

# Maximum entropy models as a tool for building precise neural controls

Cristina Savin<sup>1,2</sup> and Gaper Tkačik<sup>2</sup>



Neural responses are highly structured, with population activity restricted to a small subset of the astronomical range of possible activity patterns. Characterizing these statistical regularities is important for understanding circuit computation, but challenging in practice. Here we review recent approaches based on the maximum entropy principle used for quantifying collective behavior in neural activity. We highlight recent models that capture population-level statistics of neural data, yielding insights into the organization of the neural code and its biological substrate. Furthermore, the MaxEnt framework provides a general recipe for constructing surrogate ensembles that preserve aspects of the data, but are otherwise maximally unstructured. This idea can be used to generate a hierarchy of controls against which rigorous statistical tests are possible.

## Addresses

<sup>1</sup>Center for Neural Science and Center for Data Science, New York University, 10003 New York, USA

<sup>2</sup>IST Austria, A-3400 Klosterneuburg, Austria

Corresponding author: Savin, Cristina ([cs5360@nyu.edu](mailto:cs5360@nyu.edu))

**Current Opinion in Neurobiology** 2017, **46**:120–126

This review comes from a themed issue on **Computational neuroscience**

Edited by **Adrienne Fairhall** and **Christian Machens**

<http://dx.doi.org/10.1016/j.conb.2017.08.001>

0959-4388/© 2017 Elsevier Ltd. All rights reserved.

## From descriptive statistics to probabilistic models

Neural computation arises through the collective behavior of neural populations [1]. What ‘collective behavior’ means in statistical terms is that the distribution of activity configurations (or states) explored by the circuit has nontrivial structure that cannot be explained by statistical properties of individual cells alone. One fundamental challenge for systems neuroscience is to characterize what are the key regularities in neural activity and how they relate to circuit function.

From the data analysis perspective, finding signatures of collective behavior in data is critical for understanding neural coding and information processing [2]. Examples include functional couplings between neurons [3–5], the prevalence of redundancy and synergy in the neural code [6,7], the large-scale organization of neural activity [8,9], or comparisons of patterned variability in spontaneous and stimulus evoked activity [10].

Statistics that reflect regularities in neural activity can also be used as a bridge between theory and experiments, since models of neural computation predict patterns of collective activity which can then be compared against data. For instance, the structure of pairwise correlations has been investigated for a range of network models [11,12]. Another classic example is attractor dynamics, a core concept of computational models describing neural pattern completion, memory recall, and working memory activity whose statistical signatures were found in various circuits [13,14,15,30]. More generally, identifying model-specific regularities in population activity provides useful ways to validate theories of neural computation and mechanistic models of network dynamics [16].

As the role of correlated activity for neural coding and its mapping to underlying circuitry have recently been reviewed elsewhere [5,17,16], we focus our attention on the more pragmatic question of how to quantify statistical regularities in neural data. This is becoming an increasingly critical issue with the advent of recording technologies that allow tracking the joint activity of large groups of neurons during behavior [18,19]. The combinatorial explosion of possible population activity patterns, paired with relatively limited recording time, massively complicates data analysis [20,21]. Nonetheless the analysis tools are catching up to the technical developments [22,23]. One such tool is the maximum entropy (MaxEnt) framework, an approach which links measurable statistics of neural activity to corresponding probabilistic models of such data.

While MaxEnt models have a long history in statistics [25], physics [24], or ecology [26], they have only recently gained popularity in neuroscience. We argue that MaxEnt models offer a natural bridge between two different traditions for investigating collective behavior in neural activity. On the one hand, an approach with roots in frequentist statistics, and often favored by experimentalists, is to

identify and report single, strong, salient signatures of the neural computation (e.g. noise correlations). When the choice of summary statistic is informed by theory, this comparison becomes a hypothesis test about that theory. It usually entails a step where data is compared to controls in which the identified structure is disrupted by some form of shuffling, either to assess its statistical significance or to identify its origins. On the other hand, an approach with roots in Bayesian statistics and often favored by theorists, is to build a probabilistic model of the neural activity that predicts the probability of every possible activity configuration. Here the emphasis is on the quality of the model fit. MaxEnt models link the two approaches by being, at the same time, *bona fide* probabilistic models for neural activity, as well as generalizations of frequentist shuffles. The goal of this paper is to highlight the use of MaxEnt models as controls for more complex probabilistic models of the neural code, with a particular focus on recent maxEnt models built using simple population level constraints such as the spike synchrony distribution [27\*]. Furthermore, we will argue that sampling from MaxEnt models provides generalizations of frequentist shuffles and null distributions for hypothesis testing [28\*].

