

# Ecosystem food web constellation diagram using SAS/GRAPH® software

Rajeev Kumar and Divya Varkey

Fisheries Center, University of British Columbia, Vancouver, BC, Canada

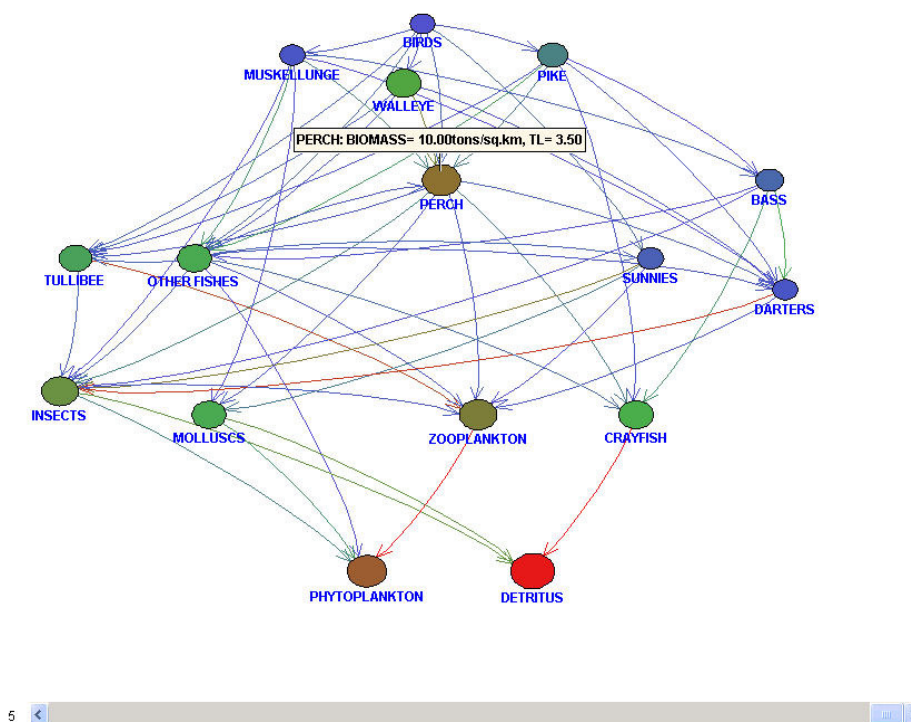
## Abstract

Ecosystem food web diagram is diagrammatic representation of trophic relationships among various organisms (plants and animals) that belong to different trophic levels in the ecosystem. It shows the direction of flow of energy among the ecosystem components. This paper presents the use of a SAS program for generating an interactive food web in Java Applets using DS2CONST macro of SAS/GRAPH® software. This interactive constellation diagram provides various user-friendly pop – up options that are helpful for easy diagnosis of trophic relationship. Diet diagrams are very useful in comparing the structure of ecosystem through time and hence the change in the system resulting from continued resource extraction.

## Diet matrix: a constellation graph

Resource management in fisheries has for decades concentrated mainly on management of single fisheries resource (for example Gulf of Maine cod, hake etc). Recent research however has highlighted the importance of interactions between species in the aquatic ecosystem (Link, 2002, Pitkitch et al, 2004, Hilborn et al, 2004). One of the most important interactions among species is their relation to each other as predator or prey. What this implies for fisheries management is that when fishing for a species, its importance to the ecosystem due to its role as predator or prey has to be quantified. Diet diagrams help quantify the predator prey interaction between species.

Constellation diagrams are very useful to 'illustrate relationships between links and nodes' (SAS Institute Inc. 2004). Many authors have used DS2CONST macro to create constellation graphs (Patel and Devarajan, 2002 and Hornibrook, 2006). This is perhaps the first time the DS2CONST macro has been used to show diet linkages in natural ecosystems. The key features of the diagram presented here are as follows: The species are arranged by trophic level. Trophic level is the position in the food chain of the species, determined by the number of energy-transfer steps to that level (Fishbase, 2009). Each species is denoted by a node and the size of the node is related to the biomass of the species in the system. The arrows between the nodes indicate the dietary links. The lines originate from the predator and the arrows terminate at the prey species. The diagram is shown in the figure.



