Using Interactive Visualization to Enhance Understanding of a Fisheries Model

BY

Carmen St. Jean

B.S., University of New Hampshire (2010)

THESIS PROPOSAL

Submitted to the University of New Hampshire
in partial fulfillment of
the requirements for the degree of

Master of Science

in

Computer Science

May 2014

This thesis proposal has been examined and approved.

Thesis Proposal director, Colin Ware

Professor of Computer Science

R. Daniel Bergeron

Professor of Computer Science

Matt Plumlee

Affiliate Assistant Professor in Computer Science

Date

Table of Contents

List of Figures

Chapter 1

Introduction

Fishery managers have only one "lever" to pull when it comes to fishery management: the ability to set harvest quotas. Fishermen work within these quotas by exerting various levels of fishing effort. Both managers and fisherman can better understand the ecosystem they work within and the implications of their decisions with the assistance of a production model. In this context, a production model is a mathematical model, based on data, that simulates interactions between species and predicts species biomass as a function of fishing effort, climate change, and other variables. MS-PROD is a multispecies production model developed by NOAA scientists Gamble and Link [8]. A visualization may enhance such a model by making its inner workings more explicit and may be useful for decision making. The goal of the proposed research is to explore design alternatives and evaluate the effectiveness of different modes of portrayal and interaction to make a visualization of the MS-PROD model that will be a valuable tool to the modelers and stakeholders alike.

Chapter 2

Background

2.1 Models

Both the short- and long-term effects of human exploitation on an ocean ecosystem, such as the species inhabiting the Gulf of Maine, are not easily understood since experiments which would allow ecosystem managers to investigate the impact of different levels of exploitation over many years are either impractical or impossible to conduct on a large scale. Fortunately, ecosystem models can be used instead to help gain a better understanding of an ecosystem.

2.1.1 Lotka-Volterra Equations

Ecosystem models are abstract representations of an ecological system, and can range from an individual species in its environment to an entire community of species. A classic example is the Lotka-Volterra model, which is a pair of differential equations for describing the non-linear interactions between a predator species and a prey species. [17, 24]:

$$\frac{dN_1}{dt} = N_1 (\alpha - \beta N_2) \tag{2.1}$$

$$\frac{dN_2}{dt} = -N_2 (\gamma - \delta N_1) \tag{2.2}$$

where N_1 is the number of prey, N_2 is the number of predator, t is time, α is the prey's growth rate, β is the rate at which the predator destroys the prey, γ is the death rate of the predator, and δ is the rate at which the predator increases from consuming the prey. The model can be generalized to discuss an arbitrary number of species rather than just a single pair.

The Lotka-Volterra model can be modified to take competition instead of predation into account, as in the Rosenzweig-MacArthur model [20] and the Leslie-Gower model [16]. These adaptations to the model also consider carrying capacity, which is the maximum number of a species that can be sustained indefinitely in a particular environment:

$$\frac{dN_1}{dt} = r_1 N_1 \left(1 - \left(\frac{N_1 + \alpha_{12} N_2}{K_1}\right)\right) \tag{2.3}$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(1 - \left(\frac{N_2 + \alpha_{21} N_1}{K_2}\right)\right) \tag{2.4}$$

where r_i is the growth rate for species i, K_i is the carrying capacity for species i, and α_{ij} is the effect species j has on species i. As with Lotka-Volterra, this model concerns only two species, but it can be generalized to include more than two.

Both Lotka-Volterra and Leslie-Gower do not incorporate a factor which is critical when discussing fisheries management: the effect of harvest. The Schaefer model adds a term to account for the effect of harvest on an individual species [22]:

$$\frac{dN}{dt} = rN \left(1 - \left(\frac{N}{K}\right)\right) - qEN \tag{2.5}$$

where N is the number (or biomass) of the species, r is the growth rate, K is the carrying capacity, q is the catchability coefficient, and E is the fishing effort.

Simple models, when available and correct, are generally preferred; since fewer components are needed to describe their real-world counterparts, they are more easily understood and implemented. All three of these models are subjectively simple in that they only consider a few ecological factors each. However, in reality, ecosystems are complex systems which require management that recognizes them as such [3]. Thus, a more holistic approach called ecosystem-based fishery management (EBFM) has been advocated [18]. However, this approach has not often been implemented due to a lack of models which consider all necessary ecological factors. Gamble and Link developed a multispecies production model (MS-PROD) to fill this gap [8].

2.1.2 The MS-PROD Model

The MS-PROD model forecasts biomass for species separated into functional groups, which are biological groupings of species that perform similar functions within their ecosystem. The model is built upon the Schaefer production model and also includes Lotka-Volterra terms for predation, Leslie-Gower terms for competition, and carrying capacities for functional groups (K_G) as well as for the entire system (K_σ):

$$\frac{dN_i}{dt} = r_i N_i \left(1 - \frac{N_i}{K_G} - \frac{\sum_{j=1}^{g} \beta_{ij} N_j}{K_G} - \frac{\sum_{j=1}^{G} \beta_{ij} N_j}{K_\sigma - K_G} \right) - N_i \sum_{j=1}^{P} \alpha_{ij} N_j - H_i N_i \tag{2.6}$$

where N_i is the number (or biomass) of species i, t is a unit of time, r_i is growth rate for species i, β_{ij} is the interaction of species j on i, α_{ij} is the predation of species j on i, H_i is the harvest rate on species i, g is the number of species within species i's group, G is the number of groups, and P is the number of predators.

This model is distinguished from other multispecies production models by describing stocks with explicit ecological and harvest factors. Each species to be included in the simulation must be specified in the parameter file by listing growth rate, functional group membership, initial biomass, carrying capacity, and catchability. Additionally, matrices representing inter-species relationships must be provided to describe the relationship between every pair of species. Such matrices are required for *predation*, where one species consumes the other, and *interaction*, where one species affects the other in any manner besides predation.

The MS-PROD authors provided us with a parameter file which listed ten key species chosen from the Northeast United States Continental Shelf Large Marine Ecosystem (NEUS LME), listed here by functional group:

- **Elasmobranchs:** Skates, Spiny Dogfish
- **Flatfish:** Windowpane, Winter Flounder, Yellowfin Tuna

- **Groundfish:** Cod, Haddock, Redfish
- **Small Pelagics:** Herring, Mackerel

The MS-PROD model runs simulations for 30 years with an annual time step to predict individual biomasses. While this outputted information is potentially valuable to fishery managers, it was lacking an interactive graphical user interface.

2.2 Visualization Methods

When designing a visualization, there can be several techniques for visualizing a feature that are worth comparing. This section reviews existing research of visualization methods which have particular relevance to the problem of creating an effective interface to a fisheries ecosystem model.

2.2.1 Time Series

Fisheries management is focused on the sustainability of choices concerning fish stocks. A main purpose of ecosystem management is to ensure that future generations can enjoy the same natural resources [3]. As such, the MS-PROD model provides biomass forecasts for 30 years. Therefore, time-oriented visualization techniques must be explored.

