# A Toolbox for the Choice of Indicator Classes for Ranking of Watersheds

K. Sham Bhat
Penn State University
Department of Statistics
Center for Statistical Ecology and Environmental Statistics
University Park, PA 16802 USA
Tel: 001 814 865 5212
kgb130@psu.edu

## ABSTRACT
There has been considerable work on determining a suitable method to accomplish a satisfactory ordering of a group of objects, when there are multiple evaluation criteria. A weighted index can be used to combine the opinion of all stakeholders to obtain a criterion for ranking the objects. Data from 21 watersheds of the Atlantic Slope Consortium (ASC) has been examined with the goal of determining an accurate ranking by overall watershed condition. In particular, there are three groups of indicators that range from Level I to Level III, increasing in the quality and accuracy of the data as well as the cost and effort needed to obtain the data. Due to the high cost, Level III data is available only for six watersheds; however the interest is in ranking all 21 watersheds. The investigators using their expertise have developed indices for the Level I and Level II indicators, and we would like to determine if those indices are effective in ranking the watershed condition. In addition, we introduce a method of selecting effective weights using POSAC (Partial Order Scalogram Analysis), which gives us insight for improving the weights already selected by the stakeholders. We also use elements of poset (Partial order set) theory as a foundation for our analysis. The Poset linear extension method can be used to find rankings without using an index, relying only on pairwise comparisons of the objects.

## Categories and Subject Descriptors
H.4 [Information Systems Application]: types of systems

## General Terms
Multicriterion decision making, weights, aggregations, posets.

## Keywords
Multicriterion prioritization, differential weightings, cumulative rank frequency distribution operators, Atlantic slope consortium, performance metrics.

## 1. DATA AND OBJECTIVES

### 1.1 Motivation and Data Description

Water affects everything that can think about, and we require clean water for not just our quality of life but our survival. Watersheds are regions of land that drain into a waterway, and to retain a healthy water supply, we need healthy watersheds. The task of maintaining healthy watersheds as well as remedying unhealthy ones is a often a monumental task, as such the large scale problem of watershed management is one that is primarily the task of the government. Federal, state, and local governments spend millions of dollars every year in an effort to promote good watershed management and with limited resources there is need to ensure that resources go to the watersheds where it would be most effectively utilized. In order to aid the government agencies in determining where to best allocate resources, we wish to obtain a solid method to prioritize the watershed which results in a lesser cost to the relevant agencies.

The data at issue in this paper is a set of 21 watersheds from the Atlantic Slope Consortium (ASC), which have the goal for determining an accurate ranking of the health of the watersheds. [1] The watersheds are from the mid-Atlantic region, which is shown in Figure 1. There are 15 indicators each of which measures a facet of the watershed and are divided into three groups. In addition, each watershed is identified with one of four physiographic regions, which describe the terrain of the land, and one of six social choices, which describes the use of the land (urban, agriculture, forest, mixed etc.).

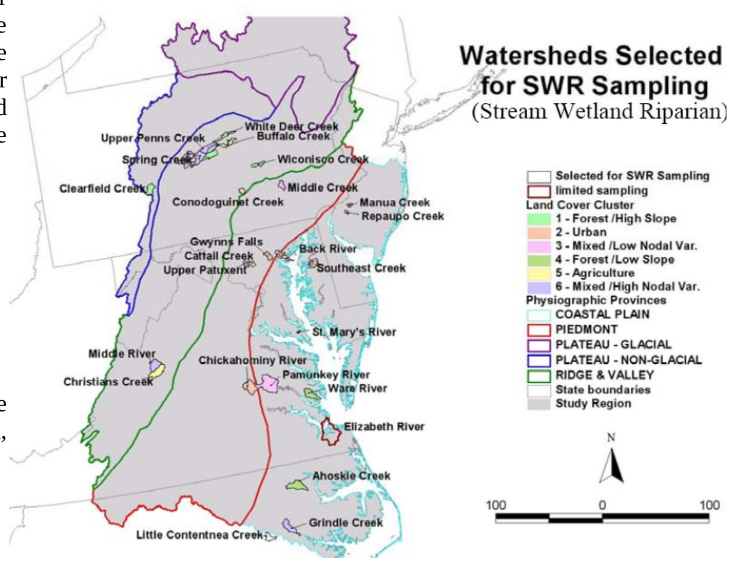


**Figure 1- The 21 watersheds from the Mid-Atlantic area**

This data set has three different groups of indicators, from Level I to Level III, increasing in the quality and accuracy of the data as well as the amount of cost and effort needed to obtain the data. For all 21 watersheds, we have data for seven Level II indicators, and for five Level I indicators. The data matrix for Level II indicators has 21 rows and 7 columns, and the Level I indicator data matrix has 21 rows and 5 columns, and the data was scored between zero and one. The three groups are as follows:

Level III – Intensive Field Assessment
Level II – Rapid Field Assessment

Level I – Landscape Assessment from GIS

The indicators are categorized in part by the amount of accessibility and availability. The Level III Intensive Field Assessment is be purchased from the U.S. Environmental Protection Agency (EPA). We consider the indicators of Level III, the best quality of data we have regarding the watershed, however, it is the most expensive data among the three groups. Due to the money and effort in the procedure of obtaining this data, it was collected for only six watersheds.

The three Level III indicators are collected from the Index of Biological Integrity (IBI) developed by EPA [1]. They are benthic IBI, fish IBI, and $NO_3$. The two IBI indicators are biological indicators, while the $NO_3$ is considered a chemical indicator. The data set of the three Intensive Field Assessments for six selected watersheds is shown below in Table 1. The $NO_3$ indicator has been reoriented in order to ensure that watersheds with the lower amounts of $NO_3$ receive higher values, e.g. subtracting each value from 5, in order to ensure that values remain positive.

**Table 1- Level III Data and Indicator Definitions**

| WATERSHED | BIBI | FIBI | $NO_3$ |
|---|---|---|---|
| Back River | 1.515 | 2.215 | 1.318 |
| Cattail Creek | 3.756 | 3.268 | 4.332 |
| Gwynn Falls | 1.938 | 2.445 | 1.362 |
| Saint Mary's A | 2.858 | 3.875 | 0.196 |
| Southeast Creek | 2.687 | 3.444 | 3.008 |
| Upper Patuxent | 3.750 | 4.113 | 2.836 |

| Indicator | Definition |
|---|---|
| BIBI | Benthic Index of Biological Integrity |
| FIBI | Fish Index of Biological Integrity |
| $NO_3$ | Nitrate content |

