PENNSTATE

Center for Statistical Ecology and Environmental Statistics

Evaluation of Multiple Indicators for Stream Channel Stability Near Bridges in the United States: Multi-Criteria Prioritization and Ranking with Differential Weights, Stepwise Aggregations, Hasse Diagrams, Poset Cumulative Rank Frequency Operators, and Markov Chain Monte Carlo Methods

By G.P. Patil[1], Jessica Newlin[2], K. Sham Bhat[1], and Michael Kase[1]

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

NSF Project Officer: Larry Brandt
EPA Project Officer: Barbara Levinson

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.

[DRAFT]
Technical Report Number 2006-0504
TECHNICAL REPORTS AND REPRINTS SERIES
May 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

Table of Contents

1 Introduction

1.1 General description of bridge program

The data set we use for this paper is a data set of the Federal Highway Administration with 49 bridges with the goal of determining the main causes of bridge failure. 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 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 overall goal of this analysis is to investigate an existing stream stability assessment method to determine if the indicators are necessary and sufficient for the ranking of bridge crossings based on the stability of the channel.

The primary causes of bridges failure are foundation and pier scour, channel movement, and hydraulic forces (Richardson and Huber, 1991). There are many ways to assess the overall risk of bridge failure. In the most recent works, Johnson (2005) proposed an assessment tool for determining the condition of channel stability near bridges. This method is based on a rapid assessment method proposed by Johnson et al. (1999). It provides a systematic, preliminary assessment of stream channel stability with respect to bridge crossings.

1.2 Description of indicators – Methods of data collection

Here we have 13 indicators, which the investigators group into four different groups, a two-level hierarchy with four subindices. The investigators experimented with two indices, one, which was an average of the all 13 indicators, the other the average of the four subindices, which were in turn the average of the indicators in the respective subindex. Our main goals for this data set are to identify which indicators are most influential in determining the rankings of bridges in susceptibility to bridge failure, and ranking of the bridges themselves. 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.

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.

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

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 matrix used in the analysis consists of 49 stream-bridge intersection sites and 13 channel stability indicators.

The data set used in this analysis consists of 49 bridge sites and 13 indicators that are described by Johnson (2005). 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 in Table 1. 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 summing 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.

The thirteen indicators can be categorized by level of expertise and necessary time for evaluation. For different indicator, varying levels of stream channel behavior knowledge is required. Three levels of expertise are assigned in the following manner: (1) 'observation and a moderate level of expertise are required,' (2) 'observation and a high level of expertise are required,' and (3) 'measurement and a high level of expertise are required.' The conceptual model describing the level of expertise is shown in Figure 1.

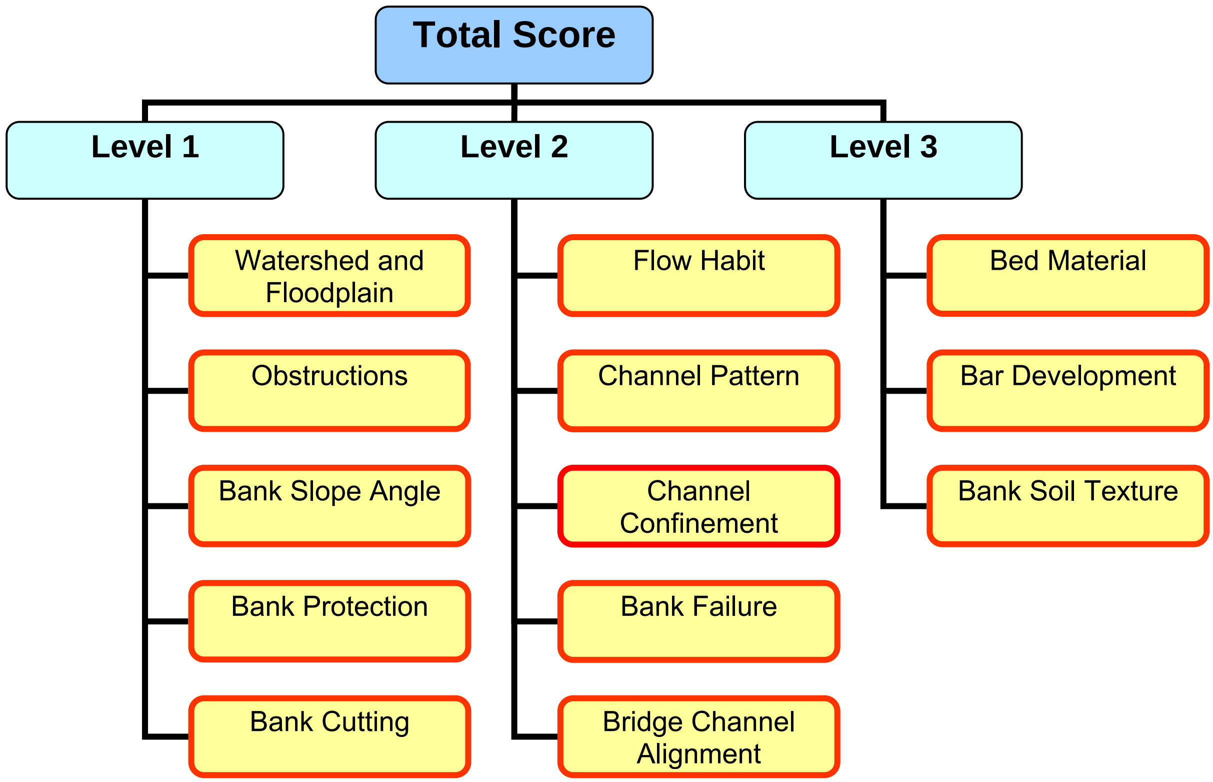


Figure 1 – The conceptual model of three levels of expertise

The indicators can also be grouped based on the particular feature of the stream-bridge intersection that it characterizes. The thirteen indicators describe watershed/regional scale, local channel conditions, bank stability, or bridge alignment. Specifically, the first four indicators: watershed and floodplain activity, flow habit, channel pattern, and channel confinement characterize the stream-bridge intersection on a larger watershed or regional scale. The next three indicators: bed material, bar development, and obstructions, describe the local channel characteristics. Bank soil texture, average bank slope angle, bank protection, bank cutting, and mass wasting describe bank stability. The alignment of the bridge with the stream channel that it crosses is the only indicator describing the alignment of the bridge with the stream channel. The conceptual model of the classification is shown in Figure 2.

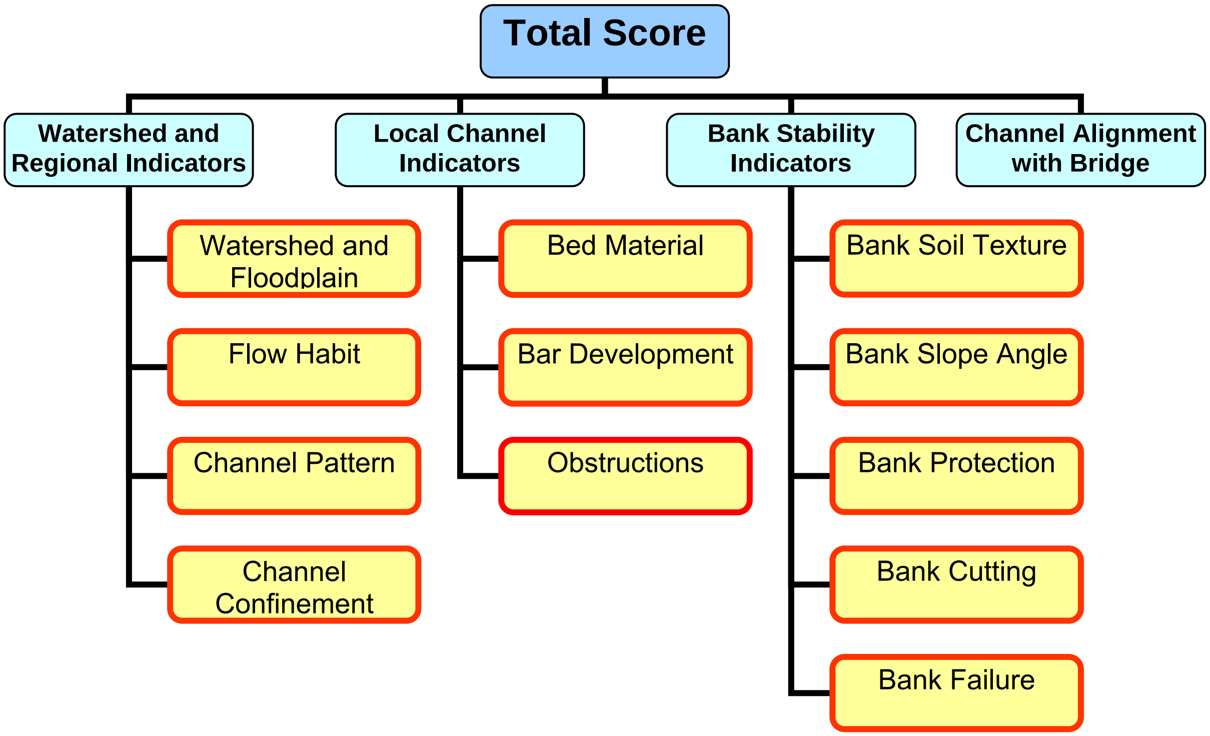


Figure 2 – The Conceptual Model of the Four Bridge Feature Categories

In order to further investigate the stream channel stability in relation with particular bridge features, the thirteen indicators are reduced to four composite indices according to Figure 2. Each of the four indices is a weighted average of the contributing indicators with equal weights. The indices are standardized into a 0-1 scale. The values closer to 0 indicate a more stable stream and the values close to 1 suggest an unstable stream channel near the bridge crossing. The new 49x4 data matrix is evaluated and compared with the original 49x13 data matrix in later sections.

1.3 Research Questions

1. Johnson (2005) uses the total score to rank the stream channel stability near a bridge crossing. How does it compare with the ranking based on the four composite indices? Are there other ways of ranking and do they agree with each other?

The poset prioritization method is able to provide a permissible ranking based on multiple indicators. The method can be applied to both the full data matrix and the reduced data matrix. The ranking results of the two data matrices can be compared with each other by calculating their correlation. The poset rankings can also be used to compare with the rankings based on the total score.

2. The stream channel stability is assigned to one of the four conditions 'Excellent,' 'Good,' 'Fair,' or 'Poor' based on threshold values of the total score. Is this a good classification? Can it be supported by other methods?

The Hasse diagram technique can best visualize the order relations among objects. As one of the most important outputs, Hasse levels can be produced to represent the relative position of an object against others. The Hasse levels can be compared with the four condition intervals to see whether there is a need to adjust the threshold values defining the condition intervals.

3. In general, not all of the indicators are equally important. We are interested to know which indicators are more influential overall, which are more influential for specific stream types, and whether it is appropriate to reduce the number of variables.

The influential indicators can be detected by using one of two methods: calculating a distance matrix or applying the POSAC method. The distance matrix we use is the W-matrix, which measures the change in the number of comparabilities among Hasse diagrams. The POSAC method reduces an N-dimensional data matrix into a two-dimensional space where the two-dimensional coordinates best preserve profile order relations in the original data matrix.

4. The thirteen indicators can be further classified into three groups based on the level of expertise. Can we reduce the cost of implementing a stream stability assessment method by collecting less expensive indicators?

The problem can be solved by applying poset prioritization to the data matrix of different levels of expertise and comparing their results to see whether the indicators requiring lower levels of expertise are adequate to represent the stream channel condition.

5. Given that the analysis of the stream channel stability assessment will be served as a tool for restoration decision-making, how can we make use of the stream stability indicators in determining the appropriate maintenance or mitigation efforts for a specific site?

The specific stream site condition of whether or not mitigation is needed can be obtained by examining the consistency of the stream site with respect to the indicators. The information of consistency can be revealed by doing the rank range run analysis. The rank range run analysis is also a good visualization tool that will provide useful information for restoration decision-making.

2 Description of Analysis Methods

Several multivariate statistical methods have been employed in this analysis to evaluate the data matrix.

2.1 Hasse diagrams

The Hasse diagram technique is used to visualize the order relations of bridges. Visualization of a poset is possible by the technique of Hasse diagrams. Hasse diagram is named after Professor Helmut Hasse, a well-known mathematician in Marburg, Germany. The benefits of the Hasse diagram technique (often named HDT for short in references) come from the assumption that a ranking can be performed while avoiding the use of an ordering index (Halfon and Reggiani, 1986).

Rules to read Hasse diagrams are simply described as following:

- Each object is represented by a small circle
- Objects in an order relation are connected by a straight line
- The object a is located higher than b if $a \leq b$
- a and b are not connected by a line if neither $a \leq b$ nor $b \leq a$ (a and b are incomparable)
- Objects are arranged in levels in order to improve the clarity of the diagram

These rules tell us that a Hasse diagram consists of three parts: circles, levels, and lines.

Circles (or vertexes) are in one-to-one correspondence with the objects in the dataset. Although it is rather flexible in locating the circles, identifying the positions of circles is important in understanding the data structure within a certain data matrix.
Levels are determined in the following fashion. An object of a poset is called maximal if no other elements are strictly greater than it (recall that two objects are comparable if all the attributes of one object are greater than or equal to the other object). The collection of maximal objects forms the top level of the Hasse diagram. Removing the top level, the maximal objects of the remaining poset determine level 2. Level 3 is obtained by removing both level 1 and level 2 and assigning the maximal element of the resulting poset. This procedure is repeated until all the levels are established.

A straight line (or edge) connects a pair of ordered objects (circles). Hasse diagrams are oriented graphs. Objects are in an order relation if they are mutually comparable. In other words, a vertical line connecting a pair of circles implies that all the attributes of the upper circle are greater than or equal to that of the lower circle. On the contrary, lack of connecting lines indicates that there are contradictions in the scoring between objects.

The Hasse diagram technique is best explained by using examples. An arbitrary data set is given in Table 2.2. The example consists of four objects and two attributes. It is obvious that object A is greater than all the other objects. Objects B and C are better than D. However, B and C are incomparable. The Hasse diagram is shown in Figure 2.1. There are three levels in the diagram. Object A is in the top level, B and C are in the middle level, and D is in the bottom level. One may notice that circle A and circle D are not connected by a line, although they are comparable. In a Hasse diagram, to avoid overwhelmed crossing lines, such lines are omitted if the order relations can be implied by transitivity. In this example, D ≤ C (or B) and C (or B) ≤ A implies that D ≤ A. Therefore, the line between A and D is redundant and thus omitted.

Sensitivity analysis can be applied to Hasse diagrams by computing a W-matrix or call it dissimilarity-matrix (Voigt et al, 2004). The W-matrix identifies the influential level of each attribute on the Hasse diagram. The W-matrix can also be seen as a distance matrix. There are many ways to define a distance matrix. The W-matrix measures the change in the number of comparabilities among Hasse diagrams. Brüggemann et al. (2001) explains the idea in more detail.

Attributes / Objects	P	Q
A	1	1
B	0	1
C	1	0
D	0	0

Table 2.2 – HDT Example I

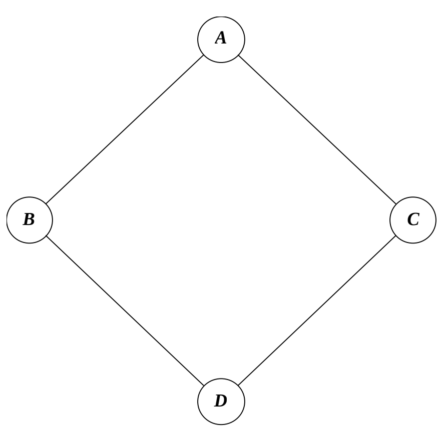


Figure 2.1 – Hasse diagram of example data matrix

2.2 Poset Prioritization

Poset prioritization method is applied to rank the stream sites. The ranking is then compared with the total score which is the summation of all the indicators in the data set. The theory of partially ordered sets and its graphical representation, Hasse diagram, are the solution to a multi-criteria analysis. The methods perform a ranking based on multiple indicators while avoiding having to integrate indicators into an index. They are data driven techniques and can be considered as parameter-free methods.

To compare objects of a data matrix, one needs to evaluate the attributes (often written as a tube *q*(.)) of these objects. Two objects are comparable if all the attributes of one object are greater than or equal to the other objects. For example, we wish to compare object *a* and *b* in a data matrix that consists of *n* attributes, we say $b \geq a$ if $q\ (i,a) \leq q\ (i,b)$ *for all* $i = 1,2,\ldots,n$. If $q\ (i,a) \leq q\ (i,b)$ not *for all i*, the two objects are incomparable. Incomparability implies contradiction in the data of the objects.

Three conditions: reflexivity, antisymmetry, and transitivity must be satisfied at the same time for objects to be set in an order relation.