Frank categorized time-oriented data as follows [6]:

1. Linear vs. cyclic
2. Time points vs. time intervals
3. Ordered time vs. branching time vs. time with multiple perspectives

Aigner et al. point out that techniques for visualizing branching time and time with multiple perspectives are unfortunately uncommon [1], but a line chart—a technique frequently used for ordered time data—serves as a good starting point.

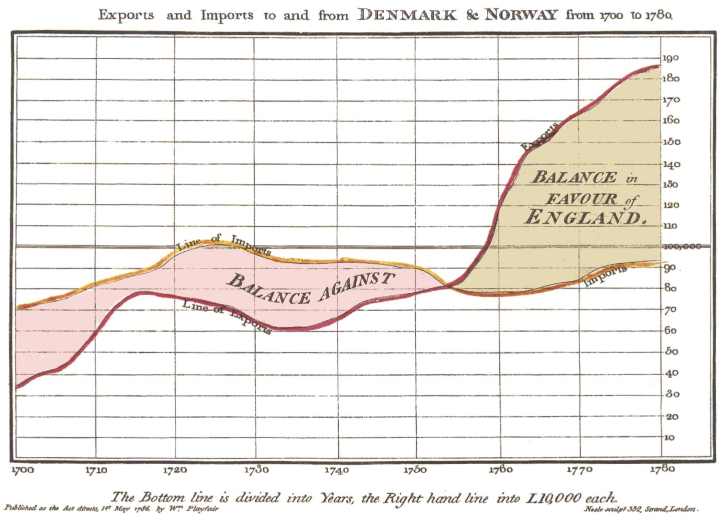
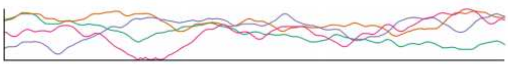
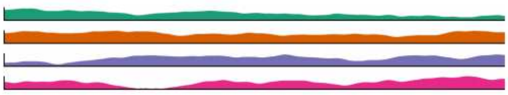


Figure 2-1: Playfair's original time series chart [19].

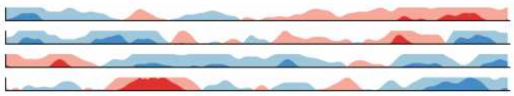


(a) A simple line graph.

(b) A braided graph.



(c) Small multiples.



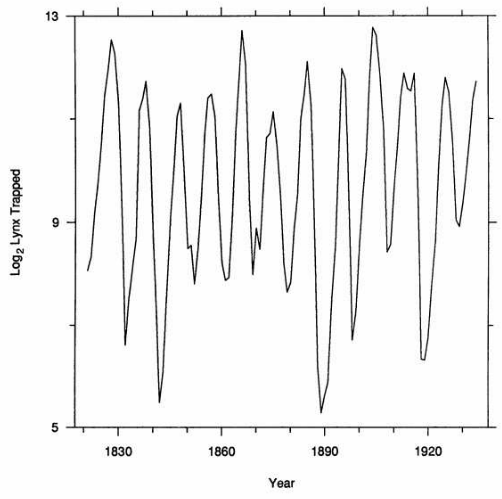
(d) Horizon graphs.

Figure 2-2: Four possible methods for visualizing multiple time series [14].

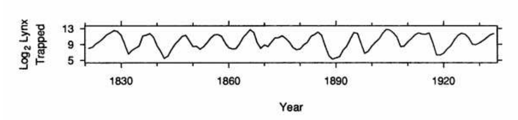
Line Charts

The line chart was first invented by William Playfair in 1786 to communicate time series data, seen in Figure 2-1 [19]. Today, it remains a common method for visualizing time-oriented data in many fields, including science, economics, planning, and engineering to name a few. Line charts typically encode time on the horizontal axis, progressing from left to right, and some time-varying value on the vertical axis. Points in the chart are connected by line segments such that the slope of the line indicates the rate of change between time steps.

Multiple time series can be part of a single line chart; each series needs only to be



(a) Shape parameter of 1.0.

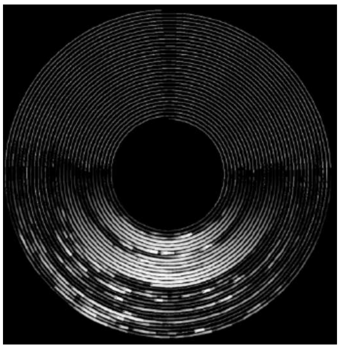


(b) Shape parameter of 0.074.

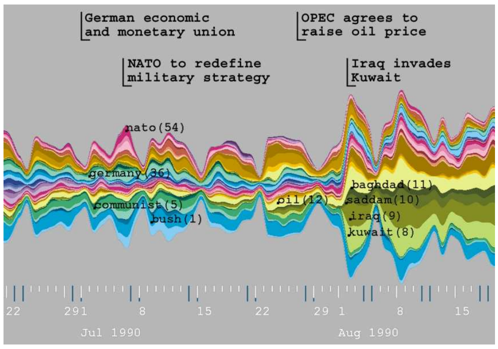
Figure 2-3: Two effect of chart shape on Canadian lynx data [4].

distinguished by a color and/or line style. However, as the number of time series on a single line chart increases, it becomes more difficult to identify an individual series. Javed et al. evaluated the four different plotting techniques for multiple time series illustrated in Figure 2-2 [14]. The first of the techniques is the "simple line chart," which was Playfair's original line chart with all series plotted together. A slight variation on that is "small multiples," where each series had its own line chart though all charts share the same axis scales. Horizon graphs, originally developed by Saito et al., wrap around a baseline in two color tones to save space [21]. Lastly, braided graphs feature all series on one chart with the coloring under the curves alternating as series intersect each other. The user evaluation by Javed et al. revealed that a simple line graph with all time series on one plot or a single graph for each time series is better suited to a variety of tasks than a horizon graph or a braided graph. They also found that users complete tasks more correctly when there is more display space allocated to the graphs. They did not recommend using a higher number of simultaneous time series—their study used eight at the most—because it also leads to a decline in correctness of task completion.

Some line charts are more effective at conveying the nature of the data than others



(a) Spiral time series.



(b) ThemeRiver time series.

Figure 2-4: Two alternative time series visualizations.

because of the way different drawing techniques affect interpretability. Cleveland et al. found the shape of a line chart—defined as the height of the chart divided by the width of the chart—to be a critical factor [4]. Shape of the chart directly impacts the slopes of line segments, which viewers interpret in order to understand the dependence of the y variable on the x variable. Figure 2-3a, a time series of Canadian lynx trapping data, features a shape of 1.0 and seems to imply rapid increases and decreases in the population. On the other hand, Figure 2-3b has a shape of 0.074 and shows more clearly that the population rises somewhat steadily and declines somewhat rapidly, which Figure 2-3a failed to show. Their user evaluation found that judgment of two slopes is influenced by the orientation mid-angle, defined as the average of the minimum slope orientation and the maximum slope orientation. They proposed line chart shape should be selected such that orientations are as close to $\pm 45°$ as is possible, like in Figure 2-3b.

