PENNSTATE

Center for Statistical Ecology and Environmental Statistics

Unmasking Weight Camouflage of a Composite Index based on Multiple Indicators

By G.P. Patil and K. Sham Bhat

[2]Center for Statistical Ecology and Environmental Statistics, Department of Statistics,
The Pennsylvania State University, University Park, PA 16802, USA

This material is based upon work partially supported by (i) the National Science Foundation under Grant No. 0307010, (ii) The United States Environmental Protection Agency under Grant No. CR-83059301 and (iii) The United States Environmental Protection Agency under Grant No. R-828684
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the agencies.

[In Preparation]

Technical Report Number 2006-0803
TECHNICAL REPORTS AND REPRINTS SERIES
August 2006

Department of Statistics
The Pennsylvania State University
University Park, PA 16802

G. P. Patil
Distinguished Professor and Director
Tel: (814)865-9442   Fax: (814)865-1278
Email: gpp@stat.psu.edu
http: //www.stat.psu.edu/~gpp
http://www.stat.psu.edu/hotspots
DGOnline News
Environmental and Ecological Statistics-Springer

Unmasking Weight Camouflage of a Composite Index based on Multiple Indicators

By G.P Patil and K. Sham Bhat
Center for Statistical Ecology, Department of Statistics
Pennsylvania State University , University Park, PA, 16802 USA

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 may be used to combine the scores of multiple investigators/indicators to serve as a criterion for ranking the objects. Our goal is to understand the effect of the indicators and of the weighting schemes on the ranking. We compare data based weighting schemes with investigator based weighting schemes through their comparability acquisition profiles. We use elements of Poset (Partial order set) theory and the Hassediagrams as a basis for this analysis.

We use two data analysis methods to better understand the meaning behind the weights. One is POSAC, a method generally used to reduce the dimensionality of the indicators, while preserving as many of the comparabilities from the data matrix as possible. POSAC plots the data on a two-dimensional plane, and thus produces two latent order variables (LOV). Due to the interest in understanding the strength of the influence of the indicators on the LOVs as well as to obtain the data based weights for each indicator, we compute "loadings". We use the concordance between the LOV and the indicator to compute these loadings and weights. The indicators are discretized to ensure robustness and weights are obtained by normalizing the loadings to sum to one. The second data analysis method, METEOR, uses step-by-step aggregation, with the number of comparabilities of the Hassediagram as a performance metric. We use METEOR for comparing two sets of weights, by comparing the comparability acquisition profiles of the two weighting schemes relative to the nature of aggregation desired. A comparability acquisition profile plots the number of comparabilities at each step of the aggregation. In addition, different scenarios of aggregation for data with hierarchical indicators are explored in the paper.

Four interesting live data sets give us unique insight into the weighting schemes. These data sets include a set of 21 watersheds from the Atlantic Slope Consortium (ASC), with the goal for determining an accurate ranking for their environmental condition based on intensive, rapid, and landscape indicators. The second such data set is from the Federal Highway Administration with 49 bridges with stream crossings with the goal of determining the main causes of bridge failure based on 13 indicators for stream stability at bridge crossings. The third data set is the Environmental Performance Index (EPI) of 102 countries which rates the countries based on their strength of environmental health as well as their ability to manage their ecosystem and natural resources. The fourth data set is from the Eurobarometer survey which estimates the satisfaction of the people in the country for 18 European countries for the performance of services such as telephone services, power utilities, and transport systems. Each one of these data sets has its own challenges, and we critique both the data-based weights as well as the investigator based weights for these data sets.

We finally ask whether there exists a composite index that results in the same ranking of the objects stipulated by one method or the other which does not integrate the multiple indicators of the given data matrix.

Introduction:

Often there is a need to rank objects, whether they are people, cities, watersheds, countries, or bridges when there are multiple criteria of importance whose viewpoints must be considered. 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. The main issue in this approach is determining an appropriate weighting scheme for the model, there being the possibility of underweighting or overweighting particular indicators.

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 Patil and Taillie (2004). Object A is considered intrinsically "better" than object B if all indicators in the model rate object A greater than or equal than object B, with at least one indicator considering object A strictly greater than object B, we denote A>B. If object A is neither intrinsically better than object B, nor intrinsically worse than object B, then we consider the two objects incomparable, we denote as A||B. For every pair of objects which 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 Hassediagram. Each object is denoted by a circle. The Hassediagram technique uses several levels for clarity, object A must be in a higher level than objects that A is better than, and in a lower level 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 otherwords, we remove the maximal elements and find the maximal elements of the objects still remaining. If object A is intrinsically better than object B, then we denote this with a line between A and B.

Another concept that is of interest in this framework is the concept of influential indicators, we consider an indicator influential if it disturbs the overall consensus of the model. In other words, influential indicators often rank the objects in a different manner than the consensus of the rest of the indicators. A performance criterion for determining if an indicator is influential is the W-matrix or a dissimilarity matrix, which is a matrix of the difference of comparabilities with an indicator removed (Voigt et al., 2004). From the W-matrix, one can find the increase in comparabilities if a particular indicator is removed, if that increase in comparabilities is large, then we say that the indicator is influential, since its inclusion resulted in producing a substantial number of additional contradictions. Influential indicators produces additional challenges which need to be further explored. See Voigt et al.(2004), for more information on influential indicators and the W-matrix.

Methods:

POSAC:

A very useful method to use for analysis is the POSAC method (Voigt et al. 2004), which 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. POSAC is a kind of profile analysis. The data in each row of the data matrix is defined as the object's profile. Order relations between profiles can be determined by comparing values for each attribute. Given two different objects, object a is considered to be greater than object b only if all the attributes of object a are greater than or equal to those of object b, with at least one attribute of object a strictly greater than object b. The two profiles are incomparable when there are contradictions in the values of attributes of the objects. 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 observed profiles should best preserve profile order relations (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 . 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 not comparable with *a*.

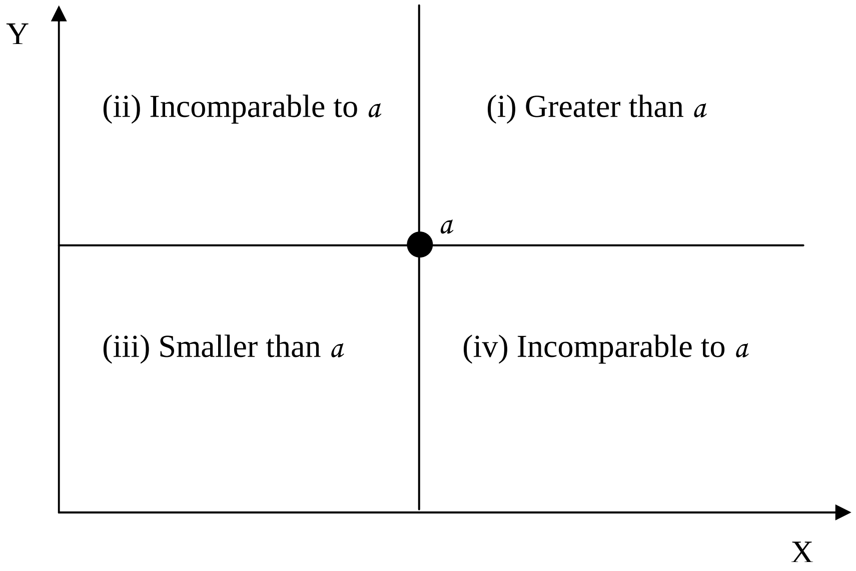


Figure 1

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 may be unordered or incomparable. Consider an example of three profiles with four attributes: 3142, 3242, and 1118.

Profiles 3142 and 3242 are ordered with 3242 greater than 3142, but 3242 and 1118 are incomparable. 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 cannot represent the fact that they are incomparable, because the set of all (single) 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, let us 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 2. 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.

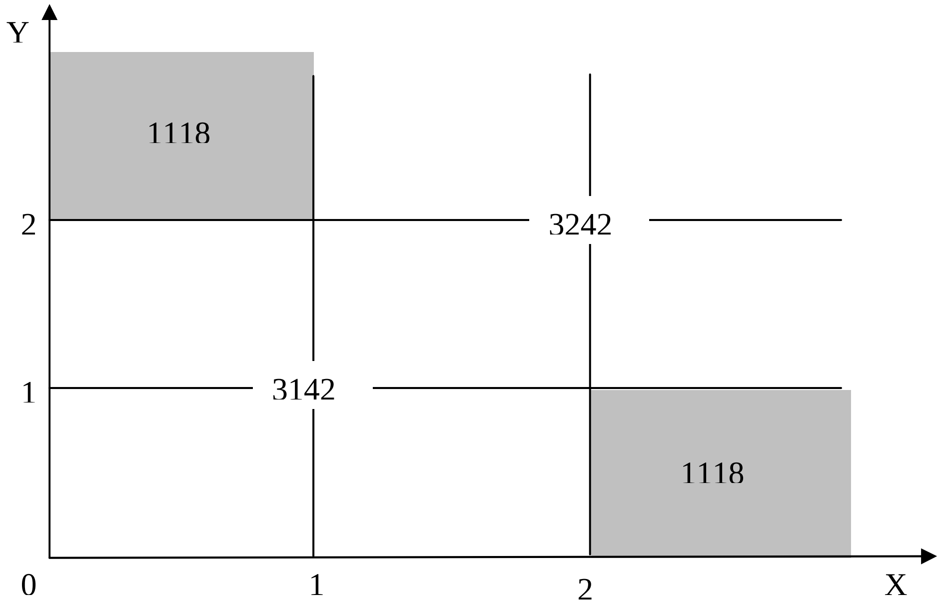


Figure 2

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, above 75% is considered good for large data sets.

For the examples in this paper we used the program package SYSTAT 11 (SYSTAT, 2004) 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 Voigt et al.(2004) as well as in Shye (1985). There is also a multivariate form of POSAC, MPOSAC (Hebrew University of Jerusalem, 2002), where more than two latent order variables can be used.

We refer to the dimensions of the POSAC diagram as latent order variables (Voigt et al. 2004), LOV1 and LOV2 for short, each object has a LOV1 and LOV2 value corresponding to the 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" by several different methods. To allow for small deviations in the POSAC algorithm, we discretized both the LOVs and the original data into several equally spaced intervals, the number of levels depends on the data set, some data sets already have divided the indicators into discrete levels. For example, if we were to discretize into 8 levels, values between 0 and 0.125 would get a score of 1, between 0.125 and 0.250 get a score of 2, etc.

We used three different methods for obtaining the loadings for the indicators as follows:

1. We use the F-value method, where we take the LOV as the response variable, and the relevant indicator for an ANOVA, and took the F-value for that test as the loading.
2. We computed a "concordance" value for each indicator by counting the number of times that the indicator and LOV gave the same scored value. If the number of different intervals that the data were discretized is large, then we may add the objects (or give a value of 0.5) when the indicator and LOV's scored values differ by one.
3. We computed Spearman's rank correlation between the LOV and the data for the indicator.

Data based weighting using POSAC:

We considered the methods for computing loadings above as well as two additional methods,

4. The p-values for each of the F-tests in method 1 were computed, and the reciprocal was taken.
5. The average of methods 1-4 were computed for each indicator.

In all the methods using POSAC, we averaged the values that we obtained from both LOVs from POSAC for each object, and normalized the resulting values to sum to one, resulting in a weighting scheme.

We can obtain an understanding for the accuracy of the different methods by considering the strengths and limitations of the methods. The F-value gives the strength of the relationship between the latent order variable and the indicator. However, since the null hypothesis is that there is no relationship, the F-value is high whether the relationship between the LOV and indicator is positive or negative. This results in high weights for indicators which have a strong inverse relationship with the LOV. The Spearman's correlation between the LOV and the indicator gives either positive or negative values, however, index weights must be non-negative. This problem may be solved by giving zero weight to indicators with negative correlation with the LOV. We are still left with the potential problem that unequally poor indicators are treated the same. The advantage for the concordance method is that it has none of these limitations.

One limitation of the concordance method is often for influential indicators, the loadings tend to understate the significant influence of the indicator. Since the results from other work we have done have also confirmed empirically that the concordance method produces the most effective data-based weighting scheme, we will use only the concordance method for the remainder of this paper.

METEOR:

In order to understand the reasoning behind the weights and to propose more effective weights, we use step-by-step aggregation using METEOR [4]. The idea behind METEOR is that at each step multiple indicators (usually two) are merged together by a weighted linear combination, i.e. $I=g_1*I_1+(1- g_1)* I_2$. Then we replace the two indicators that have been aggregated with the linear combination, i.e. we replace the indicators I_1 and I_2 with the new indicator I, reducing the number of indicators by one. The idea is to eliminate the incomparabilities among the rankings between the two indicators to be aggregated. We select the indicators to combine by choosing the least correlated indicators, again with the goal of eliminating the greatest number of incomparabilities possible. The relative weights for the aggregation step are computed by taking the weights of the two indicators to be combined and normalizing them so they sum to one. Then at each step, we combine the two indicators, including indicators that are a result of previous aggregation, which are the least correlated. The final result is an index of all the indicators. This has been called the top-down method of aggregation.

The other method of aggregation is by taking advantage of a natural hierarchy among the indicators, that is, if there are several natural groups of indicators. There may also be subgroups within these groups that contain more than one indicator. As an instructive example, consider the measurement of four chemicals in the United States. We would like to evaluate and rank the 10 strategies of reducing the amount of the four chemicals. Each township tests the 10 strategies and finds the value of each of the four chemicals for each method. The township can aggregate these four indicators to achieve a final ranking of these ten strategies. Now the county would like to find the best strategy to deal with the proliferation of these four chemicals. It would be reasonable to aggregate the indicators for all the townships in that county for Chemical 1 to get a county level index for Chemical. Having county level indices, which we can also treat as indicators, for all four chemicals, we can then aggregate these four county-level indicators to produce a full ranking of the ten strategies for the county. This approach allows for more levels of hierarchy, as each state can use the county level indicator for each of the four chemicals, aggregate the county level indicators to get state level indicators for the four chemicals, and aggregate those indicators to get a ranking for the state. The process can be continued for the nation with respect to the states, and if one wishes for the world with respect to the nations.

