

Trivial and non-trivial applications of entropy maximization in ecology: a reply to Shipley

Bart Haegeman and Michel Loreau

B. Haegeman (bart.haegeman@inria.fr), INRA research team MERE, UMR Systems Analysis and Biometrics, 2 place Pierre Viala, FR-34060 Montpellier, France. – M. Loreau, Dept. of Biology, McGill Univ., 1205 Avenue Docteur Penfield, Montreal, QC, H3A 1B1, Canada.

Entropy maximization (EM) is becoming an increasingly popular modelling technique in ecology, but its potential and limitations are still poorly understood. In our previous contribution (Haegeman and Loreau 2008), we showed that even a trivial application of EM can yield predictions that provide an excellent fit to empirical data. In his response, Shipley (2009) distinguishes two different versions of the EM procedure, an information-theoretical version and a combinatorial version, to justify a trivial application of EM. Here we first provide a brief user's guide to EM to clarify the various steps involved in the procedure. We then show that the information-theoretical and combinatorial rationales for EM are but complementary views on the same procedure. Lastly, we attempt to identify the conditions that lead to trivial and non-trivial applications of EM. We discuss how non-trivial applications of EM can yield valuable new insights in ecology.

Entropy maximization (EM) is an inference technique that has its origins in statistical mechanics (Jaynes 1957), and that has been applied to many other problems since (Jaynes 2003). Application of the EM formalism to ecological systems has been very popular in recent years (Shipley et al. 2006, Banavar and Maritan 2007, Pueyo et al. 2007, Dewar and Porté 2008, Harte et al. 2008), and holds promise to become an efficient modelling technique in ecology.

In our previous contribution (Haegeman and Loreau 2008), however, we argued that a blind application of the EM procedure can easily lead to wrong conclusions. As an example, we analysed the EM application of Shipley et al. (2006), and showed that the good correspondence between empirical observations and EM predictions is independent of the EM procedure in this case. In response to our study, Shipley (2009) argues that our criticism does not apply to the EM analysis of Shipley et al. (2006). He introduces two different versions of the EM procedure: an information-theoretical version, on which he claims his analysis is based, and a combinatorial version, on which he claims our analysis is based.

We feel that this distinction is obfuscating the fundamental issue, and will lead to confusion on the true potential of EM as a method of inference in ecology and other sciences. We show below that there are two rationales for the same EM procedure, not two different versions of the EM procedure. These two rationales are rooted in the frequency-based and Bayesian interpretations of probability theory, respectively. Our previous analysis of the limitations

of EM (Haegeman and Loreau 2008) applies to the EM procedure as such, independently of which rationale is used.

Rather than providing a point-by-point response to Shipley (2009), we prefer to broaden the debate and make a number of clarifications that we hope will be useful to all ecologists who are interested in applying EM to ecological problems. EM is not a magic recipe that can solve all problems; its principles are simple, and do not fundamentally differ from those of statistical modelling in general. Therefore, we first provide a brief user's guide to EM. We show that many more assumptions are built in the problem formulation than often believed, and that different formulations of the same problem are possible. We then discuss the information-theoretical and combinatorial rationales of EM, and show that they provide two complementary views on the same procedure. Lastly, we attempt to identify the conditions that lead to trivial and non-trivial applications of EM. We explain why Shipley et al.'s (2006) application was trivial, and how other applications can be non-trivial and lead to new fundamental insights in ecology.

A brief user's guide to EM

We explain the essentials of the EM procedure using the flow diagram of Fig. 1, together with a graphical representation of a basic EM application in Fig. 2. This simple representation will suffice to make some key points about the usefulness of EM in ecology.

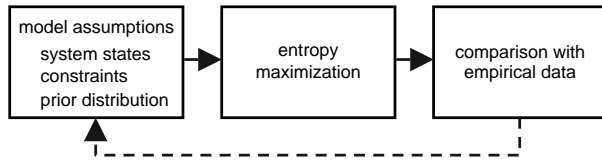


Figure 1. Flow diagram of the EM approach. To formulate an EM problem, three ingredients must be specified: (1) what are the states of the system, (2) what constraints are taken into account, and (3) what is the prior distribution. These three ingredients are then combined in a maximization problem to obtain the EM prediction. Comparing the EM prediction with empirical data either corroborates the problem formulation or shows significant differences between prediction and observation. In the latter case, the problem formulation can be changed (e.g. by adding constraints), and the EM procedure can be repeated in the hope of getting better predictions.

Model assumptions (first box in Fig. 1)

Before applying EM, we have to specify what we call a ‘state’ of the system. The variables of the EM problem are the probabilities of these states. Here we assume for simplicity that there are a finite number of states S (but see Appendices 1, 2 for more general examples), and we denote the probability of state i by p_i . We are thus looking for a probability distribution (p_1, p_2, \dots, p_S) . The choice of a set of states defines a coordinate system in which EM will be performed (Fig. 2).

