Representing Uncertainty in Ranking by Single or Multiple Indicators

[1] **Andrews, Felix**

[1] Integrated Catchment Assessment and Management Centre (iCAM),
School of Resources, Environment and Society; and Mathematical Sciences Institute
The Australian National University. E-Mail: felix.andrews@anu.edu.au

Keywords: *Ranking; Prioritisation; Indicators; Uncertainty; Partial Order*

EXTENDED ABSTRACT

This paper introduces a method for representing uncertainty in ranking by single or multiple indicators. The method can potentially integrate parametric and structural uncertainty of model outputs. It requires estimating the range of conditions over which a ranking of items should be robust. The ranking is then subjected to perturbation tests, and the results displayed graphically.

Ranking a set of measurements, or ranking a set of model outputs, is a generic task for decision support. In the case of multiple indicators, a composite index is often defined. However, as Patil and Taillie (2004) point out, “every such composite involves judgements (often arbitrary or controversial) about tradeoffs or substitutability among indicators.” These concerns are addressed by the concept of partial order.

Partially-ordered sets can be used to identify items that are objectively comparable, in the sense that all indicators favour one item over the other. If there is a tradeoff between two items (i.e. their indicators are inconsistent) then they are not inherently comparable. The concept of partial order been used recently to rank multiple indicators. For example, Hollert *et al.* (2002) used it to rank ecotoxicological contamination of small streams according to different chemical and biological tests.

This paper extends the use of partial order, from representating *ambiguity*, to also representing *uncertainty*. Outputs from perturbed models can be treated as additional indicators, alongside outputs from alternative model structures. Another possibility is the use of data resampling (jackknife or bootstrap tests) to generate perturbed indicators.

An example of a robust partial order is shown on Figure 1, where sites in a river system are ranked by their median flow magnitude. For this analysis, river flow time-series were used from 9 sites in a common 18-year period. For the ranking to be robust, it should not change when a single year is included or excluded from the common period. Additionally, it should be equivalent for any percentiles between 40% and 60% (not just the exact median, 50%). The partial order on Figure 1 shows the comparisons that are robust under these conditions – in this case, it is almost a complete order. There are only three sites with ambiguous ranks.

This paper also gives a more complex case study, combining multiple indicators.

Representing uncertainty in ranking should provide an improved basis for decision-making. The lack of agreement between indicators, or their lack of robustness, lead naturally to reconsidering and revising the modelling process.

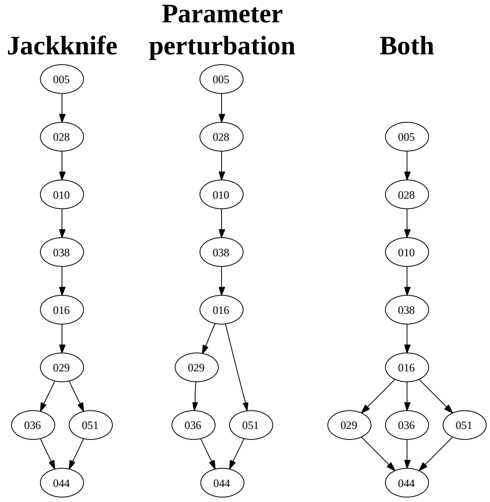

Figure 1. Partial order of sites by median flow. Magnitude of median flow increases upwards. The graph on the left shows ambiguity introduced by a jackknife procedure, where each single year of record was excluded in turn. The middle graph shows the ambiguity introduced by varying the percentile parameter from 50% (median) to between 40% and 60%. The final graph shows the case where parameter perturbations were applied to each jackknife replicate.

1 INTRODUCTION

Ranking a set of measurements, or ranking a set of model outputs, is a generic task for decision support. This is trivial when only one indicator is considered; in the case of multiple indicators, a composite index is often defined. The definition of such an index is often subjective, chosen as much for convenience as for conceptual reasons.

