

Letter to Editor

Investigating small fish schools: Selection of school—formation models by means of general linear models and numerical simulations**1. Introduction**

The aggregation of organisms is a ubiquitous biological phenomenon that occurs in a wide range of species, from bacteria to large vertebrates, and across a range of temporal stability, from ephemeral reproductive assemblages to obligatory schools or herds (Allee, 1933). In general, the processes that lead to the aggregation of individuals are driven by a complex interplay between habitat characteristics and individual behaviour. To illustrate the importance of individual behaviour, we note that a number of studies have suggested that important features of aggregations, such as size, shape, and architecture, emerge as a consequence of simple behavioural rules that govern the interactions among individuals (Parrish and Edelstein-Keshet, 1999; Niwa, 2005). Understanding aggregation therefore always entails identifying the behavioural rules that generate the observed patterns in animal aggregations.

Aggregation size, usually defined as the number of individuals that are part of an aggregation, has received considerable attention, especially in the characterization and prediction of the observed variation in school size in fish species (Bonabeau et al., 1999; Bonabeau and Dagorn, 1995; Niwa, 1996, 1998, 2003, 2004). Most of these studies have noted that the probability of finding a large school decays as flat-tailed distributions such as power-laws and power-laws with an exponential truncation for large school size (Bonabeau and Dagorn, 1995), although exponential distributions have also been recorded (see Niwa, 2003). Flat-tailed distributions are relevant because they characterize statistical properties associated with different ecological systems such as mutualistic networks (Jordano et al., 2003) and ecological communities (Hubbell, 2001).

The characterization of school-size distributions as power-laws, truncated power-laws or exponentials often focuses on fish species such as the tuna (e.g. Niwa, 2003), for which schools are very large (of the order of several hundred individuals), thereby allowing a precise statistical characterization of school-size distribution. However, many fish species do not form large schools. On the contrary, some fishes aggregate forming small schools, in

which the number of individuals is seldom larger than a few dozens. In such cases, the characterization of school-size distribution lacks statistical confidence and, as a consequence, the study of the organizing principles of school formation has been limited to fishes that aggregate in large schools (e.g. Bonabeau and Dagorn, 1995; Niwa, 2003; Powers, 2004). Statistical procedures that allow the study of fish schools with few individuals will allow the investigation of the organization of schools of many fish species and will be also important in a broader sense. As recently noted by Amaral et al. (2004) in the context of web-like systems, the characterization of the structure of natural systems containing small number of elements is one of the main challenges in the study of complex systems (see also Guimarães et al., 2005).

Here we explore a possible solution for the characterization of low dimensional systems using as reference system the Southwest Atlantic parrotfish *Sparisoma axillare* (Perciformes: Scaridae), which forms small schools at feeding sites. We combine general linear models and simulations to describe the functional form of school-size distribution and select between two recently proposed school-formation models (Powers, 2004). We demonstrated that the proposed characterization allows the selection among candidate models even for species in which school sizes are often small.

2. Material and methods*2.1. Reference system*

The study site consisted of sandy areas adjacent to the rocky shore of the Praia da Conceição, at Fernando de Noronha archipelago (03°50'S, 32°25'W), about 345 km off northeast Brazil (see Carleton and Olson, 1999 for description of the archipelago), from May to June 2003. During the study, horizontal visibility was 20 m, which allowed adequate observation of fish schools. The water temperature was 27°C and the depth of the study site ranged 2–7 m. The observations were done by diving during daylight hours, from 0900 to 1700 h in five non-consecutive days while snorkelling. No nocturnal observations were made since parrotfishes are inactive at night (Starck and Davis, 1966).

