Evaluating Design Options Against Requirements: How Far Can Statistics Help?

Ian Alexander

Scenario Plus Ltd; iany@scenarioplus.org.uk

Abstract

Trading-off candidate designs against requirements is a critical activity for many projects. This is especially so where the goals of many stakeholders conflict, and therefore cannot all be satisfied.

Traditionally, weighting has been used to try to combine scores on different criteria, so as to identify a winning design. However, this has a weak mathematical basis: criteria should be independent dimensions, and may be measured in different units.

The statistical technique of Principal Components Analysis offers a robust approach: given clear data, it gives clear guidance, of the form: "if you prefer these criteria, you should favour these candidates". Otherwise, it indicates that no guidance can be given. Either way, this rightly places responsibility for decision-making on human shoulders. The outcome is an improved trade-off process for projects.

1 Introduction

A basic challenge to projects is to identify the design that meets the requirements as well as possible. The process of matching design to requirements is called 'pushback', 'trade-off analysis', or 'optioneering' in different industries, and 'satisficing' by researchers following Herbert Simon [1].

Figure 1: Tramways in Crowded City Streets: Many Competing Constraints must be Traded-off

Requirements may be boldly written as "The system shall…" imperatives. But reality stands as the proverbial 'immovable object' between stakeholders and their goals, even if they all agree. When goals conflict, as they usually do with transport projects, trading-off alternative solutions becomes a central activity in the project.

An earlier paper [2] recounts an experience in trading-off many competing constraints while choosing a tram route. Figure 1 shows several compromises in a street in Milan: trams are obliged to proceed "at walking pace"; the street is one-way because of the tram; there are no safety barriers between the tram and either motor traffic or pedestrians. That long-standing approach might not be acceptable in cities unfamiliar with trams.

The problem of choosing a tram route under many stakeholder pressures is analogous to choosing a software architecture to simultaneously "satisfice" a set of desired qualities such as safety, availability, performance and scalability.

Paper [2] described a pragmatic approach for the following reasons:

i. The client wished to adopt a traditional approach to choosing a winning option, essentially by summing the scores awarded to an option on different criteria.
ii. The client initially favoured an indefensible approach using a very weak form of weighting. This was replaced by a combination of triage (to reject unsuitable options) and hierarchical weighting (to combine scores in an auditable way).
iii. The plausibility of the result was open to doubt, so it was checked with a sensitivity analysis, and it was found to be surprisingly stable. However, there are some persistent challenges to that approach, forming the problem that this paper aims to solve.

This paper describes work done on two transport projects – follow-up analysis on [2], and new analysis on a different route to validate the approach. It is structured as follows:

Section 2 states the problem.
Section 3 describes the approach taken on the project.
Section 4 discusses some possible conclusions.

2 The Problem

a) There is no mathematical basis for combining scores on different dimensions, especially if these are measured in different units (money, tons of $C0_2$, etc). If the assessment criteria are genuinely independent as they should be, then they represent different axes (X, Y, Z, etc in a multi-dimensional space), and cannot be added, with or without weights. More is said on this in Section 5.
b) The organisation was uncomfortable with the use of weighting, as well it might be given the approach's

weak mathematical basis.

c) Weighting has an odd status in transport planning. On the one hand it is a traditional practice. There is strong pressure in government guidance [3] to argue on the basis of objective figures – money (eg direct costs and net present values) where possible. On the other hand, there is explicit acknowledgement in the same government guidance that not everything can be reduced to costs, ie that there are multiple dimensions to an assessment.
d) The fact that the sensitivity analysis showed that the result remained practically unchanged despite large perturbations of the weights was, on reflection, somewhat puzzling. If changing the weightings dramatically did not change the output, what was happening? To put it another way, if a sum depends on the numbers you put in, but changing them makes no difference, there has to be some explanation.

Therefore, follow-up work was undertaken to re-analyse the data, avoiding the use of weights. The challenge was threefold:

a) to understand the earlier results,
b) to see if they were valid, and in particular
c) to identify an evaluation method, applicable to future projects, that was mathematically sound (and did not rely on tricks such as weighting).

