Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

Multi-neuronal activity and functional connectivity in cell assemblies

Yasser Roudi[1,2], Benjamin Dunn[1] and John Hertz[3,2]

Our ability to collect large amounts of data from many cells has been paralleled by the development of powerful statistical models for extracting information from this data. Here we discuss how the activity of cell assemblies can be analyzed using these models, focusing on the generalized linear models and the maximum entropy models and describing a number of recent studies that employ these tools for analyzing multi-neuronal activity. We show results from simulations comparing inferred functional connectivity, pairwise correlations and the real synaptic connections in simulated networks demonstrating the power of statistical models in inferring functional connectivity. Further development of network reconstruction techniques based on statistical models should lead to more powerful methods of understanding functional anatomy of cell assemblies.

Addresses
[1] Kavli Institute & Centre for Neural Computation, NTNU, Trondheim, Norway
[2] Nordita, KTH Royal Institute of Technology and Stockholm University, Stockholm, Sweden
[3] Niels Bohr Institute, Copenhagen, Denmark

Corresponding author: Roudi, Yasser (yasser.roudi@ntnu.no)

Current Opinion in Neurobiology 2015, **32**:38–44

This review comes from a themed issue on **Large-scale recording technology**

Edited by **Francesco P Battaglia** and **Mark J Schnitzer**

For a complete overview see the Issue and the Editorial

Available online 8th November 2014

http://dx.doi.org/10.1016/j.conb.2014.10.011

Introduction

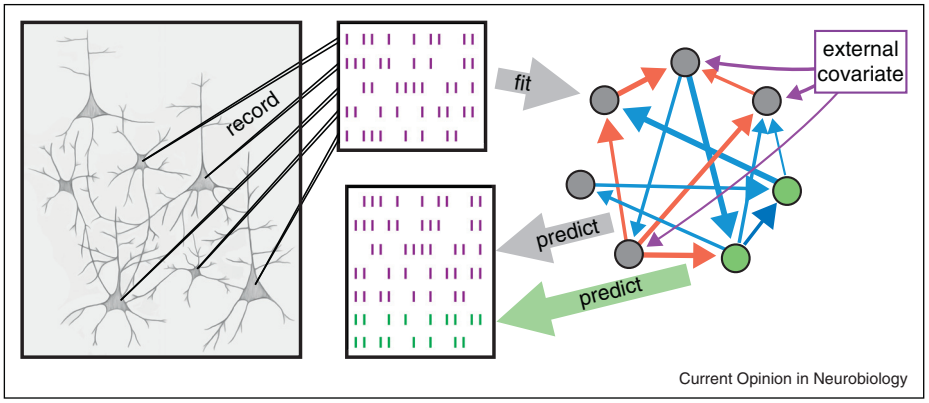
In lower species, single neurons or small circuits with stereotyped connectivity patterns are studied as computational building blocks of the nervous system. In higher species, such as mammals, on the other hand, populations of neurons, or cell assemblies, are probably the closest thing to a computational unit [1]. A familiar example is the Hebbian cell assembly [2]: a group of neurons with stronger connections between the cells within the group than with other cells. The stronger connections between the neurons in the Hebbian assembly leads to the attractor dynamics that is believed to underlie a variety of neuronal computations [3]. Other examples of cell assemblies include groups of neurons that share functional similarities, such as color, form and motion selective cells in primary visual areas [4], or the barrels in the rat barrel cortex [5]. Functional cells assemblies also exist in higher cortical areas. An example is that of grid cells in the medial entorhinal cortex [6]. As an animal runs in a two-dimensional environment, each grid cell fires maximally at locations that form a hexagonal pattern. Grid cells form a functional cell assembly and are coupled to cells in the same anatomical location that belong to other assemblies, for example, border cells [7] and head directional cells [8]. To understand computation in the mammalian nervous system, one has to characterize these assemblies and their relationship to each other and to identify the anatomical and molecular features associated with specific assemblies.