### Basic principles for constructing MaxEnt models

The MaxEnt principle provides a compressed description of data that is maximally noncommittal with respect to missing (unspecified) information [24,25]. It reproduces exactly a set of summary statistics of the data but is otherwise maximally unstructured (see Box 1). To illustrate the basic idea, consider the example of some two-dimensional real-valued data, drawn from an unknown distribution  $\text{Prob}(x, y)$  (Figure 1a). If we constrain the description of the data to the mean and the standard deviation of  $x$  and  $y$ , then we approximate the original complex distribution by two independent normal distributions. This reduced model is actually a first-order, sometimes referred to as ‘factorized’, maximum entropy model of this data. Adding more constraints, for instance by additionally constraining the correlation of the two variables, brings the reduced model closer to the true distribution. In the neuroscience context, where the data is described by patterns of neural activity, the same principle applies, but the distributions now live in a much higher dimensional binary space (for  $N$  simultaneously recorded neurons —  $2^N$  possible activity patterns). The first-order maximum entropy model in this case constrains the mean firing rate of individual neurons. The pairwise model, often referred to as the Ising model, additionally constrains correlations for all possible neuron pairs. Other higher-order constraints are also possible (Figure 1b). The main difference is that while in the Gaussian example the map between the constraints (mean, covariance) and the model parameters is very simple, this computation is much more involved for the Ising model, or other generalizations. Nonetheless, once we have constructed

#### Box 1 Maximum entropy models

The maximum entropy principle provides a strategy for constructing a probability distribution from a limited set of data points: given a set of summary statistics of the data, we seek the distribution that matches these constraints exactly, and is otherwise maximally unstructured (has maximum entropy). MaxEnt examples include well-known distributions such as the Gaussian (real valued data, constrained by mean and variance), the exponential (positive reals, constrained by mean), or the Ising model (binary vectors, constrained by mean and covariance), a popular model for patterns of neural activity.

For a set of constraints expressed as a set of expectations  $E[f_i(\mathbf{x})] = c_i$ , the corresponding MaxEnt distribution takes the form  $P(\mathbf{x}) = \frac{1}{Z} \exp\left(\sum_i \alpha_i f_i(\mathbf{x})\right)$  with normalizing constant  $Z$  and parameters  $\alpha_i$ , uniquely determined by the data statistics  $c_i$  computed as some simple functions  $f_i$  of the data. For instance, for binary patterns of neural activity  $\mathbf{x}$ ,  $f_i(\mathbf{x}) = x_i$  and  $f_{ij}(\mathbf{x}) = x_i x_j$  constrains the mean firing rate and the covariance of neural responses;  $f_k(\mathbf{x}) = \delta(\sum_i x_i - k)$  ( $\delta(0) = 1$  and 0 otherwise) constrains the spike synchrony distribution  $P(\sum_i x_i = k)$ . When chained together, more constraints bring the model progressively closer to the data. Hence MaxEnt provides a principled way to build increasingly complex structure from relatively simple (and interpretable) component constraints.

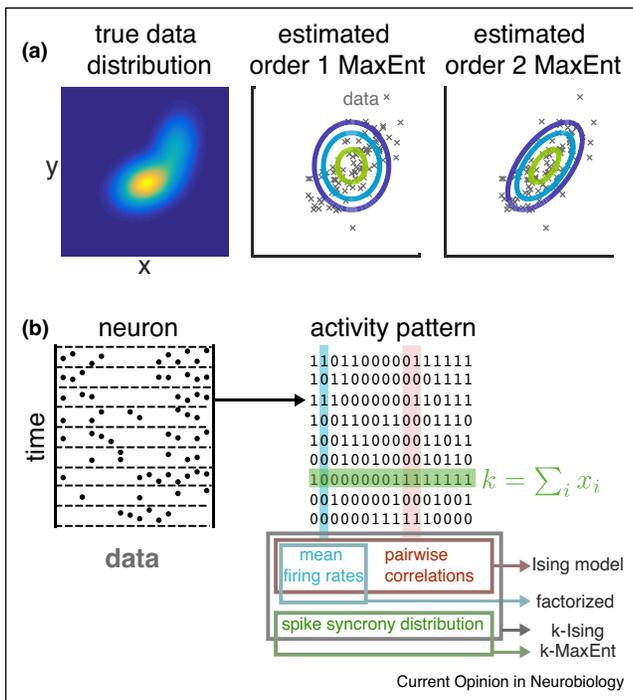
Estimating the parameters is usually done by maximum likelihood. Unfortunately, the map between parameters and the corresponding summary statistics requires the enumeration of all possible patterns. This severely restricts the number of neurons for which the sufficient statistics can be computed exactly, and additional approximations (for instance MCMC sampling) are required for more than  $N \sim 20$  neurons [57,51]. The issue of computational feasibility is particularly important in the context of large neural recordings where the practical applicability of the MaxEnt framework relies on the ability to estimate and sample quickly from the model. This motivates efforts to build tractable classes of MaxEnt models [27\*,50,48\*,51] and to use techniques that circumvent the parameter fitting procedure by directly sampling from an implicitly defined MaxEnt model (‘herding’) [58].

the MaxEnt model of the data we have a clear mathematical description for the probability of any activity pattern and related quantities. Additionally, we can generate new artificial data with the same statistics, by sampling from the model.

