI represent the direction of shared information with sign (i.e. positive/negative), we set negative values to zero and leave the opposite direction as positive for both of those comparison methods; this allowed us to continue to use NMF to identify latent networks.

We evaluated how well each model recovered the underlying network activation scores using Spearman's correlation. The model of DS features performed significantly better in this regard than the "non-linear" features (see Table 1). The PDS model also performed significantly better than the "non-linear" models, and gave comparable results to the DS model. We have visualized the estimated scores along with the true scores for each window for the DS features and unconditional Granger causality features in Figure 3. We see that the score estimates of the DS model are much more tightly

Table 1: Spearman's correlation between latent network activation estimates and true activation scores. GC: unconditional Granger causality; cGC: conditional Granger Causality; GCdiff: difference between Granger causality directions; PSI: phase slope index; DTF: directed transfer function; PDC: partial directed coherence; DS: Directed Spectrum; PDS: Pairwise Directed Spectrum. Values in *[brackets]* represent the 95% confidence interval [44].

Measure	Network 1	Network 2	Network 3
GC	0.485 *[0.468, 0.502]*	0.442 *[0.424, 0.459]*	0.281 *[0.261, 0.300]*
cGC	0.475 *[0.458, 0.491]*	0.293 *[0.274, 0.312]*	0.127 *[0.108, 0.147]*
GCdiff	0.554 *[0.539, 0.569]*	0.501 *[0.484, 0.517]*	0.444 *[0.427, 0.462]*
PSI	0.387 *[0.368, 0.405]*	0.135 *[0.115, 0.155]*	0.248 *[0.229, 0.268]*
DTF	0.426 *[0.408, 0.443]*	0.131 *[0.111, 0.151]*	0.542 *[0.526, 0.557]*
PDC	0.560 *[0.545, 0.575]*	0.154 *[0.134, 0.174]*	0.445 *[0.427, 0.462]*
DS	0.920 *[0.916, 0.923]*	0.905 *[0.901, 0.909]*	0.927 *[0.924, 0.930]*
PDS	0.908 *[0.904, 0.912]*	0.917 *[0.913, 0.920]*	0.916 *[0.913, 0.920]*

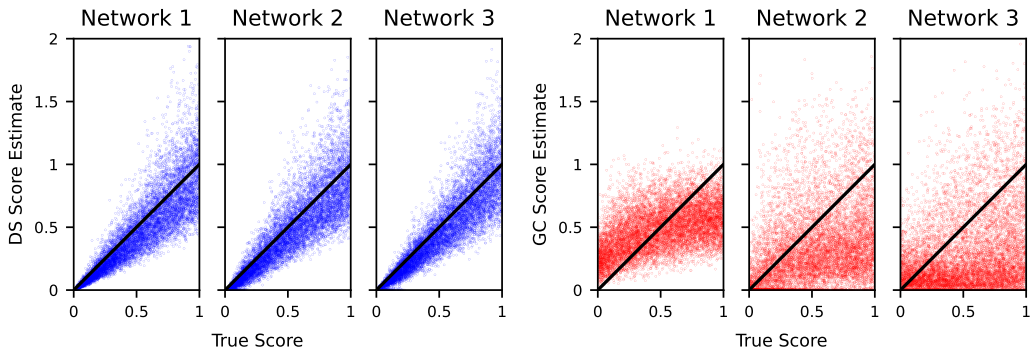

Figure 3: True vs. estimated network scores. The estimated scores for the model trained using DS features are plotted in blue, the estimated scores for a model trained on unconditional Granger causality features are plotted in red. The line in black demonstrates the expected trend for a model that perfectly recovers the latent network scores.

spread around the true score values, with the relationship between true scores and estimated scores being very weak in the model of GC features.

We also tested how robust these results were to violations of the model assumptions described in Section 3 and to variations in recording window length. Violations of the model assumptions did reduce model performance somewhat, but the DS features still performed better than the comparison methods. Shorter window lengths reduced performance in all models, but again the DS models performed better at all window lengths. Full details and results for these claims are in Supplemental Sections G and H.

Finally, we tested whether the models produced accurate and interpretable representations of the true latent networks from the dataset. The model of DS features successfully recovers all directional influences of each network in the corresponding factors (see Figures 4, S1). Spurious detection of network components was limited to connections between true nodes in the network at the frequency associated with the network, and these spurious influences were small relative to the true detected influences. While the model of GC features did also recover all directional influences of the networks underlying the simulated data, it incorrectly spread the directional influences associated with each network out over multiple estimated networks. This resulted in the GC model identifying more spurious influences with greater relative amplitudes.

