Aggregation of incomplete preference rankings: Robustness analysis of the ZMII-technique

Aggregation of incomplete preference rankings: robustness analysis of the ZM_{II}-technique

Fiorenzo Franceschini[1] and Domenico Maisano[2]

[1] *fiorenzo.franceschini@polito.it* [2] *domenico.maisano@polito.it*

Politecnico di Torino, DIGEP (Department of Management and Production Engineering),
Corso Duca degli Abruzzi 24, 10129, Torino (Italy)

ABSTRACT

A common group decision-making problem is that in which: (i) several *judges* express their *subjective* preference rankings regarding some *objects* of interest and (ii) these rankings should then be aggregated into a *collective* judgment. The authors recently developed an aggregation technique – denominated "ZM_{II}" – aggregating these rankings into a *ratio* scaling of the objects, which represents the solution to the decision-making problem of interest. This technique also includes a flexible *response mode*, which tolerates *incomplete* rankings and can therefore be adapted to various practical contexts, such as quality improvement activities, field surveys, product-comparison surveys, etc..

The aim of this article is proposing an original approach to verify the robustness of the ZM_{II}-technique under the influence of various factors, especially those concerned with the degree of "completeness" of preference rankings (e.g., number of objects identified by judges, whether these objects are ordered or not, etc.). The methodology in use relies on the simulation of several thousand decision-making problems, in order to organically study the effect of the factors of interest. Results show a certain robustness of the ZM_{II}-technique, even under relatively "unfavourable" practical conditions, characterized by very incomplete preference rankings. Description is supported by instructive examples.

Keywords: Quality improvement, Thurstone's *Law of Comparative Judgment*, Incomplete preference ranking, *Generalized least squares*, ZM_{II}-technique, Robustness analysis, Factorial analysis, Degree of agreement, Degree of completeness.

INTRODUCTION

A general group-decision problem – which is transversal to many scientific disciplines, such as Design/Manufacturing Decision Making, Quality Improvement, Quality Management, etc. – is structured as follows (Keeney and Raiffa, 1993; Das and Mukherjee, 2007; Wang et al., 2017; Franceschini and Maisano, 2019a):

- some *objects* (o_1, o_2, …) should be evaluated and compared on the basis of a certain *attribute*;
- some *judges* (j_1, j_2, …) make their *subjective* judgements on these objects. In addition, judges may refrain from evaluating a subset of the objects because of practical impediment or lack of adequate knowledge;
- The ultimate goal is to aggregate the judges' judgements into a single *collective judgment*.

Numerous aggregation techniques have been developed so far, which differ in: (i) the *response mode* through which subjective judgments are expressed (e.g., ratings, rankings, paired-comparison relationships, etc.), (ii) the type of *aggregation model* (e.g., heuristic, statistical, fuzzy models, etc.), and (iii) the nature of the *collective judgment* (e.g., object rankings, ordinal/interval/ratio scale values, etc.). For an exhaustive discussion of the existing techniques, we refer the reader to (Arrow, 1963; Nurmi, 1983; Fishburn, 1989; De Vellis, 2016; Franceschini et al., 2017).

A key element for the success of a generic aggregation technique is the simplicity of the response mode (Harzing et al., 2009; Paruolo et al., 2013; Franceschini et al., 2019), for example, some authors showed that comparative judgments of objects (e.g., "o_i is more/less preferred than o_j") are generally simpler and more reliable than judgments in absolute terms (e.g., "the degree of the attribute of o_i is low/intermediate/high") (Harzing et al., 2009; Edwards, 1957).

The authors have recently developed an aggregation technique – denominated "ZM_{II}" – which relies on the postulates and simplifying assumptions of the Thurstone's *Law of Comparative Judgment* (LCJ) (Thurstone, 1927; Edwards, 1957; Franceschini and Maisano, 2018; Franceschini and Maisano, 2020a) and embodies the *Generalized Least Squares* (GLS) method (Kariya and Kurata, 2004; Ross, 2014). This technique includes a relatively versatile response mode that tolerates "incomplete" rankings, i.e., rankings only including the objects at the top/bottom (Franceschini and Maisano, 2019a). For the sake of simplicity, a decision-making problem characterised by this type of rankings will be hereafter referred to as "incomplete ranking problem" or, even more simply, as "incomplete problem". On the other hand, a decision-making problem characterized by "complete" rankings will be hereafter referred to as "complete ranking problem" or, even more simply, as "complete problem".

The flexible response mode of the ZM_{II}-technique makes it adaptable to a variety of practical contexts, where judges do not have the concentration to formulate complete rankings; e.g., problems with a relatively large number of objects, field surveys, product-comparison surveys, customer-satisfaction survey, etc. (Harzing et al., 2009; Chen and Cheng, 2010; Lagerspetz, 2016).

Additionally, the ZM_{II}-technique allows (1) to construct a *ratio* scaling of the objects, which represents the output solution of the decision-making problem of interest, and (2) to estimate the uncertainty of this resulting scaling, by "propagating" the uncertainty of input data (Roberts, 1979; Zhang et al., 2016; Franceschini and Maisano, 2019a).

An important requirement of the ZM_{II}-technique is that, apart from the objects to be evaluated – i.e., o_1, o_2, …, o_n, which will be hereafter classified as "regular" objects – preference rankings also include two fictitious "dummy" objects, i.e., o_Z and o_M, which will be described in the next section[1].

[1] It is because of these two dummy objects that the technique is called "ZM_{II}" (Franceschini and Maisano, 2019a).

This paper aims at proposing an original approach to organically verify the robustness of the ZM_{II}-technique, depending on the "degree of completeness"[2] and other characteristic factors of the (incomplete) decision-making problems. By the term "robustness", it is meant the ability of the ZM_{II}-technique to provide a collective judgment comparable to that which would be obtained if the judges formulated complete rankings. This investigation is an essential step to substantiate the practical convenience and effectiveness of the ZM_{II}-technique in real-world practical contexts.

For the purpose of example, let us consider the two decisional problems in Table 1, both characterized by four judges (j_1 to j_4) formulating their individual rankings of four objects (o_1 to o_4). Objects are in this case alternative design concepts of some pocket projectors, which have to be ranked in terms of *ease of use* (i.e., attribute of interest).

Table 1. Example of two decisional problems (one complete and one incomplete).

	(1) Complete problem	(2) Incomplete problem	
	Rankings	Rankings	Comment
j_1:	$o_1 \succ (o_2 \sim o_3) \succ o_4$	$o_1 \succ \ldots \succ o_4$	Only the object at the top and the one at the bottom of the ranking are identified.
j_2:	$o_1 \succ o_2 \succ (o_3 \sim o_4)$	$o_1 \succ o_2 \succ \ldots$	Only the first two objects at the top of the ranking are identified.
j_3:	$o_2 \succ (o_1 \sim o_3 \sim o_4)$	$o_2 \succ \ldots$	Only the first object at the top of the ranking is identified.
j_4:	$o_1 \sim o_2 \sim o_3 \sim o_4$	$o_1 \sim o_2 \sim o_3$	Object o_4 is not assessed by the judge since it is not well-known.

In the first case, rankings are *complete* while in the second case they are *incomplete*, reflecting an "unfavourable" practical context, in which judges do not have the possibility to formulate complete rankings. It can be seen that the information content of incomplete rankings represents a subset of the information content of complete rankings. For example, o_1 is in the top and o_4 is in the bottom of the incomplete ranking by j_1, but nothing is know about the mutual relationships between the intermediate objects; on the other hand, in the corresponding complete ranking, this information is given: $o_2 \sim o_3$.

Considering that the ZM_{II} aggregation technique can be applied to both the decision-making problems, a question may arise: "*To what extent will the solution of an incomplete problem be distorted (due to the lower information content), compared to that of a corresponding complete problem?*" Or, reversing the question: "*To what extent does the ZM_{II}-technique tolerate incomplete problems, with no significant distortion of the solution compared to corresponding complete problems?*".

From a methodological point of view, a large number of decision-making problems with different "structural" factors (e.g., number of objects, number of judges, etc.) will be simulated; then, the solution of complete problems will be compared with the solutions of a number of corresponding incomplete ones (e.g., problems characterized by rankings with the more/less preferred objects only and/or without the dummy objects, etc.). The solution of a generic complete problem will be used as

[2] This concept will be formalised and clarified below.

a "gold standard" to check the goodness of the solutions of the corresponding incomplete problems. Reversing the perspective, the present study will tell us how far we can go in making preference rankings more and more incomplete – without deteriorating the solution excessively.

The rest of the paper is organized into five sections. The section "Background information" briefly recalls the response mode and the basic principles of the ZM_{II}-technique. The section "Methodology" illustrates the methodology to examine the robustness of the solution of an incomplete problem, with respect to that of a "source" complete problem; special attention will be devoted to the description of a set of (simulated) factorial experiments. The section "Results" presents and discusses the results of these experiments in detail, highlighting their practical consequences; the description is supported by an extensive use of explanatory graphs. The section "Concluding remarks" summarizes the original contributions of this paper and its practical implications, limitations and suggestions for future research. Further details on the results of this research are contained in the section "Appendix".

BACKGROUND INFORMATION

This section briefly recalls the ZM_{II}-technique and provides preparatory information to understand the rest of the paper. The presentation is organized into three subsections concerning: (1) the *response mode*, (2) the *artificial deterioration of complete rankings* (into incomplete ones), and (3) the *rationale of the aggregation technique* in use.

Response mode

A prerequisite of the ZM_{II}-technique is that each judge involved in the problem formulates a preference ranking of the objects – i.e., a sequence of objects in order of preference, with the more preferred ones in the top positions and the less preferred ones in the bottom ones.

Apart from *regular* objects (o_1, o_2, …), judges may also include two (fictitious) *dummy* objects in their rankings: i.e., one (o_Z) corresponding to the *absence* of the attribute of interest, and one (o_M) corresponding to the *maximum-imaginable* degree of the attribute (Franceschini and Maisano, 2020a). When dealing with these dummy objects, two important requirements should be considered:

1. o_Z should be positioned at the bottom of a preference ranking, i.e., there should not be any other object with preference lower than o_Z. In the case the attribute of another object is judged to be absent, that object will be considered indifferent to o_Z and positioned at the same hierarchical level.
2. o_M should be positioned at the top of a preference ranking, i.e., there should not be any other object with preference higher than o_M. In the case the attribute of another object is judged to be

the maximum-imaginable, that object will be considered indifferent to o_M and positioned at the same hierarchical level.

In the best cases, judges formulate complete preference rankings. Borrowing the language from Order Theory, any ranking including all (regular and dummy) objects, according to a hierarchical sequence with relationships of *strict preference* and/or *indifference* only, can be classified as *linear* (Gierz et al., 2003). Unfortunately, the formulation of complete preference rankings may sometimes be problematic (Harzing et al., 2009). To overcome this obstacle, a flexible response mode that tolerates incomplete preference rankings can be adopted. Below is a list of some possible types of incomplete preference rankings.

- Preference rankings including only the more preferred objects (or "*t*-objects", where "*t*" stands for "top") and the less preferred ones (or "*b*-objects", where "*b*" stands for "bottom"); these rankings will be hereafter denominated "Type-*t&b*". The *t* parameter is conventionally defined as the number of regular objects (i.e., excluding the two dummy objects) within the *t*-objects, while the *b* parameter is conventionally defined as the number of regular objects within *b*-objects. In the example in Figure 1(a), the judge merely specifies the two objects at the top and the two objects at the bottom of the ranking, therefore $t = b = 2$.
- Preference rankings including only the more preferred objects (i.e., *t*-objects) among those available; see the example in Figure 1(b), in which $t = 2$. From now on, these rankings will be denominated "Type-*t*".
- Preference rankings not including the two dummy objects (o_Z and o_M), e.g., in the case judges find it difficult to envisage them. These preference rankings will be classified as "quasi-complete" if they include all regular objects; see the example in Figure 1(c).
- Combining the previous three types of incomplete preference rankings, one can obtain Type-*t&b* or Type-*t* preference rankings that do not include the dummy objects.
- To contemplate the fact that judges may not be able to evaluate certain objects – e.g., since they are not familiar with them – preference rankings excluding some objects will also be tolerated (see Figure 1(d)).

The upper part of Figure 1 exemplifies some judges' verbal descriptions from which the previous types of incomplete rankings can be deduced. Figure 1 also shows that a generic incomplete ranking can be transformed into a "reconstructed" ranking, which includes all the (dummy and regular) objects, with the addition of appropriate incomparability relationships. Borrowing the language from Order Theory, these other rankings can be classified as *partial*, i.e., apart from *strict preference* (e.g., "$o_i \succ o_j$") and *indifference* (e.g., "$o_i \sim o_j$") relationships, they may also contain *incomparability* relationships (e.g., "$o_i \parallel o_j$") among the objects (Gierz et al., 2003). E.g., considering Type-*t&b* rankings, the objects that are not considered by judges can be allocated at an intermediate hierarchical level with respect to the *t*- and *b*-objects, with mutual *incomparability* relationships. As for Type-*t* rankings, the objects that are not considered by judges can be allocated at a lower hierarchical level with respect to the *t*-objects. As for the rankings that do not include o_Z and o_M, they can be reconstructed in compliance with the following constraints:

