

A proposal for an operational methodology to assist the rankingaggregation problem in manufacturing

*Original*

A proposal for an operational methodology to assist the rankingaggregation problem in manufacturing / Franceschini, F., Maisano, D., Mastrogiacomo, L.. - In: RESEARCH IN ENGINEERING DESIGN. - ISSN 0934-9839. - STAMPA. - 35:4(2024), pp. 373-388. [10.1007/s00163-024-00437-7]

*Availability:*

This version is available at: 11583/2993536 since: 2024-10-19T08:03:19Z

*Publisher:*

Springer

*Published*

DOI:10.1007/s00163-024-00437-7

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)



# A proposal for an operational methodology to assist the ranking-aggregation problem in manufacturing

Fiorenzo Franceschini<sup>1</sup> · Domenico A. Maisano<sup>1</sup> · Luca Mastrogiacomo<sup>1</sup>

Received: 19 March 2023 / Accepted: 28 June 2024 / Published online: 15 July 2024  
© The Author(s) 2024

## Abstract

Ranking aggregation is an ancient problem with some characteristic elements: a number of *experts*, who individually rank a set of *objects* according to a certain (subjective) *attribute*, and the need to aggregate the resulting *expert rankings* into a *collective judgment*. Although this problem is traditionally very popular in fields such as *social choice*, *psychometrics*, and *economics*, it can also have several interesting applications in *manufacturing*, e.g., for customer-oriented design, reliability engineering, production management, etc. Through a case study related to a cobot-assisted manual (dis)assembly, the paper illustrates an operational methodology and various useful tools that assist in tackling the problem practically, effectively, and with a critical mind. Some of the proposed tools allow to estimate the degree of *concordance* among experts, and the collective judgment's *consistency* and *robustness*. The paper is aimed at scientists and practitioners in manufacturing.

**Keywords** Ranking aggregation · Expert ranking · Collective judgment · Manufacturing · Ranking concordance · Consistency analysis

## 1 Introduction

Suppose you are at the helm of a small manufacturing company looking for improvements in the product packaging process. The old packaging machines are no longer up to the task, while there are five new top models on the market. The challenge? Choosing the best one, with a large investment at stake—in the order of hundreds of thousands of euros—and a major impact on the company's operations. To solve this complex decision, four experienced engineers and technicians provide assessments based on their technical expertise and (not always very in-depth) information on the five models; based on his/her own perspective, everyone formulates a preference *ranking* among the five models (e.g., the third model is preferred to the first, which in turn is preferred to

the fourth, and so on). But how to combine these individual evaluations for a collective decision on the most suitable packaging machine? This is the challenge of the so-called *ranking-aggregation problem*, in which data and science are used to guide towards a well-informed choice. Specifically, this ancient and widespread problem has three characteristic elements (Spohn 2009; Reich 2010; Saari 2011):

1. A set of *objects* to be prioritised according to a *subjective attribute*, i.e., a feature whose perception may depend on the person perceiving the stimulus, his/her technical knowledge and personal taste.
2. A set of *experts* formulating preference rankings of the objects of interest. Experts may be regarded as equally important or with a hierarchy of importance, depending on their competence regarding the evaluation they are supposed to carry out.
3. A *collective judgement* concerning the objects, resulting from the aggregation of expert rankings through a suitable *aggregation technique*. In the scientific literature, depending on the field and historical period, one can also encounter alternative expressions such as “*collective/consensus judgement/assessment/evaluation*”, which can, however, be considered interchangeable.

✉ Fiorenzo Franceschini  
fiorenzo.franceschini@polito.it  
Domenico A. Maisano  
domenico.maisano@polito.it  
Luca Mastrogiacomo  
luca.mastrogiacomo@polito.it

<sup>1</sup> DIGEP (Dept. of Management and Production Engineering), Politecnico di Torino, Corso Duca Degli Abruzzi 24, 10129 Torino, Italy

Traditional fields in which the problem is very popular are *social choice*, *psychometrics*, *economics*, and *multi-criteria decision making* (MCDM), with relevant contributions from eminent scientists (e.g., de Borda, Pareto, Samuelson, Arrow, Thurstone, Kendall, etc.) (Spohn 2009; Saari 2011; Arrow 2012; Bana e Costa, 2012; Köksalan et al. 2013; Franceschini et al. 2022). For instance, despite differences in terminology and application context, typical MCDM applications share the theoretical and methodological foundations of the ranking-aggregation problem. They begin with a finite number of *alternatives* (analogous to *objects*), each represented by its performance across multiple *criteria* (analogous to *experts*, though not necessarily flesh-and-blooded subjects). The objective is to identify the *best alternative(s)*, akin to achieving a *collective judgment* (Belton and Stewart 2002; Zeleny 1976).