**Table 2-Level II Scored Data and Indicator Definitions**

| Watershed | BUF | IR | BA | INV | SHA | SS | FPWL |
|---|---|---|---|---|---|---|---|
| Back River | 0.6047 | 0.5751 | 0.4284 | 0.2025 | 0.5215 | 0.2188 | 0.7038 |
| Cattail Creek | 0.3579 | 0.7966 | 0.5688 | 0.6525 | 0.7023 | 0.41 | 0.279 |
| Gwynn Falls | 0.3232 | 0.7676 | 0.5641 | 0.6825 | 0.6765 | 0.3113 | 0.165 |
| Saint Mary's A | 0.7774 | 0.6894 | 0.599 | 0.8375 | 0.6967 | 0.635 | 0.669 |
| Southeast Creek | 0.7701 | 0.8195 | 0.5814 | 0.5075 | 0.6108 | 0.5713 | 0.706 |
| Upper Patuxent | 0.6828 | 0.6375 | 0.7254 | 0.6125 | 0.8123 | 0.69 | 0.542 |
| Ahoskie | 0.563 | 0.3617 | 0.5195 | 0.5975 | 0.4771 | 0.4 | 0.5885 |
| Buffalo Creek | 0.3524 | 0.8753 | 0.5816 | 0.86 | 0.631 | 0.3888 | 0.2448 |
| Chickahominy | 0.5833 | 0.6598 | 0.6431 | 0.76 | 0.543 | 0.4525 | 0.498 |
| Christian Creek | 0.0948 | 0.5945 | 0.2479 | 0.92 | 0.704 | 0.3888 | 0.4668 |
| Clearfield Creek | 0.6466 | 0.837 | 0.1558 | 0.98 | 0.7033 | 0.395 | 0.6788 |
| Conodoguinet A | 0.3233 | 0.7236 | 0.3637 | 0.6474 | 0.6929 | 0.3671 | 0.2147 |
| Grindle Creek | 0.4562 | 0.3354 | 0.4679 | 0.59 | 0.4814 | 0.4475 | 0.6163 |
| Little Contentnea | 0.5199 | 0.7268 | 0.628 | 0.8225 | 0.6877 | 0.6275 | 0.7435 |
| Mantua | 0.5226 | 0.8933 | 0.6356 | 0.7594 | 0.6878 | 0.6406 | 0.345 |
| Middle Creek | 0.3611 | 0.79 | 0.4519 | 0.7947 | 0.6705 | 0.3789 | 0.2624 |
| Middle River | 0.191 | 0.5239 | 0.2664 | 0.835 | 0.6188 | 0.25 | 0.3068 |
| Pamunkey | 0.7013 | 0.573 | 0.7039 | 0.91 | 0.6712 | 0.545 | 0.6765 |
| Repaupo | 0.5313 | 0.8482 | 0.5676 | 0.6706 | 0.7109 | 0.7588 | 0.4368 |
| White Deer Creek | 0.8598 | 0.8606 | 0.8016 | 0.975 | 0.9363 | 0.795 | 0.72 |
| Wisconisco | 0.5801 | 0.8074 | 0.5249 | 0.84 | 0.7168 | 0.4363 | 0.4295 |

| Indicator | Definition |
|---|---|
| BUF | Buffer Score |
| IR | Incision Ratio |
| BA | Basal Area of Trees |
| INV | Invasive Cover Class |
| SHA | Stream Habitat Assessment Score |
| SS | No. of Stream Stressors |
| FPWL | No. of Floodplain-wetland Stressors |

The Level II Rapid Field Assessment is obtained from onsite sampling and a good degree of expertise is involved in the field assessment. Generally, however, Level II data is relatively cheap compared to the Level III data. The Level I Landscape Assessment is the satellite data, and is the easiest to access and the least expensive, and requires little effort to collect.

The scores for each of the indicators in Level I and Level II, between 0 and 1, were obtained by first reorienting the indicators to ensure that there is a positive relationship between the indicator and the quality of the watershed (i.e. higher score value for an indicator implies better condition of watershed for that indicator.) More details about the scoring are found in [1]. The Level I and Level II score data are listed below in Table 2 and Table 3.

**Table 3-Level I Scored Data and Indicator Definitions**

| Watershed | FOR | LDI | IMP | MPAT | CORFOR |
|---|---|---|---|---|---|
| Back River | 0.1407 | 0.0665 | 0.0225 | 0.0792 | 0.0703 |
| Cattail Creek | 0.3158 | 0.6609 | 0.6888 | 0.2967 | 0.2405 |
| Gwynn Falls | 0.2454 | 0.1544 | 0.0443 | 0.182 | 0.1238 |
| Saint Mary's A | 0.6947 | 0.7607 | 0.3067 | 0.6337 | 0.5737 |
| Southeast Creek | 0.3237 | 0.5293 | 0.666 | 0.4198 | 0.2629 |
| Upper Patuxent | 0.3854 | 0.6737 | 0.6441 | 0.3656 | 0.4002 |
| Ahoskie | 0.7017 | 0.7505 | 0.4717 | 0.6641 | 0.7249 |
| Buffalo Creek | 0.3038 | 0.6042 | 0.5604 | 0.3926 | 0.6536 |
| Chickahominy | 0.446 | 0.4417 | 0.1225 | 0.3841 | 0.3668 |
| Christian Creek | 0.2994 | 0.6363 | 0.4458 | 0.3856 | 0.3154 |
| Clearfield Creek | 0.7411 | 0.8386 | 0.5538 | 0.7167 | 0.6139 |
| Conodoguinet A | 0.3171 | 0.5188 | 0.2037 | 0.245 | 0.2827 |
| Grindle Creek | 0.5751 | 0.6796 | 0.6308 | 0.5418 | 0.6657 |
| Little Contentnea | 0.5853 | 0.6556 | 0.511 | 0.5361 | 0.5427 |
| Mantua | 0.3683 | 0.4168 | 0.1389 | 0.2482 | 0.12 |
| Middle Creek | 0.4281 | 0.6772 | 0.5381 | 0.4323 | 0.579 |
| Middle River | 0.3087 | 0.6037 | 0.3383 | 0.2985 | 0.2782 |
| Pamunkey | 0.6186 | 0.7651 | 0.5999 | 0.6698 | 0.5893 |
| Repaupo | 0.3548 | 0.6247 | 0.5682 | 0.3794 | 0.3074 |
| White Deer Creek | 0.9454 | 0.9607 | 0.6157 | 1 | 0.866 |
| Wisconisco | 0.8455 | 0.8879 | 0.4798 | 0.7623 | 0.7191 |

| Variable | Definition |
|---|---|
| FOR | % Forest in Watershed |
| LDI | Landscape Density Index in Watershed |
| IMP | % Impervious Surface in Watershed |
| MPAT | Mean Forest Patch Size in Watershed |
| CORFOR | % Total Forest That is Core Forest In Watershed |

The investigators made an effort to combine the seven Level II indicators into a composite index, and the five Level I variables into another index [1]. These preliminary composite indices are expected to represent the condition of the watersheds.

The composite Level II index is called Stream-Wetland-Riparian (SWR) index. The index is computed by averaging the Floodplain-Wetland(FW) index, Incision Ratio (IR) score, Stream Habitat Assessment (SHA) score, and Stream Stressor (SS) score, where the FW index is the mean of Buffer (BUF) score, Basal Areas(BA) score, Invasive (INV) score, and Floodplain-Wetland Stressor score. The conceptual model of condition used for SWR index is shown in Figure 2 [1].

