Article

A Comparison of the Maximum Entropy Principle Across Biological Spatial Scales

Rodrigo Cofré [1], Rubén Herzog [2], Derek Corcoran [3,4], Fernando Rosas [5,6,7]

[1]   CIMFAV, Facultad de Ingeniería, Universidad de Valparaíso, Valparaíso, Chile; rodrigo.cofre@uv.cl
[2]   Centro Interdisciplinario de Neurociencia de Valparaíso, Universidad de Valparaíso, Valparaíso, Chile; rubenherzog@ug.uchile.cl
[3]   Departamento de Ecología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile.
[4]   Instituto de Ecología y Biodiversidad, Santiago, Chile; dacorcov@uc.cl.
[5]   Centre for Psychedelic Research, Department of Medicine, Imperial College London, London, UK; f.rosas@imperial.ac.uk
[6]   Centre for Complexity Science and Department of Mathematics, Imperial College London, London, UK
[7]   Data Science Institute, Imperial College London, London, UK
[*]   Correspondence: rodrigo.cofre@uv.cl

1   **Abstract:** Despite their obvious differences, biological systems at different scales tend to exhibit
2   common organizational patterns. Unfortunately, these commonalities are usually obscured by the
3   parcelled terminology employed by various scientific sub-disciplines. To explore these commonalities,
4   this papers a comparative study of diverse applications of the maximum entropy principle, ranging
5   from amino acids up to societies. By presenting these studies under a common language, this paper
6   establishes a unified view over seemingly highly heterogeneous biological scenarios.

7   **Keywords:** maximum entropy principle; biological systems across scales; model-free data analysis;
8   inverse problems

1. Introduction

10   While scientific endeavor is traditionally associated with the *divide et impera* motto, the last decades
11   have witnessed a shift in many areas of research towards considering the properties of collections of
12   interacting systems such as cells, circuits of neurons, brains, species, and ecosystems [1]. This interest
13   is fostered by the growing understanding that "more is different" [2], i.e. that many of these systems
14   exhibit emergent properties that cannot be explained by the nature of their parts in isolation. Another
15   driver of this shift is the increasing amount of data available for analysis, which is enabled by novel
16   recording techniques, and the advances in technologies for information storage and transfer [3].
17   Biological research is nowadays in a peculiar situation: while there is more data available than
18   ever before, spanning all spatial biological scales, there is still a lack of an operational theory to explain
19   the collective behavior of living organisms at different scales. In order to help to build a road towards
20   such principles, an attractive method is to employ data-driven modeling techniques from the statistics
21   literature that – due to their generality – can be applied in diverse biological scenarios. While this way
22   of proceeding might go against traditional wisdom based on mechanistic considerations, a number of
23   reports have shown that collective behavior can be accurately characterized by statistical models built
24   directly from data [4–6]. Importantly, these studies show that one might successfully build models
25   *without the need to characterize the mechanistic interactions nor biological processes from first principles.*

The maximum entropy principle (MEP) is one of the statistical methods that has found applications over a surprisingly wide range of biological scenarios. The core idea of the MEP is to build statistical models that agree with data, but are otherwise as "structureless" as possible; said differently, the MEP provides a method to find the least biased model that is consistent with the experimental data [7]. The initial success of the MEP method in physics and engineering rapidly triggered a plethora of applications in biology, including DNA motifs of transcription factor binding sites [8], co-variations in protein families and amino acid contact prediction [9–11], diversity of antibody repertoires in the immune system [12,13], collective activity of neural populations [5,6,14–19], collective behavior of bird flocks [20,21], collective behavior in groups of mice [22], and ecology of abundance and distribution [23,24]. The fact that the MEP has been successfully applied in these highly heterogeneous scenarios suggests that there might exist interesting organizational commonalities across biological scales. However, these commonalities are obscured by the parcelled language and terminologies employed in the various sub-disciplines of biology, which makes comparative studies highly non-trivial.

In order to help the exploration of commonalities, in this article, we provided a unified review of recent advances related to the applications of the MEP across biological scales. There are recent reviews related to the MEP applied to different fields of biology, with a focus on e.g. parameter inference [25–27], reverse engineering [28], learning of hidden variables [29] and information-processing in biophysical systems [30]. To complement this literature, here our goal is to provide a comparative study of applications of the MEP accross biological spatial scales, providing a unified formalism, perspective, and notation that can bridge the differences between various scientific sub-fields. We compare and highlight some differences, extensions, and limitations of the MEP approach, and discuss open challenges for future research.

2. Maximum Entropy Principle: ideas and approaches

When studying living systems from experimental data, scientist are usually unable to access all the relevant information that would be required to fully characterize the system of interest. This limitation seems not to be a technological issue but a characteristic of biology (at least in the foreseeable future). For example, it is unlikely to be able to simultaneously measure the firing patterns of each of the $\approx 10^{11}$ neurons in the human brain, or quantify and classify all the insects that live at the Amazonian forest at a given time. Despite this limitation, from experimental data is usually possible to obtain accurate estimations of global properties, e.g. the average values of certain quantities of interest. Therefore, it is often relevant to find models consistent with this accurate – but partial – information. Unfortunately, there is usually an infinite number of statistical models that are consistent with these global properties measured from experimental data, and hence one needs additional criteria to decide which one to use.

The MEP provides a rational basis to choose statistical models in these situations. In particular, the MEP is special for being maximally noncommittal with regard to missing information [7]. The core of the method is based on a constrained optimization problem of a concave functional – the Shannon entropy, resulting in a unique probability distribution that is consistent with the partial information at hand, being otherwise as "random" as possible.

In the rest of this section, we introduce the MEP from a broader perspective, i.e., as an inverse statistical mechanics problem. After we introduce the basic building blocks of this principle, we set the notation and ideas to unify the approach presented in the examples.

2.1. Forward versus inverse problems

The idea of maximizing entropy has its origin in thermodynamics and statistical mechanics. In these branches of physics, there are two opposite approaches with which one can build statistical models to characterize the phenomena under study. One is assuming complete knowledge of the relevant mechanistic interactions that rules the constituents of the system, which is called *forward modeling*. This approach builds a probabilistic model, which in turn determines values for the average

of various observables of the system. The second approach works backward, hence is called *inverse modeling*: one uses experimental data to determine the average value of various observables, and then build the "less structured" model that is consistent with those values.

2.1.1. Forward approach

The primary goal of statistical physics is to derive observable quantities from microscopic laws governing the parts and interactions of a system. The standard way in which these ideas are introduced is through *Hamiltonian models* describing from first principles the interactions of a system. This approach is often called "forward", and the goal is to characterize observables representing collective phenomena such as spin magnetization, correlations, or to characterize phase transitions as a function of the physical parameters of the model.

2.1.2. Inverse approach

In statistical mechanics, the inverse problem starts taking average values of observables from experimental data generated by some system. The goal of this approach is to infer the parameters of a candidate Hamiltonian (local rules of interactions) characterizing the system only from experimental data. Although some branches of biology have rich theories [31], there are other domains for which no mechanistic approach exists, in those cases maximum entropy is usually among the best alternatives. This is particularly relevant nowadays as in almost all branches of biology at all scales there is experimental data to "recover" the parameters or interactions of a candidate Hamiltonian. In general, inverse problems are hard to solve; in particular, the application of the maximum entropy principle to biological experimental data usually depends on sophisticated numerical algorithms and computational power [5,6,18,32].

2.2. Maximum entropy principle: definitions and methods

The MEP is an inverse problem commonly employed in statistical mechanics, but has found useful applications in several other probabilistic systems. This approach can reconstruct local rules of interaction from the experimental data, this without adopting particular mechanistic assumptions. While the MEP can take more general form [33], this article focuses on the standard approach of focusing only on average values computed from data. In the next sequel, the basic building blocks of the MEP are explained, while providing an unified notation that will be use throughout the article.

