



## Limitations of entropy maximization in ecology

Bart Haegeman and Michel Loreau

*B. Haegeman (bart.haegeman@inria.fr), INRA, UR50, Laboratory of Environmental Biotechnology, Avenue des Etangs, FR-11100 Narbonne, France and INRA-INRIA 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, Montréal, QC, H3A 1B1, Canada.*

Applying ideas of statistical mechanics in ecology have recently received quite some attention. The entropy maximization (EM) formalism looks particularly attractive, as it provides a simple algorithm to infer detailed system variables from a limited number of constraints. However, we point out that a blind application of this formalism can easily lead to wrong conclusions. To illustrate this, we reanalyze an ecological data set that has been used to claim the good performance of EM in predicting species abundances from trait measurements. We show that these results are entirely due to the restrictive constraints, and do not provide any support for the applicability of EM in ecology. By comparing with a simple example from physics, we indicate which characteristic mechanism of EM, and of statistical mechanics in general, is missing for the ecological example. This analysis introduces a series of methods to evaluate future attempts to apply EM in ecology.

Ecological modelling typically follows a bottom–up approach. In community ecology, for example, models are constructed by specifying the characteristics of different species, together with their interactions, in order to obtain a description of the total community. This often leads to intricate mathematical models due to the large number of individuals and/or species, and their complicated interactions. Comparable systems in physics, where many microscopic particles interact to yield macroscopic behavior, are typically modelled based on a top–down approach. Particles and their interactions are then described only schematically, as it turns out that a lot of microscopic details are washed out when passing to the macroscopic level. The approximation techniques that enable precise macroscopic predictions are the subject of statistical mechanics.

Can techniques borrowed from statistical mechanics be of use in ecology? The situation in ecology is much more delicate as even the distinction between microscopic and macroscopic levels is not obvious. This stands in sharp contrast with statistical mechanics, where the separation between the two levels is huge: Avogadro's number ( $\approx 10^{23}$  particles mole<sup>-1</sup>) is an illustrative quantity. Therefore one should not expect that results from statistical mechanics can be immediately transferred to ecology. On the other hand, some aspects of statistical mechanics have already found wide application outside physics. The entropy maximization (EM) method, for example, provides an elegant formalism to perform statistical mechanical computations (Jaynes 1957). In this method, the microscopic degrees of freedom are described by the probability distribution that maximizes the Shannon entropy subject to a set of

macroscopic constraints. Entropy maximization has received the status of a general inference technique, which is applied to a wide range of problems (e.g. image reconstruction, Jaynes 2003), and thus it seems to be a good starting point to exploit statistical mechanics in ecology.

Entropy maximization has recently been applied to several issues in community ecology. A number of papers have used EM to derive species abundance distributions (SADs) (Banavar and Maritan 2007, Pueyo et al. 2007, Dewar and Porté 2008). These authors assumed an a priori SAD, imposed community-level constraints (e.g. the total number of individuals is fixed), and applied the EM method to generate the corresponding a posteriori SAD. This application of EM has been relatively successful since it allowed common ecological SADs to be recovered. In particular, the logseries distribution seems to be encoded most robustly in the formalism (Banavar and Maritan 2007, Pueyo et al. 2007, Dewar and Porté 2008). By slightly modifying the constraints or the a priori distribution, Pueyo et al. (2007) and Banavar and Maritan (2007) showed that the lognormal distribution and the patterns predicted by neutral community models can also be generated. These results suggest that standard SADs have a statistical basis, and contain only limited ecological information. It supports the conclusion that predicting a reasonable SAD is only a weak indication of the fruitfulness of an ecological theory (McGill et al. 2007).

These applications of the EM method do not assume differences among species, and thus are in the spirit of neutral community models. As a consequence, however,

only abundances of unspecified species can be predicted, leading to rather weak tests of the usefulness of entropy maximization. To collect stronger evidence, two recent studies incorporated differences among species in the EM formalism, and solved for the abundances of particular species (Shiple et al. 2006, Dewar and Porté 2008). Dewar and Porté (2008) assumed that various species had different resource consumption traits, and applied EM by imposing the total resource use of the community. They showed that this again leads to lognormal and neutral-like SADs, and more interestingly, to a number of diversity–productivity relationships. The latter result is remarkable as mechanistic models that predict these types of relationships are often much more complicated. It suggests that not only SADs, but also other ecological patterns could have a largely statistical origin, and could thus be prone to successful application of the EM method.

Although these theoretical insights are promising, only empirical data can ultimately provide strong evidence for entropy maximization in ecology. Dewar and Porté (2008) analyzed one plant community, showing some agreement between observed and predicted species abundances. Shipley et al. (2006) studied a more extensive data set, and claimed that entropy maximization allowed accurate prediction of species abundances based on species traits and their community-aggregated values. This result has been questioned (Marks and Muller-Landau 2007, Roxburgh and Mokany 2007), in particular by pointing out the circularity in Shipley et al.’s (2006) application of EM since they first computed constraints based on species abundances, and then predicted species abundances from these constraints. Here we reanalyze Shipley et al.’s (2006) data, and argue that a more fundamental problem invalidates their conclusion, i.e. the basic mechanism underlying any useful application of entropy maximization does not operate. To highlight this problem, we compare their ecological application with a simple system from physics, which is in form closely related to the problem of predicting species abundances, but which does exhibit the fundamental mechanism underlying entropy maximization. We then use this case study as an example to discuss more generally the limitations of EM in ecology.

### Entropy maximization in physics: a model system

Here we sketch the application of entropy maximization for a physical system, which will be compared with an ecological system in the following sections. The EM method has a long history, dating back to Boltzmann (end of 19th century), and more elaborate discussions can be found in textbooks such as Schroeder (2000) and Jaynes (2003).

Consider a box filled with non-interacting particles. Every particle occupies one out of  $S$  energy levels. The energy of level  $i$  is denoted by  $\epsilon_i$ , and the number of particles occupying that level by  $n_i$ , with  $i=1, \dots, S$  (Table 1). Suppose that at a certain time, we measure the total number of particles present in the box, denoted by  $N$ , and the total energy of these particles, denoted by  $E$ . Because the particles are non-interacting, the total energy  $E$

equals the sum of the energies of the individual particles. We then have the following relations:

$$\begin{aligned} \sum_{i=1}^S n_i \epsilon_i &= E \\ \sum_{i=1}^S n_i &= N \\ n_i &\in \{0, 1, 2, \dots\} \quad i = 1, \dots, S \end{aligned} \quad (1)$$

Let the relative occupation of energy level  $i$  be  $p_i = n_i/N$ . For large  $N$  these relative occupations can be regarded as continuous variables, and the constraints become

$$\begin{aligned} \sum_{i=1}^S p_i \epsilon_i &= \frac{E}{N} \\ \sum_{i=1}^S p_i &= 1 \\ p_i &\geq 0 \quad i = 1, \dots, S \end{aligned} \quad (2)$$

Obviously, these two equations do not allow us to reconstruct the vector  $p = (p_1, \dots, p_S)$  (if  $S > 2$ ). Nevertheless, EM provides a method to estimate these numbers. It stipulates that the Shannon entropy,

$$H(p) = - \sum_{i=1}^S p_i \ln p_i \quad (3)$$

should be maximized subject to the constraints (Eq. 2). It can be shown that this rule yields a unique vector of relative occupations, which we will denote by  $p^{\max H}$ .

