Fishery View
Improvement of Time Series Line Chart Visualization of Fishery Data
CPSC 533C Course Project Final Report

Ying Zhang and Lan Wu

{yingzhan,wlmakeit}@cs.ubc.ca

Abstract

Time series datasets are widely used in many areas. When the size of a dataset grows large or the number of datasets increases, it becomes difficult for a human to observe the structure of a dataset or to understand the underlying relationships. In this report, we present an interactive visualization tool called Fishery View to display and interact with fish population data obtained from the UBC Fishery Department. Using Fishery View, we are able to efficiently find related marine species of a specific fish, evaluate the impact of fishing strategy on fish population, and cluster fish species according to similar trends in population change. Our experimental results show that the fish population data is well organized and visualized in Fishery View.

1. Introduction

Time-dependent data and time series are ubiquitous and are widely used in many areas. "Whenever a process is monitored or an activity is recorded, the result is a time series (e.g. daily stock quotes, hourly air temperature etc.). Therefore, an important part of understanding a time-dependant process is analyzing the underlying time series"[1]. Among all the approaches we use to analyze the time series, "visualizing the analyzed series is a prominent analysis technique. The purpose of such visualizations is 'to assist the human in detecting and validating characteristic properties of an unknown function f '[1]". [2]

Visualization of time-dependent data becomes complex when we want to plot several processes along the same time axis. Curves and lines may overlap with each other as the number of processes increases. In addition, some of the questions regarding time series analysis still remain: Do any of the processes share the same pattern? Are two or more of these processes similar? How do these processes react to some outside effect? Can people do anything to optimize the trends of these processes?

In our project, we use some visualization techniques to help answer the above questions. Line layering with brushing is used to mitigate the visual effects of overlapping by allowing users' interaction to highlight only the curves under analysis or interested in. We also show the difference graph directly, which will help users to easily figure out the underlying factor causing the change of the curves. Not only can clustering decrease information density and overlapping, it can also be used to select the similar curves and find the pattern they share.

We use these visualization techniques to ease the analysis of time series for the UBC Fishery Department.

2. Related Work

There has been extensive work on visualization of time series. Many visualization techniques have been implemented.

Cleveland and McGill [3] analyzed the curve-difference chart after Playfair, where the vertical distances, or length between the two curves pictorially encoded England's balance of payments with the East Indies [4]. However, experiments showed that such length judgments could be inaccurate. People tend to evaluate the minimum distance between the curves rather than vertical distance [3]. Therefore, we should directly show the differences for curves. It is in the same situation in our project to evaluate the effect of fishing by showing the difference of fish population before and after a fishing strategy is applied. We need a difference graph to make the results more easy and direct to comprehend.

Baldonado, Woodruff, and Kuchinsky [5] considered the interaction mechanisms supported by multiple views. They proposed multiple view systems, where "two or more distinct views are used to support the investigation of a single conceptual entity" [5]. Using multiple views in information visualization offers several advantages: improving user performance, discovering unforeseen relationships and keeping all the views on the desktop updated. However, multiple views are more suitable when "there is a diversity of attributes, models, user profiles, levels of abstraction, or genres" [5]. Our project has only two views. Besides, the multiple view system is complicated to design. We choose another interaction technique linking [6], which connects data in two different views. Brushing [7] is one type of linking techniques, in which users highlight items in one view and the corresponding items in another view are highlighted by the system.

K-means clustering has the main advantage of reasonable space and time complexity [8]. In addition, the following steps of optimization can modify the improper clustering in previous steps. However, the experiments show that in some situations K-means phase takes really a long time to finish [9]. Besides, K-means assumes that clusters it is trying to find lie in a spherical Gaussian distribution [10].

3. Preliminary

Ecopath with Ecosim (EwE) is a free ecological/ecosystem modeling software suite developed by the UBC Fishery Department (source). [16]. One of its three components is Ecopath, a static, mass-balanced snapshot of the system. [16]. Ecopath uses a line chart system to illustrate the present and historical records of marine species population and biometrics. The Ecopath System can be used to explore the relationship between types of fish in a food network, analyze ecosystem effects of fishing on the marine population and evaluate the trends of a fish population and the similarity between different species.

