Number Soup: A Study of Ecosystem Collapse

Hugo Harlin

I. INTRODUCTION

The inherent complexity of nature imposes strong restrictions in the scope of any model attempting to model ecosystems, and the trade-off between predictive capability and model complexity is not trivial. [1] [2] [4] [5] [6] [7] This dilemma leads one to the question: how simple can a model be, while still exhibiting the same behaviors as the incredibly complex ecosystems found on Earth? The model proposed in the paper *Insights into resource consumption, cross-feeding, system collapse, stability and biodiversity from an artificial ecosystem* by Yu Liu and David Sumpter is an attempt address that question.[13] Their '*Number Soup*' model is very simple yet displays many characteristics of complex ecosystems. This article is a continuation of their work and focuses on one phenomenon in particular, ecosystem collapse, and attempts to answer the following questions: Is collapse inevitable regardless of parameter settings? Is the survival time correlated to a known distribution?

II. MODEL

The '*number soup*' model by Liu and Sumpter consists of metabolites represented by integers, and species, represented by specific operators that consume and excrete metabolites.[13] The metabolites are integers labelled $\bar{1}, \bar{2}, \bar{3}, ..., \bar{n}$. Species are labeled s_{ij}, where the subscripts $\bar{i}$ and $\bar{j}$ represent the metabolites that the species s_{ij} requires to reproduce. In order to reproduce, an individual must have access to one unit of each of the metabolites it requires to reproduce, and when doing so it excretes one or two new metabolites. The excreted metabolite(s) are determined by modulo-(n+1) addition. For example, if n=3 there are six species in total that consume and produce metabolites according to the following equations:

$$\left.\begin{array}{ll} s_{11}: & \bar{1}+\bar{1}\rightarrow\bar{2} \\ s_{12}: & \bar{1}+\bar{2}\rightarrow\bar{3} \\ s_{13}: & \bar{1}+\bar{3}\rightarrow\bar{1} \\ s_{22}: & \bar{2}+\bar{2}\rightarrow\bar{1} \\ s_{23}: & \bar{2}+\bar{3}\rightarrow\bar{1}+\bar{1} \\ s_{33}: & \bar{3}+\bar{3}\rightarrow\bar{1}+\bar{2} \end{array}\right\} \quad (1)$$

In general, for the *n*-metabolite system, the equations governing resource consumption and excretion are:

$$s_{ij}: \bar{i}+\bar{j}\rightarrow\begin{cases}\overline{i+j}, & \text{if } i+j<n+1\\ \bar{1}, & \text{if } i+j=n+1\\ \bar{1}+\overline{(i+j)mod(n+1)}, & \text{if } i+j>n+1\end{cases} \quad (2)$$

Note that the number of metabolites excreted is always equal or less than the number of metabolites consumed, there is no $\bar{0}$ metabolite. This corresponds to energy leaving the system during reproduction. When an individual consumes metabolites, it creates a copy of itself and excretes metabolites. For example, the equation of reproduction for the species s_{13} with $n = 3$ is the following:

$$s_{13}+\bar{1}+\bar{3}\rightarrow 2s_{13}+\bar{1}$$

In general, the equations for reproduction are as follows:

$$s_{ij}: \bar{i}+\bar{j}\rightarrow\begin{cases}2s_{ij}+\overline{i+j}, & \text{if } i+j<n+1\\ 2s_{ij}+\bar{1}, & \text{if } i+j=n+1\\ 2s_{ij}+\bar{1}+\overline{(i+j)mod(n+1)}, & \text{if } i+j>n+1\end{cases} \quad (3)$$

Replicating the work of Liu and Sumpter [13], a single influx metabolite called $\bar{u}$ is added to the system, at a constant rate μ per unit of time.