The rationale behind this application of EM goes as follows. Any vector  $n$  satisfying (Eq. 1) represents a possible distribution compatible with the measurements. However, some of these vectors can be realized in many more ways than others. Indeed, the number of ways  $N$  particles can be distributed over  $S$  energy levels to yield the vector of occupation numbers  $n = (n_1, \dots, n_S)$  is given by the multinomial coefficient

$$\binom{N}{n_1 \dots n_S} = \frac{N!}{n_1! \dots n_S!} \quad (4)$$

As soon as  $N$  is large enough, these coefficients take very unequal values. For example, the multinomial coefficient for  $n_1 \approx n_2 \approx \dots \approx n_S$  is much larger than the coefficient for  $n_1 \gg n_2, \dots, n_S$ . Therefore, distributing  $N$  particles randomly over  $S$  energy levels typically yields a more or less even distribution of occupation numbers. When constraints are taken into account, we simply look for the vector  $n$  that has the largest multinomial coefficient consistent with these constraints. It can be shown that for large  $N$ , maximizing Shannon entropy (with continuous arguments  $p$ , which are easier to manipulate) leads approximately to the same result as selecting the largest multinomial coefficient.

The nature of the EM method is thus combinatorial. However, for its application to give meaningful results, more subtle aspects have to be considered. For example, when using multinomial coefficients, we implicitly assume that the different ways a vector of occupation number  $n$  can be realized should be counted with equal weight. This is an

Table 1. Explanation of mathematical symbols.

Symbol	Explanation
S	Number of species in community
N	Number of individuals in community
$n_i, n$	Abundance of species $i$ , abundance vector $n = (n_1, n_2, \dots, n_S)$ , thus $N = \sum_i n_i$
$p_i, p$	Relative abundance of species $i$ , relative abundance vector $p = (p_1, p_2, \dots, p_S)$ , thus $p_i = \frac{n_i}{N}$
$p_i^{\text{obs}}, p^{\text{obs}}$	Observed relative abundance
T	Number of traits measured in community
$t_{ij}$	Value of trait $j$ for species $i$
$\bar{t}_j$	Community-aggregated trait, i.e. mean value for trait $j$ in community, $\bar{t}_j = \sum_i p_i^{\text{obs}} t_{ij}$
H, C, D	Functions on relative abundance vectors: Shannon entropy, Simpson concentration, and distance to observed abundance vector $p^{\text{obs}}$ , resp.
F	Set of relative abundance vectors $p$ that satisfy all the constraints, i.e. the feasible set
$p^{\text{max}}, p^{\text{avg}}$	Predicted relative abundance vector, by maximizing Shannon entropy H or by averaging over the feasible set F, resp.

assumption about the underlying system structure that is often difficult to assess, and that sometimes requires counter-intuitive hypotheses to give correct predictions. The choice of the constraints taken into account can also be particularly important. Obviously, leaving out constraints that carry crucial information will yield an inaccurate EM prediction. On the other hand, constraints should be carefully selected so that EM can be used to its full strength.

It is interesting to already note the parallels with ecological systems. Replacing ‘particle’ by ‘individual,’ and ‘energy level’ by ‘species,’ the vectors  $n$  and  $p$  represent absolute and relative species abundances, respectively. The Shannon entropy (Eq. 3) is then identical to the well-known Shannon diversity index. The multinomial coefficient (Eq. 4) has also been used in ecology, notably as the Brillouin diversity index (Magurran 2004),

$$B(n) = \frac{1}{N} \ln \binom{N}{n_1 \dots n_S}$$

### Entropy maximization in ecology: looking for empirical evidence

Although the analogy between energy level occupation numbers and species abundances looks obvious, justifying the application of EM is much more involved in ecology than in physics. The theoretical argument for the application of EM, as outlined in the previous section, is based on the separation between microscopic and macroscopic scales, and on the highly unequal values the multinomial coefficient (Eq. 4) can take. Whether the number of components in ecological systems is sufficient to apply the statistical methods of large-number systems remains unclear. Ecosystems have been characterized as medium-number systems for which both the approaches of mechanistic and statistical modelling are problematic: there are too many components to describe each of them explicitly, and there are not enough components to work with averaged properties (O’Neill et al. 1986).

As theory is lacking, empirical studies should settle the question whether the EM method is useful to predict species abundances. But the corresponding data analysis is

subtle, as the EM algorithm requires a good deal of a priori information (microscopic system structure and macroscopic constraints, see previous section). Therefore it does not suffice to evaluate the prediction performance as such, but one should also exclude the possibility that good prediction results are simply due to the a priori information rather than to EM as such. Failing to do so can lead to wrong conclusions, as we illustrate now with an ecological data set.

Shiplely et al. (2006) presented an application of the EM method to plant ecology. In a series of vineyards in the south of France, the relative abundances of 30 plant species were measured ( $p_i^{\text{obs}}$  for species  $i$ , vector  $p^{\text{obs}}$ ; experimental methods are described in Garnier et al. 2004). Eight plant traits (such as plant height, leaf thickness, perennial versus annual) were measured for each species ( $t_{ij}$  for trait  $j$  of species  $i$ ; experimental methods are described in Vile et al. 2006). So-called community-aggregated values  $\bar{t}_j$  were then computed, which average species trait values weighted by observed species abundances,

$$\bar{t}_j = \sum_i p_i^{\text{obs}} t_{ij}$$

This yields a number of constraints that predicted species relative abundances ( $p_i$  for species  $i$ , vector  $p$ ) must meet:

$$\begin{aligned} \sum_{i=1}^S p_i t_{ij} &= \bar{t}_j \quad j = 1, \dots, T \\ \sum_{i=1}^S p_i &= 1 \\ p_i &\geq 0 \quad i = 1, \dots, S \end{aligned} \quad (5)$$

There are  $S = 30$  variables to be determined, for which  $T + 1 = 9$  equality constraints are imposed. The remaining  $S - T - 1 = 21$  degrees of freedom are fixed by maximizing the Shannon entropy (Eq. 3). This procedure was repeated for 12 data sets, corresponding to 12 vineyards that were abandoned over a period ranging from 2 to 42 years. The return to natural vegetation implied a significant decrease in species diversity: the younger plots contained from 8 to 16 of the 30 species, whereas the older ones only had 4 or 6 species. Note that at most half of the 30 species were observed on any site. In fact, for 256 out of the  $12 \times 30 = 360$  observed abundances,  $p_i^{\text{obs}} = 0$ .