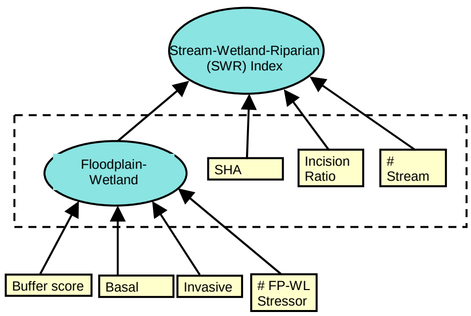


**Figure 2 – The Conceptual Model of the SWR Index**

The composite Level I index is called the Landscape index. It is computed by averaging the Forest score, Urbanization score, and Fragmentation score. Exclusively, it follows the formula Landscape index = [forest score + (IMP +LDI)/2 + (MPAT + CORFOR)/2] / 3. The conceptual model of condition used for Landscape index is shown in Figure 3. [1]

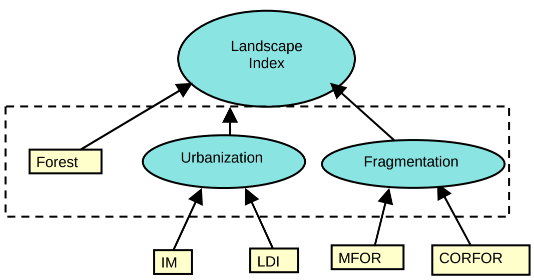


**Figure 3 – The Conceptual Model of the Landscape Index**

Both the Level I and Level II investigator based indices can be seen as a two stage hierarchy. There are three groups in the top stage for Level I; Urbanization and Fragmentization, which consist of two indicators each, and the single indicator Forest. For Level II, we have four groups in the top stage, the three single indicators SHA, Incision Ratio, and Stream Stressor, and Floodplain-Wetland Condition, which consists of four indicators.

## 1.2 Objectives

The goals of this paper are broken into the following questions:

1. Does the composite index for Level II, the SWR Index accurately represent the condition of watersheds? How well can the Level II indicators approximate the Level III indicators?
The Level III indicators are assumed to best represent the condition of watersheds. If the Level II indicators are in general agreement with the Level III indicators, then these less expensive measurements (Level II) can be considered as alternates to the expensive EPA indicators (Level III).

Scientists and investigators formulated the composite indices by taking the weighted average of individual assessments within indicator groups. Statistical methods will be applied to evaluate the quality of the composite index. Furthermore, an alternative data based method of constructing a single index will be developed, tested, and compared with the composite index.

2. Can we use POSAC to determine and assess a set of weights for an index of Level II indicators determined from the data alone?
POSAC is a method of data analysis that will give us a more thorough understanding of the data and the rankings of the objects. Using a regression approach to determine the weights for the Level II indicators is difficult here, due to fewer watersheds (6) than indicators(7). More about POSAC will be discussed later in the paper.

3. How well does our index constructed from the analysis of data alone (using Level II indicators) perform compared to the weights selected by the investigators?
We introduce a simple performance measure to understand how well the data based indices predict the true order of the watersheds. The performance measure that we use is the correlation of the ranks of the six selected watersheds derived from our index with the Level III ranks of those same watersheds.

4. How well does the Level I Landscape Index, and Level I data based index perform in comparison to Level III, or Level II results.
Level I data is primarily data from surveillance, and usually does not need any experimentation or field work. The data is also readily available, and thus it makes for a lower cost project than Level II data. We are interested in seeing how well the Level I data perform in order to determine whether it is a feasible alternative for Level III and Level II data. If Level I data is a sufficient approximation for Level III data, then it will save much trouble for scientists as well as government agencies in assessing the health of watersheds.

# 2. METHODS

## 2.1 Basics of Posets and Hasse Diagrams

There is often a need to rank objects, whether they are people, cities, watersheds, countries, or bridges when there are multiple criteria of importance. This is an especially critical problem for those involved in the government sector in allocating what are often very limited resources. One way to accomplish this task is to combine the indicators into an index, i.e. Index= $g_1 I_1 + g_2 I_2 + ... + g_n I_n$, where $I_k$ are indicators, and where the $g_k$s sum to one. The main issue in this approach is determining an appropriate set of weights for the model, acknowledging the risk of underweighting or overweighting some indicators. However, combining indicators into an index often can generate controversy, as there can be wide disagreement about the composition of the weights.

We often use partial order theory as a basis for much of the analysis for ranking objects. Poset theory is a very large field and we only go through the basics, more information about Poset theory can be found in [3]. Object A is considered intrinsically "better" than object B, which we denote A>B, if all indicators rate object A greater than or equal than object B with at least one indicator considering object A strictly greater than object B. If object A is neither intrinsically better nor intrinsically worse than object B, then we consider the two objects incomparable, denoted by A||B. For every pair of objects whose indicator values are not identical, one of the following must be true, A>B, A<B, or A||B.

A visual description of the poset formulation is through the Hasse diagram, which consists of a few simple rules. Each object is denoted by a circle. Circles are in one-to-one correspondence with the objects in the dataset. The Hasse diagram technique uses several tiers for clarity, object A must be in a higher tier than an object that A is better than, and in a lower tier than an object that is better than A. Specifically, we place objects that no other objects are better than in the top row, we call these objects *maximal elements.* For the second row, we place all objects that have no object not already placed (in the top row in this instance) better than it, and so on. In other words, we remove the maximal elements and find the maximal elements of the objects still remaining to determine each tier.

If object A is intrinsically better than object B, then we denote this with a straight line (or edge) between A and B. If A is incomparable with B, then no line between the two objects is drawn. This procedure is repeated until all of the tiers are established and lines drawn.

These rules tell us that a Hasse diagram consists of three parts: circles, tiers, and lines. Hasse diagrams are oriented graphs with objects in an order relation if they are mutually comparable. In other words, a line connecting a pair of circles implies that all of the attributes of the upper circle are greater than or equal to that of the lower circle. On the other hand, a lack of connecting lines indicates that there are contradictions in the scoring between objects.

A linear extension is any ranking of all the objects that is admissible (or in other words compatible with the poset ordering and resulting Hasse diagram). Each of these linear extensions are given one vote in ranking the objects, and the total number of votes that rank an object first, second, etc are recorded. Then the cumulative rank frequency (CRF) for an object can be computed by finding the proportion (or number) of voters who ranked the object *at most* first, second, etc. In other words, the cumulative rank frequency for an object, would be the proportion of voters who voted the site with rank 1, with rank 1 or 2, with rank 1, 2, or 3, etc. This can then be considered as another poset on the basis of the CRF with the ranks as the indicators, if object A has better or equal CRF value for every rank (between 1 and the number of objects) than object B, then A is ranked higher than B. Two sites which have identical CRF values for each rank are considered tied, receive the same rank, and are merged together. This process is repeated until the poset is a simple linear ordering (or Hasse diagram is vertical) and there is just one ranking of the sites. [3]

We use a program that has implemented the above procedure in order to compute the Poset ranking. Since the number of linear extensions can be very large, it is realistically impossible to compute all of them, and the program uses Markov Chain Monte Carlo (MCMC) methods to sample linear extensions and compute the ranking.

