

# Modeling Disruption and Importance in Food Webs

Ryan Holmdahl      Christopher Yeh

CS 224W Project, Fall 2016

## 1 Abstract

We propose an ecosystem model that uses the network structure of food webs to simulate external disruptions. We then examine how our model can be used to measure an organism’s importance to its ecosystem. Using this understanding of importance, we identify key network centrality measures that can be used to approximate organism importance.

## 2 Problem Statement

As environmental conservation becomes a more pressing issue in the face of urbanization and climate change, understanding the structure of ecological networks and how organisms interact with one another becomes more pressing as well. Efforts to maintain ecosystem biodiversity must be as efficient as possible in order to maximize returns. A key component of this biodiversity effort is an understanding of an organism’s importance to its ecosystem. We define “importance” as the severity of consequences to biodiversity an organism’s extinction would have, a definition typically implied by the literature.

With a definition of “importance” established, two problems become apparent. First, one must devise a means of predicting the consequences of an organism’s extinction. Second, one must devise a means of transforming predicted environmental impacts into a quantified importance metric.

If those problems can be addressed, then there are several promising avenues to explore. We are particularly interested in the question of how organism importance relates to the properties of its ecological network. In particular, we hypothesize that node importance may have some correlation with some or all of its network centrality measures. Within this topic, we also hope to identify importance-correlated measures that require as little information as possible about a network, facilitating understanding for resource-restricted researchers.

## 3 Approach

To identify an organism’s importance, we must have some means of determining what its impact on the ecosystem is. Our solution to this problem is the construction of a model to simulate the consequences of environmental disruptions, such as extinctions. Using an impact scoring algorithm which we define, these predictions can be used to quantify and compare organism importance. We then leverage these tools to identify network centrality measures that correlate strongly with organism importance and analyze the implications of our findings.

## 4 Related Works

Existing research has examined how an organism’s extinction impacts the stability and feasibility of its ecosystem. In [1], researchers constructed simple food web models of autotrophs, primary consumers, and secondary consumers and applied a continuous time Lotka-Volterra model to analyze the effects of extinction. The researchers found that diversity within trophic levels increased network resilience and that removing organisms of higher trophic levels was less impactful than removing organisms of lower levels. This work

presents a foundation for impact-based Ecological Network Analysis (ENA) using a predation model, but relies on unnaturally simple trophic ordering and excludes important environmental features like detritus.

Recently, throughflow centrality, the inflow of matter to an organism, has been proposed as the standard measure of an organism’s importance [2]. Its supporters argue that its simplicity and tendency to assign high importance to low-trophic-level organisms makes it a strong choice for measuring importance. However, we find the work to lack meaningful support for its claims, providing no comparisons with other centrality measures and no consistent definition of importance.

Other measures, such as Google’s PageRank algorithm, show potential as superior analyses of node importance. Originally derived to calculate the relevance of web pages, PageRank leverages network structures to assign a value to a node based on the values of the nodes feeding into it [3]. In an ecological context, one could imagine that an organism’s importance to an ecosystem is related to the importance of the organisms that consume it. In other words, it seems plausible that the food sources of important organisms are themselves important. We will test this intuition by using the PageRank algorithm as one of our measures of node centrality.

## 5 Data

We use the Pajek Food Web datasets hosted online by Professor Vladimir Batagelj of the University of Ljubljana. Nodes in these networks represent organisms, detritus “piles”, output “sinks” such as respiration, or input sources, and each node has an associated biomass. There are weighted, directed edges that indicate the transfer of biomass between nodes. Specifically, edges point from prey to predator. For this paper, we examined ten real-world ecological networks: Chesapeake Bay (aggregate, upper, lower, and middle) [4] [5], Crystal Creek (control and delta temp) [6], Mondego Estuary [7], St Marks River Estuary [8], Lake Michigan [9], and Narragansett Bay [10]. Node counts vary from 24 to 54 and edge counts from 100 to 400.

## 6 Ecosystem model

We represent an ecological network as series of species nodes connected by edges representing biomass flows. A species has some initial biomass and a set of in- and out-edges which at equilibrium are balanced.

### 6.1 Assumptions

We make the following assumptions in our model:

1. An input food web is at equilibrium.
2. A change in species biomass propagates through the network in discrete unit-less time steps.
3. Differences between organism input and output are kept as biomass.
4. Mass flow between organisms and output sinks (respiration and other losses) or detritus piles is proportional to the mass of the organism.
5. Mass flow between predators and prey – either organisms or detritus piles – is proportional to the mass of each.
6. Mass flow between input sources and organisms is proportional to the mass of the organisms.