2.2.1. State space, observables, and average values

Suppose that a scientists is interest in a particular system, from which she only have experimental data but otherwise no additional knowledge. The first step in building the maximum entropy model of the system is to describe the set of possible configurations, which is called *state space*. For concreteness, let us consider a system made of N sub-units $\mathbf{x} = (x^1, \dots, x^N)$, where each coordinate $x^i \in \chi^i$ represent the state of each sub-unit. The configuration space is denoted by $\chi = \chi^1 \times \chi^2 \cdots \times \chi^N$. It is important to note that the state space grows exponentially with the number of sub-units. Therefore, when consider systems composed by many sub-units, the configuration space is usually too large to be characterized directly from experimental data, i.e. the cardinality of the state space is usually much larger than the number of data samples.

Having a clear idea of the state space of the system, the second step is to choose the *observables* of the system. Observables are real-valued functions on the configuration space of the form:

$$f \colon \chi \to \mathbb{R}$$
$$\mathbf{x} \mapsto f(\mathbf{x}) \, .$$

Observables are random variables whose *average values* can be estimated from data. With the state space defined and the estimation of basic statistical features with sufficient accuracy, the scene is set to build the minimally structured model that is consistent with these measurements.

2.2.2. Entropy maximization under constraints

Although the concept of *entropy* was first used by Rudolf Clausius in the field of thermodynamics to study the relationship between energy and temperature, the *Shannon entropy* [34] has a much broader scope dealing with the notions of information and uncertainty. Mathematically, for a discrete random variable with discrete probability distribution q over the state space χ, its entropy is

$$\mathcal{S}[q] = - \sum_{\mathbf{x} \in \mathcal{X}} q(\mathbf{x}) \log q(\mathbf{x}). \tag{1}$$

The notion of entropy can be also applied to continuous variables via the *differential entropy* [35]. In this case, sums over the discrete state space are replaced by integrals:

$$\mathcal{S}[q] = - \int q(\mathbf{x}) \log q(\mathbf{x}) d\mathbf{x} \tag{2}$$

The Shannon entropy can be used to generalize the *principle of insufficient reason* and Laplace's *principle of indifference*. The basic formulation of these principles says that, if one is interested in assigning probabilities to events but there is no further information about them, one should model them using the uniform distribution – which is the distribution that maximizes entropy. The MEP generalize this to scenarios where there is some information about the the system is available, by stating that one should choose the distribution that maximizes the entropy while being consistent to what is known.

The inputs for the MEP are the average value of a set of observables, which represent the knowledge obtained from the data. As the empirical average of observables are usually not enough to uniquely determine a probability distribution, the MEP is used to obtain the unique probability measure p that maximizes the entropy among all the probability measures q that match the expected values of all the observables.

The MEP can be stated mathematically as the following optimization problem:

$$\max_{q \in \mathcal{M}} \quad \mathcal{S}[q]$$

$$\text{subject to} \quad \mathbb{E}_q[f_k] = \langle f_k \rangle_e = c_k, \quad \forall k \in \{1, \dots, K\}$$

where $\mathcal{M}$ is the set of probability measures, $\mathbb{E}_q[f_k]$ is the average of the observable f_k for all $k \in \{1, \dots, \mathcal{K}\}$ with respect to q, $\langle f_k \rangle_e$ is the empirical average of f_k obtained from experimental data. This problem can be solved analytically (see A).

3. Examples at different Spatial Scales

One of the most powerful features of the MEP is its generality, which enables its use over a broad range of scenarios. This section explores six case studies of application of the MEP in biology at different spatial scales, this employing an unified methodology and notation. The cases are the following: amino acids in proteins (Section 3.1), retinal ganglion cells (Section 3.2), whole brain networks (Section 3.3), plant communities (Section 3.4), macroecologic biodiversity (Section 3.5), and human vote interactions in the US Supreme Court (Section 3.6). Rather than reviewing many articles related to each scale, we follow particular articles that summarize well the application of the MEP at that particular spatial scale. For each scenario we describe the state space, observables, and the inferred information.

3.1. Amino acid interactions in proteins

Proteins are essential molecules for living systems. These are sequences of amino acids, however, is their three-dimensional structure (tertiary structure) which is critical to determine its biological function. An example are the *homolog* proteins, i.e., proteins with a common evolutionary ancestor, where their tertiary structure and biological function is highly conserved, despite the amino acid sequence may differ in many ways. A family of homolog proteins can be summarized in the so called Multiple Sequence Alignment (MSA), a matrix where all the sequences belonging to the same protein family are aligned using sophisticated sequence similarity methods [36,37]. Nowadays, huge databases of MSA for different protein families are available [38] based on the growing number of complete sequenced genomes [39]. This technological advance embrace the challenge of the determination of protein tertiary structure from the linear amino acid sequence (from a statistical point of view). In fact, if the protein function highly depends on the tertiary structure and its conservation depends on the interactions between amino acid residues, then it is expected that some sites of the sequence co-evolve.

Here, we present the results in Cocco et al [40], where the authors show a wide variety of applications of MEP to the analysis of MSA.

3.1.1. State Space

Consider a protein family composed of M homolog protein sequences (either from different species or different pathways inside the same species), each with aligned length L. In order to align sequences of different lengths, a gap, $'-'$, may be introduced in some sequences. Given that there are $r = 21$ possible values for each sequence site (20 amino acids plus a gap, $\chi^i = \{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y, -\}$). A MSA can be represented as a $L \times M$ matrix (see figure 1), where the sequence of each protein is a L-vector and each sequence site l corresponds to one of the 21 amino acids from the χ^i set. However, for technical reasons, Cocco et al [40] use a binary embedding, where each sequence is represented as a binary Lr vector, where the i-th entry of each subvector is 1 if the corresponding i-th amino acid of the χ^i set is present on that site and 0 otherwise. Thus, the state space is conformed by all the possible amino acid sequences of length L, i.e., $\chi = \{0,1\}^{Lr}$.

3.1.2. Observables and average values

The observables are single site occurrence and pairwise occurrence. Their average values consider both the single site averages and the pairwise site-correlations statistics, this is:

- $\langle f_i(a) \rangle = \langle a^i \rangle$ is the average occurrence of the amino acid a at the $i - th$ sequence site.
- $\langle f_{i,j}(a, b) \rangle = \langle a^i b^j \rangle$ is the average co-occurrence of the amino acids a at the $i - th$ site and b in the $j - th$ site.
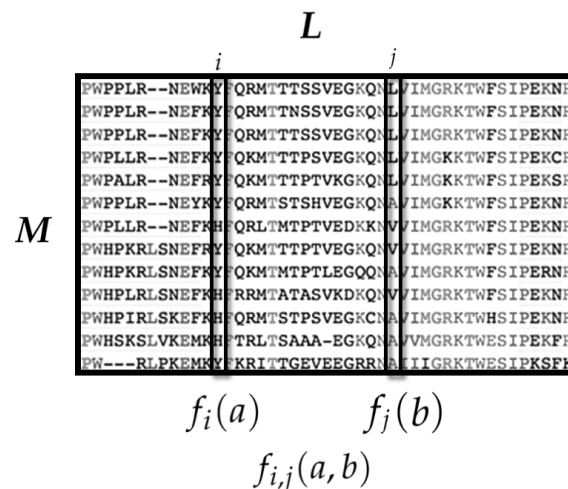
Figure 1. The amino acid sequences of M homolog proteins are aligned in the Multiple Sequence Alignment (MSA). Each MSA column is a sequence site and each row is the sequence of a member of the protein family. In order to have a fix sequence length L, a gap $('-')$ may be introduced. From the MSA, two sets of observables are considered: i) $f_i(a)$ is the occurence of the amino acid a at the site i and ii) $f_{ij}(a,b)$ is the co-ocurrence of amino acid a at the site i and amino acid b at the site j.

3.1.3. Inferred Information

The statistical models built from the MEP using MSA which consider single site and pairwise site-correlations yields valuable information about protein tertiary structure, its function and protein design. The MEP approach represent an appealing alternative to computationally expensive molecular dynamics simulations. Among the inferred information is the following.