Parrotfishes are bottom-foragers, feeding mostly on algae and dead coral (Bruggemann et al., 1994a, b). The parrotfish *S. axillare* is endemic to the tropical and

subtropical Southwestern Atlantic (Moura et al., 2001). Initial phase individuals of *S. axillare* forage alone or in schools (Bonaldo et al., 2006). At the studied site, the initial phase individuals foraged along the rocky shore and on sand flats. In sand flats, *S. axillare* gather to forage on the sparse rocks and form near motionless schools (Bonaldo and Krajewski, unpublished results). We assessed the parrotfishes distribution at Praia da Conceição with 16 transect lines (5 m × 100 m) parallel to the rocky shore, where we registered all the numbers of initial phase individual *S. axillare* in each group. The good observation conditions and the spatial homogeneity of sand flats suggest that our sample is representative. During our fieldwork, the schools observed were often small, the largest having up to 20 fish.

2.2. Characterizing the school-size distribution

The number of schools recorded for each school size formed the school-size distribution. In these analyses, solitary fish were recorded as 1-fish schools. We used the formalism of general linear models (McCulloch and Searle, 2000) to assess whether the observed school-size distribution of *S. axillare* could be characterized as an exponential or a power-law. Initially, we recall that both exponentials, $y \propto e^{\beta_1 k}$, and power-laws, $y \propto k^{\beta_2}$, can be written as linear models, respectively, as $\log(y) = \alpha_1 + \beta_1 k + \varepsilon$ and $\log(y) = \alpha_2 + \beta_2 \log k + \varepsilon$, where α_1 and α_2 are the fitted normalization constants, β_1 , and β_2 are the fitted slope constants, k is the school-size, y is the number of recorded schools with a given school-size (k), and ε is the error. If the observed school-size distribution is described by an exponential, we would expect $\beta_1 k$ to be significant, whereas if a power-law adequately describes the data set, we would expect $\beta_2 \log k$ to be significant. If both $\beta_1 k$ and $\beta_2 \log k$ are statistically significant then school-size distribution can be described by either an exponential or a power-law. If neither term is significant, the observed distribution cannot be described by either an exponential or a power-law.

2.3. School-formation models

Recently, different types of splitting/coalescing models were proposed to explain school-size distributions (Bonabeau et al., 1999; Bonabeau and Dagorn, 1995; Niwa 1996, 1998, 2003). However, the assumptions of splitting and coalescing do not adequately describe the natural history of *S. axillare* because these fishes aggregate at feeding sites where they form small groups that may remain nearly motionless and seldom coalesce with other schools or split into new schools (Bonaldo, unpublished results). Single individuals may join schools, suggesting that schools may increase in size during the feeding period (morning to afternoon). In fact, larger schools were recorded at afternoon (Bonaldo et al., 2006). The two models used here are adapted versions of the recently proposed models inspired in complex networks (Powers, 2004). The two

models used here assume that: (1) at $t = 0$ there was a single 1-fish school in the sand flats; (2) at each time interval, a fish moved from the rocky shore to the sand flats; (3) this new individual in the sand flats had a constant probability, $1 - p$, of joining a school already present, or remained alone, creating a new 1-fish school with probability p ; (4) if an individual joined a school already in the area, it could choose among any of the fish schools present—this assumption is supported by the spatial homogeneity of sand flats; (5) individuals did not leave the schools until the end of the simulation, and (6) the simulation ends when the number of fishes in the sand flats is equal to the observed number of fishes in the sand flats. The only difference between the two models was how a fish chooses among the schools already present: the first model, referred to as the random (R) model, is based on the assumption that, if a fish joined a previously established school, the probability of it choosing the i th school is

$$\Pi = \frac{1}{N}. \quad (1)$$

In this equation, N is the number of previously established schools and Π is independent of school size. In contrast, the second model, referred to as the preferential attachment (PA) model, is based on the assumption that, if the fish joined a school, the probability of its choosing among one of the already existing schools is not uniform. Rather, fish have a tendency to preferentially join larger schools. The tendency of large schools to be more attractive to solitary individuals can be modelled using a concept derived from network theory, the PA rule (Barabási and Albert, 1999), in which the probability Π that a fish joins school i depends on school size, k_i , in a linear functional form

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \quad (2)$$

In this equation, $\sum_j k_j$ is the sum of the sizes of all schools already present.