Currently the fish population is presented in forms of a line chart with fish groups represented in different colors along the time axis. However, as the number of lines drawn on the screen increases, the chart suffers from the problem of visual information cluttering and overlapping, as shown in Figure 1. Thus, it becomes difficult to retrieve any comprehensive information from the line chart system.
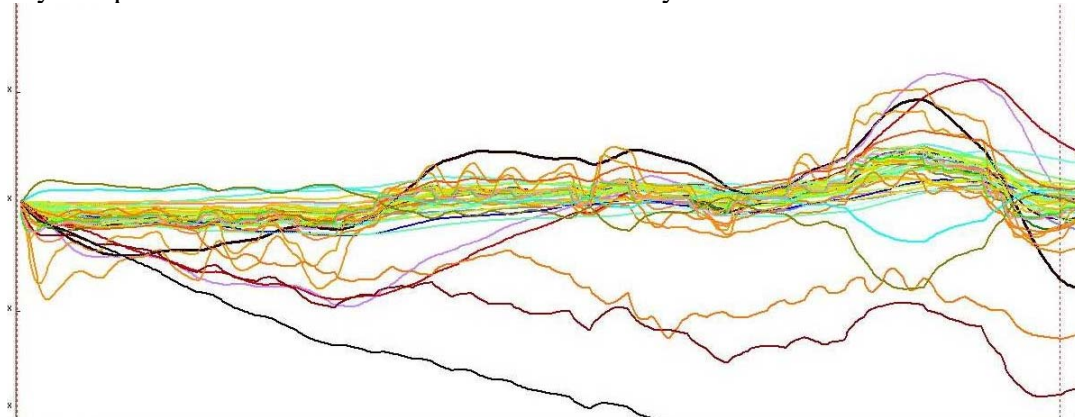

Figure 1: Original line chart from UBC Fishery Department

4. Description of Solution:

4.1 Data set:

We focus on the analysis of 45 fish species and their interconnection. We get the dataset from UBC Fishery Department as follow:

1. **The pre-fishing matrix** (45 by 648), which shows the discrete record of 45 fish populations at 648 time points before human's fishing strategy is applied.

2. **The post-fishing matrix** (45 by 648), which shows the discrete record of 45 fish populations at 648 time points after human's fishing strategy is applied under otherwise same conditions.

3. **The diet matrix** (45 by 45) of a marine food network, which describes how much each pair of fish consume each other among the 45 types of fish

4.2 Tasks:

We provide the visualization of information and trends from the following three aspects:

1. Find related marine species of a specific fish.
2. Evaluate the ecological effect of human fishing strate gy.
3. Discover the patterns that certain fish populations follow.

4.3 Information visualization techniques:

Fishery View consists of three information visualization techniques described as follows:

Visualization Solution 1: Linked Brushing/Layering view over a time series:

We first bind each fish with a position in a list box and a color. Fishery View allows a user to select one fish by brushing the mouse through the name of a fish species in the list box. The position is passed by the mouse, from which we will get which fish species the user has selected and draw its corresponding curve in red on the left panel. Then, by searching through the diet matrix, we can find its related fishes and plot their corresponding curves in their binding colors on the left panel. With these fish population curves drawn on the foreground, all the unrelated fish are faded to the background.

Visualization Solution 2: Difference graph over a time series:

With the pre-fishing matrix and the post-fishing matrix, we calculate the difference at each of the 648 time points for each fish population and then plot the difference in its corresponding color.

Visualization solution 3: Clustering of biomass groups over a time series:

Our aim is to merge the similar curves into clusters, such that the fish population patterns within one cluster are more similar than the fish population patterns in other clusters. Thus, a straightforward bottom-up clustering algorithm suffices.

Various distance measures are being used now. The Fishery Department recommended using the average geometric distance, or root-mean-square distance to describe how similar two fish population patterns are.

From the discrete records of N types of fish populations at M time points, we sample every other time point along the time axis and record its fish population. Therefore, we get an N by M/2 matrix.