Mutations are added to the model by changing the species of one of the individuals in the right hand side of equation 3, to another species with probability p. the Gillespie algorithm [8] is used to simulate the model in time:

i Set $t = 0$. Set $R_i(0) = 0$ and $S_{ij}(0) = 1$, this corresponds to zero metabolites and one individual of each species at t=0.

ii For each timestep of the model we generate a time interval δt until the next event from an exponential distribution with parameter $N(t)$ i.e.

$$P(\tau\leq\Delta t\leq\tau+d\tau)=N(t)e^{-N(t)\tau}d\tau$$

iii The number of metabolites in the system is increased due to inflow according to

$$R_u(t)=R_u(t)+\mu\Delta t$$

Note that the number of metabolites in the system are updated at the current time rather than at the next time-step, this allows for metabolites to be generated and immediately consumed at the same time-step.

iv One individual organism is selected from the entire population, so that the probability of selecting species S_{ij} is proportional to $S_{ij}/N(t)$.

v The selected organism either reproduces with probability $q(R_t(t), R_j(t))$ or dies with probability $1 - q(R_t(t), R_j(t))$. This probability is a function of the number of available metabolites, as follows:

$$q(R_t(t),R_j(t))=\begin{cases}\frac{R_i(t)}{a+R_i(t)}+\frac{R_j(t)}{a+R_j(t)}, & \text{if } i\neq j\\ \frac{R_i(t)}{a+R_i(t)}+\frac{R_i(t)-1}{a+R_i(t)-1}, & \text{if } i=j\end{cases} \quad (4)$$

Where a is a constant measuring the difficulty of obtaining metabolites. Here the Monod equation is used to describe the probability of the uptake, which is commonly used to describe the growth of microorganisms. An exception to this rule occurs when either $R_i(t) < 1$ or $R_j(t) < t$ (if $i \neq j$), or $R_i(t) < 2$ (if$i = j$). In these cases the organism dies as there are not metabolites present in sufficient amounts. If the organism dies, set

$$S_{ij}(t + \Delta t) = S_{ij}(t) - 1$$

If the selected organism reproduces, update the metabolites and the organisms according to (3).

vi Update the time, $t = t + \delta t$.

vii iterate from step (ii).

The updating of time and selection of individual in the Gillespie algorithm results in all organisms making one action each per unit time, on average. As in the work by Liu and Sumpter[13], we call one unit type cycle a generation.

III. Results

Liu and Sumpter[13] note that their '*number soup*' model tends to collapse regardless of initial values and parameter settings. To explore this phenomenon, the model was run 1000 times with varying parameter settings and the time of collapse was recorded.

Due to computational limitations, an upper simulation limit of 10000 generations was set on each run. For all simulations the following settings were used: $p = 0.001$, $\mu = 500$, and $a = 5$. The model was simulated with n=2 and a=1,2, n=3 and a=1,2,3, n=4 and a=1,2,3,4, and n=5 and u=1,2,3,4,5 where n is the number of metabolites and u is the influx metabolite. The ecosystems didn't all die out before the 10000 generation limit, and the portion of collapse for the runs can be seen in the table I.

n	u	collapsed
2	1	8.3%
2	2	100%
3	1	6.2%
3	2	23.2%
3	3	100%
4	1	8.0%
4	2	100%
4	3	24.9%
4	4	100%
5	1	14.7%
5	2	97.6%
5	3	82.2%
5	4	41.9%
5	5	100%

TABLE I: Percentage of collapsed simulations over 1000 runs.

A. Number of Metabolites n = 2

1) Influx metabolite u = 1: For n=2 and u=1 almost all of the simulations survived the 10000 generation limit, and the 8.3% that did collapse did so before reaching six generations in age. It seems that if the populations survive the earliest stages of existence and grow sufficiently large to not be vulnerable to single individuals randomly dying instead of reproducing (step v of the model algorithm), then they will survive at least 10000 generations. Individual runs with particular seeds that survived the 10000 generation limit where run with an upper limit of 50 million generations without collapsing, indicating that is potential for extremely stable ecosystems for n=2 and u=1.

2) Influx metabolite u = 2: For n=1, u=2 all 1000 simulations collapsed. Similarly to the u=1 case, some simulations collapsed very early and there is a clear divide between early collapses and simulations that survived longer than 150 generations. Figure 1 shows an enlarged section of the histogram for n=2, u=2, with the enlarged are marked in red. The separation between collapses due to early random deaths and collapses after the populations have grown large is evident, and are treated as two unrelated causes of collapse.