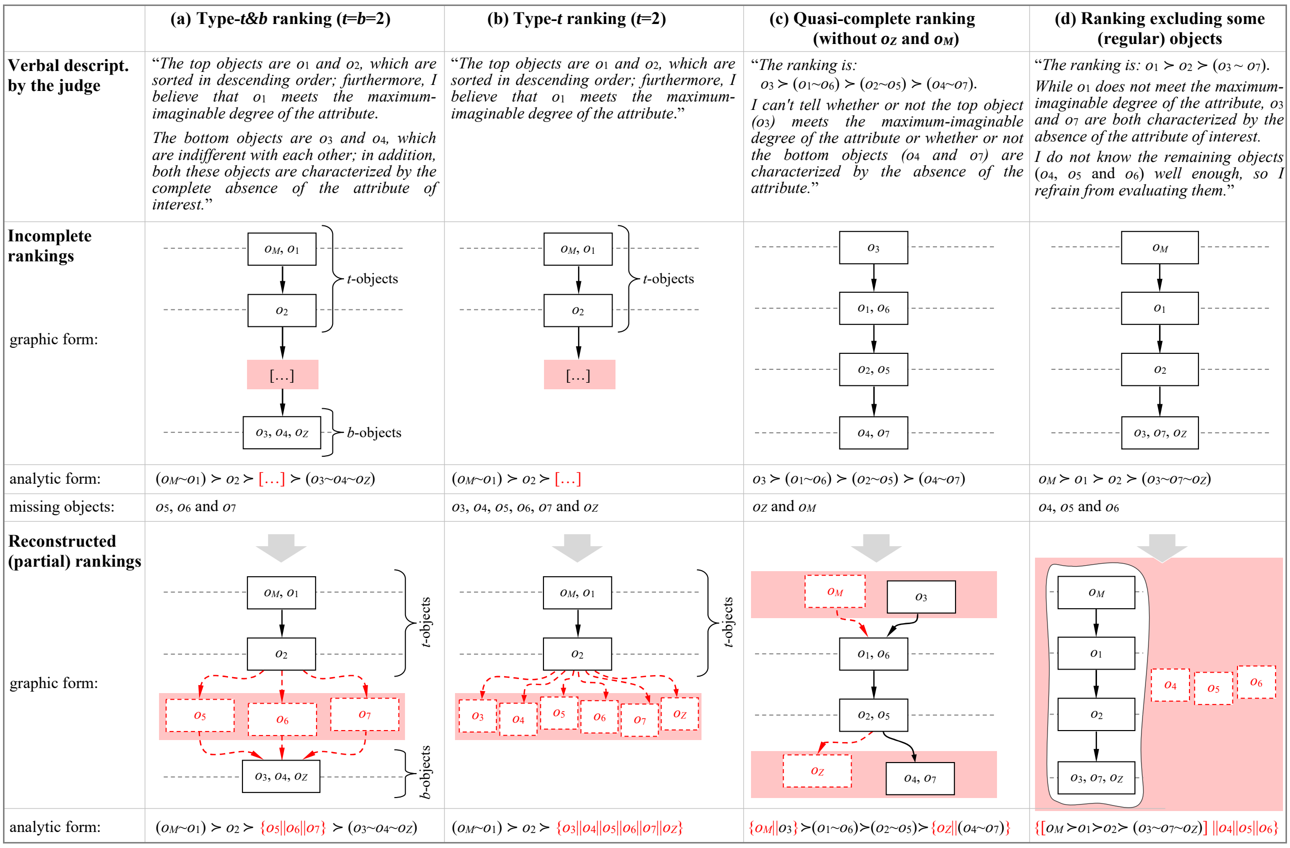


Figure 1. Example of four different types of incomplete rankings, including seven regular objects (o_1 to o_7) and two dummy objects (o_Z and o_M). These rankings can be turned into reconstructed (partial) rankings, which include all the (regular and dummy) objects; for ease of understanding, the reconstructed parts are marked in red.

1. the dummy object o_M should – by definition – be positioned at the same hierarchical level of the top object(s) or above. To take this hesitation into account, a relationship of incomparability between o_M and the top object(s) can be introduced. E.g., when reconstructing the ranking in Figure 2(b), o_M is inserted at the top of the ranking, with a relationship of incomparability with respect to the top object o_1;

2. the dummy object o_Z should – by definition – be positioned at the same hierarchical level of the bottom object(s) or below. To take this other hesitation into account, a relationship of incomparability between o_z and the bottom object(s) can be introduced. E.g., when reconstructing the ranking in Figure 2(b), o_z is inserted at the bottom of the ranking, with a relationship of incomparability with respect to the bottom objects o_5 and o_6.

Let us now make a brief digression on the meaning of the term "incomparability". In the Multiple Criteria Decision Analysis (MCDA) field, it generally depicts a comparison between two objects, in which a judge opts neither for a strict preference nor indifference relationship (Bana e Costa and Vansnick, 1999). This hesitation is generally due to lacking and/or contradictory information available, concerning the two objects. Table 2 summarizes a plurality of possible practical situations that lead to incomparability.

Table 2. Possible situations occurring when comparing two generic objects (o_i and o_j), depending on the *amount* and the *congruence* of the information available. This scheme was constructed by reworking the content of (Tsoukiàs and Vinke 1997, section 2).

Practical situation	Information available		Relation
1. o_i is preferred to o_j as the attribute of o_i is judged definitely better than that of o_j, from any point of view.	Proper amount	Congruent	$o_i \succ o_j$
2. o_i and o_j are indifferent because their attributes are indifferent from any point of view.	Proper amount	Congruent	$o_i \sim o_j$
3. o_i and o_j are in a conflicting position due to the clearly contradictory information about them.	Proper amount	Contradictory	$o_i \parallel o_j$
4. o_i and o_j could be either indifferent or conflicting, but the lack of relevant information in any of the two directions leads to hesitation.	Lacking	?	*idem*
5. o_i and o_j could be either indifferent or conflicting, but excessive and contradictory information in any of the two directions leads to hesitation.	Excessive amount	Contradictory	*idem*
6. o_i could be preferred or equivalent to o_j, but the lack of relevant information precludes determining the most appropriate relationship.	Lacking	Congruent	*idem*
7. o_i could be preferred or equivalent to o_j, but excessive and contradictory information precludes determining the most appropriate relationship.	Excessive amount	Contradictory	*idem*
8. o_j cannot be preferred to o_i, but the lack of relevant information precludes determining whether (i) o_i is preferred to o_j or (ii) they are conflicting.	Lacking	?	*idem*
9. o_j cannot be preferred to o_i, but excessive and contradictory information precludes determining whether (i) o_i is preferred to o_j or (ii) they are conflicting.	Excessive amount	Contradictory	*idem*
10. o_i could be preferred to o_j, but bad information leads to hesitation.	Lacking	Contradictory	*idem*

Firstly, we notice that there is no incomparability if-and-only-if the available information meets the following two requisites at the same time: it should be (i) in the appropriate *amount* and (ii) *non contradictory* (see the practical situations at points 1 and 2 of Table 2). Since the above requisites are not met in the remaining eight practical situations (at points 3 to 10), they all result in incomparability.

For the sake of simplicity, the present study will be limited to situations of incomparability that can be ascribed to that at point 6: "o_i could be preferred or equivalent to o_j, but the lack of relevant information precludes determining the most appropriate relationship". The incomparability of two objects will therefore be seen as a sort of hesitation between the relationships of strict preference and indifference, excluding conflicting situations. Of course, the authors are aware that this meaning of the term "incomparability" is narrower than that in other MCDA contexts.

Artificial deterioration of complete rankings

Let us now make a brief digression to show a contrivance that will be used later in our analysis: i.e., the artificial "deterioration" of a complete ranking to generate a set of incomplete preference rankings that are compatible[3] with it. Precisely, this contrivance will be used to generate incomplete rankings, artificially reproducing practical circumstances where the formulation of complete ones can be problematic.

Let us focus on the example in Figure 2, in which an initial complete preference ranking is given; Figure 2(a) shows the decomposition of this ranking into paired-comparison relationships of strict preference and indifference. Next, the initial complete ranking can be gradually deteriorated and turned into several incomplete preference rankings; e.g., consider the quasi-complete ranking in Figure 2(b), the Type-*t&b* ranking in Figure 2(c), the Type-*t&b* ranking in Figure 2(d), etc.. It can be noticed that for these incomplete rankings, new paired-comparison relationships of incomparability gradually replace those of strict preference and indifference in the complete ranking.

The *compatibility* between the initial complete ranking and the respective incomplete rankings is given by the fact that – excluding the paired-comparison relationships of incomparability – the remaining ones are identical. In general, a generic complete ranking can be deteriorated in different ways, generating a large set of incomplete rankings that are compatible with it.

Returning to the artificially-generated incomplete rankings, the *degree of completeness* of a generic *k*-th ranking can be quantitatively described by the synthetic indicator:

$$c = \frac{\text{No. of "usable" paired comparison relations in the preference ranking}}{\binom{n}{2}}, \tag{1}$$

which expresses the fraction of "usable" paired-comparison relationships – i.e. of strict preference or indifference – with respect to the total ones: $\binom{n}{2} = n \cdot (n-1)/2$, where n is the total number of objects of the problem; the adjective "usable" indicates that these are the only relationships that contribute to the solution of the decision-making problem of interest. By way of example, we determined the c values related to the rankings exemplified in Figure 2 (below the tables containing the paired-comparison relationships). This indicator tends to increase while rankings become more and more complete; for complete preference rankings, c is obviously 1.

Interestingly, even rankings that are apparently very incomplete may contain a relevant portion of usable paired-comparison relationships. E.g., consider the Type-*t* ranking in Figure 2(f), in which

[3] The concept of *compatibility* will be clarified below.

only the two more preferred regular objects are selected but not ordered; despite the apparently high degree of incompleteness, half of the usable paired-comparison relationships are still preserved (c = 50%).

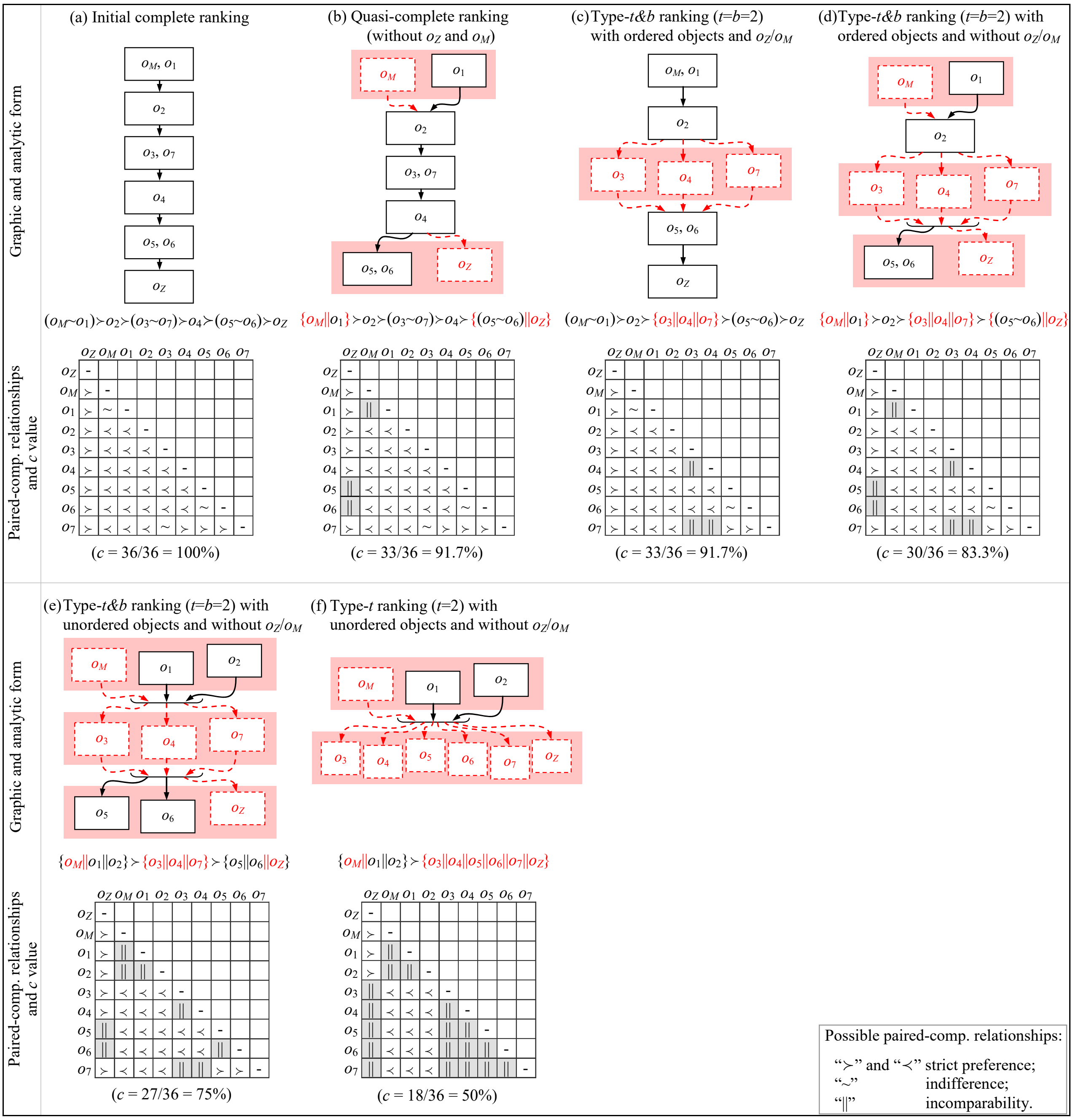


Figure 2. Example of gradual deterioration of a (complete) preference ranking (a), generating several incomplete rankings: (b), (c), (d), (e) and (f); for ease of understanding, the "reconstructed" parts are marked in red.

The indicator c can be extended from a single preference ranking to sets of m preference rankings – such as those characterizing a decision-making problem with m judges. We thus define a new aggregated indicator ($\bar{c}$), which depicts the overall degree of completeness:

$$\bar{c} = \frac{\sum_{k=1}^{m} \text{No. of "usable" paired comparison relations in the } k^{\text{th}} \text{ preference ranking}}{\sum_{k=1}^{m} \text{Total no. of paired comparison relations in the } k^{\text{th}} \text{ preference ranking}} = $$

$$= \frac{\sum_{k=1}^{m} c_k \cdot \binom{n}{2}}{m \cdot \binom{n}{2}} = \frac{\sum_{k=1}^{m} c_k}{m}, \qquad (2)$$

c_k being the degree of completeness of a generic k-th ranking.

Eq. 2 also shows that $\bar{c}$ can also be interpreted as the arithmetic mean of the c values related to the set of preference rankings under consideration.

For more examples about the artificial deterioration of complete rankings into incomplete ones, we refer the reader to the section “Example of generation of incomplete rankings” (in the Appendix).

Rationale of the aggregation technique

The mathematical formalization of the problem relies on the postulates and simplifying assumptions of the LCJ by Thurstone (1927), who postulated the existence of a *psychological/psychophysical continuum*, in which objects are positioned depending on the degree of a certain attribute. The position of a generic i-th object (f_i) is postulated to be distributed normally, in order to reflect the intrinsic judge-to-judge variability: $f_i \sim N(x_i, \sigma^2_{x_i})$, where x_i and $\sigma^2_{x_i}$ are the unknown mean value and variance related to the degree of the attribute of that object. Additionally, the distributions of different objects are considered equally dispersed and equally correlated with each other (Thurstone, 1927; Edwards, 1957). Considering two generic objects, o_i and o_j, and having introduced further simplifying hypotheses (Thurstone, 1927; Edwards, 1957), it can be asserted that:

$$p_{ij} = P[(f_i - f_j) \mathrm{V}\, 0] = 1 - \Phi[-(x_i - x_j)], \qquad (3)$$

which expresses the probability (p_{ij}) that the position of f_i is higher than that of f_j, Φ being the cumulative distribution function of the standard normal distribution $z \sim N(0, 1)$. Although p_{ij} is unknown, it can be estimated using the information contained in a set of judgments expressed by a number (m) of judges (Thurstone, 1927; Edwards, 1957). For more information on the estimation of the p_{ij} values, based on the positioning of the objects in the (reconstructed) rankings of judges, see (Franceschini and Maisano, 2019a).