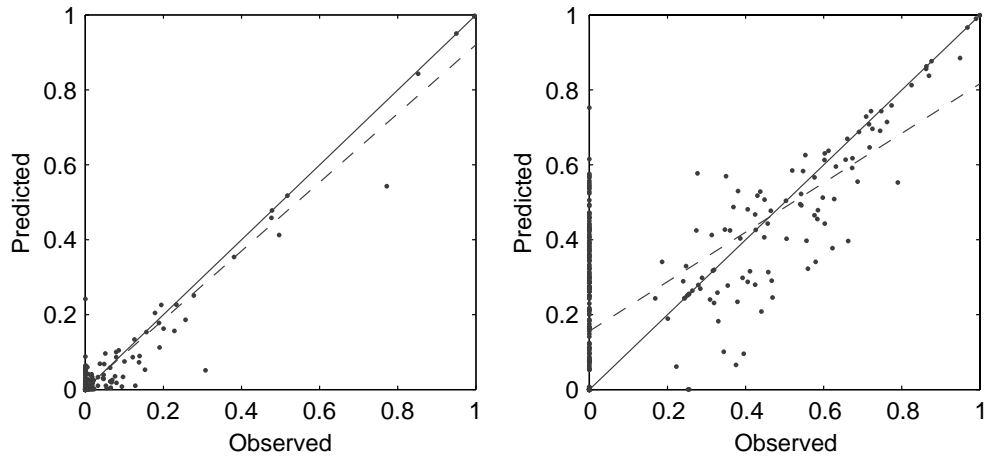


Figure 1. Predicted vs observed relative abundances for 12 plant communities with 30 species each. Observed abundances on x-axis, predicted abundances on y-axis. The prediction is obtained by maximizing the Shannon entropy conditional on community-aggregated trait values. Left: untransformed data, with correlation coefficient  $r=0.97$  and RMS prediction error  $e=0.03$ . Right: fifth root transformed data, with correlation coefficient  $r=0.71$  and RMS prediction error  $e=0.22$ . Full line: first bisector, corresponding to a perfect prediction. Dashed line: linear fit as used in the computation of the correlation coefficient.

Formally, Shipley et al.'s (2006) approach has the features of the EM method: an indeterminacy is lifted by maximizing the Shannon entropy of a probability distribution. Due to the concavity of Shannon entropy, the unique maximum can be found by standard numerical techniques (Appendix 1). We used the constrained nonlinear optimization algorithm `fmincon` in Matlab.

Figure 1 compares the observed and predicted species abundances for the 12 sites. The left panel is essentially a reproduction of Fig. 2 from Shipley et al. (2006). The Pearson correlation coefficient  $r=0.97$  is identical to the value reported in Shipley et al. (2006). Note, however, that this coefficient measures how well the data are approximated by the best linear fit, shown in Fig. 1 as a dashed line. Here we are interested instead in the prediction error of the EM solution. This can be quantified by the root mean squared (RMS) error of the predicted species abundance  $p_i^{\max H}$  with respect to the observed species abundance  $p_i^{\text{obs}}$ ,

$$e = \sqrt{\frac{1}{S} \sum_{i=1}^S (p_i^{\max H} - p_i^{\text{obs}})^2}$$

For our data set, the RMS prediction error is  $e=0.03$ .

But the left panel of Fig. 1 shows clearly that most of the data points are concentrated in the bottom left corner, i.e. there are too many species with low abundances  $p_i^{\text{obs}}$ , which precludes the use of a linear fit and makes it impossible to assess how well the prediction performs for rare species. To zoom in on the region of small abundances, we applied a fifth root transformation to both observed and predicted data. This transformation is more convenient than a logarithmic transformation to handle the numerous zeroes in the observed data (256 out of 360). We also checked that other root transformations led to the same conclusions. The correlation coefficient for the transformed data, shown in the right panel of Fig. 1, is now  $r=0.71$ , and the RMS

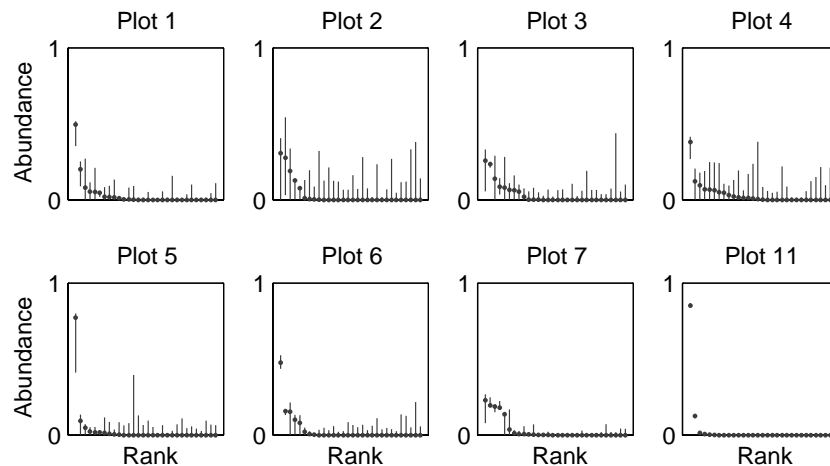


Figure 2. Rank–abundance curves, together with smallest and largest species abundance consistent with the constraints. Only 8 of 12 plots are shown, because for the other plots, the feasible set  $F$  consists of a single point. The dots give the rank-abundance curve of the observed abundance vector  $p^{\text{obs}}$ . The vertical lines through the dots connect  $p_i^{\min}$  and  $p_i^{\max}$  for the corresponding species in the corresponding community. The lower the diversity of the community, the smaller the feasible set  $F$ .

error of predicted vs observed species abundance after fifth root transformation,

$$e = \sqrt{\frac{1}{S} \sum_{i=1}^S ((p_i^{\max H})^{0.2} - (p_i^{\text{obs}})^{0.2})^2}$$

is  $e = 0.22$ .

The plot reveals that only the abundances of the commonest species are accurately estimated, a point that was also made by Marks and Muller-Landau (2007) after applying a logarithmic transformation to the data. Predictions become imprecise for species with  $p_i^{\text{obs}} < 0.1$ . It is also interesting to note that on average species abundances are underestimated for common species and overestimated for rare species. This is already visible in the linear fit for the untransformed data (Fig. 1, left), and is obvious in the linear fit for the fifth root transformed data (Fig. 1, right). This indicates that maximizing the Shannon entropy (Eq. 3) is too strong a condition to fill in the remaining degrees of freedom after imposing the constraints (Eq. 5).

### Entropy maximization in ecology: overly restrictive constraints

Although closer scrutiny shows that the predictive power of the EM formalism in the above example is much less impressive than claimed by Shipley et al. (2006), even good predictions would not be enough to prove the usefulness of the EM method in ecology. In particular, imposing such restrictive constraints that only vectors close to the experimentally observed one are possible, will lead to excellent results by any optimization method. In this case, EM would have no contribution at all in the good prediction performance.