The few differences between our models and those of Powers (2004) are related to particulars of the natural history of *S. axillare*. These differences include the assumptions that: (i) there is no recruitment or death during simulation. This assumption is reasonable because schools of *S. axillare* are temporary since they form at feeding sites (rocks on sand flats) and (ii) only 1-fish schools aggregate to previous schools. This assumption is based on naturalistic observations that indicate that schools seldom aggregate, but solitary fish usually join schools and remain in the school while feeding (Bonaldo, unpublished results). It is also important to notice that the PA model is similar to the well-known Simon model of growing aggregations, which has been used to study abiotic, biotic and social systems (see Albert and Barabási, 2002 for further discussion).

2.4. School-formation model selection and numerical simulations

The R and PA models differ in their asymptotic behaviours: the R model generates exponential distributions, in which all schools have similar sizes, whereas the PA model generates power-law distributions, in which a great proportion of individuals are part of a few large schools and the majority of schools are very small (Barabási and Albert, 1999; Powers, 2004). In contrast, in systems with a small number of component elements, such as that analysed here, characterization of the distribution may not be so easy (Guimarães et al., 2005). Indeed, we found that the school-size distributions generated by both models at small-size situations can be described by power-laws (results not shown). To circumvent this problem, we used the estimated slope constant, β_i , of the significant general linear model rather than the qualitative (“exponential” or “power-law”) description of the functional form, to select between the two models of fish aggregation. The ability of the models to predict the observed value of the slope constant was examined by using simulations to

generate the empirical distribution of β_i for each model (4000 simulations). We considered that the model reproduced the field data if the observed value lies within the 95% confidence interval of the empirical distribution.

3. Results and discussion

Neotropical parrotfish *S. axillare* at Fernando de Noronha archipelago often form small schools (often < 20 individuals). We sampled 65 schools of *S. axillare*, which corresponded to a total of 124 individuals ($\bar{X} \pm SD$, $\bar{k} = 1.9 \pm 3.0$ individuals per school). The value of \hat{p} was directly estimated from field data, simply by dividing the number of schools recorded minus one ($N - 1 = 64$), by the number of fish recorded minus one ($n - 1 = 123$), which yielded $\hat{p} = 0.52$. To remove first order bias of \hat{p} we used jackknife procedures (Davison and Hinkley, 1997), and simulations are performed using the corrected value $\hat{p}^* = 0.50$. The exponential function did not adequately describe the observed school-size distribution of *S. axillare* ($d.f. = 6$, $R^2 = 0.47$, $F = 4.46$, $P = 0.1$). In contrast, the power law distribution generated a good description of