6 Directed Spectrum improves brain state decoding of neural recordings

We next tested how effectively networks identified by the Directed Spectrum decode behaviorally relevant variables from neural recordings, as a proxy for determining whether they reflect "real"

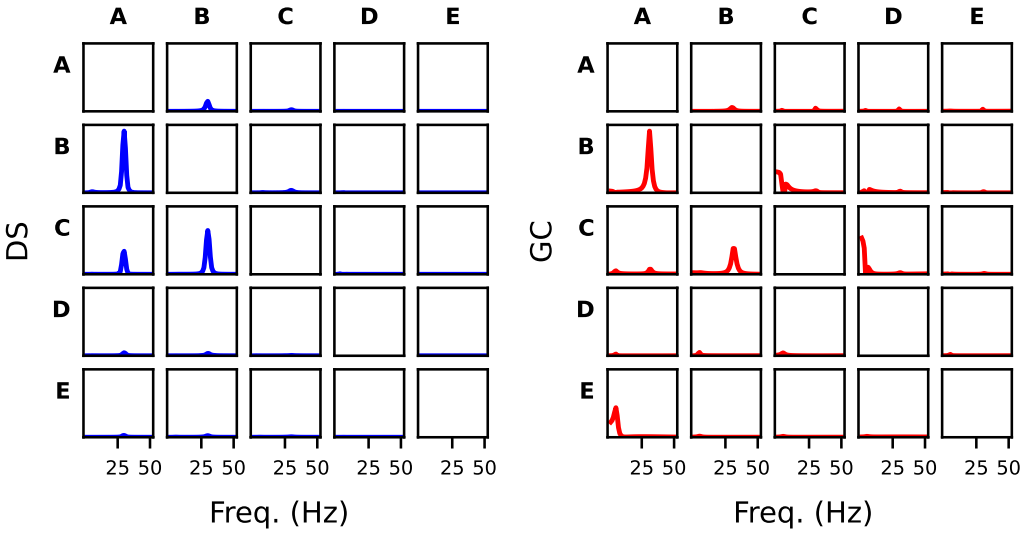

Figure 4: Estimated directional influences in Network 2 for a model of DS features (blue) and unconditional Granger causality features (red). Within each grid, a plot corresponds to signal that is being transmitted from the channel listed on the left to the channel listed above.

networks. The dataset consists of local field potentials simultaneously recorded from 11 brain regions in 26 mice (originally published in [27]). In each recording session local field potentials were recorded from a mouse while it was exposed to three different behavioral contexts of successively increasing stress levels: resting in the home cage, exploring an open field, and a tail suspension test. The tail suspension test is commonly used to investigate learned helplessness [45].

In order to train networks that could be used to decode behavioral context, we first divided the dataset into time windows with a duration of 1 second. For each time window, we calculated the Directed Spectrum for all directed pairs of brain regions, for frequencies from 1 to 56 Hz at 1 Hz intervals. A nonnegative matrix factorization model with L1 regularization was trained using the IS divergence objective [46] in order to identify latent brain networks from the DS measures. A multinomial logistic regression classifier with L1 regularization was then applied to the latent factor scores for decoding the behavioral context. In order to obtain estimates of the spread of the decoding performance and choose optimal hyperparameter values, a 5-fold nested cross-validation procedure was used, where each mouse was associated with only one split. The average one-vs-all area under the reciever operating characteristic curve (AUC) across all three behavioral contexts was used as the evaluation metric. The hyperparameters tuned during cross-validation were the number of NMF factors (20, 40, 80), NMF regularization strength (1000, 100, 10, 0), and logistic regression regularization strength (10, 1). Similar to the experiment described in Section 5, the procedure was repeated using the Pairwise Directed Spectrum (PDS) and the comparison measures listed in that section. The one comparison method that was not used here is conditional Granger causality; in our tests, obtaining stable conditional spectral Granger causality estimates on our rodent LFP data via the methods outlined by Barnett and Seth [24] requires too much computation time to be practical for this application.