We start with N clusters, and each cluster contains only one fish population pattern. Next, with the N by M/2 matrix and the root-mean-square distance measure we described above, we calculate the mutual difference between all clusters. We use an N by N matrix, or the clustering matrix, to record these results. Actually, we only use half of the matrix due to the interchangeability of distance. Then we find the most similar clusters by searching in the matrix and merge them into one cluster, followed by calculating and recording the average of the two fish populations. As a result, we have M-1 active clusters. The step of merging small clusters into larger clusters is repeated until the number of clusters is reduced to what we need, with each cluster containing the average of all fish populations falling into that cluster. In this way, we can control the number of clusters as required. To speed up the clustering procedure, we reuse the N by N matrix in each loop of clustering rather than construct a new matrix. What we need to do is only to update that matrix for new-clusters. The process of clustering is shown in Figure 2.

We also track the process of clustering to see which fish is classified into which clusters. This result is also visualized by using the same color to plot the curve representing that cluster and list the names of fishes classified into that cluster.
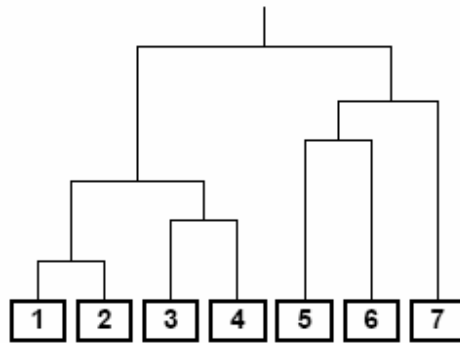

Figure 2: Clustering Algorithm

4.4 Visualization Details:

4.4.1 Information Density

"The empirical findings of Somervell [11] suggest that low information density visualizations are preferable to high information density visualizations in an 'attention-limited' environment where the visualization is not the primary focus of the user"[14]. With clustering, visual information density in the line graph system is reduced. This enables users to focus on the overall relationship between fish while utilizing a small information density. In this way, users can have a view of marine biomass groups and maintain their primary focus on the correlation between each fish population.

4.4.2 Color

Color highlighting assists viewers in perceiving interesting information and still keeps a consistent and connected view with the overall context [12]. The result of clustering requires one biomass group to be visually distinguishable from other groups. Thus, by assigning a specific color to a group of fishes according to the color of the cluster that it belongs to, we can draw the viewer's direct attention to its line graphs and its grouping in the list box. In addition, when fish are clustered they are assigned the same color for means of comparison.

4.4.3 Linked Highlighting and Layering/Brushing

The application of layering is to make relatively important information become obtrusive and differentiable from the rest of the context [15]. Both the line graph window and the list box window are tightly coupled, and any interaction initiated on one control is reflected in terms of visual changes of information in the other control. Linked brushing/layering of fish species by mouse hovering realizes the information of a specific marine biomass by highlighting.

4.5 High-level Implementation:

Fishery View was implemented using Visual Basic.Net. The framework is provided by our project consultant, Sherman Lai. The tool contains about 1000 lines of code, organized in several packages and classes. The user interface of the visualization tool was created using Visual Basic.Net. The visualization/drawing module was implemented using the graphics package, ZedGraph, provided by Visual C#.Net. The main modules are data preparation package, the drawing module of fish species, user interface components modules, and the interface manager module.

Distance Measures in Clustering algorithm:

As we described before, various distance measures are used. A simple measure is the average geometric distance, or root-mean-square distance.

Suppose there are two fish population patterns X_i and Y_i, $i = 1, 2, 3, \ldots$, M, representing the M time points. The simple root-mean-square-distance [13] is:

$$D_{rms} = \sqrt{\sum (Y_i - X_i)^2 / M}$$

When the two clusters are merged, their values should be averaged. It guarantees that the created clusters have the same value range as the original time series. The measure is robust and usually yields good results. The second distance measure is:

$$D_{nm} = \sqrt{\sum (Y_i / Y_{max} - X_i / X_{max})^2 / M}$$

Using this distance is equivalent to first normalizing each fish population pattern with its own maximum value and the use root-mean-square-distance. The clustering is based on the shape of the series. The third distance function is to first normalize the pattern using maximum and minimum values,

$$X_i = (X_i - X_{min}) / (X_{max} - X_{min}), Y_i = (Y_i - Y_{min}) / (Y_{max} - Y_{min})$$

It is used to create a better spread of the time series shape.