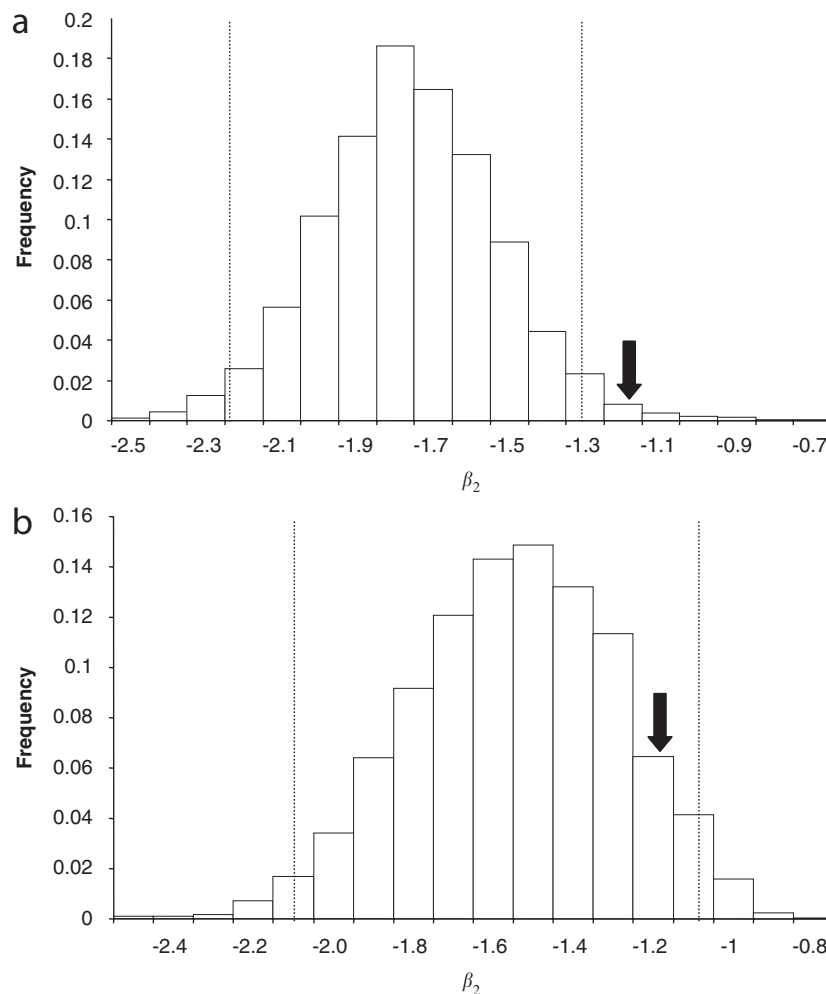


Fig. 1. The empirical distribution of β_2 generated by: (a) the R model and (b) the PA model. The dashed lines indicate the lower and upper limits of the 95% confidence interval. The black arrow indicates the observed value for *Sparisoma axillare*.

the school-size distribution of this species. The estimated value of the slope constant was $\beta_2 = -1.21$ and the linear model $\log y = 2.99 - 1.21 \log k$ was highly significant ($d.f. = 6$, $R^2 = 0.80$, $F = 20.25$, $P < 0.01$). Therefore, we used β_2 in the procedure of model selection.

Although the schools are small, the organization of these schools is not random. The average β_2 predicted by the R model was lower than the observed $\beta_2 = -1.21$ for *S. axillare* (Fig. 1a). In contrast, the empirical distribution generated by the PA model reproduced the observed β_2 for *S. axillare* (Fig. 1b). Therefore, our analysis suggests that the model that assumes that fishes preferentially join larger schools reproduces the features of the school-size distribution.

We also note that the exponent generated by the numerical simulations of the PA model is very different from the analytic prediction of the original network model (Barabási and Albert, 1999). In complex networks, PA predicts that the degree (k) of the node i varies with time following the dynamical equation:

$$\frac{\partial k_i}{\partial t} = m \frac{k_i}{\sum_{j=1}^{N-1} k_j} = m \frac{k_i}{2mt} = \frac{k_i}{2t}. \quad (3)$$

The solution of the above equation is

$$k_i = m \left(\frac{t}{t_i} \right)^a \quad (4)$$

in which $a = \frac{1}{2}$. The predicted exponent for the probability of a node showing k links derived from (4) is $\beta_2 = -1/a - 1 = -3$ (Albert and Barabási, 2002). In the PA model for fish schools, the probability that the size of the school i increases with time is slightly different from (3), depending on p

$$\frac{\partial k_i}{\partial t} = (1-p) \frac{k_i}{\sum_{j=1}^{N-1} k_j} = (1-p) \frac{k_i}{t}. \quad (5)$$

Therefore, $a = (1-p)$ and the predicted exponent is $\beta_2 = [-1/(1-p)] - 1$. In the case of the studied parrotfishes, $p = 0.5$ and, consequently, the above analytic approach predicts $\beta_2 = -3$. Additional numerical simulations demonstrate that the analytic predictions are valid for large systems (Fig. 2). Therefore, the differences between the result of numerical simulations based on the reference system and the analytic prediction resulted from the small number of fishes. These results reinforce the importance of the introduced approach for the study of small complex systems.