Due to the great generality, transversal disciplinary nature and multiplicity of potential applications, the ranking-aggregation problem has become of interest to many other scientific disciplines and operational contexts, including *manufacturing*. Some of the many possible manufacturing applications are as follows:

- *Production management*, regarding the selection of the most appropriate production system on the basis of productivity, flexibility or other performance attributes (Chuu 2009; Chatterjee and Chakraborty 2014; Nestic et al. 2019; Hakimi-Asl et al. 2018; Qin et al. 2020).
- *Procurement*, in which some managers have to identify the most appropriate suppliers or materials for a certain manufacturing system (Giachetti 1998; Yu and Hou 2016).
- *Conceptual design*, regarding the opinions of different designers about alternative design concepts, from the perspective of specific technical features (Franceschini and Maisano 2019);
- *Quality control*, regarding the prioritization of defects on manufactured parts, aggregating expert judgments by visual inspection (Franceschini and Maisano 2018b);
- *Reliability engineering*, regarding the aggregation of the opinions of maintenance/reliability experts on the criticality of the (potential) failures in a production equipment (Geramian et al. 2019);
- *Customer-driven design*, regarding the opinions of a panel of (service/product) customers on the degree of importance of a set of customer needs (Nahm et al. 2013);
- *Analysis of market demand*, regarding the opinions by marketing experts about the most appropriate actions for the promotion of a new product/service (Franceschini and Maisano 2018a).

The analyst's focus is often directed to the aggregation technique, which can be interpreted as a "black box" transforming *input* data (i.e., experts' rankings and importance hierarchy) into *output* data (i.e., collective judgement) (Franceschini et al. 2022). However, this may lead to overlooking other important methodological aspects that characterise the ranking-aggregation problem, such as preliminary assessment of the degree of concordance among experts, verification of the consistency and robustness of output data, etc.

Aimed at scientists and practitioners in the manufacturing field, this work provides a set of useful tools to tackle the ranking-aggregation problem in a practical and effective manner, addressing the following research question: "*How can the ranking-aggregation problem be effectively handled in the manufacturing field and what methodologies and tools can enhance the plausibility and robustness of the solution obtained through the expert-ranking aggregation?*". It is hypothesized that manufacturing scientists are often unfamiliar with the problem of interest, even though they may occasionally be dealing with it. Therefore, the article tries to bridge this knowledge gap by providing a relatively straightforward and effective operational methodology. The innovative aspect of this work lies in the integration of tools, within the proposed methodology, which are individually available in the scientific literature but are combined here in an organic manner. Furthermore, the proposed methodology is flexible, adapting to problems with different characteristics and incorporating different tools interchangeably. It is also iterative, including intermediate verifications that allow for adjustments and corrections while addressing the ranking-aggregation problem.

The remainder of this work is organised in three sections. Section 2 briefly introduces a real-world case study, concerning cobot-assisted manual (dis)assembly, which accompanies the description of the proposed methodology. Section 3, which is the heart of the article, provides a step-by-step description of the operational assisted methodology based on three phases, namely (i) *problem formulation*, (ii) *collection of expert rankings*, and (iii) *collective judgment and validation*. Section 4 summarises the original contributions of this work, its implications, limitations and insights for future research.

## 2 Case study

A company in the automotive industry reconditions different types of electrical components, mainly starter motors and alternators. Although the operations required are mostly manual and specific to each component, they can be divided into the following groups:

### Disassembly

- Disassembling any external coverings and shells (to access the internal parts);
- Removal of electrical connectors and cables;
- Unfastening bolts, screws, and other fasteners;
- Separation of any electronic circuits (from the motherboard or main body);
- Extraction of internal components (sensors, relays, transistors, capacitors, diodes, etc.).

### Reconditioning

- Identification of parts to be replaced or repaired;
- Repairing/replacing these parts;
- Intermediate testing.

### Reassembly

- Mounting internal components: repaired or replaced in their respective housings;
- Reconnecting electrical cables and connectors;
- Fastening with bolts, screws, or other fastening elements;
- Ensuring that electrical connections are securely fastened;
- Reassembling external shells and coverings;
- Testing and diagnostics to verify the proper functioning of the reconditioned unit;
- Cleaning, polishing, and final marking.