Our training procedure resulted in a one-vs-all AUC for each of the behavioral contexts, for each of the 5 splits, for each of the measure types (see Table 2). We performed a 2-factor repeated measures ANOVA and determined that there was a significant difference associated with the measure types ($F = 92.9$, $p < .001$). A Tukey's HSD post-hoc test revealed that there was not a significant difference between models of the pairwise and non-pairwise versions of DS ($p = .90$), but that both performed significantly better than models of the other measures ($p \leq .001$ for all).

7 Discussion

We have shown that our novel measure of directed communication, the Directed Spectrum (DS), is a linear function of latent brain networks under reasonable assumptions, and so is compatible with

Table 2: Behavioral Context Decoding Performance. The columns 'HC AUC', 'OF AUC', and 'TS AUC' report the mean and standard error of the one-vs-all AUC across 5 splits for the homecage, open field, and tail suspension behavioral contexts, respectively. The 'Mean AUC' column reports the average across the mean AUCs reported for each behavioral context.

Measure	*Mean AUC*	HC AUC	OF AUC	TS AUC
GC	0.828	0.825 ± 0.019	0.824 ± 0.022	0.835 ± 0.010
GCdiff	0.795	0.795 ± 0.013	0.798 ± 0.023	0.791 ± 0.007
PSI	0.674	0.676 ± 0.014	0.684 ± 0.015	0.661 ± 0.016
DTF	0.755	0.774 ± 0.017	0.774 ± 0.015	0.717 ± 0.022
PDC	0.717	0.733 ± 0.021	0.778 ± 0.014	0.639 ± 0.008
DS	$\mathbf{0.908}$	0.894 ± 0.016	0.916 ± 0.012	0.915 ± 0.007
PDS	$\mathbf{0.919}$	0.909 ± 0.014	0.915 ± 0.014	0.932 ± 0.005

LFMs for characterizing latent brain networks. We saw drastically improved recovery of networks in simulated data with a known ground truth. In real neural recordings, the Directed Spectrum improved decoding of neurally relevant environmental variables. These results demonstrate that the Directed Spectrum allows for more accurate and interpretable models of latent networks defined by directed communication.

A limitation that could be explored in future work is that the Directed Spectrum assumes that the network states are stationary over a single time window (see Supplemental Section B). This limitation is common to almost all methods that assess spectral content of neural activity. Because of this, the Directed Spectrum is only theoretically grounded when applied to time windows that are shorter than the time expected for substantial change to be occur in the latent brain state. In practice, time window lengths of one to five seconds have been considered a relatively stable period in studies of emotional processing [47, 48].

We have only explored LFMs here, but a variety of nonlinear latent factor models exist for modeling latent brain networks [5, 49]. Nonlinear models are especially useful in applications where predictive performance is the only priority, such as brain-computer interfaces. When the primary goal is to drive scientific understanding of the brain, nonlinear models are less desirable because it is challenging to relate the parameters of such models to relevant conclusions about brain function [8]. We believe that LFMs and the Directed Spectrum are an appropriate, efficient, and reproducible baseline approach in both scientific and prediction-based applications, while noting that other more expressive models may lead to even better predictive performance at the cost of interpretability.

The development of the Directed Spectrum provides a straightforward way to generate linear models of latent brain networks defined by directed communication. We view this as an important advancement because directed communication between neural populations is a critical component of the way neuroscientists understand brain networks. Furthermore, the Directed Spectrum is likely useful for studying latent networks in many other fields as well (e.g. latent networks of directed internet traffic). We view the major strength of the Directed Spectrum is that it enhances the expressiveness of LFMs by allowing them to accurately model directed communication within networks. Thus, we have expanded the capabilities of LFMs while retaining the level of interpretability that make them attractive models in practice.

Acknowledgments and Disclosure of Funding

We thank the anonymous reviewers for their suggestions to evaluate how robust the methods are to assumption violations, which led to the additional experiments of Supplemental Sections G and H.

Research reported in this manuscript was supported by the National Institute of Biomedical Imaging and Bioengineering and the National Institute of Mental Health through the National Institutes of Health BRAIN Initiative under Award Number R01EB026937, the National Institute of Mental Health of the National Institutes of Health under Award Numbers R01MH125430 and R01MH120158, and the Hope for Depression Research Foundation.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] Bratislav Mišić and Olaf Sporns. From regions to connections and networks: new bridges between brain and behavior. *Current Opinion in Neurobiology*, 40:1–7, October 2016. ISSN 0959-4388. doi: 10.1016/j.conb.2016.05.003. URL https://www.sciencedirect.com/science/article/pii/S095943881630054X.