In our project, as recommended by the UBC Fishery Department, we first normalize all the time series using their values at 0 time point.

$$D_{n0} = \sqrt{\sum ((Y_i - Y_0) - (X_i - X_0))^2 / M}$$

Just to shift these time series and let them begin to spread from the value 0.

5. Results

5.1 Scenarios of use:

The following are the three cases when Fishery View tool can be useful.

Scenario 1: Finding related marine species of a specific fish

A common task when marine biologists view a line graph system of fish species is finding how a specific fish species is related to others. This task is performed when checking fish population over a certain period of time.

Using Fishery View, one can view the correlation of one marine species with others by brushing mouse through each of the fish names in the list box on the left and immediately see a group of related fish lines being highlighted on the right panel as shown in Figure 3. Any fish that is unrelated to the highlighted fishes is grayed out to the background.

By linking the list box with the line graph system panel, Fishery View provides researchers with direct relation with one species and its correlated fish populations.
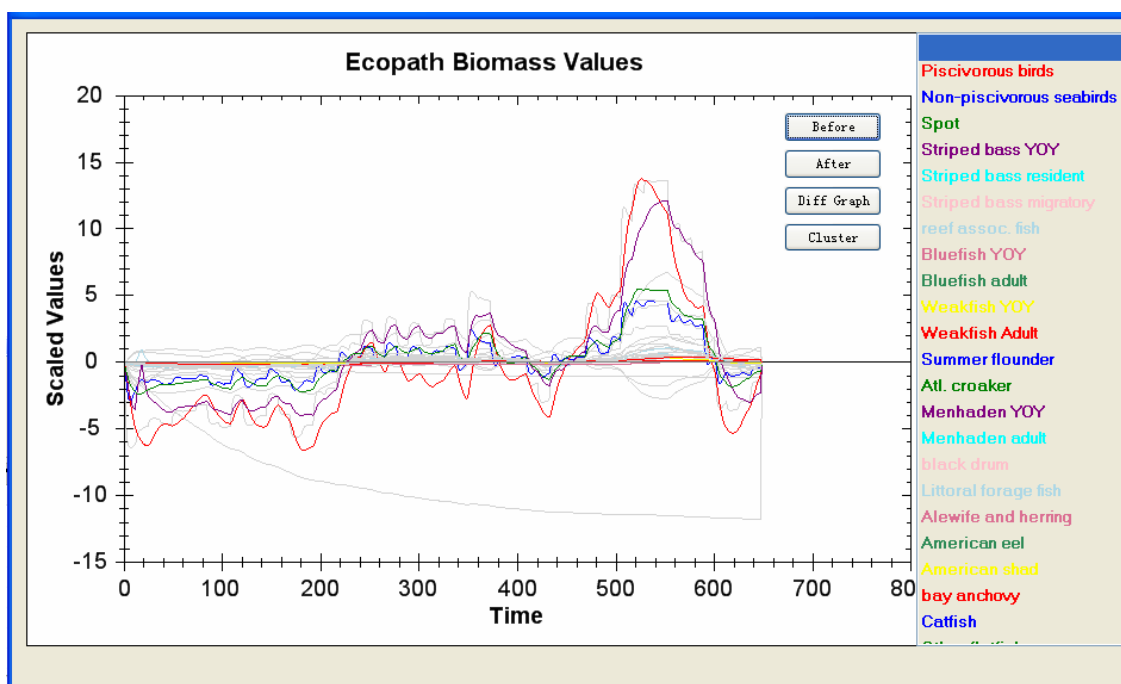

Figure 3: Finding related marine species of a specific fish

Scenario 2: Observing the impact of fishing strategy on fish population:
Estimating the impact of certain fishing strategies on marine species is another useful approach to observe the change of certain marine biomass over a time series. Fishery View allows researchers to import two sets of data indicating the population of a fish before and after a certain fishing strategy is applied, and then generate the difference graph showing the amount of change in the fish populations. Figure 4 and 5 show the fish population curves before and after the fishing strategy is applied. Figure 6 shows the difference graph.