- Reflexivity: $a \leq a$ for all $a \in S$. (an object can be compared with itself)
- Transitivity: $a \leq b$ and $b \leq c$ implies that $a \leq c$. (if *a* is smaller than *b*, and *b* is smaller than *c*, then *a* is smaller than *c*)
- Antisymmetry: $a \leq b$ and $b \leq a$ implies that $a = b$. (if *a* is not greater than *b*, then the reverse is only true if the two objects are equal)

Generally, if all objects can be mutually compared then they are in a total order (or linear order). However, if not all objects belong to such an order relation, or there is a presence of ambiguity (incomparable objects) within an ordering scheme, it is explicitly denoted by the term partial ordering. Neggers and Kim (1998) have more detailed information on partially ordered sets.

One of the most exciting applications of the partially ordered relation is the poset prioritization and ranking system provided by Patil and Taillie (2004). The system brings together the Hasse diagram, linear extension decision trees, and the cumulative rank function. A linear extension is a total order, which is reproduced from the partial order of the data matrix. The linear extensions quantify the ambiguity of the poset by deducing all the possible comparabilities from the partial order. For the example of the Hasse diagram in Figure 2.1, a set of linear extensions includes two possible total orders (see Figure 2.2). Accordingly, there are two possible rankings: $A \geq B \geq C \geq D$, and $A \geq C \geq B \geq D$.

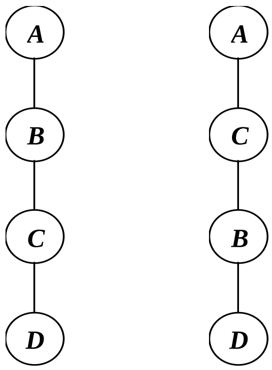


Figure 2.2 – Linear extension of example 1

Following Patil and Taillie's (2004) idea of poset prioritization, the Hasse diagram visualizes the partial order relation of the data so that the possible rankings of objects can be assigned and a linear extension decision tree is formed as a collection of these admissible rankings. Each object has a corresponding interval of possible ranks and the interval is assumed in the *rank-frequency* distribution. A new partial order is defined when stochastic ordering of the *rank-frequency* distribution is imposed and is called the *cumulative rank-frequency* (CRF) ordering. It is then the CRF operator that determines the ranking. Iterations may be needed until the CRF operator reaches linear orderings.

Table 2.3 illustrates a more complex data matrix. The example consists of five objects and two attributes. It is not easy to compare the objects by merely looking at the data matrix even in this small sample. The Hasse diagram helps us to visualize the order relations among objects (see Figure 2.3). In the diagram, A is alone in the top level since it is greater than any other objects in both attributes. Objects B and C come after A in the second level while they are incomparable to each other. D and E sit in the bottom level, whereas E is comparable to both B and C in its upper level but D is only comparable with B. Figure 2.4 illustrates the linear extensions in the form of a decision tree. There are five paths in this tree. Every downward path through the decision tree determines a linear extension. Solid lines in the decision tree connect comparable objects, whereas dashed lines mean that the linked objects are not implied by the partial order. Examining this decision tree, we see that the object A is ranked 1 in all five linear extensions. Object B is ranked 2 in three linear extensions and is assigned rank 3 in the other two. Object C is ranked 2 in two linear extensions, 3 in two linear extensions, and 4 in one linear extension. The rank distributions for other objects are summarized in Table 2.4. Note that the summation of each row should be equal to the number of linear extensions, which in this example, is five.

The cumulative rank frequency (CRF) is defined to further investigate the order relation of the data matrix. The concept of cumulative rank frequency, as the name suggests, is the accumulation of the relative rank frequency. The CRF of the example is shown in Table 2.5. We can also derive the ordinal ranking probability by dividing the cumulative rank frequency over the total number of linear extensions (see Table 2.6). The cumulative rank frequencies (same as the ordinal ranking probabilities) for this poset example are plotted in Figure 2.5. It is evident that the CRF curves are stacked one above the other, which gives a total order of the objects: A > B > C > D > E.

Objects \ Attributes	P	Q
A	5	5
B	4	3
C	2	4
D	3	1
E	1	2

Table 2.3 – Poset example II

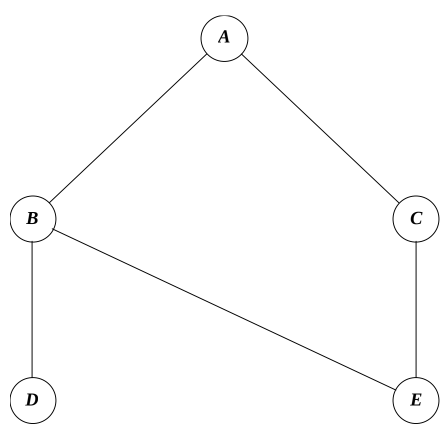


Figure 2.3 – Hasse diagram of example 2

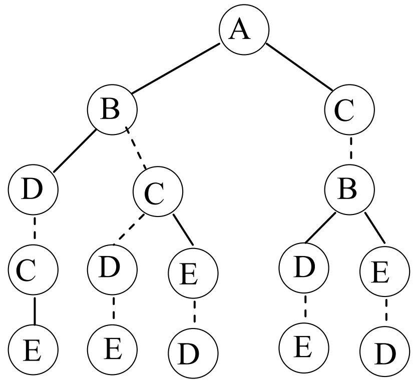


Figure 2.4 – Linear extension decision tree

	Rank					
Element	1	2	3	4	5	Totals
A	5	0	0	0	0	5
B	0	3	2	0	0	5
C	0	2	2	1	0	5
D	0	0	1	2	2	5
E	0	0	0	2	3	5
Totals	5	5	5	5	5	

Table 2.4 – Rank-frequency table for the poset example

	Cumulative Rank				
Element	1	2	3	4	5
A	5	5	5	5	5
B	0	3	5	5	5
C	0	2	4	5	5
D	0	0	1	3	5
E	0	0	0	2	5

Table 2.5 – Cumulative Rank-frequency table for the poset example

	Ordinal ranking probability				
Element	1	2	3	4	5
A	1	1	1	1	1
B	0	0.6	1	1	1
C	0	0.4	0.8	1	1
D	0	0	0.2	0.6	1
E	0	0	0	0.4	1

Table 2.6 – Ordinal ranking probabilities

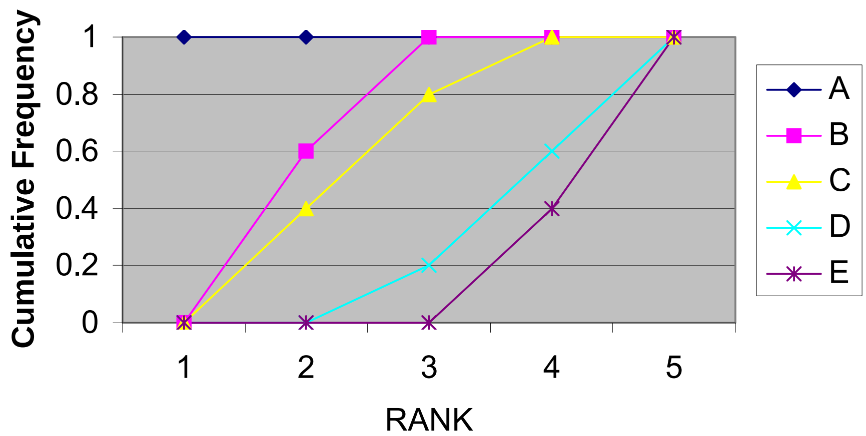


Figure2.5 – CRF plot

It is satisfactory if the cumulative rank frequency can provide a total order for every poset. However, in most cases, the CRF is not adequate to produce a linear ordering in the first attempt. Patil and Taillie (2004) demonstrated an example where two CRF iterations are needed to generate a total order (see Figure 2.6). In the case that a single CRF procedure fails to produce a vertical linear order (total order), the CRF should be applied repeatedly until the iteration determines a linear order.

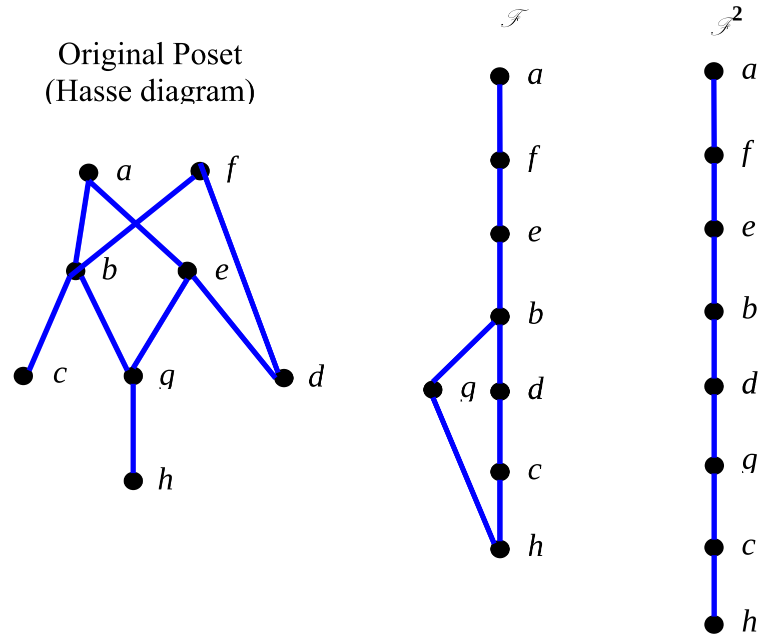


Figure 2.6 – Two CRF iterations

In order to produce the final total order of a poset, one needs to enumerate all the linear extensions of the poset. However, the number of linear extensions grows dramatically as the number of objects or the number of dimensions increases. In an extreme case, the number of linear extensions depends on the number of objects in the data matrix. By applying the combinatorial rule, the magic number reaches N! permutations. Even in a small data set of five objects, the number of all possible linear extensions is 5! = 120. This fact makes it practically impossible to identify all possible linear extensions for a larger poset. When the poset is too big for the cumulative rank function to handle, Markov Chain Monte Carlo (MCMC) method can be used to estimate the rank frequency distribution.

2.3 POSAC

Influential indicators have been identified by investigating the W-matrix and by performing POSAC.

Patially Ordered Scalogram Analysis with Coordinates (POSAC) is a kind of profile analysis. The data in each row of the data matrix is defined as the subject's profile. Order relations between profiles can be determined by comparing values in each attribute. Given two different profiles, one profile is said to be greater than the other only if all the attributes of one object are greater than or equal to those of the other object. The two profiles are incomparable when there are contradictions in the values of attributes. 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.)

There are three possible order relations in a two-dimensional Cartesian coordinate space. The possibilities are indicated in Figure 2.7. 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*.'

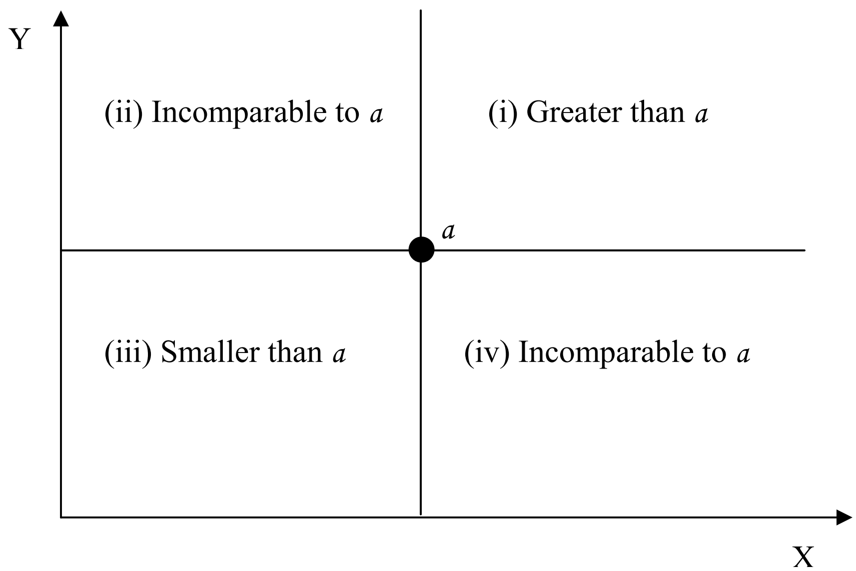


Figure 2.7 – Order relations in two-dimensional space

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. In an example of three profiles with four attributes:
3142
3242
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 profile 3142 and 3242, it must be assigned within the intersection of regions that incomparable to both (1, 1) and (2, 2). That is the shaded area in Figure 2.8. 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.

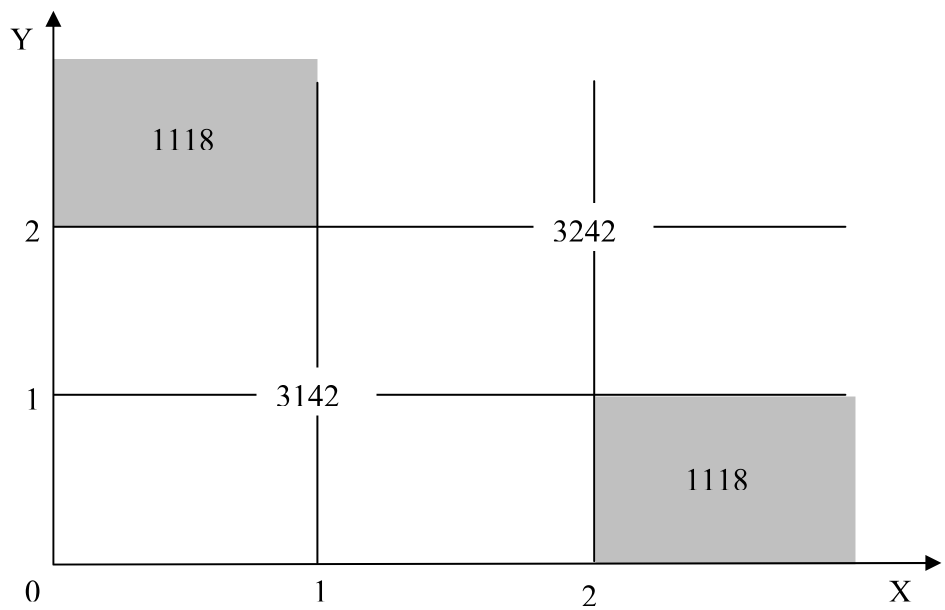


Figure 2.8 – Three profiles in a two-dimensional space

The algorithm of the POSAC method determines that not every profile can be accurately located in the two-dimensional coordinate space. With increasing number of profiles, misrepresentation becomes a problem of POSAC. The case study in this paper applied the program package SYSTAT 11 (SYSTAT, 2004) under the feature of analysis, scale. The POSAC program provides a coefficient, which describes the proportion of comparability that is correctly represented. The program is aimed to minimize the loss of comparability. The detailed background and mathematics of POSAC is described in Shye (1985).

2.4 2-Dimensional MPOSAC

There is also a multivariate form of POSAC, MPOSAC (Hebrew University of Jerusalem, 2002), where more than two latent order variables can be used. Only the two-dimensional MPOSAC will be used however, as a check and balance on POSAC results. Results from MPOSAC will be shown beside the results of POSAC.

2.5 One-Way ANOVA F-scores

Due to the interest in understanding the strength of the influence of the original indicators on the LOVs, approximate "loadings" are found by computing the F-value of the ANOVA test with the LOV as the response variable and the indicator as the factor. For the ASC Level II data, these F scores were computed for all seven indicators for both LOV variables. To allow for small deviations in the POSAC algorithm, we discretized both the LOVs and the original data into eight equally spaced intervals, and gave an appropriate score, for example, values between 0 and 0.125 would get a score of 1, between 0.125 and 0.25 get a score of 2, etc

We computed approximate loadings for the two latent order variables in order to understand which indicators most influence the latent order variables in POSAC. The method used was the F-value of the latent order variable with the respective indicator as a factor. A query is whether this would be an effective method, since POSAC does not necessarily have to be unique for a particular set of data, and we may get slightly different values for LOV1 and LOV2. If the POSAC diagrams are different, the loadings might not be robust to deviations of the latent order variable values if the F-test method is used. We might need to consider methods that are more robust.

2.6 Spearman's Correlation Coefficient

The bridge ranks computed from the latent order variables are compared to the bridge ranks from each indicator and the Spearman's rank correlation coefficient is calculated.

2.7 Concordance

The concordance is a ratio obtained from the number of bridges where the latent order variable correctly corresponds with each individual indicator value over the total number of bridges.

2.8 METEOR

When the need is to combine a final linear ordering with the understanding of the effect of each indicator's effect, we can use METEOR. METEOR allows for systematic aggregation of the indicators using user-defined weights. In an attempt to find the influence of the weights during aggregation of indicators on the Hasse Diagram, we can use stability field analysis, which finds all possible Hasse Diagrams for two pairwise aggregations. The triangular plot stability fields is an improvement in this regard as we can aggregate three indicators into one.

This method has the goal of step-by step aggregation of all the indicators with the eventual result of a linear ordering of all the objects in the model. The idea is that at each step multiple indicators (often two) are merged together by a weighted linear combination, i.e. $I=g_1*I_1+(1- g_1)* g_2$. The aggregation of the two indicators increases the number of comparabilities in the Hasse diagram as contradictory rankings between the two indicators are eliminated. The eventual result appears to be equivalent to using a weighted linear combination or an index of all the indicators.

