How to support government procurement by a pairwise comparisons model

Waldemar W. Koczkodaj[1], William O. Mackasey[2], Ireneusz Smolewski[3], Ryszard Tadeusiewicz[4]

[1] Computer Science, Laurentian University, Sudbury ON P3E 2C6, Canada, wkoczkodaj@cs.laurentian.ca
[2] Consulting geologist, WOM Associates
[3] Independent ICT procurement expert
[4] AGH University of Science and Technology Mickiewicza Av. 30, 30-059 Krakow, Poland, rtad@agh.edu.pl

Abstract: This study demonstrates how a government procurement process could be improved by the pairwise comparisons method. A case study, related to assessment of project proposals is used for demonstration purpose. The project proposals were requested by a Canadian government agency to assess the environmental and public safety hazards of abandoned mines. However, the presented model is applicable (with easy-to-implement modifications) to any other case of government procurement.

Keywords: Government procurement, project proposal assessment, pairwise comparison, distance-based inconsistency, convergence, knowledge management

1 Introduction

Government procurement, also called public tendering or public procurement, is the procurement of goods and services on behalf of a public authority, such as a government agency. According to [11], it accounts for more than EUR 2 trillion, or 18% of the EU GDP and it is a substantial part of the global economy. Evidently, government procurement is important to the EU economy as the "new EU-rules on public procurement" have been recently updated.

Laws of most countries regulate government procurement to prevent fraud, waste, corruption or local protectionism. It usually requires the procuring authority to issue public tenders if the value of the procurement exceeds a certain threshold. In the USA, the principal set of rules [13] is the Federal Acquisition Regulation System. Government procurement in the European Union has been regulated and harmonized by laws since the 1970s. In Poland, government procurement is regulated by the Public Procurement Law, an act of parliament of 29 January 2004 (see [12]). However, government procurement in Poland accounts for only 8% of GNP and it is less than 18% of the EU average.

The most recent events related to public procurement in Poland are living testimony that the "legal solution" is not efficient. Evidently, there is a need for additional research of applied nature since the strict sciences have mostly been focused on processing quantitative (or objective) data rather than qualitative (subjective) data, which we use more frequently in daily life. The importance of subjectivity processing is expressed by the idea of *bounded rationality*, proposed by Herbert A. Simon (a Nobel prize winner), as an alternative basis for the mathematical modeling of decision making.

Objective (or measurable by instruments) data, perceived as precise data, are always preferred over subjective data. Subjective data are often based on professional knowledge, experience, or even feelings. In general, they are not regarded as precise as the measurable objective data. The lack of proper methods for processing subjectivity adds to the problem. However, objectivity is often illusive since there is a fine line between objectivity and subjectivity, more often than we realize it. For example, let us assume that a contractor hopes for $1,000 for his/her service. A customer might put forward a slightly lower offer, such as $999 and the seller would most likely accept. If it is accepted, the next correction could be $998. The customer could continue offering a dollar lower, but only to a certain point, because the contractor has his/her own subjective assessment of the offered service. So, what we are really often prepared to accept is highly dependent on our subjective assessment. There are not as many numbers "carved in stone" in real life as we would like to believe.

Assessing environmental damage is expensive and government agencies in Canada must use tendering processes (known as tendering) for the selection of the contractor. Although the concept of the total cost comes into play, other factors such as reliability, quality, flexibility and timing, are considered in the procurement process. Our approach was to compute a proposal quality index to short list them for the final selection by the panel of experts. Pairwise comparisons allow us to express preferences more easily. These preferences can be highly subjective (e.g., likes/dislikes).

Pairwise comparisons were most likely used even before numbers were invented. We can easily envision that "weighting" took place during the Stone Age to decide if a fish, in one hand, can be bartered for a bird in the other hand. A comprehensive introduction to pairwise comparisons is in [9]. It includes mathematical basics and can be downloaded for free from http://arxiv.org/abs/1307.6272 repository. It is also included in [4–6,8].

The distance-based inconsistency was introduced in 1993 (see [7]) and independently analysed in 2008 (see [2]). It was validated by numerous applications and Monte Carlo experimentations.

2 The project assessment model

The mine hazard model was proposed in [1,3] for abandoned mines. It has required data collection about these mines often located in the wilderness. The challenge is to select the best proposal, i.e., to choose the most appropriate projects to do a certain task. In our case, it was for the assessment of abandoned mines in a Canadian mining district. We deliberately avoid using "the best" for the selected bid since it calls for the definition of being "the best". Certainly, it indicates "the highest quality of any kind" but within a reason.