Scenario 3: Clustering fish species sharing similar trends:
To discover the trends of marine biomasses residing in the same habitat area over a time period, Fishery View offers the feature of clustering as Figure 7 shows. This approach

allows researchers to aggregate species into one group and watch for the overall trend of the combination.
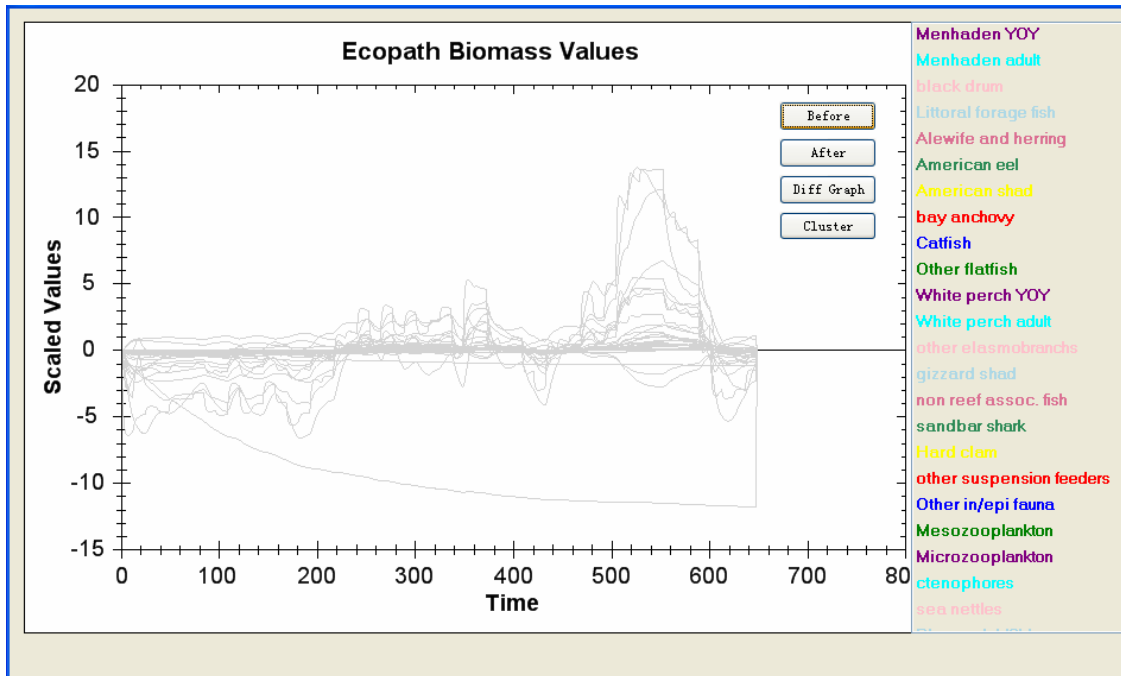

