

# Network Analysis in a Natural Perturbed Ecosystem

Nelson F.F. Ebecken, Gilberto C. Pereira

**Abstract**—The objective of this work is to explicit knowledge on the interactions between the chlorophyll-a and nine meroplankton larvae of epibenthonic fauna. The studied case is the Arraial do Cabo upwelling system, Southeastern of Brazil, which provides different environmental conditions. To assess this information a network approach based in probability estimative was used. Comparisons among the generated graphs are made in the light of different water masses, application of Shannon biodiversity index, and the closeness and betweenness centralities measurements. Our results show the main pattern among different water masses and how the core organisms belonging to the network skeleton are correlated to the main environmental variable. We conclude that the approach of complex networks is a promising tool for environmental diagnostic.

**Keywords**—Coastal upwelling, Ecological networks, Plankton - interactions, Environmental analysis.

## I. INTRODUCTION

THE Brazilian coast presents a large variety of ecosystems but little is known about its biodiversity, degree of connectivity and behavioral patterns. Nowadays, these systems are subjected to a large number of anthropogenic pressures without either knowing the load processing ability of the biological networks and its structural stability. These issues make any initiative in coastal management hard and complex for decision making. One of the major problems is the habitat change, destruction or loss [1]-[2]. Despite their adaptive character [3] and often redundant linkages, marine ecosystems are vulnerable to rapid changes in diversity and function [4]. The widespread decline of species, habitats, and ecosystem function have led to calls for ecosystem based management (EBM) as a solution for what ails the oceans [5]. Recent legislative instruments have been approved worldwide addressing the need to assess the ecological status [6]. In this way, many initiatives can be found in literature [7]-[8]. According [9], environmental factors regulate biodiversity through species sorting processes. Species distributions in communities affect ecosystem processes and environmental factors. These dynamics are determined by the traits of species in the community. The question of how changes in biodiversity will affect the ecosystem functioning, the so-called biodiversity-ecosystem function (BEF) debate, is clearly not easy to answer. However, it has long been recognized that species interact in ecosystems with other species and with abiotic factors in many ways, of which pairwise interactions are only one possibility. In this context, [10] argued that a system approach is necessary to address issues involving changes in biodiversity and function of natural ecosystems. Although pairwise interactions have always had a key role in ecology, a new focus on complex networks has been placed [11].

N.F.F.Ebecken. Author is with the Civil Engennering Program – COPPE/UFRJ; CEP: 21941-914 Brazil (phone 55.21.2562-7389; fax 55.21.2562-8464; e-mail : nelson@ntt.ufrj.br).

Gilberto C.Pereira Author is with the Civil Engennering Program – COPPE/UFRJ; CEP: 21941-914 Brazil (phone 55.21.2562-7389; fax 55.21.2562-8464; e-mail: gcp@coc.ufrj.br)

Several studies have shown how the structural characteristics of complex networks are related to their stability and dynamic [12]-[13]-[14]-[15]. The application of centralities indices to the network component can identify the keystone species [16], and the role they play in a network [17]. So, the aim of this paper is (i) to use a network approach to investigate differences between different water masses based on graphs generated from chlorophyll-a, merozooplankton larvae of epibenthic fauna and some environmental parameters; (ii) establish differences in biodiversity; (iii) apply the closeness and betweenness centralities measures in order to determine the positional importance of each specie or node; (iv) identify which set of n node belongs to the core skeleton of the network, (v) examine how the core organisms are correlated to the main environmental variable.

## II. MATERIAL AND METHODS

### A. Studied Area

The studied plankton community is found in a small (45 km<sup>2</sup>), shallow (10m depth), wind-driven and upwelling-influenced Anjos Bay, which is formed by Cabo Frio Island (23 S, 42 W) in the state of Rio de Janeiro, southeastern Brazil (Fig. 1).