Although the theoretical concept of cell assemblies is not new, tools for analyzing them have only recently emerged in systems neuroscience. Experimentalists can now record the activity of many cells at the same time, and the spatial and temporal resolution with which these recordings can be done is increasing rapidly [9,10]. With new recording technology, even areas previously inaccessible to simultaneous multi-cell recording are becoming available. In addition, optogenetic [11] and other molecular and genetic techniques [12,13] now allow experimentalists to stimulate specific kinds of cells during their recordings. All these advances have shifted the focus of efforts to understand neural computation from single-neuron recordings to simultaneous recordings of many neurons. In parallel, there has been significant progress on theoretical and computational tools for analyzing such recordings. Although, so far, these methods have been applied almost exclusively to data from sensory or motor areas, we can anticipate their exploitation in higher cortical areas, leading to new ways of thinking about information processing at the population level. In this paper, we review the main modern approaches for modeling multi-unit recordings and discuss future avenues that can be explored using these methods.

Statistical modeling

Understanding a complex system is achieved through models, and the high variability of neuronal data requires that these models be statistical ones. Here, we describe how to build statistical models of multi-neuronal activity and show, using several examples, how they can help us understand the computational and physiological properties of cells assemblies, as well as the relationship between them.

Figure 1



Using spike trains recorded from neurons in a population (upper middle panel), one can learn a statistical model and the functional or effective connections (arrows in the right panel) between neurons (circles in the right panel) and between neurons and external factors that may influence the neuronal spike trains. This process can also include learning and inference of hidden variables, for example, unrecorded neurons (green circle in the right panel). The functional connections do not in general correspond to actual physical connections, though in some cases they may be very informative the presence or absence of connections [15,16•]; see also Figure 2. The inferred model can be used, for example, to generate synthetic data (lower middle panel) or to assign quantitative values for external covariates in explaining the data.

A statistical model is based on an assumed, parameterized form of some distribution, and its parameters are found by maximizing the likelihood of the data, that is, finding which model, among all those obtained by varying the parameters, is most likely to have generated the available data. The data we are thinking of here — spike trains, calcium imaging data, local field potential signals, or combinations of these — are very high-dimensional. Nevertheless, conceptually, this problem is no more abstract than the elementary one of fitting a Gaussian distribution to a set of measurements of a single variable. It is just of higher dimensionality (and technical issues therefore arise in making the fit), because the models have many parameters. Whatever the method for fitting the model, the outcome of this process is a statistical model with a set of functional connections (described in more details in the next section) which can be used for network reconstruction, generating synthetic data and/or assigning quantitative values to the role of unrecorded (hidden) neurons or external covariates in shaping multi-neuronal activity; see Figure 1.

Of course, there are many statistical models one can fit to data, and the choice of the model depends largely on the goal of the modeling effort and the available data. Common choices usually rely on prominent physiological features, such as the fact that single neurons usually integrate the input they receive over tens of ms, theoretical concepts such as the maximum entropy principle [14], or a combination of these. We will focus on two classes of models: so-called generalized linear models (henceforth abbreviated GLMs), and maximum-entropy (abbreviated max-ent) models. We describe their main features here; more details and key equations can be found in the accompanying Box 1.

GLMs assume that every neuron spikes at a time-varying rate which depends on earlier spikes (both those of other neurons and its own) and on 'external covariates' (such as a stimulus or other quantities measured in the experiment). The influence of earlier spikes on the firing probability at a given time is assumed to depend on the time since they occurred. For each 'pre-postsynaptic' pair i, j, it is described by a function $J_{ij}(\tau)$ of this time lag. In addition, there are more functions describing the effects of the external covariates. To fit the model, then, one finds those functions, out of all possible ones satisfying some reasonable smoothness constraints, for which the actual recorded spike history has the highest probability [17–19].

Statistical models of the max-ent type are different: One does not consider the likelihood of the recorded history. Rather, one takes, as the data, the set of observed simultaneous (i.e. within a single time bin) spike patterns, without regard to their temporal order. One then finds the distribution, within the class of distributions that have maximum entropy, given (for example) the measured firing rates and pairwise correlations [20–22], that maximizes the likelihood of finding these patterns. The reason for choosing the estimated distribution among the maximum-entropy class, is that this procedure, uniquely, makes minimal assumptions about the distribution of patterns (see Box 1).

Thus, one can say loosely that GLMs focus on predicting future spikes of a given neuron from past spikes of all neurons, while max-ent models aim at predicting spikes of a given neuron from the spike pattern of other neurons at the same time. Regardless of the chosen model, with the increasing availability of large data sets we are able to