Partially-ordered sets have been used recently to rank multiple indicators objectively (e.g. Patil and Taillie, 2004; Hollert *et al.*, 2002; Weigert and Steinberg, 2002). They allow some conclusions to be drawn about the order of items, but refrain from comparing items where their indicators are inconsistent.

This paper introduces a further use for partial order: to represent uncertainty in ranking. When ranking model outputs, there is often considerable uncertainty in the model parameters, inputs and assumptions. These can be *perturbed* (varied within a plausible range), testing the robustness of comparing items. Any robust comparisons can be displayed as a partial order.

Representing uncertainty in ranking should provide an improved basis for decision-making.

2 RANKING AND PRIORITISATION

One of the most fundamental uses of decision support systems is in ranking a collection of items. Such a ranking or prioritisation might apply to:

- alternative management scenarios, with respect to their predicted outcomes. *For example, which dam management scenario gives the best combination of agricultural profit and floodplain ecosystem health?*
- a set of locations, with respect to how severe some problem is, or how great an opportunity is presented. *For example, where is dryland salinity predicted to be most severe?*
- a set of types, species, etc. *For example, which weed species are most likely to invade some region?*

Management decisions typically must take multiple criteria into account. Many decision support systems generate a unique ranking by combining different indicators into a composite index (e.g. Brans and Vincke, 1985; Young *et al.*, 2003). However, as Patil and Taillie (2004) point out, "every such composite involves judgements (often arbitrary or controversial) about tradeoffs or substitutability among indicators." These concerns are addressed by the concept of partial order.

2.1 Partial Order in Decision Support

A different approach to multi-criteria ranking uses the theory of partially ordered sets (*posets*). This works with items that are objectively comparable, in the sense that all indicators favour one item over the other. If there is a tradeoff between two items (i.e. their indicators are inconsistent) then they are not inherently comparable.

Partial orders can be visualised with Hasse diagrams. These show the *cover relationships* (adjacent comparable relationships) between items on a graph. To construct a Hasse diagram, the following procedure is sufficient:

1. Draw lines from each item to those others that it is objectively greater than (where all indicators agree).
2. Position the items so that all lines point downwards: the greatest items are then at the top of the diagram. This implies that the diagram has a certain number of *levels*.
3. Remove those lines implied by transitivity, that is, if $A > B$ and $B > C$ then there is no need to also show that $A > C$. This step simplifies the diagram.

An example of a Hasse diagram will be shown in the next section.

Hasse diagrams must be interpreted with care. Only those items that are directly connected, or connected by a chain of cover relationships, are comparable. Incomparable items cannot be positioned uniquely on the diagram, so there is not necessarily a simple scale from "good" at the top to "bad" at the bottom. Nonetheless, the information in a partial order is often useful for decision making.

Partial order has been applied to several multi-criteria problems. For example, Hollert *et al.* (2002) used it to rank ecotoxicological contamination of small streams according to different chemical and biological tests. This helps to prioritise sites for management. Patil and Taillie (2004) applied the theory to UNEP environmental quality indicators (land, air and water) for each of 106 countries. Their approach avoids unsupportable assumptions about the relative importance of component indicators. Weigert and Steinberg (2002) demonstrated the use of partial order in urban water resource management scenarios.

Where there are many different indicators, or the indicators disagree, many items may be incomparable. In these cases, Weigert and Steinberg (2002) suggest that "they should be subject to public discussions and subsequent political decisions." These authors also suggest selectively combining indicators or taking subsets of indicators to resolve ambiguities.

Patil and Taillie (2004) discuss the complete orders that are consistent with a given partial order (*linear extensions*). They identify the range of possible ranks for each item. Furthermore, they compute the rank probability distribution of each item, and work out the complete order that is most probable, in some sense. Some authors also use the average rank of each item to generate a complete order. My personal opinion is that, if a complete order is required, indicator *values* should be combined (such as a traditional weighted sum), rather than extending the partial order.