One major motivation for aggregation is to deal with situations with a natural hierarchical ordering. 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 chemical for each method. The township can aggregate these four indicators to achieve a final ranking of these ten methods. Now the county would like to find the best method to deal with the proliferation of these four chemicals. It would make sense to aggregate the for the cities for Chemical 1 to get a county level value for Chemical 1 for each of the 10 methods. Having county level values for all four chemicals, we can then aggregate these four county-level indicators to produce a full ranking of the ten methods 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.

The method of step-by-step aggregation appears to have the advantage of understanding the effect of the relative weights of the indicators, specifically one can understand the influence of the indicator through the increase of the number of comparabilities in the Hasse diagram. 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.

2.9 Rank Range Run

Furthermore, the rank range run method is applied in determining sites that require future mitigation and restoration.
The rank range run (RRR) method was developed and programmed by Myers et al. (2005). The method can be considered as a supplement of poset prioritization and the Hasse diagram technique. The RRR method converts the indicators to ranks and gets some descriptive statistics from the new dataset (ranks-wise). The method is motivated to make sense of ranking by using direct evidence (actual ranks of indicators) as opposed to the idea of poset prioritization, where ranking is determined by computing the linear extensions. The linear extension is a kind of indirect evidence and is sometimes hard to interpret.

As its name suggests, the rank range ran method particularly concerns a descriptive statistic called 'rank range.' Rank range is defined as the difference of the best rank and the worst rank of a unit. Units with narrow range of ranks reflect consistency among the indicators, while a wide range of ranks indicates disagreement among the indicators. The rank range run is very powerful in visualizing the order relations of a data matrix. A standard figure of the rank range run is shown in Figure 2.9. Series of RRR are developed in an order of increasing rank range. Within series, units are arranged in an order of increasing minimum ranks. The length of each vertical line represents the rank range of the unit. Median rank of each unit is pointed as a dot on the line. The algorithm of the rank range run determines that the figure shows less consistency from left to right in between series (the rank range increases as sequence of runs is increasing). Within series, the units go from the best to the worst in the same level of consistency.

Combining the rank range run with the Hasse levels is found to be very useful in environmental science, especially when environmental mitigation and restoration is of concern.

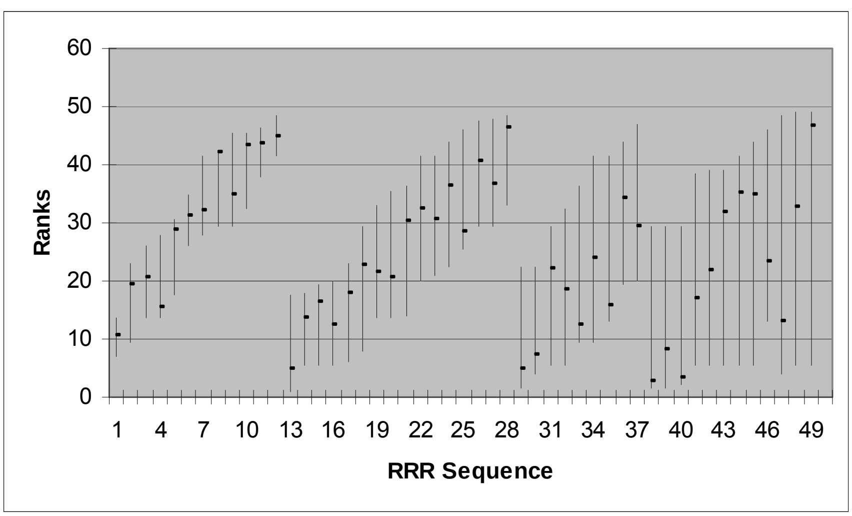


Figure 2.9 – Example Figure of 'Rank Range Run'

3 Analysis Results

Before applying the poset prioritization method, we rank the stream sites according to their total score. The rankings of total score are compared between the 13-indicator data matrix and the 4-index data matrix. The scatter plot (see Figure 3) shows a fairly good match between the two rankings, and a correlation of 0.94 suggests that grouping the thirteen indicators into four feature categories does not affect the order relations of the total score.

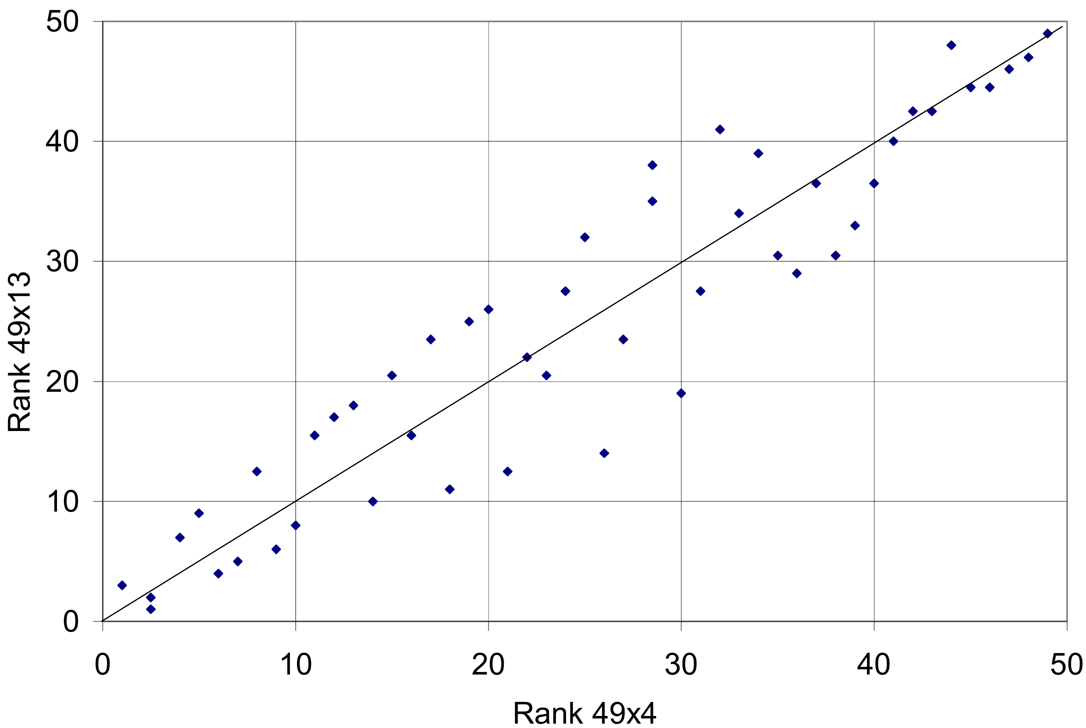


Figure 3 – Comparison of ranking based on total score for the 13-indicator data set and the 4-index data set. Correlation = 0.94

3.1 Hasse diagrams

The Hasse diagram is a graphical representation of a partially ordered set. Figure 6 is the Hasse diagram for the 13-indicator data set, labeled by overall stream condition as determined from the total score (sum of all 13 indicator scores), according to Johnson (2005). The top level on the diagram consists of 28 objects. They represent the best sites and are called maximal objects in the Hasse diagram technique. They are “better” than those objects that are in the lower levels. Similarly, the minimal object means that no other object is worse than it. There are sixteen minimal objects in this diagram. Among the sixteen objects, three stream sites, 15, 17, and 36, are represented in the bottom level and are therefore the worst sites. Six out of the sixteen minimal are isolated objects shown in the top level. It implies that these stream sites are incomparable with any other sites. The distribution of the 49 stream sites in the three Hasse levels is illustrated in Figure 7.

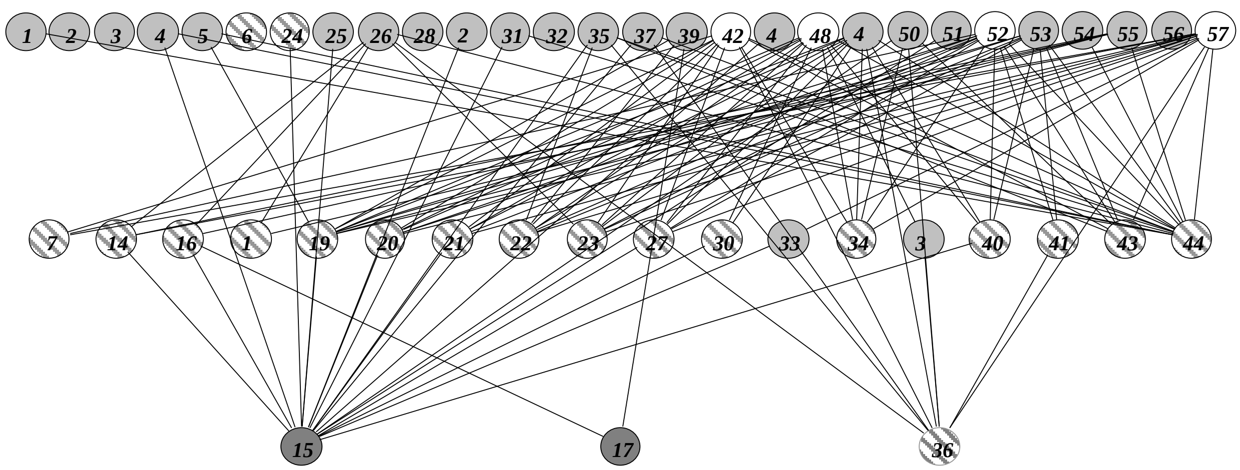


Figure 6 – Hasse diagram for 49 sites with 13 indicators. The fill patterns represent the stream condition assigned based on the total score of summing all 13 indicators. (– *Poor*, – *Fair*, – *Good*, – *Excellent*.)

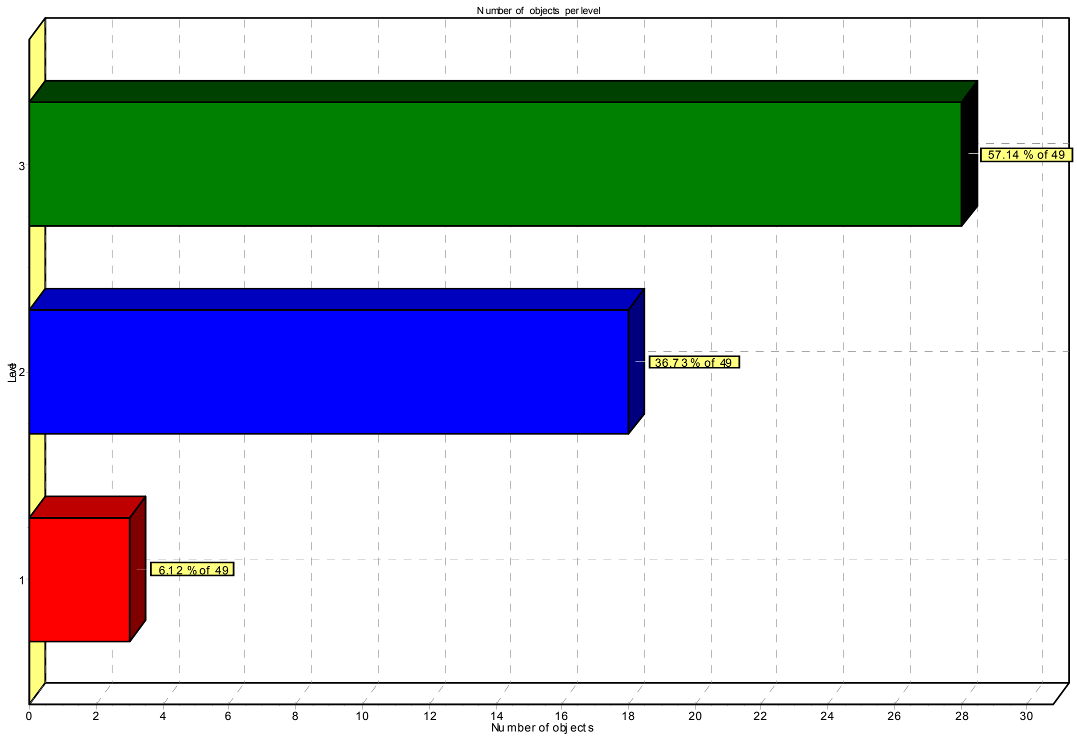


Figure 7 – Level structure of the Hasse diagram for 49 sites with 13 indicators

Figure 8 is the Hasse diagram for the 4-index data matrix. It consists of seven Hasse levels and no isolated objects. There are nine maximal objects in the top level. Six objects are minimal, among which is stream site 15. The distribution of the 49 stream sites in the seven Hasse levels is illustrated in Figure 9. Generally, the decrease in the number of indicators from 13 to 4 composite indices has resulted in an increase in comparability between study sites with 133 possible comparisons in the 13-indicator data set and 460 possible comparisons in the 4-index data set.

The assignment of an overall stream condition based on the total score (sum of all 13 indicator values) seems to be supported by the Hasse diagram technique. In Figure 6, all of the 'Excellent' sites are on the top row, while all of the 'Poor' sites are on the bottom row. Neither Figures 6 nor 8 show any case where a site with a better stream condition is considered better than a site with a worse stream condition based on the displayed comparabilities.

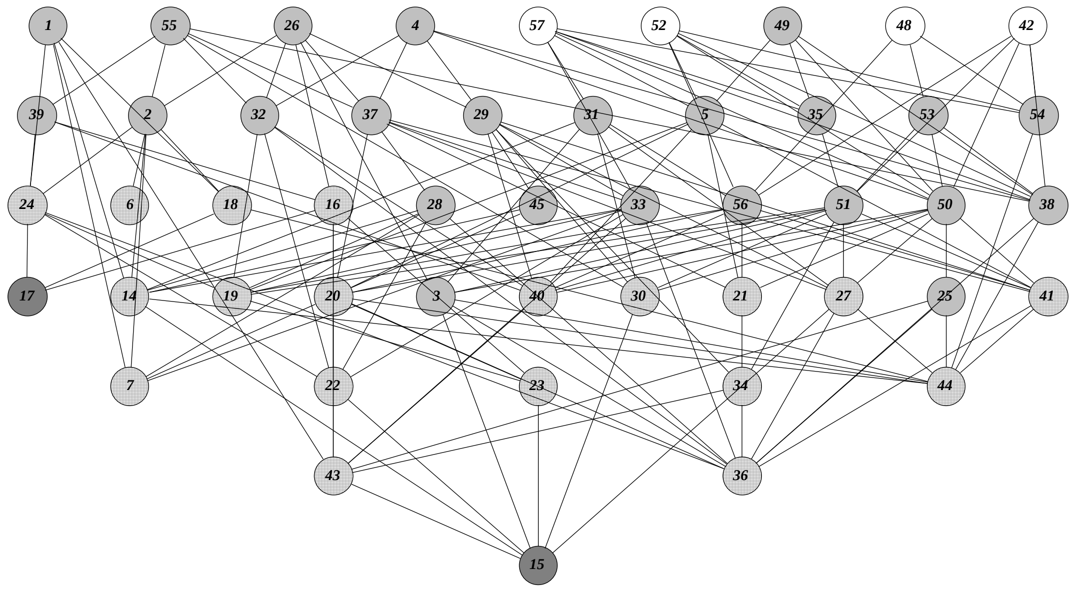


Figure 8 – Hasse diagram for 49 sites with 4 indices. The fill patterns represent the stream condition assigned to the sites based on the total score of summing all 13 indicators. (– *Poor*, – *Fair*, – *Good*, – *Excellent*)