From Eq. 3 it can be inferred that:

$$x_i - x_j = -\Phi^{-1}(1 - p_{ij}). \qquad (4)$$

Extending the reasoning to all possible pairs of objects, among which relationships of strict preference or indifference can be deduced (Franceschini and Maisano, 2019a), and introducing further simplifying assumptions, an over-determined system of (q) equations (like Eq. 4) can be then obtained. Since this system consists of linear equations with respect to the unknowns (i.e., x_i values), it can be expressed in matrix form as:

$$\begin{cases} \vdots \\ \sum_{k=1}^{n}(a_{hk} \cdot x_k) - b_h = 0 \quad \forall h \in [0, q] \\ \vdots \end{cases} \Rightarrow \quad \boldsymbol{A} \cdot \boldsymbol{X} - \boldsymbol{B} = \boldsymbol{0}, \tag{5}$$

$\boldsymbol{X} = [\ldots, x_i, \ldots]^T \in R^{n\times1}$ being the column vector containing the unknowns of the problem, a_{hk} being a generic element of matrix $\boldsymbol{A} \in R^{q\times n}$, and b_h being a generic element of vector $\boldsymbol{B} \in R^{n\times1}$. For details on the construction of $\boldsymbol{A}$ and $\boldsymbol{B}$, see (Gulliksen, 1956).

Then, this system can be solved by applying the GLS method (Kariya and Kurata, 2004), which allows to obtain an estimate of the mean degree of the attribute of each object:

$$\boldsymbol{X} = \left(\boldsymbol{A}^T \cdot \boldsymbol{W} \cdot \boldsymbol{A}\right)^{-1} \cdot \boldsymbol{A}^T \cdot \boldsymbol{W} \cdot \boldsymbol{B}, \tag{6}$$

where $\boldsymbol{W}$ is a (squared) matrix encapsulating the uncertainty related to the equations of the system; a practical way to define $\boldsymbol{W}$ is to apply the *Multivariate Law of Propagation of Uncertainty* (MLPU) to the system in Eq. 6, referring to the input variables affected by uncertainty (Kariya and Kurata, 2004), i.e., the p_{ij} values; for details, see (Franceschini and Maisano, 2019a).

The scale values in $\boldsymbol{X}$ are expressed on an arbitrary *interval* scale (Thurstone, 1927). The uncertainty of the solution can be estimated through a covariance matrix $\Sigma_{\boldsymbol{X}}$, which can be obtained by propagating the uncertainty of input data (i.e., p_{ij} values), through the following relationship:

$$\Sigma_{\boldsymbol{X}} = \left(\boldsymbol{A}^T \cdot \boldsymbol{W} \cdot \boldsymbol{A}\right)^{-1}. \tag{7}$$

Through the following transformation, the scale value of a generic i-th object (x_i) is transformed into a new scale value (y_i), which is defined in the conventional range [0, 10]:

$$\boldsymbol{Y} = \boldsymbol{Y}(\boldsymbol{X}) = [\ldots, y_i(\boldsymbol{X}), \ldots]^T = \left[\ldots, 100 \cdot \frac{x_i - x_Z}{x_M - x_Z}, \ldots\right]^T, \tag{8}$$

where: x_Z and x_M are the scale values of o_Z and o_M in the initial interval scale; x_i is the scale value of a generic i-th object in the initial interval scale; y_i is the scale value of a generic i-th object in the new scale (Franceschini and Maisano, 2019a). Since scale y "inherits" the *interval* property from scale x and has a conventional zero point that corresponds to the absence of the attribute (i.e.,

$y_Z = 0$), it can be reasonably considered as a *ratio* scale, without any conceptually prohibited "promotion" (Franceschini et al., 2019).

Next, the uncertainty related to the elements in $\boldsymbol{Y} = [\ldots, y_i, \ldots]^T \in R^{n\times 1}$ can be determined by applying a classic approach borrowed from Metrology to Eq. 8: the so-called *Delta Method*, also referred as *Law of Propagation of Uncertainty* or *Error Transmission Formula* (JCGM100:2008 2008). It is thus obtained:

$$\Sigma_{\boldsymbol{Y}} = \boldsymbol{J}_{\boldsymbol{Y}(\boldsymbol{X})} \cdot \Sigma_{\boldsymbol{X}} \cdot \boldsymbol{J}^T_{\boldsymbol{Y}(\boldsymbol{X})}, \tag{9}$$

where $\boldsymbol{J}_{\boldsymbol{Y}(\boldsymbol{X})} \in R^{n\times n}$ is a Jacobian matrix containing the partial derivatives related to the equations of the system in Eq. 8, with respect to the elements of $\boldsymbol{X}$; for details, see (Franceschini and Maisano, 2019a).

Combining Eqs. 7 and 9, $\Sigma_{\boldsymbol{Y}}$ can be expressed as:

$$\Sigma_{\boldsymbol{Y}} = \boldsymbol{J}_{\boldsymbol{Y}(\boldsymbol{X})} \cdot [(\boldsymbol{A}^T \cdot \boldsymbol{W} \cdot \boldsymbol{A})^{-1}] \cdot \boldsymbol{J}^T_{\boldsymbol{Y}(\boldsymbol{X})}. \tag{10}$$

Assuming that the p_{ij} and y_i values are normally distributed, a 95% confidence interval related to each y_i value can be computed as:

$$y_i \pm U_i = y_i \pm 2 \cdot \sigma_i \quad \forall i, \tag{11}$$

U_i being the so-called *expanded uncertainty* of y_i with a coverage factor $k = 2$ and $\sigma_i = \sqrt{\Sigma_{\boldsymbol{Y},(i,i)}}$ (JCGM 100:2008, 2008).

METHODOLOGY

General approach and response indicators

With the aim of investigating the robustness of the solution provided by the ZM_{II}-technique for incomplete problems, the methodological approach is articulated into several general points (see the flowchart in Figure 3):

- Numerous complete problems are randomly generated, determining the relevant solutions.
- These complete problems are then artificially deteriorated into incomplete ones, determining the new corresponding solutions (see the example in Figure 2). It is assumed that each complete problem has the same number of incomplete problems (i.e., nineteen, as explained in detail below).

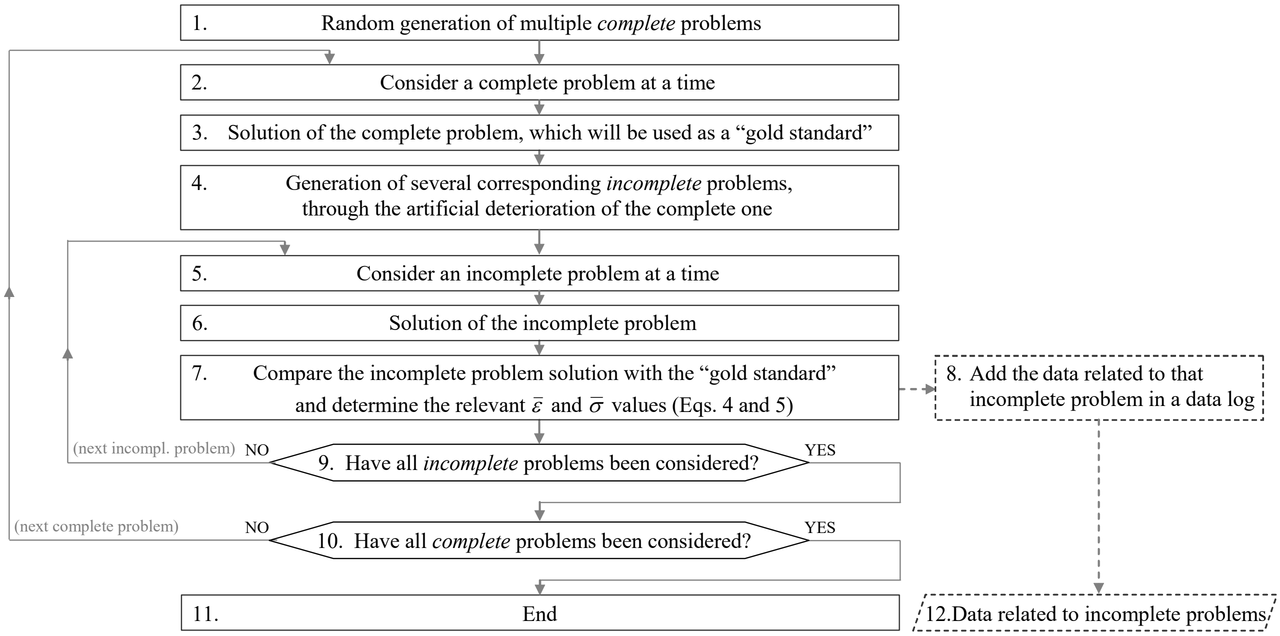


Figure 3. Flowchart representing the proposed methodological approach. $\bar{\varepsilon}$ and $\bar{\sigma}$ are two response indicators that will be defined later on.

- The solution of each incomplete problem is compared with the solution of the corresponding (source) complete problem, which can be interpreted as a sort of "gold standard". In fact, it can be demonstrated that the ZM_{II}'s solution to a complete problem coincides with that one provided by Thurstone's LCJ (Thurstone, 1927; Gulliksen, 1956; Edwards, 1957; Arbuckle and Nugent, 1973; Franceschini and Maisano, 2019a), which is a very consolidated aggregation technique, only applicable to complete problems. This is a sort of guarantee of *plausibility* of the ZM_{II}'s results.

  The robustness analysis is performed using two appropriate response indicators. The first one ($\bar{\varepsilon}$) expresses the deviation of the solution of a certain incomplete problem from that of the corresponding (source) complete problem, and it is structured as a *Root-Mean-Square Error* (RMSE) of the deviations between the y_i values resulting from these two problems (Ross, 2014):

$$\bar{\varepsilon} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - y_{i_{\text{(complete)}}}\right)^2}{n}}, \tag{12}$$

y_i being the scale value of the i-th object, resulting from the solution of the incomplete problem (i.e., $\boldsymbol{Y} = [\ldots, y_i, \ldots]^T$);

$y_{i_{\text{(complete)}}}$ being the scale value of the i-th object, resulting from the solution of the complete problem (i.e., $\boldsymbol{Y}_{\text{(complete)}} = [\ldots, y_{i\text{(complete)}}, \ldots]^T$);

n being the number of (dummy and regular) objects.

The closer the $y_{i_{\text{(complete)}}}$ values get to the y_i values, the more $\bar{\varepsilon}$ will tend to decrease; in this sense, this indicator is a measure of *(in)accuracy* (JCGM 200:2012, 2012). The calculation of this response indicator for complete rankings will obviously "degenerate" into $\bar{\varepsilon} = 0$.

The second response indicator is:

$$\bar{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} \sigma_i^2}{n}}, \tag{13}$$

σ_i^2 being the variance related to the scale value ($\hat{y}_i$) of the *i*-th object, i.e., one of the elements contained in the diagonal of the covariance matrix of ***Y*** (i.e., Σ_Y, defined in Eqs. 9 and 10). This response indicator – which can be calculated for both complete and incomplete problems – depicts the average *dispersion* of the solution.

Factorial simulations

This section describes the structured factorial (simulated) experiments that have been carried out. The scheme in Figure 4 summarizes the (sub-)factors considered, dividing them into three families:

a. *Structural factors*, which characterize a complete problem and the corresponding incomplete problems derived from it;
b. *Deterioration sub-factors*, which determine the way a complete problem is artificially "deteriorated", originating several incomplete problems.
c. *Completeness factor*, which quantifies the degree of completeness of each problem.

(a) *Structural factors*:
- Number of regular objects (n_{reg});
- Number of judges (m);
- Qualitative degree of agreement among judges;
- Kendall's concordance coefficient (W).

(b) *Deterioration sub-factors*:
- Type of rankings (hereafter abbreviated as "Ranking type");
- Ability of the judge to manage o_Z and o_M (hereafter abbreviated as "Manage o_Z/o_M?");
- Value of t and/or b (hereafter abbreviated as "t/b value");
- Ability of the judge to order the t- and/or b-objects (hereafter abbreviated as "Order t/b-objects?").

(c) *Completeness factor*:
- Overall degree of completeness of the problem (c).

Figure 4. Scheme of the (sub)factors characterizing the proposed factorial simulations. The dashed arrow indicates that the degree of completeness of a problem ($\bar{c}$) is influenced by the four deterioration sub-factors.

Several complete problems are initially generated, according to the logic illustrated in the following three points.

1. The number of *regular* objects (n_{reg}, which does not take into account the two dummy objects[4]) is varied over nine levels: n_{reg} = 4, 5, 6, 7, 8, 9, 10, 11 and 12.
2. For each of the above n_{reg} values, the number of judges (m) – which obviously coincides with the number of preference rankings – is varied over six levels: m = 5, 10, 15, 20, 25 and 30.
3. For each combination of n_{reg} and m values, four complete decision-making problems are randomly generated; these problems will be associated with one category expressing the qualitative degree of agreement of judges: *very low*, *low*, *intermediate*, *high*. Below is a detailed description of the procedure for the random generation of the complete problems and their classification into the above four categories.
   (a) A random scaling of the regular objects is generated, assigning a (random) scale value between 0 and 100 to each object.
   (b) The scale value of each object is randomly distorted by introducing a uniformly distributed error in the range $\pm\Delta$; the value of Δ is conventionally set to 10, 20, 30, and 40, respectively for a *very low*, *low*, *intermediate*, and *high* degree of inter-judge agreement. The resulting scale value is then rounded to the nearest ten (e.g., 20, 30, 40, etc.). Of course, this scale value can be outside the range [0, 100], with no effect on the subsequent steps.
   (c) The latter scaling is then translated into a new complete preference ranking.
   (d) Steps (b) and (c) are repeated m times, obtaining a complete problem with m randomly generated complete preference rankings.

   The inter-judge degree of agreement of a resulting complete problem can be better quantified through the so-called Kendall's *coefficient of concordance* (Kendall, 1962; Legendre 2010; Franceschini and Maisano, 2019b):

$$W = \frac{12 \cdot \sum_{i=1}^{n} R_i^2 - 3 \cdot m^2 \cdot n \cdot (n+1)^2}{m^2 \cdot n \cdot (n^2 - 1) - m \cdot \sum_{j=1}^{m} T_j}, \tag{14}$$

   where

   - n is the total number of (dummy and regular) objects;
   - R_i is the sum of the rank positions for the i-th object (o_i), i.e., $R_i = \sum_{i=1}^{m} r_{ij}$, in which terms r_{ij} represent the rank of o_i according to the j-th judge;
   - n is the total number of objects;
   - m is the total number of rankings;

[4] $n_{reg} = n - 2$, n being the total number of (regular and dummy) objects.