2.2 Partial Order to Represent Uncertainty

A major challenge in environmental management (as well as in other areas) is to integrate uncertainty into decision support (Refsgaard *et al.*, 2004). When predictive models are used, sensitivity and uncertainty assessment can be used to evaluate whether decisions based on them are adequately robust to minor changes in parameters, inputs or assumptions (Saltelli and Scott, 1997). The outputs of interest vary depending on the application, but Norton *et al.* (2003) list the rank position of items as a common basis for decision-making.

When the rank order of model outputs is of interest, the results of sensitivity assessment can be represented as a partial order. Outputs produced from perturbed models (e.g. varying one or several parameters through a plausible range) can be treated as additional indicators (equally valid with the base case), by analogy with the approach described in Section 2.1. Accordingly, Hasse diagrams can be used to represent uncertainty in rankings.

This approach differs from a naive Monte-Carlo assessment, where a range of inputs is translated into a range of outputs. In that case, if two outputs have an overlapping range they might be thought to be indistinguishable. However, even though the values of both vary substantially, one might be always greater than the other (i.e. they vary in a correlated fashion). Using partial order reveals this.

The proposed use of partial order to represent uncertainty can apply to ranking by a single indicator, or to ranking by multiple indicators. In the latter case, ambiguity and uncertainty are combined. If the multiple indicators are alternative model formulations, the ambiguity may represent (a part of) structural uncertainty.

3 CASE STUDY: RANKING RIVERS BY FLOW INDICES

In this section the proposed method will be demonstrated with a simple case study. Considering statistics derived from river flow time series as simple models, they are used to rank a set of rivers.

3.1 The Namoi Dataset

Rivers were selected from the Namoi River basin, within the Murray-Darling Basin, in northern NSW, Australia. The region was chosen partly because it has a relatively large number of stream gauging stations with long records.

The focus here is on natural flow features, so only unregulated rivers were considered. Furthermore, rivers subject to substantial groundwater extraction were excluded: in this case, the Mooki and Cox's Creek systems were excluded.

River flow data was from a common period at all sites, in order to compare them independently of climatic differences. The period was chosen so that as many sites as possible had adequate data: at least 15 available years from an 18-year period (this was an arbitrary criterion). The chosen period covered the 18 water years 1969-1986 (inclusive). Water years were defined as beginning 1 October in the named calendar year, and were "available" if fewer than 10% of their daily values were missing. Data was obtained from Pinneena 8 (DIPNR, 2004).

Nine sites (stream gauging stations) were selected, the details of which are shown in Table 1. Additionally, Figure 2 gives a visual display of the available data years.

Table 1. Stream gauging stations selected from the Namoi basin. Sites were selected if they (a) were unregulated and free from large-scale groundwater extraction; and (b) had no more than 3 missing data years in the common 18-year period covering water years 1969-1986 (inclusive).

station ID	river name	station name	catchment (km^2)	period of record	latitude	longitude
419005	Namoi	North Cuerindi	2538	1915-	-30.68	150.78
419010	Macdonald	Woolbrook	844	1927-	-30.97	151.35
419016	Cockburn	Mulla Crossing	900	1936-	-31.06	151.13
419028	Macdonald	Retreat	1760	1965-1987	-30.63	151.11
419029	Halls Ck	Ukolan	376	1965-	-30.71	150.83
419036	Duncans Ck	Woolomin	93	1965-1986	-31.32	151.16
419038	Macdonald	Cobrabald	358	1965-1987	-31.19	151.45
419044	Maules Ck	Damsite	45	1968-1992	-30.53	150.3
419051	Maules Ck	Avoca East	667	1972-	-30.5	150.08