We call this concept as the bottom-up method of aggregation. This method has the advantage of understanding the impact of the group indices and weights and the individual indicators and weights within each group. However, one big drawback of the bottom-up method is that often the indicators within a group or subgroup are often correlated, which results in little reduction in the incomparabilities among the indicators since there is very little incomparability to begin with.

We can use the Hasse Diagrams to see the differences at each step of aggregation. Since Hasse diagrams can be quite large, number of comparabilities of Hassediagrams are used as a performance metric to determine how much agreement in the ranking there exists among the indicators in the model. Two objects are comparable if one object is greater than or equal to the other for all the indicators in the model, and in the Hassediagram, we denote this with a connection. The total number of comparabilities is the number of pairs which are comparable. The aggregation of the two indicators increases the number of comparabilities in the Hassediagram as contradictory rankings between the two indicators are eliminated.

To compare two different sets of weights, we consider their comparability acquisition profiles. A comparability acquisition profile indicates the number of comparabilities at each step of aggregation for a particular weight scheme. The advantage of this profile is that it gives the comparabilities at each aggregation step, which can be used for comparison between two sets of weights. Specifically, we can compare weighting scheme w_1 and w_2, and look at their comparability acquisition profiles at each aggregation step. We prefer the weighting scheme w_1 over w_2 if the comparability acquisition profile of w_1 gives a higher value (number of comparabilities) than the comparability acquisition profile of w_2 at every step of the aggregation. As an example, consider the Bridge data set with the equal indicator weighting scheme and the equal group weighting scheme whose comparability acquisition profile is in Figure 3. It is noticed that the equal indicator weighting scheme’s comparability acquisition profile results in a higher value at each step than the comparability acquisition profile of the equal group weighting scheme. Often, however, we may choose to be more flexible with the criteria and only require that the later steps of aggregation must have greater value as opposed to all steps in the comparability acquisition profile.

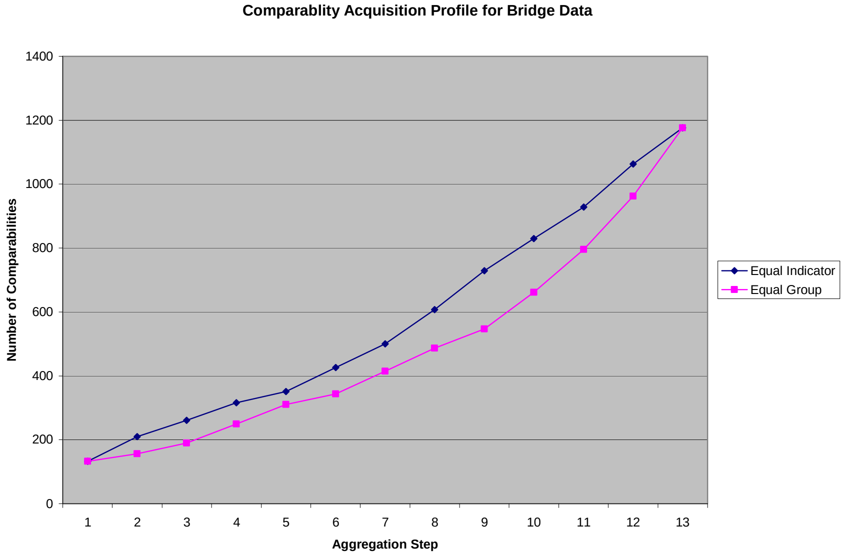


Figure 3

For the application of step-by-step aggregation to the four example data sets in this paper, we used the programming language R to code and run METEOR.

When an indicator is aggregated with the set of already aggregated indicators, we compute the difference of number of comparabilities with the last indicator aggregated from the number of comparabilities before aggregation of the last indicator. From this information, one can understand whether a particular indicator is influential or dissimilar to the other indicators. If the indicator is influential, then it might be worth some effort to question whether the relative weight given that indicator during aggregation is appropriate, and whether the weight should be increased or decreased.

Data Sets:

ASC Watershed:

The first data set that we use for exploration of these methods is a data about a set of 21 watersheds from the Atlantic Slope Consortium (ASC), with the goal for determining an accurate ranking of the health of the watersheds. . In addition, each watershed was identified with one of four physiographic regions, which describes the terrain of the land, and one of six social choices, which describes the use of the land (urban, agriculture, forest, mixed etc.).

This data set has three different levels of indicators, grouped into 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. We have for all 21 watersheds, 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, the data was normalized so every value in the data matrix is a “score” between zero and one. . The three levels are as follows:

Level III – Intensive Field Assessment
Level II – Rapid Field Assessment
Level I – Landscape Assessment from GIS

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

The Level II Rapid Field Assessment is obtained from onsite sampling. Certain level of expertise is involved in the field assessment. Generally, 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.

The analysis done in this paper is mostly done on the Level I and Level II indicators. The Level I indicators and Level II are listed below in Table 4 and Table 5. The data for each of the indicators were normalized to a score between 0 and 1.

Level II:

Variable	**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

Table 4

Level I:

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

Table 5

An effort was made by the investigators to combine the seven level II variables into a composite index, and the five level I variables into an index. These preliminary composite indices are expected to represent the condition of the watersheds in certain degrees.

The composite Level I index is called the Landscape index. It is computed by averaging the land cover 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 Land index is shown in Figure 6.

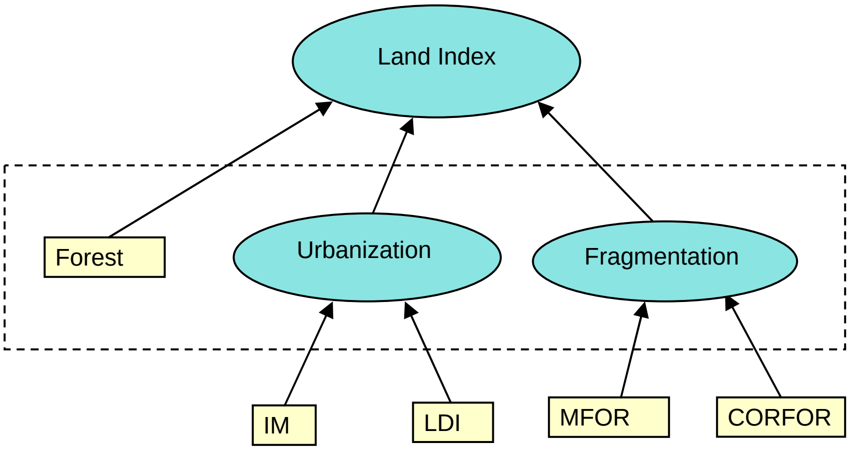


Figure 6

The composite level II index is called stream-wetland-riparian (SWR) index. The index is computed by averaging Floodplain-Wetland(FW) index, incision ratio score(IR), stream habitat assessment score, and stream stressor score, where the FW index is the mean of buffer score(BUF), basal areas score(BA), invasive score(INV), and floodplain-wetland stressor score. The conceptual model of condition used for SWR index is shown in Figure 1.2.

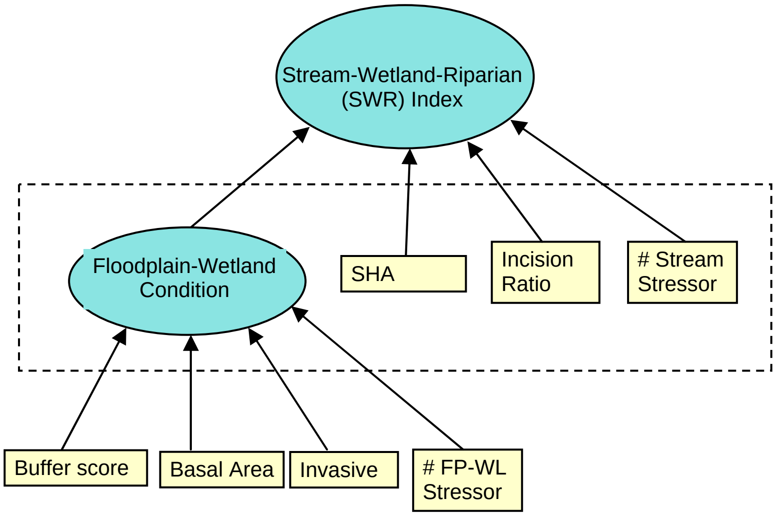


Figure 7

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 level for Level I, Urbanization, Fragmentization, which consists of two indicators each, and the single indicator Forest. For Level II, we have four groups in the top level, the three indicators SHA, Incision Ratio, and Stream Stressor, and Floodplain-Wetland Condition, which consists of four indicators. The data for the Level I indicators is given in Table 8, and the data for the Level II indicators are given in Table 9.

Watershed	Number	FOR	LDI	IMP	MPAT	CORFOR
Back River	1	0.140715	0.066475	0.022456	0.079241	0.070260
Cattail Creek	2	0.315800	0.660926	0.688812	0.296692	0.240528
Gwynn Falls	3	0.245400	0.154413	0.044251	0.182014	0.123783
Saint Mary's A	4	0.694700	0.760678	0.306705	0.633718	0.573722
Southeast Creek	5	0.323720	0.529329	0.665981	0.419773	0.262891
Upper Patuxent	6	0.385400	0.673689	0.644129	0.365625	0.400212
Ahoskie	7	0.701700	0.750497	0.471692	0.664144	0.724929
Buffalo Creek	8	0.303800	0.604162	0.560414	0.392625	0.653580
Chickahominy	9	0.446000	0.441737	0.122522	0.384063	0.366770
Christian Creek	10	0.299400	0.636326	0.445786	0.385581	0.315367
Clearfield Creek	11	0.741100	0.838575	0.553802	0.716695	0.613931
Conodoguinet A	12	0.317100	0.518752	0.203652	0.245001	0.282654
Grindle Creek	13	0.575100	0.679615	0.630803	0.541836	0.665654
Little Contentnea	14	0.585300	0.655572	0.511001	0.536146	0.542742
Mantua	15	0.368300	0.416782	0.138940	0.248248	0.119976
Middle Creek	16	0.428100	0.677209	0.538128	0.432322	0.579040
Middle River	17	0.308700	0.603747	0.338287	0.298532	0.278166
Pamunkey	18	0.618600	0.765106	0.599850	0.669773	0.589294
Repaupo	19	0.354800	0.624655	0.568231	0.379430	0.307367
White Deer Creek	20	0.945400	0.960655	0.615706	1.000000	0.865975
Wisconisco	21	0.845500	0.887929	0.479786	0.762334	0.719116

Table 8

Watershed	Number	BUF	IR	BA	INV	SHA	SS	FPWL
Back River	1	0.604700	0.575100	0.428444	0.202500	0.521500	0.218750	0.703750
Cattail Creek	2	0.357850	0.796550	0.568842	0.652500	0.702250	0.410000	0.279000
Gwynn Falls	3	0.323200	0.767632	0.564050	0.682500	0.676500	0.311250	0.165000
Saint Mary's A	4	0.777400	0.689364	0.599000	0.837500	0.696667	0.635000	0.669000
Southeast Creek	5	0.770100	0.819450	0.581350	0.507500	0.610750	0.571250	0.706000
Upper Patuxent	6	0.682800	0.637450	0.725421	0.612500	0.812250	0.690000	0.542000
Ahoskie	7	0.562965	0.361736	0.519488	0.597500	0.477083	0.400000	0.588500
Buffalo Creek	8	0.352403	0.875346	0.581555	0.860000	0.631000	0.388750	0.244750
Chickahominy	9	0.583337	0.659803	0.643118	0.760000	0.543000	0.452500	0.498000
Christian Creek	10	0.094771	0.594500	0.247863	0.920000	0.704000	0.388750	0.466750
Clearfield Creek	11	0.646642	0.837002	0.155808	0.980000	0.703250	0.395000	0.678750
Conodoguinet A	12	0.323320	0.723568	0.363688	0.647368	0.692895	0.367105	0.214737
Grindle Creek	13	0.456235	0.335350	0.467854	0.590000	0.481353	0.447500	0.616250
Little Contentnea	14	0.519905	0.726832	0.628001	0.822500	0.687745	0.627500	0.743500
Mantua	15	0.522608	0.893311	0.635562	0.759375	0.687813	0.640625	0.345000
Middle Creek	16	0.361110	0.790042	0.451871	0.794737	0.670526	0.378947	0.262368
Middle River	17	0.190991	0.523930	0.266439	0.835000	0.618750	0.250000	0.306750
Pamunkey	18	0.701264	0.573006	0.703947	0.910000	0.671248	0.545000	0.676500
Repaupo	19	0.531330	0.848151	0.567568	0.670588	0.710882	0.758824	0.436765
White Deer Creek	20	0.859772	0.860582	0.801578	0.975000	0.936250	0.795000	0.720000
Wisconisco	21	0.580086	0.807446	0.524934	0.840000	0.716750	0.436250	0.429500

Table 9

Bridge:

One data set used in this analysis consists of 49 bridge sites and 13 indicators that are described by Johnson (2005). There is an important need to determine the causes of the failure of bridges over waterways, and the determination of the indicators most influential in stream stability is vital. The observations and measurements of 49 bridges/streams over 13 indicators have been determined, with each bridge/stream getting a score between 1 and 12 (with 1 best and 12 worst), and the bridges can be ranked by using the Hasse diagram and Poset methods. The principal investigator of this project conjectured that two indicators, watershed and floodplain activity (indicator 1), and bridge channel alignment (indicator 13) would result in the maximum impact on the ranks (Newlin 2006). The 49 bridge sites cross response-type streams. Four different stream classifications are considered response-type streams. They are dune-ripple, riffle-pool, plane-bed, and modified channels. The 13 indicators describe watershed-scale factors, floodplain function, bank stability, and channel features. A list of indicators and their brief descriptions are shown below in Table 10.