Next, we have to specify the constraints that we want to take into account. These constraints must be formulated in

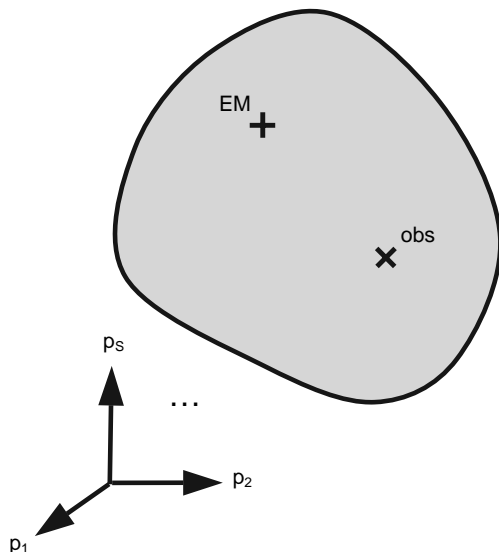


Figure 2. Graphical representation of the EM procedure. By choosing the set of states $\{1, 2, \dots, S\}$, the variables (p_1, p_2, \dots, p_S) of the problem are determined. In the coordinate system (p_1, p_2, \dots, p_S) the constraints delimit the feasible set of states (filled in light grey). Solving an optimization problem yields the EM probability vector (indicated by the + cross). In the simplest EM applications, the EM prediction can be compared directly with an observed vector (indicated by the x cross).

terms of the variables p_i . Imposing the constraints delimits a region in the set of all possible vectors (p_1, p_2, \dots, p_S) , which we call the feasible set (Fig. 2). Inside the feasible set, all constraints are satisfied; outside the feasible set, at least one of the constraints is violated. The EM solution thus belongs to the feasible set.

It is also possible to incorporate a prior probability distribution into the EM procedure. This prior distribution is necessary if we want to assign a larger probability to some states a priori, before we take the constraints into account. It could be represented as a vector (q_1, q_2, \dots, q_S) in our graphical representation, but for simplicity we leave it out of Fig. 2. In the case of a finite number of states, one can often do without a prior distribution, which amounts to assuming a uniform prior distribution.

It is important to realize that the set of states, the constraints and the prior distribution are closely interconnected. Indeed, the same EM problem can often be formulated using different choices for the set of states, the constraints and the prior distribution. Some useful examples are given in Appendix 1. But modifying the set of system states can lead to completely different EM predictions. It is therefore crucial to make all the model assumptions explicit before applying the EM algorithm as such.

Entropy maximization (second box in Fig. 1)

Once the set of system states, the constraints and the prior distribution are chosen, we are ready to apply the EM procedure. This procedure stipulates that the vector (p_1, p_2, \dots, p_S) that maximizes the entropy function H ,

$$H(p_1, \dots, p_S) = - \sum_{i=1}^S p_i \ln p_i \quad (1)$$

should be selected within the feasible set. In other words, we have to find the probability distribution (p_1, p_2, \dots, p_S) that maximizes the entropy H while respecting the constraints. If a prior distribution (q_1, q_2, \dots, q_S) has been specified, we should maximize the generalized entropy function H_R (where R stands for relative),

$$H_R(p_1, \dots, p_S | q_1, \dots, q_S) = - \sum_{i=1}^S p_i \ln \frac{p_i}{q_i} \quad (2)$$

The existence and uniqueness of the EM solution can be shown mathematically under very general conditions. Numerical algorithms are available to solve this optimization problem efficiently.

Comparison with empirical data (third box in Fig. 1)

If we use the EM formalism for modelling purposes (which should cover all cases of interest in ecology), we want to compare the EM prediction with empirical data. In the simplest applications, the probability distribution predicted by EM can be compared directly with the observed probability distribution. In more advanced applications, however, the comparison between prediction and observation will often be indirect. In such a case, the experimenter

does not have direct access to the probability distribution of the EM prediction, but average quantities computed from the EM probability distribution can be compared with experimentally accessible quantities.

Whether the comparison between prediction and observation is direct or indirect, the way to interpret it is similar to any other modelling approach. If the EM prediction fits the data well, we can conclude that the information taken into account in the EM inference (set of states, constraints, prior distribution) suffices to reproduce the data. If, on the contrary, the EM prediction differs significantly from the data, we can conclude that other mechanisms operate in the observed system. We might then be able to formalize such mechanisms, change the problem formulation (different set of states, other or additional constraints, different prior distribution), and apply the EM procedure again.

The rationale behind EM

The previous section specifies *how* the EM procedure is applied, but it does not explain *why* the EM procedure works. Why is it that we should prefer the probability distribution that maximizes the entropy H subject to the constraints to describe the system? This question can be answered in two ways. A first answer has a combinatorial nature, and was discussed in the recent ecological literature by Shipley et al. (2006) and Haegeman and Loreau (2008). A second answer is based on information theory, and is discussed by Pueyo et al. (2007), Dewar and Porté (2008) and Shipley (2009). Shipley (2009) boldly opposes the two approaches and defends the second against the first, which is all the more surprising since he used the first approach in his initial contribution (Shipley et al. 2006). Rather than

