Prioritization of national road projects in Saudi Arabia: weighting of multiple criteria

R. Abbyad[1] & I. Kaysi[2]
[1]*SETS sarl Transportation Consultants, Beirut, Lebanon*
[2]*Department of Civil and Environmental Engineering,*
American University of Beirut, Lebanon

Abstract

The Ministry of Transport of the Kingdom of Saudi Arabia is in the process of developing a methodology for the prioritization of proposed primary and secondary road projects in the Kingdom. This process was carried out using a multi-criteria analysis model. The model, called the **Sa**udi **Ro**ad **P**rioritization model (SaRoP model), consists of a linear additive model where the overall performance of a proposed road project is measured by the sum of its weighted performance on a set of criteria. With such a model, a low performance of the project on one criterion may be offset by a high performance on another. The degree of compensation between the criteria is determined by the relative weights. These weights are derived using the Analytic Hierarchy Process (AHP) developed by Saaty. Ministry experts and KSA academics, as well as international consultants, were asked to fill out a questionnaire, which consists of providing their pair-wise comparisons of all criteria within a certain level of the hierarchy. The results indicated the importance that decision-makers and experts place on socio-economic development and the improvement of the transportation network over other criteria. In order to reach these results the data was organized according to consistency and homogeneity and a set of weights was derived by checking the results from several different computational models. This paper describes the research and weighting methodology of the decision-making model. It is concluded that AHP is a useful and efficient tool for organizing priorities and aggregating decisions.

Keywords: analytic hierarchy process, multi-criteria analysis, linear additive model, consistency ratio, aggregation of individual preferences, aggregation of individual judgements, vector space AHP, singular value decomposition AHP.

1   Introduction

This paper describes the research and application of the weighting methodology used in the Saudi Road Prioritization model (SaRoP model). The aim of the SaRoP model is to use a hierarchy of criteria in order to pursue the goal of prioritizing alternatives of primary and secondary road projects in the Kingdom of Saudi Arabia (KSA). The top level of the hierarchy includes the broadest criteria to be taken into consideration. These are macro-criteria that reflect the government's development goals. For instance, socio-economic development is seen as a key criterion, which contains sub-criteria, such as the extent to which the road projects serve the population, existing economic activities, planned economic development initiatives, etc. Within this *primary level* are additional subdivisions that describe in more detail certain aspects of the previous level. For example, existing economic activities can be subdivided into categories, such as agriculture, tourism, etc. The hierarchy created facilitates decision making by comparison of criteria. The hierarchy is presented in table 1.

During a workshop in 2009, Ministry of Transport (MoT) directors and engineers involved in SaRoP (as well as from other divisions within the MoT) were asked to fill out a questionnaire which consisted of providing their pair-wise comparisons of all criteria within a certain level of the hierarchy (top, primary, or secondary). The comparisons were made when each decision maker (DM) marked the intensity of their preference for one criterion over another (less important, same, more important, etc). After the data was collected, the Analytic Hierarchy Process (AHP) was used to derive a set of numerical weights which represent the criteria's relative weights within the hierarchy. Once the weight set was calculated for each DM, an aggregation procedure was chosen in order to develop the entire group's judgments.

During the application of the AHP, a method for dealing with inconsistent preferences was employed. Several other computational models were used to verify the results of Saaty's eigenvector method.

Section 2 provides an overview of AHP as proposed by Saaty, and discusses in particular the process of consistency checking and the possible methods for aggregating individual preferences and judgments. Section 3 describes the application of AHP to the SaRoP model including the manner in which data was purged to address consistency constraints and the aggregation approaches that were considered. Finally, section 4 provides some concluding remarks.

2   Saaty's AHP

In the preliminary analysis of an appropriate prioritization framework, Saaty's AHP was chosen over other weighting methods because it is systematic, easy to implement, and transparent. It also facilitates the development of a timescale for the implementation of said road projects. Other methods considered were the swing weighting method and trade-off method. The trade-off method is not very transparent, can lead to inconsistencies, and is difficult to implement, whereas

Table 1:   SaRoP hierarchy.