Alternatives

There are many alternatives to and variations of Playfair's original time series. One example is Weber et al.'s spiral time series, seen in Figure 2-4a [26]. The spiral time series was designed for cyclic data. Cycles are emphasized in a properly-parameterized spiral visualization, however it may be difficult to describe periodic behavior in unknown datasets or determine if that behavior even exists. Another example is the ThemeRiver by Havre et al, seen in Figure 2-4b [11]. Each "current" in the ThemeRiver represents an entity or subject and must be of a distinctive color. Positioning along the y-axis is meaningless, instead the abundance of the entity or subject over time is indicated by the width of the current.

2.2.2 Networks

The input parameters to the MS-PROD model includes predation and interaction matrices. These are relationships which may be better understood if incorporated into the visualization. Relationships are often visualized through a node-link diagram, which typically represents entities as nodes and links as relationships between the nodes they connect. There are many types of node-link diagrams used for illustrating networks; those which are relevant to our research are discussed in the following sections.

Force-Directed Layouts

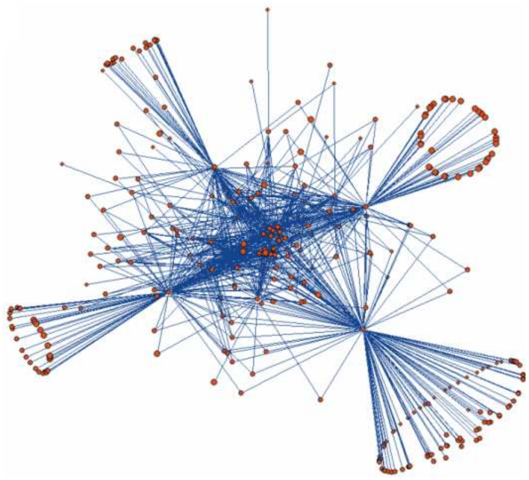


Figure 2-5: A force-directed visualization of a food web of Gulf of Alaska data [7].

One option for showing fish species interactions would be to use a force-directed layout as Gaichas and Francis did, seen in Figure 2-5 [7]. Here, the nodes represent an individual species in the Gulf of Alaska, while the links represent a predator-prey interaction. In a force-directed layout, nodes repel each other, while related nodes become pulled toward each other by links [12]. The result is an aesthetically pleasing layout where there are relatively few link crossings and links are of approximately similar length. The color of the node can be used to indicate group membership, while the size can represent the magnitude of some property of the node. Likewise, the drawing style of the link can be varied to encode different types of relationships.

Arc Diagrams

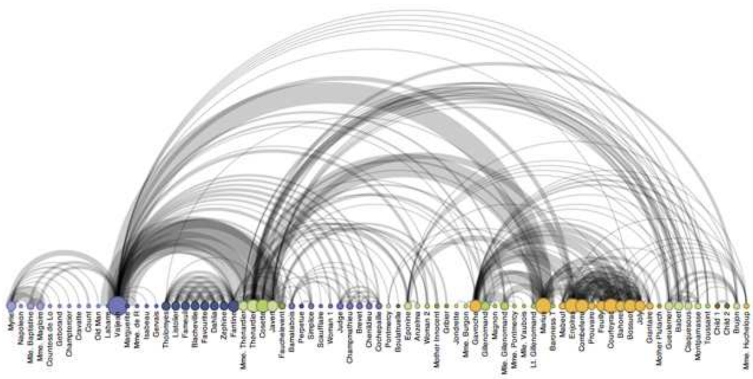


Figure 2-6: Knuth's arc diagram of *Les Misérables* characters [15].

An alternative for force-directed layout is an arc diagram. The name arc diagram was coined by Wattenberg [25], though they were invented earlier. Knuth used arc diagrams to illustrate interaction of characters in Victor Hugo's novel *Les Misérables*, seen in Figure 2-6 [15]. Each character is represented with a circular node, where size indicates the number of appearances. The nodes are arranged linearly, colored and ordered according to clusters of characters that appear together frequently. Semi-transparent arcs are drawn between the characters which appear in the same chapter, with the thickness of the arc representing the number of such appearances. While the arc diagram may fail to properly depict the structure of a network, Heer et al. point out it is advantageous because the one-dimensionality allows for other features to be easily displayed near the nodes [12], such as text labels.

Directed Edges

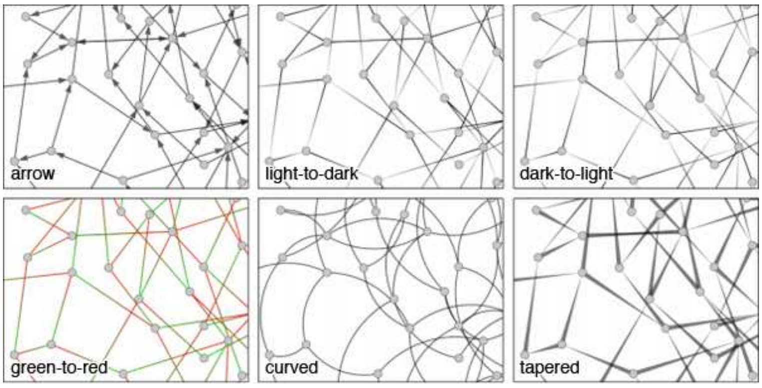


Figure 2-7: Different types of directed edges [13].

Relationships in a network may be directional, such as the predator-prey relationship. In a visualization of such a network, the direction of the edges must be encoded so these relationships can be understood. Holten and van Wijk studied the effectiveness of different techniques for indicating directionality of edges in a graph, seen in Figure 2-7 [13]. The traditional arrowhead was found to perform poorly, while tapered edges performed best. As for an intensity-based direction cue, a dark-to-light representation was found to be clearer than light-to-dark.