Traditionally the MaxEnt framework has been used as a prescription to build accurate probabilistic models of neural activity [3,29,30\*]. However, MaxEnt models can be useful even when they do not match the data well. In this case, they signal the fact that there are critical missing features in our description of the data. For instance, we can build a MaxEnt model consistent with the (stimulus dependent) firing rates of individual cells (Figure 2a) and find that it poorly explains observed population level statistics, such as the pattern of pairwise correlations (Figure 2b), or the spike synchrony distribution (Figure 2c). This suggests that there are significant dependencies in neural activity, beyond shared stimulus preference.

In the example above, the MaxEnt model consistent with the measured firing rate of individual cells (or their

Figure 1



The maximum entropy principle. **(a)** True data distribution from which a set of data points (gray crosses) are drawn; the corresponding first-(second-)order MaxEnt model is a Gaussian with the same mean and (co-)variance. **(b)** A summary of common constraints for MaxEnt models of neural activity.

tuning) corresponded simply with the independent (or conditionally independent) model of neural spiking. These can be directly obtained by appropriately shuffling the original data. For the independent model, one can break all the correlations between neurons by preserving the number of spikes for each neuron but randomly reassigning spike times; for the conditionally independent model, the same can be achieved by randomly permuting trial indices for each neuron (Figure 2a). More generally, neural activity shuffles usually have a direct correspondence to simple MaxEnt models. The crucial difference is that MaxEnt models provide a much richer set of null distributions, for which there exist no known shuffles. For instance, we can jointly constrain the tuning of individual cells and their total correlations (Figure 2b), something hard to achieve by a shuffling procedure. Iteratively adding more constraints results in a sequence of surrogate ensembles that are increasingly similar to the original data, but whose structure can be still traced back to a set of meaningful sufficient statistics. In this sense, MaxEnt shouldn't be seen as providing a single null distribution, but rather a hierarchy of null distributions that attempt to explain high-order statistical regularities in the data as resulting from an increasing number of simple statistics.