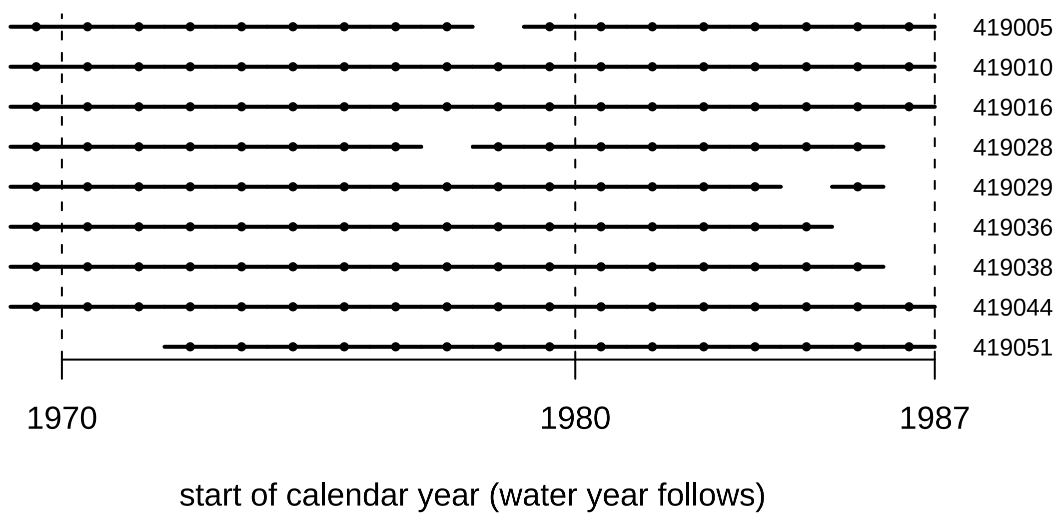

Figure 2. Available flow data in the common 18-year period covering water years 1969-1986 (inclusive). Years were defined as being available if fewer than 10% of their daily values were missing. The point following a marked year (such as "1980") represents the water year beginning 1 October in that calendar year.

3.2 Ranking by Median Daily Flow

The first case simply involves ranking the rivers by their typical size: the median flow magnitude. That is, the flow exceeded 50% of the time. It is trivial to compute the median for each site and to rank these. However, the results might depend on specific years of the flow record. For the ranking to be robust, it should not change when a single year is included or excluded from the common period. This can be tested with a jackknife procedure (Efron and Gong, 1983): each year of the common period was excluded (one at a time), and the statistics recomputed. In some of these *replicates*, two sites switched their rank (036 and 051), and therefore their rank is ambiguous. The partial order produced by this method is shown at the left of Figure 3.

Note that the first-order jackknife (removing one at a time) can be extended to higher orders. For instance, the second-order jackknife removes every combination of two years at a time (with 18 years, 153 replicates). This provides a more demanding test of robustness.

To test parameter sensitivity, the percentile parameter (50% for the median) can be varied. Subjectively, I would like the ranking to be valid not just for the exact median; anywhere between the 40% and 60% percentiles should be equivalent. Accordingly, the statistics were recomputed at several percentiles in that range. The partial order produced by this method is shown at the middle of Figure 3.

Combining the two types of perturbation, several percentiles between 40% and 60% were computed for each jackknife replicate. The partial order produced by this method is shown at the right of Figure 3. It is interesting to note that this gives extra information beyond either of the single tests: that site 029 is incomparable with site 036.

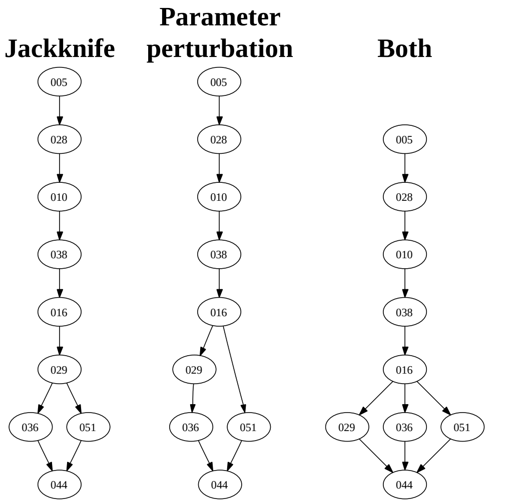

Figure 3. Partial order of sites by median flow. Magnitude of median flow increases upwards. The graph on the left shows ambiguity introduced by a jackknife procedure, where each single year of record was excluded in turn. The middle graph shows the ambiguity introduced by varying the percentile parameter from 50% (median) to between 40% and 60%. The final graph shows the case where parameter perturbations were applied to each jackknife replicate.