Matrix Representations

Node-link diagrams can have occlusion problems when they are highly-connected, so a matrix-based representation of a network is a possible alternative [12]. In many cases, networks are stored as an adjacency matrix, so all that needs to be done is visualize that matrix as a grid, where the cell at the ith row and the jth column represents the relationship from entity i to entity j. Figure 2-8 shows Knuth's visualization of *Misérables* characters in matrix-form [15]. The color of the cell indicates the presence or type of a relationship, with some neutral color indicating the lack of a relationship. Ghoniem et al. showed that a matrix-based view is suitable for large or dense networks for tasks that involve finding or

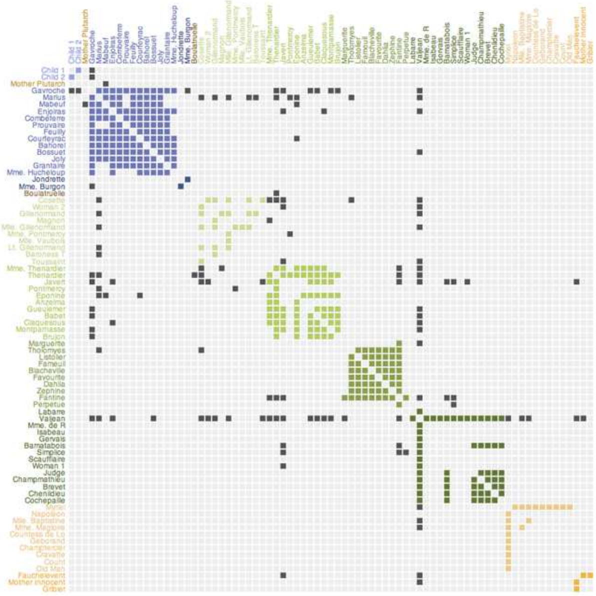


Figure 2-8: A matrix-based visualization of an adjacency matrix [15].

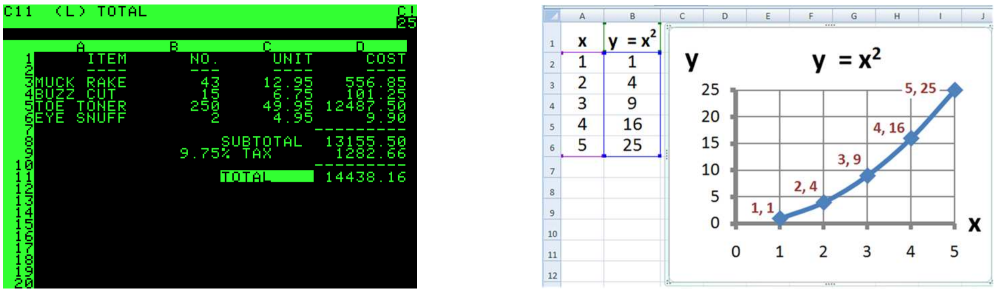
counting links or nodes [9]. With proper ordering of the rows and columns, the structure of the network can be effectively displayed, however path-finding tasks may be difficult.

2.3 Understanding Models

Users studying models can benefit from the aid of a visualization, because patterns and trends may be difficult—if not impossible—to discern from only a table of numerical values. The learning process can be even further enhanced through interaction with the model. If interactivity is supported by the visualization, then users can adjust parameter values, perceive a change (or perhaps no change) in the results, and begin to understand the degree of influence different parameters possess.

2.3.1 Spreadsheet Programs

VisiCalc was a very early example of software assisting the understanding of models [10]. As a business student, Bricklin wished there was a faster way to change the input or fix mistakes when working out financial models by hand [2]. To address this, he worked with Frankston



(a) A screenshot of VisiCalc (GNU General Public License).

(b) A chart made using Microsoft Excel (public domain).

Figure 2-9: Two examples of spreadsheet applications.

to develop VisiCalc, seen in Figure 2-9a. As the world's first electronic spreadsheet, VisiCalc consisted of rows and columns containing either text, numerical values, or formulas. Result cells were instantly updated according to changed inputs or adjusted formulas, allowing a user to work with models in a more efficient and dynamic manner. VisiCalc was superseded by Lotus 1-2-3, which was in turn supplanted by Microsoft Excel. Microsoft Excel remains popular and features graphing tools which can generate charts, such as in Figure 2-9b.

2.3.2 The Influence Explorer

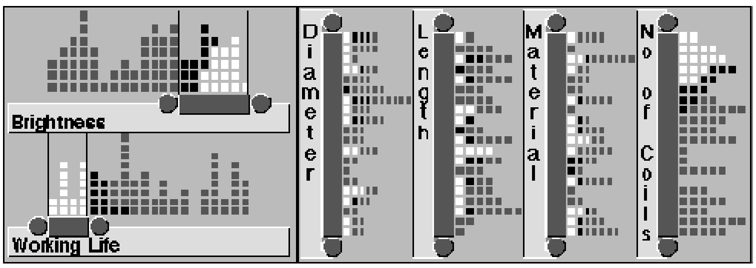


Figure 2-10: A screenshot of the Influence Explorer [23].

The Influence Explorer by Tweedie et al. is a good example of a more complex interactive visualization [23]. They developed an interface for understanding the relationships between different attributes in a design process. Parameter values of the Influence Explorer are initially randomly selected to represent different possible items. For each attribute, there is a histogram including each of the items. The attribute ranges are controlled by sliders. When the user adjusts the slider of a given attribute, all items that are within that range are highlighted on all of the histograms. Figure 2-10 is a screenshot of the Influence Explorer being used to test the performance of different light bulb designs; white indicates the design passed, black it failed one specification, and grey it failed two specifications. In a user evaluation, industrial designers found the ability to interactively explore the effects of different parameter ranges to be valuable.

2.3.3 Vensim

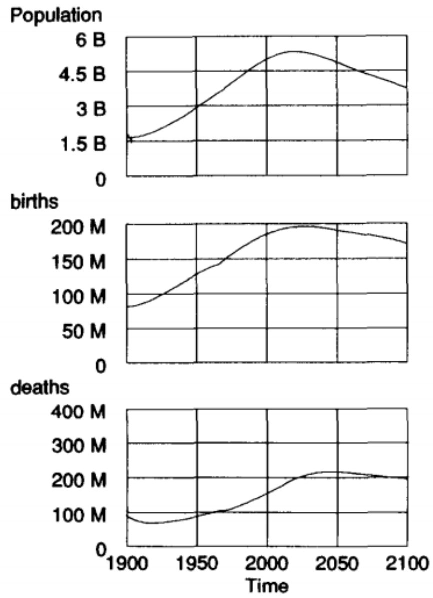


Figure 2-11: A screenshot of the strip graphs in Vensim [5].

Eberlein and Peterson recognized that both unskilled and skilled modelers have the same need: to quickly obtain a thorough understanding of a model and its implications [5]. This motivated their development of Vensim, which is a commercial tool for visualizing and analyzing simulation results. Vensim allows users to run a model under different conditions with a simple mouse click, enabling the user to learn the effects of different actions with ease. Various features enhance this learning process—e.g., "causal tracing" strip graphs, shown in Figure 2-11. Rather than simply seeing a chart of the projected population, a user exploring the causal tracing feature can begin to understand the various components that contributed to the population simulations—births and deaths—by seeing each component on its own chart. A major emphasis of Eberlein and Peterson is that the same visualization tool can and should be used for both the development and teaching phases of a model, especially since those phases may not be discrete.

Chapter 3

Proposed Work

The goal of this research is to provide an interactive interface to the NOAA MS-PROD model so as to help fishermen, fisheries managers, and other stakeholders understand the implications of decisions, such as changing catch quotas for particular kinds of fishing activity—e.g., bottom trawling versus mid-water trawling. From using our interactive interface, users should gain insights into:

- **Implications of the model:** E.g., how do two different sets of fishing effort values affect the biomass predictions of the ten species?
- **The model itself:** E.g., why does the abundance of one species increase when another species is caught?