Dominant E-NE winds are influenced by tropical maritime anticyclones due to the Coriolis Effect and Ekman transport, which shunt nutrient-depleted surface water (Brazil Current) offshore [18]. This water body is followed by up-flowing, nutrient-rich (12  $\mu$ MLNO<sub>3</sub>-N), deeper South Atlantic Central Water (SACW), which comes from around 200–300m depth and reaches the surface sporadically. An inverse pattern can be caused by S-SW winds because cold fronts drive the oligotrophic Brazil Current (<1  $\mu$ M-LNO<sub>3</sub>-N) toward the coast. As SACW is heated in the euphotic layer, nitrate declines more rapidly than phosphate, and the N/P ratio declines [19]. These processes generate different habitat conditions that influence at the same time changes in community and trophic structure [20].



Fig. 1 The studied site showing the sampling location

### B. Available data

The available data (Table I), is a matrix of 18 variables and 512 samples concerning to a weekly harvested medium-term time series (10 years) of physical, chemical and biological gradients coming from November of 1994 to December of 2005.

The physical and chemical variables demonstrate the hydrologic variability of the environment as a function of interchangeable periods of upwelling and downwelling events. The water mass identification was made through temperature and salinity gradients according [8]. The biological variables are the chlorophyll-a as estimation of phytoplankton biomass, a single food resource, and 9 merozooplankton taxa representing consumers. Every variable were categorized into five classes: zero which means no occurrence, low, mean, high and extremely high.

### C. Network Generation

In this paper we present an approach for the discovery of community structure in networks with only a single type of vertex (although they represent biotic and abiotic variables we considered them the same type) and a single type of undirected and unweighted edge, although generalizations to more complicated network types are possible. Our divisive algorithm focus is not on removing the edges between vertex pairs with low similarity, but on finding edges with the highest values of occurrences, i.e. we focus on finding community structure based on the values of the edges and not on the attributes of the vertices, as is more usual.

TABLE I  
BASIC STATISTICS OF THE AVAILABLE DATA

Variables	Min	Max	Mean	Std dev
Temperature (°C)	15,88	29,40	22,66	1,84
Salinity (g/L)	32,13	39,78	35,82	0,86
Oxygen (O <sub>2</sub> )(mg/L)	2,58	8,79	5,29	0,49
Phosphate (PO <sub>4</sub> ) (µg/l)	0,00	3,69	0,26	0,21
Nitrite (NO <sub>2</sub> ) (µg/l)	0,00	0,64	0,08	8,08
Nitrate(NO <sub>3</sub> ) (µg/l)	-0,09	10,19	0,68	0,95
Ammonium (NH <sub>4</sub> ) (µg/l)	0,07	7,85	1,26	0,88
PH	6,39	10,44	8,13	0,42
Chlorophyll-a (mg/m <sup>3</sup> )	0	11,94	0,99	1,17
Cirripedia (Org/m <sup>3</sup> )	0	3641	205	355
Mytilidae (Org/m <sup>3</sup> )	0	2636	92	170
Decapoda (Org/m <sup>3</sup> )	0	437	20	35
Polychaeta (Org/m <sup>3</sup> )	0	1683	20	89
Ostreidae (Org/m <sup>3</sup> )	0	1132	31	91
Cypris (Org/m <sup>3</sup> )	0	5192	21	248
Asciacea (Org/m <sup>3</sup> )	0	1115	14	66
Isogonomon (Org/m <sup>3</sup> )	0	2342	31	161
Bryozoa (Org/m <sup>3</sup> )	0	101	2	6

Each variable is a vertex or node in a graph whose edges represent the interaction between them. We were not interested to know only which variables each population interacts with, but to measure the simultaneity intensity of this interaction. To quantify these interactions, it was considered the probability of presence of variable Bi given the presence of a variable Bj, and is thus a measure of the statistical association between Bi and Bj, represented by P (Bi|Bj) which measures the strength of the association between Bi and Bj. As P (Bi|Bj) does not take into account statistical confidence, we

considered the equation 1, proposed by Stephens et. al. (2009), which also measures the degree of confidence one can have in the statistical association between Bi and Bj relative to the null hypothesis, P(Bi), that the distribution of Bi is independent of Bj and distributed with this probability over the region of interest (in our study, just one geographic position).

$$E (Bij | Bj) = NBj (P( Bi | Bj) - P( Bi)) / (NBj (P( Bi) (1 - P(Bi)))^{1/2} \quad (1)$$

Essentially, it is a one-sided binomial test where the null hypothesis is that the distribution of Bi is random over the collected data.