### 6.2 Model Overview

Our model operates as follows:

1. An empirical network is input to the model as a set of nodes and edges.
2. The model calculates the necessary predator-prey consumption rates required to put the input web at equilibrium (see next section).

3. A species biomass is changed to reflect some event.
4. For some number of iterations:
  - (a) Mass flows are updated according to the modeling assumptions and compartment biomasses.
  - (b) Compartment biomasses are updated based on the mass flow.

### 6.3 Consumption Equations

Our assumptions lend themselves to simple mathematical interpretations. For a prey  $n_1$  – either an organism or a detritus pile – and predator  $n_2$ , the mass flow  $m_{n_1, n_2}$  is defined as:

$$m_{n_1, n_2} = c_{n_1, n_2} b_{n_1} b_{n_2}$$

where  $b_n$  is the biomass of node  $n$  and  $c_{n_1, n_2}$  is a consumption constant. Since our input networks are at equilibrium, we can use this equation and the given mass flows and biomasses to calculate our constants:

$$c_{n_1, n_2} = \frac{m_{0, n_1, n_2}}{b_{0, n_1} b_{0, n_2}}$$

where  $b_0$  is the given biomass and  $m_0$  is the initial mass flow.

When  $n_1$  is an input source – therefore having no biomass – and  $n_2$  is an organism, the equations are as follows:

$$\begin{aligned} m_{n_1, n_2} &= c_{n_1, n_2} b_{n_2} \\ c_{n_1, n_2} &= \frac{m_{0, n_1, n_2}}{b_{0, n_2}} \end{aligned}$$

When  $n_1$  is an organism and  $n_2$  is an output sink or detritus pile, the equations are as follows:

$$\begin{aligned} m_{n_1, n_2} &= c_{n_1, n_2} b_{n_1} \\ c_{n_1, n_2} &= \frac{m_{0, n_1, n_2}}{b_{0, n_1}} \end{aligned}$$

For each iteration in the simulation, we first update the mass flows  $m$  according to these rules. When all of the flows have been updated, we update the biomass of each node  $b_n$  using the mass flows and the biomass in the previous iteration  $b_{prev_n}$ :

$$b_n = \max\{0, b_{prev_n} + \sum_{n_1} m_{n_1, n} - \sum_{n_2} m_{n, n_2}\}$$

The simulation then continues to the next iteration.

### 6.4 Assumption Analysis

To assess the strengths and limitations of the model, it is useful to examine in more detail its assumptions.

1. The initial equilibrium assumption allows us to construct models based on empirical data, but is inaccurate if the environment is undergoing change during the measurement period. Those using the model should be careful to ensure that their target environment matches this assumption. However, consumption of prey by predators certainly does occur in discrete events, so discretizing population changes does somewhat approximate our understanding of nature.
2. Time does not flow in discrete time steps, and growth of different species may occur on different intervals. Moreover, the ability of a predator to respond to a change in a prey's population could vary significantly from one species to the next.

3. Aggregate across species and conditions, it is reasonable to expect a population to grow as it has access to more nutrients; however, factors besides biomass intake could dictate a population's ability to grow. Limitations not accounted for in our model, such as space, micronutrients, water, and other resources, can restrict reproduction.
4. Natural death is not necessarily a constant slice of the population, and not all of it necessarily goes into detritus. Detritus contributions could also be a product of predation. Similarly, respiration could be based on the availability of prey and the energy required to capture it. These concerns, however, should be relatively minor when considering an ecosystem as a whole, and should all tend to scale with population size in some respect, as assumed.
5. Predator-prey consumption is not necessarily linear with respect to populations. It is feasible for some prey to be sufficiently well protected that they are always inaccessible to predators, while it is also plausible that all prey beyond some critical threshold are far more available than the others. In addition, it is assumed that all predators benefit from the increase in a prey population, whereas in actuality it is plausible that growth of prey could increase availability only to a subset of its predators. Such factors, if known, could be incorporated into the model, but without this information the basic assumption should somewhat reflect typical predator-prey relationships.
6. Space and other environmental limitations could place some limit on the growth of autotrophs, and their ability to gather mass from input sources could certainly scale nonlinearly. Lacking this information, unrestricted growth is useful to avoid arbitrary limits and will likely not prove problematic unless the ecosystem reaches near-total collapse, at which point the explosive biomass of autotrophs is the least important information gained.