Research for this thesis is being carried out in two phases, 1) a design and implementation phase and 2) an evaluation phase. In the first phase, various design alternatives have been implemented with some feedback from the originators of the fisheries model (Robert Gamble in particular). A substantial part of the research, particularly relating to design and implementation, has already been completed; most of what remains is the evaluation component. The following section discusses completed work and the section following that discusses work that remains to be done.

Chapter 4

Work Completed

The bulk of the design and implementation of an interactive interface to the MS-PROD model has been completed. However, this is not a final design; in many cases, several design alternatives have been implemented and await the evaluation study prior to final implementation. These alternatives are described below.

Time series line charts—of both the simple line chart and small multiples varieties—were chosen to display the biomass forecast data output by the model. A major reason line charts were chosen is that casual viewers can understand them without further instructions, as opposed to, say a horizon graph. Another advantage to line charts is that they tend to have a reasonable amount of whitespace where additional information can be displayed, such as uncertainty or alternate forecasts. A key design issue concerned the representation of change; line charts provide the ability to display this change. A number of alternatives for change were implemented and are described in the sections that follow.

Interaction with the model parameters is done by means of a set of sliders with the goal of allowing the user to immediately see the impact of management decisions on fisheries. The user can adjust sliders which represent harvest effort and watch the line charts change instantaneously as the model is re-run according to the new effort values. Like Eberlein and Peterson, we aimed to turn the "time consuming and tedious" task of working with a model into a "fast and fun" interactive experience [5]. Different views and features of the model visualization are described and discussed in the sections below.

4.1 Alternative Screen Layouts

Two alternative screen layouts were made for displaying these time series on line charts: a "four panel" view in Figure 4-1 and a "small multiple" view in Figure 4-2, described in further detail below. With both views, the time series for the species are organized by functional group. A functional group is a biological grouping of species which perform similar functions within their ecosystem—e.g., mackerel and herring are both members of the "small pelagic" group since they live in the water column

There were several reasons for the decision to arrange by functional group rather than placing all time series on a single line chart. First, harvest effort is controlled by functional group, so it must be somehow indicated which species are part of which functional group. With multiple time series, this can be encoded through positioning by arranging the slider of a functional group to be near the time series of that group's species. Second, it would have been impractical to simply place all time series on a single line chart because would have been difficult to select enough distinct colors to represent each species. Furthermore, the biomass of some species is significantly larger than others. By scaling the y-axis according to the largest biomass value of all species in the entire 30-year time span, the lines representing some species would have been crowded at the bottom of the chart and seemed to be flat even when they were not.

4.1.1 Four Panel View

It was observed that species of similar functional groups tend to have biomass values in similar numeric ranges, so a decision was made to make one line chart per functional group for displaying biomasses. This view is called the "four panel" view and is shown in Figure 4-1.

The major advantage of this "four panel" view is that comparisons of species within a group are easy. For MS-PROD, there are only two or three species per functional group, so the line chart for each group tends to not suffer from occlusion problems. Direct and indirect effects of changes in harvest effort are easily differentiated with the "four panel"

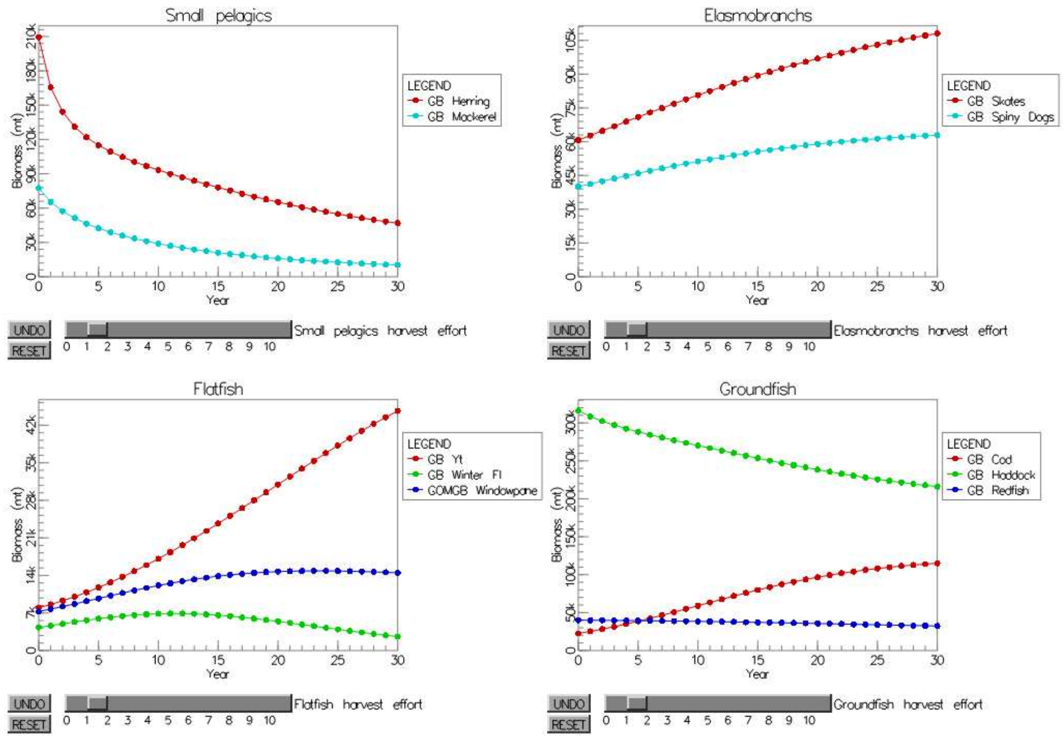


Figure 4-1: The “four panel” view of our MS-PROD visualization.

approach—e.g., if the user adjusts the effort slider only for elasmobranchs, yet sees the biomasses change on the groundfish chart, then the user can begin to understand there is some kind of relationship between elasmobranchs and groundfish.

4.1.2 Small Multiples View

The alternative to the “four panel” view is to view each species on its own plot, which we call the “small multiple” view. The main purpose of this view was to support the addition of arc graph connections between species, as are discussed later in Section 4.3. Shown in Figure 4-2, each plot is sorted and colored according to functional group membership, using Cleveland’s recommendations for chart size [4]. The harvest effort sliders are also colored by functional group and positioned near the plots of the corresponding group. This allows for the ability to differentiate between direct and indirect effects of changes in harvest effort.