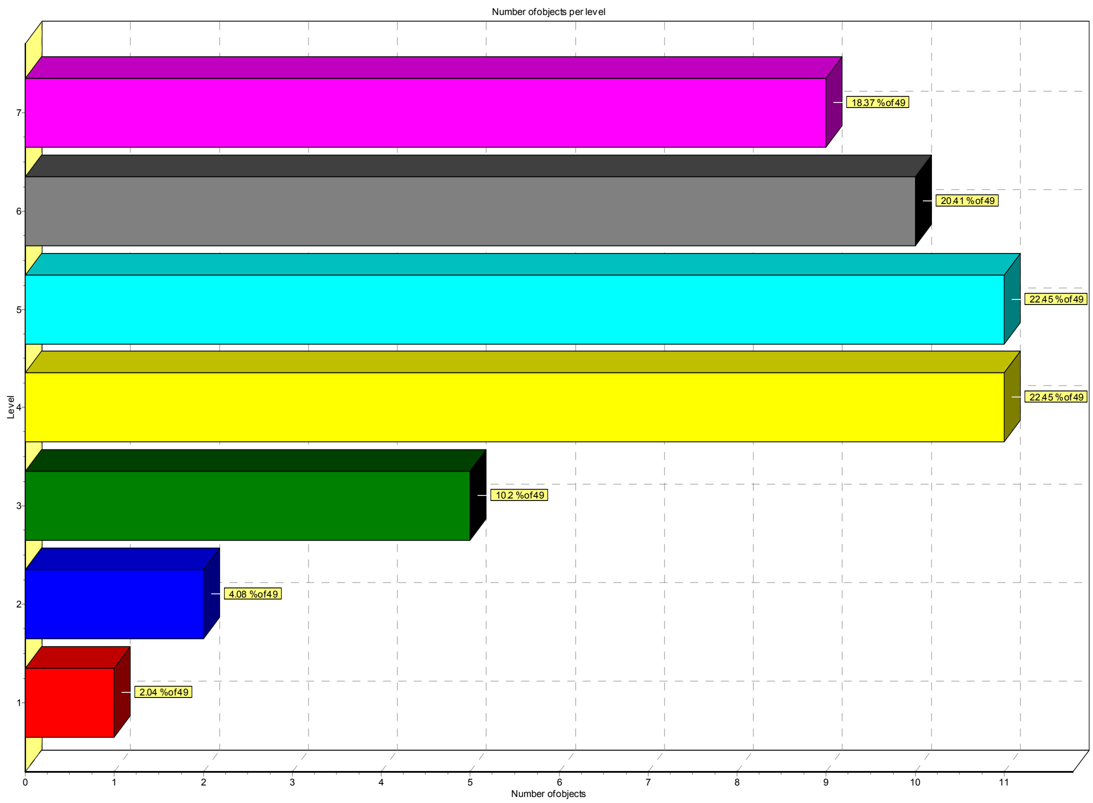


Figure 9 – Level structure of the Hasse diagram for 49 sites with 4 indicators

The W-matrix is the key for the sensitivity analysis of the Hasse diagram technique. It is used to determine the influence of indicators by quantifying the dissimilarity of different Hasse diagrams. The W-matrix is an N x N symmetric matrix where N represents the number of cases and is equal to the number of attributes plus one. The rows and columns of the matrix represent the Hasse diagrams of the data with a particular indicator removed, with exception that row and column 0 represent the full Hasse diagram with no indicators removed. The entries in the matrix are the difference in the number of possible comparisons between the Hasse diagram of the column and the Hasse diagram of the row of that entry. The W-matrix was computed using the WHasse program (Voigt 2004). Table 2 shows the W-matrix for the 13-indicator data set. In this table, case 0 indicates that the Hasse diagram is plotted using all of the 13 attributes, whereas case *i* represents the Hasse diagram with the *i*-th attribute removed. Again, the W-matrix is a large symmetric matrix which does not always need to be analyzed in its entirety. For our purpose, we are more interested in the first row of the W-matrix rather than the full matrix. This is because the first row tells us the difference in comparabilities between the full Hasse diagram and each of the Hasse diagrams with the particular indicator removed. The first row of this W-matrix tells us that the removal of indicator 13, channel alignment, gives a Hasse diagram with a difference of 85 comparabilities compared to the Hasse diagram that includes all of the indicators. This is greater than the difference between the Hasse diagram with any other indicator removed and the Hasse diagram with no indicator removed. So, channel alignment is considered the most influential indicator using the W-matrix method. The next most influential indicator is indicator number 8, Bank Soil Texture, which causes a change of 32 comparabilities from the Hasse diagram that includes all of the indicators.

The W-matrix method has also been applied to the 4-index data set (Table 3). Channel alignment again comes out to be the most influential indicator, followed by local channel characteristics, watershed and regional characteristics, and bank stability.

The changes in the number of possible comparisons are also visualized in Figure 10 and Figure 11 for both data sets.

In order to identify the most influential indicators for specific stream types, the full data matrix has been divided according to the following stream classifications: dune-ripple, riffle-pool, plane-bed, and modified (Montgomery and Buffington 1997; U.S. Army Corps of Engineers 1994). The W-matrix results are summarized in Table 4.

The indicator for channel alignment appears as the most or second most influential indicator for all data sets summarized in Table 4. Properties of the bank, bank soil texture, bank protection, and bank cutting are influential indicators for riffle-pool type streams. This is consistent with the description of riffle-pool streams having banks that act as a main roughness feature for this channel type (Montgomery and Buffington 1997). The obstructions indicator is influential for the plane-bed type streams. This is also consistent with Montgomery and Buffington's (1997) description of plane-bed streams, which states that the introduction of flow obstructions may force local pool and bar formations that would otherwise not be present in this stream type.

The influence, as given by the W-matrix method, appears to be related to the correlation between indicators. Indicators that are the least correlated with other indicators appear to be more influential according to the W-matrix. For example, the most influential indicator, channel alignment, is not correlated with any of the other indicators (Table 5).

W ij	Case 0	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10	Case 11	Case 12	Case 13
Case 0	0	6	4	7	11	9	6	25	32	1	18	7	4	85
Case 1	6	0	10	13	17	15	12	31	38	7	24	13	10	91
Case 2	4	10	0	11	15	13	10	29	36	5	22	11	8	89
Case 3	7	13	11	0	18	16	13	32	39	8	25	14	11	92
Case 4	11	17	15	18	0	20	17	36	43	12	29	18	15	96
Case 5	9	15	13	16	20	0	15	34	41	10	27	16	13	94
Case 6	6	12	10	13	17	15	0	31	38	7	24	13	10	91
Case 7	25	31	29	32	36	34	31	0	57	26	43	32	29	110
Case 8	32	38	36	39	43	41	38	57	0	33	50	39	36	117
Case 9	1	7	5	8	12	10	7	26	33	0	19	8	5	86
Case10	18	24	22	25	29	27	24	43	50	19	0	25	22	103
Case11	7	13	11	14	18	16	13	32	39	8	25	0	11	92
Case12	4	10	8	11	15	13	10	29	36	5	22	11	0	89
Case13	85	91	89	92	96	94	91	110	117	86	103	92	89	0

Table 2 – W-matrix for 13-indicator data set

W ij	Case 0	Case 1	Case 2	Case 3	Case 4
Case 0	0	88	99	63	324
Case 1	88	0	187	151	412
Case 2	99	187	0	162	423
Case 3	63	151	162	0	387
Case 4	324	412	423	387	0

Table 3 – W-matrix for 4-index data set

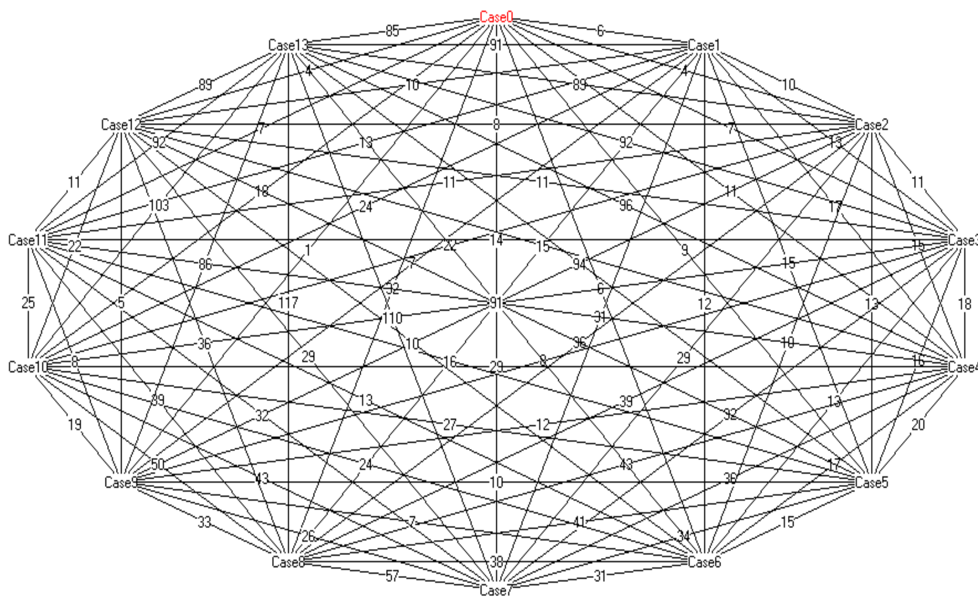


Figure 10 – Change of comparabilities within Hasse diagrams for 13-indicator data set

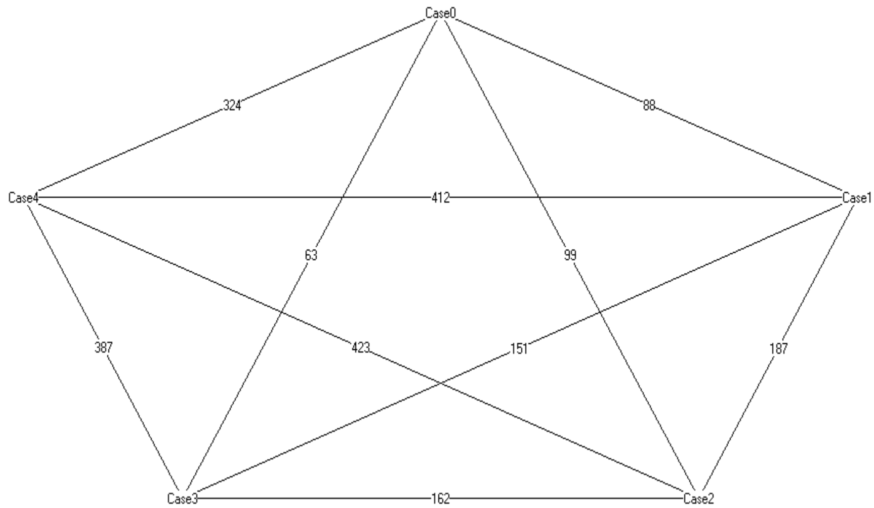


Figure 11 – Change of comparabilities within Hasse diagrams for 4-index data set

Data Set	**Order of Influence**				
	1	**2**	**3**	**4**	**5**
13-indicators; All Sites	**Channel Alignment**	**Bank Soil Texture**	**Obstructions**	**Bank Protection**	**Channel Confinement**
4-index; All Sites	**Channel Alignment**	**Local Channel Characteristics**	**Watershed and Regional Characteristics**	**Bank Stability**	**--**
13-Indicators; Riffle-pool	**Bank Soil Texture**	**Channel Alignment**	**Bank Protection**	**Bank Cutting**	**--**
13-Indicators; Dune-ripple	**Obstructions**	**Channel Alignment**	**Bank Soil Texture**	**--**	**--**
13-Indicators; Plane-bed	**Channel Alignment**	**Channel Pattern**	**Bed Material**	**Obstructions**	**Channel Confinement**
13-Indicators; Modified	**Channel Alignment**	**Bar Development**	**Obstructions**	**--**	**--**

Table 4 – Influential parameters for various data sets in order of influence with 1 being the most influential (If five influential parameters were not found, a '--' is entered in the table.)

	Watershed and Floodplain Activity	*Flow Habit*	*Channel Pattern*	*Entrenchment or Channel Confinement*	*Bed Material*	*Bar Development*	*Obstructions*	*Bank Soil Texture*	*Average Bank Slope Angle*	*Bank Protection*	*Bank Cutting*	*Mass Wasting or Bank Failure*	*Bridge-Channel Alignment*

Watershed and Floodplain Activity	*1.000*												
Flow Habit	*0.596*	*1.000*											
Channel Pattern	*0.533*	*0.641*	*1.000*										
Entrenchment or Channel Confinement	*0.374*	*0.280*	*0.218*	*1.000*									
Bed Material	*0.530*	*0.572*	*0.493*	*0.401*	*1.000*								
Bar Development	*0.474*	*0.286*	*0.341*	*0.554*	*0.589*	*1.000*							
Obstructions	*0.192*	*-0.158*	*0.219*	*0.467*	*0.190*	*0.399*	*1.000*						
Bank Soil Texture	*0.473*	*0.551*	*0.598*	*0.204*	*0.551*	*0.165*	*0.174*	*1.000*					
Average Bank Slope Angle	*0.470*	*0.261*	*0.231*	*0.804*	*0.372*	*0.553*	*0.421*	*0.148*	*1.000*				
Bank Protection	*0.426*	*0.378*	*0.374*	*0.570*	*0.363*	*0.429*	*0.245*	*0.234*	*0.702*	*1.000*			
Bank Cutting	*0.486*	*0.207*	*0.263*	*0.574*	*0.328*	*0.488*	*0.454*	*0.152*	*0.729*	*0.570*	*1.000*		
Mass Wasting or Bank Failure	*0.385*	*0.198*	*0.328*	*0.705*	*0.397*	*0.610*	*0.505*	*0.274*	*0.795*	*0.722*	*0.703*	*1.000*	
Bridge-Channel Alignment	*-0.062*	*-0.009*	*0.026*	*-0.035*	*-0.119*	*0.165*	*0.051*	*-0.134*	*0.169*	*0.079*	*0.277*	*0.132*	*1.000*

Table 5 – Correlation matrix for 13-indicator data set

3.2 Poset Prioritization Ranking

Since assigning equal weight to each indicator is rather arbitrary, the poset prioritization is applied to the two data matrices instead of summing all of the indicator scores into a single index (total score). Figure 4 compares the ranking based on total score with the ranking using poset prioritization for the 13-indicator data matrix. The two rankings show a good consistency with a correlation of 0.92. The figure also shows that the total score tends to give higher rankings to the stream sites than the poset prioritization does, especially for those sites with poorer stream channel conditions (the sites with ranks above 20). The result shows a similar pattern for the 4-index data matrix (Figure 5). The two rankings of the 4-index data matrix show more consistency than the 13-indicator data matrix (correlation = 0.98), and again the ranking based on total score is found to be consistently greater than the poset prioritization ranking for the sites that are in worse stream channel conditions.

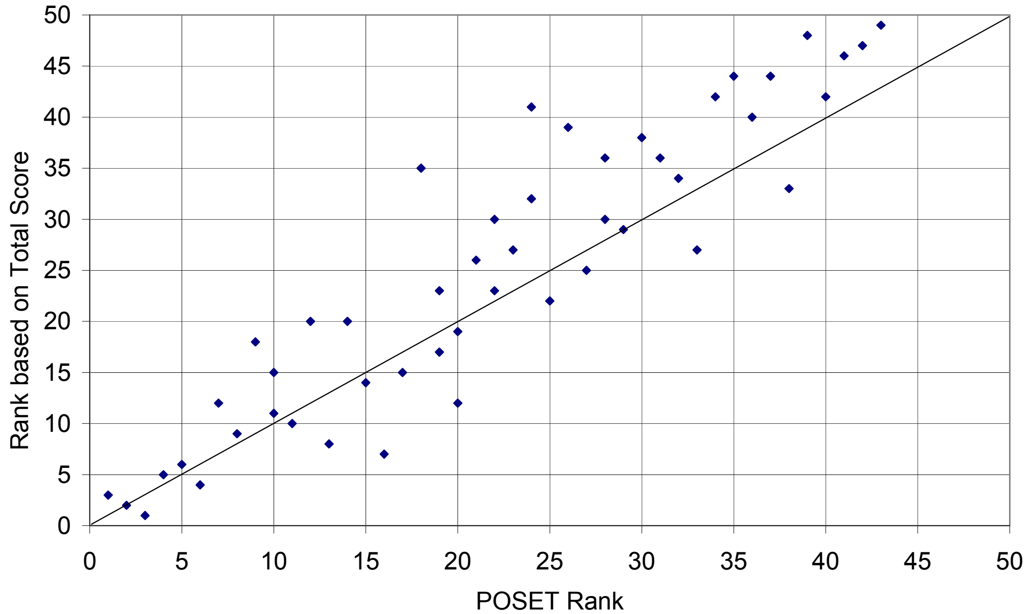


Figure 4 – Comparison of the rank of a site based on total score and based on poset prioritization for the 13-indicator data matrix.

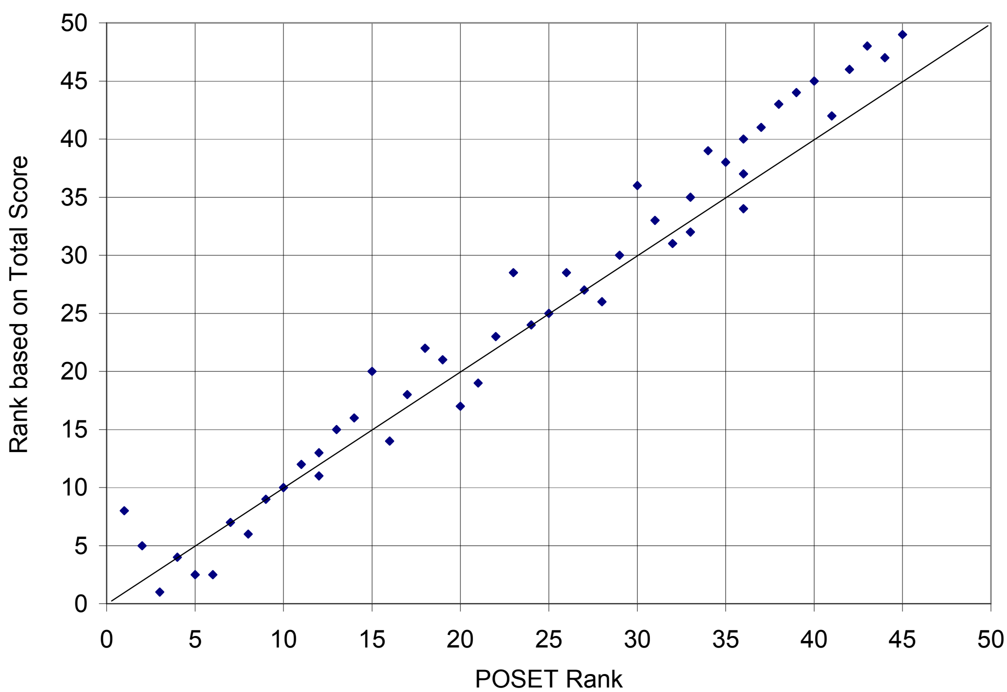


Figure 5 – Comparison of the rank of a site based on total score and based on poset prioritization for the 4-index data matrix.