## 6.5 Comparison to Lotka-Volterra

Perhaps more useful than analysis in a vacuum is to compare our model to standards in the field of ENA. The classical model of competitive predator-prey relationships is the continuous time Lotka-Volterra model, which models the relationships between species  $S$  as follows:

$$\frac{\partial N_i}{\partial t} = N_i f_i(N)$$

where

$$f_i(N) = b_i + \sum_{j \in S} a_{ij} N_j$$

and  $N_i$  is the density of species  $i$ ,  $f_i$  is its per-capita rate of increase,  $b_i$  is the per-capita base change of species  $i$ , and  $N$  is the vector of all species densities. In this model, autotrophs continue to grow uninhibited in the absence of predators ( $b_i > 0$ ) and predators die without prey ( $b_i < 0$ ). For all  $i, j$ ,  $a_{ji} = -ea_{ij}$ , where  $e$  is the efficiency by which prey are converted into predators [1].

Our model shares many assumptions in common with Lotka-Volterra. Both assume that predators will grow uninhibited with sufficient prey. Both also assume that predator-prey consumption scales linearly with population density and that loss by respiration and death is a constant portion of population. Our simultaneous biomass update implementation is one means by which Lotka-Volterra can be implemented in practice.

We argue that our model modifies Lotka-Volterra such that it is useful for real-world ecological networks. In particular, our model solves two major problems of standard Lotka-Volterra. First, our model's incorporation of detritus piles and output sinks enables the existence of detritivores and better captures the realities of organism birth, death, and respiration. The unified birth-death constant  $b$  in Lotka-Volterra implies that autotrophs do not respire or die of natural causes and assumes that all dead matter leaves the ecosystem. Our model allows all organisms, even autotrophs, to contribute to reservoirs of detritus and to do so in whatever proportions appear in nature. In this way, we allow for far more complicated trophic structures than Lotka-Volterra.

Second, our model better preserves the flow of mass through the ecosystem. Lotka-Volterra imagines some constants  $e$  that define the efficiency by which prey are transformed into predators, but doing so

requires more complex knowledge of each predator-prey relationship than is often provided in ecological network datasets, such as the ones used here. We resolve this issue by assuming conservation of mass within the ecosystem (aside from its entry as input and exit as output or respiration) and structuring predator-prey relationships accordingly. This satisfies the law of conservation of mass while also allowing empirical mass-flow networks to be modeled without additional information.

## 7 Importance Modeling

We use our model to explore the concept of “importance” in an ecosystem. Specifically, we wish to see how the removal of a particular organism from an ecosystem impacts the population of other organisms. Removing a more important organism would cause larger population changes in the rest of the ecosystem than removing a less important organism.

### 7.1 Impact Score

We calculate the impact score of a node  $n$  as follows:

1. Initialize the network model.
2. Set the biomass of event node  $n$  to 0, effectively removing it from the ecosystem.
3. Run the biomass exchange procedure described in section 6.3 for a fixed number of iterations  $B$ , e.g.  $B = 100000$ .
4. Calculate the entries of the final biomass vector  $b_{final}$  for each node  $i$ :

$$b_{final_i} = \begin{cases} b_{avg_i} & \text{if } b_{end_i} > 0 \\ 0 & \text{if } b_{end_i} = 0 \end{cases}$$

where  $b_{avg_i}$  is the average biomass of node  $i$  across the modeling iterations and  $b_{end_i}$  is the biomass of node  $i$  on the last iteration.

5. Calculate the relative change in biomass for each node  $i$  after  $B$  iterations, where  $b_{0_i}$  denotes the initial biomass of node  $i$ :

$$b_{rel_i} = \frac{b_{avg_i}}{b_{0_i}}$$

6. Calculate the impact score  $s$ :

$$s = \frac{1}{|N|} \sum_{i, i \neq n}^{i \in N} (\min\{0, b_{rel_i} - 1\})^2$$

where  $N$  is the set of nodes in the network.

It is necessary to justify a few of our choices here. We use average masses of nodes instead of final masses to account for the possibility of dynamic equilibria. If the biomasses of nodes in the ecosystem are fluctuating around a set of equilibrium biomass values, then the average biomass is more representative of the equilibrium state than any single snapshot. If the ecosystem is at a static equilibrium, then, assuming sufficient iterations are used, the average should be only slightly different from the final equilibrium. We make an exception and use the ending mass for nodes with final mass 0, which are presumed to be extinct and have no chance of recovering.