## Salient features of the neural code

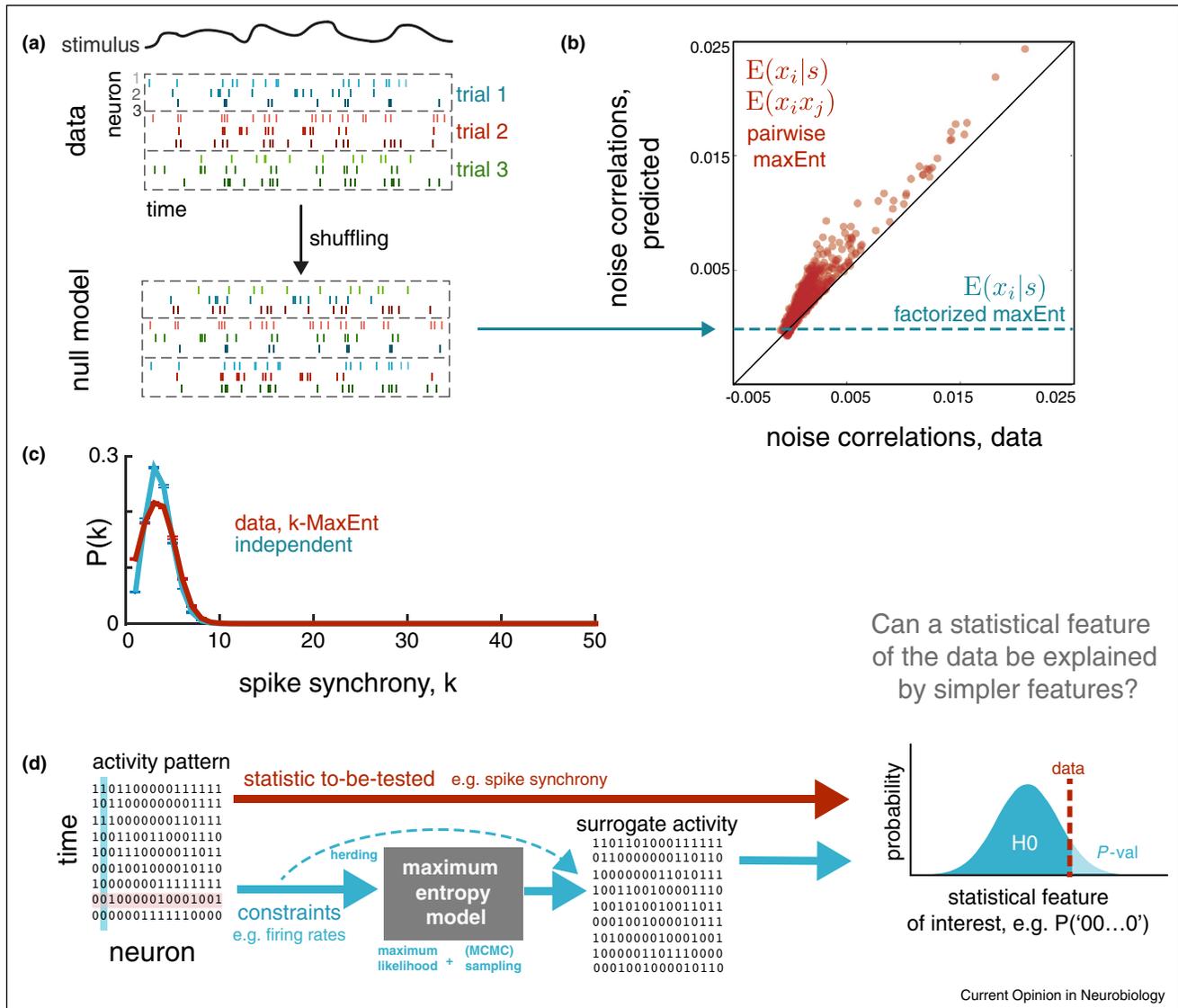
### Pairwise correlations

One key signature of the collective behavior is the presence of nontrivial pairwise correlations. These are particularly important as they constrain how much information can be encoded in neural activity and the neural readout required to retrieve it [31]. Additionally, since Hebbian forms of synaptic plasticity pick up and positively reinforce correlations in presynaptic and postsynaptic activity [32], changes in correlation structure are a natural measure of circuit-level effects of learning [33,34]. Because of this, an overwhelming part of the MaxEnt literature focuses on modeling and interpreting neural correlations (see also recent reviews [5,17]). The Ising model has proven successful in capturing the statistics of moderately sized subpopulation of neurons in the retina [3], in cortical activity [3], and the hippocampus [35]. However, there are aspects of population activity which pairwise modes do not capture well. For instance, higher order moments are required when considering larger populations (100 cells) [29]. Nonetheless, pairwise MaxEnt models have been instrumental in investigating the way neural populations encode information about sensory stimuli [36], in particular for modeling the distribution of neural responses conditioned on a stimulus [37,15]. These works have shown that while noise correlations can be small *on average*, they are strong precisely during those times when the circuit is highly active, thereby strongly reshaping the population response. These stimulus-dependent patterns of noise correlations organize neural responses into sets of semantically similar activity patterns [15], also reflected in overall probability distributions of activity patterns with multiple peaks [13,30]. These results raise the hypothesis that stimulus information is encoded not by precise microscopic patterns of neural activity, but rather by the attractor basins of the probability around these peaks (later tested by [14]).

### Constraints on global activity

Another interesting feature of cortical activity is the fact that the probability of quiescence (no neuron firing) is systematically higher in data compared to controls (independent, or Ising, see Figure 2c) [27,38,39,30,40]. Since this separation of periods of activity and silence, likely owed to local inhibitory influences, occurs across circuits and conditions, it is likely a salient property of the neural code. Adding this one simple constraint dramatically improves the ability of the factorized MaxEnt model to capture the neural activity statistics [40]. More generally, it is increasingly recognized that low-dimensional fluctuations in populations activity arising due to internal dynamics (oscillations, top-down modulatory influences, fluctuations in arousal or attention) explain an important part of higher order dependencies in neural activity [41–43]. Moreover, since global fluctuations in neural activity increase pairwise correlations, they need to be explicitly

Figure 2



MaxEnt models as controls. **(a)** Shuffle that destroys all dependencies but preserves neural tuning (equivalent to a stimulus dependent factorized MaxEnt models). **(b)** Discrepancies between measured noise correlations and those predicted by different MaxEnt models. **(c)** The frequency of population events involving a given number of spikes (spike synchrony distribution) for recorded CA1 neural activity compared to the corresponding factorized MaxEnt null model. **(d)** Hypothesis testing using MaxEnt models: the goal is to determine how likely it is that a particular statistical regularity of the data (e.g.  $P(00 \dots 0)$ ) could arise by chance, achieved by comparing data with an ensemble of surrogates obtained by sampling from the corresponding MaxEnt model. The test statistic is computed for each of these surrogate datasets and its distribution is compared to the value obtained from the data to assess significance.

controlled for when assessing and interpreting neural correlations in relation to learning [10,44\*,33].

