

Identification of Interacting Neural Populations: Methods and Statistical Considerations

Robert E. Kass^{1,2,3*}, Heejong Bong³, Motolani Olarinre^{1,3}, Qi Xin^{2,3}, and Konrad Urban^{2,3}

Abstract

As improved recording technologies have created new opportunities for neurophysiological investigation, emphasis has shifted from individual neurons to multiple populations that form circuits, and it has become important to provide evidence of cross-population coordinated activity. We review various methods for doing so, placing them in six major categories while avoiding technical descriptions and instead focusing on high-level motivations and concerns. Our aim is to indicate what the methods can achieve and the circumstances under which they are likely to succeed. Toward this end we include discussion of four cross-cutting issues: definition of neural populations; trial-to-trial variability and Poisson-like noise; time-varying dynamics; and causality.

Keywords: Cross-population coupling; information flow; latent drivers; statistical issues.

1 Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

2 Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA

3 Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA, USA

* email: kass@stat.cmu.edu

1 Introduction

Neural circuits are described in terms of anatomical areas and sub-areas, including, when relevant, consideration of cortical structure within and across areas (Bassett et al., 2010; Glasser et al., 2016; Goulas et al., 2018; Harris & Shepherd, 2015; Markram et al., 2015; Potjans & Diesmann, 2014). Although early investigations tended to focus on the function of neurons within a particular area or sub-area, multi-electrode data recordings have opened up new possibilities for scientific investigation (Jun et al., 2017; Raducanu et al., 2017; Rios et al., 2016; Steinmetz et al., 2021; Steinmetz et al., 2018), and emphasis has shifted from individual neurons to populations (Jiang et al., 2020; Kohn et al., 2016; Ni et al., 2018; Saxena & Cunningham, 2019). Thus, brain physiology is widely conceptualized in terms of interacting neural populations, and it has become important to be able to provide statistical evidence of cross-population neural activity. Here we review methods for identifying interactions across two or more populations, which are typically populations residing in distinct areas or sub-areas. We concentrate on spike trains and local field potentials (LFPs), which represent local bulk population activity (Buzsáki et al., 2012; Einevoll et al., 2013; Luo, 2020, p. 588), but many of the techniques are used (or could be used) in optical imaging or neuroimaging. Throughout, we discuss stereotypical, repeated-trial experiments, though the methods may also apply to peri-event time intervals in experiments involving freely-behaving animals (Datta et al., 2019; Juavinett et al., 2019; Pereira et al., 2020).

Several useful overviews have appeared already, each with its own focus (Kang & Druckmann, 2020; Keeley et al., 2020; Kohn et al., 2020; Semedo et al., 2020). Such articles serve differently the diverse audiences in neurophysiology, as authors strike some balance of emphasis between conceptual underpinnings and technical description. We have chosen to organize this article around a small number of analytical frameworks that are useful in examining the rich data sources now available, but while we classify methods for identifying cross-population interactions, our main purpose is to highlight the relevance of several fundamental statistical issues. Every method we mention assumes that neural populations are part of a circuit that accomplishes something; yet each time a circuit does so, measured values are different. These two aspects of experimental reality are often described in terms of “signal” and “noise,” with circuit accomplishment being portrayed evocatively using concepts such as “coding” and “information” (historical remarks may be found in Kass et al., 2018, Section 1). As a consequence of physiological variation, conclusions are based on some definition of consistency in the patterns of neural activity across repetitions, which requires both a notion that the patterns are relevant (e.g., to the flow of information) and a demonstration that the consistency passes a statistical standard for evidence. Our discussion aims to guide thinking about the linked criteria of physiological relevance and statistical standards.

The many ways that noisy data can provide evidence about the function of neural circuits is likely to be of interest to readers with a wide range of backgrounds; this includes a range of comfort with mathematically technical material. Although targeting an audience with a specific assumed level of familiarity with mathematical and statistical concepts would make exposition much easier, we believe our overview can be of use to diverse readers. By emphasizing high-level motivations and concerns, we hope to indicate what the methods can achieve, and the circumstances under which they are likely to succeed, so that experimental investigators can identify scientific questions

where these tools can be used to advantage. To maintain wide accessibility we have chosen not to use mathematical equations. Without equations and more extensive explanations of procedures, those having technical backgrounds may find our very brief summaries inadequate, but rigorous definitions can be retrieved quickly from cited references and other sources.

As an example, some of our own recent research began with the assumption, made by numerous studies, that behaviorally-relevant information is transmitted across parts of the brain through transient bursts of activity in neural populations. If population bursts are indeed important, their timing should reveal coordinated activity: on a trial-by-trial basis, the time of a burst in one population should be related to the time of a corresponding burst in a downstream population. Chen et al. (2022) computed, across multiple simultaneously-recorded populations, on a trial-by-trial basis, the times at which the population maximal firing rates occurred and found these trial-varying time values to be sensitive indicators of interaction, as shown in Figure 1. Although we will say more about this work in Section 3.6, we will not describe it in detail. Our aim, instead, is to direct readers to the underlying presumption that population bursts represent propagation of information; the methods’s suitability for describing timing with millisecond precision; its use of a broadly applicable statistical framework involving latent variables; and three key components of the approach that apparently drive its ability to find strong representations of interaction.

We emphasize these sorts of considerations by sandwiching our list of methodological approaches, in Section 3, between general statistical discussions in Sections 2 and 4. In Section 2 we present brief preliminary comments about statistical models and their parameters, including latent variables, which appear in many models of cross-population coupling to represent idealized, shared “drivers” of activity (illustrated in Figure 5). In Section 3, we categorize data analytic methods used to describe interacting populations according to six major frameworks. In summarizing them we aim to capture the main ideas behind their use so that their purpose in this context can be grasped, in rough terms, by non-experts. The first four general frameworks are linear multivariate analysis; graphical models; autoregressive models, especially for Granger causality; and latent dynamic models. Our last two general categories of methods are more-specialized approaches based on frequency analysis and point processes, the former being applied mainly to LFPs (and similar continuously varying quantities measured repeatedly across time) and the latter being suitable for modeling spike trains. In Section 4 we highlight a series of important issues that every investigator should consider, starting with the definition of a neural population, which involves neural diversity, combined with haphazard neural sampling, and the redundancy of neural signals. There are also statistical difficulties associated with large numbers of signals. We then discuss trial-to-trial variability, where Poisson-like variation can easily obscure interesting relationships, while statistical techniques can uncover them (as Figures 1B,C and 2 illustrate). A ubiquitous concern is time-varying dynamics, which can invalidate many common methods in principle, suggesting care in practice. We end our call to attention with remarks on causality, and share a few closing thoughts in Section 5.

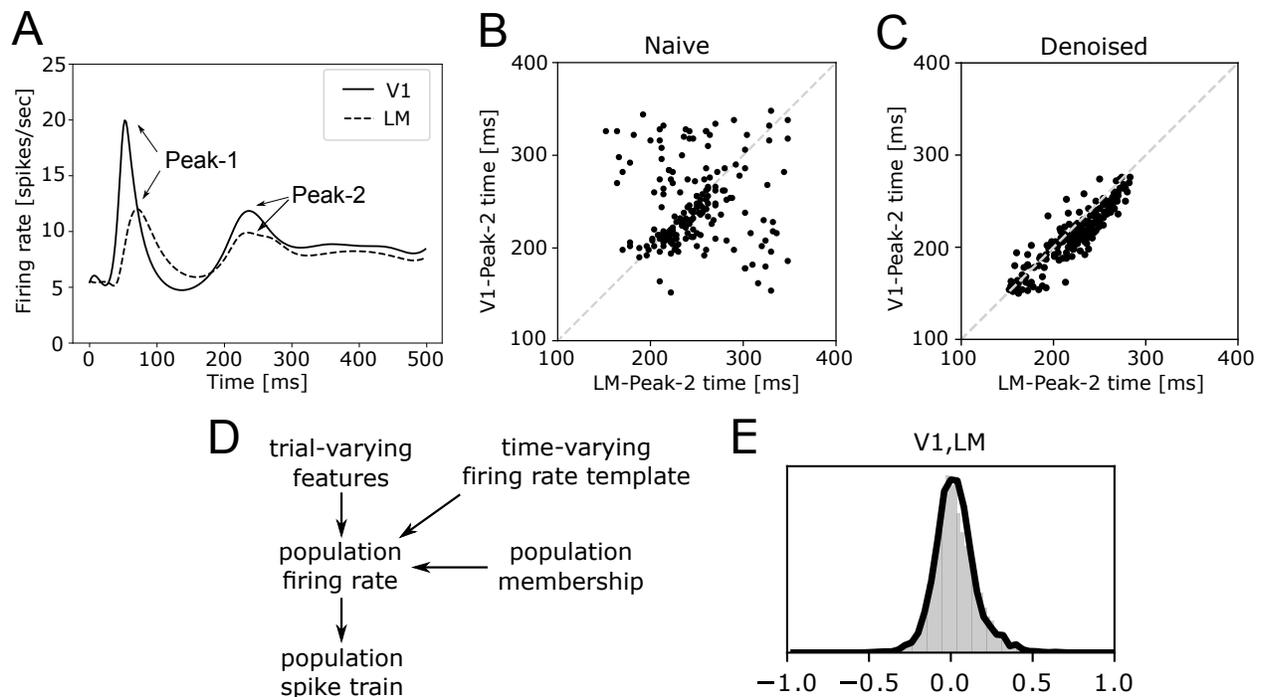


Figure 1: Cross-area coupling of population peak firing rate timing. **A** Smoothed population PSTHs in response to full-field drifting gratings for mouse visual areas V1 and LM. The two population PSTHs have similar features: the first peak (Peak-1) appears early, soon after stimulus onset while the second peak (Peak-2) appears much later. **B** Plot of the time of Peak-2 in V1 against the time of Peak-2 in LM based on smoothing the two PSTHs on a trial-by-trial basis (naïve method). Each data point represents the pair of peak times on a particular trial. There is no visible relationship (and no statistically significant correlation). **C** Results plotted as in **B** except now based on the model of Chen et al. (2022). In this plot the correlation is strong ($.87 \pm .05$; 95% confidence interval), with Peak-2 in LM tending to occur later than Peak-2 in V1. In analysis not shown here Chen et al. found the time lag from V1 to LM to be, on average, 9.6 milliseconds with standard deviation across trials less than 2 ms; this small variation in lag time contrasts with the large trial-to-trial variation in Peak-2 timing for both visual areas, seen in the plot; standard deviations across trials were roughly 30 ms. **D** Schematic of the statistical model, discussed in Section 3.6. **E** Spike count correlation histogram when one neuron is in V1 and the other neuron is in LM, which is mildly skewed toward positive correlations. The model prediction (dark line) fits the histogram well, even though the only sources of correlation in the model are the trial-by-trial shifts in timing of the population peaks and trial-by-trial changes in overall firing rate of each population; the model effectively assumed the neurons were otherwise independent. Figure modified from Chen et al. (2022).

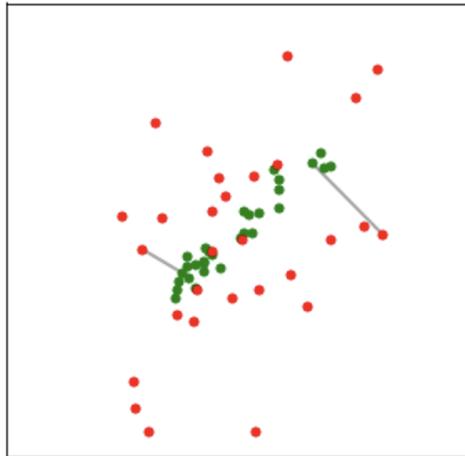


Figure 2: **Attenuation of correlation.** Green dots show data sampled from a distribution for which there is relatively high correlation between two variables (confer Figure 1C). Red dots are generated by adding noise to each of the variables represented by green dots (conceptualized as similar to Figure 1B, but less extreme). Grey lines show the mapping from the original green dot to the corresponding red dot, once noise has been added. The correlation between green dots is much higher than the correlation between red dots.

2 Statistical Models

2.1 Models and parameters

It is not too hard to summarize several important teachings from the field of statistics, as in the *Ten Simple Rules* of Kass et al. (2016) or the 14 page explanation of the “statistical paradigm” in Kass, Eden, and Brown (2014, Chapter 1). Arguably the most fundamental quality of data analytic methods informed by the field of statistics is that they are defined and studied by introducing statistical models. That is, rather than summarizing data based solely on intuition, or heuristics, according to the statistical paradigm one begins by defining a model that uses probability distributions to describe the regularity and variability in the data. This may be written, in schematic short-hand as

$$\text{observation} \stackrel{\text{combines}}{\equiv} \text{signal} \stackrel{\text{with}}{+} \text{noise} \quad (1)$$

where the way signal and noise are combined, through the statistical model, may be either simple or complicated. Importantly, when a statistical method is derived by assuming a particular model, it can be evaluated using a variety of other, different models. This is a reliable way to check the robustness of a method against departures from assumptions. In the machine learning literature, statistical models are often called “generative models” (indicating that they can generate artificial data presumed to be similar to real data).

The probability distributions used in statistical models are specified by probability density functions, which depend on theoretical quantities, known as parameters. For example, a normal (Gaussian) distribution has a mean and standard deviation as parameters. When there are only a few

parameters, a model may be relatively easy to understand but it restricts possibilities for data variation. On the other hand, when a very large number of parameters is used (in theory, even infinitely many) models can be very flexible.

It is important to distinguish between data-based quantities and their model-based theoretical counterparts. For example, a sample mean is computed from data, and its model-based counterpart could be the parameter representing the mean of a normal distribution; the latter is often called a population mean, but here we will use the term *theoretical* instead of “population” for reasons articulated in Kass et al. (2014, Section 3.2.2), and at greater length in Kass (2011). Statistically, the sample mean would be considered an estimate of the theoretical mean. Many neurophysiologists become comfortable with data-based quantities, such as a firing rate or a spike count correlation, without considering a statistical model. This is fine until the variation in the data summary is sufficiently large that it creates non-trivial uncertainty—but immediately a question arises: uncertainty about what? The answer that has remained useful since its introduction by R.A. Fisher one hundred years ago invokes a theoretical counterpart to the data summary, and it is defined in terms of a statistical model. Evaluation of the strength of evidence, usually based on confidence intervals or significance tests, relies on this fundamental conception. Concerns about the validity of some statistical manipulation of the data can only be addressed by an examination of the procedure under conditions specified by a model together with a data-informed judgment of the suitability of that model.

2.2 Dimensionality and multivariate models

Because parameters provide numerical labels to identify specific members of a family of probability distributions that make up a model, there is (for most standard models) a well-defined notion of the dimension of the model: it is the number of freely ranging univariate parameters used to identify a member of the family (Kass and Vos, 1997, Theorem 2.3.1 and Section 7.3; subtleties can arise, Hodges and Sargent, 2001). Similarly, sets of random variables that represent data have an associated dimension. Activity across n recordings would, unless somehow constrained, typically be assumed to vary freely in an n -dimensional space, that is, a set of values represented by a vector with n coordinates. When a variable X_1 can be written in terms of two other variables X_2 and X_3 , the collection (X_1, X_2, X_3) is not freely ranging in a 3-dimensional space; instead, it has 2 dimensions (at most; if the three variables are multiples of each other then there is only 1 dimension of variation). This is the source of the terminology “degrees of freedom” in statistics. That kind of mathematical redundancy is not difficult to accommodate.