3 Approach Taken on the Project

The starting-point was the re-examination of the data used in [2]. One source of discomfort was that any differences in outcome that would be predicted by individual criteria had been submerged by grouping criteria into standard "Objectives". Therefore the criteria were used individually (Table 1).

The scores were awarded by human experts (eg in town planning and the environment) on a scale from -3 to +3 (very poor to very good). For ease of analysis the scores are normalised to run from 0 to 6. Criteria that do not discriminate between the options are excluded.

These data were then analysed using Principal Component Analysis (PCA). PCA is a proven statistical method for exploring the causes of variation in non-parametric data [4].

The basic idea of PCA is to treat each criterion as a separate, independent variable. Two variables can be plotted against each other on a sheet of paper, giving each variable an independent direction: one across the page (the X-axis); one up the page (the Y-axis). Twenty-six variables require a 26-dimensional space, which is hard to visualize but simple for PCA to analyse. It does not matter if some dimensions are less important than others: no weighting is involved. If, for instance, several Construction criteria were merged into one, detail would be lost but the results would otherwise be unaffected. The choice of criteria naturally reflects the concerns of informed stakeholders.

The goal of PCA is to identify a small number of new variables that can more readily be understood than the large number of input variables – perhaps only two or three. This only makes sense when a few variables can be found that explain most of the variance in the input data.

Criterion Option	A	B	C	D	E	F	G
Landscape	4	2	2	1	2	0	0
Townscape	1	1	6	5	5	2	1
Heritage	5	0	2	1	1	2	0
Biodiversity	1	2	5	4	1	3	1
Ambience of Tram Stops	4	5	1	2	2	4	4
Whole-Life-Cost	2	6	4	0	0	5	5
Business	4	5	0	2	2	4	4
Value-to-Passengers	5	6	0	2	4	5	5
Reliability	4	5	4	4	4	5	5
Regeneration	5	5	4	5	5	5	5
Section-12-Fire-Regs-Costs	6	6	6	0	0	6	6
Safety	3	4	4	4	4	4	4
Security	4	4	0	2	2	4	4
Access	6	6	2	3	4	6	6
Participation	6	6	0	5	6	6	6
Interchange	6	6	0	4	4	6	6
Land-Use-Policy	5	5	6	5	5	5	5
Other-Policy	4	5	5	5	5	5	5
Independence-of-Other-Projects	2	2	1	1	1	5	2
Extensibility	6	6	2	0	0	5	5
Site-Access	5	4	4	3	3	4	4
Effect-on-Existing-Structures	5	2	3	4	4	4	4
Construction-Disruption	4	2	2	0	0	2	2
Effect-on-Utilities	5	2	3	1	1	3	3
Construction-Duration	6	3	2	0	0	4	2
Construction-Complexity	6	4	5	1	1	4	4

Table 1: Raw Scores: Each Option on Each Criterion

3.1 Principal Components Analysis (PCA)

The procedure works like this: first, examine the cloud of points in multi- (in this case 26-) dimensional space. Draw a new line (an eigenvector) through the cloud's longest axis to explain as much as possible of the variation (using the method of 'single value decomposition'). The percentage of the total variation is the eigenvalue. Then, draw another new line, at right angles to the first new line, to explain as much as possible of the remaining variation. Continue in this way until effectively all the variation is explained.

Draw a "scree plot" graph to show how much of the variance is explained by your eigenvectors (Figure 2). The eigenvectors are called "principal components".

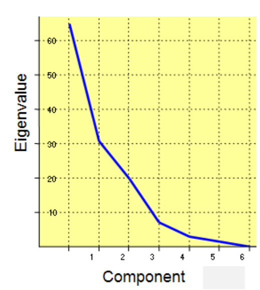

Figure 2: Scree Plot

Subjectively, statisticians simply discard any principal components after the "knee" in the graph, here clearly near the 3rd component. A rough-and-ready statistic called the Jolliffe cut-off can also be calculated to identify how many components can safely be used. Here the Jolliffe cut-off is 3.4, so there is probably little to be gained by examining more than the first 3 principal components. Table 2 shows how much of the variance is explained by the first 6 components.