About fifty companies submitted their proposals. The information given in the proposals was studied to select the winner, i.e., the company which meets criteria and expectations set by the expert panel called to conduct the selection. For this, we needed to assign a "goodness value" to each proposal and rank all of them using such a value.

We started with a certain number of criteria, proposed by the expert panel from the government agency, which each company should meet to be qualified for the contract. The main problem is to prioritize, i.e., find the contribution of each criterion to the overall score. After a certain number of meetings and discussions with the government agency representatives, the hierarchy (depicted by Fig. 1) was designed.

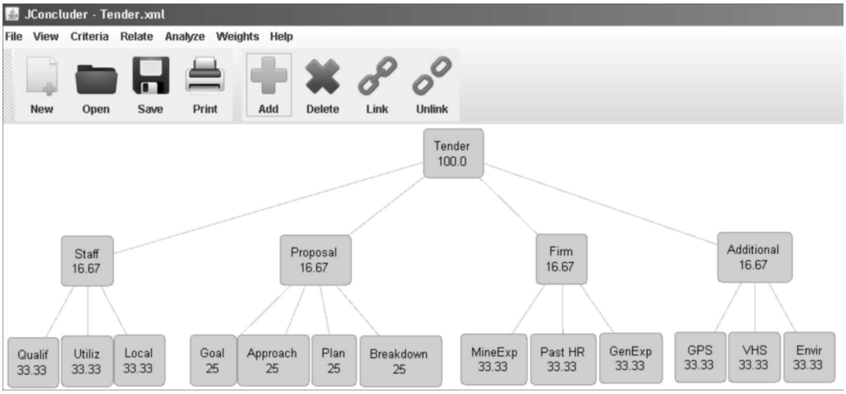

Figure 1. Tendering process model

The lowest hierarchy level in Fig. 1 consists of thirteen criteria. Subsequently, we would need to assign the priority (the weight) to each criterion by comparing them in pairs. However, the number of all pairs criteria is 78 $(13 * (13 - 1)/2)$ and too large for comparing them. For this reason, we cluster similar criteria to build the next higher level of the hierarchy. Groups of criteria are compared in pairs. We divided the criteria into four groups:

1. "staff" (to be assigned by the firm to the project, Fig. 2),
2. "proposal" (how the firm is going to manage the task, Tab. 2),
3. "firm" (experience and reputation of the firm),
4. "additional qualifications" (Tab. 3).

Next, the elements in a level are compared against all others in the same level, with respect to the elements on the preceding level. These comparisons are arranged at the table (matrix). Table 1 shows a rating scale assumed for pairwise comparisons. After a number of experiments, the scale in Tab. 1 seems to be the most effective.

Table 1. Rating scale

Rating scale	meaning
1	equal or uncertain
2	moderate
3	extreme
1.5, 2.5, 2.8	itermediate values
1/x	reciprocal of x

The matrix of pairwise comparisons is constructed by comparing criteria in pairs according to their relative importance. Say that "staff" is moderately more important (according to the expert panel subjective assessment) than "proposal", then we would place a "3" in the entry (row-staff, column proposal). Say, that we assess the group staff to be by far more important than group firm, we place a "5" in the entry (row: "staff" and column: "firm"). A similar pairwise comparisons process takes place for all other entries. The system requirement is to fill in only the "upper matrix triangle" (above the main diagonal) of the matrix since the lower triangle is reciprocal ($a_{ij} = 1/a_{ji}$). After the expert panel entered its pairwise comparisons, the final results were computed as the geometric means of rows and normalized to 1.

Table 2. Group 3

Criterion		A	B	C	D	E
Cost break-down	A	1	1	2	4	5
Goal under-standing	B	*	1	1	5	7
Approach	C	*	*	1	7	7
Work plan schedule	D	*	*	*	1	5
Final report design	E	*	*	*	*	1

Once the matrices in each level are completed, the relative importance of the elements in the level is given by geometric means of rows in the PC matrix. The weights are normalized so the sum is equal to 100%. Thus, we can effectively attribute a percentage of priority to each factor at each level. The priority of each criterion at the lowest (third) level is obtained by multiplying the priority of the criterion under the proper group by the priority of this group. The priorities of all criteria sum up to 1. The computed weights are presented in Tab. 4.