The situation becomes more subtle when statistical redundancy is considered. Statistical redundancy would occur if a variable X_1 (a random variable, following a probability distribution) were highly correlated with some combination of variables X_2 and X_3 . In this case, it is common to introduce an empirical rule that uses some measure of correlation together with a threshold which, when exceeded, lumps the high correlation case together with exact mathematical dependence by deeming the collection (X_1, X_2, X_3) again to be lower-dimensional. Similar reasoning applies when there are n variables having variation concentrated predominantly in a space of lower di-

mension k . Statistical redundancy arises in neural activity when the n recordings can not vary separately in response to a stimulus or in contributing to a behavior. This is analogous to finger kinematic redundancies during reaching tasks, where hand physiology constrains the fingers so that they can not move independently (e.g., Castellanos et al., 2015; Mason et al., 2001). As the number of recordings increases, some degree of statistical redundancy becomes likely. In the case of individual neurons, this is likely due to redundancy of inputs across a population combined with limited sampling of stimuli or behaviors in the experiment. For LFPs there is also the well-documented spatial correlation in neighboring recordings (Einevoll et al., 2013; Lindén et al., 2011). Consideration of dimensionality has been used to interpret population responses (Babadi & Sompolinsky, 2014; Cunningham & Yu, 2014; Fusi et al., 2016; Gao & Ganguli, 2015) and cross-population communication (Duncker & Sahani, 2021; Shenoy & Kao, 2021, and references therein).

The formal framework for analyzing the covariation among n variables begins with their variance matrix, which is defined by the n standard deviations together with the correlations of all pairs of variables. The most common method of dimensionality reduction involves a revealing re-expression of that matrix: the coordinates representing the n variables are rotated to define new, uncorrelated variables; the coordinate direction corresponding to the new variable that has the largest standard deviation represents the direction of maximal variance among linear combinations of the original variables; the representation of data along this coordinate is called the first principal component; and the complete decomposition using multiple components becomes Principal Component Analysis (PCA). Factor analysis relies on similar re-expressions (it is illustrated conceptually in Figure 5). It is easy to find good explanations of these methods, and basic intuitions appear in Cunningham and Yu (2014). As we discuss in Section 4.1.2, redundancy tends to become more pronounced when a variance matrix is estimated from data and n is large.

To highlight and clarify the distinction between pairwise and multivariate dependence, let us step through the case of 3 variables X_1, X_2, X_3 where it is possible to have the correlations of X_3 with each of X_1 and X_2 “explain” a correlation between X_1 and X_2 in the sense that if X_3 were held constant the correlation between X_1 and X_2 would disappear, that is, the partial correlation of X_1 and X_2 after conditioning on the value of X_3 would be zero. Panel C of Figure 3 illustrates this situation, while panel D adds arrows to show what is usually called a spurious correlation between X_1 and X_2 , due to the confounding variable X_3 . In the case of Gaussian (normal) random variables, just as X_1 and X_2 are independent when their correlation is zero, X_1 and X_2 become *conditionally independent* after conditioning on X_3 when their partial correlation given X_3 is zero (panel C). When the variables X_1, X_2, X_3 represent summaries of neural activity in three areas, if we were to consider all other areas to be irrelevant, this would indicate that an observed interaction of X_1 and X_2 (in the sense of correlated activity) would be due entirely to the separate interactions of X_1 with X_3 and X_2 with X_3 . More generally, Figure 3C represents the simplest non-trivial *probabilistic graphical model*, where the circles representing variables are called “nodes” and the lines connecting them, which indicate dependence, are called “edges.” In general, for a probabilistic graphical model, the absence of an edge between two nodes indicates independence of the two corresponding variables after conditioning on all the rest of the variables (Murphy, 2022, pp. 429-440, 939-940). As reviewed in Section 3.2, graphical models can represent interesting dependence relationships across multiple populations. They are also a starting point for network analysis.

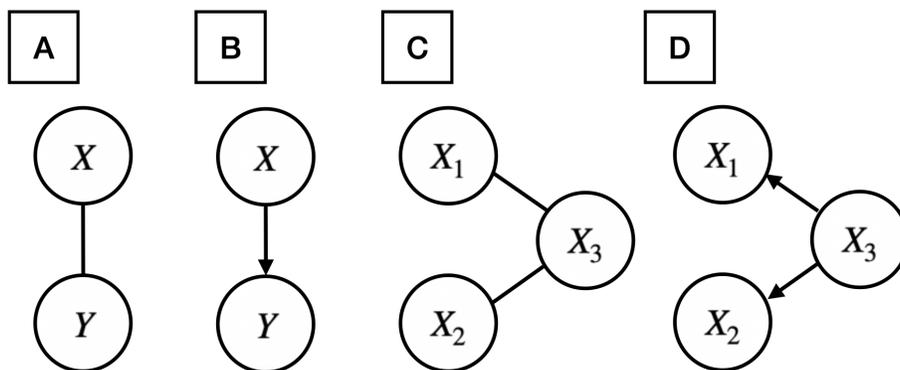


Figure 3: **Correlative and predictive graphs.** **A** The random variables X and Y are correlative, in the sense of the ordinary (Pearson) correlation or some other measure of association. **B** The variable X can be used to predict Y . **C** In general, when a pair of nodes representing two random variables are connected by an edge (a line), the two random variables remain dependent after conditioning on all other random variables. Here, X_1 and X_2 are conditionally independent given X_3 . If these variables represented activity in different brain areas, and if we were to assume that no other brain areas are relevant, then, based on these activity variables, any apparent interaction between areas 1 and 2 would be due to the separate interactions of area 1 with area 3 and area 2 with area 3. **D** Variable X_3 predicts each of X_1 and X_2 , and these predictive relationships could result in an observed “spurious” association between X_1 and X_2 . Although pictorial diagrams such as in panels B and D are commonly used in neurophysiology (where the variables would be replaced by brain locations) to suggest explanatory relationships, they are infrequently based on causal evidence (in the sense specified by statistical causal inference) and are more commonly predictive. In general, each variable depicted in the figure could itself be a random vector rather than a scalar random variable.

The graphical model in Figure 3 is *undirected*. When the edges are replaced by arrows the graph becomes *directed*. Directed graphs are highly suggestive, but converting an undirected relationship based on correlation structure to a directed relationship requires additional information. In fact, it is not hard to create an example in which Figure 3C represents the correlation structure but, rather than the correlation of X_1 and X_2 being due to X_3 the relationship is reversed: the combination of X_1 and X_2 produces X_3 , resulting in a graph as in panel D with the arrows pointing toward, rather than away from X_3 . In neuroscientific applications, area-to-area arrows are generally based on some assessment of association (correlation and its generalizations) supplemented by consideration of timing. In most cases, however, arrows representing timing should be interpreted as predictive rather than causal, a point mentioned in Section 3.3 and discussed in Section 4.4.

Although only under rare circumstances can causal claims be justified based on partial correlation (and analogous assessments of association) in conjunction with timing information, they remain useful more broadly: when the partial correlation of X_1 and X_2 given X_3 is much smaller than the original correlation of X_1 and X_2 (before conditioning on X_3), the interpretation is that the separate interactions reflected in the statistical associations of X_3 with X_1 and X_3 with X_2 contribute substantially to the interaction reflected in the observed association of X_1 with X_2 .

Chen et al. (2022) analyzed data from visual area AL, in addition to V1 and LM as shown in Figure 1. The three variables denoted here as X_1, X_2, X_3 , became summaries of activity in three populations for Chen et al. This is one common strategy, and it often employs latent variables, discussed in Section 2.4. An alternative is to consider the predictability of each recording (spike train or LFP) from the many recordings in the other populations. This becomes a regression approach, but it is usually carried out using some form of generalized regression (illustrated conceptually in Figure 4). A simpler possibility, for two populations, is to examine pairs of individual recordings, one in each population, for many pairs. The methods we review include examples of all three approaches.

Concerning terminology, it can be cumbersome to distinguish between scalars and vectors. Thus, in some cases, when we speak about a variable, as in referencing a node in a graphical model like those shown in Figure 3, the “variable” will actually be a vector. We trust this will not cause confusion.

2.3 Stochastic processes

In probability theory, a stochastic process (in time) is a collection of variables X_t , with one variable for each value of time t . When t takes values $1, 2, 3, \dots$ the process is called a time series (Section 3.3). Thus, LFPs are analyzed as time series. For point processes (Section 3.6), in theory time is continuous, but in practice spike trains are observed at discrete time values, with resolution such as 1 millisecond, so that the spike train data form binary time series (spike or no spike in each time bin). Similarly, in theory, Gaussian processes (Section 3.4.2) are continuous, but they are applied to analyze time series of spike counts. A technical point is that a probability distribution for a stochastic process must specify all possible multivariate joint distributions, which leads to simplifying assumptions.

2.4 Latent variables

Especially important in population analysis is the idea that activity within a population may have a commonality conceptualized as a “hidden driver,” often considered to represent a “hidden state.” These are known as latent variables. The adjectives “latent” and “hidden” are used because latent variables do not correspond to any measured quantities; rather, they are theoretical abstractions introduced to construct a concise and intuitive model of covariation (Bollen, 2002; Kass et al., 2014, Section 16.2; Murphy, 2022, pp. 337-420). Latent variables may appear in models as parameters, or their effects may depend on parameters through some mathematical relationship.

2.4.1 Theoretical firing rates as latent variables

In spike train analysis, spike counts can be converted into observed firing rates. Within a statistical model for neuron spiking, corresponding to an observed firing rate there is a theoretical firing rate, determined by one or more parameters; typically, this theoretical firing rate, and its dependence on experimental variables characterizing a stimulus or behavior, would be the focus of attention. When the fluctuation in a theoretical firing rate is considered to be stochastic, depending on some set of random variables (representing the variation in stimuli or behaviors, etc.), it becomes a latent variable. A simple example is the use of “high” and “low” firing rates in a single-neuron model to capture bursting (Tokdar et al., 2009, which is Example 16.3 in Kass et al., 2014). Studies of neural variability sometimes examine theoretical firing rates instead of spike counts because the latter includes Poisson-like noise. The Poisson-like noise can be effectively reduced through the use of a statistical model that separates the theoretical firing rate from the Poisson-like noise and reports estimates of theoretical quantities, such as firing rate correlation instead of spike count correlation (Goris et al., 2014; Vinci et al., 2016). This use of firing rates typically results in much higher values of correlation between pairs of neurons, which in statistics is called “correction for attenuation” (Behseta et al., 2009). We return to Poisson-like noise in Sections 3.6 and 4.3, but the idea behind attenuation of correlation is illustrated in Figure 2 and Figure 1C illustrates the benefit of removing Poisson-like noise for identifying cross-area interactions.

2.4.2 Latent drivers and hierarchical interactions

A prominent use of latent variable models is to provide a parsimonious representation of the redundancy in neural signals mentioned in Section 2.1 by restricting the variation of interest to a subspace of lower dimension than that of the observations themselves. Such dimensionality reduction is the purpose of factor analysis (see Figure 5), where every measured variable is assumed to arise as a linear combination of some comparatively small number of latent variables (assumed independent and standard normally distributed), plus noise. In addition, PCA, and the closely-related canonical correlation analysis (CCA, see Section 3.1), can be derived as estimation procedures for latent variable models known as probabilistic PCA and CCA (pPCA, pCCA; Murphy, 2022, Sections 20.2.2 and 20.2.8.3) The resulting latent variables become conceptualized as latent drivers of neural activity. As we review in Sections 3.4.1 and 3.4.2, cross-area analysis may be simplified by assuming one or more drivers of activity in each area: although the number of distinct spike train or LFP recordings may be large, the number of separate populations (typically from anatomically distinct areas), is much smaller, which suggests identification of cross-population interactions *per se* ought to be more manageable.

To describe situations in which interactions among neurons, or field potentials, might occur either within a particular population or across two or more populations, it may be helpful to use the term *hierarchical interactions*. Here we refer to statistical behavior rather than the widely-discussed anatomical and physiological hierarchical structure of the cortex (Felleman & Van Essen, 1991; Harris & Shepherd, 2015; Hilgetag & Goulas, 2020; Vezoli et al., 2021). In this conception, cross-population interactions are summarized with low-dimensional latent variables (latent vec-

tors), while within-population activity is determined partly by one or more relevant latent variables and partly by other factors. Two approaches that have been used to capture hierarchical interactions are Gaussian Process Factor Analysis (GPFA), which is discussed in Section 3.4.2, and analysis using Bayesian hierarchical models (Gelman et al., 2013; Hu et al., 2020; Lee and Mumford, 2003; additional references in Kass and Steffey, 1989 and Kass et al., 2014, Chapter 16). An example of the latter is the model in Chen et al., 2022, which produced Figure 1. The general idea, however, is not tied to a particular methodology.

Regardless of the modeling details, it is worth noting that when cross-population interactions are the focus of interest, parameters representing within-population interactions may be treated differently than parameters that identify cross-population interactions. In the statistical literature, necessary but uninteresting parameters are often called “nuisance parameters” (Schervish, 2012, Chapter 4; Van der Vaart, 2000, Chapter 25). In the case of a normal distribution, interest typically focuses on the mean, with the standard deviation becoming a nuisance parameter. This piece of terminology is helpful conceptually, and in practice statistical methods focused on parameters of interest are, in some circumstances, comparatively insensitive to errors in identifying values of nuisance parameters. In cross-population studies, parameters controlling variation *within* populations (as opposed to *across* populations) would typically be considered nuisance parameters. For example, in their assessment of beta-frequency amplitude coupling between PFC and V4 during a visual memory task, Bong et al. (2020, described briefly in Section 3.5) treated the correlation structure among LFPs within each area as a high-dimensional nuisance parameter: even though the estimate of a general correlation structure among nearly 100 LFPs can not be expected to be highly accurate, probable inaccuracies were considered unlikely to undermine the evaluation of cross-population coupling.

3 Analytic Frameworks

The methods we describe may be considered, interpretively, to be ways of arriving at pictorial representations like those in Figure 3. One broad distinction is between directed and undirected graphs (B,D versus A,C in Figure 3). The simplest case, in Figure 3A,B, distinguishes regression from correlation, with X being used to predict Y in Figure 3B. In fact, in this context, X variables are often called “predictors.” This choice of words suggests that X precedes Y , and regression across time, carefully defined, can give directional arrows in the sense of prediction (though the “predictor” terminology is used even in cases that do not involve time to indicate that the X variables could be used predictively). Figure 4 provides a more detailed set of diagrams for regression.