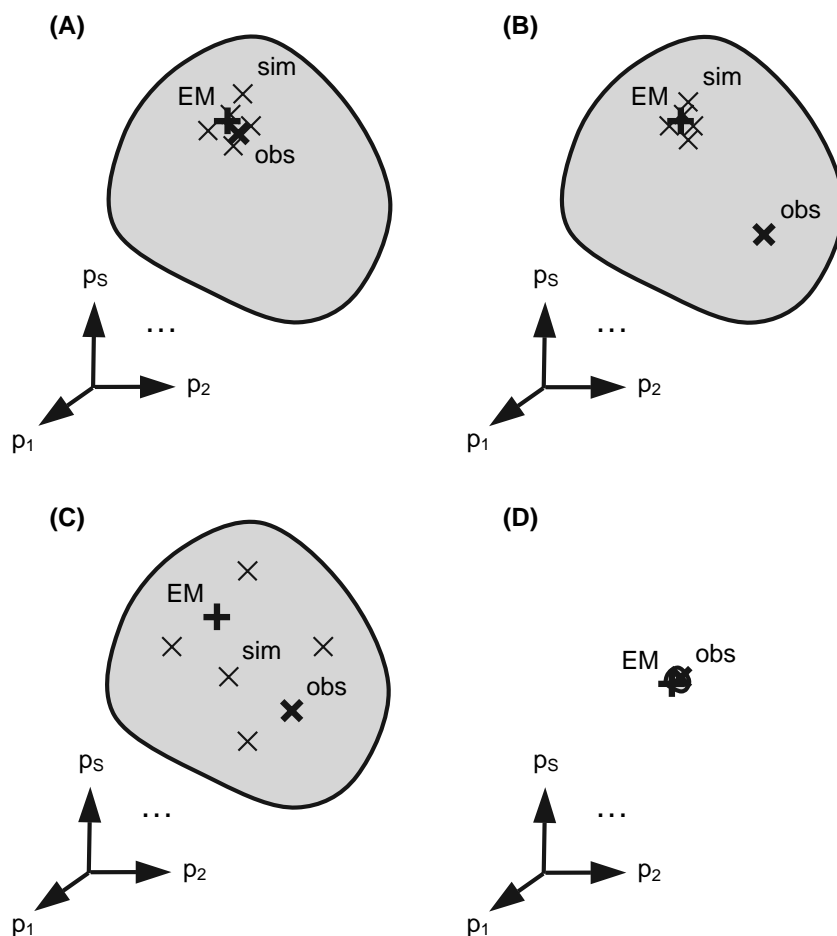


Figure 3. Graphical representation of different applications of the EM procedure, using the framework presented in Fig. 2. Here we add a number of simulated probability vectors, which can be considered as possible microscopic realizations of the system (indicated by thin x crosses). Each of these realizations corresponds, for instance, to allocating N particles to energy levels or throwing a dice N times under the EM model assumptions. (A) all the realizations are concentrated around the EM prediction, and the observed vector is situated inside this scatter of realizations. We conclude that the assumptions underlying the EM model can accurately reproduce the observations. (B) all the realizations are concentrated around the EM prediction, but the observed probability vector is situated outside this scatter of realizations. We conclude that the mechanisms at work in the observed system are not described correctly by the EM model. (C) the realizations are scattered over the entire feasible set. We cannot make any definite conclusion, even though the difference between prediction and observation may be large. (D) the feasible set is so small (i.e. the constraints in the EM model are so restrictive) that we cannot make any definite conclusion, even though the correspondence between prediction and observation may be excellent.

opposing these two rationales, we feel that both provide an interesting, if different, perspective on the EM procedure. The two rationales are but expressions of the two interpretations of probability theory, i.e. the first is rooted in frequency-based probability theory, while the second is rooted in Bayesian probability theory.

The information-theoretical justification for EM states that the information contained in the problem formulation (set of states, constraints, and prior distribution) is represented in the least biased way by the EM probability distribution. Since a distribution with lower entropy encodes more information by definition, any probability distribution that satisfies the constraints but has lower entropy than the EM probability distribution encodes information that is not contained in the problem formulation. The EM solution takes into account all we know about the problem, and maximizes the uncertainty regarding all we do not know about the problem.

The combinatorial justification adds some additional structure to the problem, and, borrowing terminology from statistical mechanics, distinguishes a microscopic and a macroscopic level. Consider the example of Shipley (2009) of repeatedly throwing a dice. The variables at the microscopic level are the outcomes of all the $N = 20\,000$ dice throws. The variables at the macroscopic level are the relative frequencies of the six outcomes. The combinatorial argument is based on the fact that the microscopic states are not evenly distributed over the macroscopic states. Some macroscopic states contain many more microscopic states than others. The EM solution gives the macroscopic state that contains the largest number of microscopic states under the specified constraints.

The information-theoretical rationale is not based on a separation between microscopic and macroscopic scales, and is therefore more general. On the other hand, the combinatorial rationale can be used to illustrate a crucial property of the EM solution, which is illustrated graphically in Fig. 3. For a large scale separation, almost all microscopic states have their macroscopic state close to the EM macroscopic state (Jaynes 1979). Stated in terms of the dice throwing example, almost all series of $N = 20\,000$ dice throws that satisfy the constraints have relative outcome frequencies close to the EM prediction. This implies that an observed macroscopic state that differs significantly from the EM prediction has a very low probability of being the outcome of the mechanisms encapsulated in the model assumptions.