It is not difficult to see that the latter problem could be an issue in the above ecological data set. Indeed, there is an alarming circularity in first computing  $\bar{t}_j$  from  $p^{\text{obs}}$ , and next imposing that average trait values should equal  $\bar{t}_j$  for all communities  $p$ . These conditions could very well select for vectors  $p$  close to  $p^{\text{obs}}$ . Our observation that only the commonest species are well predicted already suggests that the good fit between observed and predicted data might be simply due to the major influence common species have in the average trait values  $\bar{t}_j$ . This potential problem of circularity was also raised by Marks and Muller-Landau (2007) and Roxburgh and Mokany (2007).

To go further and understand the role played by the macroscopic constraints, we performed a detailed study of the set of abundance vectors compatible with community-aggregated traits. We call a vector  $p$  feasible if it satisfies all the constraints (Eq. 5). We call the set of all feasible points the feasible set, and denote it by  $F$ ,

$$F = \left\{ p \left| \begin{array}{l} \sum_i p_i t_{ij} = \bar{t}_j \text{ for all } j, \\ \sum_i p_i = 1, \quad p_i \geq 0 \text{ for all } i \end{array} \right. \right\}$$

This set, defined by a number of linear equalities and inequalities, is a so-called polyhedron, and therefore is convex (Appendix 1). Because its elements are abundance

vectors  $p$ , it is a subset of the simplex with dimensionality  $S = 30$ , and hence is bounded. Moreover, as the inequalities are not strict, it is also closed. Note that the feasible set is not empty because  $p^{\text{obs}} \in F$ .

The relatively good correlation in Fig. 1 might be explained by a small set  $F$ , so that any vector  $p \in F$  is necessarily close to  $p^{\text{obs}} \in F$ . To get a rough idea of the extension of  $F$ , we proceeded as follows. For every species  $i$ , we looked at the smallest and largest abundance  $p_i^{\text{obs}}$  compatible with  $p \in F$ . In other words, we solved optimization problems of the form

$$\begin{aligned} p_i^{\min} &= \min \{ p_i \mid p \in F \} \\ p_i^{\max} &= \max \{ p_i \mid p \in F \} \end{aligned} \quad (6)$$

Optimizing a linear function over a bounded, closed, convex set can be efficiently done by the simplex method. We used the Matlab implementation `linprog`.

Surprisingly, the feasible set  $F$  was found to consist of a single point for 4 of the 12 plots, i.e.  $F = \{p^{\text{obs}}\}$ . These plots (numbers 8, 9, 10 and 12) are all old and species poor. As a consequence, all possible methods to select a vector in the feasible set will yield  $p^{\text{obs}}$ . Figure 2 shows the extension of the feasible set  $F$  for the other 8 plots. It plots rank-abundance curves (dots), together with extremal species abundances  $p_i^{\min}$  and  $p_i^{\max}$  (vertical lines). We noted already that older plots are less diverse. Now we see that older plots also have a smaller feasible set  $F$ , and thus more restrictive constraints.

The latter observation can be understood as follows. Communities that are dominated by a few species have community-aggregated traits  $\bar{t}_j$  close to the traits  $t_{ij}$  of the dominant species. Therefore, any other feasible community, i.e. any community with the same aggregated traits  $\bar{t}_j$ , will necessarily contain the same common species. Similarly, species that are rare in the observed community can only have small abundances in other feasible communities, otherwise the aggregated traits would shift towards their own traits. Therefore less diverse communities impose more restrictive constraints on the feasible set  $F$ . This explains the small extensions of the feasible set in plot 11 and the singleton feasible sets in plots 8, 9, 10 and 12.

On the other hand, abundances  $p_i$  consistent with the constraints still vary widely in diverse plots, e.g. plot 2. Note, however, that these graphical representations give an overestimation of the extent of the set  $F$ , as the extrema  $p_i^{\min}$  and  $p_i^{\max}$  cannot be attained simultaneously for several species. To further zoom in on  $F$ , we take random samples from the convex set  $F$  with a uniform distribution. The algorithm used to do so is described in Appendix 1. For every sample, the root mean squared (RMS) distance  $D$  to the observed abundance vector  $p^{\text{obs}}$  is computed,

$$D(p) = \sqrt{\frac{1}{S} \sum_{i=1}^S (p_i - p_i^{\text{obs}})^2} \quad (7)$$

The distribution of this distance in the feasible set  $F$  provides an idea of the geometry of  $F$  with respect to  $p^{\text{obs}}$ .

Figure 3 shows the frequency distribution of  $D$  for the 8 plots that have a non-trivial feasible set. These distributions are wider in more diverse communities. A vertical line indicates the EM solution. Clearly the EM solution  $p^{\max H}$  does not hold a special status among the feasible abundance

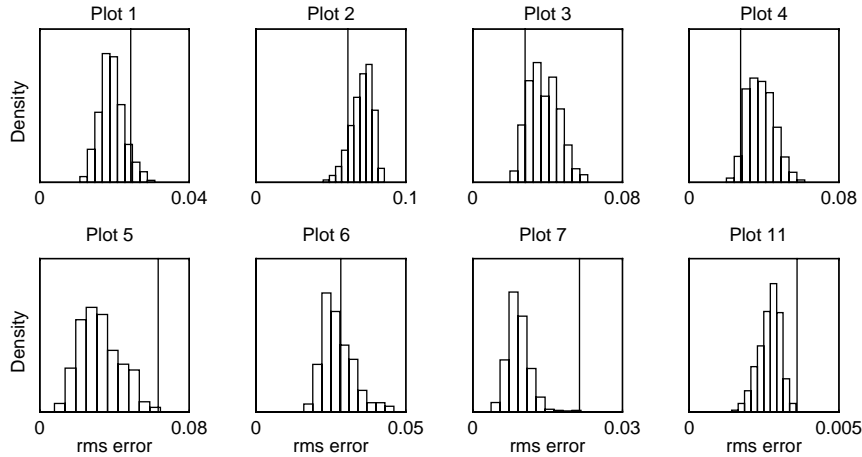


Figure 3. Distribution of the distance to the observed abundance vector for uniform samples in the feasible set. Eight of 12 plots were considered, because for the other plots, the set  $F$  consists of a single point. The histograms were constructed from  $10^4$  passages of a random walk on  $F$ , see Appendix 1 for details. The vertical line corresponds to the maximum entropy solution  $p^{\max H}$ . This abundance vector is not significantly closer to  $p^{\text{obs}}$  than any other guess in the feasible set  $F$ .

vectors  $p$ . The value  $D(p^{\max H})$  seems to be picked rather randomly from the distance distribution. This sheds quite a different light on the excellent correlation in Fig. 1, which now appears to be entirely due to the restrictive constraints, not to EM.

Shipley et al. (2006) justified the application of the EM method to this problem by an argument borrowed from statistical mechanics. They assumed that, to a first approximation, resources are attributed independently to different species. Interestingly, we can compare their EM solution with another prediction that does not rely on this hypothesized microscopic structure. Indeed, if only trait constraints are taken into account, the species abundance vector most indicated by the data is simply the center of the feasible set  $F$ . This center vector, denoted by  $p_{\text{avg}F}$ , can be obtained by averaging abundance vectors, uniformly sampled from the feasible set  $F$ . Table 2 compares the performance of the estimated abundance vectors  $p^{\max H}$  and  $p^{\text{avg}F}$ . The distance  $D(p^{\max H})$  is not significantly smaller than  $D(p^{\text{avg}F})$ . Therefore, this data set does not provide evidence for any usefulness of EM in the present problem.