In thinking about the many analytical approaches that can provide evidence for either correlative or predictive statistical dependence, it is important to remember that there are many ways multiple variables might interact, and as the number of variables grows it quickly becomes impossible to learn solely from data an accurate probability model that describes all the potentially complicated interactions. To make progress, some simplifying assumption is necessary. In the most common statistical contexts it is assumed that all the variables are independent of each other: in this case, the probability density function for all of them together becomes the product of the individual-variable

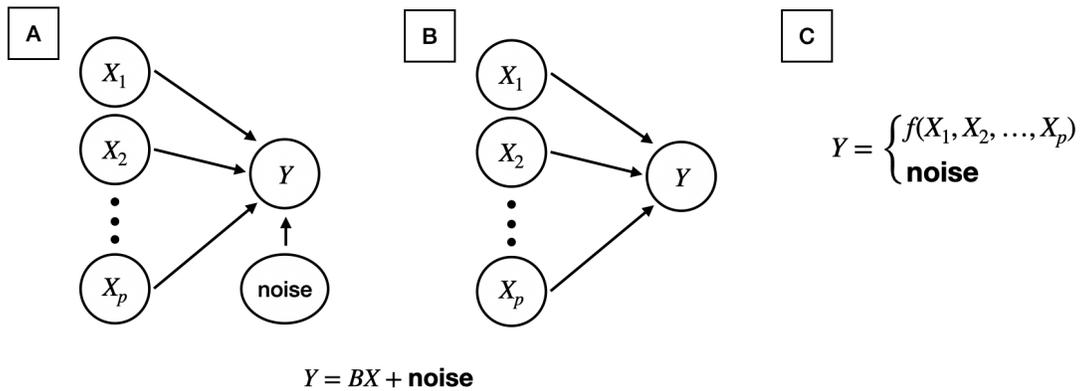


Figure 4: **Regression models.** **A,B** Two pictorial representations of the linear regression model, having the equation entered below the two diagrams. In panel A the noise variable is entered explicitly, but diagrams like that in B are often used, with the noise entering implicitly. Both diagrams show the dependence of Y on the set of variables X_1, X_2, \dots, X_p . Note that panel B would be considered the same as Figure 3B if X in that figure were the vector $X = (X_1, X_2, \dots, X_p)$ used here. **C** Panel C summarizes generalized regression by emphasizing that a function of the X variables is combined, somehow, with noise to produce Y . In the linear regression model noise is added to the linear relationship that determines the regularity in Y , i.e., the “signal.” In panel C the combination can be nonlinear and non-additive, but a probability distribution represents variation that would be considered noise. The diagram in panel B may also be used for the generalized regression case when the arrows indicate dependence but need not represent linear dependence.

probability density functions. That, however, does not allow for dependence. An intermediate simplification is to assume some form of conditional independence, as illustrated in Figure 3C. When graphs are used to supply dependence structure (as in Figure 3C), they define *probabilistic graphical models*. The resulting probability density function, for all the variables together, can then be factored into simpler components based on conditional independence, which makes computation and statistical inference much more manageable.

The methods we summarize make use of these considerations: they are based on statistical models that define correlative or predictive relationships by imposing structure that often makes use of conditional independence (either assuming it explicitly or implicitly, or by learning it from data). Our decomposition into six categories reflects in part the structure of the data, and in part the relationships a method seeks to uncover. Autoregressive models, latent dynamic models, and frequency analysis (our third, fourth, and fifth categories) are all relevant to time series, i.e., substantial numbers of repeated measurements made at different times, sequentially, such as every millisecond (LFPs are time series). Point processes (the sixth category) are specific to spike trains, and are especially valuable when precise timing of changes in firing rate could be important, so that counts in wide time bins might miss interesting effects. The first collection of methods we describe fall under the heading of linear multivariate analysis, which (usually without the modifier “linear” we have inserted for emphasis) is a branch of statistics first developed before high-speed computers, when it was essential to have mathematically tractable solutions to problems. In addition to remaining relevant, it serves to introduce some more general ideas. Graphical models (the second category) could be considered part of multivariate analysis but, importantly, these models can be applied to many kinds of data, including those beyond the scope originally envisioned for multivariate analysis.

Within each category we will list several main approaches and tools, but can describe them here only very briefly. Some of the different methods may have different purposes as, we hope, will be clear. It is very common, however, to find several different statistical methods that claim to solve the same underlying problem because they impose different structural assumptions, as we will indicate; the choice among the alternatives becomes a judgment of which structures seem most relevant, or most intuitive, or most convenient in a given context. Our primary goal is to provide an organizational and conceptual view of the data-analytic questions being answered by available methodologies (see Rule 1 of Kass et al., 2016).

3.1 Linear multivariate analysis

When activity is recorded by multiple electrodes (e.g., yielding multiple spike counts) in two areas, an initial question is whether an increase in activity in one area tends to correspond to an increase in activity in the other. The simplest assessment of covariation across recordings, such as spike counts, from only two electrodes, one in each area, is the ordinary (Pearson) correlation. This is depicted in Figure 3A. To summarize the covariation across many recordings, a standard approach represents the variables in one area by a vector X and those in the other by a vector Y and then finds the linear combinations of components of these vectors that have the greatest correlation, known

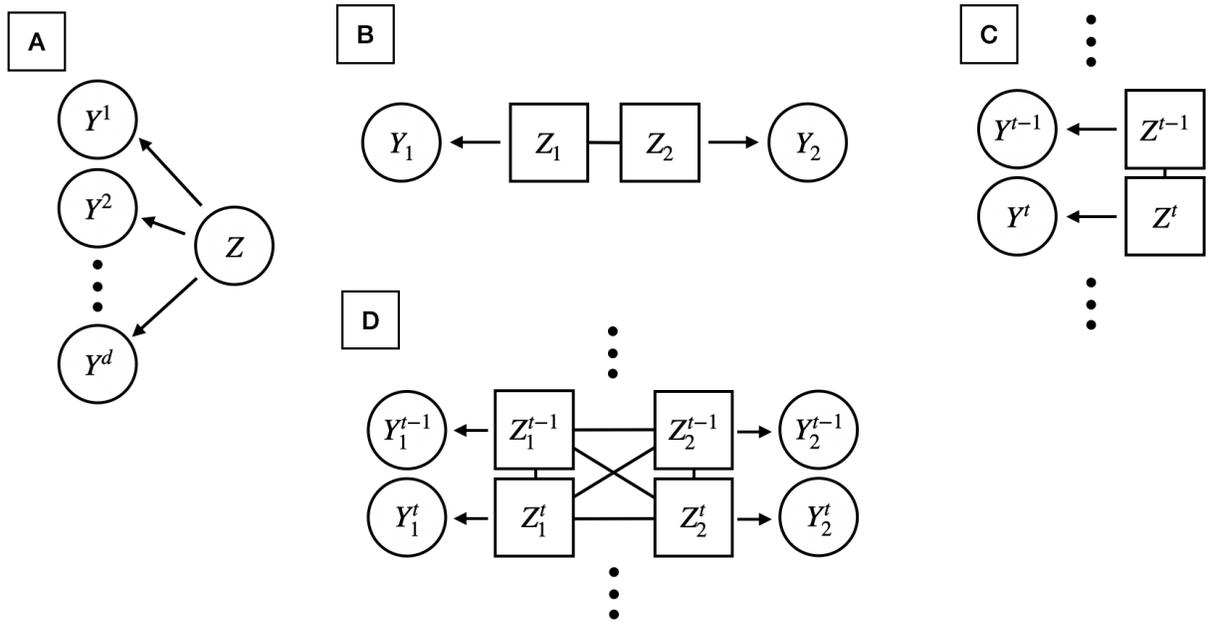


Figure 5: **Latent factor models.** **A** The latent variable Z affects each of the scalar components of $Y = (Y^1, Y^2, \dots, Y^d)$. In a one-dimensional factor analysis model each component of Y is a scalar multiple (often called a weight) of the scalar random variable Z (which is normally distributed), and noise is added, independently, to each component. More generally, factor analysis takes Z to be a random vector. **B** In pCCA (in the version used by Bong et al., 2023) a pair of correlated factor variables Z_1 and Z_2 affect the respective vectors Y_1 and Y_2 as in factor analysis; the estimated correlation of Z_1 and Z_2 is equal to the canonical correlation. **C** In GPFA a latent time series of multivariate Gaussian processes $Z^1, Z^2, \dots, Z^{t-1}, Z^t, \dots$ affects the vector time series $Y^1, Y^2, \dots, Y^{t-1}, Y^t, \dots$, with each Z^t mapping to Y^t as in factor analysis, independently and using the same weights across time. The triple dots are used to indicate the time series nature of the variables. **D** Bong *et al.* extended pCCA (panel B) to time series. The diagram takes a pair of diagrams as in C and allows the latent variables to be correlated, as in B, both within and across time series at various time lags.

as the canonical correlation; the vectors formed by the linear combination coefficients are the canonical vectors. When the canonical correlation is sufficiently large (and statistically significant) this can give a definition for the connecting line in Figure 3A when X and Y are vectors rather than scalars. The procedure is elaborated by removing from each of X and Y the variation due to their respective canonical vectors (by orthogonalization) and then treating the pair of residual variation vectors as before yielding a second correlation, which becomes the second canonical correlation, and a second canonical vector that is orthogonal to the first. The process may be repeated some number of times to produce Canonical Correlation Analysis (CCA). Semedo et al. (2022) used CCA to analyze interactions between both V1 and V2 and V1 and V4, based on spike counts, applying time delays between areas to draw conclusions about feedforward and feedback interaction. When applied at multiple time points, the linear combinations in CCA, determining the relative weights given to particular neurons, typically vary from time point to time point, which seems undesirable. Rodu et al. (2018) developed a windowing method to stabilize the solutions; they applied it to describe time-varying interactions, including time lags, between PFC and hippocampus during a memory task based on multiple LFPs recorded from each area. Multiset CCA (Kettenring, 1971) extends CCA to multiple groups (corresponding to multiple areas).

An alternative to CCA, known as reduced-rank regression (RRR), treats the two areas asymmetrically, labelling the activity in the two areas by Y and X and regressing Y on a dimension-reduced version of X . Ordinary least-squares regression finds the closest representation of Y within the space defined by linear combinations of the variables making up X (closest in the Euclidean sense of least squares). RRR takes the additional step of requiring the solution (defined by the X variables) to be of lower dimension. Semedo et al. (2019) used RRR to predict the activity of V2 neurons from a low-dimensional representation of activity among V1 neurons, which they called a “communication subspace” (confer Semedo et al., 2020). Partial Least Squares (PLS; Murphy, 2022, Section 20.2.8.2) is similar to RRR, but in determining a low-dimensional version of X to predict Y it takes into account the variability of X . Ames and Churchland (2019) suggested that PLS is a suitable tool for analyzing the degree to which contralateral and ipsilateral brain activity predict arm motions.

In applying these linear multivariate techniques, a concern, ubiquitous in statistics, is the inherent arbitrariness in the choice of a criterion that produces a method. In the case of canonical correlation, the maximization of correlation is clearly simple and intuitive, but should not be considered compelling. On the other hand, for the communal enterprise of science, there are big advantages to using well-established methods. Similarly, although linearity is, in principle, restrictive, the noisiness of the data together with modest sample sizes often precludes identification of nonlinear relationships; and even when some nonlinearity is detected, its effect on results may not change conclusions.

A more modern development represents classical methods such as PCA and CCA as latent variable models, as mentioned in Section 2.4, and then goes on to show that these alternative versions can have practical advantages. Murphy (2022, Section 20.2) uses the model-based framework to provide a unified treatment of many methods, including PCA, CCA, factor analysis, PLS, and a variety of nonlinear extensions. As described in Section 3.5 (see Figure 5D), Bong et al. (2023) used such a model-based framework to summarize interactions across multiple time series of LFPs

from PFC and V4.

3.2 Graphical models

Graphical models represent relationships among variables using nodes connected by edges. When such diagrams represent relationships according to *probabilistic* graphical models, the absence of an edge between two nodes corresponds to conditional independence (of the variables for those two nodes) given all the variables represented by all the other nodes. The simplest case was discussed with reference to Figure 3C in Section 2.2; as explained there, for Gaussian graphical models the absence of an edge corresponds to zero partial correlation.

When constructing a graphical model from data, it is important to remember that it is essentially impossible to distinguish a zero partial correlation from one that is very close to zero; the situation is analogous for other probabilistic graphical models. Thus, in employing a graphical model there is an implicit assumption that distinctions between small and zero partial correlations are not scientifically meaningful for a specified purpose, a point we repeat in Section 4.4. With this reality in mind, the construction of a Gaussian graphical model may be considered a way of estimating a variance matrix. When the matrix is large it is difficult to estimate accurately (because of the cumulative effects from inaccurate estimation of large numbers of correlations), and setting many partial correlations to zero (which amounts to determining a graphical model) effectively removes many of the parameters and can stabilize estimation. Vinci et al. (2018) combined a technique for high-dimensional problems, L1 regularization, discussed in Section 4.1.2, with correction for attenuation, discussed in Section 2.4.1, to analyze spike counts from V4 and PFC during an attention task. As discussed in Section 4.1.2, Vinci et al. modified the usual L1 regularization procedure for variance matrix estimation, known as graphical Lasso. Not only did this produce a striking improvement in performance, shown in Figure 6A, but it also produced some interesting results shown in Figure 6B: firing rates of neurons within V4 or PFC that were predicted well by the activity within the same area (V4 or PFC) were also predicted well (though not quite as well) from the activity within the other area; the effect was strongest for the “attend out” condition, in which the subject focused attention away from the V4 receptive field.

In analyzing recordings across many populations, each node might signify an individual recording, or a population summary of them, or a latent variable representing a population summary. For example, in their analysis of phase coupling of beta oscillations recorded from PFC and the hippocampal areas dentate gyrus (DG), subicullum (Sub), and CA3 during a memory retrieval task, Klein et al. (2020) presented two different but closely related graphical models based on 24 phase angles across the four areas, as shown in Figure 7. The method was based on a multivariate assessment of phase locking, which is discussed in Section 3.5. In Figure 7 each node represents phase activity in a particular area and the edges were based on a global significance test that could get evidence of area-to-area phase locking. In Figure 7B each node represents the phase from a single LFP, for a single electrode. Both kinds of graphs supported the conclusion that PFC had coordinated oscillatory activity with the output areas Sub and CA3, while the apparent phase coupling between PFC and DG was due to connections of each of these areas with Sub and CA3.