Top level	Primary level	Secondary level
Transport network (A)	Traffic demand and supply (A1)	Traffic demand (A1-1)
		LOS without the project (A1-2)
		Additional capacity (A1-3)
		Freight traffic (A1-4)
	Connectivity (A2)	Multi-modal connectivity (A2-1)
		Connectivity within national road network (A2-2)
		International connectivity (A2-3)
Socio-economic development (B)	Population (B1)	
	Support for existing economic activities (B2)	Manufacturing (B2-1)
		Agriculture (B2-2)
		Tourism (B2-3)
		Other services (B2-4)
		Mining (B2-5)
	Support for economic development initiatives (B3)	Manufacturing (B3-1)
		Agriculture (B3-2)
		Tourism (B3-3)
		Economic cities (B3-4)
		Mining (B3-5)
	Support for urban development strategy (B4)	

Table 1:   Continued.

	Time savings (C1)	
Economic feasibility (C)	Savings in VOC (C2)	
	Cost/km driven (C3)	
	Hajj traffic volume (D1)	
	LOS during Hajj (D2)	
Hajj and Umrah (D)	Umrah traffic volume (D3)	
	LOS during Umrah (D4)	
	National security (E1)	
	Safety (E2)	
Other (E)	Environment (E3)	
	Political commitment (E4)	

Table 2:   Saaty's comparison scale.

Intensity of Importance	Description
1	Equally important
3	Moderately more important
5	Strongly more important
7	Very strongly more important
9	Extremely or overwhelmingly more important

the swing method is not flexible to new options. Within AHP itself, two theoretical aspects can be implemented: a metric that expresses the reliability of the derived weights in an objective way, and an appropriate procedure for aggregating the data from individual DMs.

2.1 Consistency of pair-wise comparisons and derivation of weights

Table 2 provides the 1-9 scale Saaty [1] proposes for judging the relative importance of two criteria. This scale is based on two factors: 1) the way the mind responds to stimuli; 2) the division of qualitative distinctions into nine intensities.

After DMs complete the questionnaire, all pairs of criteria within a given branch of the hierarchy have been judged. These judgements should be consistent: if criterion A is 3 times more important than criterion B, and criterion C is 3 times more important than criterion B, then C is 3x3=9 times more important than A. In addition, if A is 9 times more important than B, then this is equivalent to saying B is 1/9 as important as A. Thus, not only must judgments be transitive, they should also lie within the 1-9 scale and be proportional. This implies that the criteria in the hierarchy being compared must belong to this magnitude scale. In other words, two elements in the hierarchy must not be compared to one another if one is more than nine times important than the other.

A metric for checking the consistency of the preference data and the relative weights of the criteria in the hierarchy can be derived by considering each branch of the hierarchy as a matrix of preference values. If the consistency condition holds then there exists a unique vector of weights for the criteria.

2.1.1 Pair-wise comparison matrices (PCMs)

The metric for consistency and the relative weights of the criteria can be derived by organizing the pair-wise comparisons in the form of a matrix as given by Saaty [1]:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ \frac{1}{a_{1n}} & \cdots & a_{nn} \end{pmatrix}. \tag{1}$$

Here n is the number of criteria being compared and a_{ij} is the magnitude from the Saaty scale comparing the importance of criterion i to criterion j.

Note that:

$$a_{ij} = \frac{1}{a_{ji}}. \tag{2}$$

In addition, the consistency condition can be stated as:

$$a_{ik} = a_{ij}a_{jk}. \tag{3}$$

2.1.2 The ratio scale

The ratio scale is best described by using an example. Suppose one wishes to compare the weights of three objects. Let w_1, w_2, w_3 represent the weights of objects A, B, and C. Then the ratio $\frac{w_i}{w_j}$ is unit-less. Suppose that all such ratios are known. Let,

$$v_i = \frac{w_i}{w_1 + w_2 + w_3} = \frac{1}{\frac{w_1}{w_i} + \frac{w_2}{w_i} + \frac{w_3}{w_i}}. \tag{4}$$

Thus, v_i is also unit-less and represents the proportion of w_i to the total weight of the group. The vector $v = (v_1, v_2, v_3)$ represents a unique quantity, whereas $w = (w_1, w_2, w_3)$ is not unique since the magnitudes are unknown. Therefore, if all ratios are known a weight set can be derived expressing objects' relative importance.