## 2.2 POSAC

POSAC [6] is a method to reduce the indicators into a smaller number of dimensions, with the goal of correctly preserving as many of the comparabilities that existed in the original model as possible.

The goal of the POSAC method is to reduce an N-dimensional data matrix by plotting it into two-dimensional space. The two-dimensional coordinate representation of objects with observed profiles should best preserve profile order relations as POSAC constructs new axes, which correctly presents as many of the order relations as possible. POSAC is similar to Principal Components Analysis (PCA) in that they are both dimension reduction methods, but while PCA tries to preserve distances, POSAC tries to preserve ranks.

There are three possible order relations in a two-dimensional Cartesian coordinate space. The possibilities are indicated in Figure 4. A given object a divides the indicator space into four quadrants. The objects that fall in the first quadrant are intrinsically better than a, and those that fall in the third quadrant are intrinsically worse than a. The second and fourth quadrants are regions of ambiguity, objects falling here are incomparable with a.

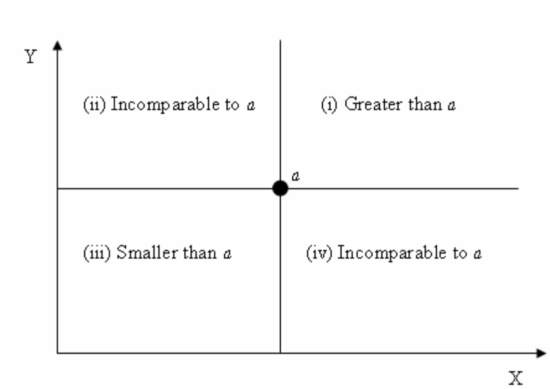


**Figure 4 – Two dimensional ordering**

In an N-dimensional data matrix, we want to form a partially ordered set by replacing the original dataset using their ordering before comparing profiles. In the partially ordered set, some pairs of profiles may be ordered or comparable while some pairs of profiles are incomparable. Consider an example of three profiles with four attributes: 3142, 3242, and 1118. Here 3142 implies that the first attribute has value 3, the second attribute has value 1, etc.

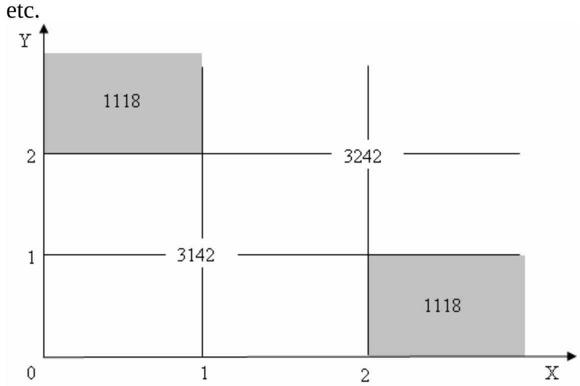


**Figure 5- Ordering of three objects**

Profiles 3142 and 3242 are ordered with 3242 greater than 3142, but 3242 and 1118 are incomparable, since the first attribute is better for the first profile (as 3>1), and the fourth attribute is better for the fourth profile (as 2<8). If two profiles are comparable, say 3142 and 3242, then it can be represented (or preserved) if we assign just a single score to every profile in the pair. For example, let us assign 1 to 3142 and 2 to 3242. Then 2 > 1 reflects the fact that 3242 > 3142. If two profiles are incomparable, say 3242 and 1118. Assigning just one numerical value to each profile cannot represent the fact that they are incomparable, because the set of all numerical values is totally ordered. However, the set of all pairs of numerical values is a partially ordered set. So, let us assign two values to each profile of the incomparable pair to represent their incomparability. Let us first locate the comparable profiles in the plot. For example, assign to 3142 the shorter profile (1, 1), to represent that 3242 is greater than 3142, it needs to be assigned somewhere in the upper right square to (1, 1), say (2, 2). Now, we add profile 1118 to the plot. Since this profile is incomparable to both profiles 3142 and 3242, it must be assigned within the intersection of regions that are incomparable to both (1, 1) and (2, 2). That is the shaded area in Figure 5. For example, we can pick the point (3, 0) to represent profile 1118. The incomparability of (3, 0) with (1, 1) and (2, 2) represents that of 1118 with 3142 and 3242.

The POSAC algorithm usually results in some profiles being unable to be accurately located in the two-dimensional coordinate space. With a large number of profiles, misrepresentation becomes a potential liability of POSAC. In order to measure how well POSAC retains comparabilities from the original data set, we compute the proportion of comparabilities correctly represented, if a pair of objects were comparable in the original data set, then they would have to be comparable with the correct orientation in the POSAC diagram in order to be considered correctly represented. Similarly if a pair of objects is incomparable in the original data set, then they would have to be incomparable in the POSAC diagram as well. We would like the proportion of comparabilities correctly represented to be as high as possible, and a proportion above 0.75 is considered rather good for large data sets.

For the examples in this paper we used the program package SYSTAT 11 [5] in the feature of Analysis in the toolbar, under Scale. The POSAC program produces a two-dimensional diagram with the objects represented and also provides the proportion of comparabilities that are correctly represented. The program is aimed at minimizing the loss of comparabilities. More details of the theory of POSAC can be found in [4] and [6].

## 2.3 Data based weighting using POSAC:

We refer to the dimensions of the POSAC diagram as latent order variables [6], or LOV1 and LOV2 for short, and each object has a LOV1 and LOV2 value corresponding to the POSAC diagram. Due to the interest in understanding the strength of the influence of the original indicators on the LOVs, as well as the need to compute data based weights, we compute loadings. The loadings are computed for each indicator, and a loading gives a measure of similarity between the LOV and the data from a particular indicator. To allow for small deviations in the POSAC algorithm, we discretize both the LOVs and the original data into several equally spaced intervals. In general, the number of intervals depends on the data, some data sets already have discretized the indicators. For the ASC data set, we chose to discretize the data values and the LOVs into 8 intervals, such that values between 0 and 0.125 would get a score of 1, between 0.125 and 0.250 get a score of 2, etc.

In this paper, we compute a "concordance" value for each indicator by counting the proportion of watersheds that the indicator and LOV give the same scored value. If the number of different intervals that the data were discretized is large, then we may add the watersheds where the indicator and LOV's scored values differ by one, which we do for our data.

Using the concordance method to determine loadings for the POSAC LOVs, we generate weights for the indicators which are solely based on the data without any input from the scientists. To determine these data based weights, we use the loadings using POSAC (using two LOVs), which gives two values for each indicator. We then average these values and normalize the resulting average values to sum to one, resulting in a set of weights for the indicators for the concordance method.

## 2.4 Performance measures using Level III

One of our goals in this study, as stated above, is to determine whether we can use Level I or Level II data instead of collecting expensive Level III data. For the purposes of this study, we consider some combination of the Level III values to be the "truth", as this is the most accurate data that we have available.