Poset prioritization was applied to the level 1 indicators and also the combination of the level 1 and level 2 indicators. Figure 17 shows the agreement between using only the level 1 indicator for ranking and using all 13 indicators for ranking. Figure 18 is the relationship between using only level 1 and 2 indicators for rank and using all 13 indicators for ranking. The correlation of 0.92 for rank based on all indicators and levels 1 and 2 indicators is fairly high. The relationship between the ranking using only the level 1 indicators and the ranking using all 13 indicators is not as good as using both level 1 and 2 indicators (correlation = 0.79). A stream assessment team may consider reducing the cost of stream stability assessment by including only level 1 and 2 indicators in their analyses.

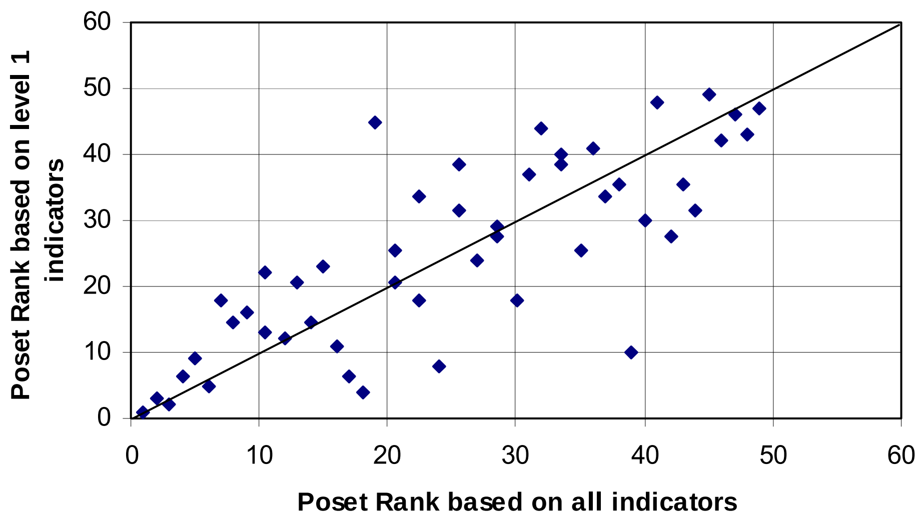


Figure 17 – Rank based on level 1 indicators versus rank based on all 13 indicators (Correlation = 0.79)

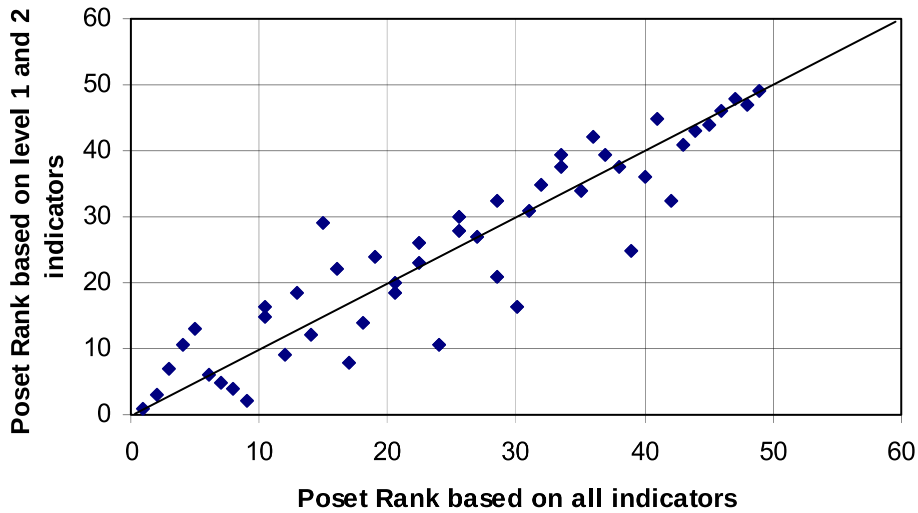


Figure 18 – Rank based on level 1 and 2 indicators versus rank based on all 13 indicators (Correlation = 0.92)

3.3 POSAC

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 (LOV1 and LOV2) found by POSAC. From Figure 3.1.1, we see that most of the bridges have either high LOV1 values and medium to low LOV2 values or medium to low LOV1 values and high LOV2 values. Bridge 57 appears to have both high LOV1 and LOV2 values, at about 0.8, while bridge 15 appears to have both LOV1 and LOV2 values of less than 0.5. Both of these bridges are consistent with the other analysis done on these bridges, bridge 57 has HasseDiagram maximality level 1, is rated as an "excellent" bridge by its index ranking, and has a low Poset ranking. Bridge 15, on the other hand, has HasseDiagram maximality level 3 (the lowest level for this model), is rated as "poor" by its index ranking, and has a high Poset ranking.

This brings us to the latent order variables themselves, as we wish to explore the relationship of LOV1 and LOV2 with the 13 indicators. We computed approximate loadings, analogous to the loadings in PCA (Principal Component Analysis) by computing the F-values of an ANOVA test of LOV1 (and LOV2) as the dependent variables, and the 13 indicators, one-by-one, as the independent variables (or factors). The loadings tell us the strength of the impact on the latent order variable by that indicator. In addition, the correlations between LOV1 and LOV2 and the indicators were computed, which gives us information about the direction of the relationship between the latent order variable and the indicator. The F-values and the correlations can be found in Table 3.1.2.

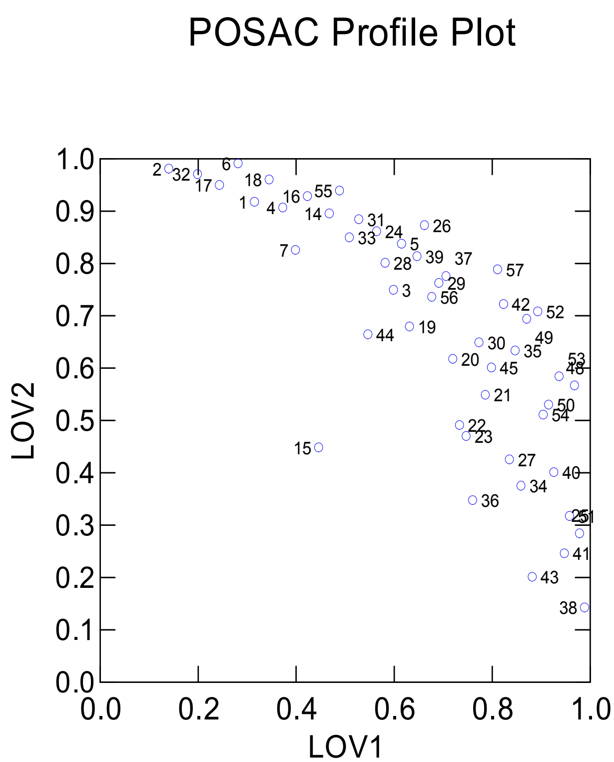


Figure 3.1.1. POSAC Profile Plot

Indicator	F-values LOV1	LOV2	Corr LOV1	LOV2
1	2.11	1.67	0.443	-0.105
2	2.68	1.04	0.476	-0.161
3	2.18	0.99	0.522	-0.159
4	1.23	1.5	0.159	0.053
5	3.81	0.86	0.561	-0.153
6	1.17	1.46	0.109	0.317
7	1.85	0.57	0.13	0.07
8	9.22	3.03	0.802	-0.48
9	0.66	0.61	-0.013	0.277
10	0.82	1.23	0.104	0.223
11	0.83	1.41	-0.004	0.35
12	0.92	0.71	0.105	-0.26
13	4.49	17.11	-0.636	0.815

Table 3.1.2. POSAC F-values and loadings

The POSAC method is another way of identifying the influential indicators. The POSAC method is aimed at reducing an N-dimensional coordinate data matrix by plotting it into a two-dimensional space. The two-dimensional coordinate is constructed by optimizing the proportion of order relations in the original data matrix.

The major outputs from the POSAC program include the two-dimensional POSAC plot and a coefficient, which describes the proportion of comparabilities that are correctly preserved. For the data matrix of 13 indicators, 78.5% of the partial order relations are correctly represented, which means 21.5% of the original information is lost. This is not a great number when considering the large number of dimensions that have been reduced. We can say that it is a good approximation for the 49x13 data matrix.

The POSAC plot is shown in Figure 12. The new coordinates are called LOV1 and LOV2, which are abbreviation for latent order variables. From the diagram, we see that most of the stream sites are clustered in the upper right corner as a zonal shape that is parallel to the diagonal of Y=1-X. Recall that the POSAC plot arranges the objects from the northeast corner to the southwest corner in an order so that the best objects come out first. The fact that the stream sites are scattered along the counter diagonal suggests that a fair amount of stream sites are incomparable with each other. Looking at the plot more closely, stream site 57 is found on the upper right of the plot with both high LOV1 and LOV2 values at about 0.8. Stream site 15 is located at the lower left side of the plot, with both LOV1 and LOV2 values of less than 0.5. Both of these stream sites are consistent with the other analyses where bridge 57 is identified as a maximal object of the Hasse diagram, is rated as an "excellent" bridge by its index ranking, and has the lowest poset ranking. Bridge 15, on the other hand, is found at the bottom level of the Hasse diagram, is rated as "poor" by its index ranking, and has the highest poset ranking.

As its name suggests, the latent order variables are considered meaningless to start with. To identify their underlying meanings, we explore the relationship of LOV1 and LOV2 with the 13 indicators. An ANOVA test is performed with LOV1 and LOV2 as dependent variables and the 13 indicators as factors. The F-values of the test are considered loadings of the indicators on the latent variables. The loadings tell us the strength of the impact on the latent order variable by that indicator. In addition, the correlations between the latent variables and the 13 indicators are computed, which provides us with information about the direction of the relationship between the latent order variable and the indicator. The F-values and the correlations can be found in Table 6.

Consistent with the previous analysis, two indicators (indicator 13, Channel Alignment, and indicator 8, Bank Soil Texture) highly dominate the latent variables. LOV1 was most heavily influenced by indicator 8, followed by indicator 13, with slight influences from indicators 3 and 5. LOV2 was overwhelmingly influenced by indicator 13, followed by a much smaller influence from indicator 8. The scatter plots of indicator 13 and 8 are shown in Figures 13 and 14 to further explore the correlations between the two influential indicators and the latent variables. In the scatter plots, the stream sites are classified as 'Excellent,' 'Good,' 'Fair,' or 'Poor' based on their score of the relevant indicator and are given a different symbol for better presentation. The plots show a strong positive relationship between LOV1 and indicator 8, and LOV2 and indicator 13, a negative relationship between LOV1 and indicator 13, and a weaker negative relationship between LOV2 and indicator 8. This explains the high negative correlation of -0.791 between LOV1 and LOV2 because the two most influential indicators are correlated in different orientations with the two latent order variables. Thus, we would expect that the two latent order variables would be negatively correlated. The concave shape of the graph is also explained by understanding these correlations.

Since the latent variables cannot be fully represented by a single indicator, we also consider weighted linear combinations (LC1 and LC2) of indicators 8 and 13 (denoted as V8 and V13) as a way to understand how well a combination of these two most influential indicators can approximate the two latent order variables. Let LC1 = 2*V8 + (-V13) and LC2 = 5.5*V13 + (-V8). The weights are selected based on the approximate F-test loadings, and we represent the negative of an indicator by the opposite value on the scale (for example if V8 had value 1, then (-V8) would take value 12 = 13-1). The weighted combinations are then divided into quartiles again, representing Excellent, Good, Fair, or Poor. The results are shown in Figures 15 and 16. The figures show a very clear division across the LOV1 spectrum for the four quartiles of LC1 and for LOV2 with respect to LC2. It is thus observed that the weighted linear combinations LC1 and LC2 are sound approximations for LOV1 and LOV2.

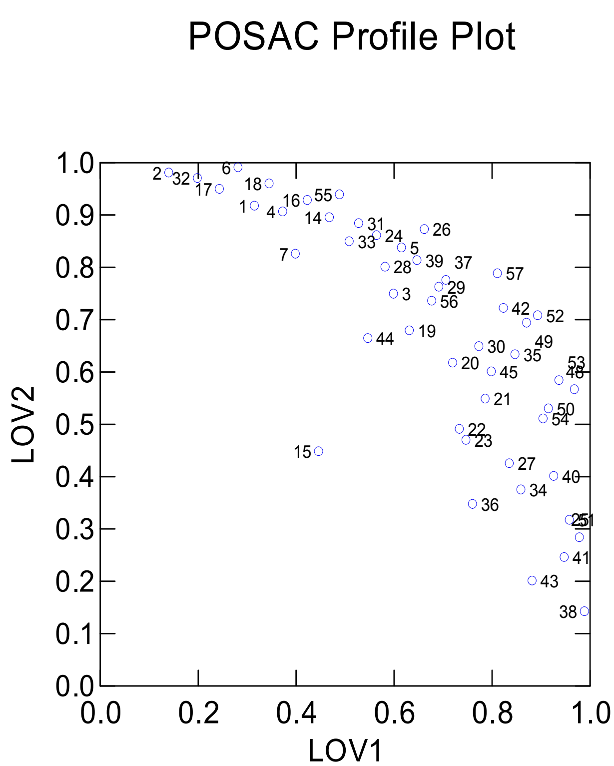


Figure 12 – POSAC plot for the 13-indicator data matrix

	F-values		Correlation	
Indicator	LOV1	LOV2	LOV1	LOV2
1	2.11	1.67	0.443	-0.105
2	2.68	1.04	0.476	-0.161
3	2.18	0.99	0.522	-0.159
4	1.23	1.5	0.159	0.053
5	3.81	0.86	0.561	-0.153
6	1.17	1.46	0.109	0.317
7	1.85	0.57	0.13	0.07
8	9.22	3.03	0.802	-0.48
9	0.66	0.61	-0.013	0.277
10	0.82	1.23	0.104	0.223
11	0.83	1.41	-0.004	0.35
12	0.92	0.71	0.105	-0.26
13	4.49	17.11	-0.636	0.815