PC	Eigenvalue	% Variance	Cumulative % Variance
1	64.8625	51.330	
2	30.6805	24.280	75.610 % explained
3	19.6643	15.562	91.172 % explained
4	6.9589	5.507	Below Jolliffe cut-off
5	2.82429	2.235	
6	1.37228	1.086	

Table 2: Explaining the Variance

In this case, PCA has "worked": nearly all the variance is explained by the first 3 components.

	Eigenvectors		
Criterion	PC1	PC2	PC3
Landscape	-0.02542	0.2406	-0.3303
Townscape	0.2903	0.05298	-0.02737
Heritage	-0.0724	0.2882	-0.3098
Biodiversity	0.1771	0.1549	0.2317
Ambience	-0.268	-0.1063	0.06726
Whole-Life-Cost	-0.1661	0.06827	0.4377
Business	-0.259	-0.1556	0.01166
Value-to-Passengers	-0.2527	-0.1944	-0.04621
Reliability	-0.128	-0.1127	0.2424
Regeneration	-0.1167	-0.2321	-0.1581
Section-12-Fire-Regs	-0.2001	0.2409	0.2751
Safety	0.07079	-0.1091	0.1807
Security	-0.2737	-0.1563	-0.03905
Access	-0.2791	-0.1128	-0.0086
Participation	-0.1841	-0.2962	-0.1962
Interchange	-0.2445	-0.2151	-0.09936
Land-Use-Policy	0.08455	0.1682	0.1146
Other-Policy	0.1097	-0.1691	0.2799
Independence	-0.1848	-0.03844	0.1624
Extensibility	-0.2801	0.1005	0.1333
Site-Access	-0.1599	0.2122	-0.0092
Effect-on-Existing	-0.04227	0.03422	-0.3839
Construction-Disruption	-0.2144	0.2845	-0.01234
Effect-on-Utilities	-0.1836	0.3008	-0.06917
Construction-Duration	-0.2404	0.2331	-0.04483
Construction-Complex	-0.1731	0.3258	0.1003

Table 3: Eigenvectors Correlated with the Criteria

The next step is to look at the first 3 components, and try to understand what (if anything) they mean in the real world (Table 3).

It is difficult to pick out correlations from such a table. When there are only a few criteria, you can plot a bar graph of each one, and by lining up the graphs, may spot the winning option visually.

A better approach for many criteria is to cluster the data by similarity. Figure 3 shows clusterings both of the Criteria, on the X-axis, and of the Options on the Y-axis. (Standard Criteria like Noise that do not discriminate between options are shown here but not elsewhere.)

The clustering of Options is useful, as it effectively gives guidance like "if you like Option E, you should also like Option D – closely similar to it – and Option C – the next closest". In the same way, Options B, F, and G are seen to be similar to each other, while A is a little further away. But it remains hard to pick out the reasons for these similarities and differences in terms of the Criteria. What we need is a reliable way to reduce the number of criteria without making dangerous assumptions.