Since microscopic states are densely concentrated around the EM solution, this property allows us to make precise statistical inferences based on the EM prediction. The larger the scale separation, the more concentrated the microscopic states around the EM vector, and the more statistically significant a given difference between an observed probability distribution and the predicted EM probability distribution. Thus, the EM procedure allows us to make precise predictions only when scale separation is large. Whether these predictions match empirical data is an entirely different matter (Fig. 3A–B). Without such a scale separation, the EM solution is still the ‘best’ probability distribution we can infer from the available information (as information theory tells us), but the corresponding predictions will have a large uncertainty. In this case, differences

between predictions and observations will not be statistically significant and will not allow any strong conclusion to be drawn (Fig. 3C).

Therefore we feel that the information-theoretical and combinatorial rationales are complementary, and that both give valuable insights into the mechanisms underlying the EM procedure. Shipley’s (2009) attempt to turn them into two fundamentally different modelling procedures is unfortunate. He seeks to illustrate the supposed difference between the two procedures by opposing the example of repeatedly throwing a dice (2009) and the example of allocating particles to energy levels (Haegeman and Loreau 2008). But these two examples are strictly equivalent: simply replace ‘face of dice’ by ‘energy level’ (there are S of them) and ‘dice throw’ by ‘particle’ (there are N of them). In both examples, the number N acts as the scale separation parameter, which has to be large to obtain precise predictions. If predictions do not match observations, the EM model can be changed to get a better match. Shipley (2009) illustrates this iterative modelling approach with the dice throwing example, but the particle allocation example could be used just as well.

A trivial EM application in ecology

Classical applications of EM, such as repeatedly throwing a dice and allocating particles to energy levels, are representative of the situations depicted in Fig. 3A and 3B. But Shipley et al.’s (2006) application of EM corresponds to the situation depicted in Fig. 3D. In our contribution (Haegeman and Loreau 2008), we showed that the constraints they used are so restrictive that the feasible set is extremely small, so small that any feasible vector is very close to any other one. Since the observed vector of species abundances belongs to the feasible set by construction of the EM problem, any other feasible vector, including the EM prediction, is necessarily very close to it. In fact, we showed that the feasible set reduces to a single point in one third of the communities studied by Shipley et al. (2006), which makes the EM procedure superfluous.

Therefore Shipley et al.’s (2006) application of the EM algorithm is largely trivial, and does not bring any new insight compared with the initial data. The equations constraining species abundances are no longer underdetermined, and we could directly invert these equations to find the empirical species abundances again. This problem concerns the EM procedure itself, and is completely independent of the rationale one prefers to use. Put in an information-theoretical context, our critique states that there is no point in maximizing uncertainty if the information we have about the problem is such that there is no (or very little) uncertainty left. Put in a combinatorial context, our critique states that there is no point in looking for the macroscopic state that is realized by the largest number of microscopic states if there is only one (or very few similar) macroscopic state compatible with the problem assumptions.

By applying the EM method to a small feasible set, the basic mechanism allowing inference illustrated in Fig. 2 no longer works. Shipley (2009) argues that a small feasible set indicates that the constraints contain a lot of information,

so that the problem is increasingly better determined. But what do we learn from this kind of EM application? Whether a number of arbitrarily chosen traits uniquely determine species abundances is a mathematical question that does not specifically require the application of EM, but can be addressed with other, better suited techniques (Haegeman and Loreau 2008). Shipley et al.'s (2006) EM application does not allow any biologically relevant hypothesis to be supported or rejected.

Shipley (2009) claims that the EM problem is justified by information theory, and does not rely on empirical assumptions. But this statement is true only for the EM algorithm as such (Fig. 1, second box), and does not apply to model assumptions (Fig. 1, first box). He justifies the assumptions made by Shipley et al. (2006) as if they would be unique. But we prove in Appendix 2 that a different, and probably more appropriate, EM problem formulation is possible for the same system. Our alternative analysis uses a more detailed set of system states, in which not only expected species abundances are described, but also the abundance distributions of all species and the correlations between them. At this more detailed level of description, we show that Shipley et al.'s (2006) EM problem formulation implies an unrealistic Poisson species abundance distribution. By contrast, we obtain a more realistic log-series species abundance distribution using recent work on prior species abundance distributions (Banavar and Dewar 2007, Pueyo et al. 2007, Dewar and Porté 2008, Harte et al. 2008). Our analysis leads to EM predictions that differ from Shipley et al. (2006). Note, however, that, because of the small size of the feasible set, we cannot test the model assumptions against the empirical data in this particular case.

Toward non-trivial EM applications in ecology

The power of the EM mechanism, as illustrated in Fig. 3A and 3B, consists in obtaining precise predictions from very little information. This dramatic predictive power relies on the fact that macroscopic system variables are governed by statistical mechanisms, so that the modelling problem can be simplified to a description of the microscopic (set of states and prior distribution) and macroscopic (constraints) system structure, without taking into account all the detailed processes that occur within the system. The examples we have discussed, such as throwing a dice a large number of times and allocating a large number of particles to energy levels, indicate how and why EM then works. There is a huge unevenness in how microscopic states are distributed over macroscopic states. In fact, for a large scale separation, almost all microscopic configurations will lead to almost the same macroscopic behaviour.