Table 6 – The F-values and the correlations with the latent variables

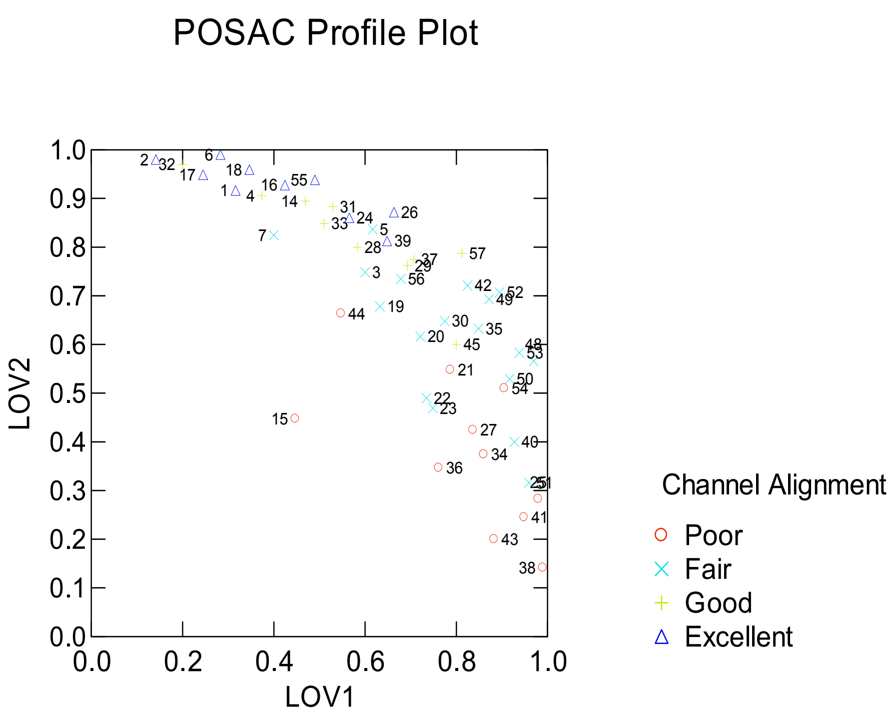


Figure 13 – Scatter plot of the indicator, Channel Alignment

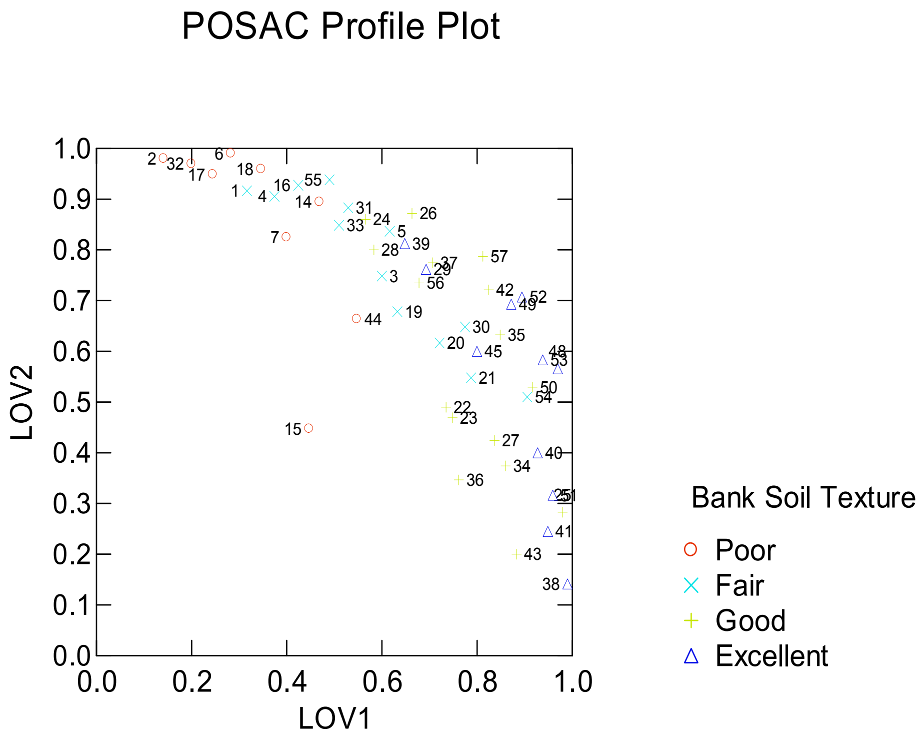


Figure 14 – Scatter plot of the indicator, Bank Soil Texture

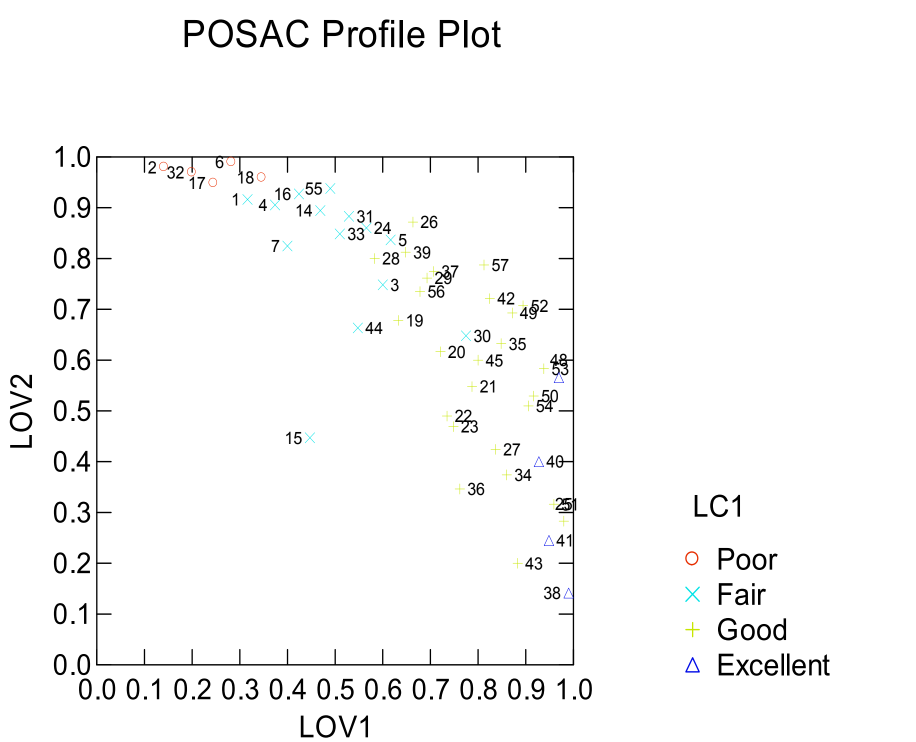


Figure 15 – Scatter plot of Linear Combination 1

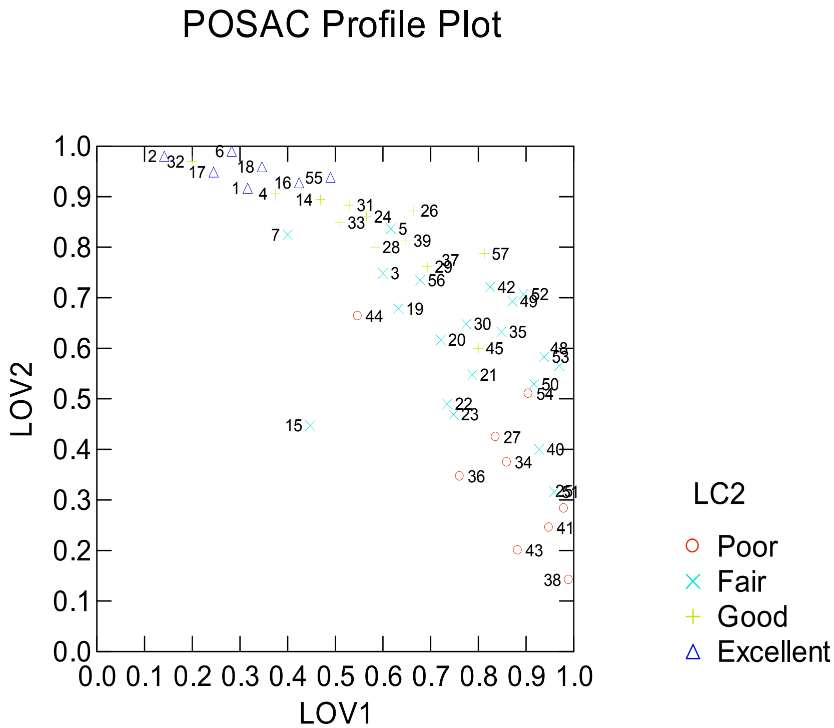
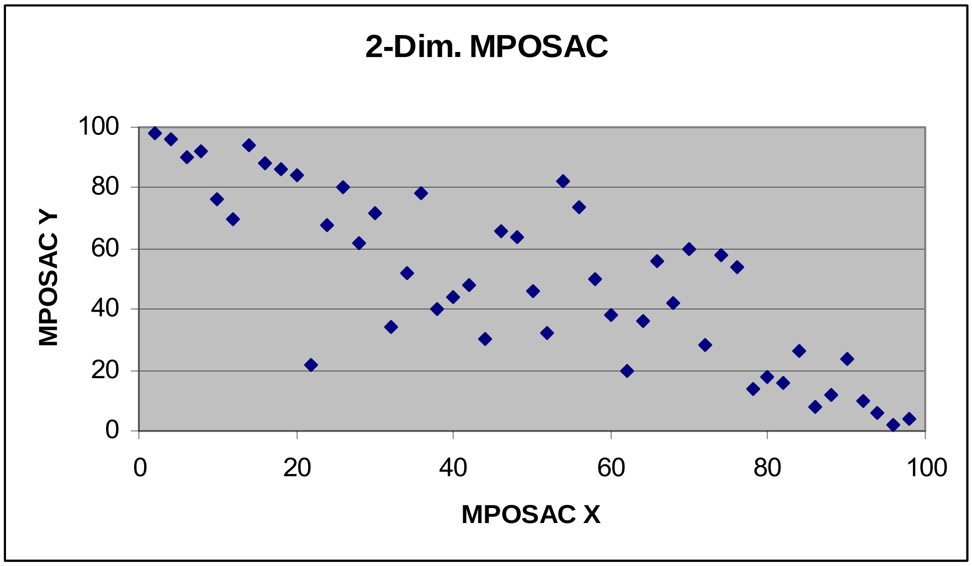


Figure 16 – Scatter plot of Linear Combination 2

Including stream stability assessment in the bridge inspection protocol would add to both the amount of time it would take inspectors to complete a site and the amount of stream channel behavior training time required for each inspector. In Table 1, each of the 13 indicators was assigned one of the three levels of expertise.

3.4 2-Dimensional MPOSAC

The graph below depicts the 2-dimensional MPOSAC performed on the 49 bridge points. The LOV values found from MPOSAC X and MPOSAC Y will be used in comparison to the LOV values found from POSAC.



3.5 One-Way ANOVA F-scores

The table below shows the corresponding F-scores from the one-way ANOVA performed between the 13 indicators and the latent-order-variables from both POSAC and MPOSAC. Their averages are also shown. Indicator 7 is clearly the most significant when comparing to the MPOSAC values, whereas for POSAC it is indicator 8.

	F-values				Average F-values	
Indicator	POSAC1	POSAC2	MPOSAC1	MPOSAC2	POSAC	MPOSAC
1	2.11	1.67	1.03	3.44	1.89	2.235
2	2.68	1.04	2.82	11	1.86	6.91
3	2.18	0.99	1.14	0.85	1.585	0.995
4	1.23	1.5	2.76	1.12	1.365	1.94
5	3.81	0.86	0.63	3.03	2.335	1.83
6	1.17	1.46	1.67	2.83	1.315	2.25
7	1.85	0.57	13.8	3.42	1.21	8.61
8	9.22	3.03	0.35	0.89	6.125	0.62
9	0.66	0.61	2.51	2.07	0.635	2.29
10	0.82	1.23	0.11	4.68	1.025	2.395
11	0.83	1.41	2.15	0.32	1.12	1.235
12	0.92	0.71	2.33	1.18	0.815	1.755
13	4.49	17.11	0.51	1.34	10.8	0.925

3.6 Spearman's Correlation Coefficient

The table below displays the results of comparing the 13 indicators to both X and Y coordinates of MPOSAC and POSAC. Indicator 13 has the highest correlation (0.838) with POSAC2. Indicator 8 follows with a correlation of 0.772 with POSAC1.

		Spearman's Corr. Coeff.		
Indicator	MPOSAC X	MPOASC Y	POSAC 1	POSAC 2
1	-0.064	0.422	0.467	-0.151
2	-0.358	0.699	0.448	-0.109
3	0.126	0.243	0.506	-0.149
4	0.221	0.109	0.214	0.043
5	-0.034	0.366	0.553	-0.252
6	0.143	0.236	0.164	0.196
7	0.748	-0.415	0.217	0.074
8	0.121	0.149	0.772	-0.498
9	0.254	0.143	0.059	0.254
10	-0.024	0.383	0.069	0.224
11	0.317	0.044	-0.011	0.304
12	0.252	0.153	0.107	0.234
13	0.104	0.089	-0.627	0.838

3.7 Concordance

Below is an example of a concordance table; specifically between Indicator 1 and MPOSAC2. We have used two measurements of concordance ratios for this analysis. Te first (Concordance I) is the count of the bridges that have corresponding values (the ones within the grey squares) divided by the total number of bridges. Concordance II is a more conservative measure in that it will only count half of those bridges that fall outside of direct correspondence (e.g. squares that are off center: (1,2), (2,1), (2,3), (3,2), (3,4),…).

Count of MPOSAC2	MPOSAC2			
I1	1	2	3	4
4			4	4
3	6	2	9	4
2	5	6		3
1	1	4		1

Concordance I	Concordance II
0.796	0.602

Indicator	Concordance 1				Concordance 2			
	POSAC		MPOSAC		POSAC		MPOSAC	
	LOV1	LOV2	LOV1	LOV2	LOV1	LOV2	LOV1	LOV2
1	0.837	0.735	0.694	0.796	0.592	0.531	0.429	0.602
2	0.878	0.776	0.571	0.837	0.704	0.510	0.378	0.622
3	0.898	0.796	0.694	0.776	0.663	0.551	0.449	0.449
4	0.755	0.694	0.816	0.694	0.500	0.510	0.582	0.500
5	0.776	0.653	0.633	0.755	0.571	0.429	0.480	0.592
6	0.796	0.878	0.755	0.735	0.551	0.612	0.541	0.500
7	0.816	0.776	0.939	0.633	0.571	0.571	0.704	0.429
8	0.918	0.592	0.735	0.735	0.694	0.357	0.500	0.520
9	0.531	0.653	0.633	0.796	0.327	0.449	0.490	0.531
10	0.694	0.633	0.694	0.857	0.429	0.429	0.500	0.684
11	0.714	0.816	0.837	0.694	0.500	0.602	0.622	0.510
12	0.714	0.755	0.755	0.653	0.561	0.510	0.571	0.500
13	0.429	0.898	0.612	0.714	0.276	0.633	0.418	0.469

Indicator	Average			
	Concordance 1		Concordance 2	
	POSAC	MPOSAC	POSAC	MPOSAC
1	0.786	0.745	0.561	0.515
2	0.827	0.704	0.607	0.500
3	0.847	0.735	0.607	0.449
4	0.724	0.755	0.505	0.541
5	0.714	0.694	0.500	0.536
6	0.837	0.745	0.582	0.520
7	0.796	0.786	0.571	0.566
8	0.755	0.735	0.526	0.510
9	0.592	0.714	0.388	0.510
10	0.663	0.776	0.429	0.592
11	0.765	0.765	0.551	0.566
12	0.735	0.704	0.536	0.536
13	0.663	0.663	0.454	0.444

The table above shows the average concordance values between the LOVs. For both Concordance I and Concordance II, indicator 3 had the highest values for POSAC. Whereas, for MPOSAC, indicator 7 had the highest value for Concordance I and indicator 10 had the highest value for Concordance II.

3.8 Rankings

The table below displays the rankings applied to the 49 bridges based on the above four methods and their average along with the ranking from the original poset. The correlations are shown for the comparison between POSAC and MPOSAC, all of which are highly significant showing much similarity. Also the correlation is shown between the each of the rankings based on each individual method with the original poset ranking, all of which are above 9.7 which shows that these other parameter-free methods did very well alongside that of the poset ranking.

Bridge	Differential Poset Rankings										POSET Rank
	F-values		Spearman's Corr. Coeff.		Concordance 1		Concordance 2		AVERAGE		
	POSAC	MPOSAC	MPOSAC	POSAC	POSAC	MPOSAC	POSAC	MPOSAC	POSAC	MPOSAC	
1	14	17	15	15	15	15	15	15	14	15	17
2	32.5	32	33	33	32.5	33	33	32.5	31.5	33	21
3	31	31	29	29	30	30	30	30	31.5	29	33
4	15	14	14	14	14	14	14	14	15	14	16
5	12	13	13	13	12	12	12	13	13	13	11
6	32.5	34	33	33	32.5	33	33	32.5	33.5	33	24
7	40	40	40	40	40	40	40	40	40	40	34
14	35	36	36.5	36.5	37	35	35	36	37	36.5	26
15	49	49	49	49	49	49	49	49	49	49	43
16	20	22	22	22	22	22	22	22	20.5	22	18
17	44	44	44	44	44	44	44	44	44	44	39
18	37	38	38	38	38	37.5	37.5	38	38	38	30
19	42	42	42.5	42.5	42	42	42	42	42	42.5	36
20	38	39	39	39	39	39	39	39	39	39	32
21	36	37	36.5	36.5	36	37.5	37.5	35	35.5	36.5	29
22	46	46	46	46	46	46	46	46	46	46	40
23	41	41	41	41	41	41	41	41	41	41	37
24	25	25	24.5	24.5	25	24.5	24.5	25	24	24.5	24
25	17	18	18	18	18	18	18	18	18	18	19
26	5	8	8	8	7	7	7	8	7	8	7
27	39	35	35	35	35	36	36	37	35.5	35	28
28	26	27	27	27	26	27	27	26	25.5	27	27
29	19	19	19	19	19	19	19	19	19	19	19
30	30	30	31	31	31	31	31	31	30	31	28
31	16	16	16	16	16	17	17	16	16	16	13
32	24	24	24.5	24.5	24	24.5	24.5	24	25.5	24.5	25
33	28.5	28.5	28	28	29	28	28	29	28	28	23
34	45	45	45	45	45	45	45	45	45	45	38
35	9	9	9	9	9	9	9	9	9	9	10
36	47	47	47	47	47	47	47	47	47	47	41
37	13	12	12	12	13	13	13	12	12	12	12
38	22	20	20	20	20	20	20	20	22	20	20
39	11	11	11	11	11	11	11	11	11	11	9
40	34	33	33	33	34	33	33	34	33.5	33	31
41	27	26	26	26	27	26	26	27	27	26	22
42	8	6	5	5	6	6	6	6	8	5	6
43	43	43	42.5	42.5	43	43	43	43	43	42.5	35
44	48	48	48	48	48	48	48	48	48	48	42
45	21	23	23	23	23	23	23	23	20.5	23	22
48	1	1	1	1	1	1	1	1	1	1	3
49	6	5	6	6	5	5	5	5	5	6	4
50	10	10	10	10	10	10	10	10	10	10	10
51	23	21	21	21	21	21	21	21	23	21	14
52	2	2	2	2	2	2	2	2	3	2	2
53	4	4	4	4	4	4	4	4	4	4	5
54	18	15	17	17	17	16	16	17	17	17	15
55	7	7	7	7	8	8	8	7	6	7	8
56	28.5	28.5	30	30	28	29	29	28	29	30	20
57	3	3	3	3	3	3	3	3	2	3	1
Cor. Between POSAC & MPOSAC	0.996045616		1		0.999209051		0.998953903		0.9974482		
Cor. w/ Orig. Poset	0.973	0.977	0.973	0.973	0.976	0.975	0.975	0.975	0.974	0.973	1.000

3.9 Rank Range Run

The consistency of the study sites, with respect to the indicators and indices, can reveal information about whether or not the stream stability assessment method might perform better for a particular stream type of physiographic region. Rank intervals represent all the possible ranks as determined from the poset prioritization procedure. The interval is from the minimum possible rank to the maximum possible rank. Figures 19 and 20 plot the rank interval as a function of the midpoint of the rank interval for the 13-indicator data seta and the 4-index data set. The rank intervals are smallest for the low midpoint ranks and for the high midpoint ranks. Figures 21 and 22 plot the rank interval as a function of the rank based on the total score value for the 13-indicator data set and the 4-index data set. The total score rank tends to be optimistic when compared to the midpoint of the rank interval for better ranked sites and pessimistic for poorer ranked sites. This is a trend similar to that shown in Patil and Taillie (2004) for the human-environment index (HEI).