The performance of other estimates of the vector  $p^{\text{obs}}$  is also shown in Table 2. Instead of maximizing the Shannon entropy  $H$  we computed the abundance vector that minimizes the entropy (Marks and Muller-Landau 2007). We also minimized and maximized Simpson concentration  $C$ ,

$$C(p) = \sum_{i=1}^S p_i^2 \quad (8)$$

and, as worst-case scenario, maximized the distance  $D$  to the observed abundance vector  $p^{\text{obs}}$  (Table 1). Whereas Shannon entropy  $H$  is concave, both Simpson concentration  $C$  and the distance function  $D$  are convex. Globally maximizing  $H$  or minimizing  $C$  reduces to a local optimization problem, performed in Matlab with the algorithm `fmincon`. On the other hand, minimizing  $H$ , maximizing  $C$  or maximizing  $D$  requires the knowledge of

the extreme points of  $F$  (Appendix 1). To compute them, we exploited a link between ecological trait constraints and (metabolic) reaction network equilibria (Appendix 1). The intricate structure of the feasible set  $F$  (excluding the 4 plots with  $F = \{p^{\text{obs}}\}$ ) is illustrated by the thousands of extreme points, with a maximum of 23044 points for plot 5.

The estimate  $p^{\text{avg}F}$  is on average the best predictor, followed by  $p^{\max H}$  and  $p^{\min C}$ . Note that these optimization problems generally possess solutions in the interior of the set  $F$ , whereas the other three (min  $H$ , max  $C$  and max  $D$ ) attain their optimum in an extreme point, which on average lies further from the observed abundance vector  $p^{\text{obs}}$ . This explains the poorer performance of the latter three estimates.

In short, we presented evidence that the EM solution has no special properties to estimate the vector  $p^{\text{obs}}$  in the given context. Any vector arbitrarily picked from the feasible set  $F$  has comparable predictive power. This shows that the correlation as presented in Fig. 1 cannot be attributed to entropy maximization, but is entirely due to the constraints. Indeed, they are so restrictive that all feasible vectors  $p$  are contained in a small set around the sought vector  $p^{\text{obs}}$ .

### Entropy maximization in physics: sharply peaked distribution

We now return to the example from physics introduced earlier. Analyzing this system in the same way as the vineyard data set reveals a mechanism that is absent from the ecological system. We claim that this mechanism is characteristic of EM in statistical mechanics, and should be exhibited by any successful application of EM in ecology.

We used the following parameters. We took  $S=30$  energy levels, i.e. a number equal to the number of species in the plant plots. The energy levels were chosen as  $\epsilon_i=i$ , for  $i=1, \dots, 30$ , thus corresponding to an equally spaced spectrum. However, the results below do not depend on this choice. The measurable quantities are the number of particles  $N$  and the total energy  $E$ , which we assumed to be

Table 2. RMS errors for different criteria to select an abundance vector  $p$ . Only the 8 plots with non-trivial feasible set  $F$  were considered. The last row gives the RMS errors for our example from physics. The first two columns (maximizing  $H$  and minimizing  $C$ ) correspond to optimization problems with a solution in the interior of  $F$ ; the next three columns (minimizing  $H$ , maximizing  $C$  and maximizing  $D$ ) correspond to optimization problems with an extreme point of  $F$  as solution; the last column is based on the estimated center of the set  $F$ .

	max $H$	min $C$	min $H$	max $C$	max $D$	avg $F$
Plot 1	0.024	0.035	0.032	0.024	0.056	0.016
Plot 2	0.061	0.066	0.101	0.101	0.107	0.068
Plot 3	0.028	0.036	0.089	0.089	0.091	0.030
Plot 4	0.028	0.038	0.075	0.076	0.080	0.037
Plot 5	0.064	0.087	0.032	0.030	0.100	0.033
Plot 6	0.028	0.031	0.024	0.024	0.055	0.025
Plot 7	0.021	0.022	0.010	0.012	0.047	0.010
Plot 11	0.004	0.005	0.005	0.005	0.005	0.003
Physics	0.004	0.018	0.176	0.176	0.176	0.068

$N = 1000$  and  $E = 4000$ . The average energy per particle in the system was therefore  $E/N = 4$ , comparable to a community-aggregated trait value.

The EM problem (Eq. 2) has  $S = 30$  variables (the relative abundance vector), one normalization constraint (the sum of the relative abundances equals one), and one trait constraint (the average energy per particle). The solution  $p^{\max H}$  can be computed analytically (Schroeder 2000),

$$p_i^{\max H} = \frac{e^{-\beta \epsilon_i}}{\sum_j e^{-\beta \epsilon_j}} \quad (9)$$

where  $\beta$  is a Lagrange multiplier that has to be determined from the condition

$$\sum_i p_i^{\max H} \epsilon_i = \frac{E}{N}$$

For our parameter values,  $\beta = 0.2872$ .

The left panel of Fig. 4 shows the rank–abundance curve of this abundance vector, as well as the smallest and largest species abundances consistent with the constraints. These were computed using the optimization problems (Eq. 6). The extension of the feasible set turns out to be larger than

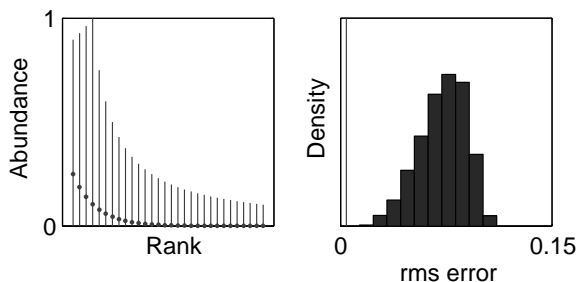


Figure 4. Structure of the feasible set for the example from physics. Left: Rank–abundance curve, together with margins for the abundances compatible with the constraints. Compared with the ecological data sets (Fig. 2), the feasible set is now more extended. Right: distribution of the distance to a simulated abundance vector  $p^{\text{sim}}$  for uniform samples in the feasible set. The vertical line corresponds to the maximum entropy solution  $p^{\max H}$ . Compared with the ecological data sets (Fig. 3), the performance of the maximum entropy estimate is now much better.

for the ecological data sets, which is not surprising as there are fewer constraints in this case.

Next, we randomly sampled abundance vectors  $n$ . It is important to note that there are two distinct ways to achieve this:

1. A first procedure consists in randomly generating particle configurations that satisfy the constraints. We attributed any of the  $N$  particles to one of the  $S$  energy levels, such that the total energy equals  $E$ . We thus sampled from the probability distribution where energy levels are a priori occupied with equal probability under the total energy constraint. An efficient algorithm to sample particle configurations from this distribution is outlined in Appendix 1.
2. A second procedure consists in uniformly sampling from the feasible set  $F$ , as already used in the previous section and described in Appendix 1.