1. **Co-evolving site pairs:** The interaction strength between site i and site j can be obtained as a function of the model parameters $J_{i,j}(a,b)$, i.e. the interaction between amino acid a in site i with the amino acid b in site j. Thus, a *coupling strength* is obtained, identifying evolutionary constraints on the protein family site-interactions.
2. **Contact Prediction:** The protein tertiary structure is associated to a topology of contacts between far amino acids residues. This topology can also be inferred as a function of $J_{i,j}(a,b)$. For predicting the tertiary structure of proteins, interactions between sites with a minimum separation of five sites on the linear sequence are studied (equivalent to one turn in an α-helix). The MEP approach outperforms the pairwise site contact prediction compared to standard correlation-based methods (e.g. mutual information).
3. **Protein Design and the Effect of Mutations:** According to the energy landscape theory of protein folding [41], proteins conserved along evolution tend to minimize their free energy in their folded state. Using the MEP, the energy of each amino acid sequence can be computed (see equation A4), which allow scoring each sequence according to their energy. This results in a set of non-naturally occurring proteins that minimizes the energy and, possibly, preserve the same functions than the original protein family. This inferred information has been applied to test and predict the effect of mutations [42,43].

3.2. Retina

The retina is a part of the brain which is located in the back part of the eye. Its main function is to encode different aspects of the visual stimulus and convey information to the retinorecipients brain areas about the visual stimulus: size, color, and movement through the retinal ganglion cells as sequences action potentials (spikes) and silences. The integration and transformation of the stimulus from retinal ganglion cells constitutes the first stage of our visual perception. Multi-electrode arrays

203 allow to record spikes from hundreds of retinal ganglion cells responding simultaneously to light
204 stimuli. Since the seminal work of Schneidman *et al* [14] and Pillow *et al* [44] to characterize the
205 spike train statistics of the vertebrate retina responding to natural stimuli, a succession of research
206 efforts have helped to better understand some aspects of the so called "retinal code" (the input-output
207 relationship). Since then, the MEP approach has become a standard tool to build statistical models
208 in this field [5,14,44–46], and it is more and more clear that genuine collective behavior in the retinal
209 network can be characterized using the MEP from relatively weak correlations among pairs of neurons
210 (Ising model) [4].

211   Here we present results reported in Ref. [5], where the authors build an accurate maximum
212 entropy model that match the firing rate of each retinal ganglion cell, their pairwise correlations, and
213 the distribution of summed spikes in the network at each time bin.

214 3.2.1. State Space

215   Consider a time discretization or "binning"Δt_b, in which for each time bin, each neuron can take
216 only two values, either zero or one. Consequently, in a network of N neurons, we denote x_t^k the binary
217 variable which take value 1 whenever the k-th neuron emits a spike during the t-th time bin, and 0
218 otherwise. Therefore $\chi^i = \{0,1\}$ for all $i = 1,...,N$. This standard procedure transforms experimental
219 data into sequences of binary patterns **(see figure 2)**. A *spike pattern* is the spike-state of all the network
220 at time bin t, denoted by $x_t := [x_t^n]_{n=1}^N$. Finally, a *spike train* or data-set is a finite sequence of spiking
221 patterns. The state space is formed by all the spike patterns $\chi = \{0,1\}^N$.

222 3.2.2. Observables and average values

223   The following is the list of observables and their average values used in this study, where $\langle \cdot \rangle$
224 means temporal average taken from data **(see figure 2)**:

225 • $\langle f_i(\mathbf{x}) \rangle = \langle x^i \rangle$ firing rate of neuron i, for all neurons.
226 • $\langle f_{ij}(\mathbf{x}) \rangle = \langle x^i x^j \rangle$ synchronous pairwise correlation between neuron i and neuron j, for all pairs
227   of neurons.
228 • $\langle f_K(\mathbf{x}) \rangle = \langle \delta(\sum_{i=1}^N x_t^i, K) \rangle$, for $K = 1,..,N$. Where δ represent the Kronecker delta.

229 3.2.3. Inferred Information

230   The inferred maximum entropy distribution is used to investigate the *principles underlying*
231 *population coding*, i.e., how groups of neurons orchestrate their spiking activity characterizing the
232 neural code. In this article, the authors use the maximum entropy distribution $p(\mathbf{x})$ to investigate the
233 following:

234 1. **The entropy:** To characterize the size of the neural vocabulary (the effective number of
235   configurations is reduced to 2^S) (1). This quantity constrains the capacity of the population
236   to represent visual information. In this case shows that the spike patterns are extremely
237   inhomogenous.
238 2. **Classification of activity patterns into set of meta-stable collective modes:** The energy
239   landscape inferred from the maximum entropy method presents valleys which resembles a
240   "clustering of patterns" of neural activity, but obtained without a particular metric for similarity
241   among patterns.
242 3. **Redundancy:** From the inferred joint distribution $p(\mathbf{x})$ the authors compute the conditional
243   marginal distributions $p(x^i = 1 \mid \mathbf{x}_{\backslash i})$, where $\mathbf{x}_{\backslash i}$ means all $\mathbf{x}$ except i. They show that the state of
244   individual neurons is highly predictable from the rest of the population, characterizing in this
245   way the level of redundancy. This property would allow error correction.
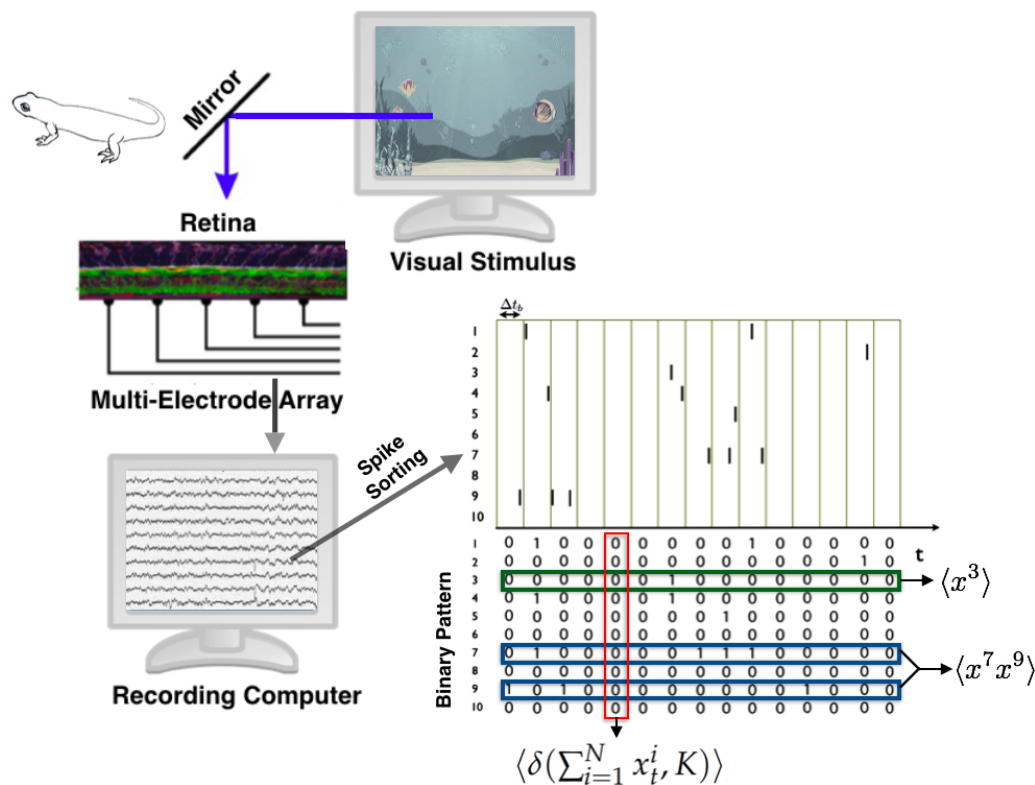
Figure 2. The retina of a vertebrate animal is extracted and mounted on the Multi-Electrode array in order to obtain the extracellular potential of the retinal ganglion cells responding simultaneously to natural stimuli. A signal processing procedure called spike sorting leads to the detection of spikes of each cell. A binning procedure is applied to obtain binary patterns of activity, form which the average values of the observables are computed.