Note that w is unique up to a constant multiple.

2.1.3 The eigenvector/ eigenvalue formulation of the problem

Saaty [1] proved that the consistency condition for a PCM is equivalent to its being expressible in the following form:

$$W = \begin{pmatrix} \dfrac{w_1}{w_1} & \cdots & \dfrac{w_1}{w_n} \\ \vdots & \ddots & \vdots \\ \dfrac{w_n}{w_1} & \cdots & \dfrac{w_n}{w_n} \end{pmatrix}. \tag{5}$$

Furthermore W is consistent if and only if:

$$Ww = nw \quad \text{where } w = (w_1, w_2, \ldots, w_n)^T. \tag{6}$$

Moreover, W is consistent if and only if the maximum eigenvalue is n and all other eigenvalues are equal to 0 (note that W has rank equal to one).

Any (PCM) can then be derived from a consistent matrix W by perturbing its elements by an appropriate quantity. It is known that small perturbations in W lead to small perturbations in the eigenvalue (Saaty [1]). With that result, the nearer the eigenvalue to n, the more consistent the matrix, the more accurate the resulting weight set v, which is derived from w. Thus, the maximum eigenvalue and corresponding eigenvector completely classify the problem of measuring inconsistency and forming a unique quantity v that measures the relative weights of a set of criteria. If the maximum eigenvalue differs greatly from the number of criteria n, the derived weights must be discarded as inaccurate.

2.1.3.1 Consistency ratio (the AHP metric) The maximum eigenvalue of any PCM, λ_{max}, serves as a metric for measuring consistency. It has been shown that, $\lambda_{max} \geq n$. Consider the equation:

$$CR = \frac{\lambda_{max} - n}{\lambda'_{max} - n} \tag{7}$$

where λ'_{max} is an average value taken from randomly generated PCM's of size n, then the CR will be a metric for the inconsistency of the PCM. Saaty [2] has suggested that 10% randomness (CR= 0.10), or at most 20% is acceptable.

2.1.3.2 The derived ratio scale Once acceptable consistency has been achieved, the eigenvector w corresponds to a nearly consistent matrix. Each component of this w over the sum of all components of w results in a meaningful weight set v. Furthermore, the components of v sum to one (see eqn (4)).

2.2 Aggregation method

Two aggregation methods are discussed in this section: aggregation of individual preferences (AIP) and aggregation of individual judgments (AIJ). Each method produces an aggregate set of weights for the entire group of DMs from the individual weights derived from each PCM. Regardless of the aggregation method, all averages are taken using the geometric mean at Saaty's [3] suggestion. The geometric mean is the only mean that satisfies all of the social choice axioms (decisiveness, unanimity, independence of irrelevant alternatives, no dictator). The next step is to describe AIP and AIJ in order to determine which one is more appropriate for the SaRoP model.

2.2.1.1 AIP method This method uses the pair-wise comparisons provided by all DMs which, after averaging, are entered into a single PCM. A single calculation of the eigenvector of this composite matrix produces the group weight set. However, such an aggregation is precarious since it ignores individual consistency ratios. Suppose, for instance, as is often the case with the data collected for the SaRoP model, that an individual has a weight set for a set of criteria which is extreme when compared to others. If this individual has a high enough consistency ratio, say his PCM is 50% random, far greater than the 10% allowed by Saaty, then his judgments are still given equal weight in calculation and affect the components of the eigenvector. Yet, when combined with a matrix of near perfect consistency the CR of the aggregate PCM can be quite low. The following example illustrates the shortcomings of this procedure.