The impact score of removing an event node is the sum of square decrease in relative biomass of each non-event node. This captures the change in biodiversity of the graph, not counting the initial removal of the event node. If other nodes do not lose biomass after removing the event node, then biodiversity is not necessarily impacted and we do not consider that node important. If some other nodes lose a significant amount of biomass, then biodiversity has been reduced and we consider the event node important. Also, we believe the presence of low-mass species is as important to biodiversity as the presence of high-mass species, so we treat the relative change in biomass equally for all nodes. Our metric is simple, but it captures the essence of biodiversity and should be sufficient for our purposes.

## 7.2 Centrality Measures

Our objective is to determine if a node’s measures of centrality at equilibrium correlate with its impact score. We present the measures of centrality we use and their significance in an ecological network:

- In-degree: the number of compartments from which the organism receives input.
- Out-degree: the number of compartments to which the organism provides output.
- Degree: the number of compartments with which the organism exchanges mass.
- Closeness (directed): the average path distance between the organism and all other compartments. Intuitively, the number of compartments through which matter must pass to get from the organism to another compartment.
- Betweenness (directed): the number of shortest paths between compartments that pass through the organism. Intuitively, a measure of the node’s role as a “hub” of mass exchange.
- PageRank: a measure of the organism’s importance as a function of the importance of the nodes on which it feeds (see [3]).
- Weighted PageRank: similar to vanilla PageRank, but edges are weighted based on the quantity of mass flow.
- Reversed Weighted PageRank: similar to weighted PageRank, but the edges are reversed. Measures importance as a function of the importance of nodes which are fed by the organism.
- Throughflow: mass which flows into and out of the organism.
- Biomass: mass of the organism.

## 7.3 Methodology

We examined ten real-world ecological networks, as described in the Data section (5) above. For each network, we calculated centrality measures and the impact score for each organism. For each network and for each centrality measure, we calculate the  $R^2$  correlation value between the nodes’ centrality measures and impact scores. We do not look for correlations across plots, as the magnitude of typical impact scores and centrality measures could vary significantly between networks of different structures.

After calculating an  $R^2$  value for each network-centrality pair, for each centrality measure, we calculate the mean value and coefficient of variation of  $R^2$  across the networks. The formula for a coefficient of variation is:

$$c_v = \frac{\sigma}{\mu}$$

where  $\sigma$  is the standard deviation and  $\mu$  is the mean. Since  $c_v$  is normalized to the mean, we can compare coefficients of variation between measures directly despite their potentially different averages.

Thus, for each centrality measure, we have a mean  $R^2$  across all 10 networks and a coefficient of variation of  $c_v$ . The mean  $R^2$  is an estimate of how strongly correlated the impact score and the centrality measure are across networks. However, a measure that is strongly correlated on some networks and weakly correlated on others is not particularly useful, since it would be unclear whether the measure will be applicable for a given network. Therefore, we use the coefficient of variation to quantify how varied the  $R^2$  values are. An ideal measure of node centrality would have a high mean  $R^2$  value and a low coefficient of variation, as this would indicate that it is consistently a good proxy for a node’s importance.