3.3. Resting State Networks in the Human Brain

As presented previously, the pairwise maximum entropy model (MEM) accurately describes firing patterns in the retinal ganglion cells [4,5], but additionally firing patterns and local field potentials (LFPs) in human cortical tissues in vitro [15] and large-scale firing patterns in the visual cortex of monkeys and cats in vivo [47,48]. These findings suggest the idea that the human brain activity patterns during rest may accurately be described by pairwise MEMs.

Here we discuss results reported in Ref. [49]. In this article, the authors study spontaneous brain activity (in the absence of a task, but awake) using functional magnetic resonance imaging (fMRI) data. This technique has revealed that different brain regions interact with each other during rest forming several resting-state networks (RSNs) [50]. The RSNs, including the default mode network (DMN) and the fronto-parietal network (FPN), are highly reproducible across different healthy individuals and are considered to underlie cognitive processes.

3.3.1. State Space

Blood Oxygen Level-Dependent (BOLD) signals are extracted from anatomically defined regions. They consider 12 regions for the DMN and 11 regions for the FPN and binarized data. To binarize, the authors choose a threshold meaning that, at each time step, if the BOLD signal is below the threshold is considered 0 (inactive) otherwise 1 (active) (see figure 3). The threshold is set to maximize the accuracy of fit of the maximum entropy distribution.

264 The authors consider two separate data sets, one for the DMN and one for the FPN. Each data set
265 is a sequence of patterns of zeros and ones. The state space as in the case of retinal ganglion cells is
266 $\chi = \{0,1\}^N$.

267 3.3.2. Observables and average values

268 The following is the list of observables and average values used in the article (see figure 3):

269 • $\langle f_i(\mathbf{x}) \rangle = \langle x^i \rangle$ activation rate of region i, 12 for the DMN and 11 for the FPN.
270 • $\langle f_{ij}(\mathbf{x}) \rangle = \langle x^i x^j \rangle$ synchronous pairwise correlation between region i and region j, for all pairs of
271 regions of the DMN and FPN.

272 3.3.3. Inferred Information

273 The authors show that the pairwise MEM accurately characterize the statistical behavior of
274 discretized BOLD signals in the human RSNs. Functional interactions (coupling parameters) from
275 the pairwise MEM were similar to the anatomical connections. The authors show that the agreement
276 between the estimated matrices of functional interactions and the anatomical connections is better
277 than others methods such as 1) Pearson's correlation coefficients, 2) Inverse Gaussian model, 3) Partial
278 correlation, and 4) Mutual Information method. These findings suggest that the large-scale human
279 brain networks of resting state can be accurately modeled by a synchronous pairwise model MEM.
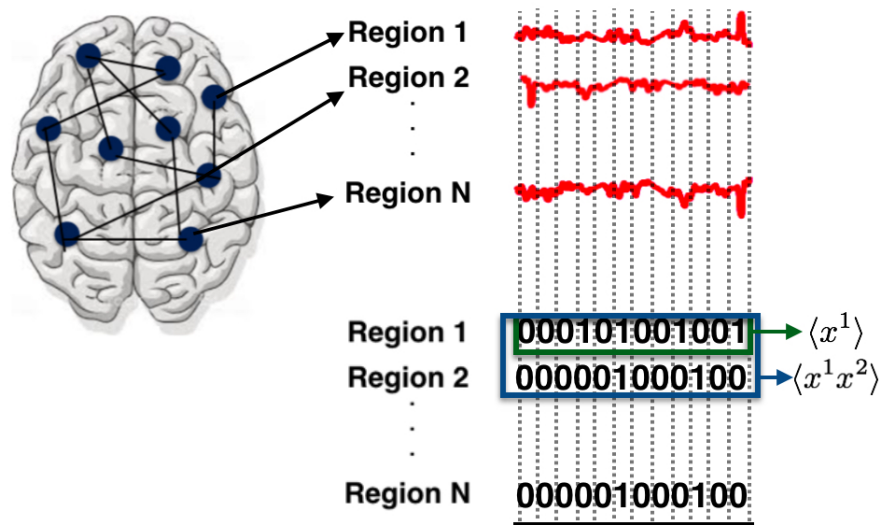
Figure 3. Regions of interest in the brain (represented as circles) corresponding to the DMN and FPN
are selected and their BOLD signals (continuous) are binarized. From the binary data the average
values of observables are computed. The maximum entropy principle is used to find the least biased
joint probability distribution consistent with data.

280 *3.4. Plant communities relative abundance*

281 For a long time, ecologist have tried to build models to estimate the relative abundance of species
282 in a community [51]. This has been tried using different alternatives such as population biology
283 approaches, and lately there has been more attempts using neutral theory [52]. One of the most
284 successful approaches has been the one developed by Shipley *et al.* [53] using MEP to estimate relative
285 abundance of species, predicting community assemblies even for a 3,000 square kilometer area with
286 a pool of over 500 species [54]. This approach brought a substantial improvement in understanding

287 plant relative abundance, without considering explicitly in the model any biological or ecological
288 process [55].

3.4.1. State Space

290 We present the approach used in [54]. Consider a community of plants in a landscape A_0
291 composed of a pool of S_0 species. Within that landscape we could have any combination of proportions
292 of each species $i \in S_0$ given by x^i. The state space is:

$$\chi = \{(x^1, x^2, \ldots, x^{|S_0|}) \mid x^i \in [0,1] \text{ and } \sum_{i=1}^{|S_0|} x^i = 1\}.$$

293 Properly defining the species pool is a crucial step as determine the state space.

3.4.2. Observables

295 The observables are the traits and represent features such as: diameter at breast height, whole
296 plant height, leaf area per leaf dry mass and seed mass among others.
297 From a restricted area A_0 traits are measured from all the plants found *without identifying the*
298 *plants*. The average values of the observables (traits) are the constraints of the maximum entropy
299 problem. The optimal proportion consistent with the observed data is found using the MEP and
300 known databases where average traits are available for several plant (see figure 4). Currently in the
301 Botanical Information and Ecology Network (BIEN) database there are 53 traits that can be extracted
302 for a large number of plant species, and if one of the traits is not available for the species level it can be
303 inferred as the mean of the next available phylogenetic level (e.g. genus, family) [56]. There are also
304 over 17 million of plot observations where every plant has been recorded in a given area, thus there is
305 a big database with more than 485 thousand species to test this method.
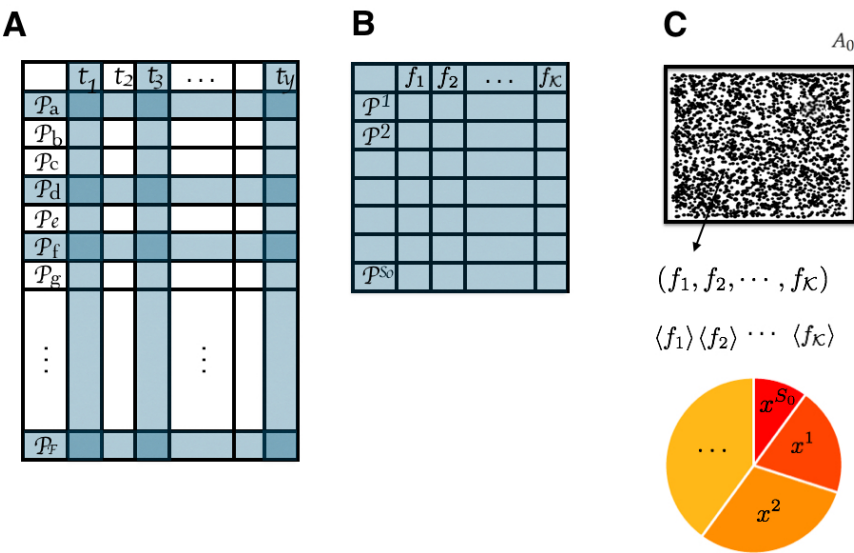
Figure 4. A) From a large datasets where several plant species have their recorded mean trait value, we extract B) a reduced database with the possible plant species present in A_0 (blue rows selected in A) and traits possible to be measured (blue columns selected in A). C) Then traits are measured in the field for all possible plants without specifying the species. The average values of these traits are the constraints for the maximum entropy problem of finding an estimate for the proportion of each plant species in A_0.