3.3 Ranking by "Harshness" Indices

The next case is more complex, combining three indicators and representing uncertainty in each. The three statistics were selected as potential indicators of "harshness" (referring to stream habitat conditions) from the list given by Fritz and Dodds (2005). The statistics were:

- Annual coefficient of variation (CV): standard deviation divided by mean of total flows in each available water year.
- Frequency of zero flow conditions (ZERO): the average number of days per year with flow less than a low threshold (by default, 0.1 ML/day).
- Flood-flow index (FFI): the proportion of total flows that are not base-flow, according to a base-flow separation filter. This has the reverse order to a base-flow index. Baseflow separation used the minimum filter (Croke *et al.*, 2002), with a default window size of 5 days.

Each indicator was subjected to perturbations as described in the previous section. These comprised a first-order jackknife resampling of the flow data, and a simple parameter perturbation for each of these replicates. The parameter perturbations were as follows:

- for annual coefficient of variation: the year boundary was varied from 1 October by +/- 37 days (10% of days in the year);
- for frequency of zero flow conditions: the threshold was varied from 0.1 ML/day to be 0.01 and 1 ML/day;
- for flood-flow index: the window size of the base-flow filter was varied from 5 days to be 3 and 7 days.

The partial orders produced by these perturbations are shown on Figures 4, 5 and 6, referring to CV, ZERO and FFI respectively.

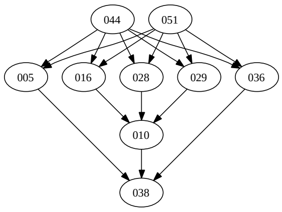

Figure 4. Partial order of sites by annual coefficient of variation. The ambiguity represents sensitivity to the following perturbations. A jackknife procedure was applied, excluding each year of record in turn. For each of these replicates, the year boundary was varied from 1 October by +/- 37 days.

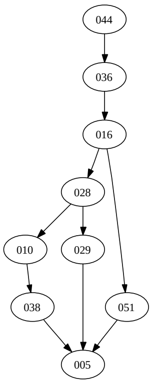

Figure 5. Partial order of sites by frequency of zero flow conditions. The ambiguity represents sensitivity to the following perturbations. A jackknife procedure was applied, excluding each year of record in turn. For each of these replicates, the threshold for "zero" flow was varied from 0.1 ML/day to be 0.01 and 1 ML/day.

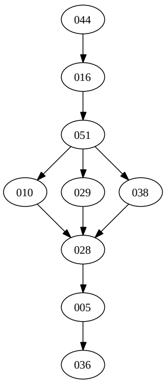

Figure 6. Partial order of sites by flood-flow index. The ambiguity represents sensitivity to the following perturbations. A jackknife procedure was applied, excluding each year of record in turn. For each of these replicates, the window size of the base-flow filter was varied from 5 days to be 3 and 7 days.

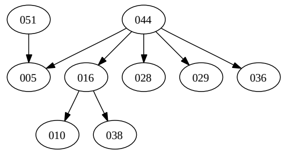

Figure 7. Partial order of sites by three potential indicators of "harshness". Harshness increases upwards. The three component indicators are those shown on Figures 4, 5 and 6, and include the uncertainty of each according to perturbations of parameters and input data.

The partial order for each of these indicators is based on a matrix of indicator values, with sites against replicates. To combine them into an overall partial order is as simple as joining the matrices, so that all replicates are considered together. The corresponding Hasse diagram is shown on Figure 7.