Indicator	Description	Level of Expertise	Feature Category
Watershed and Floodplain Activity	Surrounding land use; forested, grazing, urbanization, logging, etc.	1	Watershed or Regional
Flow Habit	Perennial, intermittent, ephemeral streams, flooding behavior, stream order	2	Watershed or Regional
Channel Pattern	Straight, engineered, meandering, braided	2	Watershed or Regional
Entrenchment or Channel Confinement	Connectivity of floodplain with channel, evidence of infrastructure undercutting	2	Watershed or Regional
Bed Material	Sediment size, packed or loose, fraction of sand	3	Local Channel
Bar Development	Narrow or wide, vegetated or newly deposited, grain size of deposited sediment	3	Local Channel
Obstructions	Bedrock outcrops, amour layer, LWD, grade control structures, revetment, vanes	1	Local Channel
Bank Soil Texture	Clay, silt, loam, sand; cohesive or noncohesive	3	Bank Stability
Average Bank Slope Angle	Bank slope for unconsolidated and consolidated materials	1	Bank Stability
Bank Protection	Vegetative (riparian zone width), engineered revetment	1	Bank Stability
Bank Cutting	Percentage of raw banks, undercutting	1	Bank Stability
Mass Wasting or Bank Failure	Scalloping of banks, slumping, tension cracks	2	Bank Stability
Bridge-Channel Alignment	Upstream distance to bridge from meander impact point, bridge alignment with channel flow direction	2	Alignment

Table 10

Each indicator is assigned with a score between 1 and 12 in the following manner: (1-3) ‘Excellent,’ (4-6) ‘Good,’ (7-9) ‘Fair,’ and (10-12) ‘Poor.’ After a score is assigned for each of the 13 indicators, a total score is obtained by a summation of the individual scores. This assumes that each of the indicators has equal weighting and that they independently describe channel stability. The total score is then given a classification of ‘Excellent,’ ‘Good,’ ‘Fair,’ or ‘Poor’ based on threshold values.

Since our goal was to rank the bridges from best to worst, we converted the scores that each indicator was given by subtracting the score from 13, so that the modified scores would go in this order: (1-3) ‘Poor’ (4-6) ‘Fair,’ (7-9) ‘Good,’ and (10-12) ‘Excellent.’ The data is shown below in Table 11

Bridge	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13
1	7	4	6	10	4	7	9	4	10	9	11	11	10
2	7	3	5	9	2	8	11	1	6	5	11	10	10
3	8	4	7	6	4	11	11	4	7	3	8	8	4
4	9	10	8	11	8	11	8	4	11	7	10	9	9
5	6	10	8	11	4	10	9	6	8	8	10	11	6
6	1	1	4	6	2	8	11	2	4	3	5	6	10
7	4	1	3	6	2	5	10	2	5	3	7	6	5
14	5	6	9	4	4	9	7	3	3	4	4	4	7
15	5	6	7	3	2	3	4	1	1	2	1	1	3
16	6	9	8	5	3	7	4	6	2	5	4	3	10
17	1	9	7	1	2	3	3	1	1	2	1	1	10
18	3	7	7	3	5	9	6	2	3	4	6	5	10
19	6	6	7	5	2	3	6	6	5	6	5	6	5
20	7	8	7	8	8	7	4	6	4	4	7	3	5
21	7	6	10	7	4	9	5	5	3	6	8	10	3
22	5	7	7	2	3	5	6	7	1	6	1	4	5
23	8	9	6	1	3	2	4	7	1	4	6	1	5
24	6	8	7	3	2	7	7	8	3	4	8	7	10
25	9	10	11	7	8	9	9	8	3	4	8	2	3
26	10	10	10	6	6	11	9	9	5	5	7	7	10
27	5	8	7	6	7	9	8	8	3	2	7	8	3
28	4	8	7	6	10	9	5	9	3	5	6	9	9
29	10	10	8	6	10	10	6	10	4	2	6	7	9
30	10	10	7	4	11	9	4	4	2	2	2	2	4
31	5	11	9	8	10	8	8	5	8	11	11	11	7
32	9	11	5	11	1	8	5	2	8	8	9	6	9
33	6	9	8	5	6	6	8	6	3	5	9	6	8
34	8	10	5	6	8	5	5	8	5	3	9	4	3
35	8	11	8	9	10	10	9	8	4	8	5	6	6
36	2	4	5	5	4	4	6	7	3	4	3	4	2
37	8	7	8	7	7	8	9	8	5	5	7	7	9
38	10	10	9	7	6	5	8	11	6	8	5	9	2
39	7	9	10	6	3	5	8	10	5	6	9	9	10
40	9	10	9	3	3	4	6	10	2	4	4	3	4
41	5	8	8	9	4	8	9	10	4	4	3	7	2
42	8	11	10	11	11	10	10	9	8	10	9	11	5
43	2	8	8	9	3	1	7	8	3	1	1	3	3
44	3	4	3	6	4	7	8	1	1	5	2	5	1
45	6	9	7	8	8	9	8	11	4	2	4	6	9
48	11	10	10	8	12	11	9	12	8	12	9	11	5
49	10	10	10	8	12	9	9	10	7	8	9	11	6
50	9	10	9	9	10	8	9	8	6	4	9	11	4
51	7	9	9	7	10	8	10	10	5	3	7	5	5
52	9	10	10	10	9	10	10	10	7	12	11	12	6
53	9	10	10	9	12	11	9	10	8	8	8	9	5
54	10	11	8	10	7	11	9	5	7	7	7	11	1
55	8	10	10	8	6	9	9	6	8	8	7	11	10
56	7	10	6	10	8	11	2	7	10	10	7	12	5
57	10	11	10	10	9	10	10	9	8	9	9	12	8

Table 11

EPI:

The Environmental Performance Index (EPI) is focused on two major goals in environmental protection, 1) the reduction of environmental stressors for the improvement of human health, and 2) the support of ecosystem vitality and better management of our natural resources. (Yale Center for Environmental Policy 2006) This first goal is covered by the inclusion of the Environmental Health group and its indicators. The second goal is measured by the five categories that are representative of policies that support ecosystem vitality and resource management. These five categories that help measure attainment of the second goal are Air Quality, Water Resources, Biodiversity and Habitat, Productive Natural Resources, and Sustainable Energy.

The performance of the goals is measured by 16 different components, which were selected by the investigators through a review of existing environmental policy literature, through discussion from the Millennium Development Goal, and the advice of experts in the field. These 16 indicators span the major environmental issues that can be effectively quantified and have available data. We list the indicators with their identification number and group in Table 12 .Unlike the other data sets in this paper, not all the indicators belong solely to one group. I5, or Urban Particulates belongs to both Environmental Health and Air Quality, I8, Water Consumption belongs to groups Water Resources and Biodiversity and Habitat, and I11, Timber Harvest, belongs to Biodiversity and Habitat and Productive Natural Resources.

Indicator	Code	Group(s)
Child Mortality	I1	Environmental Health
Drinking Water	I2	Environmental Health
Adequate Sanitation	I3	Environmental Health
Indoor Air Pollution	I4	Environmental Health
Urban Particulates	I5	Environmental Health/Air Quality
Regional Ozone	I6	Air Quality
Nitrogen Loading	I7	Water Resources
Water Consumption	I8	Water Resources/Biodiversity and Habitat
Wilderness Protection	I9	Biodiversity and Habitat
Ecoregion Protection	I10	Biodiversity and Habitat
Timber Harvest Rate	I11	Biodiversity and Habitat/Prod Nat Resources
Overfishing	I12	Productive Natural Resources
Agricultural Subsidies	I13	Productive Natural Resources
Energy Efficiency	I14	Sustainable Energy
C02 per GDP	I15	Sustainable Energy
Renewable Energy	I16	Sustainable Energy

Table 12

There are 102 countries for which we have data for all 16 indicators, the other countries in the study have data missing for one or more indicators, and we do not include them for now. Thus we have a data matrix of 102 countries (objects) by 16 indicators. Raw data was collected by the investigators for each country for each indicator, and were then scaled on the scale such that a score of 100 would be given if the performance of the country on that indicator met or exceeded the target. (EPI reference) The target was designed to be the consensus agreed by the investigators or at the fifth percentile of the values for that indicator, whichever led to more countries meeting or exceeding the target. A score of 0 was given to the country or countries with the worst performance for that indicator, and the other countries scores were determined by interpolation. The data matrix for the EPI can be found in Table.13

ISO3Cod	#	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16
AGO	1	0.0	9.7	14.9	0.0	18.3	84.1	94.0	90.0	10.6	56.8	100.0	66.7	100.0	87.4	77.8	9.7
ALB	2	91.0	94.6	86.6	85.0	0.0	28.8	99.5	100.0	0.3	0.7	100.0	66.7	100.0	78.7	80.4	41.0
ARE	3	99.0	96.4	100.0	100.0	51.4	25.6	100.0	24.1	0.0	100.0	100.0	33.3	100.0	0.0	73.7	0.0
ARG	4	97.7	89.2	78.1	100.0	56.6	69.6	86.9	55.9	10.0	69.4	100.0	50.0	93.1	81.3	88.7	13.8
AUS	5	98.8	100.0	100.0	100.0	93.9	0.0	78.0	16.6	14.0	71.5	100.0	83.3	100.0	69.4	81.7	3.7
BEL	6	98.3	93.6	100.0	100.0	87.1	31.0	97.5	9.0	0.3	2.0	100.0	33.3	88.6	70.0	92.3	0.7
BEN	7	36.6	42.2	17.3	12.0	73.7	84.2	98.8	100.0	69.8	100.0	100.0	16.7	100.0	92.4	85.2	0.1
BGD	8	78.8	54.9	36.8	4.0	2.5	11.4	98.8	83.9	14.1	36.6	0.0	16.7	100.0	96.3	85.7	1.8
BGR	9	96.6	100.0	100.0	94.0	67.5	30.2	98.2	33.3	7.5	27.2	100.0	50.0	82.6	43.4	19.6	2.5
BRA	10	91.9	80.1	69.6	73.0	83.7	44.3	99.6	95.8	15.6	58.7	100.0	50.0	92.8	80.1	90.7	37.0
CAN	11	98.9	100.0	100.0	100.0	91.2	21.2	99.8	97.0	9.9	76.5	100.0	66.7	55.0	47.4	85.3	25.9
CHL	12	98.3	91.0	90.3	85.0	60.7	66.7	97.6	69.9	47.5	76.9	100.0	0.0	88.9	78.3	82.4	23.8
CHN	13	94.1	58.5	31.9	70.0	44.7	0.0	35.0	64.3	41.2	84.3	98.7	0.0	100.0	77.3	36.0	6.3
CIV	14	17.8	71.1	27.1	7.0	61.4	91.0	99.4	96.6	46.1	79.7	100.0	33.3	100.0	90.1	80.8	16.6
CMR	15	23.4	33.2	36.8	23.0	46.9	88.0	99.2	100.0	20.8	61.8	100.0	33.3	100.0	97.3	84.4	38.6
COD	16	0.0	2.5	13.7	0.0	70.9	93.8	99.4	100.0	17.5	66.0	100.0	16.7	100.0	98.0	85.9	76.6
COG	17	61.7	2.5	0.0	0.0	42.8	100.0	99.7	100.0	28.8	79.9	100.0	50.0	100.0	89.2	81.9	22.9
COL	18	92.8	85.6	83.0	64.0	89.4	49.4	99.9	94.8	21.4	78.4	100.0	50.0	98.4	91.0	85.6	32.1
CRI	19	98.3	94.6	90.3	42.0	80.2	41.0	99.9	100.0	50.0	100.0	100.0	50.0	99.4	88.2	91.3	52.2
CUB	20	98.5	83.8	97.6	58.0	89.3	11.1	97.5	47.6	32.8	89.6	100.0	50.0	100.0	44.3	77.0	1.8
CYP	21	98.9	100.0	100.0	76.0	67.8	29.3	96.4	100.0	24.1	79.1	100.0	50.0	0.0	78.8	85.6	0.0
DEU	22	99.0	100.0	100.0	100.0	91.3	31.4	98.2	70.9	1.0	2.2	100.0	33.3	34.2	80.2	93.0	3.8
DNK	23	98.7	100.0	100.0	100.0	90.9	32.9	98.4	95.9	11.9	38.0	100.0	16.7	87.1	84.4	94.8	9.2
DOM	24	83.8	87.4	47.7	52.0	79.3	11.0	98.9	62.7	32.4	100.0	100.0	50.0	100.0	87.3	66.9	4.9
DZA	25	96.7	76.5	90.3	96.0	53.3	25.3	0.0	55.3	16.0	39.5	82.3	33.3	100.0	78.5	56.3	0.0
ECU	26	95.0	74.7	66.0	72.0	87.3	69.2	99.7	64.9	34.7	81.0	100.0	16.7	100.0	72.4	71.3	18.8
EGY	27	93.7	96.4	61.1	92.0	0.0	29.5	89.5	53.5	6.1	46.4	0.0	16.7	100.0	70.4	56.4	5.7
ESP	28	99.0	100.0	100.0	100.0	78.4	20.0	92.4	32.3	18.5	96.5	100.0	16.7	49.2	80.9	90.4	9.4
FIN	29	99.2	100.0	100.0	100.0	92.5	38.1	99.7	99.2	24.1	59.4	100.0	50.0	94.6	72.0	92.2	16.3
FRA	30	99.1	100.0	100.0	100.0	95.2	27.7	98.6	84.7	7.1	70.4	100.0	33.3	5.5	79.0	95.1	5.7
GAB	31	61.0	76.5	22.2	66.0	92.2	100.0	99.9	100.0	5.0	100.0	100.0	66.7	100.0	87.8	84.5	25.4
GBR	32	98.9	100.0	100.0	100.0	93.6	29.6	99.2	84.7	28.9	68.7	100.0	50.0	64.9	83.2	89.6	1.2
GEO	33	97.2	56.7	79.3	29.0	37.4	29.0	98.8	87.2	4.3	23.6	100.0	66.7	100.0	58.9	58.8	52.7
GHA	34	58.3	62.1	48.9	5.0	83.4	91.1	98.9	100.0	8.6	69.9	85.9	16.7	100.0	95.4	82.4	36.7
GIN	35	32.3	11.6	0.0	1.0	58.2	81.0	99.0	100.0	13.8	95.6	100.0	50.0	100.0	100.0	93.2	19.9
GMB	36	43.8	67.5	42.9	2.0	41.0	68.8	99.2	100.0	0.5	10.7	67.2	33.3	100.0	100.0	86.6	0.0
GNB	37	0.0	26.0	19.8	5.0	45.8	77.5	99.8	100.0	21.4	44.0	100.0	83.3	100.0	86.2	75.0	0.0
GRC	38	98.8	100.0	100.0	100.0	73.4	28.4	97.9	91.9	4.1	11.5	100.0	33.3	85.3	80.9	84.6	4.3
GTM	39	86.9	91.0	52.6	27.0	65.2	0.0	99.8	100.0	53.2	67.7	100.0	50.0	100.0	92.5	86.7	17.4
HND	40	83.5	81.9	61.1	34.0	72.6	6.3	99.8	95.8	45.3	100.0	99.8	66.7	100.0	84.5	75.1	17.5
HTI	41	48.6	47.7	19.8	18.0	71.6	9.7	98.9	97.2	1.3	27.1	0.0	66.7	100.0	99.3	88.1	9.5

