Statistical Mechanics of Complex Systems
Emergent Patterns in Ecosystems

Alberto Chimenti - 1210360;   Anna Braghetto - 1205200;   Davide Maniscalco - 1212063

1   Dataset overview

The dataset reports a census (from 2005) of **S=299** alive plant species from Barro Colorado Forest in Panama.
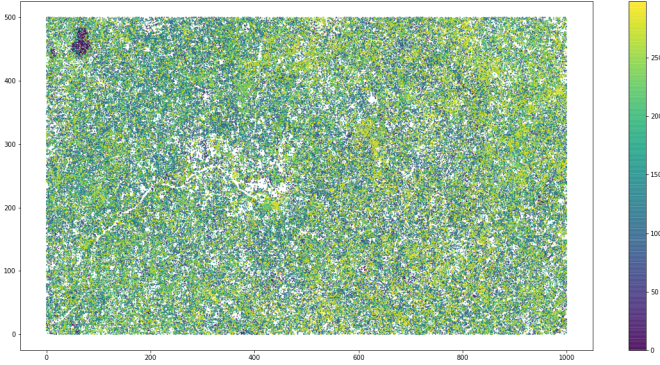

Figure 1: Dataset plot (the color grading is associated with species label)

The data was loaded in python using `pandas` library and only the alive species were selected among the full dataset. Information about species labels and their spatial location in the 50 hectare field were kept from the full dataset.

1.1   Sub-Sampling

The dataset was divided into **N=800** regions, each 0.0625 hectares big, using boolean masks (see `code`). Successively, considering the various subplots as independent from each other, we constructed the abundances vector **X** (of length $N*S$) by counting the number of individuals for each species in each subplot. Furthermore, all the "non-present" species samples (zero elements) were removed from our statistics.

1.2   Relative Species Abundance

The Relative Species Abundance (RSA) is the "abundance" of an organism of a particular kind relative to the total number of
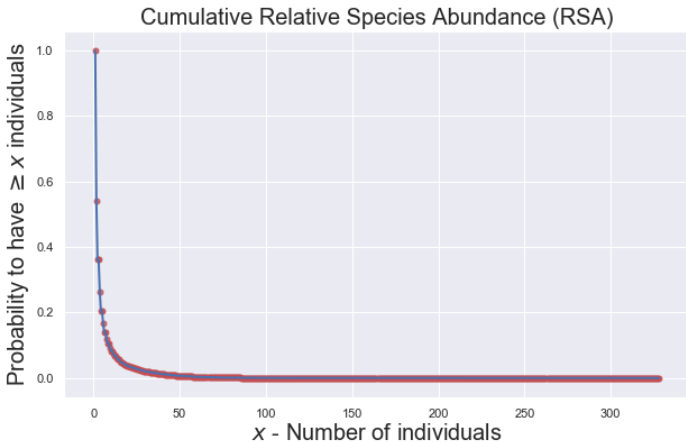

Figure 2: Empirical Cumulative Relative Species Abundance plot

organisms in the area. We calculated the Empirical Cumulative RSA distribution which tells us the probability that by picking a species at random among our samples it has abundance greater or equal than x. The latter was calculated as:

$$Cumulative\ RSA(x) = \frac{\#\mathbf{X}[n \geq x]}{\#\mathbf{X}[n]}$$

obtaining the result reported in *Figure* (2).

2   Birth and Death Master Equation

For testing purposes we tried to compare the data results by using a simple model looking at the system as composed by one unique species. In this framework we define the transition rates for *birth* and *death* processes as:

$$T(n \rightarrow n+1) = b_n = b*n\,, \qquad T(n \rightarrow n-1) = d_n = d*n\,;$$

$$b_0 = m$$

where **b**, **d** and **m** are the *birth probability*, the *death probability* and the *migration probability*, respectively.

One can find the detailed explanation of the model and further calculations in Appendix. Finally we obtain the following result for the stationary solution of the birth and death Master Equation:

$$P_n = -\frac{1}{n\,log\left(1 - \frac{b}{d}\right)} * \left(\frac{b}{d}\right)^n \qquad (1)$$

Since we are considering a *zero-sum game* model, for every death, one always has to replace the individual with either a migration or a birth. Therefore, we set:

$$b = 1 - m\,, \qquad d = 1$$

The latter can be interpreted as if we are considering death processes as the "time step" of the model evolution.