Because of the wide variety of components and the complexity of (dis)assembly and repair operations, the company has been assisting human operators with *collaborative robots*, or simply *cobots* (see Fig. 1), which are particularly useful for assisting manual operations that require great precision, dexterity and strength (Gervasi et al. 2022). Cobots are extremely versatile for multiple tasks, such as (i) picking up, clamping, handing the tools and parts to be machined/assembled, (ii) supporting dimensional inspection, online quality control, etc., and (iii) guiding less experienced operators, like virtual tutors.

The current market includes a relatively wide range of cobot models, which could be adapted to the context of interest. The company management decided to identify the most appropriate cobot depending on the *programming-practicality* attribute, which is crucial in making task preparation faster and easier, while reducing the level of technical skills required by operators (El Zaatari et al. 2019). The following five cobot models were selected from those



**Fig. 1** Cobot-supported operator for a manual assembly task

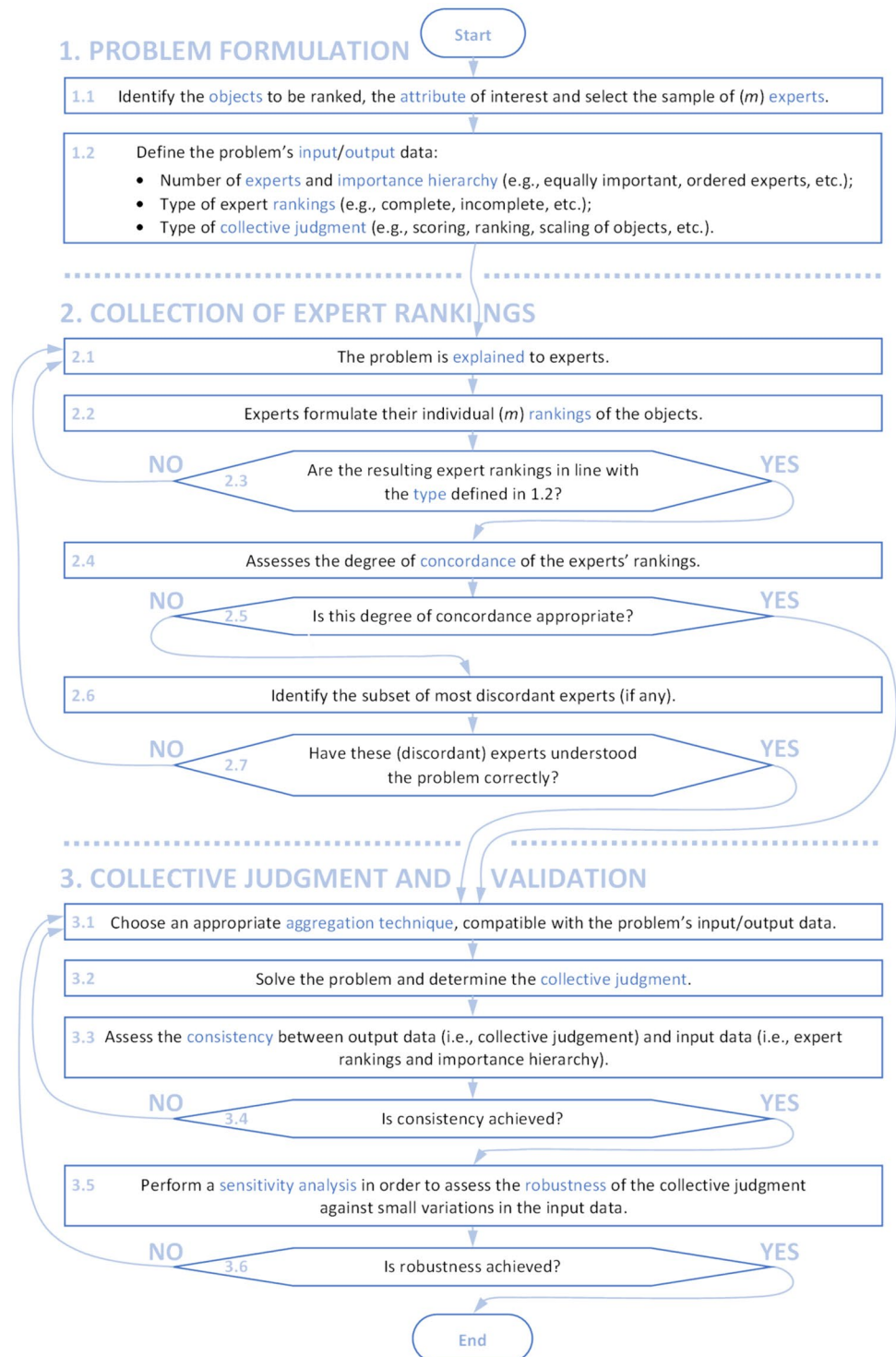
at the forefront of the market, as they all (i) have a similar payload (around 5–10 kg), (ii) are designed for precision assembly and machining applications, and (iii) are relatively cost-effective:

- (o<sub>1</sub>) Techman Robot TM5-700;
- (o<sub>2</sub>) ABB GoFa10;
- (o<sub>3</sub>) Universal Robots UR10E;
- (o<sub>4</sub>) Yaskawa Motoman HC10DTP Classic;
- (o<sub>5</sub>) Kinova Link 6.