[2] Javier Cudeiro and Adam M. Sillito. Looking back: corticothalamic feedback and early visual processing. *Trends in Neurosciences*, 29(6):298–306, June 2006. ISSN 0166-2236. doi: 10.1016/j.tins.2006.05.002. URL https://www.sciencedirect.com/science/article/pii/S0166223606000907.

[3] Dale Purves, George J Augustine, David Fitzpatrick, William C Hall, Anthony-Samuel LaMantia, James O McNamara, and Leonard E White. Neuroscience. 4th. *Sunderland, Mass.: Sinauer. xvii*, 857:944, 2008.

[4] Heejong Bong, Zongge Liu, Zhao Ren, Matthew Smith, Valerie Ventura, and Kass E. Robert. Latent Dynamic Factor Analysis of High-Dimensional Neural Recordings. *Advances in Neural Information Processing Systems*, 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/beb04c41b45927cf7e9f8fd4bb519e86-Abstract.html.

[5] Shu Zhang, Qinglin Dong, Wei Zhang, Heng Huang, Dajiang Zhu, and Tianming Liu. Discovering hierarchical common brain networks via multimodal deep belief network. *Medical Image Analysis*, 54:238–252, May 2019. ISSN 1361-8415. doi: 10.1016/j.media.2019.03.011. URL https://www.sciencedirect.com/science/article/pii/S1361841518308119.

[6] Alex H. Williams, Tony Hyun Kim, Forea Wang, Saurabh Vyas, Stephen I. Ryu, Krishna V. Shenoy, Mark Schnitzer, Tamara G. Kolda, and Surya Ganguli. Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. *Neuron*, 98(6):1099–1115.e8, June 2018. ISSN 0896-6273. doi: 10.1016/j.neuron.2018.05.015. URL https://www.sciencedirect.com/science/article/pii/S0896627318303878.

[7] Mai-Anh T. Vu, Tülay Adalı, Demba Ba, György Buzsáki, David Carlson, Katherine Heller, Conor Liston, Cynthia Rudin, Vikaas S. Sohal, Alik S. Widge, Helen S. Mayberg, Guillermo Sapiro, and Kafui Dzirasa. A Shared Vision for Machine Learning in Neuroscience. *Journal of Neuroscience*, 38(7):1601–1607, February 2018. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.0508-17.2018. URL https://www.jneurosci.org/content/38/7/1601. Publisher: Society for Neuroscience Section: TechSights.

[8] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, May 2019. ISSN 2522-5839. doi: 10.1038/s42256-019-0048-x. URL https://www.nature.com/articles/s42256-019-0048-x. Number: 5 Publisher: Nature Publishing Group.

[9] Madeleine Udell, Corinne Horn, Reza Zadeh, and Stephen Boyd. Generalized Low Rank Models. *Foundations and Trends® in Machine Learning*, 9(1):1–118, June 2016. ISSN 1935-8237, 1935-8245. doi: 10.1561/2200000055. URL https://www.nowpublishers.com/article/Details/MAL-055. Publisher: Now Publishers, Inc.

[10] John P. Cunningham and Byron M. Yu. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509, November 2014. ISSN 1097-6256. doi: 10.1038/nn.3776. URL http://www.nature.com/neuro/journal/v17/n11/full/nn.3776.html.

[11] Vince D. Calhoun, Jingyu Liu, and Tülay Adalı. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage*, 45(1, Supplement 1):S163–S172, March 2009. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2008.10.057. URL https://www.sciencedirect.com/science/article/pii/S1053811908012032.

[12] Vesa Kiviniemi, Tapani Vire, Jukka Remes, Ahmed Abou Elseoud, Tuomo Starck, Osmo Tervonen, and Juha Nikkinen. A Sliding Time-Window ICA Reveals Spatial Variability of the Default Mode Network in Time. *Brain Connectivity*, 1(4):339–347, September 2011. ISSN 2158-0014. doi: 10.1089/brain.2011.0036. URL https://www.liebertpub.com/doi/abs/10.1089/brain.2011.0036. Publisher: Mary Ann Liebert, Inc., publishers.