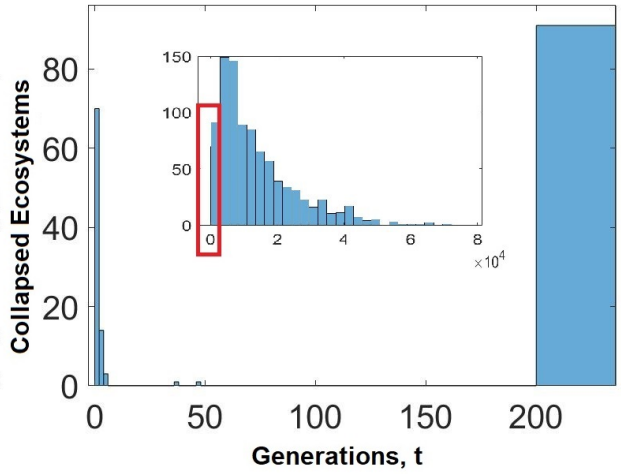

Fig. 1: Enlarged section of histogram for n=2, u=2, marked in red in figure. The clear separation in collapses for low values of t indicates that there are two separate mechanisms of collapse at play.

After filtering out early collapses due to initial random deaths, the weibull [9], log-logistic [10], gamma [11], and Burr [12] distributions where fitted to the data using the Distribution Fitter toolbox in Matlab. The Burr distribution was chosen as the best fitting distribution by visual inspection and comparison of the variances. The probability density function of the Burr distribution:

$$f(x; c, k, \alpha) = \frac{ck}{\alpha}\left(\frac{x}{\alpha}\right)^{c-1}\left[1 + \left(\frac{x}{\alpha}\right)^{c}\right]^{-k-1} \quad (5)$$

Where c and k are the first and second scale factors, and alpha is the scale factor.

With n=2 and u=2 all ecosystems collapsed within the 10000 generation time limit, and the distribution of the collapse times formed a skewed bell cure to which the Burr[12] was a good fit. See figure 2.

B. Number of Metabolites n = 3

1) Influx metabolite u = 1: For n=3 and u=1, only 63 in 1000 simulations collapsed before the 10000 generation limit, all within 45 generations. This behavior is similar to the early collapses observed for n=2 and u=1.

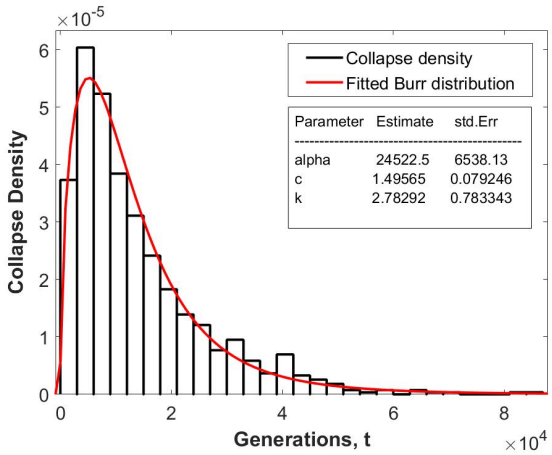

Fig. 2: System collapse density histogram for n=2, u=2. The Burr probability density function is fitted to the data in red (equation 5), where alpha is the scale factor and c,k are the first and second shape parameters respectively.

2) Influx metabolite u = 2: Filtering out early collapses, and plotting the collapses in a histogram yields figure 3. Due to the strange shape of the histogram for 1000 simulations, another 5000 simulations where run. The figure below is based on both data sets, a total of 6000 simulations.

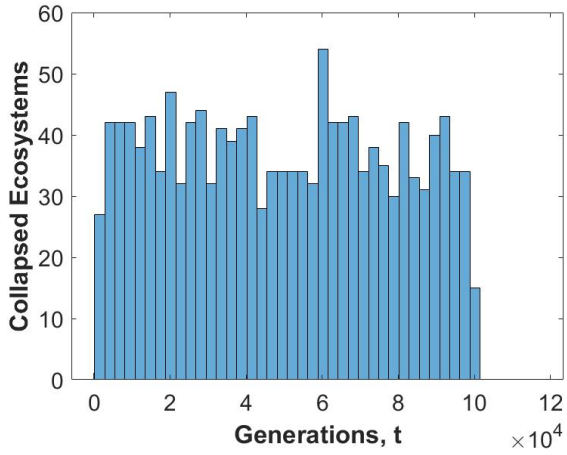

Fig. 3: System collapse density histogram for n=3, u=2. The Burr probability density function is fitted to the data in red (equation 5), where alpha is the scale factor and c,k are the first and second shape parameters respectively.

3) Influx metabolite u = 3: For n=3, u=3 the system collapses behaved similarly to the case n=2, u=2 and the Burr distribution was fitted to the data, see figure 4.