# Methodology explained

## Data

The diet matrix is part of an ecosystem model built using Ecopath and Ecosim software, a software that is widely used for building models of aquatic ecosystems (Christensen and Walters. 2004, Christensen and Walters, 2005). The idea for the diet diagram is based on the flow diagram which can be created within the Ecopath with Ecosim software. The data used in this paper is loosely based on the species composition of a North-American lake ecosystem. It is to be noted that the data are not an accurate representation of any lake ecosystem. The data presented here should therefore not be used for any scientific study of a lake ecosystem. The authors intention is to show an example of the diet diagram.

### Diet Matrix

In a diet matrix, the predators are commonly denoted by columns and the preys are presented in rows. Each column therefore shows the proportion of each prey item in the diet of the respective predator species. For example, in the diet matrix shown in table 1, 'Prey2' and 'Prey5' constitute 20% each and 'Prey3' and 'Prey4' constitute 30% each in the diet of predator 'PD1'.

| Table 1. Diet Matrix example |             |     |     |     |     |     |
|------------------------------|-------------|-----|-----|-----|-----|-----|
| SrN                          | PREY        | PD1 | PD2 | PD3 | PD4 | PD5 |
| 1                            | WALLEYE     |     |     |     |     |     |
| 2                            | PERCH       | 0.2 |     | 0.2 | 0.5 |     |
| 3                            | PIKE        | 0.3 | 0.5 |     |     |     |
| 4                            | MUSKELLUNGE | 0.3 | 0.5 | 0.4 | 0.2 | 1   |
| 5                            | TULLIBEE    | 0.2 |     | 0.4 | 0.3 |     |

When formatting the data table, care should be taken to name the predator columns as PD1 to PDn, where n indicates the maximum number of predators in the diet matrix. The diet matrix thus shows the interactions among predators and preys. Each species will therefore be represented as nodes in the desired diagram, and the lines leading out from each node show the predator prey interactions.

### Biomass and trophic level

The importance of a species in any ecosystem is also dependent on its biomass in the system. The size of each node is therefore set to represent the biomass. A bigger size node indicates a higher biomass and vice-versa. In the desired diagram, the vertical position of the nodes is represented by the trophic level. An example of a raw datafile (referred to as BMass) is shown in table 2.

| Table 2. BMass example |             |          |      |
|------------------------|-------------|----------|------|
| SrN                    | GROUP_NAME  | BIOMAS S | TL   |
| 1                      | WALLEYE     | 3        | 4.12 |
| 2                      | PERCH       | 10       | 3.5  |
| 3                      | PIKE        | 0.5      | 4.3  |
| 4                      | MUSKELLUNGE | 0.1      | 4.3  |
| 5                      | TULLIBEE    | 1.5      | 3    |

The data file should include a serial number as this number is used later in the program to recognize the nodes and the nodes which should be linked. Care should be taken that in both the data files diet matrix and Bmass, the species are arranged in the same order so that a particular serial number indicates only any one of the species in the system. It is therefore important that in the arrangement of predators, the PD1 refers to 'WALLEYE', PD2 refers to 'PERCH' and so on. In both the example tables 1 and 2, only the first 5 observations are shown.

## Program

### Importing the data files

Data files can be imported from the SAS GUI interface or a separate program can be written to import the datafiles.

In the example the diet matrix is labeled as sheet 'dmatrix' in an excel file and the data for biomass and trophic level is labeled as "bmass" in the same excel file (change file pathname or other specifications according to your system – highlighted in grey throughout the program). After submitting the programs below, two files 'dietmatrix' and 'rawnode' are imported in the temporary 'work' folder of sas system from excel file.

```

PROC IMPORT OUT= WORK.dietmatrix
    DATAFILE= "C:\SAS PAPER\DIET MATRIX.xls"
    DBMS=EXCEL REPLACE;
    SHEET="DMATRIX$";
    GETNAMES=YES;
    MIXED=NO;
    SCANTEXT=YES;
    USEDATE=YES;
    SCANTIME=YES;
RUN;
PROC IMPORT OUT= WORK.rawnode
    DATAFILE= "C:\SAS PAPER\DIET MATRIX.xls"
    DBMS=EXCEL REPLACE;
    SHEET="BMASS$";
    GETNAMES=YES;
    MIXED=NO;
    SCANTEXT=YES;
    USEDATE=YES;
    SCANTIME=YES;
RUN;

```

For drawing the constellation diagram, the DS2CONST Macro uses 2 datasets, one to draw the links among the nodes (example 'linkdata') and the other to draw the nodes (example 'nodedata').