These two examples are but the simplest instances of the EM mechanism that we are advocating. In the dice example, all throws are assumed to be independent from one another, so that a large number of independent events contribute to the relative frequencies of the various outcomes. Similarly, in the energy level example, all particles are assumed to be distributed independently over the various energy levels, so that a large number of independent

events contribute to the relative occupancies of these energy levels. Applications to more complex systems, in which interactions between system components are important, will typically lack this neat separation between microscopic and macroscopic levels. Nevertheless, EM can yield precise predictions even in the case of strongly interacting systems.

There may be little hope to ever find an ecological system in which a large number of independent events can be averaged out to describe overall system behaviour as in statistical mechanics. Nevertheless, statistical mechanisms also operate in ecological systems, and these mechanisms should be amenable to non-trivial applications of EM. We understand that this was the commendable ambition of Shipley et al. (2006), and we regret very much that Shipley (2009) now wants to reduce the application of EM to a mere analysis of how the constraints determine the prediction. This analysis can be performed using more powerful techniques such as those we presented in Haegeman and Loreau (2008). By contrast, the EM technique is most useful in predicting a system's macroscopic behaviour arising from statistical mechanisms.

Harte et al. (2008) recently presented an interesting EM application in ecology along these lines. They followed a two-step approach to predict the spatial structure of communities. First, they took the total number of individuals and the total number of species as macroscopic constraints, and used EM to predict the species abundance distribution. This is a log-series distribution, in agreement with previous work (Banavar and Maritan 2007, Pueyo et al. 2007, Dewar and Porté 2008). Next, they applied EM a second time to predict the spatial distributions of the various species, using their predicted total abundance as constraints. Lastly, by combining the species abundance distribution and the spatial distributions of the various species, they derived predictions for species–area relationships, which fit empirical data very well. In this case, the EM procedure yields a number of interesting ecological patterns, such as species abundance distributions and species–area relationships, after imposing a limited number of basic constraints, such as the total number of individuals, the total number of species and the total area occupied by the community. This disproportion between constraints and predictions, which is strikingly different from Shipley et al.'s (2006) application, strongly suggests a non-trivial application of EM. Although quantification of the number of independent events that have contributed to these patterns is un-feasible, we hypothesize that a large scale separation lies at the origin of these successful predictions. An interesting topic for future research would be to evaluate the extent to which scale parameters in the EM approach are related to prediction precision.

We believe that the most promising application of EM in ecology will be to assess whether and which ecological patterns are amenable to a statistical mechanistic approach. Whether one prefers to use a combinatorial or an information-theoretical rationale to justify EM will then be a minor issue. A critical ingredient of successful applications of EM seems to be the large separation between the microscopic scale on which ecological processes operate and the macroscopic scale on which the ecological system is observed. Another key requirement is that the problem be formulated in the most efficient form, by appropriately

choosing the system states, the constraints, and the prior distribution. Applied in this judicious way, EM will likely yield valuable, highly non-trivial new ecological insights.

Acknowledgements – We thank Claire de Mazancourt, Vincent Calcagno, Roderick Dewar, Andrew Gonzalez, Etienne Rampal and Dimitri Vanpeteghem for helpful discussions and suggestions. Michel Loreau acknowledges a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

References

- Banavar, J. R. and Maritan, A. 2007. The maximum relative entropy principle. – <<http://arxiv.org/abs/cond-mat/0703622>>.
- Dewar, R. C. and Porté, A. 2008. Statistical mechanics unifies different ecological patterns. – *J. Theor. Biol.* 251: 389–403.
- Haegeman, B. and Loreau, M. 2008. Limitations of entropy maximization in ecology. – *Oikos* 117: 1700–1710.
- Harte, J. et al. 2008. Maximum entropy and the state variable approach to macroecology. – *Ecology* 89: 2700–2711.
- Jaynes, E. T. 1957. Information theory and statistical mechanics. – *Phys. Rev.* 106: 620–630.
- Jaynes, E. T. 1979. Concentration of distributions at entropy maxima. – Reprinted in: Rosenkrantz, R. D. (ed.) 1983. *E. T. Jaynes: papers on probability, statistics and statistical physics*. Reidel, pp. 315–336.
- Jaynes, E. T. 2003. *Probability theory: the logic of science*. – Cambridge Univ. Press.
- Pueyo, S. et al. 2007. The maximum entropy formalism and the idiosyncratic theory of biodiversity. – *Ecol. Lett.* 11: 1017–1028.
- Shiple, B. et al. 2006. From plant traits to plant communities: a statistical mechanistic approach to biodiversity. – *Science* 314: 812–814.
- Shiple, B. 2009. Limitations of entropy maximization in ecology: a reply to Haegeman and Loreau. – *Oikos* 118: 152–159.