In order to determine the effectiveness of the data based weights, we need a performance measure that will allow us to compare the rankings derived from the Level II rankings to the "truth". The performance measure is the correlation of the ranking obtained by the Level II indices with the ranking obtained by the Level III indicators of the six watersheds which have Level III data.

We can use the composite indices using the data based weights as found above to obtain a ranking of the 21 objects, and in turn, an induced ranking of the six Level III watersheds. Then we can measure the performance of the SWR index and the data based indices, in their similarity with the Level III data.

There is a need for some discussion on how to determine the "true ranking" from the Level III data. Specifically, the weights of the three indicators to determine an index for the Level III data are an issue. Two indicators measure biological condition, and one measures chemical condition. It may be the case that only the biological indicators are relevant, or that the chemical indicator needs to be included as well. The scientists who collected the data have indicated some skepticism about the relevance and accuracy of the indicator that measures chemical condition in determining the health of the watershed, while not rejecting the indicator outright [1]. Due to this ambiguity, we have decided to consider both approaches, using the biological indicators alone, and using all three indicators. Obtaining additional insight on the relevance of this indicator would be quite useful to scientists and government agencies to determine if this indicator is to be used.

**Table 4 – Weight selection of Level III**

| Scheme | Benthic IBI | Fish IBI | $NO_3$ |
|---|---|---|---|
| 1 | 1 | 0 | 0 |
| 2 | 0 | 1 | 0 |
| 3 | 0 | 0 | 1 |
| 4a | ½ | ½ | 0 |
| 4b | 2/3 | 1/3 | 0 |
| 4c | 1/3 | 2/3 | 0 |
| 5a | 0.45 | 0.45 | 0.1 |
| 5b | 0.4 | 0.4 | 0.2 |
| 5c | 0.35 | 0.35 | 0.3 |

We combined the three Level III indicators using various weights to see what effect of different sets of weights have on the indicators, and the performance of the data based weights. We discuss a total of nine schemes to determine the "true rankings" which are given as follows (see Table 4). In scheme 1, 2, and 3, we give the full weight to one of the indicators. For scheme 4 (4a, 4b, and 4c), we are only interested in using the two biological indicators, and we treat the chemical indicator as unimportant.

We use equal weight between the two biological indicators as well as a two to one ratio in weighting between Benthic IBI and Fish IBI. In scheme 5 (5a, 5b, and 5c), we started with the default in scheme 4a of giving equal weight to the two biological indicators and then chose to "inject" an increasing weight of the $NO_3$ indicator to observe how much the addition of the chemical indicator changes the performance of the Level II indices.

In all these methods, we compute the values of the indices for the six watersheds, and then rank the watersheds using the index values to obtain a "true" ranking of the six watersheds.

Another way to deal with this issue is to avoid combining the indicators into an index and to simply use the Poset rankings for the Level III data. This approach is explored as a possible alternative to compute the "gold standard" due to the controversial nature of selecting weights of the index, as often there can be wide disagreement in selecting weights [3]. We can compute the Poset rankings for all three Level III indicators, or for just the two biological indicators, benthic IBI and fish IBI (see Table 5).

**Table 5–Poset rankings for Level III indicators**

| WATERSHED | Poset (full) | Poset (biological indicators) |
|---|---|---|
| Back River | 4.5 | 6 |
| Cattail Creek | 3 | 2 |
| Gwynn Falls | 4.5 | 5 |
| Saint Mary's A | 1 | 3 |
| Southeast Creek | 6 | 4 |
| Upper Patuxent | 2 | 1 |

# 3. RESULTS

## 3.1 Level III rankings and the schemes

Using the WHASSE program [6], we compute the Hasse diagram for the Level III data. From the Hasse diagram in Figure 6, we can see that the maximal watersheds are Cattail Creek, St. Mary's A, and Upper Patuxent. The second tier of the Hasse diagram consists of the remaining three watersheds, Back River, Gwynn Falls, and Southeast Creek.

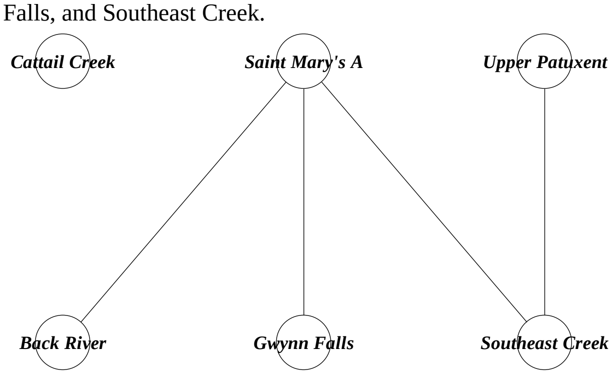


**Figure 6- Hasse diagram for all 3 Level III indicators**

If we only considered the two biological indicators, benthic IBI and fish IBI, then we have a Hasse diagram with two indicators, as shown in Figure 7. The watersheds Cattail Creek and Upper Patuxent are the maximal elements in this Hasse diagram, while the other four watersheds are in a linear ordering.

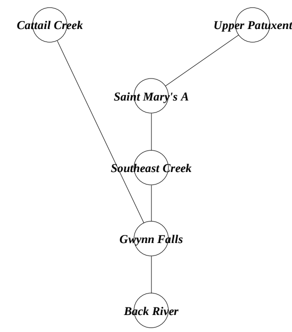


**Figure 7- Hasse diagram for the biological Level III indicators**

We use the Poset prioritization program to obtain the Poset rankings for the six watersheds for both of these situations above. In addition, we compute the index values and the ensuing rankings of the six watersheds for the different schemes in Table 4. The rankings computed by these schemes are in Table 6, while the Poset rankings are in Table 5.

**Table 6–Level III Rankings for different schemes**

| WATERSHED | 1 | 2 | 3 | 4a | 4b | 4c | 5a | 5b | 5c |
|---|---|---|---|---|---|---|---|---|---|
| Back River | 6 | 6 | 2 | 6 | 6 | 6 | 6 | 6 | 6 |
| Cattail Creek | 1 | 4 | 6 | 2 | 2 | 3 | 3 | 3 | 4 |
| Gwynn Falls | 5 | 5 | 3 | 5 | 5 | 5 | 5 | 5 | 5 |
| Saint Mary's A | 3 | 2 | 1 | 3 | 3 | 2 | 2 | 1 | 1 |
| Southeast Creek | 4 | 3 | 5 | 4 | 4 | 4 | 4 | 4 | 3 |
| Upper Patuxent | 2 | 1 | 4 | 1 | 1 | 1 | 1 | 2 | 2 |

## 3.2 POSAC results

We apply the POSAC method to both Level I and Level II data sets. For the Level II data, the POSAC diagram is in Figure 8. 84.6% of the comparabilities are correctly represented by the reduced POSAC model, which indicates that POSAC is a good reduction for the full seven indicator data set.