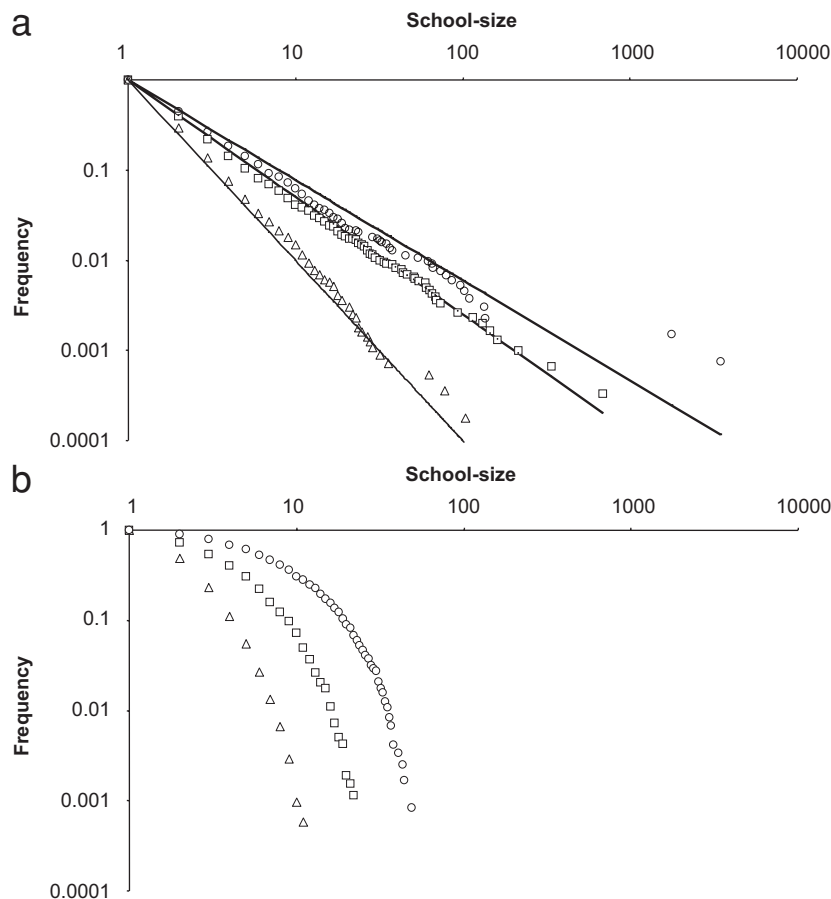


Fig. 2. (a) Agreement between numerical simulations and analytic predictions (lines) for large systems ($n = 10,000$ fishes) for preferential attachment model. School-size distributions were plotted as cumulative distributions to reduce finite-size fluctuations (Strogatz 2001): $p = 0.10$ (circles), $p = 0.25$ (squares), and $p = 0.5$ (triangles); (b) school-size distributions generated by numerical simulations using random attachment are plotted for comparisons.

The approach used here allows model selection, leading to particular hypotheses of how small schools are formed. In the context of the reference system analysed here, a key question is: what is the biological basis for the tendency of fishes to join larger schools in *S. axillare*? For *S. axillare* it is reasonable to assume that solitary fishes more easily detect larger schools than small ones. Alternatively, it is well-known that mechanisms that are apparently non-related to PA may induce it, such as copying the behavior of other individuals (see Albert and Barabási, 2002; Powers, 2004 for a further discussion). In addition, higher mortality rates associated with smaller schools, fine scale attractiveness of habitat, fusion–fission processes, and the geometric morphology of the schools are examples that might induce equivalent results. Although these processes may not explain the patterns observed in the short-term schools that *S. axillare* forms in homogeneous habitats, they may generate highly skewed frequency distributions in other species. Thus, future studies should focus in how to separate these different ecological and behavioral mechanisms. By now, our study demonstrates that it is possible to investigate the underlying behavioral rules of small aggregations of animals based on a rigorous statistical formalism.