3.4.3. Inferred Information

The results of this model successfully predict the relative abundance of plant species along the landscape and through time [53]. This prediction is extremely useful in ecology. The MEP outperforms alternative methods such as population dynamics.

3.5. Macroecology and biodiversity

Macro-ecology is a field of ecology that studies ecosystems at a global scale looking for universal patterns and relationships. The objects of study are species/area relationship, species abundance distributions among others [57]. The MEP provides the theoretical framework for unifying and study community patterns of macro-ecology through four variables which are assumed to be known: number of species S_0, number of individuals N_0, total metabolic rate E_0, and the total area A_0. Here we present the Maximum Entropy Theory of Ecology (METE) introduced in chapter 7 of the book [58], and discussed in [59–61], but using our notation and approach.

3.5.1. State Space

One of the main problems stated by METE is to estimate of the probability that a species picked at random in a chosen area A_0, belongs to a species that has a total population of n and with metabolic rate ϵ, given that in that area in known that there is a pool of S_0 species, N_0 total individuals and a total metabolic rate of E_0. This probability is denoted by:

$$p(n, \epsilon | A_0, S_0, N_0, E_0) \tag{3}$$

p is a mixed discrete distribution over n (discrete number of individuals) and continuous over ϵ (real valued metabolic rates). The state space is the product space between the abundance of each species and the total metabolic rate of each of them. Mathematically in our notation $\chi = \mathbb{N} \times \mathbb{R}_+$
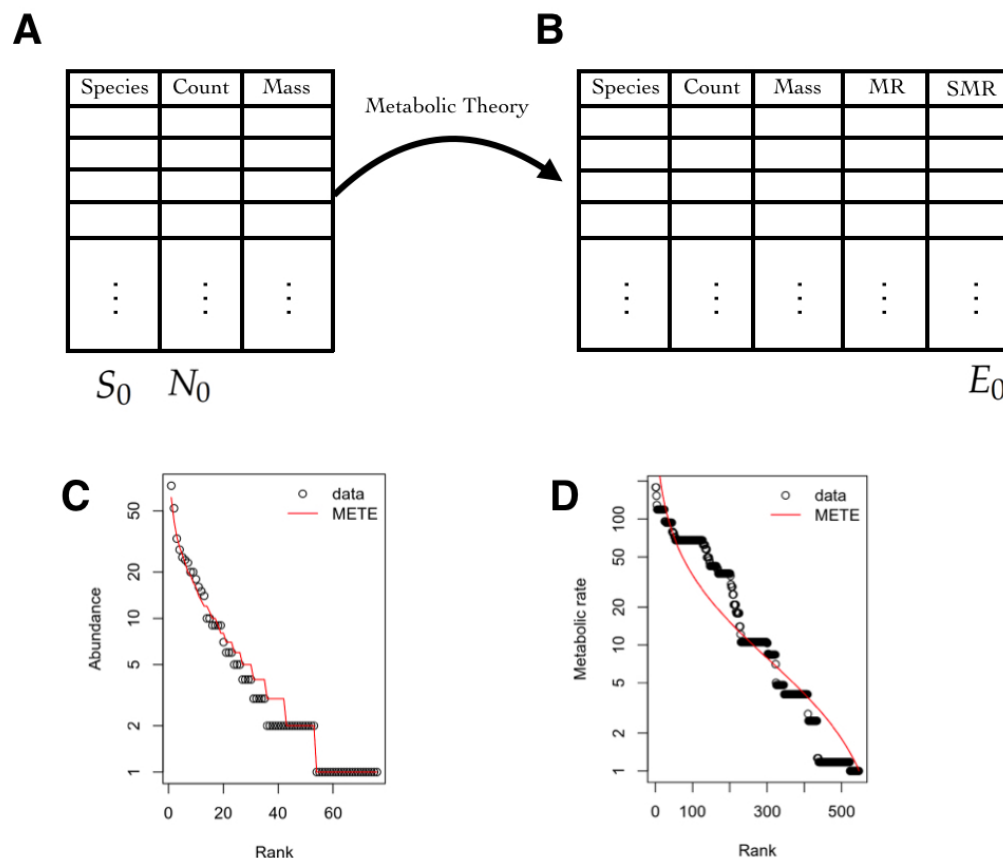
3.5.2. Observables

The observables are the abundance per species n, whose average is estimated by the fraction $\frac{N_0}{S_0}$ where numerator and denominator are assumed to be known, and $n\epsilon$ the total metabolic rate of the individuals within the species, whose average over species of the is $\frac{E_0}{S_0}$, again both quantities are assumed to be known (these quantities can be extracted or estimated from databases). Both average values serve as the constraints of the MEM:

$$\sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} n \cdot p(n, \epsilon | A_0, S_0, N_0, E_0) \, d\epsilon = \mathbb{E}_p[n] = \frac{N_0}{S_0} \tag{4}$$

$$\sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} n \cdot \epsilon \cdot p(n, \epsilon | A_0, S_0, N_0, E_0) \, d\epsilon = \mathbb{E}_p[n\epsilon] = \frac{E_0}{S_0}. \tag{5}$$

3.5.3. Inferred Information

Once the joint probability distribution p (3) is fitted by the MEP, many ecological relationships can be derived. For instance, one can obtain the marginal distributions. Integrating over ϵ, the species-abundance distribution is obtained (usually denoted by $\phi(n_0 | A_0, S_0, N_0)$.) Summing over n, the metabolic rate distribution over all individuals is obtained (usually denoted by $\psi(\epsilon | A_0, S_0, N_0, E_0)$). Similarly, the Species-Area relationship, Endemics-Area relationship among other key features in macro-ecology can be derived [62]. Recently this methodology has been used to estimate p using data from 60 different forest communities, with more than 2000 species, and it was proven to successfully explain both the Species-Area relationship, and the Individual-Size distribution [63].

Figure 5. A) Database with the species, their counts and the mass of that species for a given area A_0. From here the quantities used to compute the average value of the observables S_0 and N_0 are obtained. B) Using the metabolic theory the metabolic rate (MR) of each species are estimated. The quantity E_0 is computed from the standardized metabolic rate (SMR), which is obtained dividing all the MRs the the minimum MR. C) The Species-abundance distribution $\phi(n_0|A_0, S_0, N_0)$ is computed from the joint maximum entropy distribution and a graph of rank versus abundance is plotted. D) The metabolic rate distribution over all individuals is obtained $\psi(\epsilon|A_0, S_0, N_0, E_0))$. A graph of rank versus metabolic rate is shown. Images C) and D) were obtained from the maximum entropy distributions fitted to data available in the R package meteR [64].

3.6. Humans interactions in the US Supreme Court

In Ref. [65] the authors considered the voting on the Supreme Court of the United States (SCOTUS), which is the highest court in the US government. The article show that the structure of coalitions among multiple justices can be studied using a pairwise maximum entropy model. Moreover, is shown that a maximum entropy distribution provide surprisingly accurate descriptions of collective behavior of voting patterns.
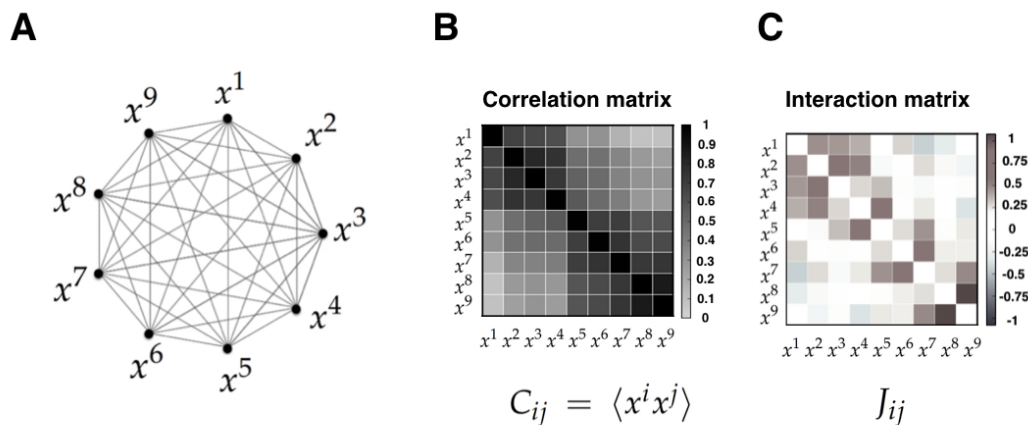
3.6.1. State Space

The data consist of $N = 9$ justices who vote on the constitutionality of legislative and executive actions. The article consider data from second Rehnquist Court (1994-2004, 895 votes) in a period of time during which the membership stays constant. The Court issues majority and minority opinions, and these can be supplemented with other opinions; although opinions can be nuanced, each justice casts a yes ($x^i = +1$) or no ($x^i = -1$) vote, and the majority of votes decides the fate of each case. The state of the whole system can be represented by $\mathbf{x} = (x^1, \dots, x^9)$. The state space of each variable is $\chi^i = \{-1, 1\}$ and the state space of the whole system $\chi = \chi^1 \times \chi^2 \times \cdots \times \chi^9$.