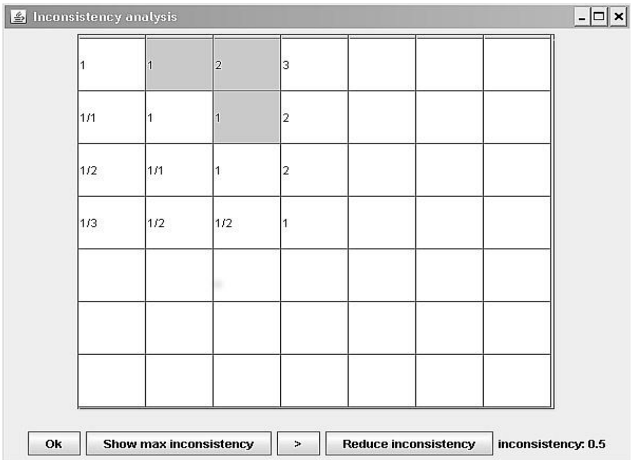

Figure 2. PC matrix for “staff”

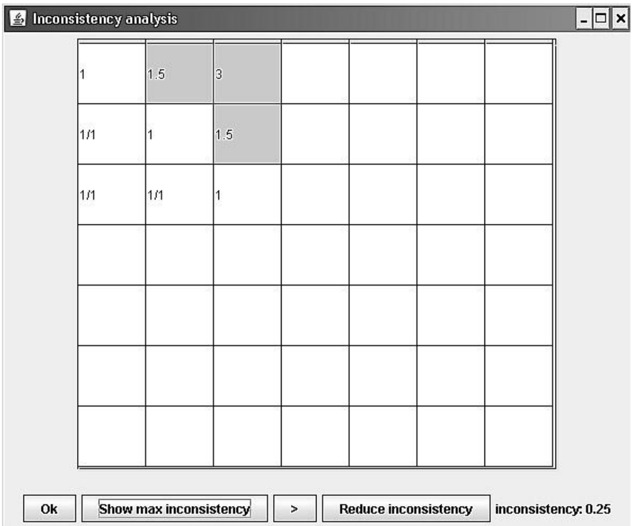

Figure 3. PC matrix for “proposal”

Table 3. Qualification

Criterion		A	B	C
GPS experience	A	1	1	2
Envirn. Knowledge	B	*	1	1
VHS/35mm camera exper.	C	*	*	1

Table 4. Results

Project = 1			
Staff=0.568	Proposal=0.252	Firm=0.120	Additional=0.059
Qual=0.649	CostDistr=0.310	PastPr=0.550	GPS=0.655
Expl=0.279	Appr=0.287	MineEx=0.368	Envir=0.290
Person=0.072	Underst=0.286	GenExp=0.082	VHS=0.055
	Sched=0.081		
	Rep=0.360		

The expert panel evaluated each company in regards to the assumed criterion. They were using the scale from 0 to 7. According to the experts' assessments and weights computed by the system the contribution of the first six criteria to the final score was equal to 81.7%. We recommended to start the assessment of all proposals using these six criteria. The final results are given in Table 5. In the first line, we have the weight (priority) of each criterion (in percentage). The first column consists of the total score of the company whose number is given in the second column. The assessment of the company (in the scale from 0 to 7). The last column, named "cost" consists of the costs given by the company. It has been concealed for confidentiality reason.

The general perception was that the estimation process was not only considerably shorter, but more accurate, since the existing panel of experts concentrated its attention on the most important factors of the proposals.

Figure 4 shows the assessment results. "Missing value" entries in Fig. 4 have a meaning of "it does not matter" since it was evident for the assessment committee that only the two top contenders had realistic chance of winning the competition.

Weights		37	15.9	7.8	7.2	7.2	6.6	4.4	4.1	3.9	2	1.7	1	0.9	0.3	Subtot
Total	#	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	
71.26	16	6.5	6	5.5	6.5	6.5	4	4	7	4.5	3.5	4	5	5	6.5	12.7
70.03	21	6	6	6	6.5	5.5	6	5.5	5	0.5	4.5	5.5	5	5	4	10.8
66.26	22	6	5	6	6	5	5.5									0
65.74	50	6	6	4	5.5	4.5	6									0
62.96	36	6	5	4	4	5.5	6									0
62.74	35	5.5	5.5	6	4.5	5	5									0
62.66	13	6	3	5.5	6	6	6									0
62.36	37	6	5	3	4.5	5.5	6									0
62.33	49	5.5	6	5	5.5	4.5	4									0
61.84	4	6	4	4.5	6	5	5									0
60.44	41	5.5	4	6.5	5	5.5	4.5									0
58.04	44	5.5	4	4	5	4.5	6									0
56.54	3	5.5	3	6	4	5	5									0
55.71	31	4.5	5	6	4	4	6									0
54.50	5	5	5	5	4	5	2									0
54.41	14	5	4	3.5	5	5	5									0
53.01	48	5.5	2.5	2.5	4.5	5.5	5.5									0
52.37	47	4.5	4	4.5	4	5.5	5									0
51.66	27	4	4	4	5.5	5.5	6									0
42.44	46	3.5	3.5	4.5	4	3	4									0
42.11	38	3.5	3	3	4.5	4	5									0
40.18	29	4.5	2.5	3	2	1.5	4									0
39.68	25	4	4.5	2	2	3	1									0
32.76	32	3	2.5	2	3	3	3									0
30.99	24	2	2.5	5	4	4	1									0
29.97	17	2.5	1	4.5	2.5	3.5	3.5									0
29.52	2	2.5	2.5	2	2.5	2	4									0
27.87	11	2.5	2	2	2	2	4									0
26.77	6	3.5	1	2	0	0	4									0
25.43	28	2.5	2	1.5	2	2	2									0
17.42	30	1.5	1.5	1	1.5	1.5	2									0
1.11	23	0	0	1	0	0	0									0
Average		4.38	3.61	3.91	3.94	4	4.25									