IDN	42	88.6	60.3	41.6	37.0	34.9	15.4	99.9	99.6	16.8	97.1	100.0	50.0	89.6	79.8	69.2	4.6
IND	43	67.1	74.7	14.9	19.0	43.9	12.9	96.5	38.8	11.5	57.1	69.6	16.7	100.0	87.8	45.6	5.3
IRL	44	98.8	100.0	100.0	100.0	91.0	29.6	99.2	100.0	3.5	10.7	92.8	33.3	91.4	90.1	90.5	1.9
IRN	45	94.4	87.4	80.5	98.0	56.4	5.8	91.0	53.7	11.9	63.3	100.0	50.0	100.0	52.4	29.8	1.7
ISL	46	99.1	100.0	100.0	100.0	92.3	30.7	100.0	98.3	14.4	93.7	100.0	0.0	0.0	41.2	94.1	71.4
ISR	47	99.0	100.0	100.0	100.0	70.2	28.9	92.2	0.0	27.8	54.6	100.0	16.7	0.0	82.7	86.5	0.1
ITA	48	98.9	89.4	100.0	100.0	83.8	26.7	93.0	67.7	12.2	62.6	100.0	33.3	35.7	85.6	91.5	6.8
JAM	49	94.3	87.4	75.7	53.0	68.5	26.9	99.8	100.0	65.0	100.0	99.5	66.7	100.0	42.5	52.0	1.3
JOR	50	96.9	83.8	91.5	90.0	52.3	28.8	91.7	0.0	16.3	96.1	80.8	33.3	0.0	62.9	52.7	0.2
JPN	51	98.9	100.0	100.0	100.0	83.5	21.8	99.8	89.7	26.7	100.0	100.0	0.0	0.0	80.8	95.0	6.2
KEN	52	46.5	31.4	36.8	15.0	75.8	98.2	94.9	74.7	18.5	69.9	97.4	16.7	100.0	87.5	77.4	26.2
KHM	53	49.9	0.0	0.0	0.0	58.3	54.8	99.8	100.0	56.1	100.0	100.0	33.3	100.0	100.0	97.3	6.8
KOR	54	98.4	85.6	100.0	100.0	76.8	17.4	99.2	82.3	8.8	39.5	100.0	16.7	0.4	67.5	83.6	0.7
LBN	55	96.0	100.0	97.6	91.0	75.2	29.0	96.8	81.7	0.0	7.0	79.8	50.0	100.0	64.3	70.9	4.5
LBR	56	1.8	31.4	10.0	17.0	78.0	95.1	99.9	100.0	15.3	22.9	100.0	66.7	100.0	97.7	84.1	0.0
LKA	57	97.4	60.3	89.1	11.0	40.4	64.9	96.6	69.8	32.4	70.7	77.2	16.7	100.0	95.5	85.4	15.4
MAR	58	91.5	63.9	52.6	89.0	86.3	22.6	0.0	13.1	2.4	97.2	100.0	16.7	97.0	89.7	77.8	2.1
MDG	59	43.0	0.7	18.5	1.0	73.9	75.5	99.4	78.3	5.2	43.9	100.0	50.0	100.0	95.6	85.8	15.1
MEX	60	95.9	83.8	72.0	78.0	69.1	0.0	0.0	42.4	13.9	64.6	100.0	33.3	84.0	77.0	72.8	4.8
MMR	61	60.2	63.9	67.2	0.0	43.6	11.2	99.8	96.5	11.9	43.2	98.3	33.3	100.0	97.2	96.3	15.1
MOZ	62	8.1	0.0	11.2	13.0	74.2	75.1	98.0	75.4	13.6	36.4	100.0	66.7	100.0	73.7	91.7	92.3
MRT	63	33.4	20.6	29.5	31.0	26.8	35.1	0.0	71.1	0.2	2.9	0.0	66.7	100.0	69.2	43.8	0.8
MYS	64	97.1	91.0	95.1	71.0	90.0	69.6	100.0	98.6	30.2	97.8	100.0	33.3	100.0	65.7	69.2	2.6
NAM	65	64.9	63.9	14.9	17.0	69.2	69.1	78.4	5.0	35.5	99.3	100.0	50.0	100.0	89.7	89.5	28.8
NGA	66	0.0	27.8	24.6	33.0	33.2	78.7	98.2	91.5	15.8	42.0	88.6	16.7	100.0	77.9	73.3	8.4
NIC	67	89.9	65.7	58.7	27.0	77.5	23.5	99.9	100.0	46.9	75.0	100.0	66.7	100.0	91.9	65.0	14.0
NLD	68	98.3	100.0	100.0	100.0	80.8	31.0	97.0	55.9	4.3	14.2	100.0	33.3	68.1	73.5	93.3	1.3
NOR	69	98.9	100.0	100.0	100.0	92.4	33.2	99.9	100.0	8.0	28.0	100.0	0.0	0.0	62.2	93.3	60.4
NZL	70	98.6	94.6	100.0	100.0	95.1	72.2	99.7	97.9	54.7	78.0	100.0	33.3	51.0	65.8	88.9	35.3
OMN	71	97.8	62.1	86.6	100.0	32.2	24.1	100.0	31.5	11.1	50.0	100.0	66.7	100.0	61.3	66.9	0.0
PAK	72	62.1	81.9	44.1	24.0	0.0	16.4	36.7	39.1	6.0	46.5	0.0	33.3	100.0	81.1	64.9	14.0
PAN	73	93.8	83.8	66.0	63.0	69.2	48.0	99.9	95.3	58.1	100.0	100.0	33.3	100.0	66.1	84.8	14.4
PER	74	81.1	65.7	53.8	60.0	63.2	52.5	98.0	69.6	16.6	80.2	100.0	0.0	41.8	90.2	88.3	33.4
PHL	75	94.4	72.9	67.2	15.0	72.0	50.8	99.9	94.5	23.9	100.0	96.1	16.7	95.7	91.8	79.4	22.0
PNG	76	70.8	0.0	33.1	3.0	84.9	62.6	100.0	96.8	2.0	30.8	100.0	83.3	100.0	93.1	88.1	20.4
POL	77	98.4	100.0	100.0	93.0	76.1	31.8	97.6	89.8	2.0	18.6	100.0	16.7	48.0	73.3	49.4	0.7
PRT	78	98.2	100.0	100.0	100.0	83.0	17.2	96.6	81.8	11.7	76.0	99.2	16.7	91.4	83.4	89.0	16.4
ROU	79	95.9	22.4	40.4	91.0	54.6	30.5	72.6	68.5	18.6	25.2	100.0	50.0	47.6	67.3	37.2	9.7
RUS	80	95.3	92.8	84.2	99.0	88.8	22.4	99.7	96.2	10.6	90.9	100.0	50.0	100.0	12.8	20.0	6.1
SAU	81	96.0	81.9	87.8	100.0	31.7	28.7	99.5	5.7	14.9	100.0	100.0	66.7	100.0	28.5	44.7	0.0
SDN	82	49.2	44.0	19.8	0.0	0.0	49.9	89.1	80.5	6.0	48.4	97.6	50.0	100.0	97.9	87.1	9.8
SEN	83	46.5	49.5	41.6	21.0	41.5	64.3	28.3	75.6	22.0	100.0	99.5	16.7	100.0	90.6	82.8	0.0
SLE	84	0.0	22.4	25.8	8.0	62.5	88.4	99.9	100.0	11.5	77.9	97.7	50.0	100.0	92.5	82.5	0.0
SLV	85	91.7	67.5	55.0	35.0	76.4	8.6	99.6	100.0	0.5	12.2	48.4	50.0	100.0	91.7	85.6	29.7
SUR	86	94.7	85.6	91.5	31.0	71.0	76.4	99.9	100.0	16.2	100.0	100.0	83.3	100.0	0.0	32.3	39.3
SVN	87	98.2	100.0	100.0	100.0	81.9	30.4	99.0	100.0	0.0	4.3	100.0	0.0	11.5	73.9	85.0	10.8
SWE	88	99.3	100.0	100.0	100.0	96.2	36.6	99.7	99.4	14.3	72.7	100.0	50.0	93.0	72.5	96.2	28.1
SYR	89	97.2	62.1	72.0	81.0	34.5	29.1	96.5	0.0	1.6	3.2	100.0	16.7	100.0	48.1	0.0	12.6
TGO	90	51.0	11.6	19.8	4.0	74.4	90.3	98.4	100.0	32.9	100.0	67.0	16.7	100.0	97.2	70.6	0.2
THA	91	95.0	72.9	98.8	28.0	53.0	42.1	99.6	84.0	34.8	77.2	72.4	0.0	64.5	79.1	72.4	3.2

TTO	92	95.0	83.8	100.0	100.0	89.9	58.9	99.8	100.0	7.4	18.7	100.0	66.7	100.0	0.0	7.3	0.0
TUN	93	97.1	67.5	75.7	71.0	73.3	24.8	65.0	5.1	0.3	4.1	20.3	50.0	79.3	86.6	81.4	0.3
TUR	94	92.3	87.4	79.3	89.0	68.6	29.5	97.4	74.6	4.5	25.6	100.0	16.7	89.2	78.9	74.3	10.8
TWN	95	92.2	100.0	100.0	100.0	63.1	31.6	99.6	100.0	29.9	86.9	100.0	0.0	14.1	75.6	81.5	1.8
TZA	96	33.4	51.3	34.3	4.0	80.6	91.9	97.0	80.2	38.2	99.4	100.0	16.7	100.0	92.3	84.1	37.9
UKR	97	97.9	96.4	98.8	89.0	82.1	31.1	74.6	55.8	6.6	47.8	100.0	33.3	100.0	8.1	0.0	1.7
USA	98	98.5	100.0	100.0	100.0	89.3	0.1	86.6	61.1	31.8	90.6	100.0	16.7	0.0	68.8	85.1	4.0
VEN	99	88.5	69.3	61.1	100.0	95.8	58.0	99.8	82.3	72.5	100.0	100.0	33.3	0.0	12.5	52.8	20.9
VNM	100	91.4	51.3	28.3	2.0	53.7	36.9	99.6	94.5	12.0	50.1	82.0	33.3	100.0	86.5	52.7	21.6
YEM	101	72.5	44.0	14.9	34.0	37.6	41.4	100.0	0.0	0.0	0.0	90.1	66.7	100.0	67.8	64.3	0.0
ZAF	102	68.1	76.5	59.9	72.0	89.9	67.3	66.5	0.0	11.5	76.9	84.2	33.3	27.6	64.6	54.6	0.2

Table 13

The investigators developed an index of the 16 indicators by applying principal components analysis (PCA) to the indicators. Six principal components were found with eigenvalues greater than one, of which three could be effectively treated as the policy categories, Environmental Health, Sustainable Energy, and Biodiversity and Habitat. The weights of the indicators within these three categories were determined from the PCA loadings, and the weights were normalized to sum to the weight given to the category. For the other three policy categories, Air Quality, Water Resources, and Productive Natural Resources, the indicators within each group were given equal weight. We can treat the model as a two dimensional hierarchy, since Environmental Health was one of the two main goals of the project, it was given weight 0.5. The other five categories which represented ecosystem vitality and natural resource management, were given equal weight, so they each received a weight of 0.1. There is the issue of indicators which are in more than one category, so their total weight will be a sum of their weights from both categories. The weighting scheme for the EPI is shown in Table 14.

Indicator	Code	Weight
Child Mortality	I1	0.105
Drinking Water	I2	0.11
Adequate Sanitation	I3	0.11
Indoor Air Pollution	I4	0.11
Urban Particulates	I5	0.115
Regional Ozone	I6	0.05
Nitrogen Loading	I7	0.05
Water Consumption	I8	0.057
Wilderness Protection	I9	0.039
Ecoregion Protection	I10	0.039
Timber Harvest Rate	I11	0.033
Overfishing	I12	0.033
Agricultural Subsidies	I13	0.049
Energy Efficiency	I14	0.043
C02 per GDP	I15	0.047
Renewable Energy	I16	0.01

Table 14

Eurobarometer Survey:

The Health and Consumer Protection General Directorate of the European Commission requires that surveys be taken regarding the opinion of Europeans about quality of life issues. The survey that produces this data set is an Eurobarometer survey, it surveys the satisfaction of Europeans from 18 different countries about public services. (Annoni, 2006) This information for the survey was obtained in September and October 2002, and contains data for the following countries, Belgium (B), Denmark (DK), West and East Germany (West-D and East-D), Greece (GR), Italy (IT), Spain(E), France (F), Ireland (IRL), Northern Ireland (North IRL), Luxembourg(L), Netherlands (NL), Portugal (P), Great Britain (GB), Norway (N), Finland (FIN), Sweden (S) and Austria (A).

In each country, subjects were asked how they perceive the quality of eight different services. The services that were inquired are: 1. mobile telephone services (mob phone), 2. fixed telephone services (fix phone), 3. electricity supply services (electricity), 4. gas supply services (gas), 5. water supply services (water), 6. postal services (postal), 7. transport services within towns/cities (transport), 8. rail services between towns/cities (rail). For each service, five criteria are used to determine the quality of the service: a. access easiness to service (access), b. price of the service (prices), c. quality of the service (quality), d. clarity of the information aimed at consumers by service customer care (info), e. fairness of terms and conditions of service contracts (contract).

There are a total of 40 different questions asked to each interviewer, each of the questions have a categorical answer, e.g. “fair”, “good”, etc. Each of the answers can be coded as a numerical score, with the lower score corresponding to a higher value of satisfaction. For each country, all the scores were tabulated, and the median was computed for each question, and then the resulting median was rescored to yield an integer between 1 and 3 for all questions.