Note that the link between these two sampling procedures is given by the multinomial coefficient (Eq. 4). Consider two abundance vectors  $m$  and  $n$  satisfying the constraints (Eq. 1), and assume that

$$\binom{N}{m_1 \dots m_S} > \binom{N}{n_1 \dots n_S}$$

Then, in the first sampling procedure, configurations corresponding to  $m$  are expected to be sampled more frequently than configurations corresponding to  $n$ . In the second sampling procedure, both feasible vectors  $m$  and  $n$  are equally likely to be sampled.

If we assume that all constraints have been taken into account, observing the particle configuration of the physical system at a certain instant corresponds to the first sampling procedure (Schroeder 2000, Banavar and Maritan 2007). Curiously enough, this is only true for physical systems described by classical mechanics. Quantum mechanics corresponds to the second sampling procedure. Thus, this procedure replaces the simulation of the abundance vector of the physical system. We generated 100 such abundance vectors  $p^{\text{sim}}$ , and computed the distance  $D(p^{\max H})$  to the maximum entropy solution  $p^{\max H}$ ,

$$D(p) = \sqrt{\frac{1}{S} \sum_{i=1}^S (p_i - p_i^{\text{sim}})^2} \quad (10)$$

The mean distance was found to be 0.005. In fact, this distance can be estimated analytically: it scales as  $1/\sqrt{N}$  for large  $N$  (Schroeder 2000). Thus, in a larger system of  $10^5$  instead of 1000 particles, one would obtain a mean distance  $D(p^{\text{max H}})$  equal to  $5 \cdot 10^{-4}$ . Considering that physical systems are typically composed of a number of particles  $N$  of the order of the number of Avogadro (i.e.  $N \approx 10^{23}$ ), one understands easily that the overwhelming majority of vectors  $p$  generated by randomly allocating particles to energy levels while satisfying the system constraints are extremely close to the maximum entropy vector  $p^{\text{max H}}$ .

The right panel of Fig. 4 shows a histogram comparable to those of Fig. 3. It was generated by first, simulating a single abundance vector  $p^{\text{sim}}$ , i.e. the first sampling procedure; next, uniformly sampling  $10^4$  vectors from the feasible set, i.e. the second sampling procedure; and finally, computing the distance (Eq. 10) of these sampled vectors to  $p^{\text{sim}}$ . The distance  $D(p^{\text{max H}})$  to the vector maximizing entropy is also shown. We see that the two procedures to select abundance vectors lead to completely different results. Whereas the vector  $p^{\text{max H}}$  is very close to the first set of simulated vectors (mean distance 0.005), it is rather remote from the second set of sampled vectors (mean distance 0.1). This demonstrates the crucial importance of using the appropriate method to weight particle configurations.

This result is in sharp contrast with the histogram shown in Fig. 3, and further illustrated in the last line of Table 2. Whereas previously the vector  $p^{\text{max H}}$  was hardly distinguishable from any other feasible vector, here it clearly stands out from all other estimates of the vector  $p^{\text{sim}}$ . Its superior predictive performance would only increase by considering larger systems. Thus, EM is not just an algorithm to select an arbitrary vector from the feasible set, where we could as well use the minimal concentration vector  $p^{\text{min C}}$ , or the central vector  $p^{\text{avg F}}$ . Quite the contrary, it stipulates to select a vector from the very small domain in the feasible set where almost all communities that are randomly generated following the appropriate procedure and taking the appropriate constraints into account, are massively accumulated.

## Discussion

Entropy maximization is an elegant mathematical technique, capable of reproducing a good deal of statistical mechanical theory. Application of this formalism to ecological problems is promising, as suggested by the fact that common SADs can be recovered rather easily (Banavar and Maritan 2007, Pueyo et al. 2007, Dewar and Porté 2008). In this paper, we investigated whether the EM method can be used to infer practical ecological information. In particular, we tried to predict the relative abundance vector from empirical community data. We

showed that successful application of EM requires that (1) the constraints contain all the significant characteristics of the community, but, on the other hand, (2) they be not so comprehensive that only the sought relative abundance vector is compatible with the data. We have shown that the results presented in Shipley et al. (2006) are entirely due to the latter scenario, and therefore do not provide any evidence for the usefulness of EM in ecology.

Marks and Muller-Landau (2007) and Roxburgh and Mokany (2007) already pointed to the circularity in Shipley et al.'s (2006) analysis. The fact that community-aggregated trait values are computed from the observed species abundances before entropy maximization is applied to estimate the very same species abundances, is indeed problematic. But any EM analysis is circular to some extent since the constraints should contain sufficient information about the unknown distribution for the EM inference to work. Therefore the fact that constraints and solution are not measured independently when testing EM predictions may not be a critical issue. The ecological application of EM by Shipley et al. (2006) faces a much more fundamental problem. The statistical mechanism that justifies EM is not at work in their data, as our comparison with the physical model system demonstrates. Note that this problem would remain even if community-aggregated traits could be measured directly.

Entropy maximization has a statistical basis as it selects the most probable feasible vector (here, the vector of species abundances). This vector yields precise predictions only if the distribution of feasible vectors is sufficiently concentrated around the most probable one. In turn, this concentration of feasible vectors requires that the system be composed of a sufficiently large number of components. This condition is clearly satisfied for the large-number systems typical of statistical mechanics. In ecology, however, one often has to deal with 'medium-number systems' (O'Neill et al. 1986). The usefulness of statistical techniques like EM is then a delicate issue, which depends on the scales involved in the problem. A clear separation is needed between the microscopic scale (the scale at which the system components are described) and the macroscopic scale (the scale at which the ecological question is formulated). Finding the appropriate scale of system description and problem formulation is an important challenge in ecological modelling.

To illustrate this, consider again the problem of determining species abundances. If one is looking for SADs of unspecified species, the microscopic structure is encoded in the a priori distribution, whereas the constraints fix the macroscopic scale. However, this scale separation is fuzzier than that encountered in statistical mechanics, leading to sometimes subtle arguments, as both scales influence the resulting SAD (Banavar and Maritan 2007, Pueyo et al. 2007, Dewar and Porté 2008). When predicting the relative abundances of particular species, we ask for a more detailed macroscopic description, and consequently we reduce the separation between microscopic and macroscopic scales even more. We think that in this case purely theoretical arguments are dubious. Only



experimental tests can then tell whether statistical techniques such as EM are useful. The merit of Shipley et al. (2006) is to have posed the question. Although we have shown that their conclusions are wrong, we agree with them that this question can only be settled through empirical evidence. Indeed, entropy maximization is not an algorithm that can be applied blindly, but it requires a careful analysis of how microscopic and macroscopic degrees of freedom are best taken into account.