Figure 4. Assessment results

Table 5. Computed weights

#	Criteria	R-total
1	Relevant qualification of personnel to be assigned to proj.	37.0
2	Explanation of use of personnel	52.9
3	Costs breakdown	60.7
4	Approach	67.9
5	Goal understanding	75.1
6	Past projects	81.7
7	Experience in mine auditing, hazard identification	86.1
8	Local knowledge of personnel	90.2
9	Experience with GPS system	94.1
10	Work plan schedule	96.1
11	Environmental knowledge and training	97.8
12	General experience	98.8
13	Proposed format of final report	99.7
14	Experience with 35mm cameras and VHS video	100.0

Proposals 22 and 50 were too expensive and the difference in the point on criteria C1 to C6 between the winner and project 36 was too big to consider it as a potential winner. The companies #16 and #21 obtained the best scores: 71.26%, 70.03% (81.7% was the highest possible result according to the first six criteria). The remaining criteria (#7 to #14) confirmed the good position of these two companies. The final choice in favour of the company #21 was the cost they proposed for their services. We can see that criteria from #7 to #14 have turned out to be of lower significance, with each of them below 5% and some even below 1%, which could be treated (from the procurement assessment perspective) as “information noise”. In it not really a side product, but important results as such criteria should be dropped from the assessment model in the future. We have included it for illustration purpose.

3 Approximation of pairwise comparison matrices

Starting with [2], a distance-based adjective has been used by other researchers for the new inconsistency defined in 1993 in [7]. The distance-based adjective reflects the nature of the inconsistency indicator, which is defined as a minimal distance from the nearest consistent triad in matrix A. Matrix A is defined as:

$$A = \begin{pmatrix} 1 & a_{12} & \cdots & a_{1n} \\ \frac{1}{a_{12}} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{a_{1n}} & \frac{1}{a_{1n}} & \cdots & 1 \end{pmatrix}$$

In data and knowledge processing, this is expressed by the popular adage: GIGO (garbage in – garbage out). GIGO summarizes well what has been known for a long time: processing “dirty data” cannot guarantee meaningful results. The distance-based inconsistency allows decision makers to localize the most inconsistent triad (or triads) in their own assessments. It is expressed by the maximum of all triads a_{ik},a_{kj},a_{ij} of elements of A (say, with all indexes i,j,k distinct) of their inconsistency indicators, which in turn are defined as:

$$ii = \min(|1 - \frac{a_{ij}}{a_{ik}a_{kj}}|, |1 - \frac{a_{ik}a_{kj}}{a_{ij}}|) \tag{1}$$

The process of reducing global inconsistency of a pairwise comparisons matrix (PC matrix), is based on the detection of triads (say, $\{a_{ik},a_{kj},a_{ij}\}$) with the maximal inconsistency. When such a triad is located, we modify the value of a_{ik}, a_{kj} or a_{ij} in order to make the replaced triad fully consistent. It is shown by Fig. 2 and 3.

4 The inconsistency reduction

The initial PC matrix is not expected to be fully consistent. Solving real-life problems usually involves inconsistent assessments. However, a matrix with a large inconsistency is undesirable according to “the garbage in, garbage out (GIGO)” principle. Inconsistencies often reflect assessing “every criterion being more important than another”.