- T_j is a correction factor for ties[5], $T_j = \sum_{i=1}^{g_j} \left(t_i^3 - t_i\right)$, in which t_j is the number of tied ranks in the i-th group of tied ranks (where a group is a set of values having constant tied rank) and g_j is the number of groups of ties in the set of ranks (ranging from 1 to n) for judge j. Thus, T_j is the correction factor required for the set of ranks for judge j. Note that if there are no tied ranks for judge j, $T_j = 0$.

The range of W is between 0 (full disagreement) and 1 (full agreement). For more information on the construction of W, see (Kendall, 1962; Legendre, 2010; Franceschini and Maisano, 2019b). Obviously, the W value associated with a generic complete problem is likely to be correlated to the corresponding qualitative degree of inter-judge agreement; this will be empirically demonstrated in the next section.

Thus, $9 \cdot 6 \cdot 4 = 216$ (i.e., number of n_{reg} levels times number of m levels times number of levels of the qualitative inter-judge degree of agreement) complete decision-making problems were generated. For each of them, nineteen incomplete problems are then generated by changing the deterioration sub-factors, according to the contrivance anticipated in the sub-section "Artificial deterioration of complete rankings". Precisely:

(a) A single (quasi-complete) problem with m quasi-complete rankings. These incomplete rankings realistically represent practical situations in which the judges, while patiently considering all the regular objects, do find it difficult to manage o_Z and o_M.

(b) Three incomplete problems with m Type-*t&b* rankings including dummy objects, where *t*- and *b*-objects are ordered. For the first, second and third of these problems, t and b were set to 1, 2 and 3 respectively; in other words, apart from the dummy objects, the incomplete rankings include only 1, 2 and 3 more and less preferred regular objects[6]. These rankings can be appropriate when judges are not required to include all the regular objects, e.g., due to lack of time, concentration, etc..

(c) Three incomplete problems characterized by m Type-*t&b* rankings *without* dummy objects, where *t*- and *b*-objects are ordered. For the first, second and third of these problems, t and b were set to 1, 2 and 3 respectively. These rankings may be appropriate when judges are not required to include all the regular objects and find it difficult to manage o_Z and o_M.

[5] In this case, "ties" are represented by indifference relationships.

[6] Since these incomplete rankings have been obtained by deteriorating complete rankings, there may be practical cases in which it is not possible to identify *t*-and-only-*t* *t*-objects and/or *b*-and-only-*b* *b*-objects, due to indifference relationships ("~") between some of them. For example, considering the complete ranking $o_M \succ (o_1 \sim o_2) \succ o_3 \succ o_4 \succ \ldots$ and having set t=1, it is not possible to identify one-and-only-one *t*-object, since o_1 and o_2 are tied. To overcome this ambiguity, we have conventionally opted to include all the (possibly) tied objects within the *t*-objects, resulting in the following incomplete ranking $o_M \succ (o_1 \sim o_2) \succ \{o_3 \| o_4 \| \ldots\} \ldots$ and actually switching from t=1 to t=2. With simple adaptations, the same reasoning can be extended to *b*-objects.

(d) Three incomplete problems characterized by m Type-$t\&b$ rankings *without* dummy objects and with *unordered* t- and b-objects. For the first, second and third of these problems, t and b were set to 1, 2 and 3 respectively. The degree of completeness of these problems is significantly lower than that of the incomplete problems at point (c), since judges only select the t- and b-objects, without ordering them.

(e) Three incomplete problems with m Type-t rankings including dummy objects, where t- and b-objects are ordered. For the first, second and third of these problems, t was set to 1, 2 and 3 respectively. These rankings can be appropriate for decision-making problems aimed at ranking the more preferred objects only, neglecting the less preferred ones.

(f) Three incomplete problems characterized by m Type-t rankings *without* dummy objects, where t-objects are ordered. For the first, second and third of these problems, t was set to 1, 2 and 3 respectively. These rankings may be appropriate for decision-making problems aimed at ranking the more preferred objects only, in situations in which judges find it difficult to manage o_Z and o_M.

(g) Three incomplete problems characterized by m Type-t rankings *without* dummy objects and with *unordered* t-objects. For the first, second and third of these problems, t was set to 1, 2 and 3 respectively.

The section "Example of generation of incomplete rankings" (in the Appendix) further exemplifies the generation of incomplete rankings. As described before, a number of incomplete problems with variable degrees of incompleteness can be generated; for example, the "least incomplete" ones are the quasi-complete problems, while the most severe incompleteness is the one related to type-t problems with $t = 1$. The degree of completeness of the resulting incomplete problems will be qualitatively estimated using the indicator $\bar{c}$ (in Eq. 2).

The solution of each of the above incomplete problems will be compared with that of the corresponding (source) complete problem, determining the two response indicators $\bar{\varepsilon}$ and $\bar{\sigma}$ (in Eqs. 12 and 13). The total number of (complete and incomplete) decision-making problems generated will therefore be: $(9 \cdot 6 \cdot 4) \cdot (1 + 19) = 4{,}320$ (see also Figure 5).

The overall degree of completeness of a specific incomplete problem will be evaluated through $\bar{c}$ (see Eq. 2). Regarding the inter-judge degree of agreement, Eq. 14 cannot be applied directly to incomplete problems; in fact, W is only applicable to complete problems. For simplicity, it was assumed that incomplete problems "inherit" the W values of the corresponding (source) complete problems. The same can be extended to the qualitative degree of agreement (*very low*, *low*, *intermediate* and *high*). We plan to overcome this approximation in the future by developing a revised version of W, which can also be applied to incomplete rankings (Franceschini and Maisano, 2020b).

1. Change the number of regular objects (n_{reg}) at nine levels (i.e., from 4 to 12). (x9)

2. Change the number of judges (m) at six levels (i.e., 5, 10, 15, 20, 25 and 30). (x6)

3. Change the degree of agreement of judges at four levels (*very low*, *low*, *intermediate* and *high*). (x4)

4.1 Generate a complete problem with rankings matching the factors set at points 1, 2 and 3. (1)
Solve the problem; the resulting solution will be considered as a "gold standard".

4.2 Deterioration of the afore-mentioned complete problem into the following (19) incomplete problems:

4.2.1 Obtain a quasi-complete problem (i.e., with regular objects only). (1)

4.2.2 Change the t and b parameters at three levels (i.e., 1, 2 and 3).

4.2.2.1 Obtain a problem consisting of type-$t\&b$ rankings with ordered t- and b-objects, and with o_Z and o_M. (1x3)

4.2.2.2 Obtain a problem consisting of type-$t\&b$ rankings, with ordered t- and b-objects, and without o_Z and o_M. (1x3)

4.2.2.3 Obtain a problem consisting of type-$t\&b$ rankings, with unordered t- and b-objects. (1x3)

4.2.2.4 Obtain a problem consisting of type-t rankings with ordered t-objects, and with o_Z and o_M. (1x3)

4.2.2.2 Obtain a problem consisting of type-t rankings, with ordered t-objects, and without o_Z and o_M. (1x3)

4.2.2.3 Obtain a problem consisting of type-t rankings, with unordered t-objects. (1x3)

4.2.3 Solve the problems at points 4.2.1 and 4.2.2, and compare the resulting solutions with the "gold standard".

Total number of simulated problems: 9·6·4·20 = 4,320

Figure 5. Synthetic description of the factorial simulations.

RESULTS

A spreadsheet (which is available in the additional material) reports the detailed results of the 4,320 simulated decision-making problems. The following three subsections illustrate respectively (1) a qualitative and (2) a quantitative analysis of the above results, and (3) a regressive model reproducing the effects of the predominant factors.

Qualitative analysis

Figure 6 and Figure 7 contain the *main effects plots*[7], representing the effect of the major examined factors (n_{reg}, m, $\bar{c}$ and W) on the two responses, i.e., $\bar{\varepsilon}$ and $\bar{\sigma}$ respectively. For practical reasons, W was used as a quantitative indicator of the degree of inter-judge agreement. The box-plot in Figure 8 shows the relatively strong link between the W value of a generic complete problem and the respective qualitative degree of inter-judge agreement.

Additionally, the degree of completeness of a specific problem is assessed – at least in the first instance – using only $\bar{c}$ and neglecting the relevant deterioration sub-factors (which are specified in Figure 4(b)).

[7] The points in the plot are the means of a response variable at the various levels of each factor; for each level of the examined factor, the mean is calculated by averaging all the responses obtained changing the remaining factor. A reference line is drawn at the grand mean of the response data. This kind of plot is useful for comparing magnitudes of main effects (Box et al., 1978).

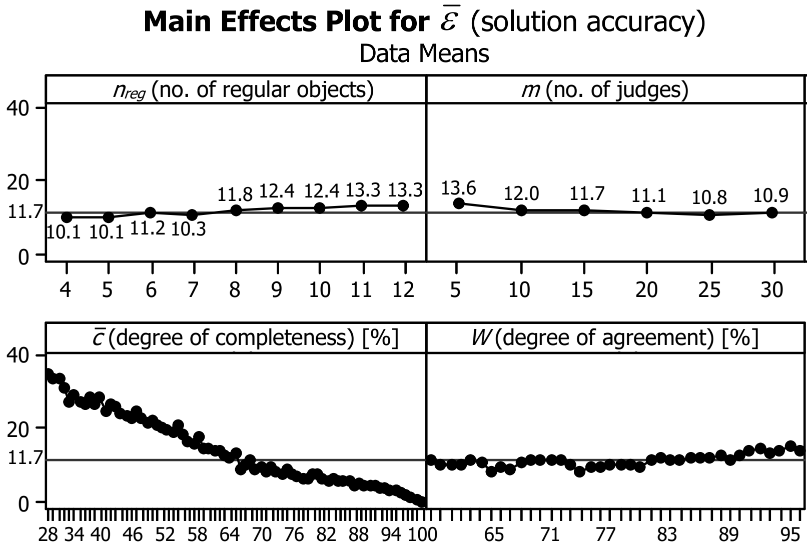


Figure 6. Minitab main effect plot of the major examined factors (n_{reg}, m, $\bar{c}$ and W) on the first response ($\bar{\varepsilon}$).

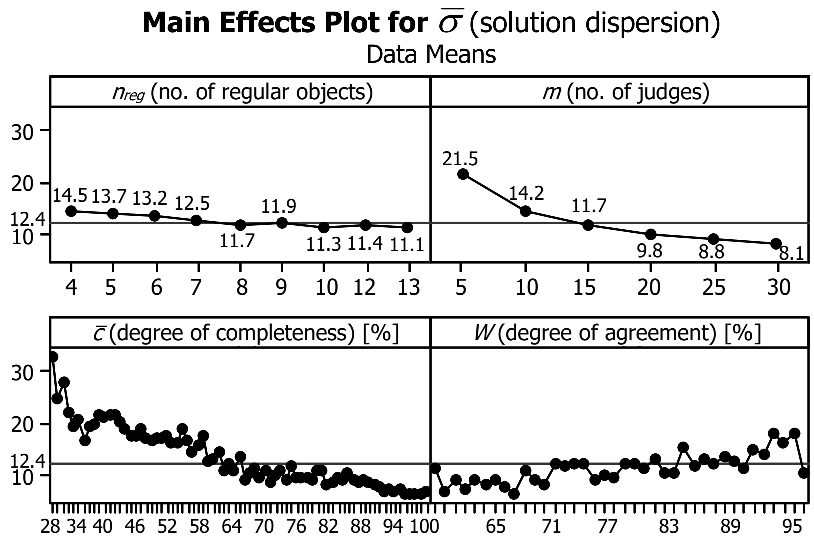


Figure 7. Minitab main effect plot of the major examined factors (n_{reg}, m, $\bar{c}$ and W) on the second response ($\bar{\sigma}$).

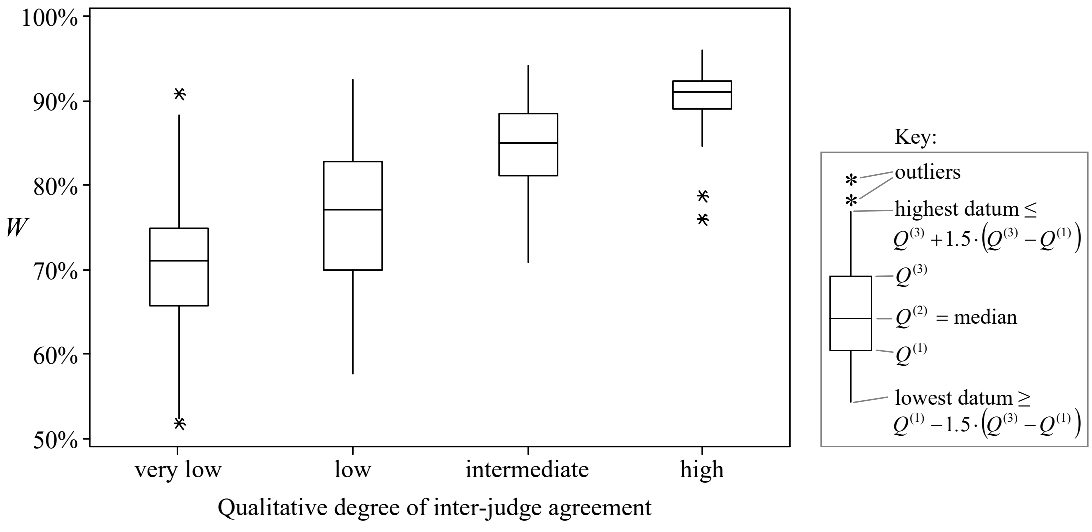


Figure 8. Minitab Box-plot related to the distributions of the W values, for the complete problems characterized by a certain qualitative degree of agreement (*very low*, *low*, *intermediate* and *high*). A relatively strong link between the two indicators can be observed.

Here are some specific comments on the graphs in Figure 6 and Figure 7.