The rank range relations for both the 13-indicator data set and the 4-index data set are shown in Figures 23 and 24, respectively. The plots are developed by converting each of the indicator scores into ranks and determining the minimum, maximum, and median ranks (Myers et al. 2005). The stream sites are sequenced according to the range of ranks for each particular site and ordered by increasing rank. The sites that are in the first run (1 through 13) of the rank range run (RRR) sequence are considered the most consistent with respect to the stream stability assessment for the 13-indicator data set (Figure 23). The sites with a RRR sequence equal to 1 through 12 are considered the most consistent with respect to the stream stability assessment for the 4-index data set (Figure 24).

The study sites that are the most consistently stable with respect to the indicators could serve as reference reaches in maintenance and restoration design efforts at stream-bridge intersections. Therefore, the study sites with both the lowest rank as indicated from the maximality level of the Hasse diagram and the highest consistency as indicated by the rank range run sequence are candidates for references reaches. Figures 25 and 26 show the relationship between the maximality level and rank range run sequence for the 13-indicator and 4-index data sets, respectively.

Investigation of the scores for each of the indicators and indices leads to some insight about the maintenance or restoration that is required at each study site. Knowledge of the indicators that result in the largest change in the range of the ranks can identify how a stream-bridge intersection should be maintained or restored. This knowledge can be used for prioritization and cost estimation in the decision-making process for mitigation efforts at stream-bridge intersections.

End-member elimination involves removing the maximum and minimum ranks separately and recalculating the range of ranks for a particular site (Myers et al. 2005). Plots of end-member elimination are included in Figures 27 and 28 for the 13-indicator data set and the 4-index data set, respectively. If removing the minimum rank results in the largest change in range of ranks, maintenance of stream-bridge intersection with respect to the indicator with the lowest rank becomes important. If this aspect of the stream-bridge intersection is not maintained, the overall rank and condition of the stream channel stability could deteriorate. Sites where this was the case are plotted as negative on the y-axis of Figures 27 and 28. If removing the maximum rank results in the largest change in the range of ranks, restoration of the stream-bridge intersection with respect to the indicator with the highest rank has the potential to greatly improve the stream channel stability. Sites where this is the case are plotted as positive on the y-axis in Figures 27 and 28. The 4-index data set (Figure 27) generally shows larger changes in rank ranges than the 13-indicator data (Figure 28). The maximum absolute change in rank range for the 4-index data set is approximately 40, while the maximum absolute change in the rank range for the 13-indicator data set is approximately 20. Focusing on the sites with the largest changes in rank range, it is apparent that the channel alignment indicator is responsible for many of the changes.

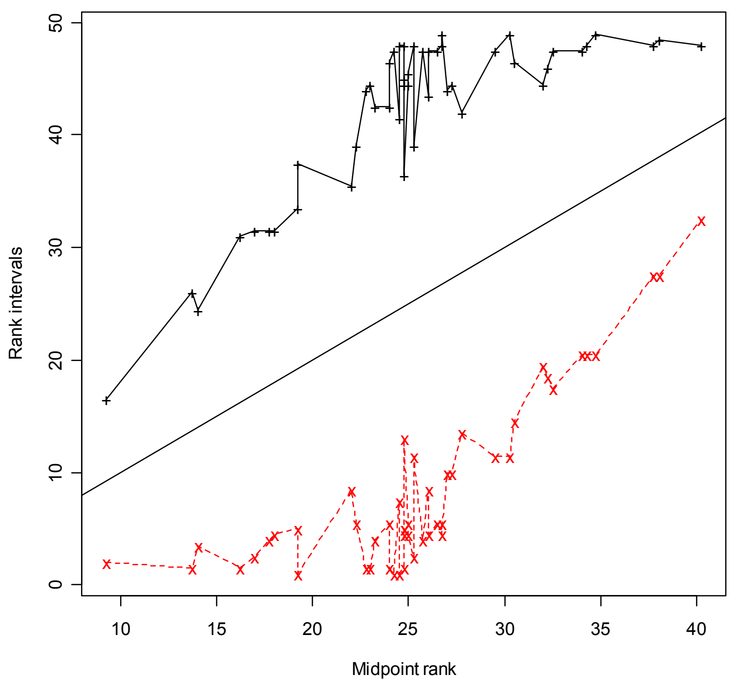


Figure 19 – Minimum, maximum, and midpoint rank vs. midpoint rank for the 13-indicator data set

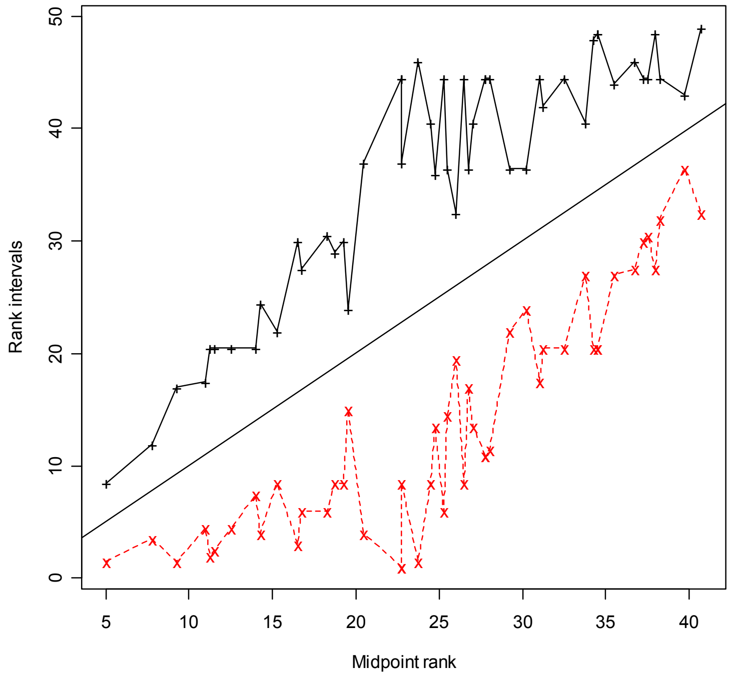


Figure 20 – Minimum, maximum, and midpoint rank vs. midpoint rank for the 4-index data set

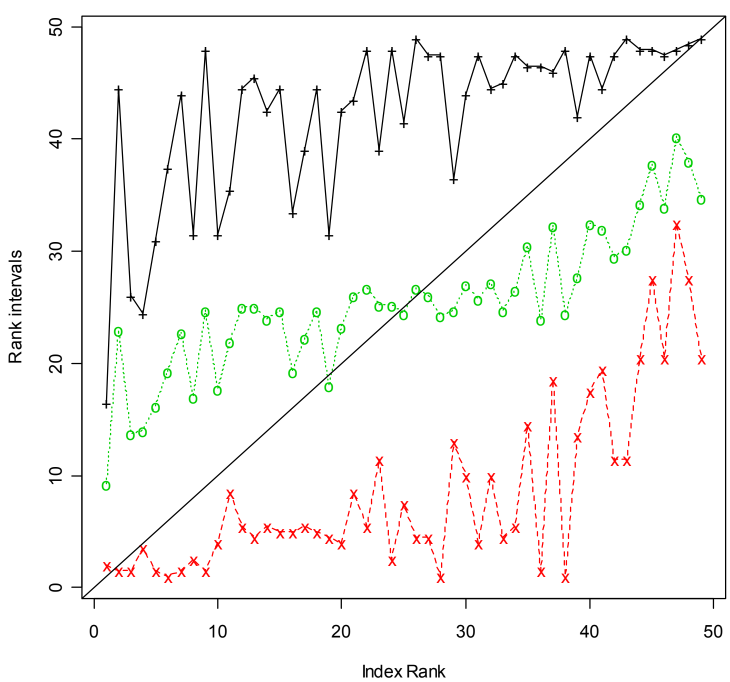


Figure 21 – Minimum, maximum, and midpoint rank compared to the rank based on total score for the 13-indicator data set.

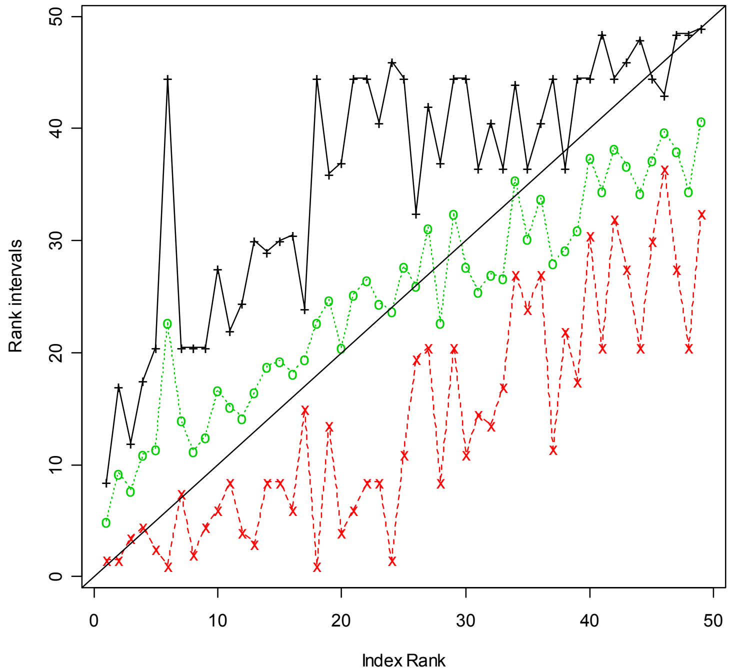


Figure 22 – Minimum, maximum, and midpoint rank compared to the rank based on total score for the 4-index data set

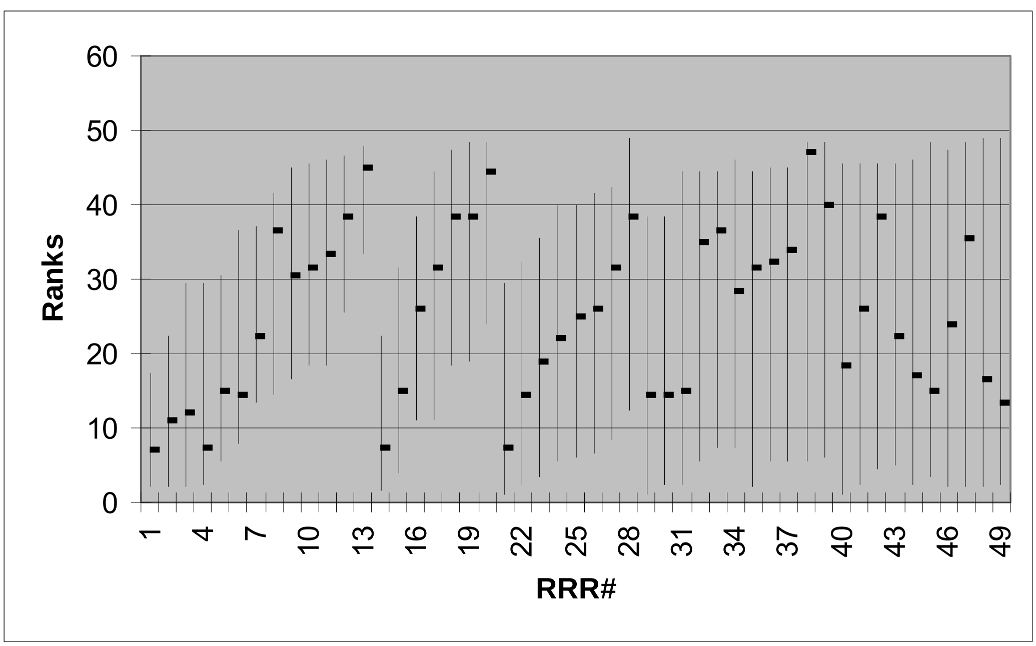


Figure 23 – Rank range run sequence for the 13-indicator data set. The bottom of each vertical line represents the minimum rank and the top of the line is the maximum rank for the indicators.

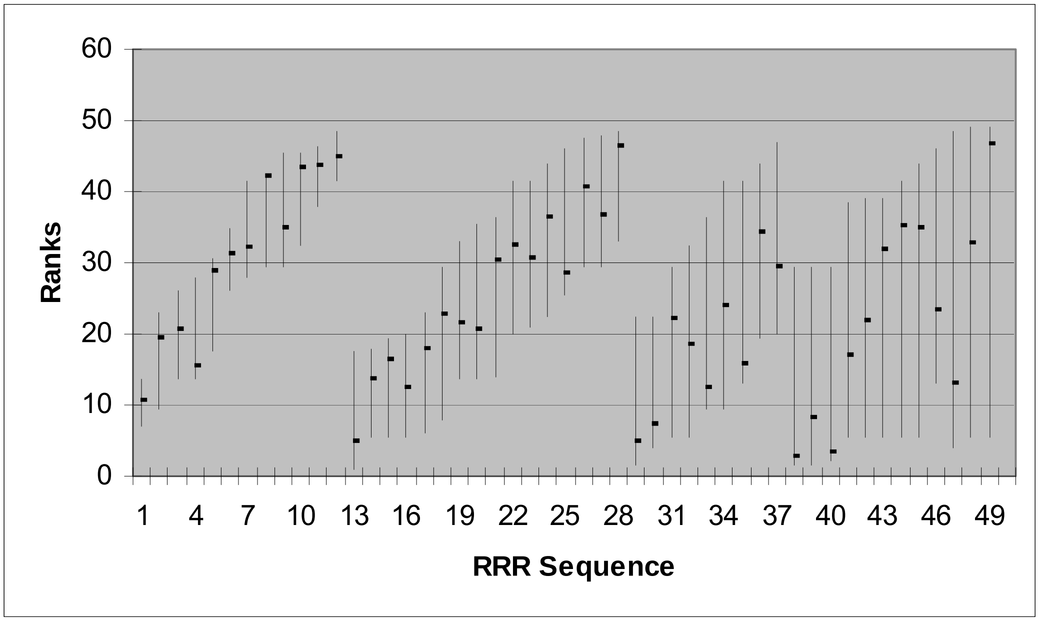


Figure 24 – Rank range run sequence for the 4-index data set. The bottom of each vertical line represents the minimum rank and the top of the line is the maximum rank for the indicators.