Appendix 1

In this appendix we illustrate how the various components of the initial formulation of the EM problem (set of states, constraints, prior distribution) are interrelated, and how different choices are possible for the same problem. We also introduce the notation used in Appendix 2, in which we present an alternative EM analysis of Shipley et al.’s (2006) data.

To state an EM problem, we first have to enumerate the possible states of the system. The set of system states is denoted by X , the individual system states by x . The set of

Table A1. The set of states for a system consisting of $N=3$ individuals and $S=2$ species. In the left column, states are described by a vector \vec{N} that specifies how many individuals each species contains. In the right column, states are described by a vector \vec{S} that specifies to which species each individual belongs. Some states \vec{N} correspond to $M(\vec{N})=1$ state \vec{S} ; other states \vec{N} correspond to $M(\vec{N})=3$ states \vec{S} . This difference in multiplicity will eventually lead to different EM predictions.

$\vec{N}=(3,0)$	$\vec{S}=(1,1,1)$
$\vec{N}=(2,1)$	$\vec{S}=(1,1,2), \vec{S}=(1,2,1), \vec{S}=(1,2,2)$
$\vec{N}=(1,2)$	$\vec{S}=(1,2,2), \vec{S}=(2,1,2), \vec{S}=(2,2,1)$
$\vec{N}=(0,3)$	$\vec{S}=(2,2,2)$

states X can be finite, denumerable infinite, or continuous. All the examples discussed in this paper have a discrete set of states X . Probability distributions on the set of states X are denoted by P , so that the probability of a state $x \in X$ is given by $P(x)$. The constraints on the system can be expressed in terms of the probabilities $P(x)$. The prior distribution on the set of states X is denoted by Q , so that the prior probability of a state $x \in X$ is denoted by $Q(x)$.

There is a close connection between the set of states X , the constraints, and the prior distribution Q . Indeed, it is possible to formulate the same EM problem by making different choices for the set of states, the constraints and the prior distribution. As an example, we consider an EM problem that consists in determining the composition of a community made up of N individuals and S species. We contrast two different choices for the set of system states.

First choice of set of states

We choose system states x that specify how many individuals belong to each of the S species. The system state x is then a vector \vec{N} with S components: $x = \vec{N} = (N_1, N_2, \dots, N_S)$. However, only the vectors \vec{N} for which

$$\sum_{i=1}^S N_i = N \quad (3)$$

holds have the correct number of individuals N . This restriction can be incorporated in the set of states, i.e., we only consider states \vec{N} that have the correct number of individuals N (alternatively, Eq. 3 could be imposed as a constraint). For example, the vector $\vec{N}=(2,1)$ tells us that two individuals belong to the first species, and 1 individual to the second species. There are four possible states \vec{N} with $N=3$ and $S=2$, listed in the left column of Table A1.

Second choice of set of states

We choose system states x that specify to which species each individual belongs. The system state x is then a vector \vec{S} with N components: $x = \vec{S} = (S_1, S_2, \dots, S_N)$. For example, the vector $\vec{S}=(1,2,1)$ tells us that the first individual belongs to species 1, the second individual belongs to species 2, and the third individual belongs to species 3. There are $2^3=8$ possible states \vec{S} with $N=3$ and $S=2$, listed in the right column of Table A1.

The correspondence between vectors \vec{S} and vectors \vec{N} is illustrated in Table A1. For example, only $\vec{S}=(1,1,1)$ corresponds to $\vec{N}=(3,0)$, but there are three vectors \vec{S} corresponding to $\vec{N}=(2,1)$. Generally, there are

$$M(\vec{N}) = \binom{N}{N_1 \dots N_S} = \frac{N!}{N_1! \dots N_S!} \quad (4)$$

vectors \vec{S} corresponding to a given vector \vec{N} . Note that the multiplicity factor $M(\vec{N})$ can take very unequal values for large values of N . Evenly distributed communities are described by a vector \vec{N} with a large multiplicity $M(\vec{N})$; communities with a few highly abundant species are described by a vector \vec{N} with a small multiplicity $M(\vec{N})$.

The latter remark has important consequences for the formulation of the EM problem. Indeed, consider the following EM problem formulations (set of states and prior distribution):

Formulation 1

We choose system states $x = \vec{N}$ (with N individuals and S species, N and S fixed), and do not specify a prior distribution $Q(\vec{N})$. Hence, all vectors \vec{N} have the same prior probability.

Formulation 2

We choose system states $x = \vec{S}$ (with N individuals and S species, N and S fixed), and do not specify a prior distribution $Q(\vec{S})$. Hence, all vectors \vec{S} have the same prior probability.

The two problem formulations are not equivalent due to the multiplicity factor $M(\vec{N})$. Indeed, vectors \vec{N} do not all have the same prior probability in Formulation 2, because some vectors \vec{N} (namely those representing evenly distributed communities) correspond to (many) more vectors \vec{S} than other vectors \vec{N} . Formulation 2 will therefore, compared to Formulation 1, favour evenly distributed communities.