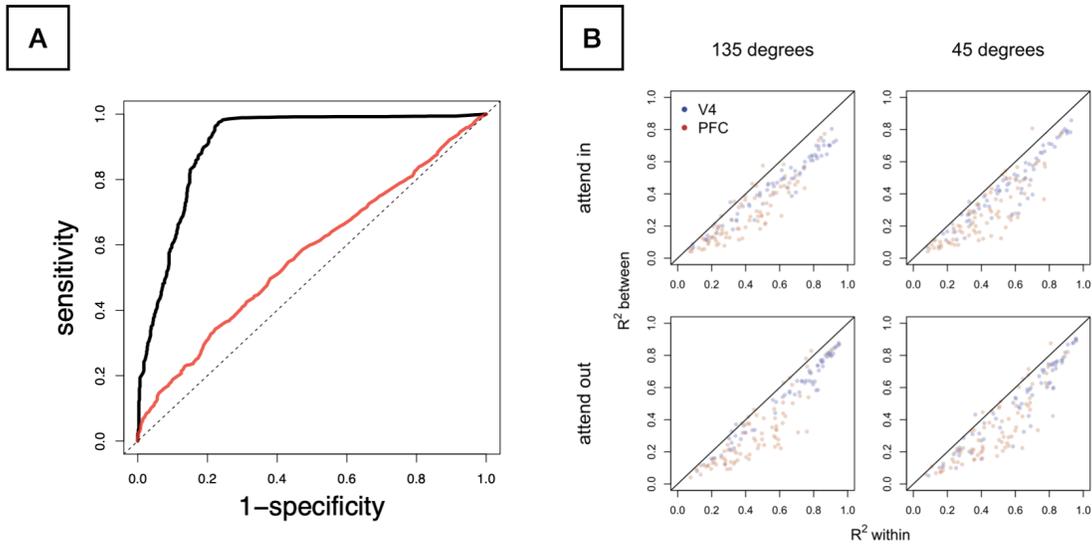


Figure 6: **Results for the graphical model of firing rates in PFC and V4.** **A** The ROC curve for the usual graphical Lasso (red) is far below the ROC curve for the method of Vinci et al. (2018, black). The ROC curves, which summarize the operating characteristics of the two procedures for correctly identifying non-zero partial correlations, are based on data simulated to be similar to real data (but with a known ground truth). The curve for an essentially perfect procedure would jump from the lower left corner, to the upper left corner, to the upper right corner; the curve for choosing by chance alone would hug the diagonal (dotted line). Thus, the graphical Lasso is only a little better than chance (coin-flipping) in identifying non-zero partial correlations while the method of Vinci et al. is much better. **B** The method of Vinci et al. estimates a variance matrix, from which regression R^2 values are easily calculated. Each dot in the four plots is the R^2 “between” versus “within” for a particular neuron’s firing rate Y (blue if that neuron is in V4, red if it is in PFC), where the regression is either on all the firing rates in the opposite area (between) or all the other firing rates in the same area (within). Four experimental conditions are shown involving two stimulus locations and either attending in or attending out according to whether the locations were in or out of the receptive fields. The plots show a wide range of values, spanning nearly the whole domain of R^2 across neurons, but those neurons whose firing rates are more (or less) predictable from the other firing rates within the same area are also more (or less) predictable from the firing rates in the other area. The method denoises to correct for attenuation, as in Figures 1C and 2. Modified from Vinci et al. (2018).

The methodology of Klein et al. (2020) was based on a class of probabilistic graphical models for which there are direct analogues to Gaussian graphical models: exponential families with two-way interactions (also known as maximum entropy models). The term “interaction” is used in the same way as in analysis of variance (ANOVA, e.g., Kass et al., 2014, Section 13.2.2). For exponential family models, the absence of a two-way interaction term in the model corresponds to conditional independence; thus, Klein et al. (2020) were also able to define an analogue of partial correlation for phase coupling. Because each phase angle lies on the unit circle and, mathematically, the product of circles is a torus, the model developed by Klein et al. (2020) for the data was an exponential family on a 24-dimensional torus; they called such models “torus graphs.”

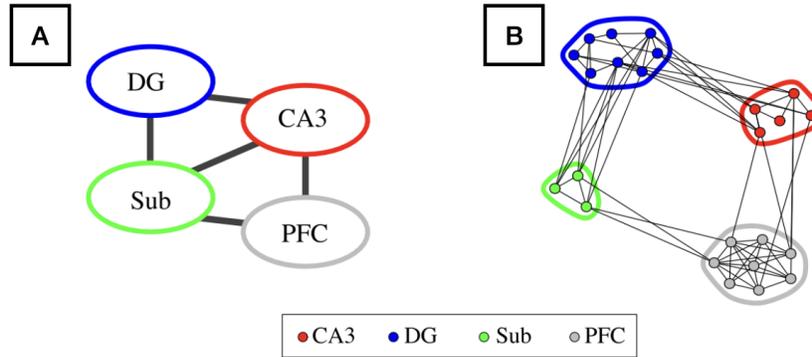


Figure 7: **Graphical models for PFC-Hippocampus phase locking during a memory task.** **A** A graph based on global testing for cross-area phase locking. **B** The graph based on post-hoc testing, after the global testing used for panel A, showing all significant edges. The main finding, in both panels, is that there was no evidence for phase locking between PFC and the dentate gyrus (DG), after conditioning on the activity in the two other areas, subiculum and CA3. A statistical quirk is that in B there are no significant edges between the subiculum and CA3, while in A that edge appears due to a cumulative effect corresponding to several edges not shown in B that were marginally significant. Modified from Klein et al. (2020). More information and analyses can be found there.

3.3 Autoregressive models and Granger causality

An important method for identifying interactions across areas from LFPs is Granger causality (Barnett & Seth, 2014; Bressler & Seth, 2011). One of the first studies that applied Granger causality to LFPs, Brovelli et al. (2004), demonstrated, first, that beta oscillations were coordinated across multiple areas during a sensorimotor task (via analysis of coherence, Section 3.5) and, second, that the activity in some areas predicted activity in other areas (Granger causality).

The idea is simple. A time series of values X_1, X_2, X_3, \dots at times $t = 1, 2, 3, \dots$, is Granger causal of a time series of values Y_1, Y_2, Y_3, \dots if, for every time value t , the past of X_t (values prior to time t) is predictive of Y_t , after taking account of the inherent dependence of Y_t on its own past. To be a bit more specific, a time series taking values Y_t at time t , is said to be autoregressive if there is a regression (a nonzero regression) of Y_t on Y_{t-1} , for all values t . This concept applies to stationary time series, meaning those having probability distributions that are time-invariant (see Kass et al., 2014, Chapter 18). More generally, a time series is autoregressive with order p if there is a regression of Y_t on the values $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$, going back in the past p time steps. Autoregression is the most straightforward way to model dependence across time. Vector autoregressive (VAR) models extend this concept to multiple time series.

With this in mind, a time series of values X_t is Granger causal of a time series of values Y_t if the regression of Y_t on the values $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ together with the values $X_{t-1}, X_{t-2}, \dots, X_{t-p}$ is stronger than the regression of Y_t on $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ alone. This verbal summary is meant to provide a succinct description of the concept for theoretical stationary time series; there are many

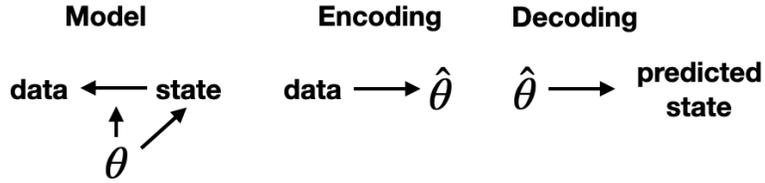


Figure 8: **Encoding and decoding.** The observation model, the first line in Equation 2, is depicted in the left-most diagram by the arrow leading from state to data. Both the observation model and the state model (in the second line of the equation) depend on parameters, collectively written here as a vector θ . Encoding uses the data to produce an estimate $\hat{\theta}$ of θ . Decoding uses the estimated model parameter vector to produce an estimate, or prediction, of the state vector. In typical brain-computer interface applications based on motor cortical activity, the state vector includes kinematic variables such as velocity that control a robot arm or a cursor on a screen.

important details in implementation and practice, and cautionary concerns about the application of Granger causality appear under time-varying dynamics, below. In addition, Granger causality is something of a misnomer because, as we discuss in Section 4.4, it is rarely causal in the usual scientific sense.

There are extensions of Granger causality based on directed information (Schreiber, 2000; Seth et al., 2015; Wibral et al., 2014), as well as extensions that apply to spike trains (Casile et al., 2021; Kim et al., 2011; Quinn et al., 2011). Furthermore, the autoregressive structure may be replaced with more flexible nonlinear relationships, including those generated by neural networks (Biswas & Ombao, 2022; Tank et al., 2022).

3.4 Latent dynamic models

In latent dynamic models a latent variable (typically a vector) takes values often known as states, which evolve dynamically, across time. The latent state variable represents theoretical quantities of interest, and its path across time is estimated from observations, often with the goal of prediction. The framework for an observation and state at time t in terms of states and noise at times t and $t - 1$ may be designated in the general form,

$$\begin{aligned}
 \text{observation}_t &\stackrel{\text{combines}}{\equiv} \text{state}_t \stackrel{\text{with}}{+} \text{observation noise}_t \\
 \text{state}_t &\stackrel{\text{combines}}{\equiv} \text{state}_{t-1} \stackrel{\text{with}}{+} \text{state noise}_t
 \end{aligned}
 \tag{2}$$

where “combines ... with” signifies that states are being combined with noise on the right-hand side to produce the left-hand side. When “combines ... with” is made precise in mathematical equations, the first line defines the way states determine observations (aside from noise) while the second line defines how the states evolve across time (subject to noise) within what is usually called a *state space*. The first line is often called the observation model while the second line becomes the state model. Important special cases are linear dynamical system models (LDS models, also

known as linear state-space models), and latent Gaussian process models, as well as the many generalizations of these two special cases; we give a few pointers to relevant literature in three successive subsections, below (see also Keeley et al., 2020).

In theory, stochastic processes can be used to formulate continuous-time state-space models, but practical implementations revert to discrete time. Although this may make the distinction between discrete and continuous time seem unimportant, some formulations have origins, and thus intuitions and useful results, grounded in continuous time. Models that are inherently discrete, and typically have additional structural simplifications, are usually called state-space models. When “combines ... with” involves linear models (and, for many purposes, Gaussian noise), the modifier “linear” may be included but, because this special case is so common, it is frequently taken as a default with “linear” omitted.

In neuroscience, state estimation is often known as “decoding,” with “encoding” referring to the process of estimating unknown parameters in the model, such as those appearing in the linear transformations and the noise distributions. See Figure 8. The first application to a decoding problem was Brown et al. (1998), who showed how the latent dynamical modeling framework could predict navigational paths based on multiple spike trains recorded from hippocampal place cells. Brain-computer interface applications to hand movement adapted this general scheme (Brockwell et al., 2004; Wu et al., 2006).

In applied mathematics the term “state space” is synonymous with “phase space,” where it refers to the possible configurations of a *dynamical system* that evolves in continuous time, moving from a state at time t to another state at time t' , according to differential equations, without the state noise term in Eq. (2). Much of theoretical neuroscience involves dynamical systems in this sense (e.g., Gerstner et al., 2014). Dynamical systems have played a prominent role in theories of motor control (Jordan, 1996; Shadmehr & Mussa-Ivaldi, 1994), and in data-analytic procedures aimed at understanding the neural control of movement (Vyas et al., 2020), with the concepts and techniques being seen as applicable to many other neural circuits (Duncker & Sahani, 2021; Shenoy & Kao, 2021). The motivation for latent dynamic models often leans on rough physical intuition, at least in the sense that relatively simple models can provide insight into dynamics. The term “dynamical system,” however, is often used in statistics and signal processing (Haykin & Moulines, 2007) to describe the noisy evolution of states in Eq. (2), and almost always refers to statistical state-space models, discussed very briefly in the following sub-section; linear state-space models are often called linear dynamical system (or LDS) models. One concept in applied mathematics of dynamical systems that might profitably receive more attention in neuroscience is the Koopman operator (Brunton et al., 2021; Lusch et al., 2018), which extends greatly a method known as dynamic mode decomposition (Brunton et al., 2016).

3.4.1 State-space models

In state-space models, the probability that a latent variable will be in a specific state at a particular time is typically restricted to depend on the most recently occurring state but not on earlier states (which is known as a Markov assumption). When the form of combination in Eq. (2) involves

a linear transformation of the state to the observations, with additive noise, in the first line (the observation model), and an autoregressive model for the states in the second line (the state model), and when the noise is Gaussian, the resulting procedure that produces state estimates is known as the Kalman filter (e.g., Kass et al., 2014, Section 16.2). In their application to place cell decoding, Brown et al. (1998) used a state-space model to accommodate Poisson spiking based on a generalized linear model (GLM, Kass et al., 2014, Chapter 14), rather than the linear Gaussian form. For general discussions, see Chen (2015) and Eden et al. (2018). Semedo et al. (2014) used a state-space model to identify the strength of cross-area interactions. By studying cross-area interactions on the level of the latent states, their method provided a succinct description of the interactions between V1 and V2 and described a time-structure distinction between within-area and cross-area interactions.

3.4.2 Gaussian processes

A Gaussian process (GP) is a stochastic process (Section 2.3) in which all the multivariate joint distributions are Gaussian. This requires specification of the covariance of variables X_{t_1} and X_{t_2} for every pair of time points t_1 and t_2 . To make analysis tractable, GP technology has emphasized particular forms of covariance, which restrict the appearance, especially the smoothness, of the resulting trajectories. Yu et al. (2009) replaced the static latent factors in ordinary factor analysis (depicted in Figure 5A) with Gaussian processes (Figure 5C) and showed how the resulting Gaussian process factor analysis (GPFA) could capture important features of multiple spike train behavior during movement, especially through visualization of latent trajectories across time. Although many possibilities for covariance functions have been discussed in the statistics and machine learning literature (Murphy, 2023, Chapter 18), the most common of these contain a single parameter that determines the smoothness of the trajectory. The implementation by Yu et al. (2009) allowed each latent factor to have its own smoothing parameter, and thereby provided multiple timescales for population representations. The latent variable structure of GPFA allows it to be considered a firing-rate model, and it reduces Poisson-like noise (Section 2.4.1) similarly to, though differently than, Poisson-based or point process models.

Gokcen et al. (2022) extended GPFA to multi-area cases by using a Gaussian process that had both area-specific and interacting components, with the interacting component including a delay variable to represent the time lag between areas. Applied to simultaneous recordings from V1 and V2, their method found bidirectional, asymmetric, and selective interactions.

3.4.3 Generalizations

The linear-Gaussian (or LDS) model and accompanying methodology has been extended widely, demonstrating the depth and breadth of the latent dynamical modeling framework (Murphy, 2023, Chapter 29). One of the basic extensions is to allow the system to evolve differently, at different times, according to one of several distinct linear dynamic regimes, leading to switching dynamical models often known as switching LDS (SLDS) models. In describing these, let us call the states

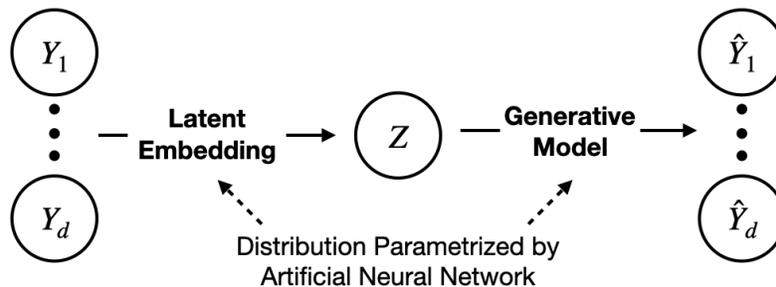


Figure 9: **Variational AutoEncoder (VAE) model.** This model combines the encoding and decoding of Figure 8 and uses artificial neural networks for each of the two parts. The labels “latent embedding” and “generative model” are terms used in the literature, but refer to the same processes and have the same general structure as encoding and decoding in state-space models. The “variational” modifier refers to variational inference, which uses a simplification to approximate the Bayesian calculations used for encoding in standard state-space models, such as linear dynamical systems.

in Eq. (2) instead *latent drivers* (they drive the population activity) and then let us refer to the different dynamic regimes as being in alternative states. With this terminology, SLDS models are called *recurrent* (rSLDS) if the latent drivers influence the probability of being in a particular state. A series of papers have developed these for spiking data. Glaser et al. (2020) included population-specific drivers and state probabilities while allowing the probabilities of staying within states to be different from the probabilities of moving to new states. The authors documented ways the dynamics within and between M1 and PMd changed as the task evolved from a delay period to a movement period during a reaching task.