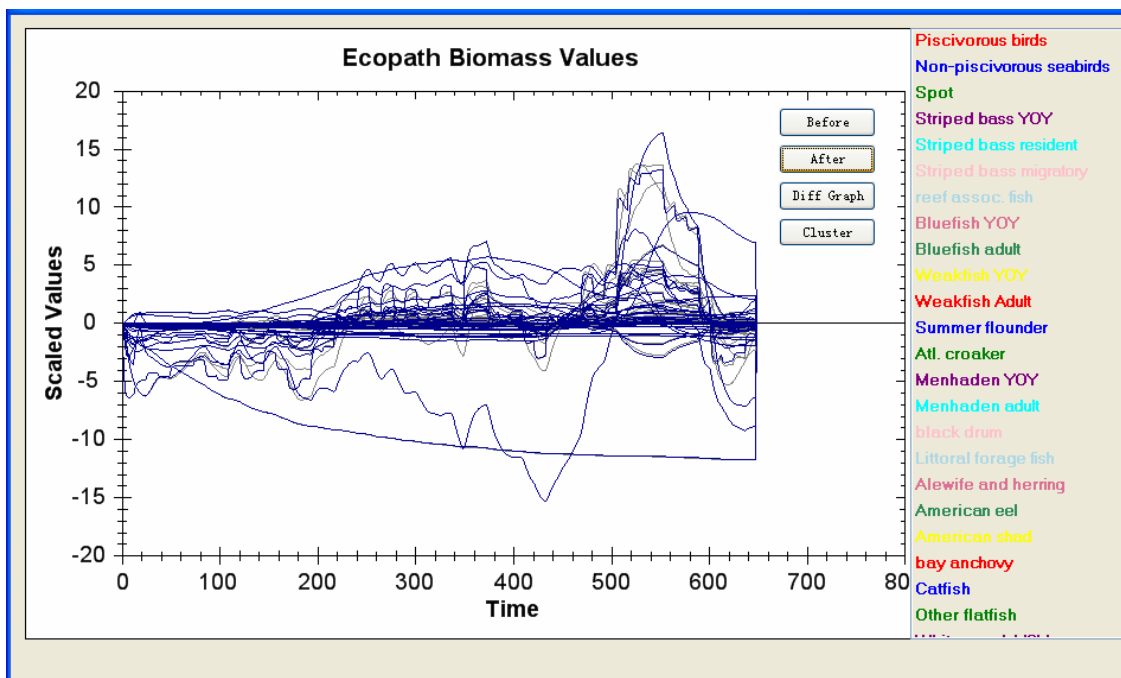**Figure 4: Fish population before fishing strategy is applied**

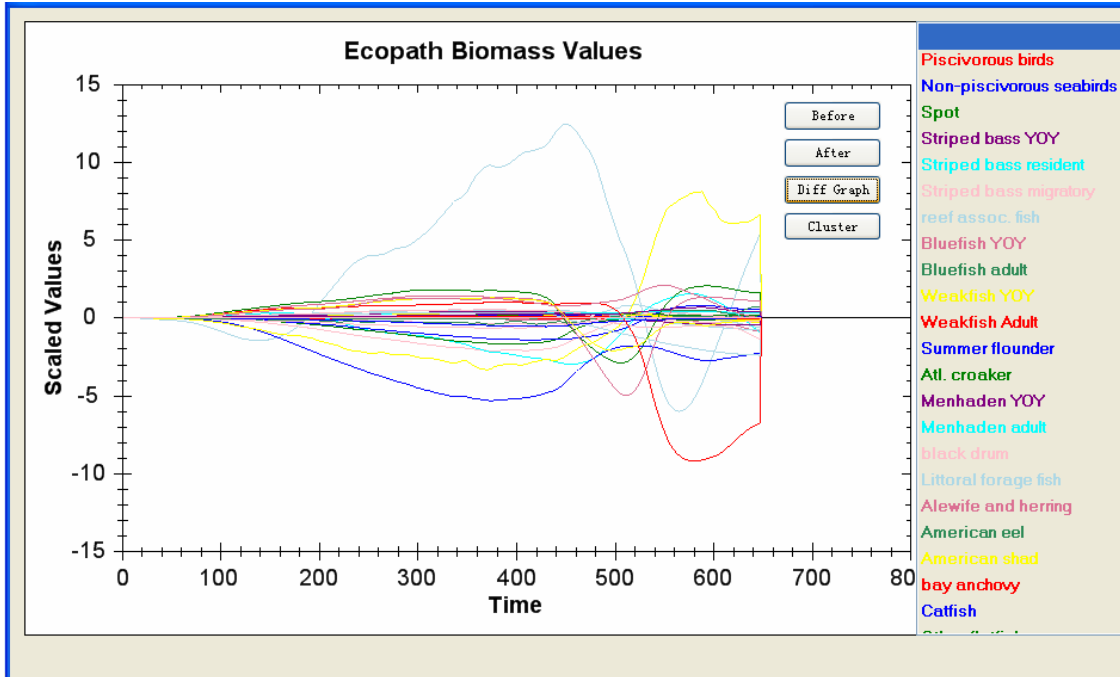**Figure 5: Fish population after fishing strategy is applied**