The population spike distribution,  $P(k)$ , provides a powerful way to characterize global population dependencies. It is simple and computationally convenient because the number of constraints is linear in the number of neurons (versus quadratic for the Ising model). It is also complex because it specifies dependencies of all orders. Although

the spike synchrony distribution has long been identified as a salient feature of collective behavior in the hippocampus, its utility as a constraint for MaxEnt models has only been recognized recently [27\*,30\*]. Constraining  $P(k)$  alone captures a surprising range of features in real data [27\*]. Furthermore, it can naturally be combined with other constraints such as mean firing or even pairwise correlations. Studying  $K$ -pairwise models, that is maximum entropy models with pairwise correlations and  $P(k)$

Table 1

## Summary comparison of maximum entropy models

Model	Constraints	Parameters	References
Factorized	$\langle x_i \rangle$	$N$	[10]
Simultaneous silence	$\langle x_i \rangle, P(k=0)$	$N+1$	[40]
Population spike distribution	$P(k), k = \sum_i x_i$	$N+1$	[27*]
Dyn. population spike distribution	$P(k^t, k^{t-1}, \dots)$	$(N+1)^{Nr}$	[45]
Pairwise/Ising	$\langle x_i \rangle, \langle x_i x_j \rangle$	$\frac{N(N+1)}{2}$	[59,3,13]
Dyn. pairwise/Ising	$\langle x_i \rangle, \langle x_i^t x_j^t \rangle, \langle x_i^t x_j^{t-1} \rangle$	$N + N^2 + \frac{N(N-1)}{2}$	[53]
Stimulus dependent Ising	$\langle x_i   s \rangle, \langle x_i x_j \rangle$	$N_s N + \frac{N(N-1)}{2}$	[37]
Stimulus dependent correlations	$\langle x_i   s \rangle, \langle x_i x_j   s \rangle$	$N_s \frac{N(N+1)}{2}$	[15*]
Population tracking	$P(k), \langle x_i   k \rangle$	$N+1$	[50]
Generalized $k$ -modulation	$P(k, x_i)$	$N(N+1)$	[48*,49]
$K$ -Ising	$\langle x_i \rangle, \langle x_i x_j \rangle, P(k)$	$\frac{(N+1)(N+2)}{2}$	[30*]

constraint, has revealed that statistical dependencies reduce the entropy by a quarter from that of an factorized model in the retina, and that these dependencies are strong enough to predict the behavior of single neurons from the rest of the population, without any knowledge of the stimulus [30\*]. More controversially, it has been suggested that neural activity distributions are close to critical [27\*,8,45], and that this property may be linked to shared latent variables [46,47]. Recent MaxEnt models further link the population spike distribution with cell-specific features [48\*], reflecting the fact that the global network state can affect individual cells to different degrees (from ‘choristers’ to ‘soloists’) [49]. Furthermore, there exist analytically tractable versions of MaxEnt models incorporating these constraints [27\*,50,48\*,51], which makes them particularly attractive for characterizing large populations (Table 1).

One aspect that has so far been largely neglected in the MaxEnt context is the temporal structure of population activity. Attempts to model transition probabilities between binary words directly have so far proved impractical [52]. A potentially more tractable approach is to restrict the modeling of temporal structure to population-level state transitions of the form  $P(k_t | k_{t-1})$  [45], or to constrain time-delayed pairwise correlations [53,54]. These approaches have been used for the better statistical characterization of responses ganglion cells of the salamander retina [54], cortical and hippocampal slices in rats [55] and of neural activity in cat parietal cortex during sleep [53]. Much-needed time dependent MaxEnt variants could provide tractable null models for the temporal dynamics of neural activity, as an alternative to latent linear dynamical systems models [56].