Figure 7 shows robust statements about the relative "harshness" of stream habitat at each site. Based on three possible indicators of harshness (these might be thought of as alternative model structures / conceptualisations) and specified tests of robustness (jackknife resampling and parameter perturbation), several conclusions can be drawn. Sites 044 and 051 are candidates for the most harsh stream. By looking at the component indicators' Hasse diagrams (Figures 4, 5 and 6), these two sites have the equally highest CV (incomparable due to uncertainty), but 044 is clearly more harsh in the other cases. On the other hand, there are six incomparable sites that are the least harsh: 005, 010, 038, 028, 029 and 036. Much of this ambiguity is introduced by CV (see Figure 4), so this indicator might be revised to allow stronger conclusions.

4 CONCLUSIONS

This paper proposed a method for representing uncertainty in ranking by single or multiple indicators. This method can potentially integrate parametric, input and structural uncertainty of model outputs. Of course, the choice of perturbation tests is somewhat subjective, and requires estimating the range of conditions over which the ranking should be robust. The lack of agreement between indicators, or their lack of robustness, lead naturally to reconsidering and revising the modelling process.

5 ACKNOWLEDGEMENTS

The analyses and visual displays in this paper were produced with open-source software tools: the *R* system for statistical computing (Ihaka and

Gentleman, 1996; http://www.R-project.org/) and Graphviz (http://www.graphviz.org/).

6 REFERENCES

Brans, J.P. and P.H. Vincke (1985), A Preference Ranking Organisation Method. (The PROMETHEE Method for Multiple Criteria Decision Making), *Management Science*, **31(6)**, 647-656.

Croke, B.F.W, A.B. Smith and A.J. Jakeman (2002), A One-Parameter Groundwater Discharge Model Linked to the IHACRES Rainfall-Runoff Model, In: Rizzoli, A.E. and A.J. Jakeman (Eds.), *Integrated Assessment and Decision Support, Proceedings of the 1st Biennial Meeting of the International Environmental Modelling and Software Society*, Volume 1, 428-433.

DIPNR (2004), Pinneena version 8 hydrological database. NSW Department of Infrastructure, Planning and Natural Resources.

Efron, B. and G. Gong (1983), A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation, *The American Statistician*, **37(1)**, 36-48.

Fritz, K.M. and W.K. Dodds (2005), Harshness: characterisation of intermittent stream habitat over space and time, *Marine and Freshwater Research*, **56**, 13-23.

Hollert, H., S. Heise, S. Pudenz, R. Brggemann, W. Ahlf and T. Braunbeck (2002), Application of a sediment quality triad and different statistical approaches (Hasse diagrams and fuzzy logic) to the comparative evaluation of small streams, *Ecotoxicology*, **11**, 311-321.

Ihaka, R. and R. Gentleman (1996), R: a language for data analysis and graphics, *Journal of Computational and Graphical Statistics*, **5**, 299-314.

Norton, J.P., R. Argent, R. Nathan, G. Podger and R. Vertessy (2003), Sensitivity-Assessment Needs of Complex Simulation Models for Integrated Catchment Management, In: Post, D.M. (Ed.), *MODSIM 2003: International Congress on Modelling and Simulation*, Townsville, Australia.

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

Refsgaard, J.C., J.P. van der Sluijs, A.L. Hjberg and P. Vanrolleghem (2004), *Harmoni-CA Guidance 1: Uncertainty Analysis*, European Commission Fifth Framework Programme.

Saltelli, A. and M. Scott (1997), Guest editorial: The role of sensitivity analysis in the corroboration of models and its link to model structural and parametric uncertainty, *Reliability Engineering and System Safety*, **57**, 1-4.

Weigert, F.B. and C.E.W. Steinberg (2002), Sustainable development – assessment of water resource management measures, *Water Science and Technology*, **46(67)**, 5562.

Young, W.J., A.C. Scott, S.M. Cuddy and B.A. Rennie (2003), *Murray Flow Assessment Tool – a technical description.* Client Report, 2003. CSIRO Land and Water, Canberra. Available from: http://www.clw.csiro.au/publications/