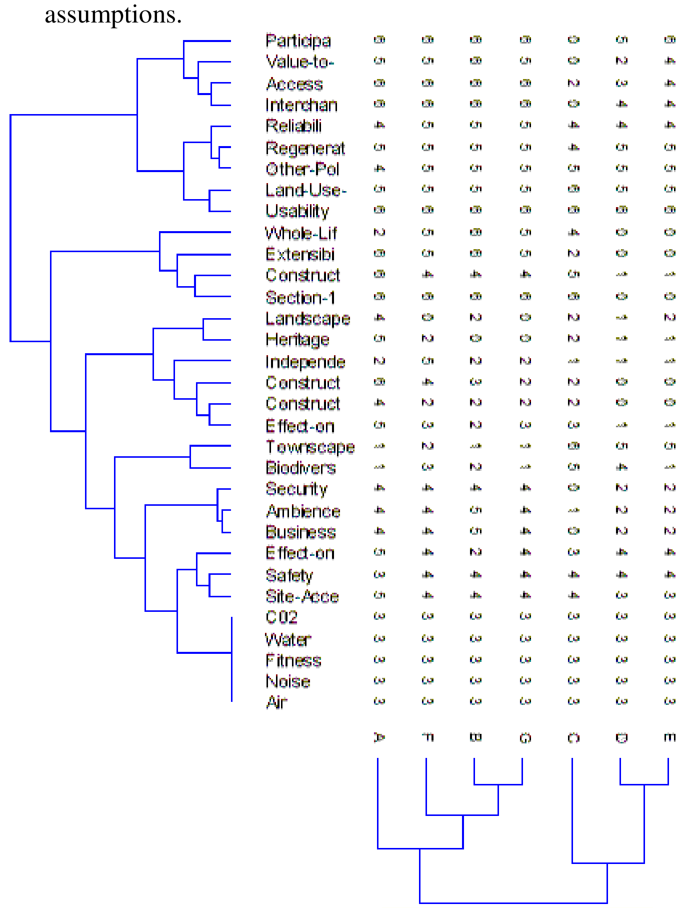

Figure 3: Cluster Analysis of Criteria and Options

Fortunately, PCA offers just such an approach. Recall that the principal components are eigenvectors, newly-drawn lines through the cloud of data points. The first two components explain three-quarters of the variation in the data. We can use them as our new X- and Y-axes, and project all the rest of the data – both Criteria and Options – on to that new plane (Figure 4). The Options are boldly labelled A, B, … G. The Criteria radiate from the origin.

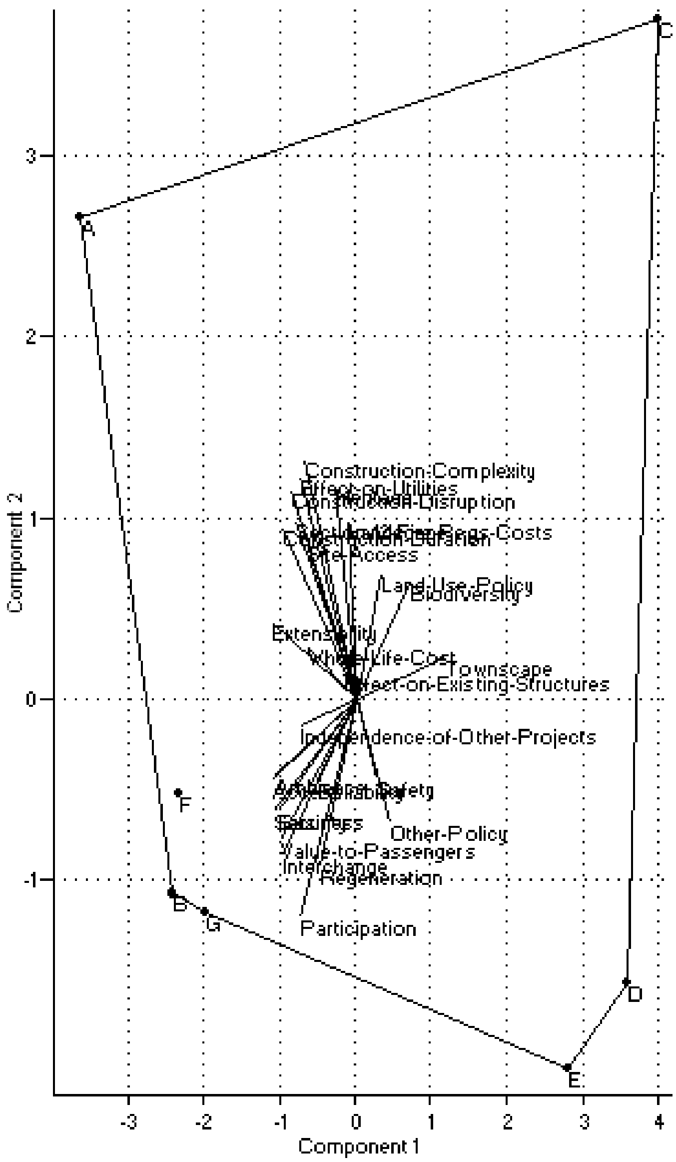

Figure 4: Criteria and Opt ions Projected on to the New Plane Defined by the First 2 Components

The locations (correlations) of the Options in the X-Y-Z space defined by the first three principal components are shown in Table 4. Only the first two principal components (PC1 and PC2) are used in the X-Y plane of Figure 4.