With each species on its own plot, it is much easier to interpret the biomass predictions of an individual species, since the y-dimension of a plot needs to be scaled to the data

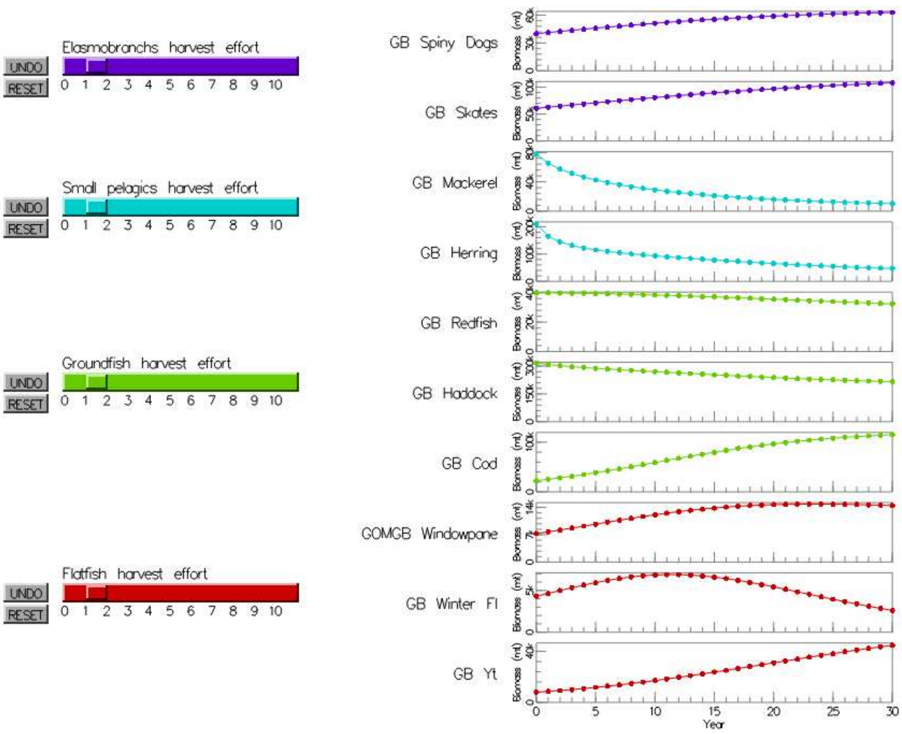


Figure 4-2: The "small multiples" view of our MS-PROD visualization.

of one species only. It is easier to perceive increases or decreases in biomass because no species suffer from the "flattening" that can occur when a series is displayed on the same plot as a series that has significantly higher values. On the other hand, this makes biomass comparison between species somewhat difficult because the user must either refer to the y-axis labels or hover over a specific point on a chart, which causes a label to appear, in order to determine the absolute value of the biomass at a point in time.

Absolute Biomass Indicators

Absolute biomass indicators, seen in Figure 4-3, were introduced to show how biomass changes over time by showing the absolute biomass of the population as the area of a circle. This makes comparison across species possible. To avoid occlusion, these indicators are drawn every five years within the thirty-year time span.

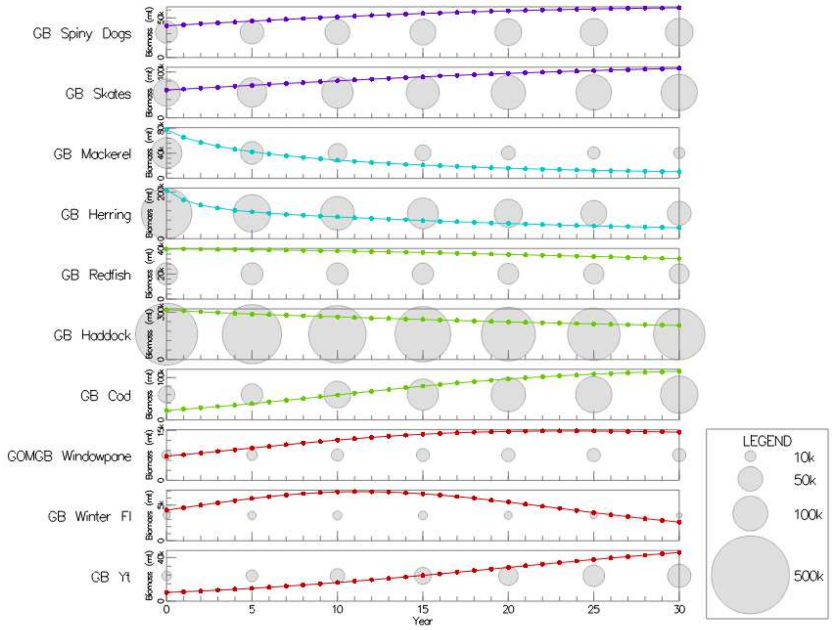


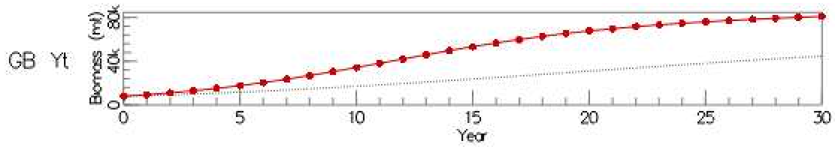
Figure 4-3: Absolute biomass indicators overlaying the “small multiples” view.

4.2 Visualization of Change

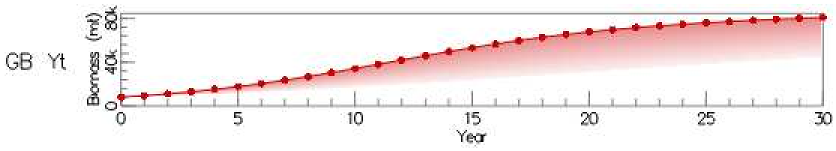
In order for modelers and other stakeholders to understand and compare decisions, the ability to perceive changes in biomass resulting from changes in the fishing effort is necessary. Therefore, we have introduced a feature that allows the user to compare the forecast effects of a change in fishing effort compared to the forecasts of the *status quo* [baseline]. There are three alternatives for displaying forecast differences with the baseline. The first is to simply have the biomass plots change instantaneously as the harvest effort sliders are adjusted, as in Figure 4-4a. The second is a dotted gray line which shows the forecast of the baseline in addition to the current forecast, shown in Figure 4-4b. The third is a shaded area originating from the curve of the current forecast that diminishes in opacity as it approaches the curve of the baseline forecast, as in Figure 4-4c.

Figure 4-5 shows all of the time series in the “small multiples” view with the effort sliders and the blended change option enabled. Again, the blended area in a line chart is colored from the current biomass line to the line as it was when the baseline was set; a colored area above the line indicates the biomass declined, while a colored area beneath

(a) Change shown by interaction.



(b) The status quo is shown as the dotted gray line.



(c) The area between the status quo graph and the new forecast is shaded.

Figure 4-4: The three options for depicting change between current biomass predictions and baseline predictions.

represents the biomass increased—e.g., the “skate” population declined dramatically with the new effort values, the “winter flounder” population increased due to the changes, and “haddock” seemed to be unaffected.