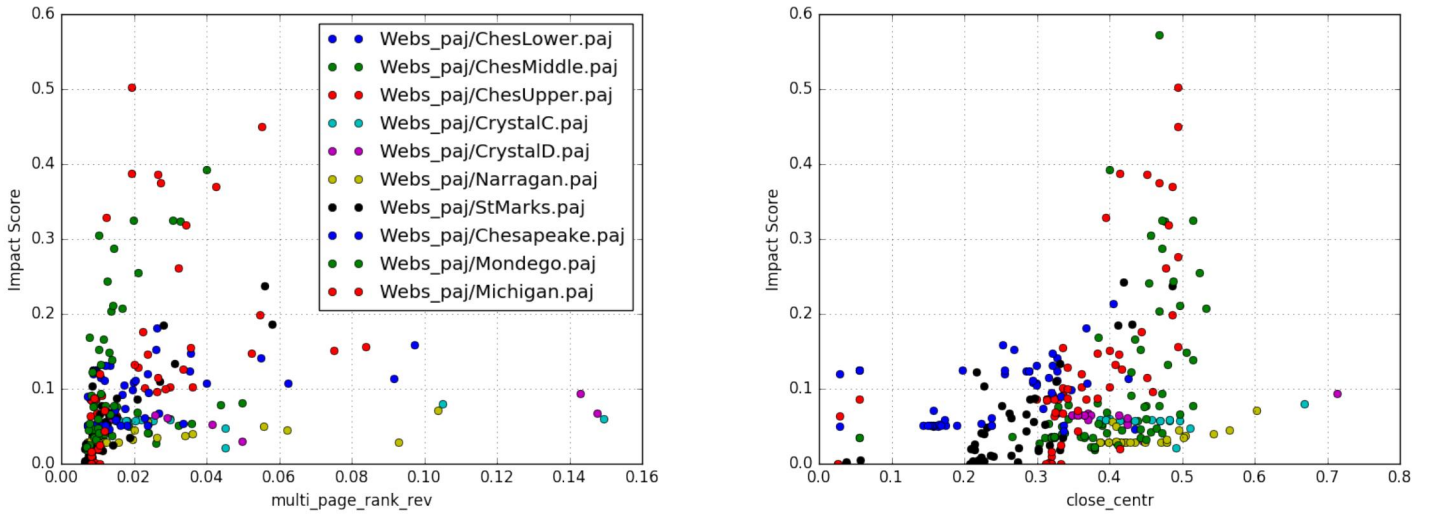


Figure 1: Plots of node centrality vs calculated impact scores for reversed weighted PageRank (left) and closeness (right). The correlation between the metrics is more clear in the PageRank than in closeness, although patterns are visible in each. Note that we are interested in intra-network correlations, not necessarily across networks, since impact score and centrality magnitudes can differ widely between networks.

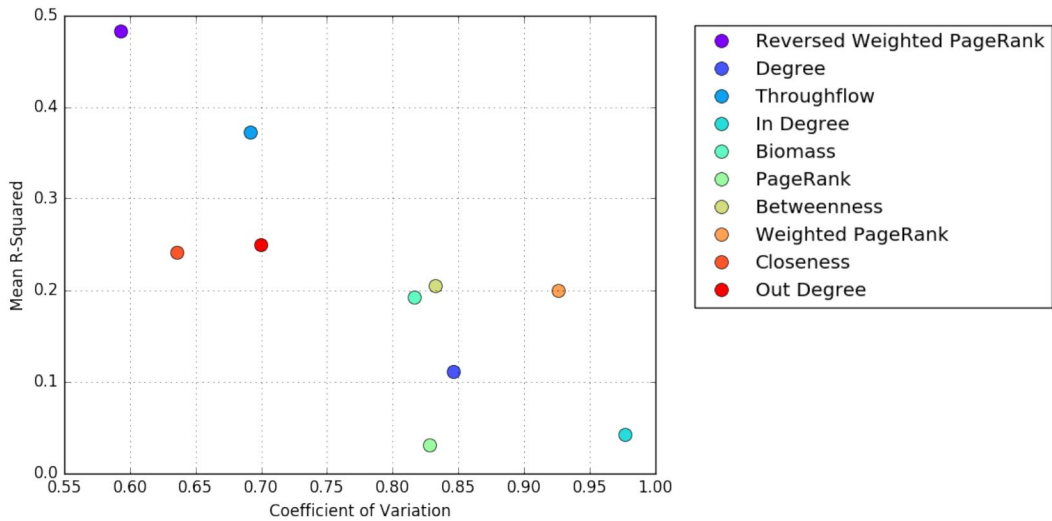


Figure 2: Plot of mean  $R^2$  vs coefficient of variation of  $R^2$  for each measure of node centrality. Reversed weighted PageRank showed the highest mean  $R^2$  while also having the lowest coefficient of variation.

## 7.4 Results

Figure 1 shows the raw centrality vs. impact scores on the nodes of each network for reversed weighted PageRank and closeness centrality. Calculating the mean  $R^2$  values for each network and their coefficients of variation of  $c_v$  gives us the results seen in Figure 2.

As one might expect from the clear correlation in Figure 1, reversed weighted PageRank has both the highest mean and the lowest coefficient of variation of  $R^2$  in Figure 2. Given this, it would seem that the fundamental assumption of reversed weighted PageRank, that the prey of important nodes is important, holds in our model.