The difference between Formulations 1 and 2 can be compensated by introducing a prior distribution Q . For example, the following formulation is equivalent to Formulation 2:

Formulation 3

We choose system states $x = \vec{N}$ (with N individuals and S species, N and S fixed), and take as prior distribution

$$Q(\vec{N}) \propto M(\vec{N}) \propto \prod_{i=1}^S \frac{1}{N_i!} \quad (5)$$

Hence, vectors \vec{N} get a prior probability proportional to their multiplicity, so that all vectors \vec{S} have the same prior probability.

Appendix 2

Here we provide an alternative EM analysis of the trait-based study of plant communities performed by Shipley et al. (2006). Our analysis applies the EM procedure on a deeper level than Shipley et al. (2006), and illustrates that other EM analyses of the same data are possible, leading to different EM predictions.

System states

In Shipley et al. (2006) the system states correspond to the $S = 30$ plant species that were observed in at least one of the 12 communities. The EM algorithm is used to determine the probability p_i that an individual (or “resource unit”, see Shipley et al. (2006)) is allocated to species i ($i = 1, \dots, S$). We call this the one-individual description.

In our alternative EM analysis of Shipley et al.’s (2006) problem, we consider a much larger set of states. We describe a system state by a vector $\vec{N} = (N_1, \dots, N_S)$ with a fixed number of individuals N and a fixed number of species S (Appendix 1). Applying the EM algorithm yields a probability $P(\vec{N})$ for every abundance vector \vec{N} . We call this the many-individuals description.

To illustrate the difference between the one-individual and the many-individuals descriptions, we compute the number of one-individual and many-individuals states for a community of $N = 1000$ individuals. In the one-individual description, the number of states equals the number of species $S = 30$. In the many-individuals description, the number of states equals the number of vectors \vec{N} having $N = 1000$ individuals and $S = 30$ species, i.e.

$$\frac{(N + S - 1)!}{N!(S - 1)!} \approx 6 \times 10^{57}$$

The many-individuals description is thus much more detailed than the one-individual description. It describes not only the expected relative abundances p_i of the various species, but also the abundance distributions of all species and the correlations between them.

Constraints

We follow Shipley et al. (2006), and impose $T = 8$ trait constraints. Rather than using the probabilities p_i for a one-individual state i , the constraints have to be formulated in terms of the probabilities $P(\vec{N})$ for a many-individuals state \vec{N} . We denote by t_{ij} the trait value of species i for trait j ($i = 1, \dots, S$ and $j = 1, \dots, T$), by \bar{t}_j the community-aggregated trait values ($j = 1, \dots, T$), and by \bar{N} the expected number of individuals in the community. The trait constraints are

$$\sum_{\vec{N}} P(\vec{N}) \sum_{i=1}^S t_{ij} N_i = \bar{t}_j \bar{N} \quad \text{for all } j = 1, \dots, T \quad (6)$$

The left-hand side equals the expected value of trait j for the total community; the right-hand side uses the community-averaged trait values \bar{t}_j to express the same quantity.

The 8 trait constraints have to be supplemented with two additional constraints. We use the expected number of individuals \bar{N} in the community, which is related to the distribution $P(\vec{N})$ by

$$\sum_{\vec{N}} P(\vec{N}) \sum_{i=1}^S N_i = \bar{N} \quad (7)$$

There is also the normalization condition given by

$$\sum_{\vec{N}} P(\vec{N}) = 1 \quad (8)$$

Prior distribution

We look for an appropriate prior distribution $Q(\vec{N})$ on the system states \vec{N} . For EM applications like throwing a dice N times or allocating N particles to energy levels, the prior distribution (5) can be used, i.e.

$$Q_{\text{dice}}(\vec{N}) = \prod_{i=1}^S \frac{1}{N_i!} \quad (9)$$

As we showed in Appendix 1, this choice assigns the same prior probability to all vectors \vec{S} . For the dice throwing example, this means that all sequences of N outcomes have

an equal prior probability, which seems a reasonable assumption.

However, some authors have argued that the prior distribution relevant for ecological communities should be different (Banavar and Maritan 2007, Pueyo et al. 2007, Dewar and Porté 2008, Harte et al. 2008). Based on quite different ideas, they all arrived at a prior distribution $Q(\vec{N})$ of the form

$$Q_{\text{ecol}}(\vec{N}) = \prod_{i=1}^S \frac{1}{N_i + c} \quad (10)$$

Although they used different values for constant c , there seems to be general agreement that the factors in $Q_{\text{ecol}}(\vec{N})$ should be approximately $1/N_i$ for large N_i . The heavy-tail distribution of $Q_{\text{ecol}}(\vec{N})$ contrasts sharply with the quickly decaying behaviour of $Q_{\text{dice}}(\vec{N})$ for large N_i .

EM problem

Solving the EM optimization problem given the set of states, constraints, and prior distribution is merely a technical issue. We have to maximize the relative entropy (2) with respect to $Q(\vec{N})$,

$$\sum_{\vec{N}} P(\vec{N}) \ln \frac{P(\vec{N})}{Q(\vec{N})}$$

subject to the $T+2$ constraints (6–8). Introducing the Lagrange multipliers λ_j for the trait constraints (6) and λ_0 for constraint (7), the EM solution reads

$$P(\vec{N}) \propto Q(\vec{N}) \prod_{i=1}^S e^{-\left(\lambda_0 + \sum_{j=1}^T \lambda_j t_{ij}\right) N_i} \quad (11)$$

The proportionality constant can be determined from the normalization constraint (8).