We emphasize that the approach described here is not restricted to the fish schools or to the particular models tested here, but can be applied to investigate any theoretical model of animal aggregation. Indeed, this approach can be used to the study of the origin of frequency distributions in other complex systems, such as the connectivity distribution in ecological networks. In this context, our approach may be a first step towards the characterization of systems with a small number of interacting elements—one of the greatest challenges in the study of complex systems (Amaral et al., 2004). Some studies with frequency distributions have tried to overcome the difficulty inherent to systems with a small number of elements by computing cumulative distributions (e.g. Guimarães et al., 2005). Although cumulative distributions are likely to be less sensitive to fluctuations (Strogatz, 2001), the effects of small size on the properties of cumulative distributions generated by theoretical models are poorly explored; the same is true for the statistical dependence of the observations in a cumulative function. As an alternative approach, we suggest focusing on the fitted parameters of the non-cumulative distribution and its comparison with the empirical distribution of parameter estimates generated by the models. The tools used in this approach, namely general linear models (McCulloch and Searle, 2000) and numerical simulations (Gotelli and Graves, 1996), are relatively simple and often used in ecological studies.

Acknowledgements

We thank J. M. Silva-Jr (Centro Golfinho Rotador) and the Projeto Tamar (C. Bellini and A. Grossman) for

logistical support at Fernando de Noronha Archipelago; the Ibama (MA Silva) for the always warm reception, logistical support and issuing study permits at Fernando de Noronha Archipelago; H.S. Niwa, I. Sazima and three anonymous reviewers for comments on the manuscript; S. Hyslop for revising the English of the manuscript; the CAPES, CNPq, FAEP- UNICAMP, FAPESP, and PROAP-UNICAMP for financial support.

References

- Albert, R., Barabási, A.L., 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97.
- Allee, W.C., 1933. *Animal Aggregations: A Study in General Sociology*. AMS Press, New York.
- Amaral, L.A.N., Barrat, A., Barabási, A.L., Caldarelli, G., de los Rios, P., Erzan, A., Kahng, B., Mantegna, R., Mendes, J.F.F., Pastor-Satorras, R., Vespignani, A., 2004. Virtual round table on ten leading questions for network research. *Eur. Phys. J. B.* 38, 143–145.
- Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. *Science* 286, 509–512.
- Bonabeau, E., Dagorn, L., 1995. Possible universality in the size distribution of fish schools. *Phys. Rev. E* 51, R5220–R5223.
- Bonabeau, E., Dagorn, L., Freon, P., 1999. Scaling in animal group-size distributions. *Proc. Natl Acad. Sci. U.S.A.* 96, 4472–4477.
- Bonaldo, R.M., Krajewski, J.P., Sazima, C., Sazima, I., 2006. Foraging activity and resource use by three parrotfish species at Fernando de Noronha Archipelago, tropical West Atlantic. *Mar. Biol.* 149, 423–433.
- Bruggemann, J.H., Begeman, J., Bosma, E.M., Verburg, P., Breeman, A.M., 1994a. Foraging by the spotlight parrotfish *Sparisoma viride*. II. Intake and assimilation of food, protein and energy. *Mar. Ecol. Prog. Ser.* 106, 57–71.
- Bruggemann, J.H., Kuyper, M.W.M., Breeman, A.M., 1994b. Comparative analysis of foraging and habitat use by the sympatric Caribbean parrotfish *Scarus vetula* and *Sparisoma viride* (Scaridae). *Mar. Ecol. Prog. Ser.* 112, 51–66.
- Carleton, M.D., Olson, S.L., 1999. Amerigo Vespucci and the rat of Fernando de Noronha: a new genus and species of Rodentia (Muridae: Sigmodontinae) from a volcanic island off Brazil's continental shelf. *Am. Mus. Nov.* 3256, 1–59.
- Davison, A.C., Hinkley, D.V., 1997. *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge.
- Gotelli, N.J., Graves, G.R., 1996. *Null Models in Ecology*. Smithsonian Institution Press, Washington.
- Guimarães, P.R., de Aguiar, M.A.M., Bascompte, J., Jordano, P., dos Reis, S.F., 2005. Random initial conditions in small Barabási–Albert networks and deviations from the scale-free behavior. *Phys. Rev. E* 71, 037101.
- Hubbell, S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton.
- Jordano, P., Bascompte, J., Olesen, J.M., 2003. Invariant properties in coevolutionary networks of plant–animal interactions. *Ecol. Lett.* 6, 69–81.
- McCulloch, C.E., Searle, S.R., 2000. *Generalized, Linear, and Mixed Models*. Wiley–Interscience, New York.
- Moura, R.L., Figueiredo, J.L., Sazima, I., 2001. A new parrotfish (Scaridae) from Brazil, and revalidation of *Sparisoma amplum* (Ranzani, 1842), *Sparisoma frondosum* (Agassiz, 1831), *Sparisoma axillare* (Steindachner, 1878) and *Scarus trispinosus* (Valenciennes, 1840). *Bull. Mar. Sci.* 68, 505–524.
- Niwa, H.S., 1996. Mathematical model for the size distribution of fish schools. *Comput. Math. Appl.* 32, 79–88.
- Niwa, H.S., 1998. School size statistics of fish. *J. Theor. Biol.* 195, 351–361.
- Niwa, H.S., 2003. Power-law versus exponential distributions of animal group sizes. *J. Theor. Biol.* 224, 451–457.