### MaxEnt as null models

As the most random distribution that satisfies a chosen set of constraints, MaxEnt models naturally provide a null distribution for quantities that are not directly constrained. The idea of using MaxEnt models as null models for hypothesis testing traces back to work by Martignon [28\*], testing whether effects of a certain order can be explained in terms of lower moments (for instance if the occurrence of triplets of activity can be explained in terms of the firing rates and covariance structure of component neurons). For a particular MaxEnt model class (log-linear), confidence bounds can be computed analytically. Unfortunately, although mathematically elegant, this approach is computationally infeasible for realistic neural populations sizes. Still, the core idea can be adapted to any MaxEnt model, at the cost of having to compute the null distribution numerically. For instance, to determine if the frequency of quiescence in data can be explained by the firing of individual neurons (Figure 2d), one would construct a factorized MaxEnt null model. Repeatedly sampling from the MaxEnt model results in surrogate ensembles which are used to estimate the frequency of the all-zeros pattern when neurons fire independently (or any other test statistic). At this point a  $P$ -value can be estimated reporting the probability of observing the effect of the size of that observed in the real data (or larger) by chance. The overall procedure is very similar to traditional strategies for building confidence bounds by shuffling. The null model is somewhat stricter (as traditional shuffles typically preserve the total number of spikes in the data, while MaxEnt shuffles do not), but it has the advantage that the same recipe can be applied for complex null models that don’t have a corresponding known shuffling procedure. In general the framework allows the choice of

constraints and test statistic to be decided in a hypothesis-driven way.

The structure of the neural activity could have been assessed by directly fitting a MaxEnt model that also incorporates the statistic of interest as an additional constraint and then showing that the extended model does a better job at explaining the data. Nonetheless, the first approach may still be preferable in practice for computational reasons: depending on the hypothesis being tested, null models may be dramatically easier to fit than the alternative. This means that the frequentist approach can be used to test a broader set of hypotheses on real data. Lastly, increasing model complexity sequentially allows information gained at one step to inform subsequent models (as priors). For instance if we discover that residual correlations that cannot be explained by global constraints are sparse, then we can incorporate a sparse prior when building a new MaxEnt model that explicitly constrains pairwise correlations. Thus, the MaxEnt framework provides a systematic way to dissect the neural code.

### Conclusions and future directions

Here we have argued that the MaxEnt framework provides general principles for constructing control ensembles for neural activity. The same principles can also be used for identifying the key regularities captured by different models. In particular, MaxEnt models could serve as a baseline comparison to increasingly popular unsupervised models from machine learning (e.g. restricted Boltzmann machines, hidden Markov models, deep nets). Although these popular techniques have more expressive power, that is, they can model complex distributions over activity patterns given sufficient data, interpreting their parameters can be difficult. Comparing their performance to simpler MaxEnt models is useful not for answering the question of ‘Which model is better,’ but to identify the additional statistical structure captured by the more expressive model. The maximum entropy principle can guide our quest for ‘the simplest possible, but not simpler’ model of neural activity, by providing a yardstick to make the necessary distinction.

### Conflict of interest statement

Nothing declared.

### Acknowledgements

This work was supported by the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme (FP7/2007–2013) under REA grant agreement no. 291734.