As another illustration, we develop the example of metabolic reaction networks in Appendix 1. In this case, the problem of finding the reaction fluxes has formally the appropriate structure for applying the EM method: we are looking for the flux distribution in the reaction network given a number of stoichiometric constraints. The network fluxes are natural microscopic degrees of freedom, whereas stoichiometry imposes hard constraints (expressing mass conservation) on the macroscopic level, thus contrasting with the arbitrary trait constraints in Shipley et al. (2006). Nevertheless, application of EM does not yield the correct result, as experimental data show that growth under available resources, not entropy, is maximized at equilibrium (Ibarra et al. 2002). This negative result, however, points to the versatility of entropy maximization. Indeed, the fact that the bacterium was maximizing its growth constitutes genuine information that was not included in the statistical algorithm of entropy maximization. But this informative conclusion can only be drawn provided that the microscopic and macroscopic structure of the problem is appropriately taken into account. For example, if arbitrary trait constraints were used, one could not determine whether the failure of EM was due to a genuine mechanism, or whether additional constraints had to be considered.

By construction, entropy maximization is potentially useful in ecology because it merely formalizes the most probable behavior of the system constituents under a set of constraints. Its application, however, requires that both system constituents and macroscopic constraints be appropriately formulated. In this paper, we introduced a set of tools to detect overly restrictive constraints, and used these tools to analyze the EM prediction of species abundances in a community. But additional tools are needed to apply the EM algorithm to realistic ecological problems. For example, the plant communities we considered were sampled almost exhaustively. Trait measurements were also assumed to be error-free. Techniques have been developed in other EM applications to deal with sampling issues and noisy data (Jaynes 2003), and should be adapted to an ecological context. Our analysis was also restricted to a community at one location and one time, thus neglecting spatial and temporal correlations. Although linking community and average trait dynamics was one of the main motivations of Shipley et al. (2006), they did not really address this issue directly. A rigorous analysis of the application of entropy maximization to dynamically evolving communities would be particularly useful. Such an analysis could clarify the status of previous attempts to apply non-equilibrium versions of entropy maximization in ecology (Fath et al. 2001, Martyushev and Seleznev 2006), which currently lack a strong theoretical basis.

*Acknowledgements* – We thank Bill Shipley Brian McGill, Alain Rapaport and Dimitri Vanpeteghem for useful 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>>.
- Boyd, S. and Vandenberghe, L. 2004. Convex optimization. – Cambridge Univ. Press.
- Dewar, R. C. and Porté, A. 2008. Statistical mechanics unifies ecological patterns. – *J. Theor. Biol.* 251: 389–403.
- Fath, B. D. et al. 2001. Complementarity of ecological goal functions. – *J. Theor. Biol.* 208: 493–506.
- Garnier, E. et al. 2004. Plant functional markers capture ecosystem properties during secondary succession. – *Ecology* 85: 2630–2637.
- Ibarra, R. et al. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. – *Nature* 420: 186–189.
- Jaynes, E. T. 1957. Information theory and statistical mechanics. – *Phys. Rev.* 106: 620–630.
- Jaynes, E. T. 2003. Probability theory: the logic of science. – Cambridge Univ. Press.
- Lovász, L. 1998. Hit-and-run mixes fast. – *Math. Prog.* 86: 443–461.
- Magurran, A. E. 2004. Measuring biological diversity. – Blackwell.
- Marks, C. O. and Muller-Landau, H. C. 2007. Comment on “From plant traits to plant communities: a statistical mechanistic approach to biodiversity”. – *Science* 316: 1425c.
- Martyushev, L. M. and Seleznev, V. D. 2006. Maximum entropy production principle in physics, chemistry and biology. – *Phys. Rep.* 426: 1–45.
- McGill, B. J. et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. – *Ecol. Lett.* 10: 995–1015.
- O’Neill, R. V. et al. 1986. A hierarchical concept of ecosystems. – Princeton Univ. Press.
- Pfeiffer, T. et al. 1999. METATOOL: for studying metabolic networks. – *Bioinformatics* 15: 251–257.
- Pueyo, S. et al. 2007. The maximum entropy formalism and the idiosyncratic theory of biodiversity. – *Ecol. Lett.* 11: 1017–1028.
- Roxburgh, S. H. and Mokany, K. 2007. Comment on “From plant traits to plant communities: a statistical mechanistic approach to biodiversity”. – *Science* 316: 1425b.
- Schroeder, D. V. 2000. An introduction to thermal physics. – Addison Wesley Longman.
- Shipley, B. et al. 2006. From plant traits to plant communities: a statistical mechanistic approach to biodiversity. – *Science* 314: 812–814.
- Smith, R. L. 1984. Efficient Monte-Carlo procedures for generating points uniformly over bounded regions. – *Operations Res.* 32: 1296–1308.
- Vile, D. et al. 2006. A structural equation model to integrate changes in functional strategies during old-field succession. – *Ecology* 87: 504–517.

## Appendix 1. Some technicalities

### Convex analysis

We recall a number of definitions from convex analysis, see e.g. Boyd and Vandenberghe (2004). A set  $C$  in  $\mathbb{R}^d$  is convex if the line segment between any two points in  $C$  lies

again in  $C$ . For a convex set  $C$ , a point  $x \in C$  is called an extreme point if it is not an interior point of any line segment in  $C$ . Intuitively, an extreme point is a corner of  $C$ .

A polyhedron is defined as the set of solutions of a finite number of linear equalities and inequalities. A polyhedron is thus the intersection of a finite number of halfspaces and hyperplanes. Polyhedra are convex sets, and have a finite number of extreme points. A bounded and closed polyhedron  $C$  with extreme points  $v_k$ ,  $k = 1, \dots, M$ , can be represented as

$$C = \left\{ \sum_{k=1}^M \lambda_k v_k \mid \lambda_k \geq 0, \quad k = 1, \dots, M, \quad \sum_{k=1}^M \lambda_k = 1 \right\} \quad (11)$$

However, finding the extreme points  $v_k$  of a polyhedron  $C$  in a high-dimensional space  $\mathbb{R}^d$  can be computationally difficult, see below.

A function  $F: \mathbb{R}^d \rightarrow \mathbb{R}$  defined on a convex set  $C$  is convex if for all  $x, y \in C$ , and for all  $\lambda$  with  $0 \leq \lambda \leq 1$ , we have

$$F(\lambda x + (1 - \lambda)y) \leq \lambda F(x) + (1 - \lambda)F(y) \quad (12)$$

A function  $F$  is strictly convex if strict inequality holds, whenever  $x \neq y$  and  $0 < \lambda < 1$ . We say  $F$  is concave if  $-F$  is convex, and strictly concave if  $-F$  is strictly convex. For an affine function we always have equality in Eq. 12, so all affine (and therefore also linear) functions are both convex and concave.

Consider a bounded, closed, convex set  $C$  in  $\mathbb{R}^d$ , and an optimization problem over  $C$ , e.g. minimizing  $F(x)$  subject to  $x \in C$ . We say  $y$  is a (global) minimum if  $y \in C$  and

$$F(y) = \min\{F(x) \mid x \in C\}$$

We say  $y$  is a local minimum if  $y \in C$  and there exists an  $\epsilon > 0$  such that

$$F(y) = \min\{F(x) \mid x \in C, \|x - y\| \leq \epsilon\}$$

Global and local maximum are defined analogously.