First, consider the prior $Q_{\text{dice}}(\vec{N})$. Substituting (9) into (11),

$$P(\vec{N}) \propto \prod_{i=1}^S \frac{1}{N_i!} e^{-(\lambda_0 + \sum_{j=1}^T \lambda_j t_{ij}) N_i}$$

we obtain independently distributed species abundances N_i . The abundance N_i of species i obeys a Poisson distribution with mean \bar{N}_i given by

$$\bar{N}_i = e^{-\left(\lambda_0 + \sum_{j=1}^T \lambda_j t_{ij}\right)} \quad (12)$$

To determine the Lagrange multipliers, we substitute (12) into (6),

$$\sum_{i=1}^S t_{ij} e^{-\left(\lambda_0 + \sum_{k=1}^T \lambda_k t_{ik}\right)} = \bar{t}_j \bar{N} \quad \text{for all } j = 1, \dots, T \quad (13)$$

Similarly, substituting (12) into (7) gives

$$\sum_{i=1}^S e^{-\left(\lambda_0 + \sum_{j=1}^T \lambda_j t_{ij}\right)} = \bar{N} \quad (14)$$

Dividing (13) by (14), we get

$$\sum_{i=1}^S t_{ij} \frac{e^{-\sum_{k=1}^T \lambda_k t_{ik}}}{\sum_{m=1}^S e^{-\sum_{l=1}^T \lambda_l t_{ml}}} = \bar{t}_j \quad (15)$$

which is a system of T equations for the T Lagrange multipliers λ_j . One can check that equations (15) are identical to the equations provided by Shipley et al. (2006). As a result, the Lagrange multipliers of both problems are identical, and the solution (p_1, \dots, p_S) of Shipley et al. (2006) satisfies

$$p_i = \frac{e^{-\sum_{k=1}^T \lambda_k t_{ik}}}{\sum_{m=1}^S e^{-\sum_{l=1}^T \lambda_l t_{ml}}} = \frac{\bar{N}_i}{\bar{N}}$$

This formula shows that the one-individual problem is embedded in the many-individuals problem with the prior $Q_{\text{dice}}(\vec{N})$. The one-individual solution (p_1, \dots, p_S) is recovered as the expected relative abundance \bar{N}_i/\bar{N} of the many-individuals solution $P(\vec{N})$.

Next, we consider the prior $Q_{\text{ecol}}(\vec{N})$. Substituting (10) into (11),

$$P(\vec{N}) \propto \prod_{i=1}^S \frac{1}{N_i + c} e^{-\left(\lambda_0 + \sum_{j=1}^T \lambda_j t_{ij}\right) N_i}$$

we obtain independently distributed species abundances N_i . The probability distribution of the abundance N_i of species i is closely related to the log-series distribution. The equations for the Lagrange multipliers λ_j can be derived in a similar way as we did for the prior $Q_{\text{dice}}(\vec{N})$. However, in the case of the prior $Q_{\text{ecol}}(\vec{N})$, these equations are not related to the equations for the Lagrange multipliers of Shipley et al. (2006). As a result, the solution of the many-individuals problem with the prior $Q_{\text{ecol}}(\vec{N})$ is not related to the solution of the one-individual problem.

Discussion

We have presented a many-individuals EM analysis as an alternative to the one-individual approach in Shipley et al. (2006). In the one-individual description, the choices for the EM model assumptions appeared to be unique. However, the analysis on the many-individuals level shows that this uniqueness is deceptive; on the many-individuals level different EM approaches are possible. In particular, we used two different choices for the prior distribution, $Q_{\text{dice}}(\vec{N})$ and $Q_{\text{ecol}}(\vec{N})$.

Using the prior $Q_{\text{dice}}(\vec{N})$, we retrieved the one-individual problem in a many-individuals setting. This result is not surprising since $Q_{\text{dice}}(\vec{N})$ yields an analogous embedding for the examples of repeatedly throwing a dice and allocating particles to energy levels. The statistics of the latter two problems are correctly described by a Poisson

distribution. However, a Poisson species abundance distribution is unrealistic from an ecological point of view.

Using the prior $Q_{\text{ecol}}(\vec{N})$, we obtained a log-series species abundance distribution, for which there is much more empirical support than for the Poisson distribution. Again, this result is not surprising, as $Q_{\text{ecol}}(\vec{N})$ was devised to obtain realistic species abundance distribution (Banavar and Maritan 2007, Pueyo et al. 2007, Dewar and Porté 2008, Harte et al. 2008). The many-individuals EM problem with

the prior $Q_{\text{ecol}}(\vec{N})$ is unrelated to the one-individual EM problem studied by Shipley et al. (2006).

We conclude that the EM analysis of Shipley et al. (2006) is not unique. The assumptions underlying their analysis are somehow hidden in the one-individual description, but these assumptions appear clearly in the many-individuals description. We have showed that these assumptions may not be most appropriate, and that other assumptions yield more realistic species abundance distributions.