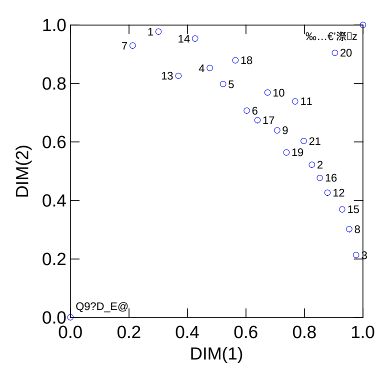


**Figure 8– POSAC Plot for Level II data**

For the Level I data set, 89.6% of the comparabilities are preserved by the two dimensional POSAC model, and the two dimensional POSAC diagram is in Figure 9. The high number of comparabilities preserved suggests that the POSAC model is a good approximation for the five indicator data set.

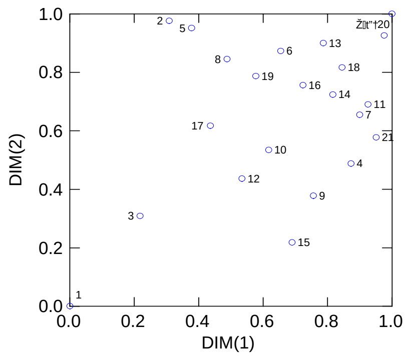


**Figure 9- POSAC Plot for Level I data**

We then compute the loadings for both LOV1 and LOV2 for both Level I and Level II. We use 8 different categories to discretize both the LOV and the indicator, and both will take an integer between 1 and 8. We use the concordance method that computes the proportion of objects which the indicator and LOV value differ by at most one. The loadings provide some insight into the LOVs and their relationship to the indicators.

**Table 7–Loadings using concordance method for Level I indicators**

| | LOV1 | LOV2 |
|---|---|---|
| FOR | 0.428571 | 0.428571 |
| LDI | 0.857143 | 0.619048 |
| IMP | 0.428571 | 0.380952 |
| MPAT | 0.571429 | 0.476191 |
| CORFOR | 0.380952 | 0.52381 |

The loadings using the concordance method described above for both LOVs for Level I are shown in Table 7. We see that LOV1 is most impacted by indicators LDI, which the Landscape Density index of the watershed, and MPAT, which is the mean forest patch size of the watershed. LOV2 is most impacted by indicators LDI and CORFOR, which is the percentage of the total forest in the watershed that is core forest. The two LOVs have a correlation of 0.33, which suggests on its face a positive correlation between the two latent order variables. However, one can see from the plot in Figure 9, that if the plot is divided into regions whose boundaries are negatively sloped, there is a clear negative correlation between LOV1 and LOV2.

**Table 8–Loadings using concordance method for Level II indicators**

| | LOV1 | LOV2 |
|---|---|---|
| BUF | 0.380952 | 0.571429 |
| IR | 0.761905 | 0.333333 |
| BA | 0.428571 | 0.380952 |
| INV | 0.619048 | 0.476191 |
| SHA | 0.571429 | 0.476191 |
| SS | 0.333333 | 0.428571 |
| FPWL | 0.190476 | 0.476191 |

For Level II, as we can see from Table 8, LOV1 is most impacted by indicators IR, which is the incision ratio, INV, which is invasive cover class, and SHA, the stream habitat assessment score. LOV2 is most impacted by BUF, which is the buffer score of the watershed, and to a smaller extent by INV, SHA, and FPWL. Unlike the Level I diagram, it appears that all the objects lie close to a negatively sloped line. LOV1 and LOV2 have a correlation of -0.813, which shows a strong negative correlation, but however it still appears that both LOVs are relevant.

## 3.3 Data based weights and rankings

As we described in section 3.2, we compute indicator weights derived solely from the data, rather than the insights of the investigators. Looking at the data based weights from concordance method from Table 9, we see that the data based index gives approximately the same weight to all the indicators with a little less weight to Stream Stressor and FPWL Stressor. When compared to the investigator-based SWR index, the concordance index gives less weight to the indicators IR, SHA, and SS than does SWR, since SWR gives all three of these indicators a weight of 0.25.

**Table 9–Weights for Level II indicators**

| Indicator | Concordance | SWR |
|---|---|---|
| *1.Buffer Score* | 0.150 | 0.063 |
| *2.Incision Ratio* | 0.154 | 0.250 |
| *3.Basal Area* | 0.137 | 0.063 |
| *4.Invasive* | 0.169 | 0.063 |
| *5 SHA* | 0.159 | 0.250 |
| *6.Stream Stressor* | 0.126 | 0.250 |
| *7.FP-WL Stressor* | 0.106 | 0.063 |

Looking at the rankings of the six selected watersheds from Table 10, we see that the rankings agree that watersheds Saint Mary's A, Southeast Creek, Upper Patuxent are the three best watersheds, and that the watersheds Back River, Cattail Creek, and Gwynn Falls are the three worst.

**Table 10–Rankings of watersheds from Level II indices**

| Watershed | Concordance | SWR |
|---|---|---|
| *Back River* | 6 | 6 |
| *Cattail Creek* | 4 | 4 |
| *Gwynn Falls* | 5 | 5 |
| *Saint Mary's A* | 1 | 2 |
| *Southeast Creek* | 3 | 3 |
| *Upper Patuxent* | 2 | 1 |

## 3.4 Performance Analysis of SWR and the Data Based Indices

We now compare the performance of the data based set of weights from the concordance method of computing loadings using POSAC. The performance measure we use is the correlation between the rankings of the six selected watersheds as derived by the Level II index, (whether from the data based weighting scheme or the investigator based SWR), and the "true ranking" derived from the Level III indicators. The correlations (and p-values) between the Level II index rankings and the Level III rankings, where "true ranking" is assumed to be the Poset ranking using the Level III indicators, or the Poset ranking using only the biological indicators are shown in Table 11.

For the Poset ranking for which all three indicators are included, SWR gives a correlation of 0.580 with a p-value 0.228. The data based concordance index performs better than SWR in this situation, with a correlation of 0.638 and a p-value of 0.173. Neither one of these correlations is significant.

If only the biological indicators are used we see that the SWR index yields the better performance with a correlation value of 0.829 and a p-value of 0.042, while the data based concordance index has a correlation of 0.714 and a p-value of 0.111. It is fairly clear that omitting the $NO_3$ indicator results in much better performance for the Level II indices.

**Table 11–Performance of Level II index rankings with Level III Poset rankings**

| Level III | Concordance | SWR |
|---|---|---|
| Poset (full) | 0.638 | 0.580 |
| p-value | 0.173 | 0.228 |
| Poset (bio indicators) | 0.714 | 0.829 |
| p-value | 0.111 | 0.042 |

Next we consider other schemes for obtaining the Level III ranking, in particular using just a single Level III indicator, using a combination of the two biological indicators, and slowly increasing the amount of the $NO_3$. From Table 12, we see that for schemes 1, 2, and 3, which are single Level III indicators, the Level II indices have the highest correlations using scheme 2, Fish IBI alone, and then with only Benthic IBI. If only the chemical indicator $NO_3$ is used as the Level III standard, as in scheme 3, then the Level II indices do a very poor job in ranking the six selected watersheds.