### References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest

1. Yuste R: **From the neuron doctrine to neural networks.** *Nat Rev Neurosci* 2015, **16**:487–497.
2. Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W: *Spikes: Exploring the Neural Code.* Cambridge, MA, USA: MIT Press; 1999.
3. Schneidman E, Berry MJ, Segev R, Bialek W: **Weak pairwise correlations imply strongly correlated network states in a neural population.** *Nature* 2006, **440**:1007–1012.
4. Pillow JW *et al.*: **Spatio-temporal correlations and visual signalling in a complete neuronal population.** *Nature* 2008, **454**:995–999.
5. Roudi Y, Dunn B, Hertz J: **Multi-neuronal activity and functional connectivity in cell assemblies.** *Curr Opin Neurobiol* 2015, **32**:38–44.
6. Schneidman E, Bialek W, Berry MJ: **Synergy, redundancy, and independence in population codes.** *J Neurosci* 2003, **23**:11539–11553.
7. Schneidman E *et al.*: **Synergy from silence in a combinatorial neural code.** *J Neurosci* 2011, **31**:15732–15741.
8. Tkačik G *et al.*: **Thermodynamics and signatures of criticality in a network of neurons.** *Proc Natl Acad Sci U S A* 2015, **112**:11508–11513.
9. Sadtler PT *et al.*: **Neural constraints on learning.** *Nature* 2014, **512**:423–426.
10. Berkes P, Orban G, Lengyel M, Fiser J: **Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment.** *Science* 2011, **331**:83–87.
11. Trousdale J, Hu Y, Shea-Brown E, Josić K: **Impact of network structure and cellular response on spike time correlations.** *PLoS Comput Biol* 2012, **8**:e1002408.
12. Helias M, Tetzlaff T, Diesmann M: **The correlation structure of local neuronal networks intrinsically results from recurrent dynamics.** *PLoS Comput Biol* 2014, **10**:e1003428.
13. Tkacik G, Schneidman E, Berry MJ II, Bialek W: *Ising Models for Networks of Real Neurons.* 2006. arXiv preprint q-bio/0611072.
14. Prentice JS *et al.*: **Error-robust modes of the retinal population code.** *PLoS Comput Biol* 2016, **12**:e1005148.
15. Ganmor E, Segev R, Schneidman E: **A thesaurus for a neural population code.** *eLife* 2015, **4**:e06134.  
The paper uses a stimulus-dependent MaxEnt model to show that retinal responses are organized in clusters of activity patterns, which share similar meaning but may differ significantly in their structure.
16. Doiron B, Litwin-Kumar A, Rosenbaum R, Ocker GK, Josić K: **The mechanics of state-dependent neural correlations.** *Nat Neurosci* 2016, **19**:383–393.  
The paper highlights the link between mechanistic models of circuit dynamics and corresponding statistical regularities in neural activity, and the utility of the latter for validating models of neural computation.
17. Schneidman E: **Towards the design principles of neural population codes.** *Curr Opin Neurobiol* 2016, **37**:133–140.
18. Lewis CM, Bosman CA, Fries P: **Recording of brain activity across spatial scales.** *Curr Opin Neurobiol* 2015, **32**:68–77.
19. Okun M, Lak A, Carandini M, Harris KD: **Long term recordings with immobile silicon probes in the mouse cortex.** *PLOS ONE* 2016, **11**:e0151180.
20. Stevenson IH, Kording KP: **How advances in neural recording affect data analysis.** *Nat Neurosci* 2011, **14**:139–142.
21. Gao P, Ganguli S: **On simplicity and complexity in the brave new world of large-scale neuroscience.** *Curr Opin Neurobiol* 2015, **32**:148–155.
22. Cunningham JP, Byron MY: **Dimensionality reduction for large-scale neural recordings.** *Nat Neurosci* 2014, **17**:1500–1509.
23. Freeman J: **Open source tools for large-scale neuroscience.** *Curr Opin Neurobiol* 2015, **32**:156–163.
24. Jaynes ET: **Information theory and statistical mechanics.** *Phys Rev* 1957, **106**:620.