For an inconsistency to occur, a minimum size of 3 for PC matrix is required, since at least one triad needs to exits. Needless to say that for two comparisons, inaccuracy (not inconsistency) takes place. We use n = 7 as the maximal PC matrix size. For a matrix with n elements, there are $n*(n-1)/2$ comparisons. It gives us 21 comparisons for n = 7 and it is what most respondents can usually tolerate (we wonder who would agree to compare 100 objects giving 4,950 pair combinations).

5 The inconsistency reduction

This study proposes an improvement to the government procurement using pairwise comparisons. Such an approach has been already used for the assessment of research entities by the Ministerial Commission (the Polish acronym KEJN). The Ministry of Education Bill, dated 2012-07-13, set the criteria and algorithm for categorizing scientific entities. The original Polish text from [10] is depicted by Fig. 5.

(Dz.U. 2012 nr 0 poz. 877):

4. Na podstawie ocen, o których mowa w ust. 1, komisja Komitetu ustala ostateczną ocenę jednostki naukowej w danej GWO, **przy zastosowaniu metody porównań parami**, *wykorzystującej ważoną relację przewyższania, według algorytmu określonego w załączniku nr 8 do rozporządzenia.*

Figure 5. The Polish text of the relevant Ministry Bill fragment

The authors' interpretation (not legally binding hence) of Fig. 5 is as follows:

4. *Using assessments set in the part 1 of the Bill, The Ministry Commission sets the category of the scientific entity in a given region by using the pairwise comparisons method. The algorithm for the method is in the Appendix 8.*

The pairwise comparisons method has been explicitly used five times in the Ministry Bill.

6 Conclusions

The pairwise comparisons method contributed to shorter processing time of the proposal assessment. In Canada, the government procurement system has been in place for a longer period of time than in Poland. Its perfection has never stopped in Canada, which is why it has been possible for our model to be used. The pairwise comparisons method contributed to shorter processing time of the proposal assessment. It provided efficiency, confidence and fairness to the tendering process.

Acknowledgement

Authors would like to thank Emalie Hendel, a philosophy student at Laurentian University, for her editorial improvements of our text.

References

1. Bolger P. M., Duszak Z., Koczkodaj W. W., Mackasey W. O, (1993). Ontario Abandoned Mine Hazards Prioritizing - an Expert System Approach. In: *Proceedings of the 15th Annual Abandoned Mine Land Conference*, Jackson, Wyoming, 370-388.
2. Bozoki S., Rapcsak T., (2008). On Saaty's and Koczkodaj's inconsistencies of pairwise comparison matrices, J. of Global Optimization, 42(2): 157-175.
3. Duszak Z., Koczkodaj W. W., Mackasey W. O., (1993). Towards Better Abandoned Mine Hazard Prioritizing - An Expert System Approach, in The Challenge of Integrating Diverse Perspectives in Reclamation, In: *Proceedings of the 10th National Meeting of ASSMR*, Spokane, Washington, 577-589.
4. Fulop J., Koczkodaj W.W., Szarek S., (2010). A different perspective on a scale for pairwise comparisons, *Transactions on computational collective intelligence I*, 71-84.
5. Grzybowski A. Z., (2012). Note on a new optimization based approach for estimating priorityweights and related consistency index, *Expert Systems with Applications*, 39(14): 11699-11708.
6. Herman, M., Koczkodaj, W. W., 1996. Monte Carlo Study of Pairwise Comparisons, *Information Processing Letters*, 57(1): 25-29.

7. Koczkodaj, W. W., (1993). A New Definition of Consistency of Pairwise Comparisons. *Mathematical and Computer Modelling*, 18(7): 79-84.
8. Koczkodaj, W. W., (1996). Statistically Accurate Evidence of Improved Error Rate by Pairwise Comparisons. *Perceptual and Motor Skills*, 82: 43-48.
9. Koczkodaj W.W., Szwarc, R., (2013). On Axiomatization of Inconsistency Indicators for Pairwise Comparisons, (accepted for a journal publication; also stored in arXiv.org/abs/1307.6272 for the free downloading).
10. Ministry Bill, Dz.U. 2012 nr 0 poz. 877, www.uzp.gov.pl/cmsws/page/?F;239; Parliament web page isap.sejm.gov.pl/DetailsServlet?id = WDU20120000877
11. New EU-rules on public procurement - ensuring better value for money (Supplement to the Official Journal - Public procurement notices), *Official Journal of the European Union*, Reference No : 20140110BKG32432, Eur − lex.europa.eu, 2014.
12. Public Tendering, (2004). *Journal of Laws in the Republic of Poland*, 19(177).
13. The Office of Federal Procurement Policy Act 41 U.S.C. 401, 1974.