[13] Fumikazu Miwakeichi, Eduardo Martinez-Montes, Pedro A. Valdés-Sosa, Nobuaki Nishiyama, Hiroaki Mizuhara, and Yoko Yamaguchi. Decomposing EEG data into space–time–frequency components using Parallel Factor Analysis. *NeuroImage*, 22(3):1035–1045, July 2004. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2004.03.039. URL https://www.sciencedirect.com/science/article/pii/S1053811904001958.

[14] Aapo Hyvärinen, Pavan Ramkumar, Lauri Parkkonen, and Riitta Hari. Independent component analysis of short-time Fourier transforms for spontaneous EEG/MEG analysis. *NeuroImage*, 49(1):257–271, January 2010. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2009.08.028. URL https://www.sciencedirect.com/science/article/pii/S1053811909009409.

[15] Neil Gallagher, Kyle R Ulrich, Austin Talbot, Kafui Dzirasa, Lawrence Carin, and David E Carlson. Cross-Spectral Factor Analysis. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 6845–6855, 2017. URL http://papers.nips.cc/paper/7260-cross-spectral-factor-analysis.pdf.

[16] C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–438, 1969. ISSN 0012-9682. doi: 10.2307/1912791. URL https://www.jstor.org/stable/1912791. Publisher: [Wiley, Econometric Society].

[17] John Geweke. Measurement of Linear Dependence and Feedback between Multiple Time Series. *Journal of the American Statistical Association*, 77(378):304–313, June 1982. ISSN 0162-1459. doi: 10.1080/01621459.1982.10477803. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1982.10477803.

[18] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Chapter 13: Linear Factor Models. In *Deep Learning*, pages 485–498. MIT Press, November 2016. ISBN 978-0-262-33737-3. Google-Books-ID: omivDQAAQBAJ.

[19] Hervé Abdi and Lynne J. Williams. Principal component analysis. *WIREs Computational Statistics*, 2(4):433–459, 2010. ISSN 1939-0068. doi: https://doi.org/10.1002/wics.101. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.101. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wics.101.

[20] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4):411–430, June 2000. ISSN 0893-6080. doi: 10.1016/S0893-6080(00)00026-5. URL https://www.sciencedirect.com/science/article/pii/S0893608000000265.

[21] Yu-Xiong Wang and Yu-Jin Zhang. Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE Transactions on Knowledge and Data Engineering*, 25(6):1336–1353, June 2013. ISSN 1558-2191. doi: 10.1109/TKDE.2012.51. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

[22] Cédric Févotte and Jérôme Idier. Algorithms for nonnegative matrix factorization with the β-divergence. *Neural computation*, 23(9):2421–2456, 2011.

[23] Mitul Kumar Ahirwal and Narendra Londhe. Power spectrum analysis of eeg signals for estimating visual attention. *International Journal of computer applications*, 42(15):22–25, 2012.

[24] Lionel Barnett and Anil K. Seth. The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *Journal of Neuroscience Methods*, 223:50–68, February 2014. ISSN 1872-678X. doi: 10.1016/j.jneumeth.2013.10.018.

[25] André M. Bastos and Jan-Mathijs Schoffelen. A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. *Frontiers in Systems Neuroscience*, 9, January 2016. ISSN 1662-5137. doi: 10.3389/fnsys.2015.00175. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4705224/.

[26] Karl J. Friston. Functional and Effective Connectivity: A Review. *Brain Connectivity*, 1(1):13–36, January 2011. ISSN 2158-0014. doi: 10.1089/brain.2011.0008. URL https://www.liebertpub.com/doi/abs/10.1089/brain.2011.0008. Publisher: Mary Ann Liebert, Inc., publishers.

[27] David Carlson, Lisa K. David, Neil M. Gallagher, Mai-Anh T. Vu, Matthew Shirley, Rainbo Hultman, Joyce Wang, Caley Burrus, Colleen A. McClung, Sunil Kumar, Lawrence Carin, Stephen D. Mague, and Kafui Dzirasa. Dynamically Timed Stimulation of Corticolimbic Circuitry Activates a Stress-Compensatory Pathway. *Biological Psychiatry*, June 2017. ISSN 0006-3223. doi: 10.1016/j.biopsych.2017.06.008. URL http://www.sciencedirect.com/science/article/pii/S0006322317316682.