We then obtain the following equation:

$$P_n = -\frac{\left(1 - m\right)^n}{n\,log\left(m\right)} \qquad (2)$$

The migration probability has been experimentally measured and provided being equal to $m = 0.05$.
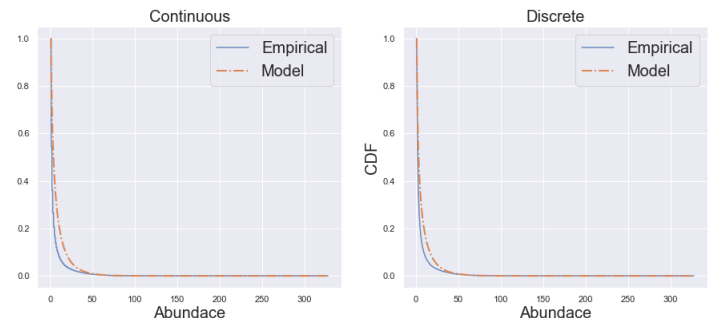

Figure 3: CDF confrontation

In order to confront the aforementioned result with our data we computed its Cumulative distribution. Since this is a confrontation simplified model, the CDF was computed considering the function as both discrete case and continuous approximation.

We obtained the results reported in *Figure* (3).

As we can see, the model doesn't fit properly the data in both cases.

It seemed reasonable to perform a curve fitting in order to see which parameter is giving the best fit to the empirical data.
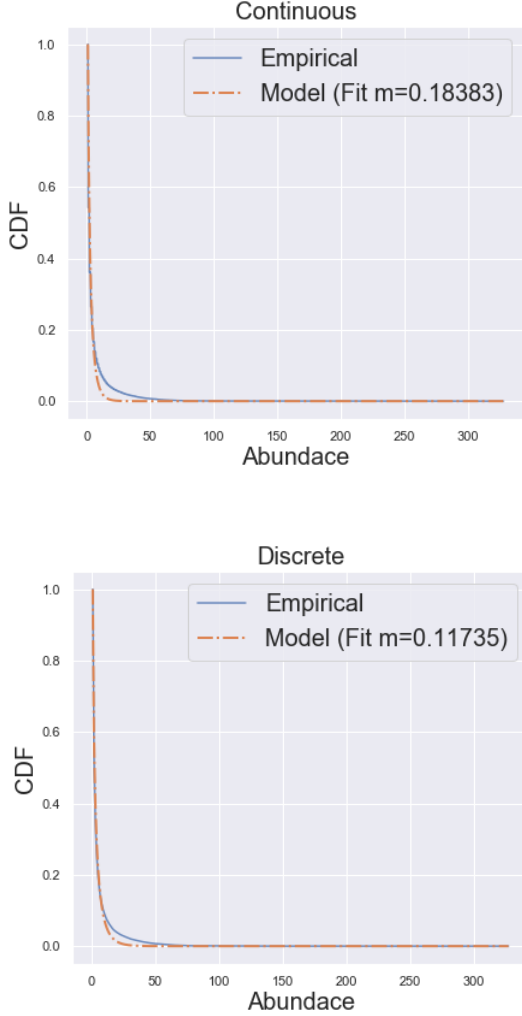

Figure 4: CDF Fit confrontation

By looking at *Figure* (4) we can tell that in both cases we are predicting the wrong parameter which suggests to better discard the model. The curves are still very different from the experimental one. Moreover the continuous approximation is acceptable as long as we have high abundances, this has probably to be considered the culprit of the bad fitting in the "low-abundance" region.

3 Species Area Relationship

The Species Area Relationship (SAR) gives the average number of present species as a function of the sampled area. Our dataset is made of 800 subplots of $a = 625m^2$ area each, for a total area $A = 5 \cdot 10^5 m^2$. In order to compute the SAR empirically, we counted the number of species present in a given number of subplots starting from one until reaching the full area. To obtain a smoother curve, the latter procedure has been performed 100 times, over randomly shuffled sample data in order to perform averages. The obtained curve is the blue dotted one shown in *Figure* (5) and (6) with semi-logaritmic scale.