C. Number of Metabolites n = 4

1) Influx metabolite u = 1: For n=4 and u=1, the 8.3% of the simulations that collapsed did so in 130 generations or less, with a vast majority collapsing in 15 generations or less.

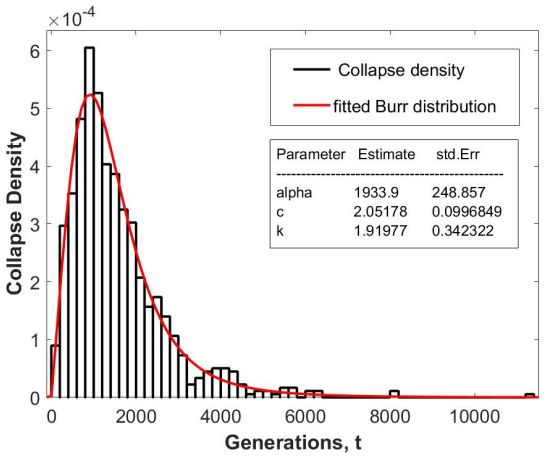

Fig. 4: System collapse density histogram for n=3, u=3. The Burr probability density function is fitted to the data in red (equation 5), where alpha is the scale factor and c,k are the first and second shape parameters respectively.

2) Influx metabolite u = 2: Filtering out the early collapses, and plotting the resulting data in a histogram can be seen in figure 5 below. A Burr distribution (equation 5) has been fitted in red.

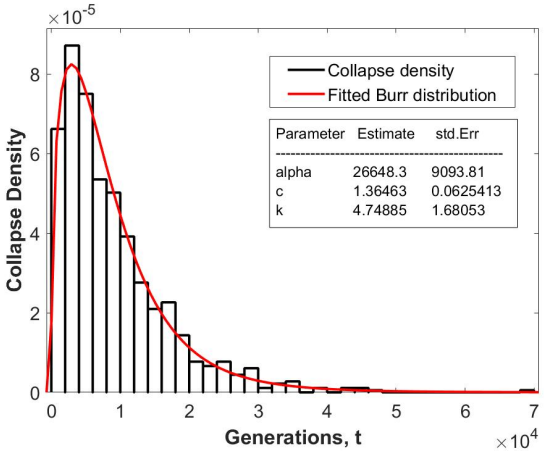

Fig. 5: System collapse density histogram for n=4, u=2. The Burr probability density function is fitted to the data in red (equation 5), where alpha is the scale factor and c,k are the first and second shape parameters respectively.

3) Influx metabolite u = 3: For n=4 and u=3, the 24.9% of the simulations that collapsed show a similar behavior to the case n=3, u=2, see figure 6.

4) Influx metabolite u = 4: Filtering out the early collapses, and plotting the resulting data in a histogram can be seen in figure 7 below. A Burr distribution (equation 5) has been fitted in red.

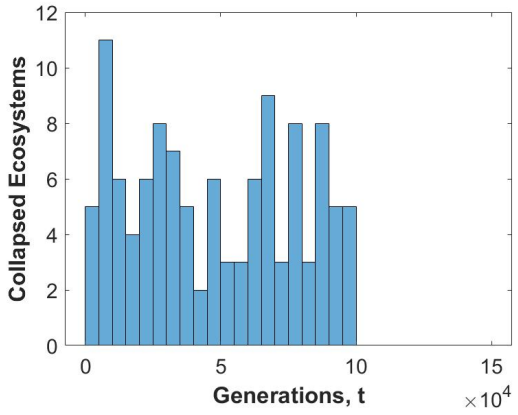

Fig. 6: System collapse histogram for n=4, u=3. Note the similarity with the case n=3, u=2, figure 3.

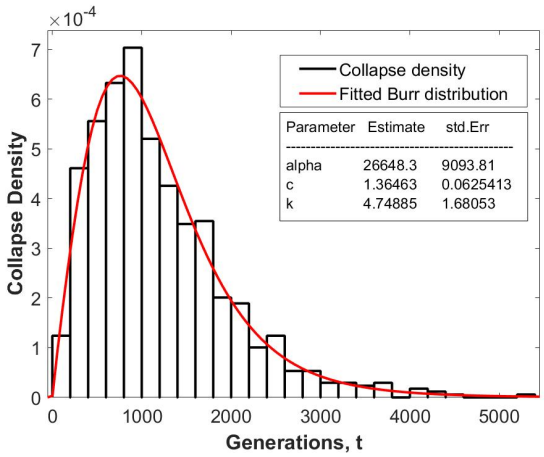

Fig. 7: System collapse density histogram for n=4, u=4. The Burr probability density function is fitted to the data in red (equation 5), where alpha is the scale factor and c,k are the first and second shape parameters respectively.