[28] Alexander Z. Harris and Joshua A. Gordon. Long-Range Neural Synchrony in Behavior. *Annual Review of Neuroscience*, 38(1):171–194, 2015. doi: 10.1146/annurev-neuro-071714-034111. URL http://dx.doi.org/10.1146/annurev-neuro-071714-034111.

[29] Raquel Prado and Mike West. *Time Series: Modeling, Computation, and Inference*. CRC Press, May 2010. ISBN 978-1-4200-9336-0. Google-Books-ID: kUhKLLdGGZ4C.

[30] G. Tunnicliffe Wilson. A convergence theorem for spectral factorization. *Journal of Multivariate Analysis*, 8(2):222–232, June 1978. ISSN 0047-259X. doi: 10.1016/0047-259X(78)90073-8. URL https://www.sciencedirect.com/science/article/pii/0047259X78900738.

[31] Mukeshwar Dhamala, Govindan Rangarajan, and Mingzhou Ding. Estimating Granger Causality from Fourier and Wavelet Transforms of Time Series Data. *Physical Review Letters*, 100(1):018701, January 2008. doi: 10.1103/PhysRevLett.100.018701. URL https://link.aps.org/doi/10.1103/PhysRevLett.100.018701. Publisher: American Physical Society.

[32] John F. Geweke. Measures of Conditional Linear Dependence and Feedback Between Time Series. *Journal of the American Statistical Association*, 79(388):907–915, 1984. ISSN 0162-1459. doi: 10.2307/2288723. URL https://www.jstor.org/stable/2288723. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[33] Guido Nolte, Andreas Ziehe, Vadim V. Nikulin, Alois Schlögl, Nicole Krämer, Tom Brismar, and Klaus-Robert Müller. Robustly Estimating the Flow Direction of Information in Complex Physical Systems. *Physical Review Letters*, 100(23):234101, June 2008. doi: 10.1103/PhysRevLett.100.234101. URL https://link.aps.org/doi/10.1103/PhysRevLett.100.234101. Publisher: American Physical Society.

[34] Luiz A. Baccalá and Koichi Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, 84(6):463–474, May 2001. ISSN 1432-0770. doi: 10.1007/PL00007990. URL https://doi.org/10.1007/PL00007990.

[35] M. J. Kaminski and K. J. Blinowska. A new method of the description of the information flow in the brain structures. *Biological Cybernetics*, 65(3):203–210, July 1991. ISSN 1432-0770. doi: 10.1007/BF00198091. URL https://doi.org/10.1007/BF00198091.

[36] Martin Vinck, Lisanne Huurdeman, Conrado A. Bosman, Pascal Fries, Francesco P. Battaglia, Cyriel M. A. Pennartz, and Paul H. Tiesinga. How to detect the Granger-causal flow direction in the presence of additive noise? *NeuroImage*, 108:301–318, March 2015. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2014.12.017. URL https://www.sciencedirect.com/science/article/pii/S105381191401009X.

[37] Jose-Manuel Alonso, W. Martin Usrey, and R. Clay Reid. Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature*, 383(6603):815–819, October 1996. ISSN 1476-4687. doi: 10.1038/383815a0. URL https://www.nature.com/articles/383815a0. Number: 6603 Publisher: Nature Publishing Group.

[38] Terry Bossomaier, Lionel Barnett, Michael Harré, and Joseph T. Lizier. *An introduction to transfer entropy: information flow in complex systems*. Springer International Publishing, January 2016. ISBN 978-3-319-43221-2 978-3-319-43222-9. doi: 10.1007/978-3-319-43222-9. URL https://researchoutput.csu.edu.au/en/publications/an-introduction-to-transfer-entropy-information-flow-in-complex-s.

[39] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables. *Physical Review Letters*, 103(23):238701, December 2009. doi: 10.1103/PhysRevLett.103.238701. URL https://link.aps.org/doi/10.1103/PhysRevLett.103.238701. Publisher: American Physical Society.

[40] K. J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *NeuroImage*, 19(4):1273–1302, August 2003. ISSN 1053-8119. doi: 10.1016/S1053-8119(03)00202-7. URL https://www.sciencedirect.com/science/article/pii/S1053811903002027.

[41] K.J. Friston, A. Bastos, V. Litvak, K.E. Stephan, P. Fries, and R.J. Moran. DCM for complex-valued data: Cross-spectra, coherence and phase-delays. *Neuroimage*, 59(1):439–455, January 2012. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2011.07.048. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3200431/.