The sum of the values of the edges was considered to identify the network structure. Thus, in our proposal, an edge is considered part of a sub-network if it connects a vertex pair in an amount equal to a defined threshold. Naturally, the vertex pair connected by this edge is also part of this sub-network. So, as an example, if a vertex is part of a 10-threshold sub-network, it is connected by an edge with value "10" to at least one other vertex. The approach we take to identify the structure of the network follows roughly these lines. Thus, the general form of our network structure finding algorithm is as follows:

1. Calculate the value of each interaction with equation 1.
2. Calculate the value of the sum of the edge (for each pair of vertices) in the network.
3. Calculate the frequency distribution of the edges over the values (each frequency distribution class is equal 1).
4. Identify the values that correspond to 25%, 50% and 75% of the frequency distribution as a reference for first pre-division of the edges (and respective vertex pair) into sub-networks. Naturally, a vertex is categorized in a frequency distribution class according the highest value of its edges.
5. Categorize each edge (and the vertex pair connected by it) in one of the 4 intervals defined above.
6. Identified the edges with the highest values (above 75% interval) and considered them part of the core network.

For graphical representation, the interactions whose values are higher than the value that corresponds to 75% of distribution (empirical threshold) appear reinforced. In fact, this threshold was set based on the average summed to one fold of the standard deviation of all water masses. On this way, the nodes connected by these stronger values of interactions will be considered as belonging to the skeletons of such networks

### D. Applied Indices

The Shannon-Wiener index was applied to each network to access differences in biodiversity. It was computed as:

$$H' = - \sum_{i=1}^S p_i \ln p_i \quad (2)$$

Where S is the total number of species, called species richness; N the total number of individuals. pi is the relative abundance of each species i, calculated as the proportion of individuals of a given species to the total number of individuals in the community: ni/N.

In order to establish the positional importance of each node we applied two common measures of centrality: closeness (CC) and betweenness (CB). The former, is based on the total distance between one vertex and all other, such that large distances yield low centrality values. In the network theory, it is defined as the mean geodesic distance between a vertex  $v$  and all other vertices reachable from it such as:

$$C_c(v) = \sum_{t \in V \setminus v} 2^{-d_G(v,t)} \quad (3)$$

Where  $n \geq 2$  is the size of the network's connectivity component  $V$  reachable from  $v$ . Closeness can be regarded as a measure of how long it will take information to spread from a given vertex to other reachable vertices in the network [21]. On the other hand, betweenness is a centrality measure of a vertex within a graph so that vertices that occur on shorter paths between others have higher betweenness than those that do not. For a graph  $G=(V,E)$  with  $n$  vertices, the betweenness  $C_B(v)$  for vertex  $v$  is computed as follows:

1. For each pair of vertices  $(s,t)$ , compute all shortest paths between them.
2. For each pair of vertices  $(s,t)$ , determine the fraction of shortest paths that pass through the vertex in question (here, vertex  $v$ ).
3. Sum this fraction over all pairs of vertices  $(s,t)$ . Such that [22]:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{ST}(v)}{\sigma_{st}} \quad (4)$$