Notably, throughflow centrality, which has been proposed as a standard centrality measure for ecological networks [2], was less strongly correlated with impact and had a higher coefficient of variation than reversed weighted PageRank. This is particularly interesting because both measures require the same information: a full understanding of the all edges (predator-prey relationships) and their weights (biomass transfer). This indicates that throughflow centrality is perhaps too naïve to adequately leverage the information inherent in the network structure, at least in our limited 10-network study, and should not necessarily be the default measure of importance.

Also interesting was reversed weighted PageRank’s performance relative to weighted PageRank and vanilla PageRank. Both weighted PageRank and vanilla PageRank had relatively poor performance – in particular, vanilla PageRank had the lowest mean  $R^2$  of any metric. This would indicate that nodes do not derive importance from their prey – that is, the predator of an important node is not necessarily important.

The difference in the mean  $R^2$  and coefficient of variation between out-degree and in-degree centrality measures further supports the intuition that more important nodes “support” higher trophic levels. Whereas out-degree was among the top 4 centrality measures for both mean  $R^2$  and coefficient of variation, in-degree was one of the least useful centrality measures. Since out-degree is the number of predators and in-degree is the number of prey, this is consistent with the idea that an organism’s relationship with higher trophic levels is more important than its relationship with lower trophic levels.

In studying ecosystems where mass flow information is less complete, the results also present some useful insights. If a researcher has an understanding about the trophic relationships between species but not necessarily the quantities of mass that flow between them, closeness centrality seems to be a decent proxy of node impact. Even simpler, a node’s out-degree is a slightly more variable but still-effective estimate.

## 8 Conclusion

We presented our ecological network model and showed how it expands upon the classic continuous time Lotka-Volterra model. Though the model suffers from some of the same flaws as Lotka-Volterra, namely its dependence on overly simple assumptions about predation and population growth, it allows existing competitive predation principles to be applied to complex empirical food webs. We then used this model to calculate a node’s importance as a function of its contribution to the structure of the ecosystem.

Using this function, we evaluated the strength of correlations between an organism’s importance and its measures of network centrality. We found that reversed weighted PageRank was highly correlated with importance and showed low variation between networks, particularly in comparison to the standard of throughflow centrality proposed by other work. Closeness centrality and out-degree proved to be decent estimators of importance for situations where full network information is not available.

Our findings can assist ecological researchers and policy makers to quickly and with potentially incomplete information identify important organisms in an ecosystem and make more educated conservation decisions accordingly. Going forward, we would be excited to incorporate additional ecosystem information into our model, such as geographic restrictions, predation habits, and reproduction cycles. Future research can build on our work to construct more advanced definitions of importance, including the impact of organisms on the environmental safety of their habitats and on nearby human developments.

## 9 Contributions

Ryan: model idea, design, and implementation, correlation calculations, writeup

Chris: data collection and processing, centrality calculations, writeup

## References

- [1] Charlotte Borrvall, Bo Ebenman, and Tomas Jonsson. Biodiversity lessens the risk of cascading extinction in model food webs. *Ecology Letters*, 3(2):131–136, 2000.
- [2] Stuart R. Borrett. Throughflow centrality is a global indicator of the functional importance of species in ecosystems. 2012.



- [3] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: bringing order to the web. 1999.
- [4] Baird D. and Ulanowicz R. E. The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecological Monographs*, 59:329–364, 1989.
- [5] J. D. Hagy. *Eutrophication, hypoxia and trophic transfer efficiency in Chesapeake Bay*. PhD Dissertation, University of Maryland at College Park (USA), 2002.
- [6] M. Homer and W. M. Kemp. Growth and Development: Ecosystems Phenomenology. Springer, New York. *Unpublished Ms. See also Ulanowicz, R. E. 1986.*, 69-79, 1986.
- [7] J. Patricio. (*In Preparation*) *Master's Thesis*. University of Coimbra, Coimbra, Portugal.
- [8] D. Baird, J. Luczkovich, and R. R. Christian. Assessment of spatial and temporal variability in ecosystem attributes of the St Marks National Wildlife Refuge, Apalachee Bay. *Florida. Estuarine, Coastal, and Shelf Science*, 47:329–349, 1998.
- [9] A. Krause and D. Mason. (*In preparation.*). A Krause, PhD.
- [10] M. E. Monaco and R. E. Ulanowicz. Comparative ecosystem trophic structure of three U.s. Mid-Atlantic estuaries. *Mar. Ecol. Prog. Ser.*, 161:239–254, 1997.