When minimizing a convex function  $F$  over a bounded, closed, convex set  $C$ , any local minimum is also a global minimum. If the function  $F$  is strictly convex, this global minimum is unique. Similar statements hold for maximizing a concave function.

When minimizing a strictly concave function  $F$  over a bounded, closed, convex set  $C$ , any local minimum is an extreme point of  $C$ . The global minima can be found by comparing the value the function  $F$  takes in the extreme points of  $C$ . Similar statements hold for maximizing a convex function.

### Analogy with metabolic networks

Consider a set of  $S$  metabolic reactions using  $T$  reagentia, the metabolites. The concentration of metabolite  $j$  is denoted by  $X_j$ . The flux of reaction  $i$ , i.e. the number of times the reaction proceeds per unit of time, is denoted by  $p_i$ . The stoichiometric coefficient  $\tilde{t}_{ij}$  gives the quantity of metabolite  $j$  produced (if  $\tilde{t}_{ij} > 0$ ) or consumed (if  $\tilde{t}_{ij} < 0$ ) when reaction  $i$  happens once.

The dynamics of metabolite  $j$  are then given by

$$\frac{dX_j}{dt} = \sum_{i=1}^S p_i \tilde{t}_{ij}$$

The steady-state condition is given by

$$\sum_{i=1}^S p_i \tilde{t}_{ij} = 0$$

All the reactions are assumed to be irreversible, and therefore  $p_i \geq 0$ . Multiplying the reaction fluxes  $p_i$  by a factor corresponds to rescaling time. We can therefore, without loss of generality, impose that  $\sum_i p_i = 1$ . This leads to the set of conditions

$$\begin{aligned} \sum_{i=1}^S p_i \tilde{t}_{ij} &= 0 \quad j = 1, \dots, T \\ \sum_{i=1}^S p_i &= 1 \\ p_i &\geq 0 \quad i = 1, \dots, S \end{aligned} \quad (13)$$

They are equivalent with the conditions (Eq. 5), as can be seen by identifying  $\tilde{t}_{ij} = t_{ij} - \tilde{t}_j$ .

As a result, methods developed to analyze metabolic networks can be used to analyze ecological trait constraints. For example, an algorithm for the computation of the extreme points of the feasible set  $F$  is implemented in the program METATOOL (Pfeiffer et al. 1999). This allows us to construct the representation (Eq. 11) for the feasible set, which can be used to solve a number of optimization problems (minimizing a concave function, or maximizing a convex function).

Finding the function  $F$  to be optimized in order to reproduce observed reaction fluxes  $p_i^{\text{obs}}$  is a central research question in metabolic network analysis. For the reconstructed metabolic network of *Escherichia coli*, reaction flux vectors were computed that maximize growth subject to the availability of external metabolites (Ibarra et al. 2002). In an experiment the bacterium was seen to undergo adaptive evolution to achieve the predicted growth (Ibarra et al. 2002).

### Uniform sampling from convex set

Sampling from a convex set  $C$  in a high-dimensional space  $\mathbb{R}^d$  is a common computational task. We want to generate an (approximately) uniformly distributed random point. The generic method to do this is to define a random walk on  $C$  with a uniform stationary distribution, and to follow this random walk for a sufficiently large number of steps. The point obtained this way will be approximately stationary and thus approximately uniform.

We used the so-called hit-and-run random walk (Smith 1984), defined as follows. If the current point is  $x$ , we generate the next by selecting a random line through  $x$  (uniformly over all directions), and choosing the next point uniformly from the segment of the line in  $C$ . Compared to other random walks, this algorithm appears to need a small number of steps to reach stationarity (Lovász 1998).

Our implementation starts the hit-and-run random walk from the abundance vector  $p^{\text{max } H}$ , i.e. the feasible point with maximum entropy  $H$ . We take  $10^4$  steps to overcome

the transient, and another  $10^6$  steps in the (approximately) stationary distribution. Points are recorded every 100 steps, which yields  $10^4$  points from which the histograms in Fig. 3 were computed. We checked that the same distributions were obtained for other realizations, or by starting from other initial points.

### Sampling particle configurations

We describe a method to randomly distribute  $N$  particles over  $S$  energy levels (level  $i$  has energy  $\epsilon_i$ ) with given total energy  $E$ . We sample  $N$  times from the set  $\{1, 2, \dots, S\}$  with probability distribution

$$\frac{e^{-\beta\epsilon_1}}{\sum_i e^{-\beta\epsilon_i}}, \frac{e^{-\beta\epsilon_2}}{\sum_i e^{-\beta\epsilon_i}}, \dots, \frac{e^{-\beta\epsilon_S}}{\sum_i e^{-\beta\epsilon_i}}$$

Denote the results by the vector  $\mathbf{i} = (i_1, i_2, \dots, i_N)$ . The probability for a configuration  $\mathbf{i}$  is

$$\mathbb{P}(\mathbf{i}) = \frac{e^{-\beta E}}{\left(\sum_i e^{-\beta\epsilon_i}\right)^N}$$

with  $E(\mathbf{i})$  the total energy of the configuration  $\mathbf{i}$ ,

$$\mathbb{P}(E) = \sum_{\mathbf{i} | \sum_k \epsilon_{i_k} = E} \mathbb{P}(\mathbf{i}) = \#\left\{ \mathbf{i} \mid \sum_k \epsilon_{i_k} = E \right\} \frac{e^{-\beta E}}{\left(\sum_i e^{-\beta\epsilon_i}\right)^N}$$

The conditional probability on the total energy  $E$  is

$$\mathbb{P}(\mathbf{i}|E) = \frac{\mathbb{P}(\mathbf{i})}{\mathbb{P}(E)} = \frac{1}{\#\left\{ \mathbf{i} \mid \sum_k \epsilon_{i_k} = E \right\}}$$

and thus all vectors  $\mathbf{i}$  are equally weighted. Therefore, by keeping only those vectors  $\mathbf{i}$  that satisfy the energy constraint  $E(\mathbf{i}) = E$ , we obtain a sampling procedure to distribute particles with given total energy. Note that we are still free to choose the parameter  $\beta$ . To maximize our chances that the sampled vector  $\mathbf{i}$  will have the required energy  $E$ , we impose

$$\mathbb{E}(E(\mathbf{i})) = N \frac{\sum_i \epsilon_i e^{-\beta\epsilon_i}}{\sum_i e^{-\beta\epsilon_i}} = E$$

For the parameter values  $S = 30$ ,  $N = 1000$ ,  $E = 4000$  and  $\epsilon_i = i$ , we have  $\beta = 0.2872$ . Interestingly, this closely resembles the maximum entropy solution (Eq. 9). We stress, however, that this sampling algorithm is nothing but a mathematical trick, and does not depend in any way on the entropy maximization method.