A further extension is to allow either the mapping from latent drivers to observations, or the evolution of the latent drivers (or both) to involve neural network models. Gao et al. (2016) showed how multiple spike trains could be analyzed using a feed-forward neural network model to map from latent drivers to observations, where the states evolved linearly. In addition, Pandarinath et al. (2018) used a recurrent neural network to replace the Gaussian processes in GPFA, and artificial neural networks for encoding and decoding in an architecture known as a Variational AutoEncoder (VAE, see Figure 9), which improved predictive performance for single populations of spiking motor cortical neurons. Karniol-Tambour et al. (2022) then extended the multipopulation approach to rSLDS models (as discussed by Glaser et al., 2020) to incorporate a recurrent neural network that represented the evolution of latent drivers (as opposed to the linear dynamics in rSLDS). Analyzing calcium image data, they described feedforward drive from V1 to mV2 and PPC and feedback from PPC to V1 and mV2. Such an approach would apply to spike trains with a relatively minor modification to the observation distributions. From efforts such as Gao et al. (2016) and Karniol-Tambour et al. (2022) it is apparent that neural networks are especially valuable when they can be used in hybrid models that incorporate both components involving interpretable parameters and components for mapping functions (represented by neural networks) that do not require interpretation. The neural networks are then controlled by high-dimensional nuisance parameters in the sense of Section 2.4.2.

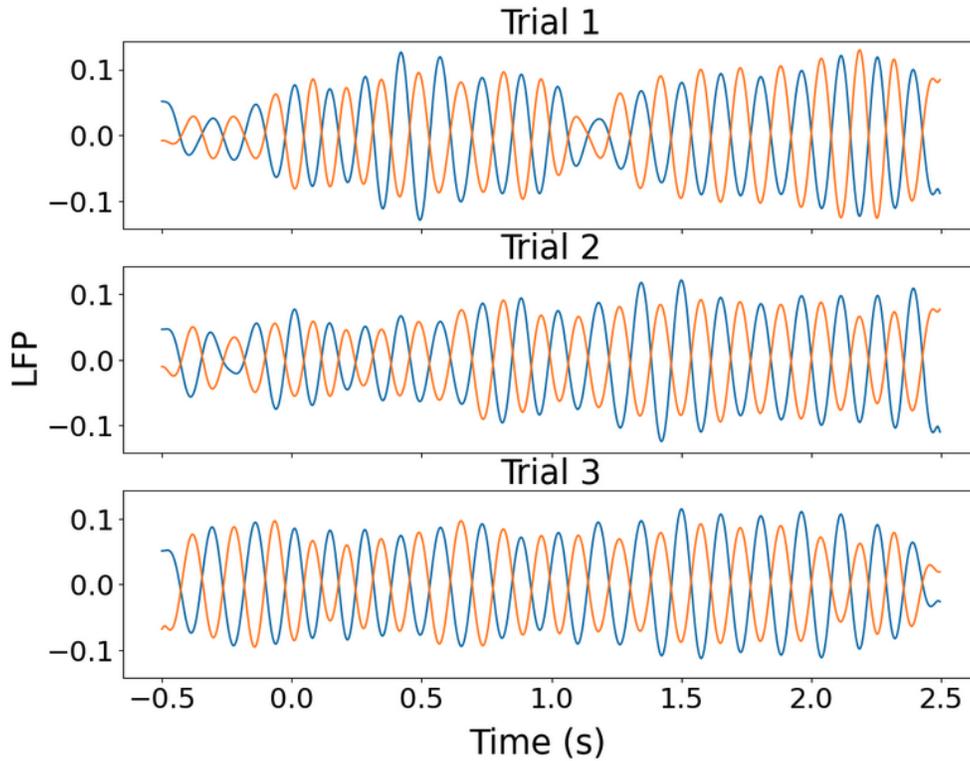


Figure 10: **Coupling of phases across trials.** Orange and blue lines denote two different theta-band filtered LFP signals, for three trials. There is a consistent pattern of phase offset between the two signals, with one signal offset from the other by about half a cycle, subject to some variation, but the phases at any specific trial time point vary across trials. The filtered LFPs also display amplitude variation, within and across trials.

3.5 Frequency analysis

Interest in neural oscillations has a long history (see, e.g., Buzsaki and Draguhn, 2004), and the potential physiological relevance of oscillatory cross-area interactions has been discussed in many places (Siegel et al., 2012; Whittington et al., 2018). Much of the literature concerns EEG and MEG recordings which, while methodologically different, produce data structured and analyzed similarly to LFP data. The analytical starting point is a decomposition of oscillatory signals into idealized frequency components. In the oldest and still-dominant case of Fourier analysis, the signal is decomposed into sinusoidal curves at many frequencies, each of which has both an amplitude and a phase (Cohen, 2014, Chapter 13; Kass et al., 2014, Chapter 18). Thus, a pair of signals from two areas can exhibit correlated behavior (across trials) in various combinations of amplitudes and phases at different frequencies, or in combinations of frequencies (see Figure 10).

Mathematically, an amplitude and phase may be considered together as defining a complex number, and when a pair of sinusoids have randomly varying amplitudes and phases, the magnitude of their complex-valued correlation is known as coherence. Coherence fits naturally into a now-classical theoretical framework for understanding coupling of oscillatory components in signals

(Shumway & Stoffer, 2017, Section 4.6). However, while complex-valued correlation is, in some respects, similar to the more familiar real-valued correlation (Pearson correlation), it is different. In particular, coherence depends not only on phase coupling but also on amplitude variation, which can sometimes make interpretation challenging. An alternative phase coupling assessment, known as Phase Locking Value (PLV), ignores amplitude and, under certain commonly-found circumstances, may be considered an analogue of Pearson correlation that applies to angles. When the amplitudes are nearly constant, coherence and PLV behave similarly, but the frequent presence of amplitude fluctuation has led some analysts to prefer PLV. For a comparative summary, and references, see Urban et al. (2023).

In the same sense that coherence is analogous to correlation, partial coherence (e.g., Dahlhaus, 2000) is analogous to partial correlation (which was illustrated in Figure 3C). The torus graphs developed by Klein et al. (2020) provided a multivariate generalization of PLV and, thus, as we mentioned in Section 3.2, the framework also provided measures analogous to partial correlation. Klein et al. discussed an important special case of torus graphs, which had been suggested earlier by Cadieu and Koepsell (2010), based on coupled oscillators. Lead-lag structures in the frequency domain can be identified using partial directed coherence (Baccalá & Sameshima, 2001) and spectral causality (Pinto-Orellana et al., 2021), the latter being a frequency-domain version of Granger causality estimated through VAR models (Section 3.3).

Cross-frequency coupling of theta and gamma oscillations in LFPs recorded from the hippocampus was described by Belluscio et al. (2012), and this is often considered a form of phase-amplitude coupling (Tort et al., 2010). Additional statistical references can be found in Ombao and Pinto (2022). General statistical considerations concerning cross-frequency coupling were discussed by Savolainen (2021). Amplitude-amplitude coupling has also been observed (Canolty & Knight, 2010). Siegel et al. (2012) reviews the different categories of interactions between phase and amplitude and their interpretation in the context of neural data. Urban et al. (2023) developed a complex-valued latent variable model to examine the combined phase-and-amplitude coupling of theta oscillations across 4 visual areas and two hippocampal areas in response to drifting gratings. Based on latent complex-valued partial correlation (latent partial coherence), only the two hippocampal areas CA1 and DG, as well as the contiguous areas AM and PM were coupled directly.

The time evolution of frequency coupling can be studied with several methods, including time windowing (e.g., Kass et al., 2014, Section 18.3.7) or wavelets (e.g., Cohen, 2014, Chapters 12 and 13). Fiecas and Ombao (2016) used time-localized spectra (periodograms) to identify time-varying interactions in LFPs recorded from hippocampus and the nucleus accumbens. Bong et al. (2020), Bong et al. (2023), and Liu et al. (2022) developed methods for identifying time-varying amplitude-amplitude interactions. Their primary model, depicted in Figure 5D, is a generalization of pCCA to time series (in a precise sense they formulated and proved). It allowed arbitrary time lags identified with partial correlation, within a fixed window (in their data analysis, lags up to 100 ms), which made it very flexible but it created a large number of parameters. They therefore developed an approach to estimation and inference (confidence intervals and significance tests) based on known strategies for high-dimensional problems, and they applied the methods to find time-varying beta amplitude coupling between PFC and V4 during a visual memory task.

When a neural spike train is driven by an oscillation, unless that oscillation is very powerful, recovering it is challenging statistically (Arai & Kass, 2017). Frequency analysis of point processes requires care. Investigators sometimes apply standard frequency analysis techniques to smoothed spike trains, but this can lead to erroneous conclusions (Brémaud & Massoulié, 2002; Kass et al., 2014, Section 19.3.7).

3.6 Point process models

A point process model describes spike trains by assuming at most one spike can occur at any point in time, and the probability of spiking is governed by a firing rate function (in continuous time). In practice, spikes are recorded in small time bins, and if the bins are small enough that at most one spike can occur in each bin, the point process defines a bin-specific probability that a spike will occur and the spike count (0 or 1) in each bin may be considered to follow a Poisson distribution. Because of possible statistical dependencies across bins, the spike counts in larger time bins need not be Poisson, which allows point processes to mimic the Poisson-like yet discernibly non-Poisson behavior of observed spike trains that is universally observed *in vivo* (with cortical spike trains exhibiting variation that is closest to Poisson; Shadlen and Newsome, 1998).

Formally, a point process statistical model for a spike train involves two things: (1) a simple, universal formula for the probability density of the spike train in terms of the theoretical firing rate and (2) a specification of the way the theoretical firing rate depends on relevant variables. The variables determining the theoretical firing rate could include measurements characterizing the stimulus, or behavior, or the past spiking of the neuron itself or of other neurons, and it often becomes a latent variable (as discussed in Section 2.4). Point process models are typically implemented using standard generalized regression methods (Kass et al., 2014, Chapter 19), where possibly nonlinear effects (e.g., due to a stimulus or the spiking of other neurons) are represented using basis functions (such as splines). Because the framework is intuitive, mathematically rigorous, and able to take advantage of modern statistical methods, point processes have long been considered the natural statistical approach to modeling spike trains (Brillinger, 1988; Brown et al., 2004; Kass et al., 2005). Increased knowledge, including documentation of their close relationship to integrate-and-fire models, has only made the theoretical case for their use more compelling (Kass et al., 2018, Sec. 2.4; Weber and Pillow, 2017), and there is a substantial literature on point process methods for spike train analysis, with historical markers including Kass and Ventura (2001), Truccolo et al. (2005), and Pillow et al. (2008). Many additional references may be found in Keeley et al. (2020) and Meyer et al. (2017). Point process models also form a statistical foundation for analysis of precise spike timing effects, as discussed in Kass et al. (2018, Section 3.4.2 and references therein), though these have been applied mainly within populations.

Although point processes may have a strong statistical motivation, they are models for spike counts in infinitesimally small time bins, usually approximated by time bins having 1 millisecond duration, and for some purposes it is entirely adequate to consider broader time windows. As a consequence, many useful analyses have applied simpler methods. Similarly, while a version of GPFA that works well for very small time bins has been articulated (Cunningham et al., 2007; Duncker

& Sahani, 2018), the original formulation, which uses Gaussian distributions rather than assuming the data to be binary, is less cumbersome. Point processes become a method of choice when it is advantageous to model the evolution of firing rate functions with millisecond precision. An example is the method of Chen et al. (2022) which, as illustrated in Figure 1, focused on the time of maximal firing rate in response to a stimulus. That method is discussed further below.

In Section 2.2 we noted that functional connectivity between two or more populations can be analyzed either by modeling directly the coupling of activity across the populations or by modeling the predictability of each recording from the many recordings across the other populations; the latter could be called a regression approach, as in Figure 4C. Both strategies have been applied to point process analysis of spike trains. The regression approach has been used by Hart and Huk (2020), Stevenson et al. (2008), and Zoltowski and Pillow (2018). For example, Hart and Huk (2020) used point process regression models to compare the strengths of interactions within and between LIP and FEF, and to show that the strengths of interactions changed from a fixation period to a delay period in an oculomotor task. At the population level, the possibility of applying point processes that include latent variables in the form of state models or Gaussian processes has been discussed by Keeley et al. (2020), based on successful implementations of such models within a single population (e.g., Linderman et al., 2016).

Chen et al. (2022) attacked population coupling by selecting neurons within each population that had similar theoretical firing rate variation across time, pooling together the selected spike trains within each population (i.e., merging all the spike times) to define a population spike train, and then modeling the covariation of particular features of the population firing rate functions, including the times of peak firing rate as shown in Figure 1A. The major components of the model appear in Figure 1D. Importantly, the neurons representing each population (“population membership”) were selected separately for each experimental condition (by incorporating model-based clustering into their Bayesian procedure). The approach then produced condition-specific PSTHs: the population firing rate function combined a condition-specific time-varying template (for the overall shape) with trial-varying features, which were the two peak firing rate times and a gain constant. The population firing rate function for a given trial (in a given brain area) then produced a population spike train by pooling together the spikes produced by all the neurons in that condition-specific population. Chen et al. focused on the time of maximal firing rate because, by definition, relatively large numbers of spikes occur near the time of maximal firing rate so that the statistical information content was high. Not only was this method strikingly sensitive (Figure 1C), but it could also recapitulate the spike count correlation histograms based on pairwise neural interactions, where all the neurons within a population are assumed to be statistically independent after conditioning on the two peak times and the gain constant, as shown in Figure 1E. That is, for these data, those three features of covariation appear to account for observed spike count correlation. Chen et al. examined covariation among the three visual areas V1, LM, and AL to find results like those in Figure 3C.

In a similar vein, as yet unpublished results by Urban, Jia, and Kass show that time-lagged neuron-to-neuron coupling across visual areas based on cross-correlograms (Jia et al., 2022) can be explained almost entirely by such population activity. In addition, other unpublished work by Xin, Siegel, and Kass demonstrates ways in which population firing rate models can be much more

accurate than individual neuron models for identifying cross-population interactions, and they use population firing rate models to document strong diminution of interactions across visual areas during locomotion. Further unpublished work by Olarinre, Siegle, and Kass used a simpler approach than that of Chen et al. (2022), corroborating results while verifying the importance of modeling (and thereby greatly reducing) the Poisson-like noise (discussed in Section 2.4.1) and fitting population firing rate functions separately for each experimental condition based on condition-specific neural populations. They used the simpler approach to analyze mouse-to-mouse variation of peak times and correlation of peak times across seven brain areas.