Option	PC 1	PC 2	PC 3
A	-3.654	2.6647	-2.8278
B	-2.404	-1.0843	1.6644
C	3.9844	3.7493	1.6366
D	3.5815	-1.578	-0.93884
E	2.7848	-2.0475	-1.7142
F	-2.3199	-0.52639	1.1981
G	-1.9728	-1.1779	0.98175

Table 4: Option Co-ordinates in the Space Defined by the First 3 Components

3.2 Interpreting the Results

The statistics have now taken us as far as they can. Views like Figure 4 spread the options apart as widely as possible. Similar views can be prepared, using principal components 1 and 3, or 2 and 3, effectively looking at the cube of data from the side or the front rather than the top.

The first task is to decide what the axes of Figure 4 mean in the world. They are by definition not the same as any of our input criteria, so we have to examine the correlations – positive or negative – of the criteria with the axes, and imagine what the axes could mean.

If we are lucky, PCA reveals something important about the world, or at least about how the experts who scored the options perceive the world. In that case, the axes may have a clear and definite interpretation.

Indeed, if several projects involve similar kinds of trade-off, as could happen with some types of transport project, then it might be possible to score the options directly and efficiently on the newly-discovered axes. This would save the trouble of scoring the options on the large number of old criteria, which often correlate only weakly with the maximally discriminative axes.

Component 1, the X-axis, is reasonably strongly correlated positively (to the right) with just one criterion, townscape. In the negative direction, there are several components like value to business, value to passengers, and interchange, which all have to do with the usefulness of an option. This could be called the "Usefulness—Townscape Axis".

Component 2, the Y-axis of Figure 4, similarly correlates negatively with several social value criteria, and positively with several buildability criteria, as shown in Table 2. This could be called the "Social Value—Buildability Axis".

Armed with this understanding, the next task is to look at the trade-offs in the light of the statistical analysis displayed in Figure 4.

Essentially the positions of the Options A..G on our two newly-named Axes say:

(i) "If you prefer townscape to usefulness, you should choose among the options C, D, or E; if not, you should choose from A, B, F, G." and

(ii) "If you prefer buildability to social value, you should choose options A or C; if not, you should choose from B, F, G, D, or E."

This places responsibility for the final decision firmly on human shoulders: there is no mathematical reason to favour townscape, buildability or any other criterion.

On the other hand, the statistical analysis does say quite clearly that if you think social value and usefulness the most important considerations, it is illogical of you to choose option C, for instance.

3.3 Reinterpreting the Results of Weighting

The final step in this analysis is to reflect the findings back on to the earlier work [1], done without the benefit of PCA, and to reinterpret what happened then. The recommendations in [1] were that Option A should win, followed by Option B and then Option C. This looks surprising in the light of PCA, for the following reasons:

- If Option A is preferred because of its buildability, then Option C should have been close to it, or indeed preferred to it; no other option should have come close, so the choice of third place could have gone anywhere (F, B, G, D, or E).
- On the other hand, if Options A and B were preferred because of their usefulness, then third place should have gone to Option F or G; Option C should have been placed last.
- Either way, if B is considered a good second-placed candidate, then F and G, which are very similar to B, should probably have been placed 3rd and 4th.

A more complex position is also possible: perhaps A won because it scored best on both Buildability and Usefulness. In that case, we can imagine contours of merit running diagonally across Figure 4: A is at the top of the slope, and D and E are at the bottom. In that case, B, F, G, and C might perhaps get similar scores, half-way down the slope. The only problem with this is that 'social value' must then be rated highly, but 'usefulness' must be low, which is not easy to reconcile as the groups of criteria feel similar.

3.4 Reinterpreting the Sensitivity Analysis

One mystery with the earlier work is resolved by the new analysis. That was, that sensitivity analysis which made very large variations in the assigned weights made little difference to the outcome – A or B came first almost regardless of changes in weights.

The explanation is that the sensitivity analysis explored the group weightings as opposed to the weights on individual criteria. Varying the weight of a whole group could reduce the positive effect of one criterion, but also the negative effect of another criterion – which would thus "cancel out". Within the Environment group, for instance, Biodiversity and Ambience pull in opposite directions on the first two principal components.