This results in a data set with 18 countries (objects) and 40 questions (indicators). However, for 16 of the questions, the median value computed is the same for all 18 countries, which adds no useful information for the goal of ranking the 18 countries. Without these 16 questions, we are left with 24 indicators which yield useful information, and the 18x24 data matrix. Since our goal is to rank the object from best to worst, and our scheme considers higher scores to be indicative of better objects, we change the orientation by subtracting each data point from 3. The modified data matrix is in Table 15, which is in three parts in order to fit it in this paper.

The investigators used a sum of the scores as a composite score for that particular country, we would like the weights to sum to one, so we give equal weight to all 24 indicators, or a weight of 1/24 for all 24 indicators.

country code	Number	prices: mob phone	prices: fix phone	prices: electricity	prices: gas	prices: water	prices: postal	prices: transport	prices: rail
B	1	3	3	3	3	3	3	3	3
DK	2	3	3	3	3	3	3	3	3
West D	3	2	3	3	2	3	3	2	1
GR	4	2	1	1	2	3	3	3	3
IT	5	1	1	1	1	2	3	2	1
E	6	1	1	2	3	3	3	3	2

F	7	1	2	3	3	1	3	2	2
IRL	8	2	3	3	3	3	3	3	3
North IRL	9	3	3	3	3	3	3	3	2
L	10	3	3	3	3	3	3	3	3
NL	11	1	3	3	3	3	3	1	1
P	12	3	1	1	3	3	3	2	2
GB	13	3	3	3	3	3	3	3	2
East D	14	2	3	2	2	1	2	1	1
N	15	2	3	1	2	3	3	2	2
FIN	16	3	3	3	2	3	3	2	2
S	17	3	3	3	2	3	3	3	2
A	18	3	3	3	2	3	3	2	1

country code	Number	access: gas	quality: fix phone	quality: electricity	quality: gas	quality: water	quality: postal	quality: rail	info: electricity	info: gas
B	1	3	2	2	2	2	2	2	3	3
DK	2	3	3	3	3	3	3	2	3	3
West D	3	3	2	2	2	2	2	2	3	3
GR	4	2	2	2	1	2	2	2	3	2
IT	5	3	2	2	2	2	2	1	3	3
E	6	3	2	2	2	2	2	2	3	3
F	7	3	2	2	2	2	2	2	3	3
IRL	8	3	3	3	3	3	3	2	3	3
North IRL	9	3	2	2	2	2	2	2	3	3
L	10	3	3	3	3	3	2	2	3	3
NL	11	3	2	2	2	2	2	1	3	3
P	12	3	2	2	2	2	2	2	3	3
GB	13	3	2	3	3	2	2	2	3	3
East D	14	3	2	2	2	2	2	1	3	3
N	15	2	3	3	1	3	2	2	3	2
FIN	16	2	2	2	1	2	2	2	3	2
S	17	2	3	3	1	3	2	2	2	2
A	18	3	2	3	2	3	2	2	3	3

country code	Number	contract: mob phone	contract: fix phone	contract: electricity	contract: gas	contract: water	contract: transport	contract: rail
B	1	3	3	3	3	3	3	3
DK	2	3	3	3	3	3	3	3
West D	3	3	3	3	3	3	3	3
GR	4	3	3	3	2	3	3	3
IT	5	2	1	2	2	2	2	2
E	6	2	2	3	3	3	3	3
F	7	2	3	3	3	3	3	3
IRL	8	3	3	3	3	3	3	3
North IRL	9	3	3	3	3	3	3	3
L	10	3	3	3	3	3	3	3
NL	11	3	3	3	3	3	3	3
P	12	3	3	3	3	3	3	3

GB	13	3	3	3	3	3	3	3
East D	14	3	3	3	3	3	2	2
N	15	3	3	3	2	3	3	3
FIN	16	3	3	3	2	3	3	3
S	17	3	3	3	2	3	3	3
A	18	3	3	3	3	3	3	3

Table 15

Results:

ASC Watershed:

We applied the POSAC method to both Level I and Level II data sets. 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 on the left in Figure . The high number of comparabilities preserved suggests that the POSAC model is an excellent approximation for the five indicator data set. For Level II, the POSAC diagram is on the right on Figure 16. 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. We then compute the loadings for both LOV1 and LOV2 using the concordance method for both Level I and Level II. We used 8 different levels to discretize both the LOV and the indicator, both will take an integer between 1 and 8. Since we have eight different level, we use the concordance method that computes the proportion of objects which the indicator and LOV value differ by at most one.

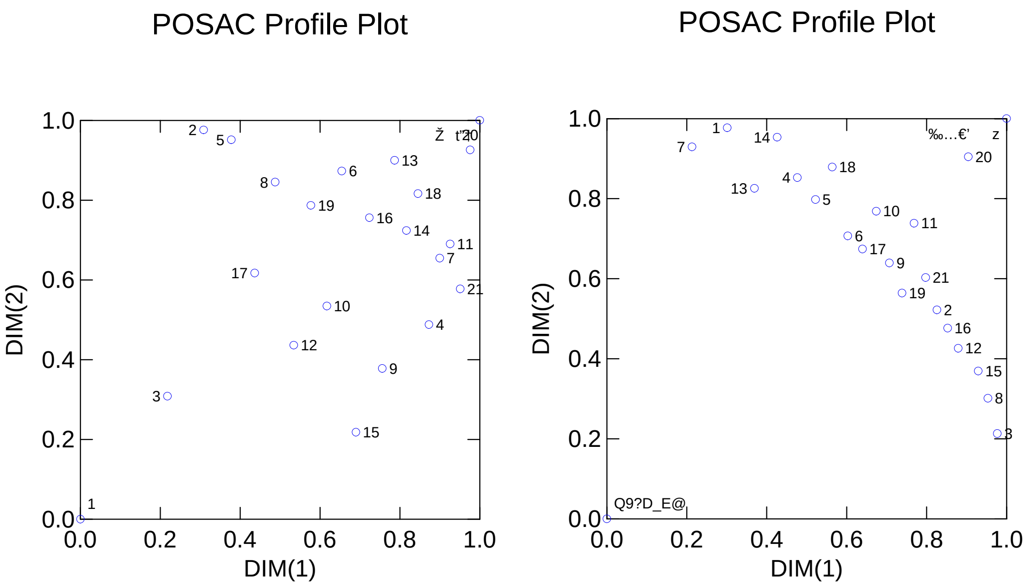


Figure 16

The loadings using the concordance method described above for both LOVs for Level I is shown in Table 17. 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 small positive correlation between the two latent order variables. However, one can see from the plot (left side of Figure), that if the plot is divided into regions whose boundaries which are negatively sloped, there is a clear negative correlation between LOV1 and LOV2. Furthermore, when the watersheds are divided into their social choices, there is a negative correlation between LOV1 and LOV2, for example, the urban watersheds (1,3,9,12,15) can be seen toward the lower left side of the graph. LOV1 is dominated primarily by LDI and MPAT, while LOV2 is dominated by LDI and CORFOR.

	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

Table 17

For Level II, we can see from Table 18, 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. 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. LOV1 seems to be dominated most by the indicators IR, INV, and SHA, while LOV2 appears to be influenced most by BUF, and to a smaller extent by INV, SHA, and FPWL.

	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

Table 18

For the ASC watershed data set, we performed step-by-step aggregation on both Level I and Level II data using three weighting schemes. We computed the data based weights by averaging the loadings for each indicator, and normalizing the resulting values to sum to one for both Level I and Level II. We also used the investigator based weights, as suggested by the Landscape index for Level I and the SWR index for Level II. We call this as the equal group weights, since each group received the same weight. The third weighting scheme we used was the equal indicator weighting scheme, where each indicator received the same weight. To compare the performance of the weighting schemes, we use the comparability acquisition profile for all three weighting schemes.

The comparability acquisition profiles for Level I of the ASC watershed data set is in Figure 19. The data based weighting scheme technically accumulates more comparabilities at each aggregation step, however the improvement in the number of comparabilities by the data based weighting scheme over the two investigator based weighting scheme is rather small. Since the difference is very small, we say that for practical purposes that the three weighting schemes yield similar performance.

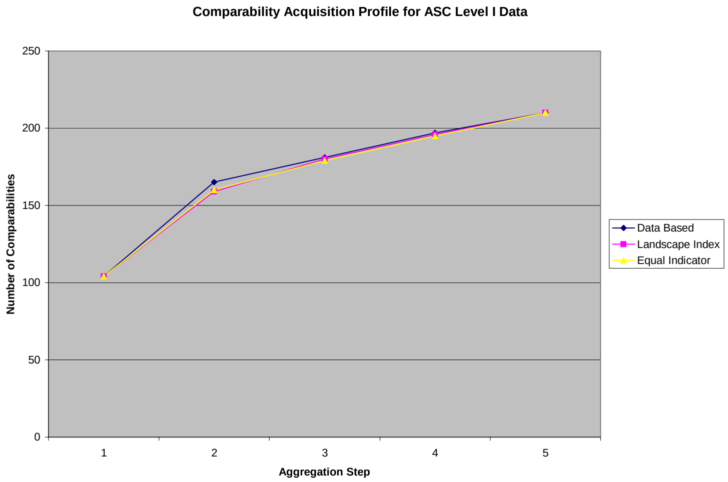


Figure 19

The comparability acquisition profiles for Level II of the ASC watershed data set is in Figure 20. It is clear from Figure that the SWR weighting scheme performs worse than both the equal indicator weighting scheme as well as the data based weighting scheme. However, it is inconclusive whether the data based weighting scheme performs better than the equal indicator weighting scheme, and any difference is small enough that it is unclear if there is any practical difference.

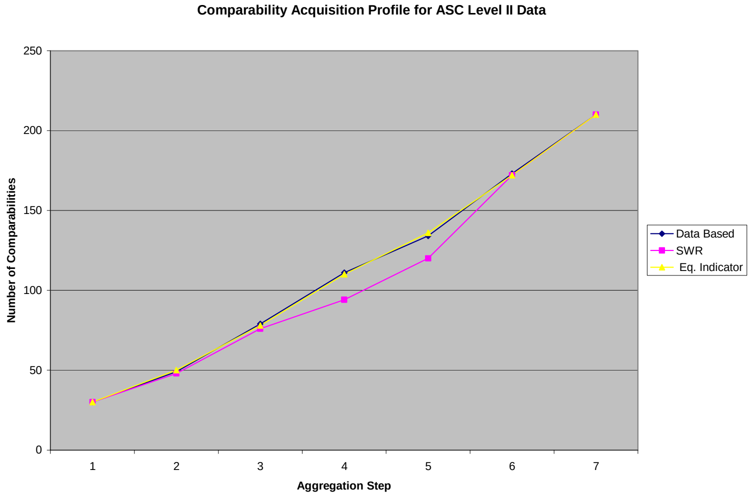


Figure 20

Bridge:

For the bridge data, POSAC correctly preserves 78.5% of the comparabilities from the original data set. We can get a reasonably good approximation for the Hasse diagram ordering by replacing the 13 indicators with the two dimensions found by POSAC, the diagram is show below in Figure 21. We also compute the loadings by both the F-value and concordance methods. For the concordance methods, we use four levels to discretize the LOVs, since the investigators divided the indicator values into four qualitative levels, ‘Excellent,’ ‘Good,’ ‘Fair,’ or ‘Poor’. Since four levels is a fairly small number of levels, we try two methods of concordance, one which finds the proportion of objects which the LOV value is exactly the same as the indicator level, and another where we compute the proportion of objects which the indicator and LOV value differ by at most one.

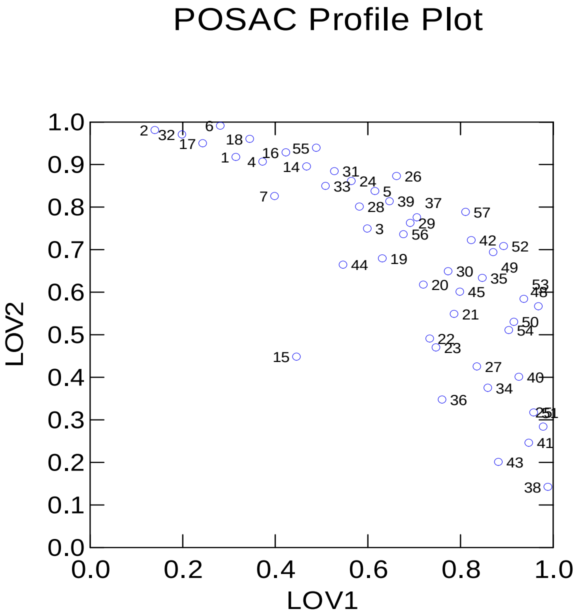


Figure 21

The F-value and concordance loadings can be found in Table 22. From the F-value and concordance loadings, we can see that LOV1 is heavily impacted positively by indicator 8, or bank soil texture, while negatively impacted by indicator 13, bridge channel alignment. Also having an impact on LOV1 are indicators 1,2, 3 and 5. LOV2 is heavily impacted positively by indicator 13, and negatively impacted by indicator 8. Also impacting LOV2 are indicators 6, 7, and 8. This is consistent with other analysis done on this data set using the W-matrix for influential indicators. (Newlin 2006, Voigt 2004) We also notice that for this data set, that the concordance loadings do not seem to clearly identify the influential indicators as well as the F-values. One possible explanation is that there appears to be few LOVs that take low values, which results in most of the LOV values having discretized values of 3 or 4 (out of a scale of 1-4).