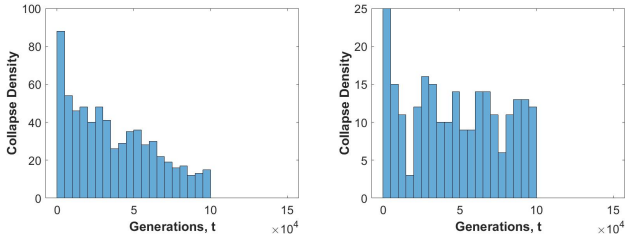

Fig. 8: Histograms of the survival times of 1000 simulations for n = 5 and u = 3 (a), and u = 4 (b).

D. Number of Metabolites n = 5

1) Influx Metabolite u = 1: For n=5 and u=1, the portion of collapsed simulations was low with only 14.7% not surviving the first 10000 generations. Similarly to the other simulations with u=1, all the collapsed simulations did so within 200 generations.

2) Influx Metabolite u = 2: For n=5 and u=2, 97.6% of the simulations collapsed before the 100000 generation time limit, and the data did not fit a Burr distribution but seemed to decay exponentially instead:

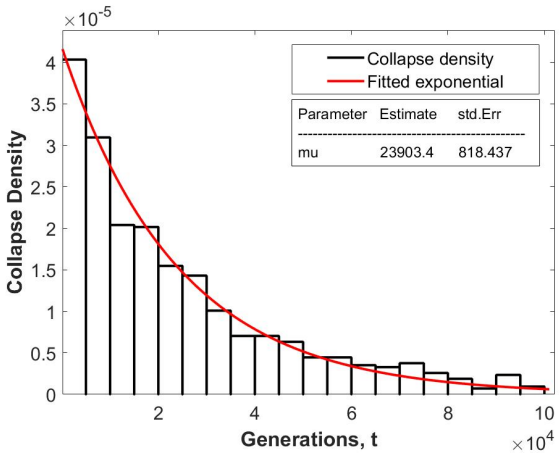

Fig. 9: System collapse density histogram for n=5, u=2. The exponential distribution is fitted to the data in red (equation 6).

$$f(x;\mu) = \frac{1}{\mu} e^{\frac{-x}{\mu}} \tag{6}$$

In figure 9 the histogram of the filtered collapse data is plotted together with a fitted exponential distribution.

3) Influx Metabolite u = 3,4: For u=3 and u=4, it is unclear what underlying distribution is governing the collapses. The histogram of the collapses can be seen in figure 8. Note that twice as many simulations collapsed for u=3 (82.2%), compared to u=4 (41.9%).

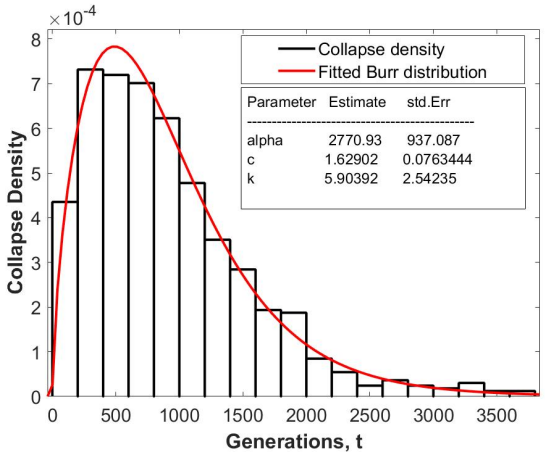

Fig. 10: System collapse density histogram for n=5, u=5. The Burr probability density function[12] is fitted to the data in red (equation 5), where alpha is the scale factor and c, k are the first and second shape parameters respectively.

4) Influx Metabolite u = 5: For n=5 and u=5, all simulations collapsed before the 10000 generation limit, and as in many of the previous cases, the Burr distribution[12] fitted the data well.