Consider two DMs and their PCMs comparing three criteria:

$$A = \begin{pmatrix} 1 & 3 & 1 \\ \frac{1}{3} & 1 & \frac{1}{5} \\ 1 & 5 & 1 \end{pmatrix} \text{CR}{=}.03 \quad B = \begin{pmatrix} 1 & 7 & 7 \\ \frac{1}{7} & 1 & 9 \\ \frac{1}{7} & \frac{1}{9} & 1 \end{pmatrix} \text{CR}{=}.48$$

Only the weight set of A is valid. The weight set of A, w_A, and the aggregate weight set using AIP, w are:

$$w_A = (.41, .11, .48) \quad w = (.64, .18, .18)$$

Apparently, the values of the weight set and even rank of the weight set is changed quite drastically. Furthermore, the aggregate PCM has CR= 0.04, which is well within the proposed benchmark.

2.2.1 AIJ method
In contrast to the AIP method, the AIJ method allows for the removal of inconsistent matrices before the aggregation procedure. The aggregate weight set w, is computed by taking the component-wise average of all individual DMs' weight sets which have acceptable CRs.

It can be demonstrated mathematically that the aggregate weights obtained by AIP and AIJ are exactly the same for the consistent case, but that they can change significantly in the inconsistent case and induce a distinct ranking of the criteria. Therefore, it appears AIJ is better suited for a heterogeneous group of decision-makers that have variable levels of consistency.

3 Application of AHP to the SaRoP model

Though the application of AHP to the SaRoP model is straightforward, many of the PCMs that resulted from the questionnaires did not have an acceptable CR. This section of the paper describes how the CR problem was mitigated, the chosen aggregation procedure, and a validation of SaRoP's initial results by crosschecking with similar models that differ in computational and theoretical structure. These techniques helped to develop a stronger and more reliable set of weights.

3.1 Initial results

The far right column of Table 3 (see 3.2.1.1) shows the initial weight set to be used in SaRoP. The top five overall weights are:
B1 – Population directly served by the proposed road project
B4 – Consistency of road project with national urban strategy
A1-2 – Deterioration of level of service in case road project is not implemented
A1-1 – Expected volume of traffic (AADT) on road project if implemented
B2-1 – Ability of road to serve existing manufacturing/industrial activities

3.2 The issue of inconsistency

Ultimately, the reliability of the weights depends on how consistent the pair-wise comparisons are. If they are highly consistent, the weight set is accepted, if not, then care should be taken. The less consistent the data, the less the eigenvector reflects the true preferences of the DM. However, the integrity of the project also depends on a sufficient amount of data. For example, setting a CR benchmark of 0.10, leaves 44% valid data (70 out of 160 PCMs have CR below 0.10). On the other hand, a benchmark of 0.20 leaves 70% valid data (112 out of 160 PCMs). The issue is resolved by taking the 0.20 benchmark and noting any deviations in the weight set when additional data in the 0.10–0.20 range is removed on the basis of extremeness of judgments (purging of outlier data).

3.2.1 Purging data on the basis of extremeness of weights and CR
It can be argued that the data which is most hazardous to the integrity of the SaRoP model is data which varies from the mean or median significantly and is inconsistent. Thus, the data which does not fall below the CR benchmark of 0.10 can be purged on this basis while keeping other data in the 0.10–0.20 range. This further limits the number of people contributing judgments and also serves as a comparative measure which reflects the stability of the 0.20 data set. If this purging influences the weight set enough so that the ranks of the criteria change, then these weights are sensitive to additional judgments. In other words, an

additional DM contributing his/her judgment could easily change the distribution of weights and ranking among them if all judgments falling below the 0.20 consistency benchmark are accepted. Finding the extreme weights can be done by the measure of cohesion introduced later in the vector space AHP model and one can check that this corresponds to the values of box plots of the weights for each criterion.

Table 3:   The effects of outlier removal on weights and their rankings.