#	Indicator Name	F-value		Concordance I		Concordance II	
		LOV1	LOV2	LOV1	LOV2	LOV1	LOV2
1	Watershed and Floodplain Activity	2.11	1.67	0.34694	0.32653	0.83673	0.73469
2	Flow Habit	2.68	1.04	0.53061	0.2449	0.87755	0.77551
3	Channel Pattern	2.18	0.99	0.42857	0.30612	0.89796	0.79592
4	Entrenchment or Channel Confinement	1.23	1.5	0.2449	0.32653	0.7551	0.69388
5	Bed Material	3.81	0.86	0.36735	0.20408	0.77551	0.65306
6	Bar Development	1.17	1.46	0.30612	0.34694	0.79592	0.87755
7	Obstructions	1.85	0.57	0.32653	0.36735	0.81633	0.77551
8	Bank Soil Texture	9.22	3.03	0.46939	0.12245	0.91837	0.59184
9	Average Bank Slope Angle	0.66	0.61	0.12245	0.2449	0.53061	0.65306
10	Bank Protection	0.82	1.23	0.16327	0.22449	0.69388	0.63265
11	Bank Cutting	0.83	1.41	0.28571	0.38776	0.71429	0.81633
12	Mass Wasting or Bank Failure	0.92	0.71	0.40816	0.26531	0.71429	0.7551
13	Bridge-Channel Alignment	4.49	17.11	0.12245	0.36735	0.42857	0.89796

Table 22

We also applied the method of METEOR for the bridge data as well. We can apply METEOR for two different sets of weights determined by the investigators for this data. One set of weights gives the equal weight for all 13 indicators, or a weight of 1/13 for each of the 13 indicators. The other set of weights gives equal weight for all four groups (of ¼), and then divides the weight among the indicators within the groups equally. For the equal indicator weighting scheme, we applied METEOR on all 13 indicators. We also applied METEOR on all 13 indicators for the equal group weighting scheme as well, in the same manner. The comparability acquisition profile is shown in Figure 23.

We see from Figure that for the bridge data, aggregation with equal indicator weights seems to result in more comparablities than with the equal group weights at each step. A large part of the explanation for this phenomenon is due to the influential nature of Indicator 13. Indicator 13 is considered to be influential, since its removal from the data set results in a large increase of comparabilities. More about Indicator 13 will be discussed in the conclusion.

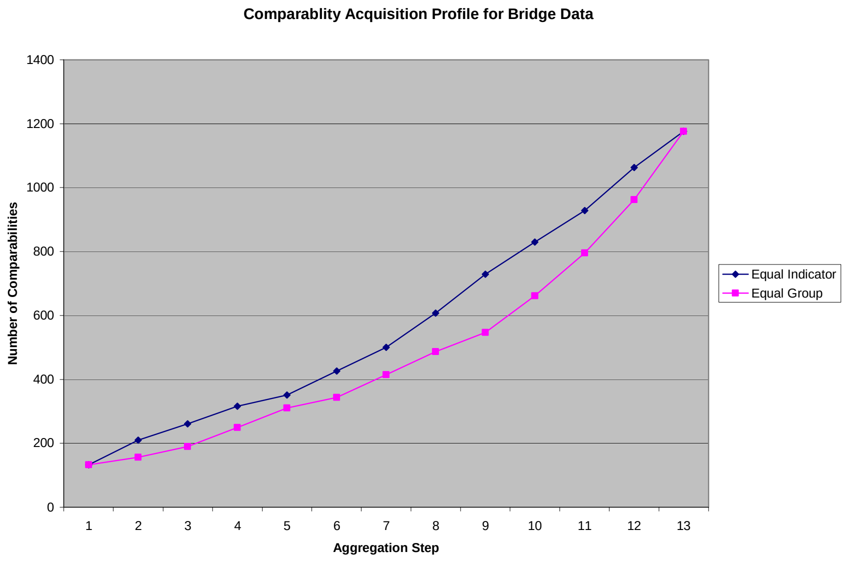


Figure 23

Step-by-step aggregation was also performed for the two sets of data based weights. For both concordance methods used for this data set, we computed the data based weights by averaging the loadings found by the concordance method for both LOVs and then normalizing the weights to add up to one. The comparability acquisition profile for all four weighting schemes, two of them are investigator based, two are data based are in Figure 24. We notice from Figure that the two data-based weighting schemes accumulate comparabilities faster at each step than the equal group weighting scheme, but the equal indicator weighting scheme accumulates comparabilities better than the data based weighting schemes. Between the two data based weighting schemes, it is undecided which weighting scheme is better since the two comparability acquisition profile graphs intersect, but it is worth noting that for the last few aggregation steps, the second data based weighting scheme performs better, yielding more comparabilities than the first data based weighting scheme. It is also noted that the two data based weighting schemes are very similar numerically, with no weight differing by more than 0.02, yet there are clear differences in their comparability acquisition profile.

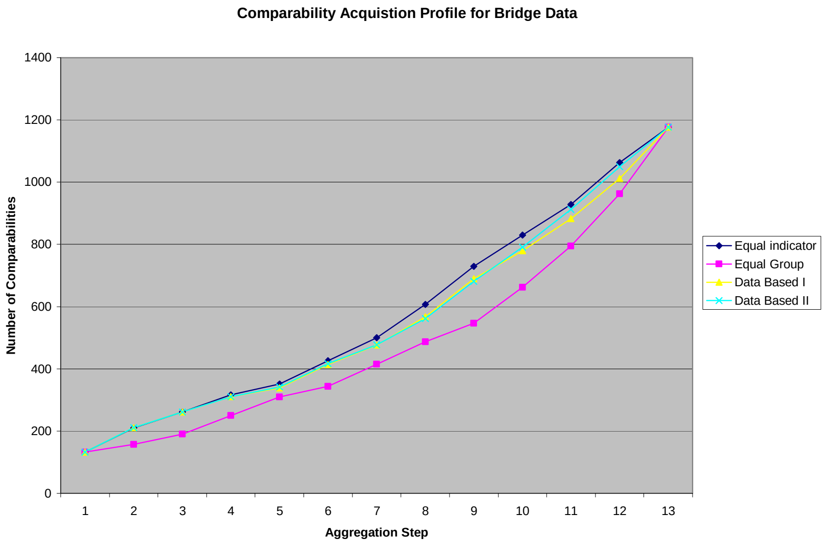


Figure 24

For the equal group weighting scheme, we can also treat the indicators as a two-level hierarchy, specifically the indicator groups as the top level, and the indicators within the groups at the lower level. We can thus perform METEOR in two steps, first aggregate all the indicators step-by-step into an index for the group, and then aggregate the indices for the groups into a final index. Below are the number of comparabilities at each step for each group (group 4 was indicator 13 by itself, and thus did not need aggregation), and for the aggregation of the indicators together.

For the two-stage hierarchy, we note that the majority of the increase in comparabilities takes place in the top hierarchy. The aggregation of the 13 indicators to the four indicies of the groups increases the comparabilities from 133 to 460, while reducing the four indicators to the index results in an increase from 460 to 1176. The number of comparabilities with four indicators in the ordinary aggregation (for the equal group weights with aggregation step 10) is 662 as opposed to 460 in hierarchical aggregation. The method of hierarchical aggregation does not increase comparabilities as fast as the ordinary aggregation.

EPI:

We first performed POSAC analysis on the EPI data, the POSAC diagram is shown below in Figure 25. The two LOVs for POSAC correctly represent 96.2% of the comparabilities of the entire model, which is an extremely high level of accuracy for the two dimensions. We also find that the correlation between the two latent order variables is -0.933, indicating a very strong negative correlation between the two variables.

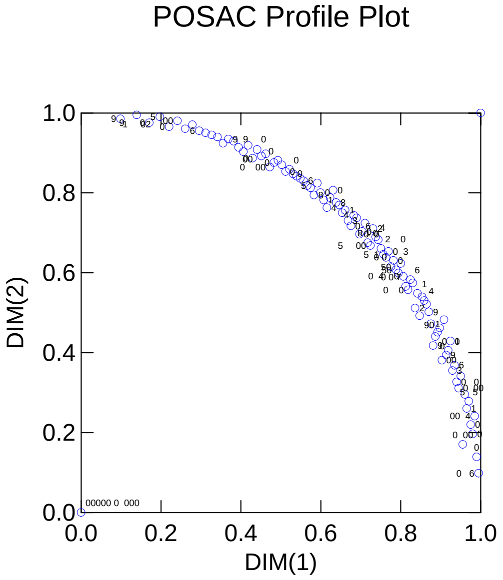


Figure 25

We also consider the "loadings" of the LOVs with respect to the 16 EPI indicators. We see that using the F-value method, both LOVs are influenced heavily by Indicators 1 to 4, which are the main indicators of Environmental Health. These four indicators have very high F-values (for the ANOVA test between the LOV and the indicator) for both LOVs. The other indicators that have significant F-values are I5, I6, I11, I13, I14. It is noted that for all the indicators, if one LOV had a high F-value for that indicator, then the other LOV also had a high value for that indicator. Since there is such a large negative correlation between the LOVs, we are not surprised by this phenomenon.

To use the concordance method, we used 5 different levels, we discretize the LOVs and the indicator values into five different values. Since five levels is a fairly small number of levels, we try two methods of concordance, one which finds the proportion of objects which the LOV value is exactly the same as the indicator level, and another where we compute the proportion of objects which the indicator and LOV value differ by at most one. Observing the other measures to obtain the loadings, we see that from the concordance method, which LOV1 has the highest concordance with the indicators of Environmental Health, I1-I5. I7, I10, I14 and I15 have concordance greater than 0.5 as well. LOV2 has the highest concordance with I13-I15, and has concordance greater than 0.5 with I6, I7, I8, I10, and I12. Using the Spearman's rank correlation method, we see that LOV1 has the highest positive correlation with the four Environmental Health Indicators, I1-I4, with a high positive correlation with I5. LOV1 also has a high negative correlation with I13 and I14. LOV2 has a high positive correlation with I13 and I14, while it has a high negative correlation with I1-4. Again this is intuitive since the two LOVs have high negative correlation.

		F-test		Conc I		Con II	
		LOV1	LOV2	LOV1	LOV2	LOV1	LOV2
1	MORTALITY	29.62	16.38	0.401961	0.088235	0.813726	0.421569
2	WATSUP	24.71	24.83	0.22549	0.137255	0.803922	0.411765
3	ACSAT	33.75	35.84	0.196078	0.127451	0.803922	0.343137
4	INDOOR	90.35	62.91	0.205882	0.078431	0.794118	0.205882
5	PM10	9.48	7.38	0.362745	0.137255	0.794118	0.5
6	OZONE	11.34	11	0.068627	0.264706	0.294118	0.598039
7	NLOAD	0.61	0.498	0.352941	0.333333	0.578431	0.568628
8	OVRSUB	1.45	0.961	0.205882	0.205882	0.490196	0.607843
9	PWI	0.382	0.734	0.117647	0.04902	0.294118	0.215686
10	PACOV	2.23	1.33	0.147059	0.088235	0.519608	0.519608
11	HARVEST	4.48	3.35	0.078431	0.117647	0.470588	0.382353
12	OVRFSH	0.88	1.989	0.176471	0.245098	0.431373	0.568628
13	AGSUB	22.89	60	0.127451	0.147059	0.22549	0.647059
14	ENEFF	7.861	5.279	0.235294	0.45098	0.588235	0.735294
15	CO2GDP	0.981	1.651	0.392157	0.362745	0.607843	0.637255
16	RENPC	0.386	1.365	0.088235	0.039216	0.254902	0.205882

Table 26

It can be clearly seen that LOV1 is positively influenced by the indicator of Environmental Health, I1-I5, while negatively influenced by I13 and I14. Similarly it appears that LOV2 is positively influenced by I13 and I14, while negatively influenced by the Environmental Health indicators I1-I5. It seems that LOV2 adds little new information that is not given LOV1, which can be explained by the high negative correlation between LOV1 and LOV2. Indicators which have a strong positive relationship with LOV1 have a strong negative relationship with LOV2, and vice versa. Indicators that have no strong relationship with LOV1 also have no strong relationship with LOV2.

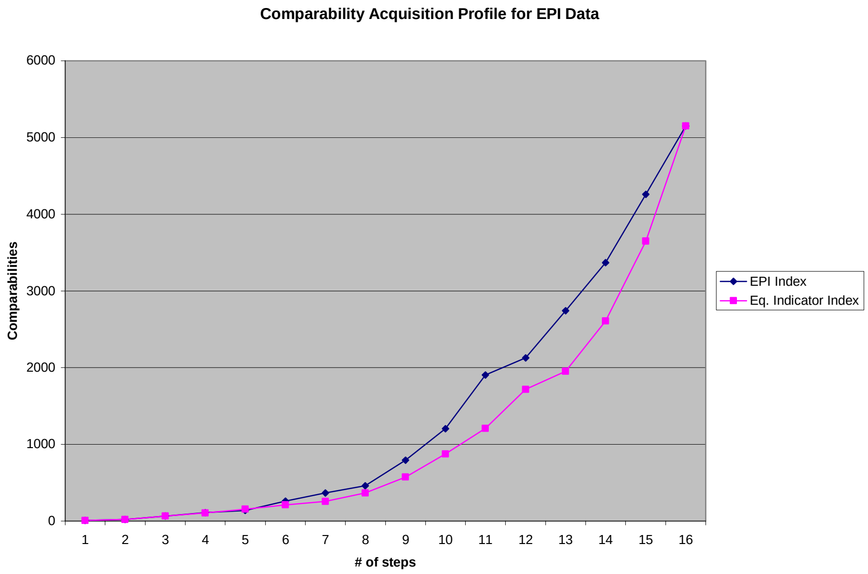


Figure 27

We next performed step-by-step aggregation using METEOR on the EPI data set. We used two investigator based weighting schemes, the EPI index derived from PCA, and the equal indicator weights. We also used the two data based weighting schemes computed by averaging the loadings found by the concordance method for both LOVs and then normalizing so the weights add up to one. The comparability acquisition profiles for the investigator based weighting schemes are shown in Figure 27. It is clear that as the number of aggregation steps becomes larger (above 6), the EPI accumulated more comparabilities than the equal indicator weight scheme, and thus outperforms it. From Figure 28, which shows the comparability acquisition profiles for all four weighting schemes, we see that the EPI outperforms the two data based weight schemes as well. It is also clear that it is inconclusive whether the data based weight schemes perform better than the equal indicator weights, or which data based weight scheme performs better.