For schemes 4a, 4b, and 4c, we use only the two biological indicators, using different weights. Using the weighting scheme for 4a and 4b results in the same ranking of the six watersheds, and thus the performance of any Level II set of weights will be the same for both scheme 4a and 4b. Scheme 4c, which gives 1/3 to benthic IBI and 2/3 to fish IBI, results in the both SWR and the data based Level II set of weights having a higher performance as compared with scheme 4a and 4b. It appears as a whole that both SWR and the index derived from the concordance method perform well in estimating the rankings of the watersheds when the Level III biological indicators are used.

For schemes 5a, 5b, and 5c, we incrementally increase the weight of the $NO_3$ indicator. From Table 12, we can see that the rank of the selected watersheds from the SWR index has a high correlation with all three schemes, but the correlation decreases as $NO_3$ indicator is given more weight. Conversely, the data based weights using concordance seem to have an increased correlation when $NO_3$ is given more weight, enough so that the data based index actually has a higher performance than does the SWR index for scheme 5c.

**Table 12–Performance correlations and p-values of Level II data based weights with Level III schemes**

| Level III Scheme | Concordance | p-value. | SWR | p-value |
|---|---|---|---|---|
| 1 | 0.6 | 0.208 | 0.657 | 0.156 |
| 2 | 0.943 | 0.005 | 1 | 0 |
| 3 | 0.086 | 0.872 | -0.086 | 0.872 |
| 4a/4b | 0.714 | 0.111 | 0.829 | 0.042 |
| 4c | 0.886 | 0.019 | 0.943 | 0.005 |
| 5a | 0.886 | 0.019 | 0.943 | 0.005 |
| 5b | 0.943 | 0.005 | 0.886 | 0.019 |
| 5c | 0.943 | 0.005 | 0.886 | 0.019 |

The SWR index, as well as the data based index show high correlations with the Level III index when the index is a combination of the biological indicators. It is clear that the $NO_3$ indicator behaves in a different manner than the two IBI indicators, and the Level II indices rank the six watersheds in a different manner than the $NO_3$ indicator. The addition of a substantial weight of $NO_3$ indicator into the index results only in small changes the ranking of the six selected watersheds. The Level III index rankings using IBI indicators appear to be mostly robust to a small increase in the weight of $NO_3$ to the index, however, when the weight of $NO_3$ is increased, the performance of data based weighting scheme improves slightly while the performance of SWR decreases slightly.

## 3.5. Level I Results

We have up to now considered only the possibility of approximating the true rankings using Level II data. However, we do have Level I data, which is even less expensive and easier to collect than Level II data. We will now apply the methods that we have described in Section 2 to the Level I data. We will also compare the Level I results to the Level II results. In Table 13, the rankings of the six selected watersheds using the indices derived from the data based concordance weights, as well as from the investigator based Landscape Index (LI) are given. From inspection, the Landscape Index differs from the data based concordance index only by ranking Southeast Creek third, and Cattail Creek fourth, which is the opposite ranking that the concordance index provides for this pair of watersheds.

**Table 13- Rankings of Level I indices**

| WATERSHED | Concordance | LI |
|---|---|---|
| Back River | 6 | 6 |
| Cattail Creek | 3 | 4 |
| Gwynn Falls | 5 | 5 |
| Saint Mary's A | 1 | 1 |
| Southeast Creek | 4 | 3 |
| Upper Patuxent | 2 | 2 |

We next use the performance measure of computing the correlation of the ranking of the six watersheds using Level I data and the ranking using Level III data for the different schemes of Level III indicators. We include the correlations and the p-value between the Level I rankings and the Poset rankings of all three Level III indicators, denoted as full, and the two biological indicators, denoted as biological. The results (correlations and p-values) are in Tables 14 and 15.

**Table 14–Performance correlations of Level I indices with Level III schemes**

| Scheme | Concordance | LI |
|---|---|---|
| #1 | 0.771 | 0.6 |
| #2 | 0.886 | 0.943 |
| #3 | 0.029 | 0.086 |
| #4a/4b | 0.829 | 0.714 |
| #4c | 0.943 | 0.886 |
| #5a | 0.943 | 0.886 |
| #5b | 1 | 0.943 |
| #5c | 0.943 | 1 |
| Poset (full) | 0.812 | 0.638 |
| Poset (bio indicators) | 0.829 | 0.714 |

For the Level III schemes that include the two biological indicators, both the Level I concordance index as well as the Landscape Index perform fairly well. It appears that much like the Level II indices, the Level I indices perform better when the fish IBI (scheme 2) is used as the Level III standard as opposed to the benthic IBI (scheme 1). The Landscape index performs better than the data based index for scheme 2, but not in scheme 1. When both biological indicators are included in the index as in schemes 4a/4b/4c, the data based index performs better than the Landscape index. Both indices have higher performance for scheme 4c as compared to 4a/4b, and this is not surprising because 4c give 2/3 weight to the Fish IBI indicator.

**Table 15–Performance p-values of Level I indices with Level III schemes**

| Scheme | Concordance | LI |
|---|---|---|
| #1 | 0.072 | 0.0208 |
| #2 | 0.019 | 0.005 |
| #3 | 0.957 | 0.872 |
| #4a/4b | 0.042 | 0.111 |
| #4c | 0.005 | 0.019 |
| #5a | 0.005 | 0.019 |
| #5b | 0 | 0.005 |
| #5c | 0.005 | 0 |
| Poset (full) | 0.05 | 0.173 |
| Poset (bio indicators) | 0.042 | 0.111 |

When increasing amounts of $NO_3$ are included along with equal weight of biological indicators in the Level III index, as in scheme 5a/5b/5c, both the Landscape index as well as the data based index actually perform even better than the level III schemes that include only biological indicators. A weight of 0.2 of $NO_3$ (scheme 5b) results in the exact same rankings for both the data based index and the Level III index, and thus a performance of 1. The same thing happens for the Landscape Index with scheme 5c, while the performance of the data based index drops slightly. This phenomenon is different than the Level II indices where the data based index performed slightly better as an increasing weight of $NO_3$ was added, while the investigator based SWR performed worse. Since all the indices do very poorly when $NO_3$ is the Level III scheme (as in scheme 3), this trend is only valid as long as the weight of $NO_3$ remains minimal, while if the weight of $NO_3$ is too large, the performance will undoubtedly become worse.

The Level I data based indices also perform better than the Landscape index when the Level III scheme is the Poset ranking, whether or not $NO_3$ is included. The data based indicators yield significant correlations with p-values of equal of less than 0.05, with the Level III Poset rankings, while the correlations for the Landscape Index are not significant with p-values greater than 0.10. Both indices also perform better without the $NO_3$ indicator.

Since we have verified that the Level II SWR and the index derived from the data based concordance method perform well in predicting Level III, we would like to see how well the Level I data based indices and the Landscape Index compares with both SWR and the concordance method. We would also like to observe if there are any differences between the data based methods for Level I. Since we have rankings for all 21 watersheds (see Table 18), and not just for six, we find the correlation between the ranks for all 21 watersheds. The results (correlation and p-values) are shown below in Tables 16 and 17.