Underneath each slider, a colored rectangle—if present—indicates differences from the baseline effort settings. Blue indicates the effort value has been increased since the baseline was set—e.g., again in Figure 4-5, the effort for “elasmobranchs” was originally set to 1.0 and now it is approximately 2.0. Red represents the effort value has been decreased since the baseline was set—e.g., the effort for “flatfish” was originally set to 1.0 and now it is approximately 0.5.

This feature is available in both the “four panel” and “small multiples” views. The baseline effort setting can be defined at any time with a simple button at the top of the screen. This resets the baseline as the current slider settings. Buttons are also available to undo or reset changes to the effort values.

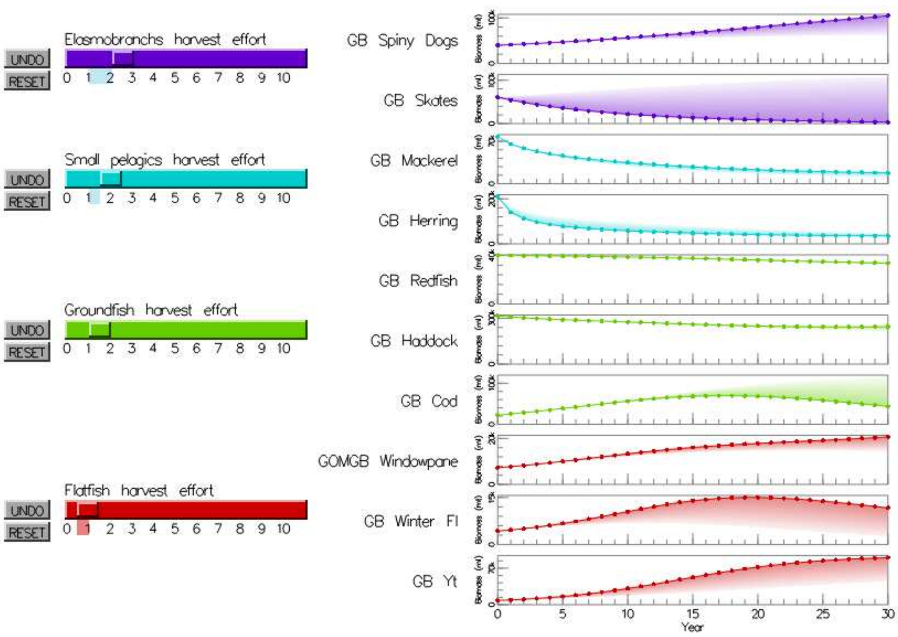
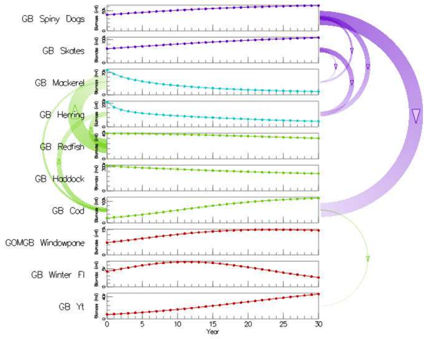


Figure 4-5: Showing change between different effort values.

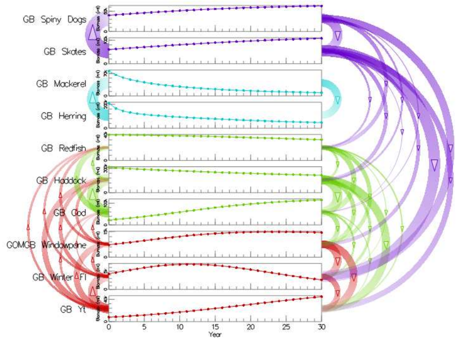
4.3 Visualization of Inter-Species Relationships

Understanding of the model requires understanding of the underlying relationships between species—namely, *predation* and *interaction*. As defined earlier, predation is when one species consumes another and interaction accounts for any way species might impact another in a way that is not predation. Our interface must explain to users which species impact each other and the magnitude of those relationships. These are illustrated with arc diagram network visualizations between time series. Firstly, users have the ability to view either predation as in Figure 4-6a, interaction as in Figure 4-6b, or both as static displays. Secondly, predation, interaction or both can be viewed dynamically, with the arcs being selectively drawn according to their impact as effort values are adjusted.

The arc diagram was selected over other network visualizations because it is easily combined with the small multiples. A separate node-link diagram, such as one with a force-directed layout, may have been confusing because it would require the user to mentally associate nodes in the network with line charts. An arc diagram does not suffer from this problem since, in our case, the time series themselves are the nodes. Furthermore, the arc



(a) Predation arcs.



(b) Interaction arcs.

Figure 4-6: Static arcs drawn between species charts to represent relationships.

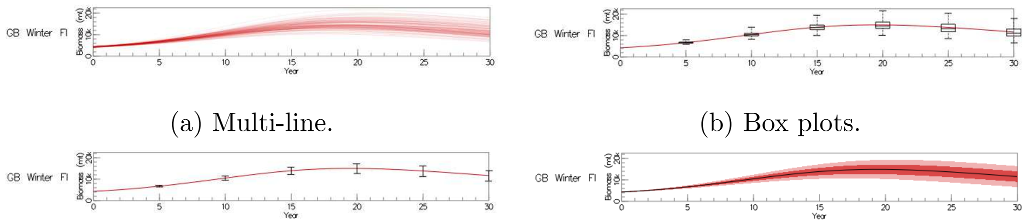
diagram enhances the entire visualization without occluding the time series. Arc diagrams are also well suited to smaller datasets with clusters of nodes, which applies to our dataset since the fish are segmented into functional groups.

4.3.1 Directionality

These relationships, predation and interaction, both imply directionality, where one species of fish is the "source" and the other species is the "recipient." Therefore, our arcs have been drawn with fading opacity to indicate the direction, since Holten and van Wijk recommended a dark-to-light representation for an intensity-based cue [13]. Additionally, our arcs follow a clockwise direction. This is necessary because there may be reciprocal "Fish A affects Fish B" and "Fish B affects Fish A" relationships, especially for the interaction type of relationship, so arcs must be drawn on both sides of the time series, as seen in Figure 4-6. Additionally, triangular marks have been drawn in the middle of the arcs to point from the originator to the recipient species.

4.4 Visualization of Uncertainty

Since models are simplifications of reality, their output is best understood as a range of expected values. It is possible that a representation of uncertainty may aid decision making.



(a) Multi-line. (b) Box plots.

(c) Error bars. (d) Error bands.

Figure 4-7: The four methods for visualizing uncertainty of MS-PROD model simulations.

To add uncertainty visualization to the MS-PROD model, our interface can perform Monte Carlo simulations by randomly jittering the non-zero input parameter values ±10%.