In order to carry out a comprehensive evaluation, the company set up a panel of eight experts (mostly engineers, technicians and external consultants) from different technical areas and with diverse and complementary skills, a brief description of which follows:

- (e<sub>1</sub>) *Industrial-automation expert* with in-depth skills in industrial process design and optimization;
- (e<sub>2</sub>) *Electrical engineer* with comprehensive knowledge of electrical components and the technical specifications required for their safe assembly and disassembly;
- (e<sub>3</sub>) *Artificial-vision specialist* capable of integrating advanced vision systems onto cobots, for precise recognition and positioning of electrical components;
- (e<sub>4</sub>) *Ergonomics expert* with skills to define ergonomic and intuitive interaction modes with cobots for operators;
- (e<sub>5</sub>) *Robot-programming specialist* with significant experience in both traditional industrial robots and collaborative robots;
- (e<sub>6</sub>) *Workplace-safety expert* with in-depth knowledge of safety regulations and protocols for safe human-machine collaboration;

**Fig. 2** Flow chart summarising the assisted operational methodology for ranking aggregation



( $e_7$ ) *Maintenance expert* with skills for planning and managing preventive and corrective maintenance activities on cobots;

( $e_8$ ) *Quality engineer* with relevant experience to ensure that the robot-assisted assembly/disassembly process complies with quality standards.

### 3 Assisted operational methodology

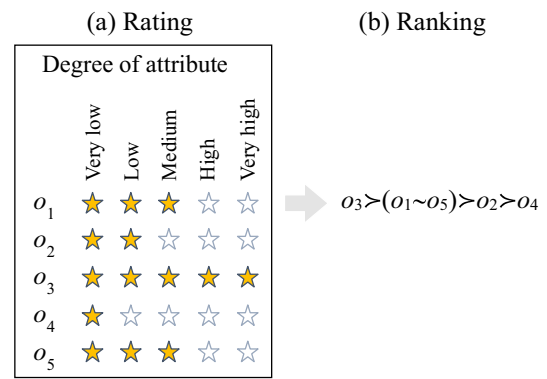
This section illustrates an assisted operational methodology for tackling the ranking-aggregation problem in a practical, comprehensive and critical manner. The flowchart in Fig. 2 summarises the proposed methodology, which is divided into three operational phases illustrated in the corresponding subsections: “problem formulation”, “collection of expert rankings” and “collective judgment and validation”. The multiple feedback loops denote the iterative nature of the proposed procedure, which includes several intermediate verifications, with possible in-progress corrections and adjustments.

#### 3.1 Problem formulation

First, the specific problem and its characteristics should be identified clearly and unambiguously. Based on the above considerations, a specific ranking-aggregation problem can be formulated. With reference to the case study, the cobot models are the  $n=5$  objects ( $o_1$ – $o_5$ , cf. Section 2) that will be evaluated in terms of *programming practicality*, i.e., the *attribute* of interest. This attribute encompasses a range of *desiderata*, many of which are related to subjective perceptions, as detailed below:

- An *intuitive and easy-to-learn programming language* will reduce development time and programming errors.
- The user interface of the *teach pendant* should be intuitive and user-friendly to simplify and speed up the programming and control phase of the cobot.
- It would be desirable to be able to programme and simulate the behaviour of the cobot even *offline*, without necessarily being connected to it.
- The cobot should be integrable with *external sensors* (such as cameras, force sensors, etc.), so as to be more versatile for complex tasks.
- The cobot programming should include *advanced safety features* to avoid accidents and ensure safe collaboration between the cobot and human operators.
- Some cobots support the use of *third-party programming languages*, such as Python or C++, making the import of external routines more versatile.
- *Tutorials, documentation and technical support* should help make operator learning quicker and easier.