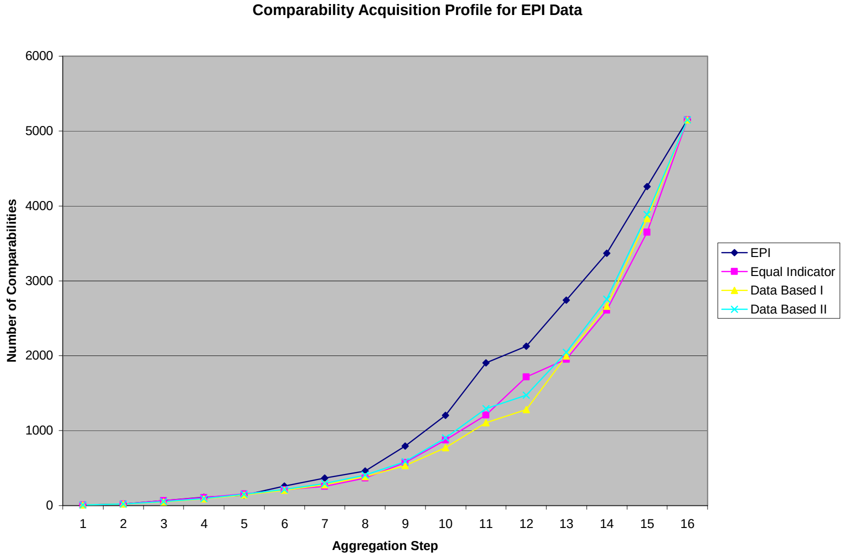


Figure 28

We also did aggregation by treating the EPI data as a two level hierarchy, the policy categories as the top level, and the indicators within the categories as the lower level. We included any indicators which are in multiple categories in all the categories that contain the indicator. For example, indicator, PM10, or Urban Particulates is in both Environmental Health and Air Quality. We used the EPI index weight, giving the appropriate weights to each indicator in each group as determined by the investigators, which are given in Table 29 . The sum of the weights in each category is one.

Indicator	Code	Weight	Category
Child Mortality	I1	0.21	Environmental Health
Drinking Water	I2	0.22	Environmental Health
Adequate Sanitation	I3	0.22	Environmental Health
Indoor Air Pollution	I4	0.22	Environmental Health
Urban Particulates	I5	0.13	Environmental Health
Urban Particulates	I5	0.50	Air Quality
Regional Ozone	I6	0.50	Air Quality
Nitrogen Loading	I7	0.50	Water Resources
Water Consumption	I8	0.50	Water Resources
Water Consumption	I8	0.07	Biodiversity and Habitat
Wilderness Protection	I9	0.39	Biodiversity and Habitat
Ecoregion Protection	I10	0.39	Biodiversity and Habitat
Timber Harvest Rate	I11	0.15	Biodiversity and Habitat
Timber Harvest Rate	I11	0.333	Prod Nat Resources
Overfishing	I12	0.333	Productive Natural Resources
Agricultural Subsidies	I13	0.333	Productive Natural Resources
Energy Efficiency	I14	0.43	Sustainable Energy
C02 per GDP	I15	0.47	Sustainable Energy
Renewable Energy	I16	0.10	Sustainable Energy

Table 29

For the two stage hierarchy, we used a bottom-up method, first aggregating the indicators within each category, and then aggregating the indices of all six categories together. One metric that we use to check how well the aggregation of the lower level indicators into categories performed in reducing incomparabilities is to compare the number of comparabilities of the six categories to the single level model with six indicators remaining. With six indicators remaining (or the 11th aggregation step) in the single level model, the number of comparabilities for the EPI index is 1905. The hierarchical approach with the indices of the six performance categories yields 295 comparabilities, which suggests that most of the incomparabilities to be resolved are between the six categories, and not within the categories. We show the comparability acquisition profiles in Figure 30.

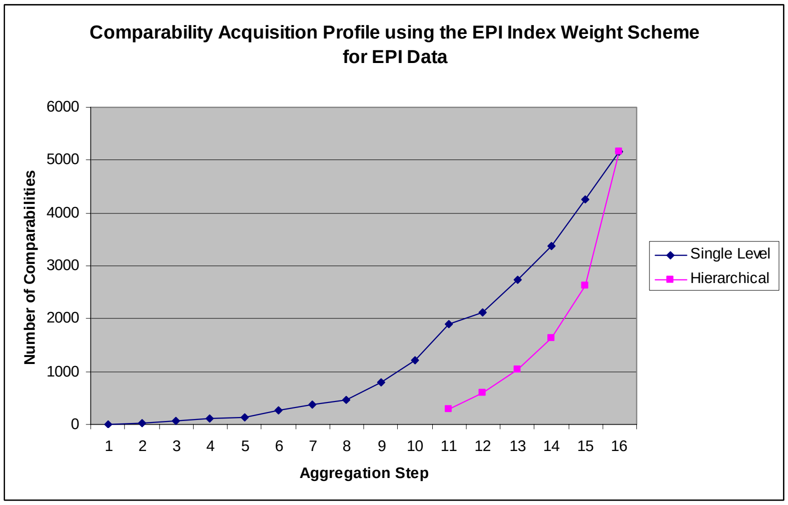


Figure 30

We also looked closely at the aggregations for the lower level of the hierarchy, that which individual indicators were aggregated into the category index. The indicators within the Environmental Health category had 2802 comparabilities out of 5151 possible pairs of countries. This implies that there was substantial consensus between the five indicators in the category. In contrast, the two indicators in the category Air Quality had 2623 comparabilities, which suggests some disagreement among the two indicators in ranking the countries. Generally, there are more comparabilities when there are fewer indicators within the category. The information for all six policy categories are in Table 31 .

Category	# Indicators	Comparabilities
Environmental Health	5	2802
Air Quality	2	2623
Water Resources	2	3956
Biodiv & Habitat	4	1959
Prod Nat Resources	3	3348
Sust. Energy	3	2165

Table 31

Eurobarometer Survey:

We apply POSAC on this data and the POSAC diagram is in Figure 32. 79.8% of the comparabilities are correctly represented by the reduced POSAC model, which indicates that POSAC is a good approximation for the full data set. We have computed the loadings using the concordance method, first we discretize the LOV into 3 levels, 1,2,3. We use 3 levels since the data is already discretized into those three categories. Since there are only three levels, we use a concordance that only counts the proportion of objects that have the same discretized value for both the LOV and the indicator.

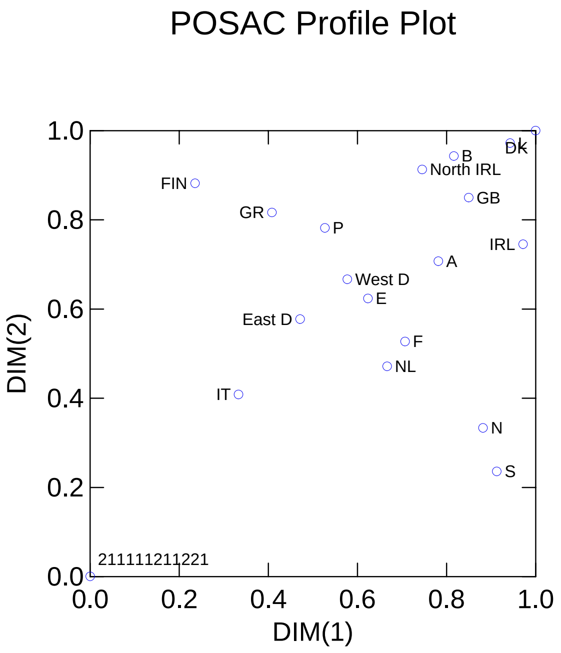


Figure 32

The concordance values for the 24 indicators for both LOV1 and LOV2 are listed in Table 31. We use the 0.7 threshold for determining which indicators are considered to have the greatest impact on the LOV. From Table 33, LOV1 is most strongly impacted by I11, which is the quality of electricity, and I23 and I24, which are the transport and rail contracts. LOV2 is most strongly impacted by I18, I23, and I24, which are mobile phone, transport, and rail contracts. The correlation between LOV1 and LOV2 is 0.09, which suggests that the two LOVs are not correlated and have their own conclusions.

Indicator		LOV 1	LOV 2	Indicator		LOV 1	LOV 2
access: gas	1	0.555556	0.555556	quality: water	13	0.666667	0.5
prices: mob phone	2	0.555556	0.555556	quality: postal	14	0.444444	0.444444
prices: fix phone	3	0.555556	0.555556	quality: rail	15	0.222222	0.166667
prices: electricity	4	0.666667	0.611111	info: electricity	16	0.555556	0.611111
prices: gas	5	0.611111	0.5	info: gas	17	0.555556	0.555556
prices: water	6	0.611111	0.666667	contract: mob phone	18	0.666667	0.777778
prices: postal	7	0.666667	0.666667	contract: fix phone	19	0.666667	0.666667
prices: transport	8	0.555556	0.555556	contract: electricity	20	0.666667	0.666667
prices: rail	9	0.333333	0.444444	contract: gas	21	0.611111	0.611111
quality: fix phone	10	0.611111	0.444444	contract: water	22	0.666667	0.666667
quality: electricity	11	0.722222	0.555556	contract: transport	23	0.722222	0.722222
quality: gas	12	0.555556	0.555556	contract: rail	24	0.722222	0.722222

Table 33

We averaged the loadings above for each indicator and normalized the resulting values to sum to one, which gives us the data based weights. We also wanted to determine how well the investigator's weights, which give equal weight to all 24 indicators, did in comparison to the data-based weights. We applied step-by-step aggregation through METEOR, and in order to consider the performance of these weight scemes, we look at the comparability acquisition profiles for the two sets of weights, which can be seen below in Figure 34. We can clearly see that the data based weight scheme outperforms the investigator's equal indicator weight scheme, as the number of comparabilities at each step of aggregation is greater for the data based weight scheme.

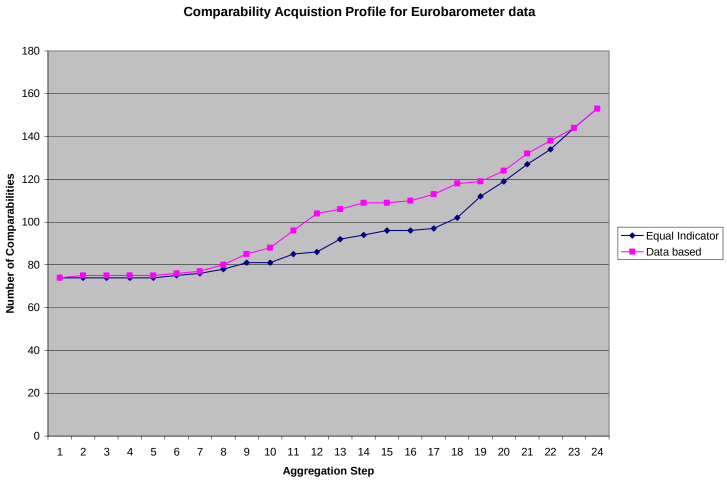


Figure 34

Conclusions and Further ideas:

For the ASC watershed Level II data, from previous work we have a clear performance criterion to compare the accuracy of the weighted index that has been obtained, in the value of rank correlation of the Level II rankings derived from the index with the Level III rankings for the six selected watersheds. But for the other datasets, no such criterion is clear and we can consider using METEOR to better understand and evaluate the two indices. We see that the equal indicator index increases comparabilities quicker than the equal group index, and if our main goal is to increase comparabilities and reduce conflicting views among the indicators, then we can choose to use the equal indicator weights.

The idea of consensus ranking both drives and explains the results of step-by-step aggregation. A general idea of a consensus ranking is the ranking that the data through the indicators seem to conclude, if there are many indicators which seem to give different rankings, then there would be little consensus among those indicators. A subset of indicators may also have some level of consensus as well. We can see that as the number of comparabilities is a measure of consensus among the rankings, as the number of comparabilities increases, there is more agreement among the indicators on the order of pairs, and thus increasing agreement among the total rankings.

By combining indicators into an index, our goal is for the data to reach a consensus ranking, and by using step-by-step aggregation, we come closer to that consensus at each step of the aggregation. All other things approximately equal, we would prefer that the indicators have greater consensus at each step, as greater consensus among the indicators have several advantages. If our goal is to identify the best object, or the best couple of objects, we do not necessarily need to go through the entire method of aggregation, we may stop once there is only one maximal element. When there is greater consensus, it takes fewer steps to reach the point that a maximal object can be identified.

The comparability acquisition profile also gives us a sense of the strength of the consensus ranking. A comparability acquisition profile that accumulates a greater number of comparabilities at each step suggests that the indicators are coming toward a consensus ranking at steps before the final step. If the number of comparabilities are large in the last few steps, this suggests that much of the ordering has been in place from early in the aggregation, and that most of the indicators generally would rank the objects in a similar manner. In particular, if objects A and B were comparable (say A>B) from an early step in the aggregation, it likely implies that most of the indicators considered A>B. On the other hand, the number of comparabilities are rather small in the last few steps, then it implies that either there is substantial disagreement between the indicators on the ranking of the objects and a lack of consensus, or there is an influential indicator.

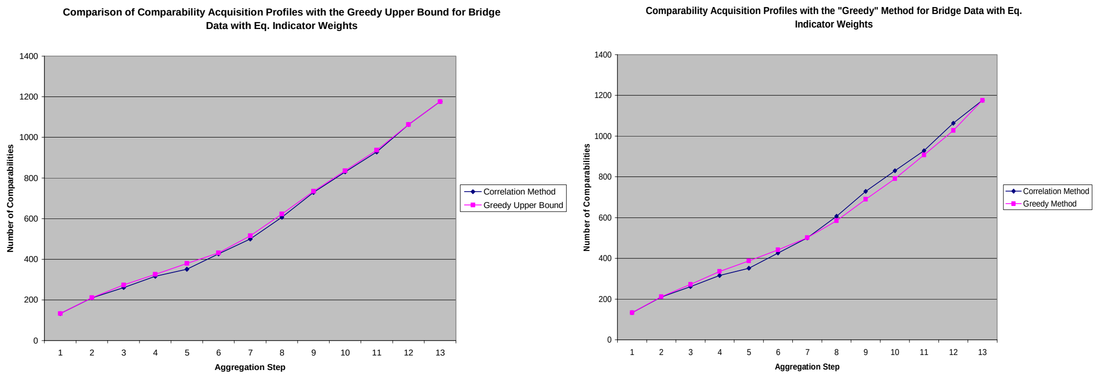
One way that we can understand that strength is by observing that when there is more of a consensus results in fewer possible admissible rankings. (An *admissible* ranking is a ranking that does not conflict with the partial order of the data set, i.e. if object A is considered greater than object B by all the indicators, then a ranking that ranks B greater than A is not admissible.) As the number of comparabilities increases, some potential rankings of the objects become inadmissible as they violate the partial order, and thus the number of admissible rankings decreases. The lesser number of admissible rankings mean that the final index ranking is more "tight", or that there are fewer other rankings that can be possible and that the potential error of the final index ranking is smaller. Thus we have more confidence in the final index ranking if there is a better comparability acquisition profile, which in turn results in the indicators reaching a stronger consensus during steps of aggregation.