### III. RESULTS AND DISCUSSION

The use of temperature and salinity data enabled us to identify different water masses such that the Coastal/Tropical mixing type corresponds to 44.80% of the occurrences followed by the Tropical water of Brazil current (25,57%), Coastal water (22.40%), SACW/Coastal (3.17%), SACW/Tropical (1.36%) and SACW with only 0.90%. However, as previously reported by [8], it was found that 1.81% of the examples do not belong to any of these ranges, suggesting another class of water, identified here as "New". Table II presents the results of the applied index related to each node and water masses. The highest biodiversity (2) occurs in the mixing of Coastal/Tropical Water Mass, while the smallest (1,47) was verified in the SACW. The centrality values provide us a good evaluation about the positional importance of these populations or nodes in each of water mass. In another way, the Fig. 2 shows an example of the representative graph topology of the Coastal/Tropical Water Mass. It shows firstly the occurrence of the lowest (L) values of these variables indicating the oligotrophic condition of this water mass. It is also possible to see that chlorophyll-a is strongly and preferably associated to the ammonium (NH4) followed by phosphate (PO4), nitrate (NO3) and nitrite (NO2) respectively. The PO4 importance to chlorophyll in this system has been previously highlighted in [19]. This figure still shows the occurrence of Cirripedia, Mytilidae and Decapoda as the main consumers. The chlorophyll-a and these three groups of consumers are present in the most of water mass (data not show) indicating they represent the skeleton of the biological network at the studied site. Differences were detected by the presence of mean values of chlorophyll-a(M) n SACW and SACW/ Tropical water mass that is the result of upwelling process and the absence of Decapoda in the class "New".

TABLE II  
THE SHANNON-WIENER, CLOSENESS AND BETWEENNESS INDICES OF THE STUDIED POPULATIONS FOR EACH WATER MASS

	Coastal Water	Coastal tropical	Tropical	Acas	Acas Coastal	Acas Tropical
Shannon Index	1.94	2.0	1.97	1.47	1.63	1.49
Closeness and Betweenness Centralities (CC-CB)						
Biotic Variables						
Asciadiacea	0.9-0	0.69-0	0.9-0	0-0	0.64-0	0.47-0
Bryozoa	0.9-0	0.75-0.02	1-0.02	0-0	0.64-0	0-0
Cirripedia	1-0.03	0.9-0.04	1-0.02	0.53-0	0.75-0.02	0.7-0.06
ClorofA	1-0.03	1-0.27	1-0.02	0.7-0.25	1-0.28	0.8-0.23
Cypris	0.69-1	0.53-0	0.69-0	0-0	0.56-0	0.43-0
Decapoda	1-0.03	0.82-0.02	1-0.02	0.38-0	0.9-0.19	0.62-0.01
Isognomon	0.82-0	0.69-0	0.9-0	0.38-0	0.56-0	0-0
Mytilidae	1-0.03	0.82-0.03	1-0.02	0.53-0	0.75-0.02	0.56-0
Ostreida	1-0.03	0.82-0.02	0.9-0	0.53-0	0.75-0.04	0.62-0.01
Polychaeta	0.82-0	0.69-0	0.9-0	0.53-0	0.6-0	0.56-0

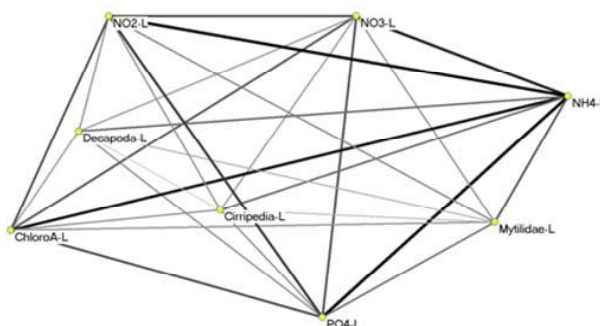


Fig. 2 The graph of the core network topology representative of the Coastal/Tropical water mass

### IV. CONCLUSION

The fundamental goals underlying community ecology is to model the distribution of biota, identify their interactions patterns and understand what drives the assemblages in order to perform predictions. The biological monitoring of the marine part of coastal zone is crucial and has become a politically as well a scientifically vital task. The main contribution of this paper is to show how the representation of biological interaction could be constructed through a network approach to discriminate those of greater influence for a specific condition. It was possible to identify the core network of each water mass and their similarities.

### ACKNOWLEDGMENT

This paper was supported by CAPES the Brazilian research agency and FAPERJ the Rio de Janeiro State research foundation.