355 3.6.2. Observables

356   The model only consider pairwise correlations between the nine variables $C_{ij} = \langle x^i x^j \rangle$ (see
357 figure 6). These correlations are the restrictions of the maximum entropy problem. The maximum
358 entropy distribution that is consistent with the observed pairwise correlations among justices' votes is
359 equivalent to an Ising spin glass with energy function (A4):

$$\mathcal{H}(\mathbf{x}) = -\frac{1}{2} \sum_{i \neq j} J_{ij} x^i x^j,$$

A

B

Correlation matrix

$$C_{ij} = \langle x^i x^j \rangle$$

C

Interaction matrix

$$J_{ij}$$

Figure 6. A) Interacting random variables x^i representing votes of the nine justices. B) Correlation matrix between random variables x^i and x^j measured directly from data. C) Interaction matrix computed from the maximum entropy principle.

360 3.6.3. Inferred Information

361   The MEM, predicts the joint distribution over voting patterns $p(\mathbf{x})$ and can be tested in various
362 ways. For example the probability that the vote is split $(k, 9-k)$, with $k = [5, 9]$ votes for the majority,
363 can be computed from data and predicted from the model. The article report small quantitative
364 discrepancies. Additionally, probability versus energy and mutual information are measured from
365 data and compared with the predictions of the model.
366   The MEM shows that voting patterns are organized in an energy landscape that is equivalent to an
367 Ising spin glass. This simple model correctly predicts the extent to which each justice is correlated with
368 the majority. A basic question about the dynamics of a court concerns the influence that individual
369 justices have on the majority decision. This dynamics can be inferred from the MEM. It can be observed
370 that a strong tendency toward unanimity emerges from the inferred probability model.
371   This investigation shows that the competition between unanimity and ideological division emerge
372 from interactions among the justices in the same way that, magnetism emerges from interaction among
373 spins.

374 **4. Discussion**

375   In the previous sections, we outline the fundamental ideas behind the MEP and explored how
376 it can be used to analyze experimental data across different biological scales, ranging from the
377 amino-acids up to macroscopic social scenarios. Key features that all these scenarios share are the
378 underlying randomness in biological phenomena, which make the probabilistic approach appropriate,
379 and the fact that it is impossible to sample the whole state space from experimental data, therefore
380 modeling and statistical inference becomes necessary.

Scenario	State space	Observables and average values
Amino acids in proteins	$\{0,1\}^{Lr}$	Average amino acid occurrence on a given site and average co-occurrences of amino acids on site-pairs
Retinal ganglion cells	$\{0,1\}^{N}$	Firing rates and pairwise correlations
Whole brain networks	$\{0,1\}^{N}$	Activation rates and pairwise correlations
Plant communities	$[0,1]^{S_0}$	Average value of traits
Macroecologic biodiversity	$N \times \mathbb{R}^{+}$	Average abundance per species and average over species of the total metabolic rate of the individuals within the species.
US Supreme Court	$\{-1,1\}^{9}$	Pairwise correlations

4.1. Lessons from the case studies

Let us summarize some key takeaway messages from the comparative studies presented in section 3.

- **The MEP can be applied in very heterogeneous systems.** From the considered case studies, some observables are directly related to causal/mechanistic interactions, while others do not. Moreover, the averages of these observables are in some cases temporal, while in other cases are spatial. The fact that the same formalism can be adapted to such different systems highlights the flexibility of the MEP approach.
- **It is highly non-trivial to choose the right state space and observables.** While different applications of the MEP do not require conceptual changes to the basic method, the results rely entirely on the definition of state variables and observables, which are both determined by the modeler. For this reason, the researcher needs to double-check that these choices are adequate, i.e., if the model is capable to predict average values of observables not included in the MEP to fit the data, and if the model address the questions that one would like to ask. It is important not to lose perspective on this as the MEP is a concave maximization problem which will always find a solution, whether is meaningful or not is the real issue.
- **Correlations versus interactions**. It is important to note that the MEP make a strong distinction between *interactions* and *correlations*. Indeed, correlations are statistical dependencies between variables, while interactions are the local rules of the system (defined in the energy function) from which correlations and collective phenomena emerge. Importantly, it has been shown that the interactions give a more useful account of the physical topology of some system than correlations. Examples of this include structural connectivity [49,66], and contacts between proteins sites [40].

4.2. Final remarks

The MEP is an attractive approach that fits well the needs of our Big Data era, where information abounds but knowledge about general principles and mechanical rules behind it are scarce. Indeed, biology is a "data-rich but theory poor" field (at least in comparison with physics), which makes the MEP so well suited for it needs. Modern approaches work without even trying to characterize the underlying mechanistic structure ruling the systems. The MEP approach is gathering momentum as Big Data becomes ubiquitous in experimental biology at all scales. Big data obtained experimentally can be accurately characterized by statistical models using a small set of coefficients as shown in the examples, and new data can be predicted with a given degree of accuracy on the basis of the older one.

We believe that the role of the MEP in biology is to make sense of this complexity, derive general properties of living systems based on the interactions, provide insight into how such systems behave collectively. It is our hope that this article might contribute to developing a deeper understanding of the MEP across multiple biological scales, which in turn might help to enable new applications in the future.

Abbreviations

426 The following abbreviations are used in this manuscript:
427

MEP	Maximum entropy principle.
MEM	Maximum entropy model.
DNA	Deoxyribonucleic Acid.
MSA	Multiple sequence alignment.
LFP	Local field potential.
fMRI	Functional magnetic resonance imaging.
BOLD	Blood Oxygen Level-Dependent signals.
RSN	Resting state network.
DMN	Default mode network.
FPN	Fronto-parietal network.
METE	Maxent Theory of Ecology.
SCOTUS	Supreme Court of the United States.
BIEN	Botanical Information and Ecology Network.

Symbol list

x_t^k	State of k-th variable at time t.
x^i	State of i-th variable on $\mathbf{x}$.
$\mathbf{x}$	Configuration of N-elements system.
χ^i	State space of a random variable x^i.
χ	State space of a N-elements system.
$\mathbb{R}$	Set of real numbers.
$q[\mathbf{x}]$	Set of probability distributions that match $\langle f_k \rangle_e$.
$p[\mathbf{x}]$	Probability distribution that maximizes entropy and match $\langle f_k \rangle_e$.
$\mathcal{S}[p]$	Entropy of the probability measure p.
$\mathcal{K}$	Number of observables.
f_k	Observable k.
$\langle f_k \rangle_e$	Empirical average value of observable k.
$\mathbb{E}_q[f_k]$	Expected value of f_k respect to q.
$\mathcal{M}$	Set of probability measures.
λ_k	Set of Lagrange multipliers.
$\mathcal{H}$	Energy function.
Z	Partition Function.