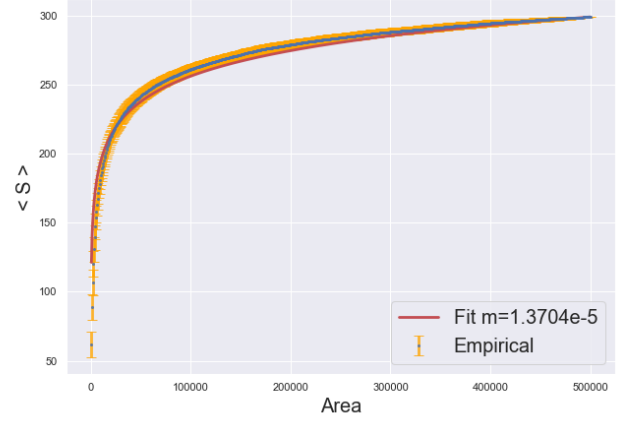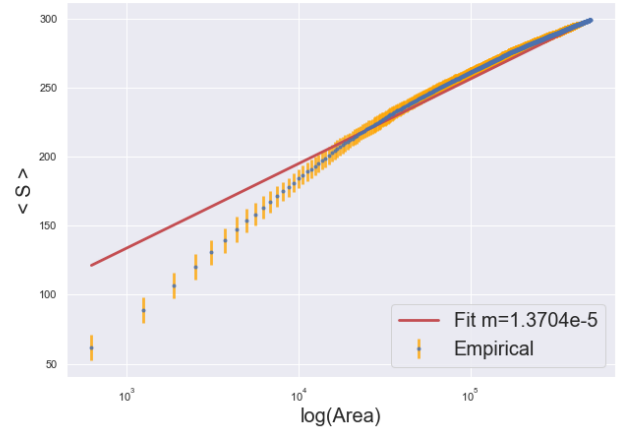Figure 5: Empirical and theoretical curve of the SAR

Figure 6: Empirical and theoretical curve of the SAR, logaritmic scale

As expected, the SAR grows quickly at the beginning slowing down in a logaritmic manner, and in the end reaches $S = 299$, the total number of species. To each point is associated an error calculated as the standard deviation over the shuffled resampling. By looking at the errors in *Figure* (7), one can see that the fluctuations over the number of species are bigger in the early stages and reduce to zero while reaching the full area (as expected).

The theoretical model used is represented in the following equation:

$$\langle S(ja)\rangle = s\left(1 - \frac{\log(\frac{ja}{A}(1-m)+m)}{\log(m)}\right) \tag{3}$$

where $j \in [1 : 800]$ is an index, S the average number of species and m is a parameter to be estimated. The estimation was performed through curve-fitting, and it has been found to be:

$$m \simeq 1.3704 \cdot 10^{-5}\,.$$

The obtained value is very small,and it is three orders of magnitude smaller than the empirical one $m = 0.05$. These considerations lead to conclude that the model is not describing the recorded data system.

The fitted curveis the red continuous one reported in *Figure* (5) and (6) with semi-logarithmic scale.
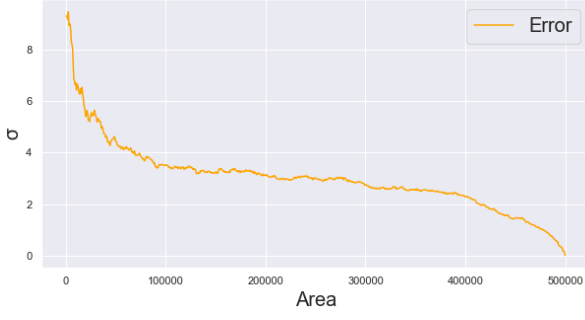
Figure 7: Error of the average number of species as a function of the area

4 Logistic Model

A model for the population dynamics is the *Logistic Model*, which is often used in ecology. In the model, the species abundance initially grows exponentially and then reaches a maximum, relative to the carrying capacity K of the system. In this task, 3 different reactions are considered:

- Birth: $A \rightarrow A + A$ with probability b, per unit of time and per particle.

- Migration: $0 \rightarrow A$ with probability m, per unit of time.

- Death: $A \rightarrow 0$ with probability $d = b\frac{A}{K}$, per unit of time and per particle.

In order to write the Birth and Death Master Equation, for a system with only the species A, the transitions rates are defined as follows.

$$W^+(A) = W(A \rightarrow A + 1) = m + bA$$

$$W^-(A) = W(A \rightarrow A - 1) = Ad = \frac{A^2}{K}b$$

where A is the abundance of the considered species. The rate $W^+(A)$ takes into account the probability of migration (m) and birth (b), which depends on the number of elements already present in the system (A). Instead, the rate $W^-(A)$ considers the probability that an element present in the system dies (d), which depends, as the birth event, on the abundance of the species (A).