A much larger sensitivity analysis, working with individual criteria, could in theory have discovered this effect. Unfortunately, with so many criteria to consider, simply varying one out of many criteria might have little effect; to explore sensitivity properly, one ought to vary many combinations of weights of criteria – a daunting task unless automated. PCA side-steps the difficulty.

3.5 Validating the Approach on a Second Project

The last step was to apply the technique to a different project, to see if it gave a sensible result there. The opportunity arose to review a rapid transit project.

That project triaged out several options as wholly unsuitable ('hard' triage) or much less good on key criteria ('soft' triage). This left just 4 routes for detailed optioneering. The project found these hard to separate, but judged routes A and B preferable to C and D.

Data preparation and PCA took only half a day, leading to the results in Figure 5. Options C and D prove to be very similar to each other; options A and B are more distinct. The criteria form no visible pattern (unlike Figure 4) but radiate around the graph, though more criteria favour Options A and B than Options C and D.

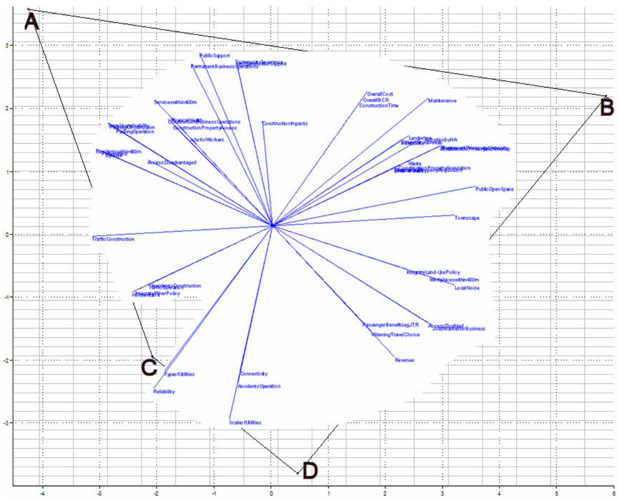

Figure 5: Applying PCA to a Second Project Data Presentation as Figure 4

These findings suggest that the project was right: the remaining options are not very different, and any of them could form an acceptable solution. However, Options A and B may be the best. Interestingly, A had most public support, as the route runs along a residential road, offering a convenient transport service. In contrast, B runs through an industrial zone, and so had the least harmful impact on townscape and public open space. Perhaps these are the decisive issues in this case. PCA appears to have homed in on the essence of the project.

4 Discussion

4.1 A Reason for Discomfort

It is satisfying to arrive at a simple, general, method that effectively explains one's discomfort with an earlier approach, mathematically and graphically. The hierarchical grouping of criteria, essential to the balanced allocation of weights, had masked genuine differences between the options. Further, the essence of any weighting method is to melt down all differences in all dimensions to one score or figure of merit. "You are blending all the colours down to a greyish-brown mush", as one colleague expressed it. This applies even to mathematically sophisticated techniques such as the Analytic Hierarchy Process (AHP) [7] which computes weights from human pairwise comparison of criteria.

PCA showed that there were real differences which had been missed by the weighting approach. If option B was considered best, then option C should not have been picked as runner-up: options F and G were closest to B, while C was far away. This was reason enough to feel unhappy with the old approach; and conversely, to feel that with PCA, it was finally possible to understand what was happening. Paradoxically, hierarchical weighting had been adopted because it seemed more objective.

4.2 PCA: A Modest Method

PCA implies a somewhat modest method.

4.2.1 Not a silver bullet – does not work every time

Firstly, PCA is not a silver bullet. It is an old and reliable statistical method. But it does not guarantee to find a result: indeed, there is no assurance that it will help at all on some evaluations, particularly where the scores pull in different directions.

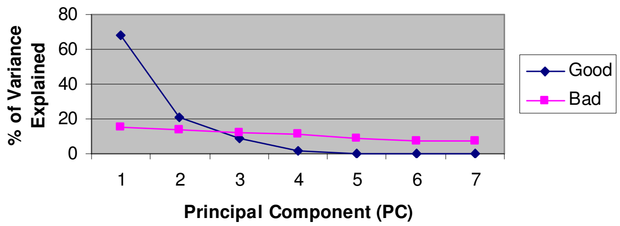

Figure 6: When PCA does not help

Fortunately, it is straightforward to tell when PCA is not working: there are many principal components of similar size, so the first few components only account for a small part of the variance (Figure 6). That is a warning that decision-making will be difficult, that many factors must be compared, and that mathematics cannot help.