- Niwa, H.S., 2004. Space-irrelevant scaling law for fish school sizes. *J. Theor. Biol.* 228, 347–357.
- Niwa, H.S., 2005. Power-law scaling in dimension-to-biomass relationship of fish schools. *J. Theor. Biol.* 235, 419–430.
- Parrish, J.K., Edelstein-Keshet, L., 1999. Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science* 284, 99–101.
- Powers, J.E., 2004. Recruitment as an evolving random process of aggregation and mortality. *Fish. Bull.* 102, 349–365.
- Starck, A.S., Davis, W.P., 1966. Night habits of fishes of Alligator Reef, Florida. *Ichthyologica* 38, 315–356.
- Strogatz, S.H., 2001. Exploring complex networks. *Nature* 410, 268–276.

Paulo. R. Guimarães Jr.
Departamento de Física da Matéria Condensada,
Instituto de Física Gleb Wataghin,
UNICAMP, 13083-970, Campinas, SP, Brazil
and
Integrative Ecology Group, Estación Biológica de Doñana,
CSIC, Apdo. 1056, E-41080 Sevilla, Spain
E-mail address: prguima@ifi.unicamp.br

Roberta M. Bonaldo
 João P. Krajewski
Programa de Pós-Graduação em Ecologia,
Instituto de Biologia, UNICAMP, 13083-970,
Campinas, SP, Brazil

E-mail addresses: robertabonaldo@yahoo.com
(R.M. Bonaldo),
jpaulokra@yahoo.com.br (J.P. Krajewski)

Paulo Guimarães
Daitan Labs, Galleria Office,
Bloco 4, cj 444, Campinas, SP, Brazil
E-mail address: prguima.pm@gmail.com

Aluísio Pinheiro
Departamento de Estatística, Instituto de Matemática,
Estatística e Ciência da Computação, UNICAMP,
13083-970, Campinas, SP, Brazil
E-mail address: pinheiro@ime.unicamp.br

Joseph Powers
2147 Energy, Coast and the Environment Building,
School of Coast and Environment,
Louisiana State University, Baton Rouge,
LA 70803, USA
E-mail address: jepowers@lsu.edu

Sérgio Furtado dos Reis
Departamento de Parasitologia, Instituto de Biologia,
UNICAMP, 13083-970, Campinas, SP, Brazil
E-mail address: sfreis@unicamp.br