- The factor with the predominant effect on both responses is $\bar{c}$, corroborating the hypothesis that the incompleteness of preference rankings contributes to deteriorate the problem solution significantly, both in terms of accuracy ($\bar{\varepsilon}$) and dispersion ($\bar{\sigma}$). Nevertheless, the aggregation technique in use proved to be robust, since it allowed to determine solutions that were not exaggeratedly different from the "gold standard", even for highly incomplete problems (e.g., with $\bar{c}$ < 50%).
- W seems to have a rather weak effect on both responses. Curiously, it seems that a certain disagreement among judges may contribute to reduce both $\bar{\varepsilon}$ and $\bar{\sigma}$. In fact, problems with a relatively low W value result in a more homogeneous distribution of the non-usable relationships of incomparability (among the possible paired comparisons), with a consequent benefit for the solution accuracy and dispersion.
- Factors n_{reg} and m seem to have not-very-relevant effects on the response $\bar{\varepsilon}$. The slight effects in Figure 6 are due to the procedure of random generation of the incomplete problems. In fact, for Type-*t&b* or Type-*t* rankings, t and b were set to 1, 2 or 3, regardless of the total number of regular objects (n_{reg}) of the problem. For a certain t/b value, the degree of completeness of rankings with relatively large n_{reg} values will reasonably be lower than that of rankings with relatively low n_{reg} values, with a consequent growth of $\bar{\varepsilon}$. In addition, as m grows (more judges) W will tend to decrease (more probability to obtain discordant rankings), with a consequent decrease in $\bar{\varepsilon}$ (see previous point).

  Regarding the response $\bar{\sigma}$, the effect of n_{reg} is irrelevant while that of m seems relevant. A plausible justification of the latter effect is that the variability of the input tends to decrease while increasing m and therefore the variability of the solution (depicted by $\bar{\sigma}$) will tend to decrease too (Franceschini and Maisano, 2019a).

Qualitative analysis

In order to qualitatively judge the presence of interactions between $\bar{c}$ and W, a plot[8] of $\bar{\varepsilon}$ as a function of these factors was constructed (see Figure 9). The curves represented in this graph – which can be approximately considered as "iso-W" – are not perfectly "parallel" to each other, denoting a slight correlation between $\bar{c}$ and W.

[8] Interaction between two factors is present when the response at a factor level depends upon the level(s) of the other factor. The greater the departure of the curves from the parallel state, the higher the degree of interaction (Box et al., 1978).

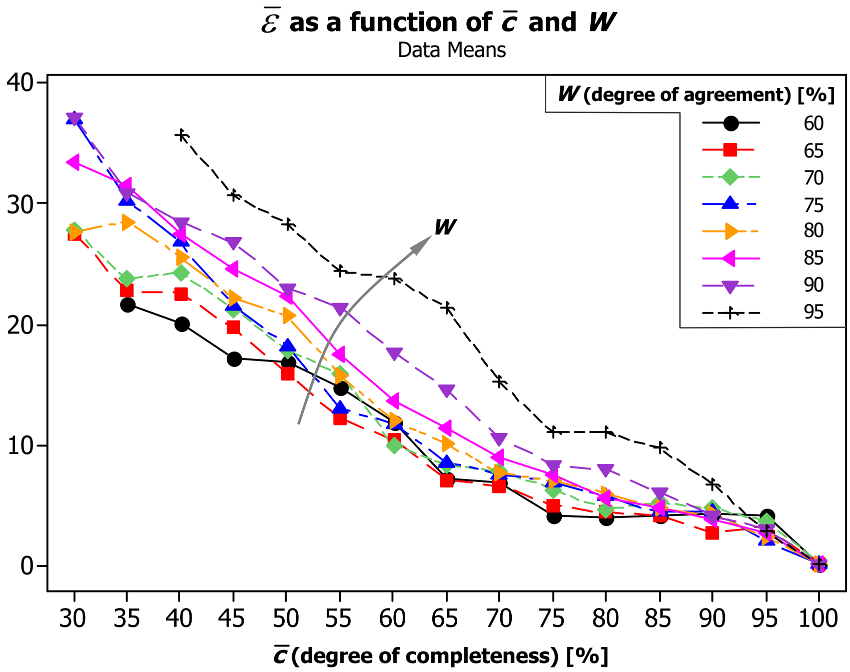


Figure 9. Minitab plot for $\bar{\varepsilon}$ as a function of the two most influential factors, $\bar{c}$ and W. The "iso-W" curves refer to problems with W values "rounded" to those reported in the legend (i.e. 60%, 65%, 70%, etc.), considering a resolution of 5%.

A further quantitative confirmation of the afore-illustrated results is given by Table 3, which contains the Pearson[9] correlation coefficients (ρ) between all the possible pairs of variables under consideration (i.e., main factors, n_{reg}, m, $\bar{c}$ and W, and responses, $\bar{\varepsilon}$ and $\bar{\sigma}$). For each possible pair, the ρ value is associated with a corresponding p-value for the hypothesis test of the correlation coefficient being zero (i.e., absence of correlation). Cases in which this value is lower than 0.001 – i.e. cases of rejection of the null hypothesis that there is no correlation – are those marked with the symbol "*".

Table 3. Pearson correlation table for the main factors (n, m, $\bar{c}$ and W) and responses ($\bar{\varepsilon}$ and $\bar{\sigma}$).

Variable	n_{reg}	m	$\bar{c}$	W	$\bar{\varepsilon}$	$\bar{\sigma}$
n_{reg}	1					
m	0	1				
$\bar{c}$	-0.266*	0.001	1			
W	-0.341*	-0.126*	0.084*	1		
$\bar{\varepsilon}$	0.119*	-0.088*	-0.843*	0.125*	1	
$\bar{\sigma}$	-0.145*	-0.561*	-0.580*	0.258*	0.732*	1

*p-value for the hypothesis test of the correlation coefficient being zero is lower than 0.001.

According to the proposed correlation analysis, the major factor affecting both responses is $\bar{c}$, followed by W. Additionally, we note that W is positively related with $\bar{\varepsilon}$ and $\bar{\sigma}$; this means that a

[9] This coefficient is a measure of the linear correlation between two variables and has a value between +1 and −1, where +1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation (Ross, 2014).

certain degree of disagreement between the rankings fosters the accuracy and precision of the solution. This behaviour certainly depends on the intrinsic characteristics of the ZM_{II} aggregation technique, with particular reference to the propagation of the uncertainty of the input data (Franceschini and Maisano, 2019a).

Excluding the correlations involving n_{reg} and m, since they are related to the way incomplete preference rankings are randomly generated[10], it can be noted that the correlation between $\bar{c}$ and W is relatively weak ($\rho \approx 0.084$), confirming the impression gained by analyzing Figure 9.

Let us now return to the major factor affecting responses, i.e., $\bar{c}$, which can be affected by the four deterioration sub-factors: "Ranking type", "Manage o_Z/o_M?", "t/b value", and "Order t/b-objects?". The interaction plot in Figure 10 shows that these sub-factors seem to be uncorrelated with each other. Not surprisingly, their mutual Pearson's correlation coefficients are null. On the other hand, the predominant sub-factor affecting $\bar{c}$ is the "Ranking type", with $\rho \approx -0.741$; in fact, Figure 10 shows that Type-t rankings tend to make $\bar{c}$ decrease dramatically, deteriorating the accuracy of the solution. We checked that the solution accuracy tends to worsen considerably for the less preferred objects especially, due to the relatively lower information content concerning these objects.

The factor $\bar{c}$ is affected by the sub-factors "t/b value" and "Order t/b-objects?", with ρ values of 0.363 and 0.369 respectively. On the other hand, the impact of the sub-factor "Manage o_Z/o_M?" is significantly lower ($\rho \approx 0.154$).

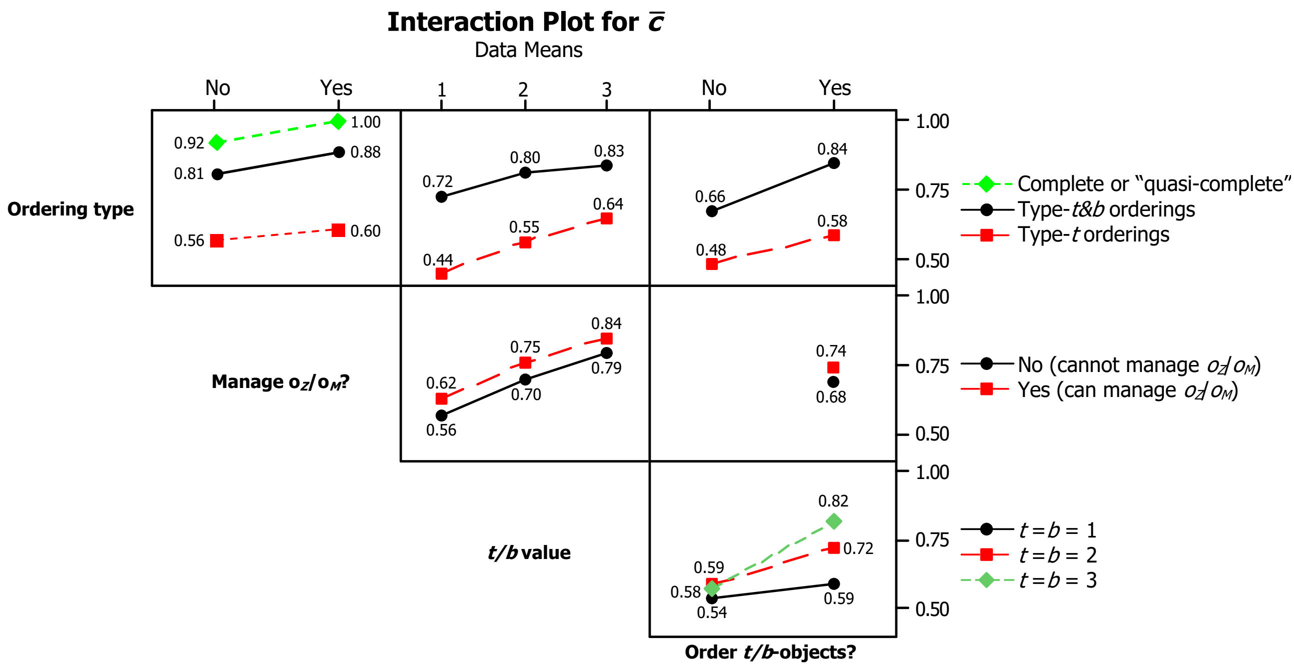


Figure 10. Interaction plot for $\bar{c}$, considering the four deterioration sub-factors: "Ranking type", "Manage o_Z/o_M?", "t/b value", and "Order t/b-objects?".

[10] E.g., for a given t/b value, the Type-t or Type $t\&b$ rankings with a relatively large number of regular objects (n_{reg}) are likely to have lower $\bar{c}$ and W values. Also, when increasing the number of judges (m), the possibility to generate discordant preference rankings will grow (and therefore W will decrease).

Regression model

To further confirm the above quantitative results, this sub-section illustrates the construction of a regression model, which links the response $\bar{\varepsilon}$ to the most influential factors ($\bar{c}$ and W). This analysis also provides an estimate of the so-called *effect size* of the factors themselves (Levine and Hullett, 2002).

This section focuses on a regression model to link the response $\bar{\varepsilon}$ with the predominant factors, $\bar{c}$ and W, for incomplete decision-making problems. This model enriches the analysis presented in the section "Results".

The data related to the decision-making problems described in the section "Factorial simulations" have been used to construct the model. Considering Figure 6 – which shows the patterns of $\bar{\varepsilon}(r)$ and $\bar{\varepsilon}(W)$ – a second order polynomial model was chosen. Being quadratic with respect to $\bar{c}$ and W, this model seems to well represent the previous graph patterns:

$$\bar{\varepsilon} = K_1 + K_2 \cdot \bar{c} + K_3 \cdot W + K_4 \cdot \bar{c}^2 + K_5 \cdot W^2 + K_6 \cdot \bar{c} \cdot W . \tag{15}$$

It is important to notice the presence of the last term ($K_6 \cdot \bar{c} \cdot W$), which accounts for the interaction between $\bar{c}$ and W.

With the support of the Minitab Best-Subsets tool, it was confirmed the importance of all terms (see results in Figure 11).

Best Subsets Regression: $\bar{\varepsilon}$ versus $\bar{c}$, W, $\bar{c}^2$, W^2, $\bar{c}$*W

Vars	R-Sq	R-Sq(adj)	Mallows Cp	S	(K_2) $\bar{c}$	(K_3) W	(K_4) $\bar{c}^2$	(K_5) W^2	(K_6) $\bar{c} \cdot W$
1	66.9	66.9	1528.3	5.4154	X				
1	62.4	62.4	2283.6	5.7763			X		
2	71.9	71.9	700.4	4.9894		X			X
2	71.9	71.9	702.1	4.9903	X			X	
3	74.8	74.8	223.2	4.7261	X		X	X	
3	74.7	74.7	239.4	4.7353	X	X	X		
4	76.0	76.0	22.7	4.6106	X	X	X		X
4	75.9	75.8	53.3	4.6284	X		X	X	X
5	76.2	76.1	$\boxed{6.0}$	4.6003	X	X	X	X	X

Figure 11. Results obtained from Minitab Best-Subsets tool. The above table suggests that the model with the five terms $\bar{c}$, W, $\bar{c}^2$, W^2 and $\bar{c} \cdot W$ is relatively precise and unbiased because its Mallows' Cp (6.0) is closest to the number of predictors plus the constant (6).

Since the variance of the response variable ($\bar{\varepsilon}$) is not homogeneous, a simple linear regression is not perfectly suitable. In particular, heteroscedasticity[11] may have the effect of giving too much weight to data subsets where the error variance is larger, when estimating coefficients. To reduce standard error associated with coefficient estimates, in regression in which homoscedasticity is

[11] Sect. 5 showed that $\bar{\sigma}^2$ is positively correlated with $\bar{\varepsilon}$. Since the variance of $\bar{\varepsilon}$ is related to $\bar{\sigma}$ (demonstration is left to the reader), it can be deduced that the variance of $\bar{\varepsilon}$ tends to grow while $\bar{\varepsilon}$ itself increases.

violated, a common approach is to weight observations by the reciprocal of the estimated point variance[12] (Box et al., 1978).
The final regression equation is

$$\bar{\varepsilon} \cong 43.1 - 103.4 \cdot \bar{c} + 22.4 \cdot W + 65.4 \cdot \bar{c}^2 + 0.00118 \cdot W^2 - 37.9 \cdot \bar{c} \cdot W \ . \tag{16}$$

The residual plots in Figure 12 show that the variance of the residuals seems rather uniform across the full range of fitted values, confirming the efficacy of the above "weighing" of the observations. Although the top-right plot denotes a slight under-fitting pattern in the bottom-left part, residuals seem globally satisfactory. Furthermore, they can be considered as randomly distributed by the Anderson-Darling normality test at $p < 0.05$.
The regression output is quantitatively examined by an ANOVA (see Figure 13). Based on a t test at $p < 0.05$, it can be deduced that all the terms in Eq. 16 are significant. The model fits the experimental data well.
In addition to the significance tests, the so-called *effect size* of the terms in Eq. 16 can be estimated (Levine and Hullett, 2002). Precisely, the ANOVA table in Figure 13 can be enriched by determining the *eta-squared* coefficient related to each term:

$$\eta^2 = \frac{SS_{term}}{SS_{Total}} \ , \tag{17}$$

being the SS and SS_{Total} values reported in the third column of the ANOVA table itself. From a practical point of view, η^2 describes the proportion of variance of the dependent variable ($\bar{\varepsilon}$), which is explained by the term of interest; according to a rule of thumb, $0 \le \eta^2 \le 0.01$ denotes a *small* effect, $0.01 < \eta^2 \le 0.06$ denotes a *medium* effect, while $0.06 < \eta^2 \le 1$ denotes a *large* effect (Pierce et al., 2004; Field, 2013). Therefore, returning to the analysis in Figure 13, the η^2 values reported in the lower right part denote a large effect of the $\bar{c}$ and W terms, a medium effect of the $\bar{c}^2$ and $\bar{c} \cdot W$ terms, and a small effect of the W^2 term.
Figure 14 graphically represents the final regression equation, confirming the previous results: the predominant effect of $\bar{c}$, the relatively lower effect of W on $\bar{\varepsilon}$, and the relatively weak interaction between $\bar{c}$ and W.
We underline that this regression model is aimed at confirming and quantifying the effects of $\bar{c}$ and W on $\bar{\varepsilon}$, which have already been highlighted in the "Results" section. Unfortunately, this model cannot be used for predictive purposes (e.g., to estimate the $\bar{\varepsilon}$ value of a certain incomplete problem), for the simple fact that W cannot be calculated directly for incomplete problems, but only for complete ones.

[12] Although being aware that $\bar{\sigma}^2$ depends on the variances of the $\hat{y}_i$ values (see Eq. 5), we have – for simplicity – considered $\bar{\sigma}^2$ as a proxy for the point variance related to each $\bar{\varepsilon}$ value.

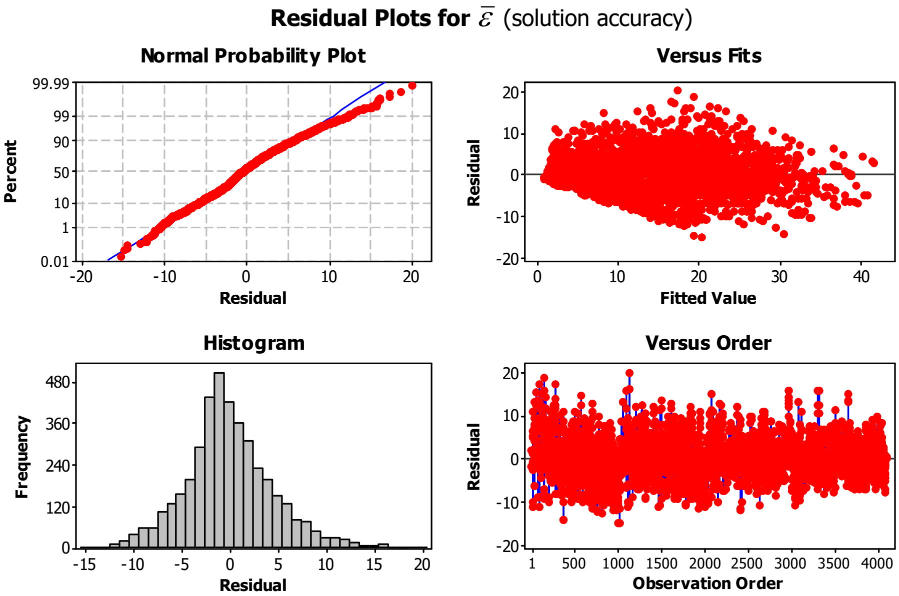


Figure 12. Minitab residual plots resulting from the (weighted) regression analysis.

(Weighted) Regression Analysis: $\bar{\varepsilon}$ versus $\bar{c}$, W, $\bar{c}^2$, W^2 and $\bar{c}\cdot W$

Regression Equation

$\bar{\varepsilon}$ = 43.1073 - 103.359·$\bar{c}$ + 22.3868·W + 65.389·$\bar{c}^2$ + 0.00117845·W^2 - 37.8818·$\bar{c}\cdot W$

Coefficients

Term	Coef	SE Coef	T	P
Constant	43.107	1.50909	28.5650	0.000
$\bar{c}$	-103.359	3.35368	-30.8196	0.000
W	22.387	3.20286	6.9896	0.000
$\bar{c}^2$	65.389	2.15008	30.4124	0.000
W^2	0.001	0.00027	4.4336	0.000
$\bar{c}$*W	-37.882	2.33708	-16.2091	0.000

Summary of Model

S = 0.415199 R-Sq = 68.86% R-Sq(adj) = 68.82%
PRESS = 684.233 R-Sq(pred) = 68.73%

Analysis of Variance

Source	DF	SS	Adj SS	Adj MS	F	P	η^2
Regression	5	320007	320007	64001.4	3261.66	0.0000000	
$\bar{c}$	1	287840	14946	14945.9	761.68	0.0000000	0.71132
W	1	16091	1344	1344.0	68.49	0.0000000	0.03976
$\bar{c}^2$	1	9614	11366	11366.0	579.24	0.0000000	0.02376
W^2	1	276	339	338.9	17.27	0.0000331	0.00068
$\bar{c}$*W	1	6185	6185	6185.3	315.22	0.0000000	0.01528
Error	4314	84651	84651	19.6			
Lack-of-Fit	3833	82686	82686	21.6	5.28	0.0000000	
Pure Error	481	1964	1964	4.1			
Total	4319	404658					

Figure 13. Results of the (weighted) regression analysis.

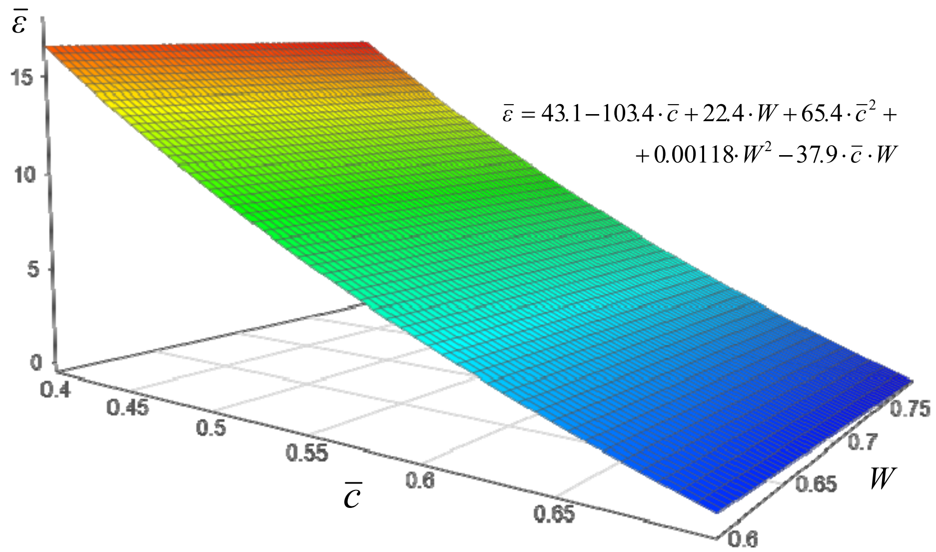


Figure 14. Graphic representation of the experimental regressive model in Eq. 16.

CONCLUDING REMARKS

The ZM_{II}-technique includes a flexible response mode that can be adapted to various practical contexts, in which (i) the concentration of judges cannot realistically be too high, or (ii) judges may find it difficult to formulate complete preference rankings. This flexibility encourages the reliability of input data, as it prevents judges from providing forced and unreliable responses. The present study proposed an original approach to verify the robustness of the ZM_{II}-technique, through massive structured experimentation.

Being based on a large number of simulated decision-making problems (i.e., 4,320), the proposed factorial experiments showed the robustness of the aggregation technique, even for significantly incomplete problems. The concept of *robustness* is here interpreted as the ability to provide a relatively stable solution, despite relatively significant variations in the type of input data. Interestingly, the technique tends to converge to a reasonable solution, even for problems where judges can identify only a few more preferred objects, neglecting the others (e.g., Type-t rankings with t =1 or 2). The solution towards which the technique converges is the one related to complete rankings; the plausibility of the ZM_{II}'s solution is guaranteed by the fact that it coincides with that of the LCJ, which is a very consolidated technique of the scientific literature. For a quantitative assessment of the plausibility of the results related to a specific complete problem – without necessarily knowing the corresponding complete problem from which it derives – it is possible to integrate the ZM_{II}-technique with some indicators present in the literature, such as those proposed in (Franceschini and Maisano, 2015) or others (Kendall, 1962; Perny, 1998; Brasil Filho et al., 2009).

The factor that mostly affects the solution accuracy ($\bar{\varepsilon}$) is $\bar{c}$, depicting the degree of completeness of a problem, while the predominant sub-factor affecting $\bar{c}$ is the "Ranking type": in fact, the solution accuracy and dispersion are significantly deteriorated in the presence of Type-t rankings. In addition, both the sub-factors "*t/b* value" and "Order *t*/*b*-block?" have a certain impact on $\bar{c}$, while the sub-factor "Manage o_Z/o_M?", depicting the ability to include o_Z and o_M within preference rankings, does not seem to be very influential.

The factor *W*, depicting the inter-judge agreement, has a less pronounced influence on $\bar{\varepsilon}$ than $\bar{c}$. Paradoxically, a relatively low degree of agreement tends to improve the solution, both in terms of accuracy and precision; however, this effect is rather weak.

The simulated (incomplete) problems are characterized by "homogeneous" preference rankings, i.e., all rankings have the same form of incompleteness (e.g., all Type-*t* rankings with t=2, unordered *t*-objects and without o_Z/o_M). Nevertheless, the aggregation technique can also be applied to problems characterized by "heterogeneous" preference rankings (e.g., partly complete, partly incomplete and/or with different forms of incompleteness). The example in the section "Application of the aggregation technique to 'heterogeneous' preference rankings" (in the Appendix) demonstrates the adaptability of the aggregation technique to problems with heterogeneous preference rankings.

The solutions of the simulated decision-making problems were determined using an *ad hoc* software application, developed in the MS Excel - Visual Basic for Applications environment, which is available on request. This software application made the generation and solution of the (thousands of) simulated problems agile.

A methodological limitation of the proposed study regards the estimation of the degree of inter-judge agreement: since *W* is applicable to complete problems only, it was assumed that one incomplete problem "inherits" the *W* value of the corresponding (source) complete problem. Regarding the future, we plan to replace *W* with another suitable indicator, which can also be applied to incomplete problems.

ACKNOWLEDGMENTS

This research was partially supported by the award "TESUN-83486178370409 finanziamento dipartimenti di eccellenza CAP. 1694 TIT. 232 ART. 6", which was conferred by "Ministero dell'Istruzione, dell'Università e della Ricerca".

REFERENCES

Arbuckle, J., Nugent, J.H. (1973), "A general procedure for parameter estimation for the law of comparative judgement", British Journal of Mathematical and Statistical Psychology, Vol. 26, No. 2, pp. 240-260.

Arrow, K.J. (1963). Social Choice and Individual Values, 2nd edition, Yale University Press, New Haven and London.

Bana e Costa, C.A., Vansnick, J.C. (1999) Preference Relations and MCDM. In: Gal T., Stewart T.J., Hanne T. (eds) Multicriteria Decision Making. International Series in Operations Research & Management Science, vol 21. Springer, Boston, MA.

Box, G.E., Hunter, W.G., Hunter, J.S. (1978). Statistics for experimenters, Wiley, New York.

Brasil Filho, A.T., Pinheiro, P.R., Coelho, A.L., Costa, N.C. (2009). Comparison of two prototype-based multicriteria classification methods. In 2009 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making (MCDM) (pp. 133-140). IEEE.

Chen, Y.L., Cheng, L.C. (2010). An approach to group ranking decisions in a dynamic environment. Decision support systems, 48(4), 622-634.

Das, P., Mukherjee, S. (2007), Modeling of Customer Preferences on Product Features and Comparing the Competitors' Performances. Quality Engineering, 20(1), 53-62.

DeVellis, R.F. (2016), Scale development: Theory and applications (4th edn.), Sage, London.

Edwards, A.L. (1957), Techniques of Attitude Scale Construction, Irvington Publishers, New York.

Field, A. (2013). Discovering statistics using IBM SPSS statistics. Sage. Los Angeles, London, New Dehli, Singapore and Washington DC.

Fishburn P.C. (1989). Foundations of decision analysis: Along the way. Management Science 35(4), 387-405.

Franceschini, F., Maisano, D. (2015). Checking the consistency of the solution in ordinal semi-democratic decision-making problems. Omega, 57, 188-195.

Franceschini, F., Galetto, M., Maisano, D., Mastrogiacomo, L. (2017). Ordinal aggregation operators to support the engineering characteristic prioritization in QFD. The International Journal of Advanced Manufacturing Technology, 91(9-12), 4069-4080.

Franceschini, F., Maisano, D. (2018), A new proposal to improve the customer competitive benchmarking in QFD. Quality Engineering, 30(4), 730-761.

Franceschini, F., Maisano, D. (2019a), Fusing incomplete preference rankings in design for manufacturing applications through the ZM_{II}-technique. The International Journal of Advanced Manufacturing Technology, 103(9-12): 3307-3322.

Franceschini F., Maisano D. (2019b), Design decisions: concordance of designers and effects of the Arrow's theorem on the collective preference ranking, Research in Engineering Design, 30(3), 425-434.

Franceschini, F., Maisano, D. (2020a), Adapting Thurstone's Law of Comparative Judgment to fuse preference orderings in manufacturing applications, Journal of Intelligent Manufacturing, 31: 387-402, DOI: 10.1007/s10845-018-1452-5.

Franceschini F., Maisano D. (2020b), Decision concordance with incomplete expert rankings in manufacturing applications, to appear in Research in Engineering Design, DOI: 10.1007/s00163-020-00340-x.

Franceschini, F., Galetto, M., Maisano, D. (2019), Designing Performance Measurement Systems: Theory and Practice of Key Performance Indicators, book series Management for Professionals, Springer Nature, Berlin, ISBN: 978-3-030-01191-8, DOI: 10.1007/978-3-030-01192-5.

Gierz, G., Hofmann, K.H., Keimel, K., Mislove, M., Scott, D.S. (2003). Continuous Lattices and Domains. Encyclopedia of Mathematics and its Applications. 93. Cambridge University Press. ISBN 978-0-521-80338-0.

Gulliksen, H. (1956), “A least squares solution for paired comparisons with incomplete data”, Psychometrika, Vol. 21, pp. 125-134.

Harzing, A.W., Baldueza, J., ... and Zander, L. (2009), “Rating versus ranking: What is the best way to reduce response and language bias in cross-national research?”, International Business Review, Vol. 18, No. (4), pp. 417-432.

JCGM 100:2008 (2008), Evaluation of measurement data - Guide to the expression of uncertainty in measurement, BIPM, Paris.

JCGM 200:2012 (2012) International vocabulary of metrology – Basic and general concepts and associated terms (VIM). International Organization for Standardization, Geneva. Retrieved January 2020 from http://www.bipm.org/en/publications/guides/vim.html.

Kariya T., Kurata H. (2004), Generalized least squares, John Wiley & Sons, New York.

Keeney, R.L., Raiffa, H. (1993), Decisions with multiple objectives: preferences and value trade-offs, Cambridge university press, Cambridge.

Kendall, M.G. (1962), Rank Correlation Methods (3rd ed.), New York, Hafner.,chap. 6.

Lagerspetz, E. (2016), Social choice and democratic values, Springer, Heidelberg.

Legendre, P. (2010). Coefficient of concordance. Encyclopedia of research design, 1, 164-169.

Levine, T.R., Hullett, C.R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. Human Communication Research, 28(4), 612-625.

Nederpelt, R., Kamareddine, F. (2004), Logical Reasoning: A First Course, King’s College Publications, London.

Nurmi, H. (1983). Voting procedures: A summary analysis. Br. J. Political Sci. 13, 181-208.

Paruolo, P., Saisana, M., Saltelli, A. (2013). “Ratings and rankings: voodoo or science?”, Journal of the Royal Statistical Society: Series A (Statistics in Society), Vol. 176, No. 3, pp. 609-634.

Perny, P. (1998). Multicriteria filtering methods based on concordance/nondiscordance principles. Annals of Operational Research 80:137–167.

Research 80:137–167

Pierce, C.A., Block, R.A., Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. Educational and psychological measurement, 64(6), 916-924.

Roberts, F.S. (1979), Measurement theory: with applications to decision making, utility, and the social sciences, Encyclopedia of Mathematics and its Applications, vol. 7, Addison-Wesley, Reading, MA.

Ross, S.M. (2014), Introduction to probability and statistics for engineers and scientists, Academic Press, Amsterdam.

Thurstone, LL. (1927), “A law of comparative judgment”, Psychological Review, Vol. 34, No. 4, p. 273.

Wang, G.X., Huang, S.H., Yan, Y., Du, J.J. (2017). Reconfiguration schemes evaluation based on preference ranking of key characteristics of reconfigurable manufacturing systems. The International Journal of Advanced Manufacturing Technology, 89(5-8), 2231-2249.

Zhang, M., Guo, X., Chen, G. (2016). Prediction uncertainty in collaborative filtering: Enhancing personalized online product ranking. Decision Support Systems, 83, 10-21.

APPENDIX

Example of generation of incomplete rankings

Table A.1 shows an example of deterioration of a single complete ranking – i.e., $(o_M\sim o_{12})\succ(o_9\sim o_8)\succ o_1\succ(o_{11}\sim o_2)\succ(o_{10}\sim o_3\sim o_6)\succ o_7\succ(o_4\sim o_5\sim o_Z)$ – into the aforementioned nineteen incomplete rankings, by changing the deterioration sub-factors in Figure 4(b). For each of these incomplete rankings, the corresponding c_k value is also determined to depict the degree of completeness (see the last column of Table A.1); of course, the c_k values may change depending on the combination of deterioration sub-factors.

Table A.1. Example of generation of nineteen different incomplete preference rankings (in the last column) by "deteriorating" a single complete ranking – $(o_M\sim o_{12})\succ(o_9\sim o_8)\succ o_1\succ(o_{11}\sim o_2)\succ(o_{10}\sim o_3\sim o_6)\succ o_7\succ(o_4\sim o_5\sim o_Z)$ – through different combinations of the four sub-factors: "Ranking type", "Manage o_Z/o_M?", "t/b value", and "Order t/b objects?". Reconstructed parts are marked in red.

Ranking type	Manage o_Z/o_M?	t/b value	Order t/b-objects?	Incomplete rankings	c
Quasi-complete	No	N/A	N/A	$\{o_M\Vert o_{12}\}\succ(o_9\sim o_8)\succ o_1\succ(o_{11}\sim o_2)\succ(o_{10}\sim o_3\sim o_6)\succ o_7\succ\{o_Z\Vert(o_4\sim o_5)\}$	96.7%
Type-$t\&b$	Yes	1	Yes	$(o_M\sim o_{12})\succ\{o_1\Vert o_2\Vert o_3\Vert o_6\Vert o_7\Vert o_8\Vert o_9\Vert o_{10}\Vert o_{11}\}\succ(o_4\sim o_5\sim o_Z)$	60.4%
Type-$t\&b$	Yes	2	Yes	$(o_M\sim o_{12})\succ(o_9\sim o_8)\succ\{o_1\Vert o_2\Vert o_3\Vert o_6\Vert o_7\Vert o_{10}\Vert o_{11}\}\succ(o_4\sim o_5\sim o_Z)$	76.9%
Type-$t\&b$	Yes	3	Yes	$(o_M\sim o_{12})\succ(o_9\sim o_8)\succ\{o_1\Vert o_2\Vert o_3\Vert o_6\Vert o_{10}\Vert o_{11}\}\succ o_7\succ(o_4\sim o_5\sim o_Z)$	83.5%
Type-$t\&b$	No	1	Yes	$\{o_M\Vert o_{12}\}\succ\{o_1\Vert o_2\Vert o_3\Vert o_6\Vert o_7\Vert o_8\Vert o_9\Vert o_{10}\Vert o_{11}\}\succ\{o_Z\Vert(o_4\sim o_5)\}$	57.1%
Type-$t\&b$	No	2	Yes	$\{o_M\Vert o_{12}\}\succ(o_9\sim o_8)\succ\{o_1\Vert o_2\Vert o_3\Vert o_6\Vert o_7\Vert o_{10}\Vert o_{11}\}\succ\{o_Z\Vert(o_4\sim o_5)\}$	73.6%
Type-$t\&b$	No	3	Yes	$\{o_M\Vert o_{12}\}\succ(o_9\sim o_8)\succ\{o_1\Vert o_2\Vert o_3\Vert o_6\Vert o_{10}\Vert o_{11}\}\succ o_7\succ\{o_Z\Vert(o_4\sim o_5)\}$	80.2%
Type-$t\&b$	N/A	1	No	$\{o_M\Vert o_{12}\}\succ\{o_1\Vert o_2\Vert o_3\Vert o_6\Vert o_7\Vert o_8\Vert o_9\Vert o_{10}\Vert o_{11}\}\succ\{o_Z\Vert o_4\Vert o_5\}$	56.0%
Type-$t\&b$	N/A	2	No	$\{o_M\Vert o_8\Vert o_9\Vert o_{12}\}\succ\{o_1\Vert o_2\Vert o_3\Vert o_6\Vert o_7\Vert o_{10}\Vert o_{11}\}\succ\{o_Z\Vert o_4\Vert o_5\}$	67.0%
Type-$t\&b$	N/A	3	No	$\{o_M\Vert o_8\Vert o_9\Vert o_{12}\}\succ\{o_1\Vert o_2\Vert o_3\Vert o_6\Vert o_{10}\Vert o_{11}\}\succ\{o_Z\Vert o_4\Vert o_5\Vert o_7\}$	70.3%
Type-t	Yes	1	Yes	$(o_M\sim o_{12})\succ\{o_Z\Vert o_1\Vert o_2\Vert o_3\Vert o_4\Vert o_5\Vert o_6\Vert o_7\Vert o_8\Vert o_9\Vert o_{10}\Vert o_{11}\}$	27.5%
Type-t	Yes	2	Yes	$(o_M\sim o_{12})\succ(o_9\sim o_8)\succ\{o_Z\Vert o_1\Vert o_2\Vert o_3\Vert o_4\Vert o_5\Vert o_6\Vert o_7\Vert o_{10}\Vert o_{11}\}^{(*)}$	50.5%
Type-t	Yes	3	Yes	$(o_M\sim o_{12})\succ(o_9\sim o_8)\succ\{o_Z\Vert o_1\Vert o_2\Vert o_3\Vert o_4\Vert o_5\Vert o_6\Vert o_7\Vert o_{10}\Vert o_{11}\}$	50.5%
Type-t	No	1	Yes	$\{o_M\Vert o_{12}\}\succ\{o_Z\Vert o_1\Vert o_2\Vert o_3\Vert o_4\Vert o_5\Vert o_6\Vert o_7\Vert o_8\Vert o_9\Vert o_{10}\Vert o_{11}\}$	26.4%
Type-t	No	2	Yes	$\{o_M\Vert o_{12}\}\succ(o_9\sim o_8)\succ\{o_Z\Vert o_1\Vert o_2\Vert o_3\Vert o_4\Vert o_5\Vert o_6\Vert o_7\Vert o_{10}\Vert o_{11}\}^{(*)}$	49.5%
Type-t	No	3	Yes	$\{o_M\Vert o_{12}\}\succ(o_9\sim o_8)\succ\{o_Z\Vert o_1\Vert o_2\Vert o_3\Vert o_4\Vert o_5\Vert o_6\Vert o_7\Vert o_{10}\Vert o_{11}\}$	49.5%
Type-t	N/A	1	No	$\{o_M\Vert o_{12}\}\succ\{o_Z\Vert o_1\Vert o_2\Vert o_3\Vert o_4\Vert o_5\Vert o_6\Vert o_7\Vert o_8\Vert o_9\Vert o_{10}\Vert o_{11}\}$	26.4%
Type-t	N/A	2	No	$\{o_M\Vert o_8\Vert o_9\Vert o_{12}\}\succ\{o_Z\Vert o_1\Vert o_2\Vert o_3\Vert o_4\Vert o_5\Vert o_6\Vert o_7\Vert o_{10}\Vert o_{11}\}^{(*)}$	44.0%
Type-t	N/A	3	No	$\{o_M\Vert o_8\Vert o_9\Vert o_{12}\}\succ\{o_Z\Vert o_1\Vert o_2\Vert o_3\Vert o_4\Vert o_5\Vert o_6\Vert o_7\Vert o_{10}\Vert o_{11}\}$	44.0%

(*) Consistently with the convention described in footnote 6, the number of t- or b-objects in these rankings is higher than the respective t and b value, due to the presence of some indifference relationships among these objects.

Application of the aggregation technique to "heterogeneous" preference rankings

This section exemplifies a decision-making problem in which preference rankings of different nature are aggregated (e.g., partly complete, partly incomplete and/or with different forms of incompleteness). Let us consider the example in Table A.2, which contains twenty not-necessarily-complete preference rankings (representing an incomplete problem), obtained by deteriorating twenty corresponding complete rankings (representing the "source" complete problem). Deterioration is performed by changing the four sub-factors in Figure 4.

The aggregation technique of interest is applied to the incomplete problem, resulting in the solution in Figure A.1. We note that the error bands of the $\hat{y}_i$ values (which represent $\hat{y}_i \pm U_{\hat{y}_i}$; see Eq. 11) are relatively wide, due to the relatively low degree of completeness of the problem ($\bar{c}$ = 62.2%). Despite this, the above solution is relatively close to that of the complete problem (which – by the way – is also characterized by W = 69.9%). Figure A.2 shows that the $\hat{y}_i$ values related to these two solutions are strongly correlated.

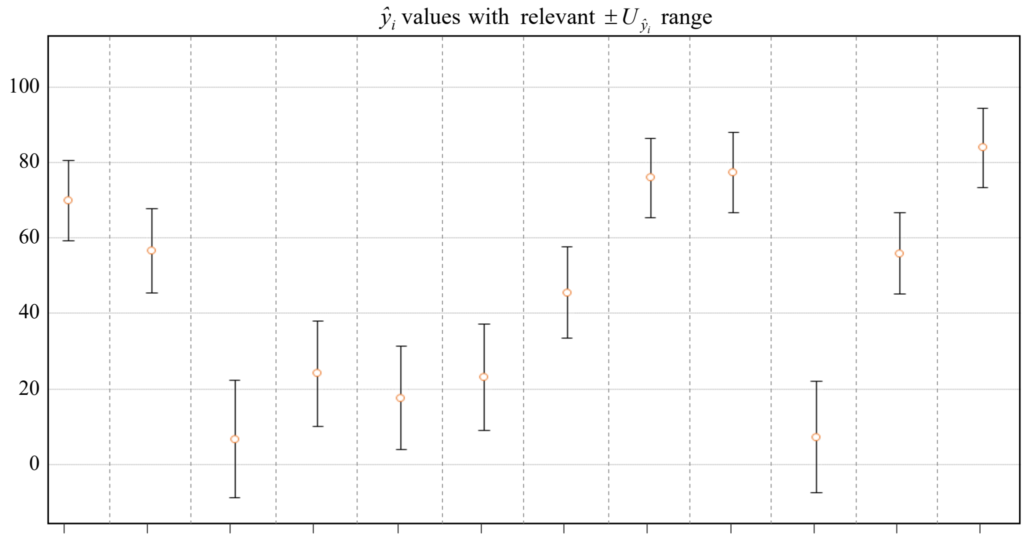
Table A.2. Incomplete preference rankings (in the last column) that are obtained by deteriorating some complete source rankings (in the second column). It can be noticed that the deterioration mechanism may differ from ranking to ranking. The reconstructed parts of the incomplete rankings are marked in red.

Judge	Complete rankings (Complete problem)	Deterioration parameters: Ranking type	Manage o_Z/o_M?	t/b value	Order t/b-block(s)?	Non-necessarily-complete rankings (Incomplete problem)
j_1	$(o_M \sim o_1 \sim o_8 \sim o_{12}) \succ o_9 \succ o_7 \succ o_6 \succ o_{11} \succ o_2 \succ o_4 \succ (o_3 \sim o_{10} \sim o_Z \sim o_5)$	Type-$t\&b$	Yes	3	Yes	$(o_M \sim o_1 \sim o_8 \sim o_{12}) \succ \{o_2 \Vert o_4 \Vert o_6 \Vert o_7 \Vert o_9 \Vert o_{11}\} \succ (o_3 \sim o_{10} \sim o_Z \sim o_5)$
j_2	$(o_M \sim o_8) \succ o_{11} \succ o_9 \succ (o_7 \sim o_{12}) \succ o_1 \succ (o_4 \sim o_2) \succ (o_{10} \sim o_6 \sim o_3) \succ (o_5 \sim o_Z)$	Type-$t\&b$	Yes	3	No	$\{o_M \Vert o_8 \Vert o_9 \Vert o_{11}\} \succ \{o_1 \Vert o_2 \Vert o_4 \Vert o_7 \Vert o_{12}\} \succ \{o_Z \Vert o_3 \Vert o_5 \Vert o_6 \Vert o_{10}\}$
j_3	$(o_M \sim o_9) \succ o_1 \succ o_{11} \succ o_{12} \succ o_8 \succ o_5 \succ (o_3 \sim o_2) \succ o_7 \succ (o_6 \sim o_{10} \sim o_4 \sim o_Z)$	Type-t	No	3	Yes	$\{o_M \Vert o_9\} \succ o_1 \succ o_{11} \succ \{o_Z \Vert o_2 \Vert o_3 \Vert o_4 \Vert o_5 \Vert o_6 \Vert o_7 \Vert o_8 \Vert o_{10} \Vert o_{12}\}$
j_4	$(o_M \sim o_9 \sim o_{12}) \succ o_8 \succ o_1 \succ (o_4 \sim o_2) \succ o_7 \succ o_{11} \succ (o_6 \sim o_3) \succ (o_{10} \sim o_Z \sim o_5)$	Complete	Yes	N/A	N/A	$(o_M \sim o_9 \sim o_{12}) \succ o_8 \succ o_1 \succ (o_4 \sim o_2) \succ o_7 \succ o_{11} \succ (o_6 \sim o_3) \succ (o_{10} \sim o_Z \sim o_5)$
j_5	$o_M \succ o_1 \succ o_9 \succ (o_8 \sim o_4) \succ (o_6 \sim o_{11} \sim o_{12}) \succ (o_2 \sim o_7) \succ o_3 \succ (o_{10} \sim o_Z \sim o_5)$	Type-$t\&b$	No	1	Yes	$\{o_M \Vert o_1\} \succ \{o_2 \Vert o_3 \Vert o_4 \Vert o_6 \Vert o_7 \Vert o_8 \Vert o_9 \Vert o_{11} \Vert o_{12}\} \succ \{o_Z \Vert (o_{10} \sim o_5)\}$
j_6	$(o_M \sim o_{12}) \succ (o_9 \sim o_8) \succ o_1 \succ (o_{11} \sim o_2) \succ (o_{10} \sim o_3 \sim o_6) \succ o_7 \succ (o_4 \sim o_5 \sim o_Z)$	Type-t	Yes	3	Yes	$(o_M \sim o_{12}) \succ (o_9 \sim o_8) \succ \{o_Z \Vert o_1 \Vert o_2 \Vert o_3 \Vert o_4 \Vert o_5 \Vert o_6 \Vert o_7 \Vert o_{10} \Vert o_{11}\}$
j_7	$o_M \succ (o_1 \sim o_8) \succ o_9 \succ o_{12} \succ o_2 \succ (o_7 \sim o_4) \succ o_{11} \succ o_5 \succ (o_6 \sim o_{10} \sim o_Z \sim o_3)$	Type-$t\&b$	No	1	No	$\{o_M \Vert o_1 \Vert o_8\} \succ \{o_2 \Vert o_4 \Vert o_5 \Vert o_7 \Vert o_9 \Vert o_{11} \Vert o_{12}\} \succ \{o_Z \Vert o_3 \Vert o_6 \Vert o_{10}\}$
j_8	$(o_M \sim o_9) \succ (o_2 \sim o_7 \sim o_1) \succ o_5 \succ (o_8 \sim o_{12}) \succ o_{11} \succ (o_6 \sim o_4) \succ (o_{10} \sim o_Z \sim o_3)$	Type-t	Yes	3	Yes	$(o_M \sim o_9) \succ (o_2 \sim o_7 \sim o_1) \succ \{o_Z \Vert o_3 \Vert o_4 \Vert o_5 \Vert o_6 \Vert o_8 \Vert o_{10} \Vert o_{11} \Vert o_{12}\}$
j_9	$(o_M \sim o_{11} \sim o_{12}) \succ o_9 \succ (o_1 \sim o_6 \sim o_7) \succ o_5 \succ (o_2 \sim o_3) \succ o_4 \succ o_8 \succ (o_Z \sim o_{10})$	Type-t	Yes	3	Yes	$(o_M \sim o_{11} \sim o_{12}) \succ o_9 \succ \{o_Z \Vert o_1 \Vert o_2 \Vert o_3 \Vert o_4 \Vert o_5 \Vert o_6 \Vert o_7 \Vert o_8 \Vert o_{10}\}$
j_{10}	$(o_M \sim o_{12}) \succ (o_{11} \sim o_5) \succ o_9 \succ o_7 \succ (o_4 \sim o_2) \succ (o_8 \sim o_6) \succ o_1 \succ (o_{10} \sim o_Z \sim o_3)$	Type-$t\&b$	No	1	No	$\{o_M \Vert o_{12}\} \succ \{o_1 \Vert o_2 \Vert o_4 \Vert o_5 \Vert o_6 \Vert o_7 \Vert o_8 \Vert o_9 \Vert o_{11}\} \succ \{o_Z \Vert o_3 \Vert o_{10}\}$
j_{11}	$(o_M \sim o_{12}) \succ (o_7 \sim o_8 \sim o_1) \succ o_{11} \succ o_9 \succ o_2 \succ o_3 \succ (o_{10} \sim o_6 \sim o_4) \succ (o_5 \sim o_Z)$	Type-t	Yes	3	Yes	$(o_M \sim o_{12}) \succ (o_7 \sim o_8 \sim o_1) \succ \{o_Z \Vert o_2 \Vert o_3 \Vert o_4 \Vert o_5 \Vert o_6 \Vert o_9 \Vert o_{10} \Vert o_{11}\}$
j_{12}	$(o_M \sim o_8 \sim o_9 \sim o_{12}) \succ (o_{11} \sim o_2) \succ (o_1 \sim o_5) \succ o_6 \succ o_7 \succ (o_3 \sim o_{10}) \succ (o_Z \sim o_4)$	Type-t	Yes	3	Yes	$(o_M \sim o_8 \sim o_9 \sim o_{12}) \succ \{o_Z \Vert o_1 \Vert o_2 \Vert o_3 \Vert o_4 \Vert o_5 \Vert o_6 \Vert o_7 \Vert o_{10} \Vert o_{11}\}$
j_{13}	$o_M \succ o_8 \succ (o_1 \sim o_{12}) \succ o_9 \succ o_2 \succ (o_{11} \sim o_5) \succ o_7 \succ o_{10} \succ (o_Z \sim o_6 \sim o_3 \sim o_4)$	Type-$t\&b$	Yes	3	No	$\{o_M \Vert o_1 \Vert o_8 \Vert o_{12}\} \succ \{o_2 \Vert o_5 \Vert o_7 \Vert o_9 \Vert o_{10} \Vert o_{11}\} \succ \{o_Z \Vert o_3 \Vert o_4 \Vert o_6\}$
j_{14}	$(o_M \sim o_1) \succ o_8 \succ o_{12} \succ (o_2 \sim o_5) \succ (o_9 \sim o_6) \succ (o_{11} \sim o_4) \succ o_3 \succ o_7 \succ (o_Z \sim o_{10})$	Type-t	No	3	No	$\{o_M \Vert o_1 \Vert o_8 \Vert o_{12}\} \succ \{o_Z \Vert o_2 \Vert o_3 \Vert o_4 \Vert o_5 \Vert o_6 \Vert o_7 \Vert o_9 \Vert o_{10} \Vert o_{11}\}$
j_{15}	$o_M \succ o_{12} \succ (o_9 \sim o_2) \succ o_1 \succ (o_8 \sim o_6 \sim o_7) \succ (o_{11} \sim o_{10}) \succ o_5 \succ (o_4 \sim o_Z \sim o_3)$	Type-$t\&b$	No	3	No	$\{o_M \Vert o_2 \Vert o_9 \Vert o_{12}\} \succ \{o_1 \Vert o_6 \Vert o_7 \Vert o_8 \Vert o_{10} \Vert o_{11}\} \succ \{o_Z \Vert o_3 \Vert o_4 \Vert o_5\}$
j_{16}	$(o_M \sim o_8 \sim o_9) \succ o_{12} \succ o_7 \succ (o_4 \sim o_5) \succ o_6 \succ o_1 \succ o_{11} \succ o_3 \succ o_{10} \succ o_2 \succ o_Z$	Type-t	No	3	Yes	$\{o_M \Vert (o_8 \sim o_9)\} \succ o_{12} \succ \{o_Z \Vert o_1 \Vert o_2 \Vert o_3 \Vert o_4 \Vert o_5 \Vert o_6 \Vert o_7 \Vert o_{10} \Vert o_{11}\}$
j_{17}	$o_M \succ o_{11} \succ o_2 \succ (o_4 \sim o_8 \sim o_1) \succ (o_5 \sim o_{12}) \succ (o_3 \sim o_9) \succ o_7 \succ o_{10} \succ (o_Z \sim o_6)$	Type-t	No	2	Yes	$\{o_M \Vert o_{11}\} \succ o_2 \succ \{o_Z \Vert o_1 \Vert o_3 \Vert o_4 \Vert o_5 \Vert o_6 \Vert o_7 \Vert o_8 \Vert o_9 \Vert o_{10} \Vert o_{12}\}$
j_{18}	$(o_M \sim o_9 \sim o_{12}) \succ o_1 \succ o_2 \succ (o_6 \sim o_8) \succ o_7 \succ (o_{11} \sim o_5) \succ o_4 \succ (o_{10} \sim o_Z \sim o_3)$	Quasi-complete	No	N/A	N/A	$\{o_M \Vert (o_9 \sim o_{12})\} \succ o_1 \succ o_2 \succ (o_6 \sim o_8) \succ o_7 \succ (o_{11} \sim o_5) \succ o_4 \succ \{o_Z \Vert (o_{10} \sim o_3)\}$
j_{19}	$(o_M \sim o_8 \sim o_{12}) \succ o_2 \succ o_7 \succ (o_{11} \sim o_1) \succ o_9 \succ (o_6 \sim o_4) \succ o_3 \succ (o_{10} \sim o_Z \sim o_5)$	Type-t	No	3	Yes	$\{o_M \Vert (o_8 \sim o_{12})\} \succ o_2 \succ \{o_Z \Vert o_1 \Vert o_3 \Vert o_4 \Vert o_5 \Vert o_6 \Vert o_7 \Vert o_9 \Vert o_{10} \Vert o_{11}\}$
j_{20}	$(o_M \sim o_9 \sim o_{11}) \succ o_{12} \succ o_8 \succ o_1 \succ o_5 \succ (o_2 \sim o_7) \succ o_4 \succ (o_6 \sim o_{10} \sim o_Z \sim o_3)$	Type-$t\&b$	No	4	No	$\{o_M \Vert o_8 \Vert o_9 \Vert o_{11} \Vert o_{12}\} \succ \{o_1 \Vert o_2 \Vert o_5 \Vert o_7\} \succ \{o_Z \Vert o_3 \Vert o_4 \Vert o_6 \Vert o_{10}\}$

Curiously enough, the value of $\bar{\varepsilon}$ for the incomplete problem (i.e., 7.98) is relatively close to the value that would be expected through the application of the experimental regression model in Eq. 16:

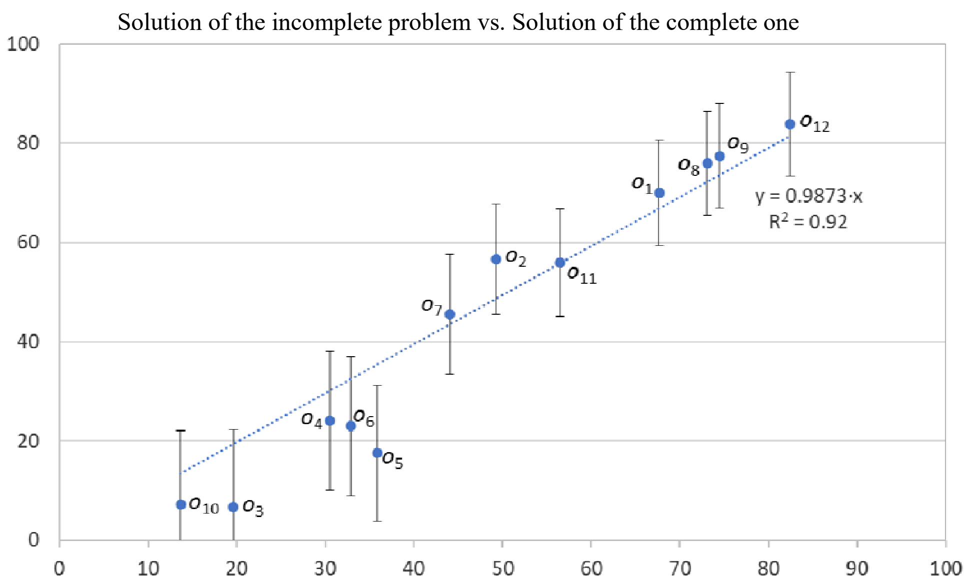
$$
\begin{aligned}
&\bar{\varepsilon}\left(\bar{c} = 62.2\%, W = 69.9\%\right) = \\
&= 43.1 - 103.4 \cdot (62.2\%) + 22.4 \cdot (69.9\%) + 65.4 \cdot (62.2\%)^2 + 0.00118 \cdot (69.9\%)^2 - 37.9 \cdot (62.2\%) \cdot (69.9\%) \cdot \\
&\cong 3.27
\end{aligned}
\tag{A.1}
$$

The above result represents a sort of plausibility check of the model in Eq. 16.



	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	o_9	o_{10}	o_{11}	o_{12}
$\hat{y}_i$	70.0	56.6	6.7	24.1	17.5	23.0	45.5	75.9	77.4	7.2	55.9	83.9
$U_{\hat{y}_i}$	10.7	11.2	15.6	14.0	13.7	14.1	12.1	10.5	10.6	14.8	10.8	10.5

Figure A.1. Graphic representation of the solution of the incomplete problem in Table A.2 (last column).



	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	o_9	o_{10}	o_{11}	o_{12}
Compl. problem	67.7	49.3	19.6	30.5	35.8	32.9	44.1	73.1	74.5	13.6	56.5	82.5
Incompl. problem	70.0	56.6	6.7	24.1	17.5	23.0	45.5	75.9	77.4	7.2	55.9	83.9

Figure A.2. Comparison between the solution of the incomplete problem and that of the complete problem in Table A.2. Error bands represent the expanded uncertainty associated with the scaling of the incomplete problem. The complete problem is characterized by $W = 69.9\%$, while the incomplete problem by $\bar{c} = 62.2\%$ and $\bar{\varepsilon} = 7.98$.