### ***Data file for drawing the links (linkdata)***

The dataset referred to as 'linkdata' is created using the dataset 'dietmatrix' imported earlier.

In the paper dataset 'linkdata' contains the columns for:

1. the node from which the link is drawn
2. the node to which the link is drawn
3. the width of the line linking the nodes
4. the popup that is displayed when the mouse is placed over the link.

The diet matrix as the name suggests contains the data in the form of a matrix of predator and prey. All the predator prey combinations need to be represented in 2 columns of predator and prey. In the example above, the predator PD1 feeds on 4 prey items, and PD2 feeds on 2 prey items. Therefore 4 links are needed to be drawn from predator PD1 and 2 links from predator PD2. To perform this operation, the diet matrix is rotated. This rotation is done by using the array (P) and do loop. The do loop loops over PD1 to PDn. Look at the program below. A variable 'predator' is created in the data set 'dmr' with the value of 1 to n. Thus, 1 represents PD1 and 2 represents PD2....., and n represent PDn. In the do loop the diet proportion is retained by creating a variable 'amount'. The purpose behind retaining diet proportions is that it can be used to set the relative width of the line linking the nodes and the colour of the link lines. Thus the thickness of the line and its color will indicate how important the prey is in the diet of the particular predator. The diagram will thus identify the important interactions in the ecosystem.

```

data dmr (keep= SrN Prey predator amount);
    set dietmatrix ;
    array P {*} PD1-PD16;
    do predator=1 to dim(p);
        Amount=P{predator};
        output;
    end;
run;

```

In an ecosystem, every organism is not directly linked with every other organism. This causes missing values in the diet matrix as can be seen from the table 1 above. The missing values exist in the data file 'dmr' created after diet matrix rotation.

To remove the missing values and retention of only non-zero values of the variable 'amount' in the data set 'dmr' which show interaction, the 'where' statement is used in 'proc sort' data step below.

```
proc sort data=dmr out=dmrs;
    by predator;
    where amount gt 0;
run;
```

DS2CONST macro provides the opportunity that when the mouse is moved over the link line, then a popup be displayed. The popup can be created using the concatenation operator (||) with the trim function (to remove the trailing blanks). Now the following program finally creates the sas data set 'linkdata' which is one of the requirements of DS2CONST macro argument.

```
data linkdata (drop= text);
set dmrs;
Text= 'FEEDS ON';
amount=amount*100;
Link_popup = trim(text)|| ' ' || trim(Prey) || ' BY ' || put(amount, 4.)||
'%';
run;
```

### ***Data file for drawing the nodes (nodedata)***

The datasets referred to as 'nodedata' is created using the data 'rawnode' imported earlier. In the paper dataset 'nodedata' contains the columns for:

1. determining the number of nodes in the graph
2. setting the size of the nodes
3. setting the position of the nodes on the y axis (when the 'layout = user' in the DS2CONST macro argument)
4. setting the position of the nodes on the x axis. (when the 'layout = user' in the DS2CONST macro argument)
5. the popup that is displayed when the mouse is placed over the nodes.

The number of nodes in the desired diagram will be equal to the number of observations in the dataset 'nodedata'.

The size of the nodes is proportional to the biomass of the species in the ecosystem. Since the biomass of low trophic species like phytoplankton and high trophic top predator varies by several orders of magnitude, the sizes of the nodes are scaled to the log of the species biomass.

The vertical position of the nodes (yloc) is scaled to the trophic level of the species to depict the flow of energy in the system. The scale of the applet on the yaxis is 0 at the top end and increases towards the bottom of the graph. The yloc is calculated by the formula: applet height – (trophic level\*applet height/maximum trophic level)

The horizontal position on the x axis is calculated such that every alternate observation occurs to the right and left of the midpoint of the x axis. The distance at which the points should fall is chosen by sampling from a uniform distribution.