TOP	No Adjustment for Inconsistency		.2 Benchmark		Outliers level 1		Outliers level 2		SaRoP weights
A	2	.28	2	.26	2	.26	2	.26	.26
B	1	.40	1	.46	1	.46	1	.46	.46
C	4	.10	4	.10	4	.10	4	.09	.10
D	3	.15	3	.11	3	.11	3	.13	.12
E	5	.07	5	.06	5	.06	5	.06	.06
PRIMARY									
A1	1	.77	1	.77	1	.77	1	.77	.77
A2	2	.23	2	.23	2	.23	2	.23	.23
B1	1	.34	2	.27	2	.27	2	.27	.27
B2	2	.26	1	.30	1	.30	1	.30	.30
B3	4	.16	4	.20	4	.20	4	.20	.20
B4	3	.24	3	.24	3	.24	3	.24	.24
C1	1	.47	1	.43	1	.43	1	.43	.43
C2	2	.31	3	.27	3	.27	3	.27	.27
C3	3	.23	2	.30	2	.30	2	.30	.30
D1	3	.25	1	.29	2	.24	2	.25	.26
D2	4	.18	3	.22	3	.24	3	.24	.23
D3	2	.26	4	.20	4	.18	4	.14	.17
D4	1	.31	2	.30	1	.34	1	.37	.34
E1	2	.19	2	.18	2	.18	2	.18	.18
E2	1	.61	1	.62	1	.62	1	.61	.62
E3	3	.11	3	.12	3	.12	3	.13	.12
E4	4	.08	4	.09	4	.09	4	.08	.08
SECONDARY									
A1-1	2	.29	2	.30	2	.30	2	.30	.30
A1-2	1	.33	1	.32	1	.32	1	.32	.32
A1-3	4	.17	4	.18	4	.18	4	.18	.18
A1-4	3	.21	3	.20	3	.20	3	.20	.20
A2-1	2	.29	2	.22	2	.22	2	.22	.22
A2-2	1	.54	1	.62	1	.62	1	.62	.62
A2-3	3	.17	3	.16	3	.16	3	.16	.16
B2-1	1	.40	1	.38	1	.41	1	.45	.41
B2-2	3	.15	4	.12	4	.14	4	.13	.13
B2-3	5	.10	5	.09	5	.08	5	.08	.08
B2-4	2	.22	2	.26	2	.23	2	.19	.23
B2-5	4	.14	3	.14	3	.15	3	.14	.15
B3-1	1	.32	1	.32	1	.29	1	.30	.29
B3-2	4	.12	4	.13	5	.11	5	.11	.10
B3-3	5	.10	5	.12	4	.11	4	.13	.13
B3-4	2	.24	3	.22	2	.25	3	.22	.22
B3-5	3	.21	2	.22	3	.24	2	.25	.26

3.2.1.1 Results of data purging Table 3 shows the weights resulting from each layer of outlier removal and the resulting rankings of their magnitudes.

The column labelled "Outliers Level 1" represents the reduced data set after purging the most extreme DM or two (depending on the size of data set). The column labelled "Outliers Level 2" represents the reduced data set after purging the next one or two most extreme individuals. The data in shaded cells indicate a shift in rank after the removal of outliers from the 0.20 benchmark data set. Thus the results for the Top level matrix, which experiences no change in rank in all three data sets, are strong (low level of volatility). The A branch of the primary level exhibits the same characteristics. C is also unchanged in the three relevant data sets (remember the first data set is unacceptable). The same holds for B. The same is true for A1, A2, and B2 in the secondary level.

The data in D exhibits weaker results for the 0.20 benchmark with the results that removing outliers changed the rankings of D1 and D4. This group exhibits the lowest level of homogeneity; its coherence is 53%. As data is removed, coherence improves to 63% which is still inadequate. The same phenomenon is observed with B3.

Note that the major changes in magnitudes of the weights occur between the data which is uncorrected for consistency and the data obeying the 0.20 benchmark and not between the various layers of data purging.

3.3 Appropriate aggregation procedure (AIJ)

Low levels of consistency among the DM's, and the discussion in 2.3.2 lead to the choice of AIJ over AIP. Interesting to note is that when using AIP to aggregate, the resulting PCMs have a CR well below 0.10.
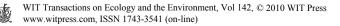
3.4 Model comparison

A brief description of the alternative implementations of AHP is now given. All use PCMs, but the difference lies in the computation of the weight set. Within each model the stability of the data (sensitivity to removal of judgments) can be noted by the impact of data removal. All models are analyzed side by side to verify the results of the conventional AHP of Saaty.