Thus, the Birth and Death Master Equation can be written as

$$\frac{dP_A(t)}{dt} = W^+(A-1)P_{A-1}(t) + W^-(A+1)P_{A+1}(t)$$
$$- \left(W^+(A) + W^-(A)\right)P_A(t)$$

The solution at the stationary state is

$$P_A = \frac{W^+(A-1)}{W^-(A)} \frac{W^+(A-2)}{W^-(A-1)} \cdots \frac{W^+(0)}{W^-(1)}$$

and by replacing the transition rates found above, it becomes

$$P_A = \frac{m + (A-1)b}{\frac{A^2}{K}b} \cdot \frac{m + (A-2)b}{\frac{(A-1)^2}{K}b} \cdots \frac{m}{\frac{b}{K}}$$

$$\longrightarrow \quad P_A = P_0 \cdot \left(\frac{K}{b}\right)^A \frac{1}{A!^2} \prod_{i=1}^{A} [m + (A-i)b]$$

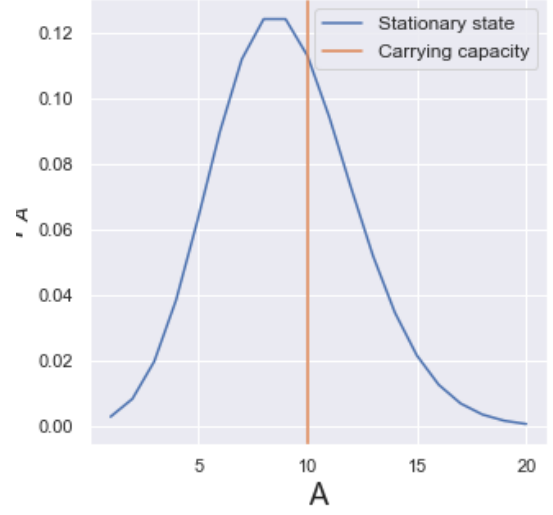where P_0 is the normalization constant which is computed by the following constraint.

Figure 8: Stationary states with m=0.1, b=1 and K=10.

$$\sum_{A=1}^{\infty} P_A = 1$$

The stationary state is then computed and shown for m=0.1, b=1 and K=10. Since the probability, for the parameters defined above, is greater than 0 just in a little interval of the domain (as shown in Figure 8), the constant P_0 is computed just for $A \in [1, 40]$.

As shown in Figure 8, the peak of the stationary state, represented by the blue curve, is relative to a value lower than the carrying capacity K=10, represented by the orange line. It was predictable since the stationary states represents the probability P_A that the system holds A elements and we expect that P_A is centered before the carrying capacity; then, once K is exceeded it decreases and goes to 0.

A Birth and Death Master Equation

In the following we report the calculations for the stationary solution reported in **Section 2**.

Once defined the transition rates as:

$$T(n \to n+1) = b_n = b * n, \qquad T(n \to n-1) = d_n = d * n;$$

with boundary conditions,

$$d_0 = 0 \quad b_{-1} = 0 \quad b_0 = m$$

We write down the Master equation in the form of:

$$\partial_t P_n(t) = b_{n-1}P_{n-1}(t) + d_{n+1}P_{n+1}(t) - (b_n + d_n)P_n(t) \tag{4}$$

By defining the *probability current*

$$J_n \doteq d_n P_n - b_{n-1}P_{n-1}$$

we can write the equation for stationary states as a *continuity equation* for the current:

$$J_{n+1} = J_n \tag{5}$$

Summing over n we find:

$$\sum_{x=0}^{n} \left[J_{x+1} - J_x \right] = 0 = J_{n+1}$$

Therefore using recursive equation (5):

$$P_n = \frac{b_{n-1}\ldots b_0}{d_n \ldots d_1}P_0 = P_0 \frac{b^{n-1}m\,(n-1)!}{d^n\,n!} \tag{6}$$

Since we are dealing with only present species, we normalize the solution for n which goes from 1 to $+\infty$.

Setting $r = \frac{b}{d}$

$$1 = \sum_{n=1}^{\infty} P_n = P_0 \frac{m}{b} \sum_{n=1}^{\infty} \int_0^r x^{n-1}dx = P_0 \frac{m}{b} \int_0^r \frac{1}{1-x}dx \quad \Rightarrow \quad P_0 = -\frac{b}{m}\frac{1}{\log(1-\frac{b}{d})}$$

We obtain finally equation (1):

$$P_n = -\frac{1}{n\,log\left(1 - \frac{b}{d}\right)} * \left(\frac{b}{d}\right)^n$$