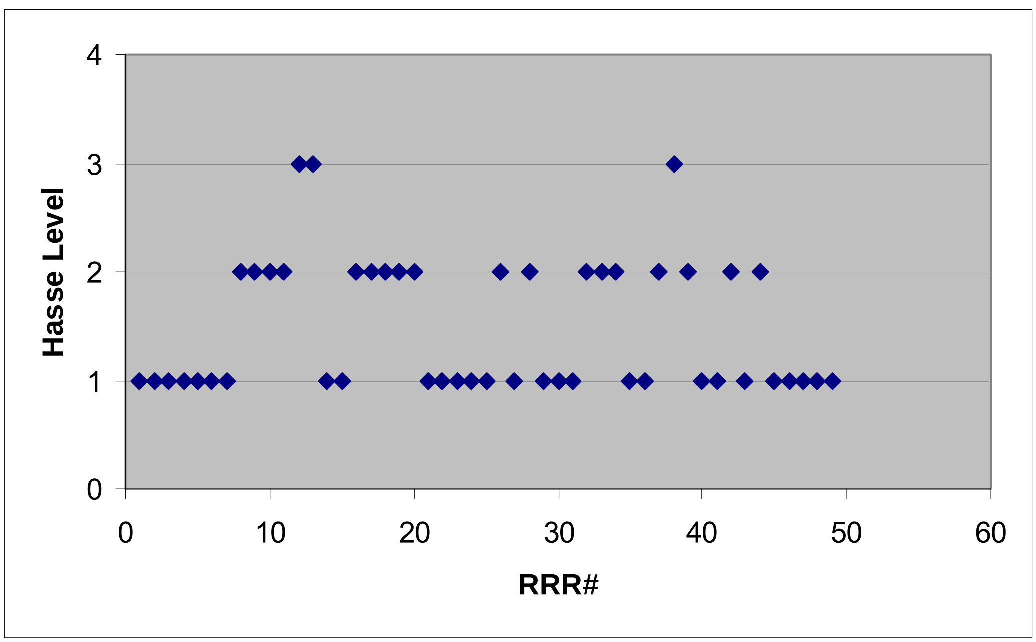


Figure 25 – Maximality level vs. consistency level for the 13-indicator data set

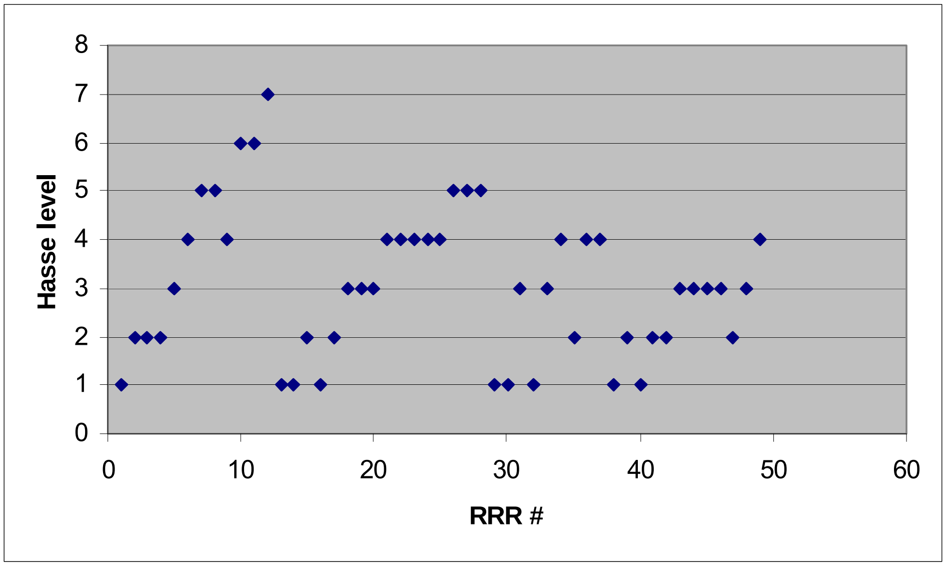


Figure 26 – Maximality level vs. consistency level for the 4-index data set

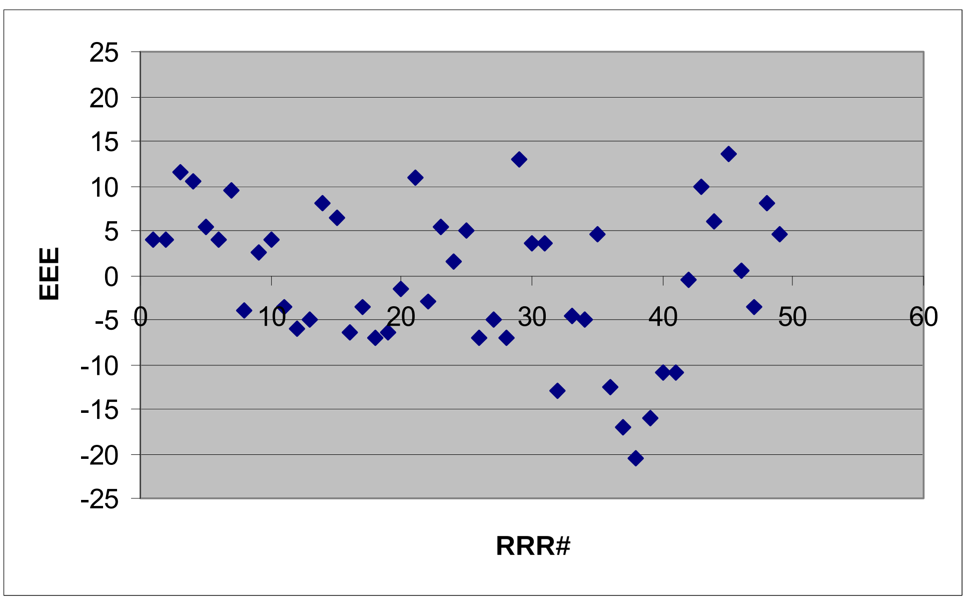


Figure 27 – End-member Elimination results for the 13-indicator data set

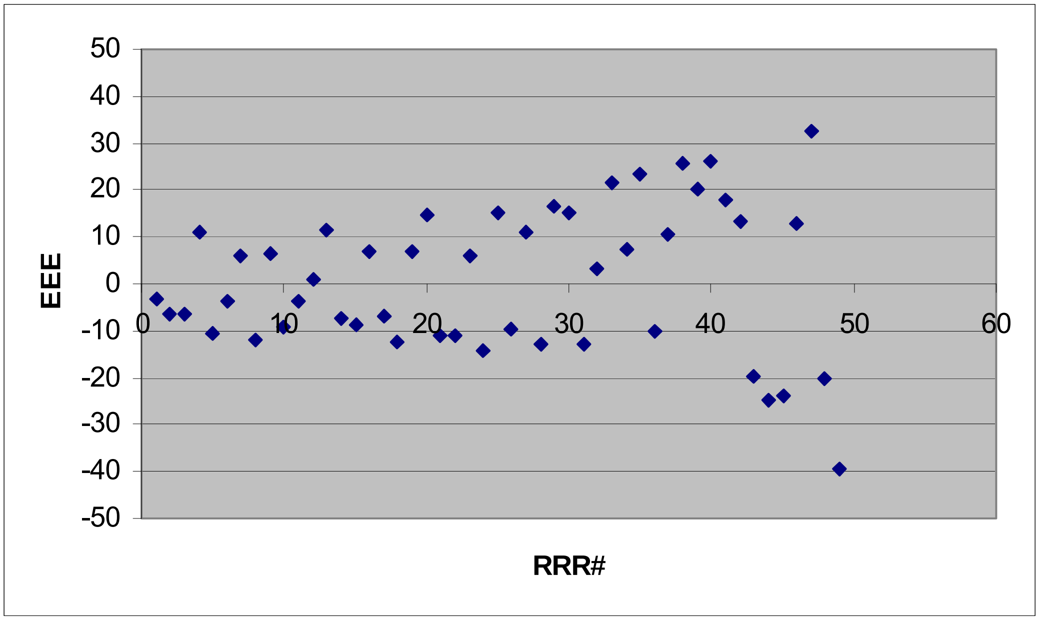


Figure 28 – End-member Elimination results for the 4-index data set.

3.10 METEOR

In the bridge data, we see a natural hierarchy determined by the principal investigator (Johnson 2005) of this project, where the 13 indicators have already been divided into four groups, with indicators 1-4 in group 1, indicators 5-7 in group 2, indicators 8-12 in group 3, and indicator 13 alone in group 4. Aggregation here would be done in two stages, aggregating the indicators from 13 indicators into the four groups, and then from the four groups into one final index.

We also applied the method of METEOR for the bridge data as well. We can apply METEOR for two different sets of differential weights determined by the scientists 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. 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 can be used as a metric to understand how well the aggregation has performed.

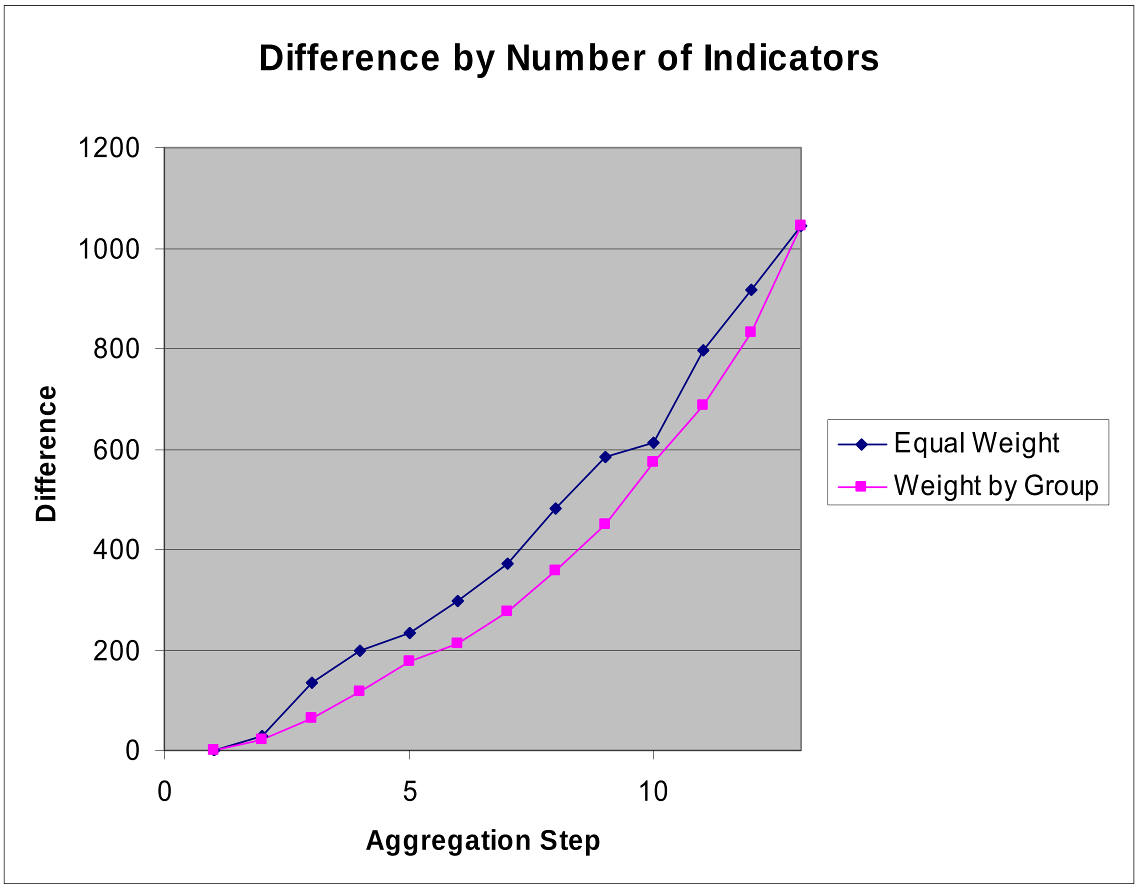
For the equal indicator weighting scheme, we applied METEOR on all 13 indicators, as opposed to treating them as a hierarchy. At each step, we aggregated the two least correlated indicators by using the relative weights for the two indicators. We also applied METEOR on all 13 indicators for the equal group weighting scheme as well, in the same manner. The number of comparabilities at each step can be seen in the table below.

13 indicator equal weights(1/13)	
Number of indicators	Comparabilities
13	133
12	161
11	266
10	333
9	367
8	430
7	503
6	614
5	717
4	746
3	928
2	1050
1	1176

13 indicator weighted by group	
Number of indicators	Comp
13	133
12	156
11	196
10	250
9	310
8	344
7	410
6	492
5	583
4	708
3	818
2	965
1	1176

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.

We see from the graph below 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.



It is noted that the indicator 13 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. Since indicator 13 gets an overall weight of 0.25, it is likely to have a strong influence or dominate any aggregated indicator, 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, which results in fewer comparabities. Thus it is plausible that the equal group weights index would be have fewer comparabilities at each aggregation step.

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) is 708 as opposed to 460 in hierarchical aggregation. The method of hierarchical aggregation does not increase comparabilities as fast as the ordinary aggregation.

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.

4 Conclusions

For the stream channel case study, assigning stream condition based on total score is satisfactory. Also, grouping all 13 indicators into four indices does not have a large effect on the order relations of the stream sites. The sites show reasonable agreement with the Hasse diagram and the poset prioritization results for both data matrices. Based on the total score, the threshold values that assign the stream sites into four categories are supported by the Hasse levels. Channel alignment is one of the most influential indicators for overall stream stability, as expected. However, the watershed and floodplain activity indicator also was expected to have greater overall influence on the stream stability condition. When considering the expense of collecting the indicator data, at least ten indicator scores (level 1 and level 2) should be collected. An analysis of consistency of study sites can reveal good candidates for reference reaches, which are often used as optimum examples in stream restoration practices. The analysis also provides information about which maintenance or mitigation practices could be most beneficial to a stream-bridge intersection. This knowledge can be used for cost estimation and prioritization of maintenance and mitigation efforts.

Working through a series of analyses, we are able to answer the research questions asked in the above objectives section for the stream channel case study. We have learned that assigning stream condition based on total score is satisfactory. Also, grouping all 13 indicators into four indices does not have a large effect on the order relations of the stream sites. The sites show reasonable agreement with the Hasse diagram and the poset prioritization results for both data matrices. Based on the total score, the threshold values that assign the stream sites into four categories are supported by the Hasse levels. Channel alignment is one of the most influential indicators for overall stream stability, as expected. However, the watershed and floodplain activity indicator also was expected to have greater overall influence on the stream stability condition. When considering the expense of collecting the indicator data, at least ten indicator scores (level 1 and level 2) should be collected. An analysis of consistency of study sites can reveal good candidates for reference reaches, which are often used as optimum examples in stream restoration practices. The analysis also provides information about which maintenance or mitigation practices could be most beneficial to a stream-bridge intersection. This knowledge can be used for cost estimation and prioritization of maintenance and mitigation efforts.

Several multi-criterion assessment methods have been evaluated and demonstrated thoroughly through this empirical study. These methods include the Hasse diagram technique, poset (partially ordered set) prioritization, POSAC (partially ordered scalogram analysis with coordinates), and RRR (rank range run) method. All of the methods considered here are data-based, in other words, parameter-free. No extra knowledge beyond the data matrix is necessary.

The Hasse diagram technique has been demonstrated as a fundamental, yet powerful, tool in understanding the intrinsic structure of a data set. It not only produces the most general rankings of objects, but also provides additional information such as maximality level, connectivity of components, and influence of indicators. These basic and essential features make the Hasse diagram a useful tool alone and with other methods. This is not to mention that the Hasse diagram is the foundation of poset prioritization as well as the fact that the maximality level is an important element in the rank range run analysis.

Based on the Hasse diagram technique, poset prioritization works further on linear extensions. The method induces randomness into the picture by assuming a probability distribution and a rank frequency distribution. The method is so unique that an exclusive ranking is provided without any arbitrary judgment involved.

The POSAC method can be seen as a combination of order theory and a data reduction technique. Similar to other variable reduction methods, such as principle component analysis, POSAC reproduces the data set in less dimensional space and provides a good visualization of the objects. Moreover, the POSAC method takes into account the order relations among objects, which makes this method more powerful when object prioritization is of concern.

The rank range run method is a type of exploratory data analysis. Serving as a supplement of poset prioritization, this method obtains information directly from the data matrix. It is particularly informative when combining with other ranking methods such as the Hasse diagram technique. The method examines the consistency of the data matrix through a set of statistics such as rank range and largest rank gap. The method is also able to reveal the information graphically.

These methods have constructed comprehensive evaluations to the data sets. The data matrices are evaluated from both the perspective of objects and indicators. On one hand, the partial orders of a data set can be visualized by Hasse diagram and its total order can be obtained by implementing poset prioritization. On the other hand, the influences of indicators can be evaluated by conducting the sensitivity (W-matrix) and consistency (RRR) analyses.

As been mentioned earlier, all of the methods used in this analysis are data driven, in other words, parameter-free. Efforts have been made to avoid arbitrary judgments. However, there is a dilemma in that there are often times when the participation of stakeholders and other participants is needed in the evaluation process. Patil and Taillie (2004) discuss the possibility of generalizing to a weighted CRF operator by giving linear extensions differential weights, either on mathematical grounds (e.g., number of jumps) or empirical grounds (e.g., indicator concordance), in the procedure of poset prioritization. Recently, a new decision support system, METEOR (Method of evaluation by order theory), was introduced by Simon et al. (2005). METEOR allows and encourages stakeholders to participate in a step-by-step indicator aggregation process. Although bias can be introduced during the process, the aggregation enables the method to provide an exclusive ranking of objects with participation and transparency. In the meantime, some mathematical and technical questions are open to be resolved. Some concerns have been raised by Brüggemann and Simon (2006).

As a major approach in detecting influential indicators, the POSAC method shows its advantage when compared to other dimension reduction methods. In the current analysis, we use POSAC to obtain two latent order variables and the loading of their indicators to get an approximate model. As a more powerful program, MPOSAC (Hebrew University of Jerusalem), which has recently become available, makes it possible to apply a multidimensional POSAC to the data matrices. This capability would allow us to obtain more than two latent order variables and also use them to get relative weights of the indicators.