4 Important Issues

4.1 Defining neural populations

4.1.1 Neural population diversity

In spike train analysis, the sampled neurons are inhomogeneous, with both excitatory and inhibitory cells varying in properties, genetic makeup, location (including depth in cortex), and function (Harris and Shepherd, 2015). Typically, only task-relevant neurons, having firing rates that increase substantially in response to a stimulus or behavior, are used in analyses, but among these considerable diversity remains (Tripathy et al., 2013). In population-level analysis, the constitution of a relevant neural population is often left vague, the apparent assumption being that the haphazardly sampled neurons do a reasonably good job of capturing important effects. In some cases, however, populations could involve only a small percentage of neurons within some designated area; they could be condition-specific or behavior-specific or state-dependent, and they could involve particular classes of neurons. For example, Chen et al. (2022) identified relatively small condition-specific groups of neurons that participated in cross-population coupling (in the sense those authors were examining). The strong correlations observed by Chen et al. (2022), illustrated in Figure 1C, suggest this kind of selectivity can be an effective analytic strategy and could inform conceptions of population interaction.

Local field potentials also arise from many diverse sources, which may have very different activation patterns. One approach to disentangling the signals is through current source density localization (Buzsáki et al., 2012; Mitzdorf, 1985), methods for which continue to be developed, sometimes with an eye toward cross-population coupling. Klein et al. (2021) used Gaussian process current source densities to map alpha-band phase coupling between lateral and medial subareas of primary auditory cortex following tone presentations.

4.1.2 High dimensionality

While large numbers of electrodes create great opportunities to attack complexities in cross-population coupling, they also create computational challenges, which a small body of work has addressed (e.g., Cunningham et al., 2008; see also Keeley et al., 2020). More fundamentally, it is hard to fit accurately the large numbers of parameters involved in high-dimensional statistical models. In practice, quantities of data that seem very large in raw size (number of bytes) may actually be small when it comes to answering, reliably, a question being posed because a new recording capability has become available. This is the problem of granularity: as new scientific queries are generated by refining old ones, the number of possible results, which are typically defined by combinations of variables (e.g., identifying which neurons across all areas are interacting), grows rapidly, making it difficult or impossible to gain knowledge without somehow limiting the possibilities under consideration. Thus, every valuable statistical approach must, explicitly or implicitly, introduce some assumption that makes the problem tractable.

One common approach to limiting the possible results in high-dimensional problems invokes the statistical assumption known as sparsity, according to which there are comparatively few large effects (such as those quantifying interactions), and the small effects are not important. In settings such as linear regression and generalized regression, this leads to a computationally efficient and well-studied method known as L1 regularization, often called Lasso, which reduces the size of parameter estimates (it “shrinks” them) and sets the smallest estimates to zero. L1 regularization has been used to analyze multiple simultaneously-recorded spike trains where a given neuron’s spiking behavior could be functionally related to the spiking behavior of large numbers of other neurons (e.g., Kelly et al., 2010).

In neural applications it often happens that there are not a small number of large effects (e.g., neurons functionally connected to each other) but rather a large number of small effects. This calls into question the assumption of sparsity. In addition, statistical inference with L1 regularization can be challenging. Two ideas have been used to improve the situation. First, it is known that Lasso is generally better at prediction (predicting firing rate of one neuron based on firing patterns of many others) than model selection (finding correctly the set of neurons that have strong interactions with a given response neuron): intuitively, a good set of “wrong” variables can predict nearly as well as the right ones; see Bühlmann (2010). Bong et al. (2023) applied general results that take advantage of this fact (Janková & van de Geer, 2018; Ren et al., 2015) to identify time-varying beta-oscillation amplitude coupling of PFC and V4 during a memory task from LFPs recorded from Utah arrays.

Often, special-purpose methods can improve on generic Lasso-based approaches by incorporating scientifically reasonable assumptions. This was the strategy of Vinci et al. (2018), discussed in Section 3.2, where the usual L1 regularization for variance matrix estimation known as the graphical Lasso was modified by allowing the shrinkage of partial correlation for each pair of firing rates to depend on the distance between the two neurons and their tuning curve correlation. This produced the striking improvement in performance shown in Figure 6A.

4.2 Time-varying dynamics

Traditionally, interactions across brain areas have been depicted as static network snapshots. In reality, brain networks are dynamic and constantly changing. In many situations, the changes across time are of central interest. Tracking such changes is the primary motivation not only for the PSTH, but also for comparatively recent methods such as GPFA.

A particular challenge, affecting both spike trains and LFPs, is large fluctuations driven by stimulus or behavior, often called evoked or transient (or phasic) responses, which are clearly visible in typical PSTHs and trial-averaged LFPs. Such pulsatile fluctuations are difficult to reconcile with the statistical assumption of stationarity (time invariance), which is fundamental to common time series procedures. Repeated trials can provide the information needed to disentangle transient fluctuations from the kind of steady-state variation standard time series methods are built to handle. In GPFA, multiple processes with different effective timescales can often accommodate transient fluctuations while also capturing background variation. However, application of Granger causality requires care, and must be modified or abandoned in the presence of transient activity (Barnett & Seth, 2014; Ding & Wang, 2019; Seth et al., 2015; Wang et al., 2008). The same is true of directed information.

At the opposite end of the spectrum, variation that is slow relative to the length of the trial creates what is known as long-range dependence, which also causes problems. In time series models, the dependence between observations at different points in time, often measured by their correlation, which in this context is called autocorrelation, tends to get smaller as the time between the two observations increases. When the autocorrelation function decays exponentially fast, the dependence is called *short range*, and when it instead decays slowly the time series exhibits *long range dependence*. Even though long-range dependence can satisfy stationarity, when long-range dependence is present typical time series procedures will behave poorly, producing effects that are similar to non-stationarity (Beran, 1992). If, for example, there are trial-dependent waves of activity that elevate the average activity on some trials compared with others, observations near the end of a trial will tend to be correlated with observations near the beginning of a trial, thus exhibiting long-range dependence. When autoregressive models are fit to data, the best-fitting model may be the one having the greatest number of autoregressive terms, meaning that it allows dependence on values at maximal distance in the past; this would be an indication of likely long-range dependence or non-stationary behavior.

A conceptual complication, which often produces some confusion, involves the characterization of signals exhibiting long-range dependence as “ $1/f$ noise,” by which is meant that the spectral density (the power spectrum), written as a function of frequency f , decreases either exactly or approximately as $1/f$, though the terminology often refers to decreases that have a more general power law form $1/f^\alpha$, with α not necessarily close to 1. The exact or approximate form $1/f$ contrasts with spectra that decrease at least as rapidly as $1/f^2$, which are seen in standard statistical models exhibiting short-range dependence, such as autoregressive models.

The confusion comes from a few sources: failure to clearly distinguish the exact $1/f$, approximate $1/f$, or more general power law cases (including $1/f^2$); the statistical difficulty of determining the

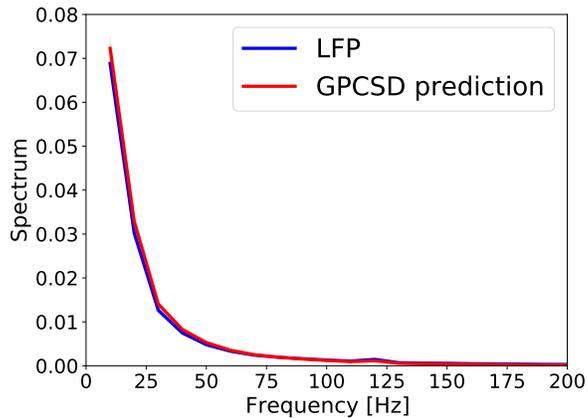


Figure 11: **LFP spectra.** Spectrum of an LFP (blue) from primate primary auditory cortex (A1) and spectrum of the fitted LFP using Klein et al. (2021)’s method (red), called GPCSD. The three-component model fits the data the LFP very well.

precise form of the spectral density for long-range dependent time series; and the conglomeration of other aspects of a signal that may accompany or explain long-range dependence, including scale invariance, chaotic evolution, and self-organized criticality. Cogent summaries may be found in Wagenmakers et al. (2004) and Ward and Greenwood (2007). There are many ways to generate time series having spectra that decrease as $1/f^\alpha$, including the special case $\alpha = 1$. In particular, beginning with a Poisson process that produces decaying voltage pulses, as in the usual conception of synaptic input, various realistic considerations (the precise form of decay; a combination of multiple independent processes; time-varying intensities) can lead to exact or approximate $1/f^\alpha$ spectra, with the value of α depending on the precise specification (Lowen & Teich, 1990; Ward & Greenwood, 2007). However, in practice, a very small number of short-range dependent processes, such as 3 processes, can produce spectra that are statistically (and visually) indistinguishable from $1/f$ (Wagenmakers et al., 2004). For example, as shown in Figure 11, Klein et al. (2021) obtained very good fits to multielectrode LFP recordings when they assumed the current source density (which produced the LFPs) had a 3-component form they labelled “transient + slow + fast,” where the transient component was a deterministic function (representing the evoked response) and the two other components were well-behaved Gaussian processes with two different time scales for “slow” and “fast” (which were learned from the data) and the fast component could accommodate alpha-band oscillations. They then used the current source densities to analyze cross-population coupling.

The two standard remedies for long-range dependence, which are equivalent conceptually but may differ in detail, are either to model it, as in Klein et al. (2021), or to fit the slow trends and then analyze residuals after trend removal (e.g., Kass et al., 2014, pages 528-529). When it is safe to assume stationarity, with short-range dependence, LFPs can be decomposed into frequency components. Instead of stationarity across the entire trial, local stationarity is often assumed, effectively meaning that stationarity holds within smaller windows of time; references were given in Section 3.5. Ombao et al. (2005) defined an efficient procedure for breaking trials into a relatively small number

of blocks within which stationarity could be assumed and Hefny et al. (2016) extended that model and applied it to multiple LFPs.

4.3 Trial-to-trial variability and Poisson-like noise

Substantial trial-to-trial variation is visible in both spike trains (where it is typically seen in raster plots of *in vivo* cortical data) and LFPs (where very large numbers of trials are usually averaged to see effects). Trial-to-trial variation is often called “noise,” especially in the context of spike count correlation (Averbeck et al., 2006; Cohen & Kohn, 2011; Smith & Kohn, 2008). A source of variability is labeled “noise” to suggest it is either not understood or thought to be irrelevant for some limited purpose, and the label may not be appropriate in a different context (we might say, “Today’s noise is tomorrow’s signal”). In particular, calling it noise is not intended to imply that trial-to-trial variation is uninteresting: when neural activity co-varies across trials in two areas, something has produced it. Co-variation indicates shared participation in circuits, at least in the sense that activity in one area affects activity in the other; or the two areas are driven to some identifiable extent by inputs not too far upstream that they have in common; or both.

There are at least four categories of sources for the variation in neural activity across trials. First, there may be subtle discrepancies in repetitions of the experimental paradigm or the manner in which the subject participates. During presentation of a visual stimulus, for example, even small deviations in foveal fixation might change responses of retinal ganglion cells, and thus the whole visual system. Second, endogenous states affecting neural activity, such as those associated with arousal or attention, may vary across trials. These first two sources may account for the large trial-to-trial variation in the mapping between experimental time and neural response seen in Figure 1 (see the legend for panel C). The third source is stochastic behavior in the flow of ions and neurotransmitters into and out of each neuron, and the fourth, at least in cortex, is the apparent widespread, low-level network activity, which has been hypothesized to benefit transmission of information (Shadlen & Newsome, 1998; Stein et al., 2005) and may arise from chaotic behavior of neurons that behave like sparsely-connected leaky integrators (Van Vreeswijk & Sompolinsky, 1996). The latter two sources, and perhaps other sources, produce what we called “Poisson-like” variation in Section 3.6, meaning that spiking patterns are similar to those produced by a Poisson process; although close examination typically reveals them to be noticeably different, neural spiking patterns are often not far different from patterns produced by Poisson processes (Chapter 19, Kass et al., 2014). Furthermore, point process models that can accommodate deviations from Poisson variation are similar mathematically to Poisson process models, which constitute an especially tractable special case. Thus, Poisson-like noise will tend to produce trial-to-trial variation that is consistent with that described by point process models.

4.4 Causality

Granger causality is a very natural and useful statistical tool, but it does not assess causality in the usual scientific sense; it is instead predictive. Experimental investigators know that correlation

does not imply causality: in general, an observed correlation between variables X and Y could be due to a confounding variable W , in which case if W were held fixed, changing the value of X would not change the value of Y . Statistical evidence that X causes Y can be gained only when confounding is highly improbable, such as when the values of X are assigned with a randomized procedure. Thus, in the absence of reasons to think confounding is highly improbable, Granger causality is predictive, but does not provide evidence of causality.

A different dichotomy that helps to clarify discussions of interacting neural populations contrasts causes of effects with effects of causes (Holland, 1988). The former aims to label the goal of reasoning from observations back to causal influences, while the latter refers to situations in which outcomes (effects) of differing experimental conditions (causal manipulations) are observed. Not only is the latter at the heart of experimental science, it is also the presumptive situation in which causal statistical reasoning can be valid. The field of statistics has developed a workable definition of causal effects, and a body of methods for establishing them. An elementary overview is given by Rosenbaum (2017); additional details may be found in work cited there and in Hernán and Robins (2023).

What should it mean to say that neural activity in one population, represented by a variable X , is a cause of activity in another population, represented by a variable Y ? Causal statements use imagined situations: a comparison is made between values of Y (or probability distributions of Y) when values of X change (e.g., a population-average firing rate goes from 2 Hz to 25 Hz), while all of the inputs to the neurons whose activity is represented by Y (those inputs that are not themselves driven by X) remain the same. Because it is not possible to observe neural activity Y under *exactly the same brain activity configurations* for two different values of neural activity X , a comparison of them involves *potential* outcomes, which are also called *counterfactual*. Statistical causal analysis thus defines causal effects in terms of potential outcomes (Hernán & Robins, 2023; Rosenbaum, 2017). Under special circumstances characterizing the design of a study (including but not limited to randomized controlled trials) it is possible to obtain causal statements. In the absence of causal manipulation, statistical analysis can not discover causes of effects while also demonstrating conclusively that they are causal (Robins & Wasserman, 1999).

One of the difficulties, mentioned in Section 3.2, arises even under relatively strong qualifying assumptions. Suppose, for example, that X_1 , X_2 , and X_3 represent measurements from interacting areas, under certain experimental conditions, and it is found that the partial correlation of X_1 and X_2 given X_3 is not statistically significantly different than zero, which could lead to the suggestion of Figure 3C. Even if it is assumed that (a) the measurements adequately represent relevant population activity and (b) no other brain areas affect this relationship, it is extremely difficult to distinguish a small partial correlation from a zero partial correlation, and the usual interpretation of Figure 3C corresponds to zero partial correlation. For this reason, interpretation of graphical models requires care: when authors use a graphical model to draw conclusions, they should keep in mind their merger of small and zero partial correlations. Helpful discussions of useful methods for determining graphical models, e.g., Glymour et al. (2019), sometimes appear to ignore this point or minimize its importance.