**Table 16–Correlations of the rankings of the Level II indices and Level I indices**

| | SWR | Concordance |
|---|---|---|
| Landscape Index | 0.349 | 0.461 |
| Concordance | 0.325 | 0.434 |

**Table 17–P-values of the rankings of the Level II indices and Level I indices**

| | SWR | Concordance |
|---|---|---|
| Landscape Index | 0.121 | 0.035 |
| Concordance | 0.151 | 0.049 |

Among Level I data based indices, the correlations with both the Level II SWR and concordance indices are very similar. The Landscape Index has slightly higher correlations with both the Level II indices. The Level I indices have higher correlations with the concordance index than SWR, and the correlations with the concordance method are significant with p-values less than 0.05, while the correlations with SWR are not significant.

## 4. CONCLUSIONS

We have explained POSAC, a very important and useful method of dimension reduction, which reduces the data to two latent order variables. One can compute loadings of the indicators using the concordance method for both of the LOVs, which in turn computes a set of weights for the indicators. These weights are completely determined by the data, without expert intervention.

We have identified a performance criterion to measure how well a Level II index will predict the true (Level III) rankings of the watersheds. The performance measure simply is the correlation between the rankings of the six selected watersheds derived by Level II index, and the ranking using the Level III data. From this measure we understand whether the rankings of the Level II index are a good approximation for the true rankings. If so, then we can use the rankings from that index rather than collect time consuming and expensive Level III data.

The SWR index of Level II indicators, determined by expert investigators, appears to approximate the two Level III biological indicators fairly well. The correlation of the rankings derived from the SWR index and from the Level III biological indicators are high, and except for scheme 1, where the benthic IBI indicator is used as the Level III index, the correlations are significant.

For the data based weighting scheme using the concordance method, we again used the above performance measure, and found that the concordance method performs quite well also.

We considered the Level I indicators as well, to see if the surveillance based indicators could perform well enough to approximate the condition of the watershed. We found that the data based Level I indices to be very similar to the rankings obtained from the Level III biological indicators. We also found a significant correlation between rankings of the Level I indicators and the Level II concordance method, and a substantive correlation with the ranks from the SWR index using all 21 watersheds that have data for the Level I and Level II indicators.

From our analysis, it is clear that the Level I and Level II indicators do a good job of approximating the biological Level III indicators. When biological indicators alone are considered, the correlation between the Level II and Level I rankings and the

Level III rankings are quite high. This observation holds for both the investigator based indices and the data based indices. For Level II, the concordance index performs nearly as well as the investigator based SWR, while for Level I, the concordance index actually performs better than the investigator based Landscape Index.

**Table 18- Rankings of all Watersheds using Level I and II**

| | Watershed | Level II | | Level I | |
|---|---|---|---|---|---|
| | | SWR | Conc | LI | Conc |
| 1 | Back River | 19 | 20 | 21 | 21 |
| 2 | Cattail Creek | 12 | 13 | 14 | 13 |
| 3 | Gwynn Falls | 15 | 15 | 20 | 20 |
| 4 | Saint Mary's A | 5 | 2 | 7 | 7 |
| 5 | Southeast Creek | 7 | 9 | 13 | 14 |
| 6 | Upper Patuxent | 4 | 5 | 10 | 11 |
| 7 | Ahoskie | 18 | 18 | 4 | 4 |
| 8 | Buffalo Creek | 11 | 12 | 11 | 10 |
| 9 | Chickahominy | 14 | 11 | 16 | 17 |
| 10 | Christian Creek | 17 | 16 | 15 | 15 |
| 11 | Clearfield Creek | 9 | 8 | 3 | 3 |
| 12 | Conodoguinet A | 16 | 17 | 18 | 18 |
| 13 | Grindle Creek | 20 | 19 | 6 | 6 |
| 14 | Little Contentne | 6 | 4 | 8 | 8 |
| 15 | Mantua | 3 | 6 | 19 | 19 |
| 16 | Middle Creek | 13 | 14 | 9 | 9 |
| 17 | Middle River | 21 | 21 | 17 | 16 |
| 18 | Pamunkey | 10 | 3 | 5 | 5 |
| 19 | Repaupo | 2 | 7 | 12 | 12 |
| 20 | White Deer Creek | 1 | 1 | 1 | 1 |
| 21 | Wisconisco | 8 | 10 | 2 | 2 |

However, Level II and Level I indicators represent the chemical indicator, the amount of $NO_3$ very poorly. The rankings of the watersheds using the Level I and II indices are effectively uncorrelated with the rankings of the watersheds using $NO_3$. When the $NO_3$ indicator is given a low weight, the Level II and Level I indicators are robust to a small amount of $NO_3$, but the ranking of the objects change slightly. The inclusion of a small weight of $NO_3$ changes the order of quality of the indices. Increasing the weight of $NO_3$ appears to result in the Level II concordance index outperforming the SWR index, and the Level I Landscape Index outperforming the Level I concordance index, while giving zero weight to $NO_3$ results in the opposite orientation in the comparison of the performance of the two indices.

In Table 18, the rankings of all 21 watersheds are listed using Level I and II indices. While there appears to be some general agreement among some watersheds between Level II and Level I rankings, there are also substantial differences.

As long as we are interested primarily in the biological condition of the watersheds, it would be advisable to use the Level II or even the Level I indices to estimate the true ranking of the 21 watersheds, which will save much time and effort for the scientists, and money from the government funding agencies. The determination of these rankings will help government agencies select the watersheds that are in good health and should be preserved, and the watersheds that are in poor health and are candidates for restoration at minimal cost.

## ACKNOWLEDGEMENTS

This material is based upon work supported by (i) the NSF Grant No. 0307010, and (ii) the US EPA Grant. Nos. R-828684, CR-83059301, and RD-8324400l. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the agencies. I give my greatest thanks to my research advisor, Dr. G.P. Patil for his guidance and support in this endeavor. In addition, I would like to thank Michael Kase, who was very helpful in computing many of the results in this paper. Lots of thanks are due to Robert Brooks and Mary Easterling for providing data and sharing insights.

## REFERENCES

[1] Brooks, R.P, McKinney-Easterling, M., Brinson, M., Rheinhardt, R. , Havens, K. , O'Brien, D., Bishop, J. Rubbo, J., Armstrong, B., and Hite, J. (2006). A Stream-Wetland-Riparian (SWR) Index for assessing condition of aquatic ecosystems in small watersheds along the Atlantic Slope of the eastern U.S. (Manuscript)

[2] Brüggemann, R. and Simon U. (2006) Simple Posetic Structures to Analyze Water Management Strategies in Urban Regions.

[3] Patil, G.P. and Taillie, C. Multiple indicators, partially ordered sets, and linear extensions: Multi-criterion ranking and prioritization. Environmental and Ecological Statistics 11 (2004) 199-228

[4] Shye S. *Multiple Scaling*. (1985) Elsevier Publishers: Amsterdam.

[5] SYSTAT (2004) http://www.systat.com

[6] Voigt, K., Welzl, G., and Brüggemann, R. Data analysis of environmental air pollutant monitoring systems in Europe. *Environmetrics* 15 (2004) 577-596