It is hard to avoid the conclusion, reflecting on the way that PCA works, and the meaning of Figure 5 for example, that approaches such as weighting, which attempt to extract an automatic answer from a complex cloud of conflicting factors, are unjustifiable. That in turn probably rules out a Shortest Path or Travelling Salesman algorithm to combine scores on route segments.

From a wider perspective, this result points to the need for a Soft Systems approach, eg Checkland's [8]. Option selection is part of a larger process of solving complex business problems. That involves human activities like consultation, reflection, explanation and decision-making. These can be supported by analysis and modelling, but never replaced by them. Engineering recommendations exist in the context of politics, economics, and social and environmental concerns.

4.2.2 Takes you so far – and no further

Secondly, using PCA recalls to mind the mining proverb: *"When the miner strikes gold, he throws away his tools"*. When PCA works well, it shows clearly what choices must be made. Then the mathematical tools must be thrown away, and human responsibility shouldered. Trade-offs are painful; every choice comes at a price in dissatisfied stakeholder goals. Perhaps uncomfortable responsibility is the true meaning of "satisficing": PCA provides evidence to support a decision. But it offers no easy "proof" to hide behind.

4.3 A Process Outcome

A desirable outcome of the work is that the organisation's process manual has been updated, downplaying the use of weights and suggesting the use of

a) selection of the most appropriate evaluation criteria for each project, by customising a standard list
b) independent scoring of each option on each criterion
c) statistical analysis to show the differences between the scored options
d) human evaluation of the (statistical) findings.

This is part of a larger process which includes an initial feasibility study and cost/benefit analysis. Possible solutions – in the case of a tram, candidate routes – are then developed. Routes which face major obstacles are triaged out, though the reasons for their rejection must still be carefully documented to demonstrate fairness.

Triage in this sense is closer to the original medical usage – sorting patients into a small number of categories to be treated urgently, later, or not at all – than to that in [5] which seems unnecessarily to equate triage with prioritisation in general. However, [6] reverts to a more traditional understanding, seeing triage as the desired sorting-into-groups outcome, for which the prioritisation of requirements may be an input.

Evaluation of many route options by experts against a battery of 30 or more criteria is slow and costly. Since, as explained in [2], the number of route options expands combinatorially, early triage to reject unworkable route segments can bring large savings.

Statistical analysis of the findings by PCA is quick and cheap. While the results of PCA require some skill to interpret [9], the approach described here is robust, widely applicable to the trade-off problem, and promises to be extremely cost-effective.

5 Acknowledgements

The author is grateful to Jane Cleland-Huang and Martin Glinz for their encouragement to write this paper.

6 References

[1] H.A. Simon, The Sciences of the Artificial, 3rd Edition, MIT Press, 1996. (1st Edition 1969)

[2] I.F. Alexander, *Choosing a Tram Route: An Experience in Trading-Off Constraints*, Proceedings of the 15th IEEE International Requirements Engineering Conference, New Delhi, October 15-19, 2007, pp 350-355.

[3] *Transport Analysis Guidance, Appraisal TAG Unit 3.2*, February 2004. (http://www.webtag.org.uk/webdocuments/3_Expert/2_Appraisal/index.htm)

[4] C. Dytham, *Choosing and Using Statistics,* Blackwell, 2003.

[5] A. M. Davis, *Just Enough Requirements Management*, Dorset House 2005.

[6] P. Laurent, J. Cleland-Huang, C. Duan, *Towards Automated Requirements Triage*, Proceedings of the 15th IEEE International Requirements Engineering Conference, New Delhi, October 15-19, 2007, pp 131-140.

[7] T.L. Saaty, *Priority Setting in Complex Problems*, IEEE Transactions on Engineering Management, Vol. EM-30, August 1983, pp 140-155.

[8] P. Checkland, *Soft Systems Methodology,* Wiley, 1990.

[9] L.I. Smith, *A Tutorial on Principal Components Analysis*, 2002. (http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf)