sciendo
Control and Cybernetics
vol. **50** (2021) No. 3
pages: 413-422
DOI: 10.2478/candc-2021-0025

On the use of data analysis and OR modelling in MCDM problems: a case analysis – a rejoinder to the paper by Sousa Ribeiro et al.*

by

Jan W. Owsiński

Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warszawa, Poland
owsinski@ibspan.waw.pl

Abstract: Taking as an example the very interesting and motivating paper by Sousa Ribeiro et al. (2021) an attempt is made of providing a couple of insights into the decision making process from the point of view of the potentially helpful aspects of data analysis and OR-related modelling. These are just hints and suggestions, meant primarily to emphasise the multifaceted character of the decision making situations and processes, especially when concerning more complex issues. While the course of the procedure proposed and exemplified in Sousa Ribeiro et al. (2021) is treated as fully correctly and successfully carried out to the end, we wish to show the potential use of information, constituting in a sense a "by-product" of such a procedure, or, actually, of any similar procedure, aimed at supporting decision making.

Keywords: location problem, variables, criteria, objectives, modelling

1. Introduction

This short note was stimulated by a very interesting and – exactly – stimulating paper by Sousa Ribeiro et al. (2021), which appeared in one of preceding issues of this journal. The paper dealt with the decision making process in the presence of multiple criteria (MCDM – multiple criteria decision making), and was primarily meant to highlight the advantages of using the FITradeoff methodology, developed, anyway, by the very same authors, in implementing a concrete MCDM process against a well documented concrete decision problem.

The problem was the one of locating a shopping mall in one of the cities of northeastern Brazil, this problem boiling, actually, down to the choice of

*Submitted: July 2021; Accepted: September 2021

one of the preselected towns. The data, used in this example by the Authors of the paper, are shown in Table 1. The variables presented were treated as criteria, meaning that some of them were being maximized (*car fleet*, *GDP* and *population*), while the other ones – minimized (*number of competitors*, *electricity price*, *Gini index*, and *land price*) in the search for the best location. It is vis a vis this set of criteria and concrete data for the potential locations of a shopping mall that the illustration was provided for the procedure, implemented with the use of the FITradeoff methodology and the respective software.

Taking this illustration as an example, and not referring at all to the FITradeoff methodology, I wish to make a number of remarks and suggestions, which, I think, are pertinent to many, if not the majority (the totality being put apart) of the multicriteria decision situations.

Two points ought to be particularly stressed with respect to the present note:

1. It is obvious that any practical MCDM procedure has to be executed according to a definite set of rules and assumptions, for which it is known (demonstrated analytically or empirically) that they ensure fulfilment of definite formal and pragmatic stipulations. Thus, when executing a procedure, one is guided by the adopted rules and assumptions, even if there exists a multiplicity of potential other pathways of proceeding, especially in view of the concrete circumstances that may arise in a given concrete decision situation.
2. At the same time, in the actual realization of any MCDM procedure the respective data appear, necessary for this realization, which not only simply serve to go on with the procedure, but represent a value in themselves. These data, actually, characterize the given decision problem and the perspectives on it. They are used to carry out the designed procedure, but they might also be used for a broader purpose, or for an additional analysis, even if this analysis does not change the outcome from the designed procedure.

This, exactly, is the situation with the study and data reported in Sousa Ribeiro et al. (2021). Side by side with effective performance of the procedure, a set of data is obtained, which also has a definite value in itself.

2. A cursory analysis

This example is a representative of a vast range of situations, in which the multiple criteria are meant to altogether represent a kind of direct, perceived or indirect and expected "benefit" to the (ultimate) decision maker ("the real patron", call it DM). This is also very clear in the here considered case:

The corporation, planning to locate a shopping mall in a definite city, wishes to gain from the decision a possibly high "profit". This means: a possibly high volume of sales and a possibly low cost. That is why the variables are included,

Table 1. The data table

	Criteria						
Cities	Car fleet	Competitors	Electricity price	GDP	Gini index	Land price	Population
Arapiraca	95,312	2	1.124	4,100,974.72	0.5589	223.49	214,006
Caruaru	150,462	3	1.158	6,877,208.88	0.5422	405.73	314,912
Imperatriz	134,918	2	1.400	6,599,566.71	0.5612	387.02	247,505
Juazeiro	89,887	2	1.207	3,700,881.80	0.5723	286	197,965
Juazeiro do Norte	112,091	3	1.114	4,427,525.37	0.5488	226.07	249,939
Mossoró	137,019	2	1.082	6,166,118.90	0.5340	300.12	259,815
Nossa Senhora do Socorro	62,793	3	1.114	2,597,290	0.4980	212.48	160,827
Parnaíba	83,475	2	1.198	2,037,540.02	0.5772	216.51	145,705
Santa Rita	34,962	1	1.052	2,222,358.73	0.4760	199.84	120,310
Vitória da Conquista	122,605	3	1.207	6,482,662.68	0.5588	238.36	306,866

Source: Sousa Ribeiro et al. (2021)

meant to represent (reflect) the potential turnover, like *car fleet*, *GDP* value, and *population* number. The bigger the values of these three variables, which might be referred to as "volume" characteristics, the higher the expected turnover. On the other hand – the DM expects the turnover to decrease and the costs to increase with the bigger numbers, characterizing the criteria of the number of *competitors*, *electricity price*, *Gini index*, and *land price*.

In an obvious manner, all these, even if rightly selected, have quite a different influence on the potential real turnover and costs, i.e. the supposed "benefit" to the DM. This applies to such characteristics, linking the variables (criteria) and the "benefit", as the strength (weight) of influence, the timing of this influence (e.g. at the time of construction or during day-to-day management, i.e. one-year or ten-year time horizon...), and also the certainty of influence.

In cases like this it is, of course, out of question to use the classical standard location theory and location models, which, even if quite elaborate and effective, account for a much narrower range of factors and their characteristics than it is possible to consider (even if only implicitly) in a MCDM procedure with the active participation of the persons "in charge". The present author is fully aware of this, having done some work on location analysis (see Owsiński, 2014), and, more generally, on the logistic systems (see, e.g., Mażbic-Kulma et al., 2021).

The analysis, provided in Sousa Ribeiro et al. (2021), here represented merely by Table 1, concerns, of course, only an excerpt from the usual complete decision making process, which forms the framework for the here quoted MCDM task. A number of issues are taken up and decided upon in this process, and I shall comment here just on some of these, treating them as potential or hypothetical, but starting from the data and the formulation as reported here.

3. Some obvious hints

The framework, mentioned in the last paragraph of the preceding section, includes, of course, variable selection, done in the context of understanding of the relation between the variables considered and the perception of the "benefit". Table 2 shows the values of simple Pearson correlation coefficients, calculated on the basis of Table 1 (notwithstanding all of the quite naturally arising reservations).

These results appear largely obvious. It could be expected that the number of cars, the GDP value and the population number are quite tightly connected. In fact, were it just for the "meaning" of these variables, it would have sufficed to use just one of them. But, of course, the effect on decision is different, when all three are used, and this is, apparently, what the DM wished to have.

On the top of this, it may justly be argued that in terms of specification of preferences by the DM, concerning the objects, described by a set of criteria, there is no sense of speaking of "correlations", especially in the sense of the "preference space". While this is certainly true, there is, beyond any doubt,

Table 2. Correlation coefficient values for the data from Table 1

	Competitors	Electricity price	GDP	Gini index	Land price	Population
Car fleet	0.483	0.426	**0.920**	0.548	0.766	**0.916**
Competitors		0.074	0.399	0.221	0.165	0.615
Electricity price			0.407	0.593	0.560	0.271
GDP				0.318	0.762	**0.941**
Gini index					0.285	0.398
Land price						0.625

the objective image of the situation, in which such correlations simply exist, and bear influence on the final outcome. We shall yet comment on this in the context of the potential "model of the decision situation (of the benefit)".

Against this background it is interesting to note that (again, formal reservations put apart) the least correlated variable is that of the number of competitors. Yet, also this variable is in itself most strongly associated with population number, followed by the car fleet and GDP... It appears that the location decisions of the competitors had been also largely influenced by these "volume" variables, with other factors having played a role, as well. Then, the fact that no negative correlation is observed for land and electricity prices, nor for the Gini index, only demonstrates that the influence of the "volume" variable(s) was altogether prevalent.

With this we go back to the framework in terms of a hypothetical (or implicit) model of the "benefit". Such a model, in quite a schematic rendition, might look like

$$B = f(D, C) \tag{1}$$

where B denotes the "benefit", D is effective (potentially realized) demand, and C is lumped cost of operation. We do not make explicit here the time horizon, for (1) holds for any time horizon selected, although the concrete formulae would be quite different. Formula (1) may also be understood as representing not just the concrete values, but rather, as this is the case here, forming the basis for a ranking.

A simple question is how, actually, the variables here considered, relate to this model? Obviously, the ranking of the potential locations with respect to C is represented through electricity and land prices. However, the effective demand would, it seems, be effectvely represented by the ratio of the "volume" variables accounted for and the number of the already operating competitors. This is, for instance, expressed through Table 3.

Of particular interest seems to be the range of values of the new variable of GDP/competitors, with the ratio between the biggest (3 299 783 for Imperatriz) and the smallest (865 763 for Nossa Senhora do Socorro) values being close to 4 (more precisely: 3.8). This is, by far, the largest scope of the variable values obtained in this exercise. It is also telling, and, indeed, worth a thought, that Nossa Senhora do Socorro features distinctly the lowest values of both this variable and Population/competitors. At the same time, GDP/capita is among the lowest ones there, similarly as the Gini index. A question that comes to mind is: why did the competitors locate as many as three facilities there? Have they, after all, used a different set of criteria (in the light of the previous remarks: in addition to the "volume" variables)? If so – what they could have been?

It is, definitely, worthwhile to ponder on this. If the "models of benefit" are different for different competitors, this can be due to a number of reasons, but, given that the data we consider here form an objective image of reality, the main reason is the difference of corporate strategies, perhaps coupled with the capacities of the individual companies.

On the other hand, the position of Imperatriz appears to be strengthened by a very high value of Population/competitors, and, quite significantly – the highest GDP per capita! Similarly, these new variables push up the position of Mossoro.

These considerations are complemented with Table 4, in which correlation coefficients for the newly introduced variables are shown. Some of the entries are obvious, like the mechanically appearing negative correlations for the number of competitors, or the very high coefficients at the bottom of the table, and, generally, it brings little new information altogether, rather confirming the observations already made, especially regarding the association of the "volume" variables and some of the other ones (like prices). Yet, it is interesting to note that the Gini index seems to be uncorrelated with the GDP per capita, which may be considered a significant information, primarily from the point of view of the "demand generation" model assumed.

In this manner, based on the same data as those given originally, one might gain a very interesting and revealing insight into the qualities of the potential locations, aiding significantly in the making of the respective decisions.

4. A potential OR / optimization oriented extension

To close this short note a comment is forwarded, broadening somewhat the scope of the problem here considered. It does not apply so much to the data / criteria side of the problem, as to the very basic formulation of it. Namely, the formulation from Sousa Ribeiro et al. (2021) implies, definitely, that we deal with a *binary decision*: to locate or not to locate (and that in only one place!).

Table 3. Characterisations of the locations with three additional variables

	Original variables							Added variables		
Cities	Car fleet	Competitors	Electricity price	GDP	Gini index	Land price	Population	Population / competitors	GDP/ competitors	GDP per capita
	1	2	3	4	5	6	7	8	9	10
Arapiraca	95312	2	1,124	4100975	0.5589	223.49	214006	107003	2050487	19.16
Caruaru	150462	3	1,158	6877209	0.5422	405.73	314912	104971	2292403	21.84
Imperatriz	134918	2	1,400	6599567	0.5612	387.02	247505	123753	3299783	26.66
Juazeiro	89887	2	1,207	3700882	0.5723	286.00	197965	98983	1850441	18.69
Juazeiro do Norte	112091	3	1,114	4427525	0.5488	226.07	249919	83306	1475842	17.72
Mossoro	137019	2	1,082	6166119	0.534	300.12	259815	129908	3083059	23.73
Nossa Senhora do Socorro	62793	3	1,114	2597290	0.498	212.48	160827	53609	865763	16.15
Parnaiba	83475	2	1,198	2037540	0.5772	216.51	145705	72853	1018770	13.98
Santa Rita	34962	1	1,052	2222359	0.476	199.84	120310	120310	2222359	18.47
Vitoria da Conquista	122605	3	1,207	6482663	0.5588	238.36	306866	102289	2160888	21.13

Table 4. Correlation coefficients with the new variables (variable numbers as provided in Table 3)

Variable numbers	New variables		
	8: population / competitors	9: GDP / competitors	10: GDP / capita
1	0.371	0.556	0.686
2	-0.519	-0.319	-0.040
3	0.136	0.344	0.477
4	0.550	0.726	0.854
5	-0.003	0.053	0.100
6	0.463	0.648	0.759
7	0.334	0.482	0.637
8		**0.938**	0.787
9			**0.946**

This kind of decision problem formulation is well justified in a very preliminary stage of the decision process, in which we wish to select one (or very few) “candidate locations” and then only to perform a more detailed analysis of both the conditions for and the nature of investment. One could, therefore, try to design or project the potential further steps in the procedure. It is, namely, rather obvious that in different cities the concrete local conditions for setting up a shopping mall would be different (exact location in the city, surface available, infrastructure, etc.). Then, even if the corporation in question has a sort of “typical design” for a shopping mall, there is always a definite possibility (and usually a must) of scaling. In some cases one can even speak of a range of scales of the located facility (e.g. “large”, “medium-sized”, and “small”).

It is true that the data at hand hardly allow for any sensible reasoning, concerning this additional dimension to the problem, but, if, say, the variables, related to cost, especially concerning construction, play an important role in formula (1), and it is expected that the respective relations are not necessarily linear, then we can imagine that, in fact, formula (1) takes the shape of a matrix of $B = \{b_{ij}\}$, where b_{ij} is the benefit from location i for the j^{th} version of the project:

$$b_{ij} = f(d_{ij},\ c_{ij}) \tag{2}$$

where $i \in I = \{\text{Arapiraca},\ldots,\text{Vitoria da Conquista}\}$, and $j \in J_i = \{1,,,.j^{max}(i)\}$, $j^{max}(i)$ being the maximum number of shopping mall project versions, possibly considered for the city i. Such a formulation would require not only having the function $f(.)$, given in some way (even if “only” an expert opinion), but also the assessments of d_{ij} and c_{ij} for the particular options of the shopping mall versus concrete location.

This formulation is in a natural manner extended through the consideration of the choice of location(s) for a definite intended volume of investment. Actually, selecting multiple locations might be even a better solution from the point of view of the aspect until now not considered at all, namely: risk / security of the potential benefit.

Yet, as mentioned already at the beginning, solving of such a problem, potentially leading to a much better solution than just the “binary” formulation, requires having at disposal much more data, most of which can be highly uncertain (even if based on a kind of experience), on both demand and cost sides. That is why the simple, even as simple as “binary”, decision, based on well-rooted data and formally justified procedure, may be a viable option, which was taken in the case here considered.

5. Some conclusions

Decisions are often made in conditions of shortage of relevant knowledge and data, lacking adequate formulation of the objective(s), model of the situation and the object(s), as well as data on the characteristics of the decision alternatives. In such – actually, usual – cases one refers to the proxy (even if correct and precise) data and to the human assessments, concerning the essential evaluations (preferences, in particular). Yet, even if we carry out the procedures in this kind of a manner, there is often a wide margin of different ways of using the data available and formulating the essential questions. This margin ought to be exploited to the maximum degree feasible.

The DM, who actually determines the course of the respective procedure by assigning values or preferences, or by other means, even if s/he considers the result obtained from the procedure to be correct and satisfying, might be interested in additional pieces of information that can be obtained from the same set of data as that used in the procedure.

References

Mażbic-Kulma B., Owsiński J. W., Stańczak J., Barski A., Sęp K. (2021) Mathematical conditions for profitability of simple logistic system transformation to the hub-and-spoke structure. In: Atanassov K. T. et al., eds., *Uncertainty and Imprecision in Decision Making and Decision Support: New Challenges, Solutions and Perspectives. Selected Papers from BOS-2018, held on September 24-26, 2018, and IWIFSGN-2018, held on September 27-28, 2018 in Warsaw, Poland*. Springer, 398-408.

Owsiński, J. W. (2014) Bi-partial version of the p-median / p-center facility location problem and some algorithmic considerations. *JAMRIS*, 3, 59-63. DOI 10.14313/JAMRIS_3-2014/28

Sousa Ribeiro, M. L., Peixoto Roselli, L. R., Asfora Frej E., de Almeida, A. T. and Costa Morais, D. (2021) Using the FITradeoff method to solve a shopping mall location problem in the northeast countryside of Brazil. *Control and Cybernetics*, **50** (1), 109-126.