3.4.1 The vector space formulation of AHP (VAHP)
Saaty's version of AHP rests on the intuitive model of ratios and weights. The weights are normalized so that they represent portions of a pie and their sum, sums to the area of the pie (which is equal to one). However, the same problem can be viewed from a geometric perspective (Zahir [4]).

In the perfectly consistent case, the weights generated are identical. The more similar the weights the more stable the conventional AHP seems to small changes in consistency (as the eigenvectors between methods are not the same when the PCM's are inconsistent).

Further, a measure of 'coherence' in a group can be defined by using the scalar product of any two preference vectors to find their angle of separation and

comparing this to the maximum angle of separation. This measure of coherence produces results that are verified by box plots and represents the homogeneity of the group. An analysis of box plots verifies the correlation between VAHP and the conventional AHP.

3.4.1.1 Results within VAHP Zahir [4] proposes that the individual weight vectors be aggregated by vector summation followed by normalization in the Euclidean norm. However, during the data purging this method resulted in more variation in the weights. In addition to changing the ranks of criteria in B3 and D, it also altered the ranks of B2. Therefore, the conventional aggregation method of AHP which uses the geometric mean was used instead. The result is that only B3 changed in rank.

3.4.2 The singular value decomposition (SVD) approach

This method as explained by Gass and Rapcsák [5] approximates inconsistent preference matrices with consistent matrices by applying the singular value decomposition (SVD). Weights are given by a combination of the values of the left and right singular vectors corresponding to the singular values of the matrix. The result is that a good approximation (using consistent matrices) is achieved for the weight set of conventional AHP. Thus, if results mimic those of conventional AHP then this is another sign of reliable results, as the SVD weights represent good approximations.

3.4.2.1 Results within the SVD approach An analysis of the rankings using the SVD approach indicates the same weight sets change rank in removing outliers as in the conventional AHP (D, and B3). However, there is the additional result that the weight set of B is ranked differently from B of the conventional AHP (although it is stable within the model itself since it does not change rank as outliers are removed). This change in ranking for B also occurs when one uses VAHP with geometric mean; however, the actual magnitudes of the differences in weight sets are small.

3.5 Strong and weak results

The changes in rank for B3 and D within each model reflect an apparent weakness (volatility) in the data itself (sensitivity to additional judgments) for these two criteria sets. On the other hand, the similarity in rankings between all the models indicates the strength of the AHP framework. There is only a slight difference in the relative weights between models.

Therefore, the conventional AHP eigenvector model is used to generate the initial weight set. To reflect the multiple levels of analysis, the geometric mean across all layers of data purging is taken and the results are normalized to represent weight sets. This is done to mitigate the problems of low consistency and data purging. The results are found in the far right column of Table 3.

4   Conclusion

The similarity of the effects of data purging on all models as well as the similarity in actual weight sets between models indicates the resilience of the AHP framework to different computational models. It is concluded that the SaRoP model exhibits reliable initial results. However, in the future the data corresponding to D and B3 should be strengthened with regard to consistency and homogeneity.

References

[1] Saaty, T.L., *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process,* RWS Publications: Pittsburgh, pp. 45-93, 2006.
[2] Saaty, T.L., Kearns, K.P., *Analytical Planning: The Organization of Systems.* Pergamon Press Inc., pp. 34, 1985.
[3] Saaty, T.L., Vargas L.G., The seven pillars of the analytic hierarchy process (chapter 2), *Models, Methods, Concepts and Applications of the Analytic Hierarchy Process.* Kluwer Academic, pp. 27-45, 2001.
[4] Zahir, S. Geometry of decision making and the vector space formulation of the Analytic Hierarchy Process. *European Journal of Operational Research,* **112**, pp. 373-396, 1999.
[5] Gass, S.I., Rapcsák, T., Singular value decomposition in AHP, *European Journal of Operational Research,* **154,** pp. 573-584, 2004.