As in dataset 'linkdata' when the mouse is moved over the nodes, then a popup be displayed. The popup can be created using the concatenation operator (||) with the trim function (to remove the trailing blanks). See the final data steps of the programs below.

Finally, the entire program for creating a dataset 'nodedata' which is the other dataset requirement of DS2CONST macro argument is given below.

```
proc sort data=rawnode;
by descending tl;
run;

data z (keep= x1);
set rawnode;
x1=uniform(150847);
run;
```

```

proc transpose data =z out=z;
run;

data z (drop=i);
set z ;
array divya (*) _numeric_;
do i= 3 to dim (divya) by 2;
divya (i-1)= divya(i);
end;
run;

proc transpose data=z out=z;
run;

data rawnode (drop = _name_);
set rawnode;
set z;
run;

data nodedata (drop=width text text1 x1 x11);
set rawnode;
text= 'BIOMASS';
text1='TL';
node_popup = trim(group_name) !! ': '!!trim(text)!! '=' !! put(biomass, 6.2)
!! 'tons/sq.km'!! ', '!!trim(text1)!! '=' !!put(tl, 4.2);
Node_size= log(biomass*10**5)*5;
yloc=700-(TL*140);/*700 is the max height of graph and 140 is 700/5.
5 is max TL to be plotted */
width=800;/* width is the desired width of graph in macro argument*/
x11=(int(x1*width/2))-20;
if mod(_n_ , 2)=0 then xloc = width/2+x11;
if mod(_n_ , 2)=1 then xloc = width/2-x11;
if _n_ =1 then xloc=width/2;
run;

```

## DS2CONST Macro

The DS2CONST macro allows the diagram to be displayed on a Java applet. The Java applet provides more interactive features with the diagram. It allows the user to zoom in or out on the diagram, to display or not display certain links, to show popups describing the links etc. The diagram is created in an html file and thus can be displayed on any computer irrespective of whether the computer has or does not have SAS program on it. In computers without SAS program an additional file “constapp.jar” is required to display the diagram. This function makes it ideal for presentations from any computer anywhere. The detailed description of the macro arguments can be found in sas documentation (SAS Institute Inc. 2004).

### *Important arguments in the DS2CONST macro*

Applet height and width: In the example, the applet height and width are set to 700 and 800 pixels respectively. It can be changed according to the convenience of the user. The user will have to be careful about rescaling the diagram to the new applet dimensions. The program has been commented so that it is easy to follow the steps. The DS2CONST macro has large numbers of arguments to enhance the graphics. The complete documentation for DS2CONST macro is available in SAS documentation (SAS Institute Inc. 2004).

```

ods listing;          /* ods listing should be open before running the macro*/

%DS2CONST
(ndata=nodedata,      /*required dataset for creating nodes*/
ldata=linkdata,       /* required dataset for links*/

codebase= C:\XXX\,    /*location of java applets of your system */
htmlfile=C:\XXX\diet.html, /* name of the html file to be created in the
                           same location as codebase */

archive=constapp.jar,
height=700,           /* applet height*/
width=850,             /* applet width*/
border=Y,             /* for border around background area. default=N */
align=absmiddle,      /* alignment of the applet window in the browser
                        window(left,right,top,bottom,middle */

ibackloc= URL,        /* URL of the picture for background of the diagram*/
ibackpos=tile,        /* Orientation of background image in ibackloc argument
                        of the diagram (Center, scale, position, tile)*/

datatype=assoc,       /* arcs(by default),assoc, hier*/
colormap=y,           /* for colour of node and link. default=N */
nvalue=node_size,     /* size of the node. It also determines default node
                        color if colormap=Y */

minlnkwt=5,           /*minimum weight of the links to be displayed. This can
                        be adjusted later using the horizontal scroll bar
                        below the diagram */

sclnkwt=N,            /*related to minlnkwt argument; variable values range
                        should be greater than or equal to 2*/

nid=srn,              /* variable which specifies the nodes */
nlabel=Group_Name,    /* labels for the nodes */
labels=y,             /* node label display in diagram(y or n) default = Y */
fntsize=12,           /* Node label font size */
fntstyl=bold,         /* Node label font style (bold, italic, plain) */
catext=blue,          /* Node label text color) */
nodeshap=circle,      /* shape of node(circle,square,diamond,triangle,none) */
layout=user,          /* auto or user. If user then position based on user
                        defined nx=xloc and ny=yloc*/

nx=xloc,              /* user defined horizontal position of the nodes */
ny=yloc,              /* user defined vertical position of the nodes */
lfrom=predator,       /* origin of link lines */
lto=SrN,              /* destination of link lines */
lvalue=amount,        /* variable determine the colour of link */
ltip=link_popup,      /* popup for link*/
ntip=node_popup,      /* popup for node*/
linktype=arrow);      /*type of link- line, arrow*/

```

## Discussion

Effort has been made to write the program so that it is as less hardwired as possible. A user only needs to create the dataset in the format shown in the examples. The code runs from importing the data files to creating the diagram in just a few seconds. The serial number in the data files 'diet matrix' and 'bmass' are key factors in the drawing the diagram accurately. The predator and prey for the nodes and links are identified by the serial number throughout the SAS program.

For a diagram which needs more than 50 nodes or more than 1000 links (for example a diet matrix of a coral reef ecosystem) the pop-ups are useful because they provide the necessary information without cluttering the diagram with labels.

The weakness of the program lies in the horizontal position of the nodes. Since these positions are decided by sampling from a uniform distribution, there are chances that the nodes might overlap over each other. It is advised that the user run the program a few times and save the diet diagram that presents the best arrangement of the nodes. Another alternative is that the users try out a number of different seeds for the random number generation and choose the seed that provides the best arrangement of the nodes on the diagram. However, if the layout of the diagram is set to auto, the SAS program automatically positions the nodes without any overlaps using "stress and strain equations" (SAS Institute Inc. 2004).

## Acknowledgements

The authors thank Prof. T. J. Pitcher and members of Fisheries Ecosystem Restoration Group, Fisheries Center, University of British Columbia, for their valuable suggestions.

## References

- Christensen, V. and C. J. Walters. 2004. Ecopath with Ecosim: methods, capabilities and limitations. *Ecol. Model.* 172:109-139
- Christensen, V., Walters, C.J., 2005. Using ecosystem modeling for fisheries management: where are we? ICES CM/M:19, Copenhagen, Denmark.
- Fishbase. 2009. Froese, R. and D. Pauly. Editors. 2009. FishBase. World Wide Web electronic publication. [www.fishbase.org](http://www.fishbase.org), version (06/2009).
- Hilborn, R., Punt, A.E. and J. Orensanz, Beyond band-aids in fisheries management: fixing world fisheries, *Bulletin of Marine Science* 74 (3) (2004), pp. 493–507.
- Hornibrook, S. Analyzing Large Social Networks with MP Connect, SAS/IntrNet®, and %DS2CONST. The SouthEast SAS Users Group (SESUG) 2006.
- Link, J. 2002. What does ecosystem-based fisheries management mean? *Fisheries* 27, (4): 10-21.
- Patel, H. and Devarajan, R. Enhancements to SAS/GRAPH® in V9. SAS Conference Proceedings: SAS Users Group International 27 April 14-17, 2002, Orlando, Florida.
- Pikitch, E.K., Santora, C., Babcock, E.A., Bakun, A. Bonfil, R. and D.O. Conover *et al.*, Ecosystem-based fishery management, *Science* 305 (2004), pp. 346–347.
- SAS Institute Inc. 2004. Macro arguments for the DS2CONST, DS2Tree, DS2Csf, and Meta2htm Macros In SAS/GRAPH® 9.2 Reference, Volumes 1 and 2, Chapter 21, pp. 535 to 566.

## Contact information

### Rajeev Kumar



PhD Candidate  
Fisheries Center  
University of British Columbia  
AERL, 2202 Main Mall  
Vancouver, BC V6T 1Z4 CANADA  
[r.kumar@live.com](mailto:r.kumar@live.com)

### Divya Varkey

PhD Candidate  
Fisheries Center  
University of British Columbia  
AERL, 2202 Main Mall  
Vancouver, BC V6T 1Z4 CANADA  
[d.varkey@live.com](mailto:d.varkey@live.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.  
Other brand and product names are registered trademarks or trademarks of their respective companies.