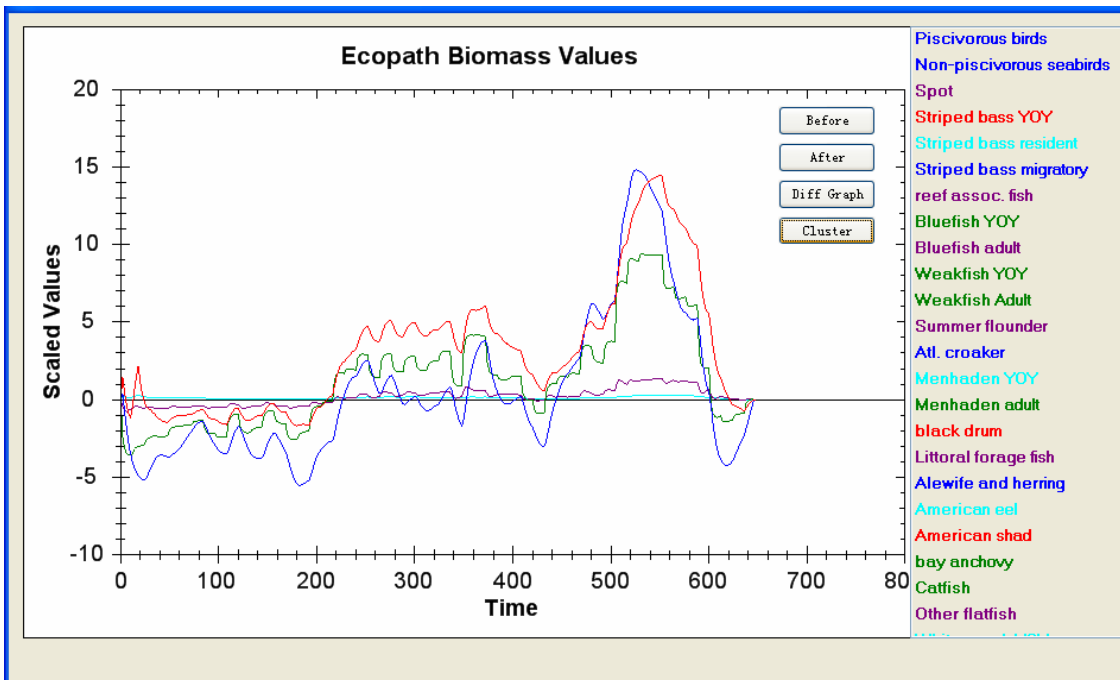**Figure 6: The Difference graph between the fish population before and after fishing strategy is applied**

Figure 7: Clustering ended up with five groups

5.2 Evaluation:

Fishery View is still in a partially implemented stage, although most intended features are working. With brushing and layering, exploration of fish species and their inter

correlation in a marine eco-system can be performed effectively. Assisted by the difference graph, users can directly observe the degree of impact that a certain fishing strategy would cause on marine biomasses and consequently be able to suggest or plan any necessary intervention to preserve or protect the balance of a marine system. After clustering fish population lines based on the shape of the underlying time series, fish population patterns are merged into five groups; users can then effectively see the patterns of biomass groups in a collected view.

6. Discussion

6.1 Future Work

Handling component interaction:
Due to time constrains and lack of knowledge of the development platform, the tool has not been optimized to enable further interactions between the line graph panel and the list box panel. At the moment, the interaction between the panels is synchronized by a timer.

Repeated colors:
Fishery View is supposed to use 45 different colors to represent the 45 fish populations. However, we proposed two coloring strategies, but both failed in the implementation stage. One way is using an existing color package with 500 colors; however, due to the limitation of human's vision system, humans' eyes can hardly detect the tiny intensity differences between two colors sharing the same hue. The other way is to iterate through a set of static system colors with obvious hue difference, but the number of colors is limited to ten. The problem is we have to use them repeatedly, which will result in clashes in color selection. After experimenting, we consider the second approach is better though it is far from perfect. Maybe we need a third method.

Overlapping can be reduced but not eliminated:
When a user moves the mouse to brush through a specific fish, the related fish species will be highlighted at the foreground with the colors we assign them at the beginning, and the unrelated fishes will grayed out to the background. However, due to the drawing functions evoked from the package of Zedgraph, all the curves will be redrawn in the predefined order, so we cannot guarantee the related fishes are in the foreground. These fish lines may sometimes be occluded by the grayed-out curves, which results in unsatisfactory results.

6.2 Lessons learned

Building an application based on any existing framework does speed up the progress of implementation, but can also limit creativity and raise frustrations. When reusing frameworks and codes of others, there is a learning curve of understanding the purpose and usage of every variable and function before the continuation of implementation can take place.

Some precautions have to be taken when using any existing software package, because it may raise frustrations. For example, in our project, we directly used the draw function in graphics package Zedgraph. It can only draw in a predefined order and limit our codes.

7. Conclusion

This paper presents Fishery View, an interactive visualization tool for displaying and interacting with fish population data. The tool uses key information visualization techniques to support effective visualization of data. With the interaction mechanism, brushing, Fishery View allows users to find the related fishes by simply moving the mouse over the fish of interest. The tool directly shows difference graph to visualize the effects of human's fishing strategy. Clustering fishes into five groups is the proper way to explore the similarity of these fish population pattern.

8. Acknowledgement

We should acknowledge Dr. Tamara Munzner. Dr. Munzner provided us with suggestions on how to improve the original graph and the reference of implementing clustering. We also thank Sherman Lai for providing the framework and his help on answering questions we have about that. He also gave us the good suggestion to use Zedgraph.

9. Reference

[1] Lior Berry, Tamara Munzner, *Binx: Dynamic Exploration of Time Series Datasets Across Aggregation Levels*, Infovis 2004

[2] M. Wolfgang, H. Schumann, *Visualization Methods for Time-Dependent Data: An Overview*, WSC 2003

[3] William S. Cleveland, Robert McGill, *Graphical Perception: Theory, Experimentation and the Application to the Development of Graphical Models* J. Am. Stat. Assoc. 79:387, pp. 531-554, 1984.

[4] Monkhouse, F.J., and Wilkinson, H.R., *Maps and Diagrams*, London: Methuen

[5] M. Q. Wang Baldonado, A. Woodruff, A. Kuchinsky, *Guidelines for Using Multiple Views in Information Visualization*, Proceedings of AVI 2000, Palermo, Italy, May 2000, pp. 110-119.

[6] Velleman, P. *The DataDesk Handbook*. Odesta Corporation, Ithaca, NY, 1988.

[7] Becker, R.A., and Cleveland, W. S. "*Brushing Scatterplots*." Technometrics, 1987, 29(2):127-142.

[8] Jain, A. K., M. N. Murty, P. J. Flynn, *Data clustering: A review*, ACM Computing Surveys, Vol. 31, No. 3, September 1999

[9] Alistair Morrison, Greg Ross, Matthew Chalmers, *Fast Multidimensional Scaling through Sampling*, Springs and Interpolation, Information Visualization 2(1) March 2003, pp. 68-77.

[10] Bradley, P. S., U. M. Fayyad, "*Refining Initial Points for K-Means Clustering*", Proceedings of the Fifteenth International Conf. on Machine Learning, (1998).

[11] Somervell, J., McCrickard, D. S., North, C., and Shukla M. *An Evaluation of Information Visualization in Attention-Limited Environments* in Proceedings of the IEEE TCVG Symposium of Visualization (VisSym '02), Barcelona, Spain, May 2002

[12] Bernice E. Rogowitz and Lloyd A. Treinish, *How Not to Lie with Visualization*, Computers In Physics 10(3) May/June 1996, pp 268-273.

[13] Jarke J. van Wijk Edward R. van Selow, *Cluster and Calendar based Visualization of Time Series Data*, Proc InfoVis 99.

[14] Bodnar, Adam and Nekrasovski, Dmitry, *LabVis: Simple and Intuitive Visualization of Medical Laboratory Results*, Infovis 2004

[15] Tufte, Edward R., Envisioning Information, Graphics Press LLC, Cheshire, Connecticut, U.S.A. 2003, pp 53.

[16] Ecopath welcome page, http://www.ecopath.org/