25. Wainwright MJ, Jordan MI *et al.*: **Graphical models, exponential families, and variational inference.** *Found Trends Mach Learn* 2008, **1**:1-305.
26. Phillips SJ, Anderson RP, Schapire RE: **Maximum entropy modeling of species geographic distributions.** *Ecol Model* 2006, **190**:231-259.
27. Tkačik G, Marre O, Mora T, Amodei D: **The simplest maximum entropy model for collective behavior in a neural network.** *J Stat Phys* 2013, **3**:3-11.  
The first paper to introduce the spike synchrony distribution as a constraint in a maximum entropy model.
28. Martignon L, Von Haselein H, Grün S, Aertsen A, Palm G: **Detecting higher-order interactions among the spiking events in a group of neurons.** *Biol Cybern* 1995, **73**:69-81.  
Key paper introducing the use of log-linear MaxEnt as null models for hypothesis testing.
29. Ganmor E, Segev R, Schneidman E: **Sparse low-order interaction network underlies a highly correlated and learnable neural population code.** *PNAS* 2011, **108**:9679-9684.
30. Tkacik G *et al.*: **Searching for collective behavior in a large network of sensory neurons.** *PLoS Comput Biol* 2014, **10**:e1003408.  
An early example for the use of MaxEnt models as a way systematically dissect the neural code, this paper is a precursor to much of the recent MaxEnt results on global activity modulation, macroscopic structure of the neural code.
31. Cohen MR, Kohn A: **Measuring and interpreting neuronal correlations.** *Nat Neurosci* 2011, **14**:811-819.
32. Abbott LF, Nelson SB: **Synaptic plasticity: taming the beast.** *Nat Neurosci* 2000, **3**:1178-1183.
33. Fiser J, Lengyel M, Savin C, Orban G, Berkes P: **How (Not) to Assess the Importance of Correlations for the Matching of Spontaneous and Evoked Activity.** 2013 arXiv:1301.6554.
34. Modi MN, Dhawale AK, Bhalla US: **CA1 cell activity sequences emerge after reorganization of network correlation structure during associative learning.** *eLife* 2014, **3**:e01982.
35. Dunn B, Morreaunet M, Roudi Y: **Correlations and functional connections in a population of grid cells.** *PLoS Comput Biol* 2015, **11**:e1004052.
36. Fairhall A, Shea-Brown E, Barreiro A: **Information theoretic approaches to understanding circuit function.** *Curr Opin Neurobiol* 2012, **22**:653-659.
37. Granot-Atedgi E, Tkacik G, Segev R, Schneidman E: **Stimulus-dependent maximum entropy models of neural population codes.** *PLoS Comput Biol* 2013, **9**:e1002922.
38. Amari SI, Nakahara H, Wu S, Sakai Y: **Synchronous firing and higher-order interactions in neuron pool.** *Neural Comput* 2003, **15**:127-142.
39. Ohiorhenuan IE *et al.*: **Sparse coding and high-order correlations in fine-scale cortical networks.** *Nature* 2010, **466**:617-621.
40. Shimazaki H, Sadeghi K, Ishikawa T, Ikegaya Y, Toyozumi T: **Simultaneous silence organizes structured higher-order interactions in neural populations.** *Sci Rep* 2015, **5**:9821.
41. Goris RLT, Movshon JA, Simoncelli EP: **Partitioning neuronal variability.** *Nat Neurosci* 2014, **17**:858-865.
42. Pachitariu M, Lyamzin DR, Sahani M, Lesica NA: **State-dependent population coding in primary auditory cortex.** *J Neurosci* 2015, **35**:2058-2073.
43. Rabinowitz NC, Goris RL, Cohen M, Simoncelli EP: **Attention stabilizes the shared gain of V4 populations.** *eLife* 2015, **4**:e08998 <http://dx.doi.org/10.7554/eLife.08998>.
44. Okun M *et al.*: **Population rate dynamics and multineuron firing patterns in sensory cortex.** *J Neurosci* 2012, **32**:17108-17119.  
The first paper to point out the importance of taking into account network fluctuations when assessing changes in population statistics due to learning.
45. Mora T, Deny S, Marre O: **Dynamical criticality in the collective activity of a population of retinal neurons.** *Phys Rev Lett* 2015, **114**:078105.
46. Schwab DJ, Nemenman I, Mehta P: **Zipf's law and criticality in multivariate data without fine-tuning.** *Phys Rev Lett* 2014, **113**:068102.
47. Humplik J, Tkačik G: *Semiparametric Energy-Based Probabilistic Models.* 2016 arXiv:1605.07371.
48. Gardella C, Marre O, Mora T: **A tractable method for describing complex couplings between neurons and population rate.** *eNeuro* 2016, **3**.  
Highly scalable version of a MaxEnt model with population constraints. It also captures cell-specific degrees of modulation of neural activity by the global network state.
49. Okun M *et al.*: **Diverse coupling of neurons to populations in sensory cortex.** *Nature* 2015, **521**:511-515 <http://dx.doi.org/10.1038/nature14273>.
50. O'Donnell C, Goncalves JT, Whiteley N, Portera-Cailliau C, Sejnowski TJ: **The population tracking model: a simple, scalable statistical model for neural population data.** *Neural Comput* 2017, **29**:50-93.
51. Nonnenmacher M, Behrens C, Berens P, Bethge M, Macke JH: *Signatures of Criticality Arise in Simple Neural Population Models with Correlations.* 2016. arXiv:1603.00097, <https://github.com/mackelab/CorBinian>.
52. Donner C, Obermayer K, Shimazaki H: **Approximate inference for time-varying interactions and macroscopic dynamics of neural populations.** *PLoS Comput Biol* 2017, **13**:e1005309.
53. Marre O, El Boustani S, Frégnac Y, Destexhe A: **Prediction of spatiotemporal patterns of neural activity from pairwise correlations.** *Phys Rev Lett* 2009, **102**:138101.
54. Vasquez JC, Marre O, Palacios AG, Berry MJ, Cessac B: **Gibbs distribution analysis of temporal correlations structure in retina ganglion cells.** *J Physiol (Paris)* 2012, **106**:120-127.
55. Tang A *et al.*: **A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro.** *J Neurosci* 2008, **28**:505-518.
56. Macke JH, Buesing L, Sahani M: **Estimating state and parameters in state space models of spike trains.** *Advanced State Space Methods for Neural and Clinical Data* . 2015.
57. Maoz O, Schneidman E: *maxent\_toolbox: Maximum Entropy Toolbox for Matlab, Version 1.0.2.* 2017 <http://dx.doi.org/10.5281/zenodo.191625>. [https://orimaoz.github.io/maxent\\_toolbox](https://orimaoz.github.io/maxent_toolbox).
58. Welling M: **Herding dynamical weights to learn.** In *Proceedings of the 26th Annual International Conference on Machine Learning.* ACM; 2009:1121-1128.
59. Shlens J *et al.*: **The structure of multi-neuron firing patterns in primate retina.** *J Neurosci* 2006, **26**:8254-8266.