### REFERENCES

- [1] Halpern, B.S., Selkoe, K.A., Micheli, F., Kappel, K., 2007. *Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats*. Conservation Biology 21, 1301–1315.
- [2] Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., Fujita, R., Heinemann, D., Lenihan, H.S., Madin, E.M.P., Perry, M.T., Selig, E.R., Spalding, M., Steneck, R., Watson, R., 2008. *A global map of human impact on marine ecosystems*. Science 319, 948–952.
- [3] Levin, S.A., Lubchenco J. 2008. *Resilience, robustness, and marine ecosystem based management*. Bioscience 58, 1-11.

- [4] Palumbi S, McLeod KL, Grunbaum D. 2008. *Ecosystems in action: Lessons from marine ecology about recovery, resistance, and reversibility*. *Bioscience* 58 (1): 33-42.
- [5] [USCOP] US Commission on Ocean Policy. 2004. *An Ocean Blueprint for the 21st Century*. Washington (DC): USCOP.
- [6] Borja, A., D.M. Dauer, 2008. *Assessing the environmental quality status in estuarine and coastal systems: comparing methodologies and indices*. *Ecological Indicators*, 8(4): 331-337.
- [7] Windhorst, W., Colijn, F., Kabuta, S., Remi Laane, R., Lenhart, H.J. 2005. *Defining a good ecological status of coastal waters — a case study for the Elbe plume*. *Environmental*
- [8] Pereira, G.C., Coutinho, R., Ebecken, N.F.F. 2008. *Data Mining for environmental analysis and diagnostic: a case study of upwelling ecosystem of Arraial do Cabo*. *Brazilian Journal of Oceanography* 56(1):1-12.
- [9] Norberg, J. 2004. *Biodiversity and ecosystem functioning: A complex adaptive systems approach*. *Limnol. Oceanogr.*, 49(4, part 2), 1269–1277.
- [10] Raffaelli, D., Van den Putten, W. H., Person, L., Wardle, D. A., Petchey, O. L., Koricheva, J., Van den Heijden, M. G. A., 2002. *Multi-trophic processes and ecosystem function*. In *Biodiversity and Ecosystem Functioning*. Ed. By M. Loreau,
- [11] Dunne, J.A., Williams, R.J., Martinez, N.D., 2002. *Network structure and biodiversity loss in food webs: robustness increases with connectance*. *Ecol. Lett.* 5, 558–567.
- [12] Huxel, G.R., Mc Cann, K. 1998. *Food Web Stability: The Influence of Trophic Flows across Habitats*. *The American Naturalist* 152 (3), 460-469. Jordán, F., W-C. Liu and A.J. Davis. 2006.
- [13] Albert R, Jeong H, Barabási AL, 2000, Error and attack tolerance of complex network. *Nature* 406: 378-382.
- [14] Strogatz SH, 2001. *Exploring complex networks*. *Nature* 420: 268-276.
- [15] Kolasa, J., 2005. *Complexity, system integration, and susceptibility to change: biodiversity connection*. *Ecol. Complex.* 2, 431–442. Levin SA,
- [16] Libralato, S., V. Christensen and D. Pauly. 2006. *A method for identifying keystone species in food web models*. *Ecological Modeling* 195:153-171.
- [17] Gonzalez AMM, Dalsgaard B, Olesen JM, 2010. *Centrality measures and the importance of generalist species in pollination network*. *Ecological Complexity* 7, 36-43.
- [18] Castelao, R. M., Barth, J. A., 2005. *Upwelling around Cabo Frio, Brazil: The importance of wind stress curl*, *Geophys. Res. Lett.*, 33, L03602.
- [19] Pereira, G. C., Evsukoff, A., and Ebecken, N. F. F. 2009. *Fuzzy modelling of chlorophyll production in a Brazilian upwelling system*, *Ecol. Model.*, 220, 1506–1512.
- [20] Pereira, G.C., Ebecken, N.F.F. 2009. *Knowledge discovering for coastal waters classification*. *Expert Systems with Applications* 36, 8604–8609.
- [21] Newman, M.E.J. 2003. *A measure of betweenness centrality based on random walks*. *Arxiv preprint cond-mat/0309045*. 1-15 pp.
- [22] Shivaram, N. 2005. *The Betweenness Centrality of Biological Networks*. Thesis of the Faculty of Virginia Polytechnic Institute and State University. 74p.