Causal manipulation can be provided by optogenetic stimulation, which has been a spectacular

advance, yet requires care to avoid misleading results (Deisseroth, 2015; Segala & Looger, 2021). One common concern is an imperfect match between the targeted population activity and the activity of cells actually affected. In causal inference, well-defined manipulations that only approximate an ideal manipulation are known as instrumental variables (Hernán & Robins, 2023; Rosenbaum, 2017). In some cases, estimated effects can be adjusted to estimate the effects that would have occurred under the ideal manipulation (here, the ideal stimulation of the targeted population). The use of optogenetic manipulations as instrumental variables has been discussed by Jiang et al. (2023) and Lepperød et al. (2018). Causal inference for neuroscience has been discussed in general terms by Biswas and Shlizerman (2022) and by Marinescu et al. (2018).

Causal effects are often depicted using directed graphs. An important caveat, however, is that in representing causal effects based on causal analysis, the directed graphs are restricted to being directed *acyclic* graphs (DAGs), which means they do not contain paths leading from a node through other nodes and back to the starting node. This restriction of causal graphs to DAGs rules out recurrent relationships.

5 Discussion

Our guide for data analysts wishing to identify interactions among neural populations has stressed analytical frameworks and statistical issues because progress requires both deployment of effective data analysis tools and appreciation of their inferential roles. There is a tension here, however, between powerful methods on the one hand, and a series of statistical cautions, on the other. The cautiousness emanating from the field of statistics, and the resulting reluctance to draw strong conclusions, can sometimes seem detrimental, especially when contrasted with the more ambitious culture of computer science that dominates machine learning (Kass, 2021), but the concerns should be taken seriously.

The prevailing agnosticism of statistics toward theoretical implications of data-driven results is most apparent in statistical causal analysis where, as described in Section 4.4, causal statements are said to concern “effects of causes” as opposed to “causes of effects.” However, the key discussion point in scientific reports, involving potential confounders, is familiar to experimental investigators. Furthermore, the flip side of the effects-of-causes versus causes-of-effects dichotomy is that effects of causal manipulations resulting from the “gold standard” of randomized controlled trials, while crucial, are rarely fully satisfactory for scientific pursuit. This is clearest in clinical studies. For example, randomized controlled trials demonstrated the effectiveness of lithium in treating bipolar disorder long before the emergence of substantial knowledge about potential mechanisms and, while evidence of effectiveness has strengthened (Severus et al., 2014), even now mechanisms are not well understood (Alda, 2015). It is thus useful to distinguish causal evidence (delivered through causal analysis, as in randomized trials) from causal explanation. Furthermore, because causal analysis rules out recurrence, causal explanation for recurrent circuits will have to go beyond standard frameworks for identifying causal evidence. Theoretical models, which have proven value in providing plausible mechanistic explanations that can inform understanding and guide future experimentation (Abbott et al., 2014), will surely continue to play a role, and we may get

closer to satisfying accounts of circuit behavior as suggested causes of effects are layered onto demonstrated effects of causes.

More generally, scientific articles often devote considerable space to *post hoc* explanatory hypotheses, stories that aim to synthesize and interpret results in terms of presumptive causal mechanisms. Such narratives are essential, even treasured components of the process. Still, it is important to recognize them as speculating about causes of effects, which is different than summarizing demonstrated effects of causes. This is a major motivation for our emphasis on statistical issues. As emerging technologies create more refined and detailed data, new statistical methods will likely continue to rely on the principles outlined here to meet the resulting challenges.

References

- Abbott, L., Miller, K., & Fusi, S. (2014). Theoretical approaches to neuroscience: Examples from single neurons to networks. In E. R. Kandel, S. J. H., T. M. Jessell, S. A. Siegelbaum, H. A. J., & M. S. (Eds.), *Principles of neural science, fifth edition*. McGraw-Hill Education.
- Alda, M. (2015). Lithium in the treatment of bipolar disorder: Pharmacology and pharmacogenetics. *Molecular psychiatry*, *20*(6), 661–670.
- Ames, K. C., & Churchland, M. M. (2019). Motor cortex signals for each arm are mixed across hemispheres and neurons yet partitioned within the population response. *Elife*, *8*, e46159.
- Arai, K., & Kass, R. E. (2017). Inferring oscillatory modulation in neural spike trains. *PLoS computational biology*, *13*(10), e1005596.
- Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature reviews neuroscience*, *7*(5), 358–366.
- Babadi, B., & Sompolinsky, H. (2014). Sparseness and expansion in sensory representations. *Neuron*, *83*(5), 1213–1226.
- Baccalá, L. A., & Sameshima, K. (2001). Partial directed coherence: A new concept in neural structure determination. *Biological cybernetics*, *84*(6), 463–474.
- Barnett, L., & Seth, A. K. (2014). The mvgc multivariate granger causality toolbox: A new approach to granger-causal inference. *Journal of neuroscience methods*, *223*, 50–68.
- Bassett, D. S., Greenfield, D. L., Meyer-Lindenberg, A., Weinberger, D. R., Moore, S. W., & Bullmore, E. T. (2010). Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS computational biology*, *6*(4), e1000748.
- Behseta, S., Berdyeva, T., Olson, C. R., & Kass, R. E. (2009). Bayesian correction for attenuation of correlation in multi-trial spike count data. *Journal of neurophysiology*, *101*(4), 2186–2193.
- Belluscio, M. A., Mizuseki, K., Schmidt, R., Kempner, R., & Buzsáki, G. (2012). Cross-frequency phase–phase coupling between theta and gamma oscillations in the hippocampus. *Journal of Neuroscience*, *32*(2), 423–435.
- Beran, J. (1992). Statistical methods for data with long-range dependence. *Statistical Science*, *7*(4), 404–416.

- Biswas, A., & Ombao, H. (2022). Frequency-specific non-linear granger causality in a network of brain signals. *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1401–1405.
- Biswas, R., & Shlizerman, E. (2022). Statistical perspective on functional and causal neural connectomics: A comparative study. *Frontiers in Systems Neuroscience*, *16*, 817962.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual review of psychology*, *53*(1), 605–634.
- Bong, H., Liu, Z., Ren, Z., Smith, M., Ventura, V., & Kass, R. E. (2020). Latent dynamic factor analysis of high-dimensional neural recordings. *Advances in neural information processing systems*, *33*, 16446–16456.
- Bong, H., Ventura, V., Yttri, E. A., Smith, M. A., & Kass, R. E. (2023). Latent cross-population dynamic time-series analysis of high-dimensional neural recordings. *arXiv preprint arXiv:2105.03508*.
- Brémaud, P., & Massoulié, L. (2002). Power spectra of general shot noises and hawkes point processes with a random excitation. *Advances in Applied Probability*, *34*(1), 205–222.
- Bressler, S. L., & Seth, A. K. (2011). Wiener–granger causality: A well established methodology. *Neuroimage*, *58*(2), 323–329.
- Brillinger, D. R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological cybernetics*, *59*(3), 189–200.
- Brockwell, A. E., Rojas, A. L., & Kass, R. E. (2004). Recursive bayesian decoding of motor cortical signals by particle filtering. *Journal of neurophysiology*, *91*(4), 1899–1907.
- Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., & Bressler, S. L. (2004). Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by granger causality. *Proceedings of the National Academy of Sciences*, *101*(26), 9849–9854.
- Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., & Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, *18*(18), 7411–7425.
- Brown, E. N., Kass, R. E., & Mitra, P. P. (2004). Multiple neural spike train data analysis: State-of-the-art and future challenges. *Nature neuroscience*, *7*(5), 456–461.
- Brunton, B. W., Johnson, L. A., Ojemann, J. G., & Kutz, J. N. (2016). Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*, *258*, 1–15.
- Brunton, S. L., Budišić, M., Kaiser, E., & Kutz, J. N. (2021). Modern koopman theory for dynamical systems.
- Bühlmann, P. (2010). Proposing the vote of thanks: Regression shrinkage and selection via the lasso: A retrospective by robert tibshirani.
- Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *science*, *304*(5679), 1926–1929.
- Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. *Nature reviews neuroscience*, *13*(6), 407–420.
- Cadieu, C. F., & Koepsell, K. (2010). Phase coupling estimation from multivariate phase statistics. *Neural computation*, *22*(12), 3107–3126.
- Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in cognitive sciences*, *14*(11), 506–515.

- Casile, A., Faghieh, R. T., & Brown, E. N. (2021). Robust point-process granger causality analysis in presence of exogenous temporal modulations and trial-by-trial variability in spike trains. *PLoS computational biology*, *17*(1), e1007675.
- Castellanos, L., Vu, V. Q., Perel, S., Schwartz, A. B., & Kass, R. E. (2015). A multivariate gaussian process factor model for hand shape during reach-to-grasp movements. *Statistica Sinica*, *25*(1), 5.
- Chen, Y., Douglas, H., Medina, B. J., Olarinre, M., Siegle, J. H., & Kass, R. E. (2022). Population burst propagation across interacting areas of the brain. *Journal of Neurophysiology*, *128*(6), 1578–1592.
- Chen, Z. (2015). *Advanced state space methods for neural and clinical data*. Cambridge University Press.
- Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature neuroscience*, *14*(7), 811–819.
- Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. MIT press.
- Cunningham, J. P., Shenoy, K. V., & Sahani, M. (2008). Fast gaussian process methods for point process intensity estimation. *Proceedings of the 25th international conference on Machine learning*, 192–199.
- Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, *17*(11), 1500–1509.
- Cunningham, J. P., Yu, B. M., Shenoy, K. V., & Sahani, M. (2007). Inferring neural firing rates from spike trains using gaussian processes. *Advances in neural information processing systems*, *20*.
- Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika*, *51*, 157–172.
- Datta, S. R., Anderson, D. J., Branson, K., Perona, P., & Leifer, A. (2019). Computational neuroethology: A call to action. *Neuron*, *104*(1), 11–24.
- Deisseroth, K. (2015). Optogenetics: 10 years of microbial opsins in neuroscience. *Nature neuroscience*, *18*(9), 1213–1225.
- Ding, M., & Wang, C. (2019). Analyzing meg data with granger causality: Promises and pitfalls. *Magnetoencephalography: From Signals to Dynamic Cortical Networks*, 647–657.
- Duncker, L., & Sahani, M. (2018). Temporal alignment and latent gaussian process factor inference in population spike trains. *Advances in neural information processing systems*, *31*.
- Duncker, L., & Sahani, M. (2021). Dynamics on the manifold: Identifying computational dynamical activity from neural population recordings [Computational Neuroscience]. *Current Opinion in Neurobiology*, *70*, 163–170. <https://doi.org/https://doi.org/10.1016/j.conb.2021.10.014>
- Eden, U. T., Frank, L. M., & Tao, L. (2018). Characterizing complex, multi-scale neural phenomena using state-space models. In *Dynamic neuroscience* (pp. 29–52). Springer.
- Einevoll, G. T., Kayser, C., Logothetis, N. K., & Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience*, *14*(11), 770–785.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, *1*(1), 1–47.

- Fiecas, M., & Ombao, H. (2016). Modeling the evolution of dynamic brain processes during an associative learning experiment. *Journal of the American Statistical Association*, *111*(516), 1440–1453.
- Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current opinion in neurobiology*, *37*, 66–74.
- Gao, P., & Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Current opinion in neurobiology*, *32*, 148–155.
- Gao, Y., Archer, E., Paninski, L., & Cunningham, J. P. (2016). Linear dynamical neural population models through nonlinear embeddings.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, B. (2013). Bayesian data analysis. 3rd edition.
- Gerstner, W., Kistler, W. M., Naud, R., & Paninski, L. (2014). *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press.
- Glaser, J., Whiteway, M., Cunningham, J. P., Paninski, L., & Linderman, S. (2020). Recurrent switching dynamical systems models for multiple interacting neural populations. *Advances in neural information processing systems*, *33*, 14867–14878.
- Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178.
- Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in genetics*, *10*, 524.
- Gokcen, E., Jasper, A. I., Semedo, J. D., Zandvakili, A., Kohn, A., Machens, C. K., & Yu, B. M. (2022). Disentangling the flow of signals between populations of neurons. *Nature Computational Science*, *2*(8), 512–525.
- Goris, R. L., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature neuroscience*, *17*(6), 858–865.
- Goulas, A., Zilles, K., & Hilgetag, C. C. (2018). Cortical gradients and laminar projections in mammals. *Trends in Neurosciences*, *41*(11), 775–788.
- Harris, K. D., & Shepherd, G. M. (2015). The neocortical circuit: Themes and variations. *Nature neuroscience*, *18*(2), 170–181.
- Hart, E., & Huk, A. C. (2020). Recurrent circuit dynamics underlie persistent activity in the macaque frontoparietal network. *Elife*, *9*, e52460.
- Haykin, S., & Moulines, E. (2007). Large-scale dynamic systems. *Proceedings of the IEEE*, *95*(5), 849–852.
- Hefny, A., Kass, R. E., Khanna, S., Smith, M., & Gordon, G. J. (2016). Fast and improved sleX analysis of high-dimensional time series. *Machine Learning and Interpretation in Neuroimaging: 4th International Workshop, MLINI 2014, Held at NIPS 2014, Montreal, QC, Canada, December 13, 2014, Revised Selected Papers 4*, 94–103.
- Hernán, M., & Robins, J. M. (2023). *Causal inference: What if*. CRC PRESS.
- Hilgetag, C. C., & Goulas, A. (2020). ‘hierarchy’ in the organization of brain networks. *Philosophical Transactions of the Royal Society B*, *375*(1796), 20190319.
- Hodges, J. S., & Sargent, D. J. (2001). Counting degrees of freedom in hierarchical and other richly-parameterised models. *Biometrika*, *88*(2), 367–379.
- Holland, P. W. (1988). Causal inference, path analysis and recursive structural equations models. *ETS Research Report Series*, *1988*(1), i–50.