IV. Discussion and Future Work

A behavior shared across all the studied values of n and u was the very early collapse of a small portion of the simulations. Upon inspection, these collapses were distributed independently of the majority of the later collapses. In figure 1 this division of collapses can be seen for n=2 and u=2. This is likely due to the possibility of individuals dying randomly (step v of the model algorithm), coupled with the relatively small population sizes early in the simulations (ten of each species at t=0).

A shared behavior across all values of n was the very high survival rate when u=1. This is to be expected because the '1'-metabolite is the most commonly created metabolite due to the modular dynamic of the resource consumption and creation, which together with u=1 ensures a high total influx.

When a large majority of the simulations collapsed, the distribution of the collapse likened a skewed normal distribution, except for n=5, u=2,3 where an exponential distribution made a better fit. There are several distributions which can model a skewed bell curve, and four common alternatives where compared in this paper: the weibull [9], the log-logistic [10], gamma [11], and Burr [12] distributions, of which the Burr distribution was found to be the best fit.

n	u	distribution
2	1	-
2	2	Burr
3	1	-
3	2	unknown
3	3	Burr
4	1	-
4	2	Burr
4	3	unknown
4	4	Burr
5	1	-
5	2	exponential
5	3	unknown
5	4	unknown
5	5	Burr

TABLE II: Distributions of survival densities for all simulations. Note that early collapses due to random deaths (part v of the algorithm) was found for all parameter settings and is excluded here. A dash "-" indicates that no collapses were observed in the filtered data.

The next step in exploring ecosystem collapse is to increase the number of simulations and parameter settings substantially. This can only be done by re-implementing the model in an efficient language such as c or fortran, in such a way that it can be run on a gpu, preferably a gpu cluster. The simulations done in this paper where run on a desktop with an Intel(R) Core™ I7-4510U CPU @ 2.6GHz with four cores, and the total simulation time was in excess of two weeks. If one makes the assumption that the simulation time of one generation is constant on average, then an increase in the maximum generation limit with a factor of ten would require roughly half a year of computation time. It is evident that the speedup enjoyed by running the simulations in parallel on a gpu instead of a cpu is a complete necessity to realize further exploration.

References

[1] Friedrichs, M.A.M., Hood, R., Wiggert, J., 2006. "Ecosystem model complexity versus physical forcing: quantification of their relative impact with assimilated Arabian Sea data". Deep-Sea Research II 53, 576–600.

[2] Pauly, D. 2000. "Ecopath, Ecosim, and Ecospace as tools for evaluating ecosystem impact of fisheries". ICES Journal of Marine Science. 57 (3): 697–706.

[3] Fasham, M. J. R.; Ducklow, H. W.; McKelvie, S. M. 1990. "A nitrogen-based model of plankton dynamics in the oceanic mixed layer". Journal of Marine Research. 48 (3): 591–639.

[4] Vasconcelos FR, Diehl S, Rodríguez P, Hedström P, Karlsson J, Byström P. 2016. "Asymmetrical competition between aquatic primary producers in a warmer and browner world." Ecology 97(10): 2580-2592.

[5] Wollrab, S., Diehl, S. & De Roos, A.M. 2012. "Simple rules describe bottom-up and top-down control in food webs with alternative energy pathways." Ecology Letters, 15, 935–946.

[6] N. Mouquet et al. 2015. Predictive ecology in a changing world. J. Appl. Ecol. 52, 1293–1310.

[7] Urban, M. C. et al. 2016. "Improving the forecast for biodiversity under climate change." – Science 353: aad8466.

[8] Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2340–2361.

[9] Kızılersü et al. 2018. "The Weibull distribution". Royal Statistical Society, Volume 15, Issue 2.

[10] Bennett, 1983. "Log-Logistic Regression Models for Survival Data". Journal of the Royal Statistical Society. Series C (Applied Statistics) Vol. 32, No. 2, 165-171.

[11] Thom et al. 1958. "A Note on the Gamma Distribution". Monthly Weather Review, Volume 86, Number 14.

[12] Tadikamalla, P.R. 1980. "A Look at the Burr and Related Distributions," International Statistical Review, 48, 337-349.

[13] Liu Y, Sumpter D. 2017. "Insights into resource consumption, cross-feeding, system collapse, stability and biodiversity from an artificial ecosystem." Journal of The Royal Society Interface. 14(126):20160816.