The resulting uncertainty can be displayed in four styles. First, a multi-line option shows the uncertainty by drawing one semi-transparent line for each run of the Monte Carlo simulation, as in Figure 4-7a. The second option displays a box every five years to indicate the boundaries of the first, second (median), and third quartiles with whiskers stretching to the minimum and maximum values, as in Figure 4-7b. Third, in Figure 4-7c, the mean of all simulations is displayed as a line, with bars every five years to represent one standard deviation above and below the mean. Lastly, Figure 4-7d shows the mean as a solid black line with bands to show one and two standard deviations above and below the mean.

Chapter 5

Work to be Completed

The remaining work consists mostly of completing a user evaluation and then finalizing the interactive interface based on the results of the evaluation. Some minor implementation changes may also be needed to support the evaluation. The purpose of the evaluation is to assess the design alternatives with the intended goal of creating a user interface to the MS-PROD model.

5.1 User Evaluation

There will be two phases to the user evaluation. First, experts will be instructed on how to use the MS-PROD visualization and questioned about their preferences in the form of a semi-structured interview. Second, novice users will be evaluated on their understanding of the model.

5.1.1 Feature Comparison by Experts

For the expert evaluation, we will visit the authors of the MS-PROD model, scientists, fishery managers, and other significant stakeholders. First, we will instruct them on how to use the various visualization alternatives. Then, we will ask them about their preferences and opinions regarding the two views (four panel grid and small multiples), the three methods for displaying change (instantaneous change only, line, and blend), and the four types of uncertainty (multi-line, box plots, error bars, and error bands). The different features will be either ranked or rated. Some questions which may be asked are:

- Which visualization method did you like the best?
- Is this feature useful?

5.1.2 Goal of Model Understanding

The second evaluation will be aimed at a more objective evaluation of how well the arc and change representations enable people to understand the underlying model. This will require undergraduate or graduate students as participants. Participants will be instructed in how to use the interactive interface to the model. Next, participants will be asked questions to guage their understanding of the model. For example, the participant may asked questions like, "What happens to X if we increase catch on Y?" More importantly, this question will be followed up with, "Why?" The purpose of these questions is to determine whether or not average users can gain an understanding of the complex relationships that result in indirect and chained effects.

Bibliography

[1] Wolfgang Aigner, Silvia Miksch, Wolfgang Müller, Heidrun Schumann, and Christian Tominski. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14:47–60, 2008.

[2] Dan Bricklin and Bob Frankston. VisiCalc: Information from its creators, Dan Bricklin and Bob Frankston. `http://bricklin.com/visicalc.htm`, 1999. [Online; accessed 2013-10-07].

[3] Norman L. Christensen, Ann M. Bartuska, James H. Brown, Stephen Carpenter, Carla D'Antonio, Rober Francis, Jerry F. Franklin, James A. MacMahon, Reed F. Noss, David J. Parsons, Charles H. Peterson, Monica G. Turner, and Robert G. Woodmansee. The Report of the Ecological Society of America Committee on the Scientific Basis for Ecosystem Management. *Ecological Applications*, 6(3):665–691, August 1996.

[4] William S. Cleveland, Marylyn E. McGill, and Robert McGill. The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83(402):pp. 289–300, 1988.

[5] Robert L. Eberlein and David W. Peterson. Understanding models with Vensim$^{\mathrm{TM}}$. *European Journal of Operational Research*, 59(1):216–219, May 1992.

[6] Andrew U Frank. Different types of "times" in GIS. *Spatial and Temporal Reasoning in GIS*, pages 40–62, 1998.

[7] Sarah K. Gaichas and Robert C. Francis. Network models for ecosystem-based fishery analysis: a review of concepts and application to the Gulf of Alaska marine food web. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(9):1965–1982, 2008-09-01T00:00:00.

[8] Robert J. Gamble and Jason S. Link. Analyzing the tradeoffs among ecological and fishing effects on an example fish community: A multispecies (fisheries) production model. *Ecological Modelling*, 220(19):2570 – 2582, 2009.

[9] Mohammad Ghoniem, Jean-Daniel Fekete, and Philippe Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *Proceedings of the IEEE Symposium on Information Visualization*, INFOVIS '04, pages 17–24, Washington, DC, USA, 2004. IEEE Computer Society.

[10] B. Grad. The creation and the demise of VisiCalc. *Annals of the History of Computing, IEEE*, 29(3):20–31, 2007.

[11] S. Havre, B. Hetzler, and L. Nowell. ThemeRiver: visualizing theme changes over time. In *IEEE Symposium on Information Visualization*, pages 115–123, 2000.

[12] Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky. A tour through the visualization zoo. *Commun. ACM*, 53(6):59–67, June 2010.

[13] Danny Holten and Jarke J. van Wijk. A user study on visualizing directed edges in graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2299–2308, New York, NY, USA, 2009. ACM.

[14] Waqas Javed, Bryan McDonnel, and Niklas Elmqvist. Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):927–934, November 2010.

[15] Donald E. Knuth. *The Stanford GraphBase: a platform for combinatorial computing.* ACM, New York, NY, USA, 1993.

[16] P. H. Leslie and J. C. Gower. The properties of a stochastic model for the predator-prey type of interaction between two species. *Biometrika*, 47:219–301, 1960.

[17] Alfred J. Lotka. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12):317–323, 1926.

[18] United States. National Marine Fisheries Service. Ecosystem Principles Advisory Panel. *Ecosystem-based fishery management: a report to Congress.* U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, 1999.

[19] William Playfair. *The Commerical and Political Atlas: Representing, by means of stained copper-plate charts, the progress of the commerce, revenues, expenditure and debts of England during the whole of the Eighteenth Century.* 1786.

[20] M. L. Rosenzweig and R. H. Macarthur. Graphical Representation and Stability Conditions of Predator-Prey Interactions. *The American Naturalist*, 97(895):209+, January 1963.

[21] Takafumi Saito, Hiroko Nakamura Miyamura, Mitsuyoshi Yamamoto, Hiroki Saito, Yuka Hoshiya, and Takumi Kaseda. Two-tone pseudo coloring: Compact visualization for one-dimensional data. *IEEE Symposium on Information Visualization*, 0:23, 2005.

[22] Milner B. Schaefer. A study of the dynamics of the fishery for yellowfin tuna in the eastern tropical pacific ocean. *Inter-American Tropical Tuna Commission Bulletin*, 2(6):243–285, 1957.

[23] Lisa Tweedie, Robert Spence, Huw Dawkes, and Hua Su. The influence explorer. In Jim Miller, Irvin R. Katz, Robert L. Mack, and Linn Marks, editors, *CHI 95 Conference Companion*, pages 129–130. ACM, 1995.

[24] Vito Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118:558–560, 1926.

[25] Martin Wattenberg. Arc diagrams: Visualizing structure in strings. In *Proceedings of the IEEE Symposium on Information Visualization*, INFOVIS '02, pages 110–116, Washington, DC, USA, 2002. IEEE Computer Society.

[26] M. Weber, M. Alexa, and W. Muller. Visualizing time-series on spirals. In *IEEE Symposium on Information Visualization*, pages 7–13, 2001.