Appendix A. Solving the discrete maximum entropy problem with constraints via Lagrange multipliers

433 Finding the maximum entropy distribution given the average value of a set of observables is
434 mathematically equivalent to finding the maximum of a concave functional (the entropy) in the space
435 of probability distributions that satisfies the constraints set by the average values of the observables.
436 This mathematical problem is solved analytically by using the method of Lagrange multipliers, as
437 explained below.
438

Input:

$$\langle f_k \rangle_e = c_k \text{ for } k \in \{1, \ldots, \mathcal{K}\}$$

439 .

440

Output:

$$\{p_1, p_2, \cdots, p_{|\chi|}\} \text{ s.t. } \sum_{i \in \chi} p_i = 1$$

t

$$\min \sum_{\mathbf{x} \in \mathcal{X}} q(\mathbf{x}) \log q(\mathbf{x})$$

441 To solve this optimization problem (which is treated as the equivalent minimization problem)
442 a common approach is to use a set of Lagrange multipliers $\boldsymbol{\lambda} \in \mathbb{R}^{\mathcal{K}+1}$ one per constraint ($\mathcal{K}$ average
443 values plus the normalization constraint), and build a function called *Lagrangian*:

$$\mathcal{L} = \sum_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}) \log(\mathbf{x}) - \lambda_0 \left(\sum_{i \in \chi} p_i \right) - \sum_{i \in \chi} p_i \sum_{k=1}^{\mathcal{K}} \lambda_k f_k(a_i)$$

$$\frac{\partial \mathcal{L}}{\partial p_j} = \frac{\partial}{\partial p_j} \left[p_j \ln p_j - \lambda_0 p_j - p_j \sum_{k=1}^{\mathcal{K}} \lambda_k f_k(a_j) \right]$$

$$= \ln p_j + 1 - \lambda_0 - \sum_{k=1}^{\mathcal{K}} \lambda_k f_k(a_j) = 0 \tag{A1}$$

444 Solving the previous equation for p_j gives the following,

$$p_j = \exp \left(\lambda_0 - 1 + \sum_{k=1}^{\mathcal{K}} \lambda_k f_k(a_j) \right). \tag{A2}$$

445 Now, fix λ_0 to get normalization,

$$\sum_{j \in \chi} p_j = \sum_{j \in \chi} \exp \left(\lambda_0 - 1 + \sum_{k=1}^{\mathcal{K}} \lambda_k f_k(a_j) \right)$$

$$= \exp(\lambda_0 - 1) \sum_{j \in \chi} \exp \left(\sum_{k=1}^{\mathcal{K}} \lambda_k f_k(a_j) \right) = 1 \tag{A3}$$

446 Calling

$$Z = \sum_j \exp \left(\sum_{k=1}^{\mathcal{K}} \lambda_k f_k(a_j) \right) \rightarrow \exp(\lambda_0 - 1) = \frac{1}{Z},$$

447 thus from (A2),

$$p_j = \frac{\exp \left(\sum_{k=1}^{\mathcal{K}} \lambda_k f_k(a_j) \right)}{Z},$$

448 or

$$p(\mathbf{x}) = \frac{\exp \left(\sum_{k=1}^{\mathcal{K}} \lambda_k f_k(\mathbf{x}) \right)}{Z}.$$

449 The term in the exponential is a linear combination of the observables and is called the *energy function*,
450 it is denoted by $\mathcal{H}(\mathbf{x})$, from which

$$p(\mathbf{x}) = \frac{e^{\mathcal{H}(\mathbf{x})}Z}{.} \tag{A4}$$

In statistical mechanics the constant Z is the partition function. Notice that the solution of the MEP is the canonical Boltzmann-Gibbs distribution. This last equation establish a bridge between information theory and statistical mechanics (canonical ensambles).

1. Waldrop, M.M. *Complexity: The emerging science at the edge of order and chaos*; Simon and Schuster, 1993.
2. Stein, R.B. More Is Different. *Science* **1972**, *177*, 393–396.
3. Rosas, F.; Mediano, P.A.; Gastpar, M.; Jensen, H.J. Quantifying high-order interdependencies via multivariate extensions of the mutual information. *Accepted in Physical Review E, arXiv:1902.11239* **2019**.
4. Schneidman, E.; Berry, M.J.; Segev, R.; Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **2006**, *440*, 1007–12.
5. Tkačik, G.; Marre, O.; Amodei, D.; Schneidman, E.; Bialek, W.; Berry, M.J. Searching for collective behavior in a large network of sensory neurons. *PLoS computational biology* **2014**, *10*, e1003408.
6. Nasser, H.; Cessac, B. Parameter Estimation for Spatio-Temporal Maximum Entropy Distributions: Application to Neural Spike Trains. *Entropy* **2014**, *16*.
7. Jaynes, E. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*.
8. Santolini, M.; Mora, T.; Hakim, V. A General Pairwise Interaction Model Provides an Accurate Description of In Vivo Transcription Factor Binding Sites. *PloS one* **2014**, *9*, e99015.
9. Weigt, M.; White, R.A.; Szurmant, H.; Hoch, J.A.; Hwa, T. Identification of direct residue contacts in protein–protein interaction by message passing. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 67–72.
10. Morcos, F.; Pagnani, A.; Lunt, B.; Bertolino, A.; Marks, D.S.; Sander, C.; Zecchina, R.; Onuchic, J.N.; Hwa, T.; Weigt, M. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, E1293–E1301.
11. Barton, J.; K. Chakraborty, A.; Cocco, S.; Jacquin, H.; Monasson, R. On the Entropy of Protein Families. *Journal of Statistical Physics* **2015**, *162*.
12. Mora, T.; Walczak, A.M.; Bialek, W.; Callan, C.G. Maximum entropy models for antibody diversity. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 5405–5410.
13. Elhanati, Y.; Murugan, A.; G. Callan Jr, C.; Mora, T.; M. Walczak, A. Quantifying selection in immune receptor repertoires. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*.
14. Schneidman, E.; Berry II, M.; Segev, R.; Bialek, W. Weak pairwise correlations imply string correlated network states in a neural population. *Nature* **2006**, *440*, 1007– 1012.
15. Tang, A.; Jackson, D.; Hobbs, J.; Chen, W.; Smith, J.; Patel, H.; Prieto, A.; Petrusca, D.; Grivich, M.; Sher, A.; Hottowy, P.; W.Dabrowski.; Litke, A.; Beggs, J. A Maximum Entropy Model Applied to Spatial and Temporal Correlations from Cortical Networks *In Vitro. The Journal of Neuroscience* **2008**, *28*, 505–518.
16. Tkačik, G.; Mora, T.; Marre, O.; Amodei, D.; Berry II, M.; Bialek, W. Thermodynamics for a network of neurons: Signatures of criticality. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*.
17. Marre, O.; El Boustani, S.; Frégnac, Y.; Destexhe, A. Prediction of spatiotemporal patterns of neural activity from pairwise correlations. *Physical review letters* **2009**, *102*.
18. Cofré, R.; Cessac, B. Exact computation of the maximum entropy potential of spiking neural networks models. *Physical Review E* **2014**, *107*, 368–368.
19. Cofré, R.; Maldonado, C. Information Entropy Production of Maximum Entropy Markov Chains from Spike Trains. *Entropy* **2018**, *20*.
20. Bialek, W.; Cavagna, A.; Giardina, I.; Mora, T.; Silvestri, E.; Viale, M.; M Walczak, A. Statistical mechanics for natural flocks of birds. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 4786–91.

21. Cavagna, A.; Giardina, I.; Ginelli, F.; Mora, T.; Piovani, D.; Tavarone, R.; M Walczak, A. Dynamical maximum entropy approach to flocking. *Physical review. E, Statistical, nonlinear, and soft matter physics* **2014**, *89*, 042707.

22. Shemesh, Y.; Sztainberg, Y.; Forkosh, O.; Shlapobersky, T.; Chen, A.; Schneidman, E. High-order social interactions in groups of mice. *eLife* **2013**, *2*, e00759.

23. Harte, J. *Maximum entropy and ecology. A theory of abundance, distribution, and energetics*; Oxford University Press, 2011.

24. Harte, J.; Newman, E. Maximum information entropy: A foundation for ecological theory. *Trends in ecology & evolution* **2014**, *29*.