Influential indicators are another thorny issue that must be dealt with, we can look at an influential indicator as one that substantially deviates from the consensus of the other indicators. That is, the influential indicator tends to rank objects in a very different manner than would the other indicators, which produces many incomparabilities between pairs of objects (by ranking the two objects in a different orientation than the other indicators) which were comparable among the remaining indicators. Influential indicators tend to have low correlation with other indicators, which makes the influential indicator a prime candidate for being selected early during aggregation.

Consider Indicator 13, channel alignment, from the bridge data set, appears to be part of one of the two indicators aggregated in many of the aggregation steps for the equal group case. This is likely due to the fact that Indicator 13 has a high weight and low correlations with other indicators, and is known to be highly influential using the W-matrix described in the introduction. Since indicator 13 gets an overall weight of 0.25, it is likely to have a strong influence or dominate any aggregated indicator which includes it, while the low correlation that Indicator 13 has with all the other indicators results in the aggregated indicator which includes (and is dominated by) Indicator 13 being selected as one indicator for aggregation at the step. Since Indicator 13 is influential, the aggregated indicator which includes Indicator 13 will give a different opinion that the remaining indicators. Thus unlike most situations, where the general consensus of all the indicators absorbs indicators which differ from the consensus, Indicator 13's influence and relative weight results in Indicator 13 imposing itself on the other indicators. The end result is that the aggregated indicator which contained Indicator 13 at any aggregation step is different from the remaining indicators, which results in fewer comparabilities. Thus it is plausible that the equal group weights index would be have fewer comparabilities at each aggregation step.

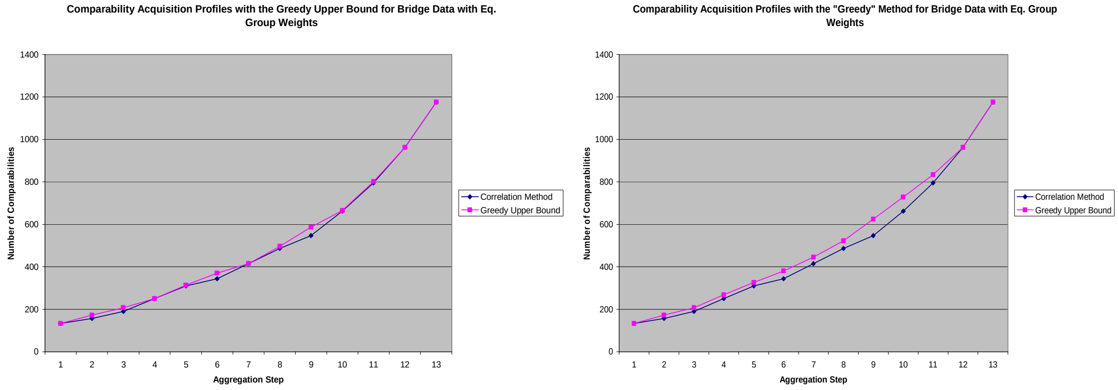
We also wanted to check how well our algorithm of combining the two indicators with the lowest pairwise correlation at each aggregation step accumulates comparabilities. Specifically, we wanted to see how well our method of combining the tow indicators with the smallest Spearman's correlation compares with the "greedy" method, which at each step starts with the indicators from the last aggregation step (using our method until this step) and computes the increase in comparabilities for every pair of indicators, and then find the two indicators which deliver the greatest increase in comparabilities. This method provides an upper bound to the comparability acquisition profile that can be achieved at any step using our method. We also considered a second greedy method, which is to use the greedy method at every step. This is an alternative method, and one which could possibly accumulate fewer comparabilities in the later steps of aggregation as it focuses on short term gain in comparabilities. We applied these performance comparisons to the bridge and EPI data sets and their weighting schemes. We also computed the time (to run the program) to get the results that it takes to actually run these aggregation methods as well to determine if slightly better performance is worth the time it takes.

For the bridge data, we did the performance analysis on the four weighting schemes, equal indicator, equal group, and the two concordance schemes. The comparability acquisition profiles for the equal indicator weighting scheme is in Figure and the comparability acquisition profiles for the equal group weighting scheme is in Figure 35. The left side of the figure is the comparison of the correlation method against the upper bound using the greedy method, while the right-side compares the correlation method with the method that uses the greedy method at every step. For both the equal indicator weighting scheme as well as the equal group weighting scheme, the correlation method of stepwise aggregation seems to come close to the upper bound of maximum comparabilities from that step, and at some steps equaling that upper bound. This is also true for the two data based (concordance) weighting schemes for the bridge data as well. In comparison to the "greedy" method for aggregation, for all of the weighting schemes except for the equal group weights, the "greedy" method yields more comparabilities than our correlation method in the early steps of aggregation, but the reverse is true for the later steps of aggregation, as can be seen on the right graph of Figure 36. The one exception is for the equal group weighting scheme, as can be seen in Figure , that the greedy method does better than our correlation method. On the time issue, the greedy method takes an average of 352 seconds, while the correlation method take an average of 13 seconds to compute the aggregation and the comparabilities of each step.


Comparison of Comparability Acquisition Profiles with the Greedy Upper Bound for Bridge Data with Eq. Indicator Weights

Comparability Acquisition Profiles with the "Greedy" Method for Bridge Data with Eq. Indicator Weights

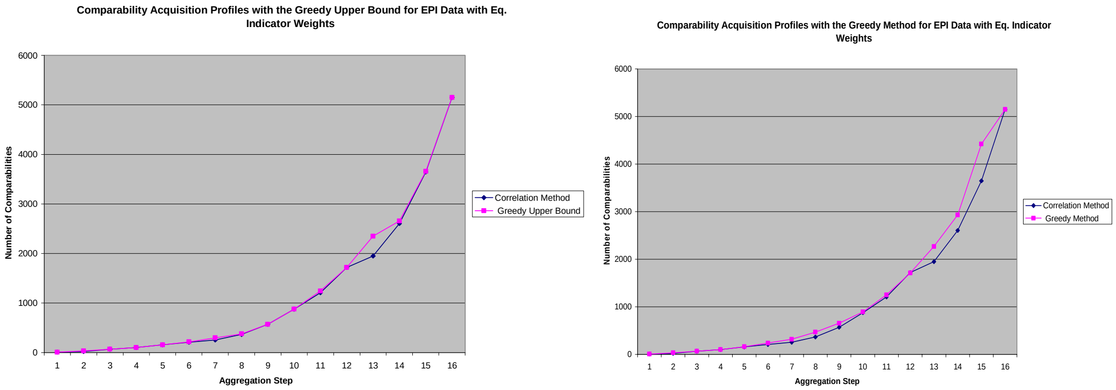
Figure 35


Comparability Acquisition Profiles with the Greedy Upper Bound for Bridge Data with Eq. Group Weights

Comparability Acquisition Profiles with the "Greedy" Method for Bridge Data with Eq. Group Weights

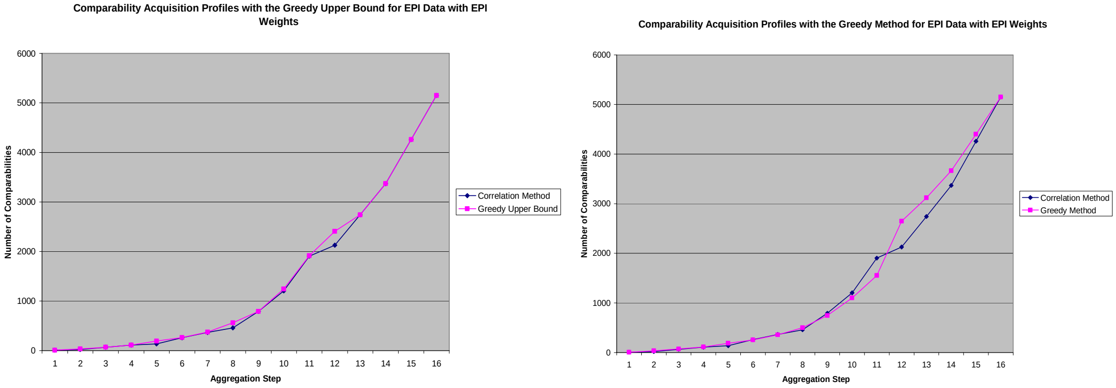
Figure 36

For the EPI data set, we also did this performance analysis on two investigator based weighting schemes, equal indicator weights and the EPI index. The comparability acquisition profiles for the equal indicator weighting scheme is in Figure 37 and the comparability acquisition profiles for the EPI scheme is in Figure 38, with the left and right graphs in each figure the same as in the bridge data. For both of the investigator based weighting schemes, the correlation aggregation method almost meets the upper bound found by the application of the greedy method at each step. For the equal indicator weighting scheme, the greedy method performs slightly better than the correlation method for aggregation. For the EPI index weighting scheme, the greedy method performs better at both the first and last few steps of aggregation, but performs worse for some of the middle steps. As far as time to run the program is concerned, the greedy method takes an average of 970 seconds, while the correlation method take an average of 15 seconds to compute the aggregation and the comparabilities of each step.


Comparability Acquisition Profiles with the Greedy Upper Bound for EPI Data with Eq. Indicator Weights

Comparability Acquisition Profiles with the Greedy Method for EPI Data with Eq. Indicator Weights

Figure 37


Comparability Acquisition Profiles with the Greedy Upper Bound for EPI Data with EPI Weights

Comparability Acquisition Profiles with the Greedy Method for EPI Data with EPI Weights

Figure 38

Using METEOR with Spearman's rank correlation performs almost as well as the greedy upper bound. Further, it is not conclusive that the using greedy method for all steps of the aggregation yields better performance in terms of a comparability acquisition profile than does the correlation method. Finally, it is clear that the greedy method takes much more time than the correlation method, anywhere from a factor 20 to 60 in our data sets, and it is clear that if the number of indicators and objects were increased this factor would be further increased. In light of these findings, we conclude the Spearman's correlation method for aggregation is a good method which obtains near optimum performance and we see no need to use the greedy method.

Further ideas:

We would like to use the results from METEOR to determine better weights. The particularly interesting question is how to deal with influential indicators. Should we increase its weight? Decrease its weight?

We use the concordance method for determining data based weights, but the number of levels in the discretization was selected either by the levels in the existing data or selected based on our intuition. How much of a difference will allowing for more or less levels of discretization make in the determination of the weights and its performance?

One of our goals of the analysis of the ASC watershed data is to find a method using Level I and Level II data to approximate the ranking of the Level III data. We consider the Level III data to be the most accurate representation of the health of the watershed, however collecting Level III data is expensive, and we have Level III data for only 6 selected watersheds. In previous work, we used the weights from data based methods as well as SWR using Level II indicators to obtain a ranking of the 21 objects, and in turn, a ranking of the six selected watersheds that have Level III data. One measure of the performance of each method for our purposes is to compute the correlation of the ranking obtained by the index with the ranking of the Level III indicators, which we consider as the gold standard. We considered several possible indices of Level III data and the rankings obtained from those indices for our gold standard.

We also consider the problem from the other direction, by asking if there exists an index using the seven Level II indicators whose rankings for the six selected watersheds match the Poset rankings of the Level III indicators. If so, we would like to find the possible weights that would produce an index as in (1) that would rank the six selected watersheds in the desired order.

$$\text{Index} = g_1 I_1 + g_2 I_2 + \ldots + g_n I_n \text{, with } g_1, g_2, \ldots g_n \geq 0,\ g_1 + g_2 + \ldots + g_n = 1 \qquad (1)$$

where I_k are indicators.

The issue of measurement error and missing data are also of great interest in unmasking the weight camouflage.

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(s) and do not necessarily reflect the views of the agencies. Our thanks are due to Robert Brooks, Mary Easterling, Jessica Newlin, Paola Annoni for providing data and sharing insights.

References:

Annoni, Paola. Europeans and services of general interest: Where is the best quality perceived? (Manuscript, 2006)

Brooks, R.P., D.H. Wardrop, J.A. Bishop. Assessing wetland condition on a watershed basis in the Mid-Atlantic region using synoptic land-cover maps. Environmental Monitoring and Assessment. 94(1-3): 9-22. (2004)

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

Brüggemann, R., Welzl, G., and Voigt, K. Order theoretical tools for the evaluation of complex regional pollution patterns. *J. Chem. Inf. Comput. Sci*. 43 (2003) 1771-1779

Hebrew University of Jerusalem, Computation Authority, the HUDAP Package. http://ca.huji.ac.il/bf/Hudap-Info.pdf

Johnson, P.A. Preliminary assessment and rating of stream channel stability near bridges. *Journal of Hydraulic Engineering* (2005) 131(10): 845-852

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

Pudenz, S., Brüggemann, R. (2002): A new decision support system:METEOR. In: Voigt, K.,Welzl, G (eds.):Order theoretical tool in evironmental science-Order Theory(Hasse Diagram Technique) meets multivariate statistics. Shaker, Aachen: 103-112.

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

Simon U., Brüggemann R., Mey S., and Pudenz S. METEOR – application of a decision support tool based on discrete mathematics. *MATCH communications in mathematical and in computer chemistry* (2005) 54:623-642

Systat (2004) http://www.systat.com

Voigt, K., Brüggemann, R., and Pudenz, S. Chemical databases evaluated by order theoretical tools. *Analytical and bioanalytical chemistry* (2004)

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

Yale Center for Environmental Policy and CIESIN, Pilot 2006 Environmental Performance Index (2006)