[42] K. E. Stephan, W. D. Penny, R. J. Moran, H. E. M. den Ouden, J. Daunizeau, and K. J. Friston. Ten simple rules for dynamic causal modeling. *NeuroImage*, 49(4):3099–3109, February 2010. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2009.11.015.

[43] Alard Roebroeck, Elia Formisano, and Rainer Goebel. Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage*, 25(1):230–242, March 2005. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2004.11.017. URL https://www.sciencedirect.com/science/article/pii/S1053811904006688.

[44] Douglas G. Bonett and Thomas A. Wright. Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika*, 65(1):23–28, March 2000. ISSN 1860-0980. doi: 10.1007/BF02294183. URL https://doi.org/10.1007/BF02294183.

[45] H. M. Abelaira, G. Z. Reus, and J. Quevedo. Animal models as tools to study the pathophysiology of depression. *Revista Brasileira de Psiquiatria*, 35:S112–S120, 2013.

[46] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu. Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis. *Neural Computation*, 21(3):793–830, March 2009. ISSN 0899-7667. doi: 10.1162/neco.2008.04-08-771. URL https://doi.org/10.1162/neco.2008.04-08-771.

[47] Rainbo Hultman, Kyle Ulrich, Benjamin D. Sachs, Cameron Blount, David E. Carlson, Nkemdilim Ndubuizu, Rosemary C. Bagot, Eric M. Parise, Mai-Anh T. Vu, Neil M. Gallagher, Joyce Wang, Alcino J. Silva, Karl Deisseroth, Stephen D. Mague, Marc G. Caron, Eric J. Nestler, Lawrence Carin, and Kafui Dzirasa. Brain-wide Electrical Spatiotemporal Dynamics Encode Depression Vulnerability. *Cell*, 173(1):166–180.e14, March 2018. ISSN 0092-8674. doi: 10.1016/j.cell.2018.02.012. URL http://www.sciencedirect.com/science/article/pii/S0092867418301569.

[48] W. Zheng and B. Lu. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175, September 2015. ISSN 1943-0604. doi: 10.1109/TAMD.2015.2431497.

[49] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, December 2000. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.290.5500.2319. URL https://science.sciencemag.org/content/290/5500/2319. Publisher: American Association for the Advancement of Science Section: Report.

[50] Steven L. Bressler, Craig G. Richter, Yonghong Chen, and Mingzhou Ding. Cortical functional network organization from autoregressive modeling of local field potential oscillations. *Statistics in Medicine*, 26(21):3875–3885, 2007. ISSN 1097-0258. doi: https://doi.org/10.1002/sim.2935. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2935. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.2935.

[51] Kyle R Ulrich, David E Carlson, Kafui Dzirasa, and Lawrence Carin. GP Kernels for Cross-Spectrum Analysis. *Advances in Neural Information Processing Systems 28*, pages 1999–2007, 2015. URL http://papers.nips.cc/paper/5966-gp-kernels-for-cross-spectrum-analysis.pdf.

[52] G. Tunnicliffe. Wilson. The Factorization of Matricial Spectral Densities. *SIAM Journal on Applied Mathematics*, 23(4):420–426, December 1972. ISSN 0036-1399. doi: 10.1137/0123044. URL https://epubs.siam.org/doi/abs/10.1137/0123044.

[53] Huifang E. Wang, Christian G. Bénar, Pascale P. Quilichini, Karl J. Friston, Viktor K. Jirsa, and Christophe Bernard. A systematic framework for functional connectivity measures. *Frontiers in Neuroscience*, 8, 2014. ISSN 1662-453X. doi: 10.3389/fnins.2014.00405. URL http://journal.frontiersin.org/article/10.3389/fnins.2014.00405/full.

[54] P. Welch. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, June 1967. ISSN 1558-2582. doi: 10.1109/TAU.1967.1161901. Conference Name: IEEE Transactions on Audio and Electroacoustics.

[55] Claude Bédard and Alain Destexhe. Macroscopic Models of Local Field Potentials and the Apparent 1/f Noise in Brain Activity. *Biophysical Journal*, 96(7):2589–2603, April 2009. ISSN 0006-3495. doi: 10.1016/j.bpj.2008.12.3951. URL https://www.sciencedirect.com/science/article/pii/S0006349509004147.