25. Nguyen, H.C.; Zecchina, R.; Berg, J. Inverse statistical problems: from the inverse Ising problem to data science. *Advances in Physics* **2017**, *66*, 197–261.

26. De Martino, A.; Martino, D. An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon* **2018**, *4*, e00596.

27. R Stein, R.; Marks, D.; Sander, C. Inferring Pairwise Interactions from Biological Data Using Maximum-Entropy Probability Models. *PLoS computational biology* **2015**, *11*, e1004182.

28. Natale, J.L.; Hofmann, D.; Hernández, D.G.; Nemenman, I. Reverse-engineering biological networks from large data sets. *bioRxiv* **2017**, [https://www.biorxiv.org/content/early/2017/05/28/142034.full.pdf].

29. Battistin, C.; Dunn, B.; Roudi, Y. Learning with unknowns: Analyzing biological data in the presence of hidden variables. *Current Opinion in Systems Biology* **2017**, *1*, 122 – 128.

30. Tkačik, G. From statistical mechanics to information theory: understanding biophysical information-processing systems. *arXiv e-prints* **2010**, p. arXiv:1006.4291, [arXiv:q-bio.MN/1006.4291].

31. Marquet, P.A.; Allen, A.P.; Brown, J.H.; Dunne, J.A.; Enquist, B.J.; Gillooly, J.F.; Gowaty, P.A.; Green, J.L.; Harte, J.; Hubbell, S.P.; others. On theory in ecology. *BioScience* **2014**, *64*, 701–710.

32. Marre, O.; El Boustani, S.; Frégnac, Y.; Destexhe, A. Prediction of spatiotemporal patterns of neural activity from pairwise correlations. *Physical review letters* **2009**, *102*.

33. Kazama, J.; Tsujii, J. Maximum Entropy Models with Inequality Constraints: A Case Study on Text Categorization. *Machine Learning* **2005**, *60*, 159–194.

34. Shannon, C.E. A mathematical theory of communication. *Bell system technical journal* **1948**, *27*.

35. Cover, T.M.; Thomas, J.A. *Elements of information theory*, second ed.; Wiley-Interscience, 2006.

36. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **2011**, *39*, W29–W37.

37. Remmert, M.; Biegert, A.; Hauser, A.; Soding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods* **2011**, *9*, 173–175.

38. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; Sonnhammer, E.L.; Hirsh, L.; Paladin, L.; Piovesan, D.; Tosatto, S.C.; Finn, R.D. The Pfam protein families database in 2019. *Nucleic Acids Research* **2018**, *47*, D427–D432.

39. Mukherjee, S.; Stamatis, D.; Bertsch, J.; Ovchinnikova, G.; Katta, H.Y.; Mojica, A.; Chen, I.M.A.; Kyrpides, N.C.; Reddy, T. Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Research* **2018**, *47*, D649–D659.

40. Cocco, S.; Feinauer, C.; Figliuzzi, M.; Monasson, R.; Weigt, M. Inverse Statistical Physics of Protein Sequences: A Key Issues Review. *Reports on Progress in Physics* **2017**, *81*.

41. Onuchic, J.N.; Wolynes, P.G. Theory of protein folding. *Current Opinion in Structural Biology* **2004**, *14*, 70 – 75.

42. Cheng, R.R.; Nordesjö, O.; Hayes, R.L.; Levine, H.; Flores, S.C.; Onuchic, J.N.; Morcos, F. Connecting the Sequence-Space of Bacterial Signaling Proteins to Phenotypes Using Coevolutionary Landscapes. *Molecular Biology and Evolution* **2016**, *33*, 3054–3064.

43. Hopf, T.; B Ingraham, J.; Poelwijk, F.; P I Schärfe, C.; Springer, M.; Sander, C.; Marks, D. Mutation effects predicted from sequence co-variation. *Nature Biotechnology* **2017**, *35*, 128–135.

44. Pillow, J.W.; Shlens, J.; Paninski, L.; Sher, A.; Litke, A.M.; Chichilnisky, E.J.; Simoncelli, E.P. Spatio-temporal correlations and visual signaling in a complete neuronal population. *Nature* **2008**, *454*, 995–999.

45. Ganmor, E.; Segev, R.; Schneidman, E. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 9679–9684.

46. Vasquez, J.; Palacios, A.; Marre, O.; Berry II, M.; Cessac, B. Gibbs distribution analysis of temporal correlation structure on multicell spike trains from retina ganglion cells. *J. Physiol. Paris* **2012**, *106*, 120–127.

47. Yu, S.; Huang, D.; Singer, W.; Nikolic, D. A Small World of Neuronal Synchrony. *Cereb. Cortex* **2008**.

48. Ohiorhenuan, I.E.; Mechler, F.; Purpura, K.P.; Schmid, A.M.; Hu, Q.; Victor, J.D. Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* **2010**, *466*, 617–621.

49. Watanabe, T.; Hirose, S.; Wada, H.; Imai, Y.; Machida, T.; Shirouzu, I.; Konishi, S.; Miyashita, Y.; Masuda, N. A pairwise maximum entropy model accurately describes resting-state human brain networks. *Nature communications* **2013**, *4*, 1370.

50. R.L Buckner, J.A.H.; Schacter, D. The brain's default network: anatomy, function, and relevance to disease. *Ann. N Y Acad. Sci.* **2008**, *1124, 1-38*.

51. MacArthur, R.H. *Geographical ecology: patterns in the distribution of species*; Princeton University Press, 1984.

52. Hubbell, S. *The Unified Neutral Theory of Biodiversity and Biogeography.*; Princeton University Press, 2001.

53. Shipley, B.; Vile, D.; Garnier, É. From plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science* **2006**, *314*, 812–814.

54. Sonnier, G.; Shipley, B.; Navas, M.L. Plant traits, species pools and the prediction of relative abundance in plant communities: a maximum entropy approach. *Journal of Vegetation Science* **2010**, *21*, 318–331.

55. Kattge, J.; Diaz, S.; Lavorel, S.; Prentice, I.C.; Leadley, P.; Bönisch, G.; Garnier, E.; Westoby, M.; Reich, P.B.; Wright, I.J.; others. TRY–a global database of plant traits. *Global change biology* **2011**, *17*, 2905–2935.

56. Maitner, B.S.; Boyle, B.; Casler, N.; Condit, R.; Donoghue, J.; Durán, S.M.; Guaderrama, D.; Hinchliff, C.E.; Jørgensen, P.M.; Kraft, N.J.; others. The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods in Ecology and Evolution* **2018**, *9*, 373–379.

57. Brown, J.H. *Macroecology*; University of Chicago Press, 1995.

58. Harte, J.; Zillio, T.; Conlisk, E.; Smith, A.B. Maximum entropy and the state-variable approach to macroecology. *Ecology* **2008**, *89*, 2700–2711.

59. Favretti, M. Remarks on the Maximum Entropy principle with Application to the Maximum Entropy Theory of Ecology. *Entropy* **2018**, *20*.

60. Harte, J. Maximum Entropy and Theory Construction: A Reply to Favretti. *Entropy* **2018**, *20*.

61. Favretti, M. Maximum Entropy Theory of Ecology: A Reply to Harte. *Entropy* **2018**, *20*.

62. Harte, J.; Newman, E.A. Maximum information entropy: a foundation for ecological theory. *Trends in ecology & evolution* **2014**, *29*, 384–389.

63. Xiao, X.; McGlinn, D.J.; White, E.P. A strong test of the maximum entropy theory of ecology. *The American Naturalist* **2015**, *185*, E70–E80.

64. Rominger, A.; Merow, C. meteR: An R package for testing the Maximum Entropy Theory of Ecology. *Methods in Ecology and Evolution* **2017**, *8*, 241–247.

65. Lee, E.D.; Broedersz, C.P.; Bialek, W. Statistical Mechanics of the US Supreme Court. *Journal of Statistical Physics* **2015**, *160*, 275–301.

66. Kadirvelu, B.; Hayashi, Y.; Nasuto, S. Inferring structural connectivity using Ising couplings in models of neuronal networks. *Scientific Reports* **2017**, *7*.