- Hu, L., Guindani, M., Fortin, N. J., & Ombao, H. (2020). A hierarchical bayesian model for differential connectivity in multi-trial brain signals. *Econometrics and Statistics*, *15*, 117–135.
- Janková, J., & van de Geer, S. (2018). Inference in high-dimensional graphical models. In *Handbook of graphical models* (pp. 325–350). CRC Press.
- Jia, X., Siegle, J. H., Durand, S., Heller, G., Ramirez, T. K., Koch, C., & Olsen, S. R. (2022). Multi-regional module-based signal transmission in mouse visual cortex. *Neuron*, *110*(9), 1585–1598.
- Jiang, X., Saggat, H., Ryu, S. I., Shenoy, K. V., & Kao, J. C. (2020). Structure in neural activity during observed and executed movements is shared at the neural population level, not in single neurons. *Cell reports*, *32*(6), 108006.
- Jiang, Z., Chen, S., & Ding, P. (2023). An instrumental variable method for point processes: Generalised wald estimation based on deconvolution. *arXiv preprint arXiv:2301.03246*.
- Jordan, M. I. (1996). Computational aspects of motor control and motor learning. *Handbook of perception and action*, *2*, 71–120.
- Juavinett, A. L., Bekheet, G., & Churchland, A. K. (2019). Chronically implanted neuropixels probes enable high-yield recordings in freely moving mice. *Elife*, *8*, e47188.
- Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, *551*(7679), 232–236.
- Kang, B., & Druckmann, S. (2020). Approaches to inferring multi-regional interactions from simultaneous population recordings: Inferring multi-regional interactions from simultaneous population recordings. *Current Opinion in Neurobiology*, *65*, 108–119.
- Karniol-Tambour, O., Zoltowski, D. M., Diamanti, E. M., Pinto, L., Tank, D. W., Brody, C. W., & Pillow, J. W. (2022). Modeling communication and switching nonlinear dynamics in multi-region neural activity. *bioRxiv*, 2022–09.
- Kass, R. (2021). The two cultures: Statistics and machine learning in science. *Observational Studies*, *7*(1), 135–144.
- Kass, R. E. (2011). Statistical inference: The big picture. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *26*(1), 1.
- Kass, R. E., Amari, S.-I., Arai, K., Brown, E. N., Diekman, C. O., Diesmann, M., Doiron, B., Eden, U. T., Fairhall, A. L., Fiddymont, G. M., et al. (2018). Computational neuroscience: Mathematical and statistical perspectives. *Annual review of statistics and its application*, *5*, 183–214.
- Kass, R. E., Caffo, B. S., Davidian, M., Meng, X.-L., Yu, B., & Reid, N. (2016). Ten simple rules for effective statistical practice.
- Kass, R. E., Eden, U. T., & Brown, E. N. (2014). *Analysis of neural data* (Vol. 491). Springer.
- Kass, R. E., & Steffey, D. (1989). Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, *84*(407), 717–726.
- Kass, R. E., & Ventura, V. (2001). A spike-train probability model. *Neural computation*, *13*(8), 1713–1720.
- Kass, R. E., Ventura, V., & Brown, E. N. (2005). Statistical issues in the analysis of neuronal data. *Journal of neurophysiology*, *94*(1), 8–25.
- Kass, R. E., & Vos, P. W. (1997). *Geometrical foundations of asymptotic inference*. John Wiley & Sons.

- Keeley, S. L., Zoltowski, D. M., Aoi, M. C., & Pillow, J. W. (2020). Modeling statistical dependencies in multi-region spike train data. *Current Opinion in Neurobiology*.
- Kelly, R., Smith, M., Kass, R., & Lee, T. (2010). Accounting for network effects in neuronal responses using l1 regularized point process models. *Advances in neural information processing systems*, 23.
- Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, 58(3), 433–451.
- Kim, S., Putrino, D., Ghosh, S., & Brown, E. N. (2011). A granger causality measure for point process models of ensemble neural spiking activity. *PLoS computational biology*, 7(3), e1001110.
- Klein, N., Orellana, J., Brincat, S. L., Miller, E. K., & Kass, R. E. (2020). Torus graphs for multivariate phase coupling analysis. *The Annals of Applied Statistics*, 14(2), 635–660.
- Klein, N., Siegle, J. H., Teichert, T., & Kass, R. E. (2021). Cross-population coupling of neural activity based on gaussian process current source densities. *PLoS Computational Biology*, 17(11), 1–24. <https://doi.org/10.1371/journal.pcbi.1009601>
- Kohn, A., Coen-Cagli, R., Kanitscheider, I., & Pouget, A. (2016). Correlations and neuronal population information. *Annual Review of Neuroscience*, 39, 237–256.
- Kohn, A., Jasper, A. I., Semedo, J. D., Gokcen, E., Machens, C. K., & Byron, M. Y. (2020). Principles of corticocortical communication: Proposed schemes and design considerations. *Trends in Neurosciences*.
- Lee, T. S., & Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *JOSA A*, 20(7), 1434–1448.
- Lepperød, M. E., Stöber, T., Hafting, T., Fyhn, M., & Kording, K. P. (2018). Inferring causal connectivity from pairwise recordings and optogenetics. *bioRxiv*, 463760.
- Lindén, H., Tetzlaff, T., Potjans, T. C., Pettersen, K. H., Grün, S., Diesmann, M., & Einevoll, G. T. (2011). Modeling the spatial reach of the lfp. *Neuron*, 72(5), 859–872.
- Linderman, S. W., Adams, R. P., & Pillow, J. W. (2016). Bayesian latent structure discovery from multi-neuron recordings. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2010–2018.
- Liu, Z., Bong, H., Ren, Z., Smith, M., & Kass, R. E. (2022). *Simultaneous inference in multiple matrix-variate graphs for high-dimensional neural recordings* [In preparation].
- Lowen, S. B., & Teich, M. C. (1990). Power-law shot noise. *IEEE transactions on information theory*, 36(6), 1302–1318.
- Luo, L. (2020). *Principles of neurobiology*. Garland Science.
- Lusch, B., Kutz, J. N., & Brunton, S. L. (2018). Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-07210-0>
- Marinescu, I. E., Lawlor, P. N., & Kording, K. P. (2018). Quasi-experimental causality in neuroscience and behavioural research. *Nature human behaviour*, 2(12), 891–898.
- Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2), 456–492.
- Mason, C. R., Gomez, J. E., & Ebner, T. J. (2001). Hand synergies during reach-to-grasp. *Journal of neurophysiology*, 86(6), 2896–2910.

- Meyer, A. F., Williamson, R. S., Linden, J. F., & Sahani, M. (2017). Models of neuronal stimulus-response functions: Elaboration, estimation, and evaluation. *Frontiers in systems neuroscience*, *10*, 109.
- Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: Investigation of evoked potentials and eeg phenomena. *Physiological reviews*, *65*(1), 37–100.
- Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT press.
- Murphy, K. P. (2023). *Probabilistic machine learning: Advanced topics*. MIT Press. <http://probml.github.io/book2>
- Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J., & Cohen, M. R. (2018). Learning and attention reveal a general relationship between population activity and behavior. *Science*, *359*(6374), 463–465.
- Ombao, H., & Pinto, M. (2022). Spectral dependence. *Econometrics and Statistics*.
- Ombao, H., Von Sachs, R., & Guo, W. (2005). Slex analysis of multivariate nonstationary time series. *Journal of the American Statistical Association*, *100*(470), 519–531.
- Pandarathna, C., O’Shea, D. J., Collins, J., Jozefowicz, R., Stavisky, S. D., Kao, J. C., Trautmann, E. M., Kaufman, M. T., Ryu, S. I., Hochberg, L. R., et al. (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods*, *15*(10), 805–815.
- Pereira, T. D., Shaevitz, J. W., & Murthy, M. (2020). Quantifying behavior to understand the brain. *Nature neuroscience*, *23*(12), 1537–1549.
- Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., & Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, *454*(7207), 995–999.
- Pinto-Orellana, M. A., Mirtaheri, P., Hammer, H. L., & Ombao, H. (2021). Scau: Modeling spectral causality for multivariate time series with applications to electroencephalograms. *arXiv preprint arXiv:2105.06418*.
- Potjans, T. C., & Diesmann, M. (2014). The cell-type specific cortical microcircuit: Relating structure and activity in a full-scale spiking network model. *Cerebral cortex*, *24*(3), 785–806.
- Quinn, C. J., Coleman, T. P., Kiyavash, N., & Hatsopoulos, N. G. (2011). Estimating the directed information to infer causal relationships in ensemble neural spike train recordings. *Journal of computational neuroscience*, *30*(1), 17–44.
- Raducanu, B. C., Yazicioglu, R. F., Lopez, C. M., Ballini, M., Putzeys, J., Wang, S., Andrei, A., Rochus, V., Welkenhuysen, M., Helleputte, N. v., et al. (2017). Time multiplexed active neural probe with 1356 parallel recording sites. *Sensors*, *17*(10), 2388.
- Ren, Z., Sun, T., Zhang, C.-H., & Zhou, H. H. (2015). Asymptotic normality and optimalities in estimation of large gaussian graphical models. *The Annals of Statistics*, *43*(3), 991–1026. Retrieved February 26, 2023, from <http://www.jstor.org/stable/43556545>
- Rios, G., Lubenov, E. V., Chi, D., Roukes, M. L., & Siapas, A. G. (2016). Nanofabricated neural probes for dense 3-d recordings of brain activity. *Nano letters*, *16*(11), 6857–6862.
- Robins, J. M., & Wasserman, L. (1999). On the impossibility of inferring causation from association without background knowledge. *Computation, causation, and discovery*, *1999*, 305–21.
- Rodu, J., Klein, N., Brincat, S. L., Miller, E. K., & Kass, R. E. (2018). Detecting multivariate cross-correlation between brain regions. *Journal of Neurophysiology*, *120*(4), 1962–1972.

- Rosenbaum, P. (2017). *Observation and experiment: An introduction to causal inference*. Harvard University Press.
- Savolainen, O. W. (2021). The significance of neural inter-frequency power correlations. *Scientific Reports*, *11*(1), 23190.
- Saxena, S., & Cunningham, J. P. (2019). Towards the neural population doctrine. *Current opinion in neurobiology*, *55*, 103–111.
- Schervish, M. J. (2012). *Theory of statistics*. Springer Science & Business Media.
- Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, *85*(2), 461.
- Segala, M. G. P., & Looger, L. L. (2021). Optogenetics. In *Molecular imaging* (pp. 283–302). Elsevier.
- Semedo, J. D., Gokcen, E., Machens, C. K., Kohn, A., & Byron, M. Y. (2020). Statistical methods for dissecting interactions between brain areas. *Current opinion in neurobiology*, *65*, 59–69.
- Semedo, J. D., Jasper, A. I., Zandvakili, A., Krishna, A., Aschner, A., Machens, C. K., Kohn, A., & Yu, B. M. (2022). Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nature communications*, *13*(1), 1099.
- Semedo, J. D., Zandvakili, A., Kohn, A., Machens, C. K., & Byron, M. Y. (2014). Extracting latent structure from multiple interacting neural populations. *Advances in neural information processing systems*, 2942–2950.
- Semedo, J. D., Zandvakili, A., Machens, C. K., Byron, M. Y., & Kohn, A. (2019). Cortical areas interact through a communication subspace. *Neuron*, *102*(1), 249–259.
- Seth, A. K., Barrett, A. B., & Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, *35*(8), 3293–3297.
- Severus, E., Taylor, M. J., Sauer, C., Pfennig, A., Ritter, P., Bauer, M., & Geddes, J. R. (2014). Lithium for prevention of mood episodes in bipolar disorders: Systematic review and meta-analysis. *International journal of bipolar disorders*, *2*(1), 1–17.
- Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of neuroscience*, *18*(10), 3870–3896.
- Shadmehr, R., & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *Journal of neuroscience*, *14*(5), 3208–3224.
- Shenoy, K. V., & Kao, J. C. (2021). Measurement, manipulation and modeling of brain-wide neural population dynamics. *Nature communications*, *12*(1), 633.
- Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: With r examples.
- Siegel, M., Donner, T. H., & Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience*, *13*(2), 121–134.
- Smith, M. A., & Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *Journal of Neuroscience*, *28*(48), 12591–12603.
- Stein, R. B., Gossen, E. R., & Jones, K. E. (2005). Neuronal variability: Noise or part of the signal? *Nature Reviews Neuroscience*, *6*(5), 389–397.
- Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., et al. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*, *372*(6539).

- Steinmetz, N. A., Koch, C., Harris, K. D., & Carandini, M. (2018). Challenges and opportunities for large-scale electrophysiology with neuropixels probes. *Current opinion in neurobiology*, *50*, 92–100.
- Stevenson, I. H., Rebesco, J. M., Hatsopoulos, N. G., Haga, Z., Miller, L. E., & Kording, K. P. (2008). Bayesian inference of functional connectivity and network structure from spikes. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *17*(3), 203–213.
- Tank, A., Covert, I., Foti, N., Shojaie, A., & Fox, E. B. (2022). Neural granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(8), 4267–4279.
- Tokdar, S. T., Xi, P., Kelly, R. C., & Kass, R. E. (2009). Detection of bursts in extracellular spike trains using hidden semi-markov point process models. *Journal of computational neuroscience*, *29*(1-2), 203–212.
- Tort, A. B., Komorowski, R., Eichenbaum, H., & Kopell, N. (2010). Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *Journal of neurophysiology*, *104*(2), 1195–1210.
- Tripathy, S. J., Padmanabhan, K., Gerkin, R. C., & Urban, N. N. (2013). Intermediate intrinsic diversity enhances neural population coding. *Proceedings of the National Academy of Sciences*, *110*(20), 8248–8253.
- Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of neurophysiology*, *93*(2), 1074–1089.
- Urban, K., Bong, H., Orellana, J., & Kass, R. E. (2023). *Oscillating neural circuits: Phase, amplitude and the complex normal distribution*.
- Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.
- Van Vreeswijk, C., & Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, *274*(5293), 1724–1726.
- Vezoli, J., Magrou, L., Goebel, R., Wang, X.-J., Knoblauch, K., Vinck, M., & Kennedy, H. (2021). Cortical hierarchy, dual counterstream architecture and the importance of top-down generative networks. *Neuroimage*, *225*, 117479.
- Vinci, G., Ventura, V., Smith, M. A., & Kass, R. E. (2016). Separating spike count correlation from firing rate correlation. *Neural computation*, *28*(5), 849–881.
- Vinci, G., Ventura, V., Smith, M. A., & Kass, R. E. (2018). Adjusted regularization in latent graphical models: Application to multiple-neuron spike count data. *The annals of applied statistics*, *12*(2), 1068.
- Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation through neural population dynamics. *Annual Review of Neuroscience*, *43*, 249–275.
- Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of $1/f\alpha$ noise in human cognition. *Psychonomic bulletin & review*, *11*(4), 579–615.
- Wang, X., Chen, Y., & Ding, M. (2008). Estimating granger causality after stimulus onset: A cautionary note. *Neuroimage*, *41*(3), 767–776.
- Ward, L. M., & Greenwood, P. E. (2007). $1/f$ noise. *Scholarpedia*, *2*(12), 1537.
- Weber, A. I., & Pillow, J. W. (2017). Capturing the dynamical repertoire of single neurons with generalized linear models. *Neural computation*, *29*(12), 3260–3289.
- Whittington, M. A., Traub, R. D., & Adams, N. E. (2018). A future for neuronal oscillation research. *Brain and Neuroscience Advances*, *2*, 2398212818794827.

- Wibral, M., Vicente, R., & Lizier, J. T. (2014). *Directed information measures in neuroscience*. Springer.
- Wu, W., Gao, Y., Bienenstock, E., Donoghue, J. P., & Black, M. J. (2006). Bayesian population decoding of motor cortical activity using a kalman filter. *Neural computation*, *18*(1), 80–118.
- Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., & Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of Neurophysiology*, *102*(1), 614–635.
- Zoltowski, D., & Pillow, J. W. (2018). Scaling the poisson glm to massive neural datasets through polynomial approximations. *Advances in Neural Information Processing Systems*, *31*, 3517–3527.