As seen in Sect. 2, the  $m=8$  experts ( $e_1$ – $e_8$ ) are technicians, engineers and external consultants who formulate their individual preference rankings of the cobot models. In general, when selecting experts (at least) two aspects must be taken into account:



**Fig. 3** Example of the conversion of judgments (on five objects:  $o_1$ – $o_5$ ) from **a** a five-level *rating* scale to **b** a (complete) ranking (Franceschini et al. 2022). The procedure was applied to expert  $e_1$  and can be extended to the other seven experts

1. The greater the number of experts formulating their individual rankings, the higher the statistical relevance of the problem output (Friedman 1940; Kendall 1962; Gibbons and Chakraborti 2010). Unfortunately, there may be practical constraints that limit the availability of the number of experts (e.g., they should have a high level of technical expertise). Pragmatically, it would be desirable for  $m$  to be no less than 5–6 in order for the results of the study to be relevant (Franceschini et al. 2022).
2. It may sometimes be appropriate to have a hierarchy of importance of experts, for instance by discriminating those with greater technical expertise. This hierarchy can be constructed in different ways, typically by associating each expert with a *weight* or defining an *importance ranking* (Gibbons and Chakraborti 2010; Leo Kumar 2019). In the case study, the technical competences of the experts are notably different and, at the same time, complementary, with no clear superiority of one over the other (cf. Sect. 2). For this reason, all these experts are regarded as equally important. From a practical point of view, this choice simplifies the handling of the problem and broadens the range of applicable aggregation techniques (cf. Sect. 3.3).

Next, the type of expert rankings can be determined depending on several factors, such as the *goal* of the problem (e.g., identifying the best/worst object(s), drawing up a complete ranking, etc.), the *data-collection* strategy (e.g., through focus groups, personal telephone/street interviews, online forms, etc.), the *literacy level* of experts, etc. *Complete* rankings—i.e., ranking in which experts order all objects by linking them with *strict preference* (“ $o_i > o_j$ ”)

**Fig. 4** **a** Complete rankings of  $n=5$  objects, formulated by  $m=8$  experts; **b** corresponding rank table.  $T_i$  is a correction factor for ties (cf. Eq. 1)

		(a) Rankings					(b) Rank table					Row totals	$T_i$
		$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$o_1$	$o_2$	$o_3$	$o_4$	$o_5$		
Experts	$e_1$	$o_3 > (o_1 \sim o_5) > o_2 > o_4$					2.5	4.0	1.0	5.0	2.5	15	6
	$e_2$	$o_5 > o_1 > (o_2 \sim o_3 \sim o_4)$					2.0	4.0	4.0	4.0	1.0	15	24
	$e_3$	$(o_1 \sim o_3) > o_2 > (o_4 \sim o_5)$					1.5	3.0	1.5	4.5	4.5	15	12
	$e_4$	$o_1 > o_5 > o_3 > o_2 > o_4$					1.0	4.0	3.0	5.0	2.0	15	-
	$e_5$	$o_4 > (o_1 \sim o_2 \sim o_5) > o_3$					3.0	3.0	5.0	1.0	3.0	15	24
	$e_6$	$o_5 > o_4 > (o_1 \sim o_3) > o_2$					3.5	5.0	3.5	2.0	1.0	15	6
	$e_7$	$o_3 > o_5 > o_4 > o_2 > o_1$					5.0	4.0	1.0	3.0	2.0	15	-
	$e_8$	$(o_3 \sim o_5) > (o_1 \sim o_2) > o_4$					3.5	3.5	1.5	5.0	1.5	15	12
Col. totals ( $R_j$ )		22.0	30.5	21.5	29.5	17.5	120						

or *indifference* relationships (“ $o_i \sim o_j$ ”)—represent a classic scenario, although their formulation requires some effort, especially if the number of objects is large (Lagerspetz 2016). On the other hand, *incomplete* rankings are more “digestible” for experts, because they can take into account possible hesitations or doubts; for instance, incomplete are those rankings in which only a small number of top or bottom objects are included (e.g., the three most/least preferred), or in which the expert decides to omit an object from his/her ranking (e.g., since he/she is not familiar with), or even rankings with *incomparability* relationships between objects (“ $o_i \parallel o_j$ ”) (Chen et al. 2012). Given the relatively small number of objects, in the present case experts are asked to formulate complete rankings of all five objects. In Sect. 3.2 we will illustrate a way to indirectly formulate complete rankings, through a simplified response mode.

Subsequently, the collective-judgment type must be defined according to the “desirable” properties for the specific problem. There is a wide range of possibilities: *rankings*, *scalings* on different scale types (e.g., *interval*, *ratio*), *clusterings*, *scorings*,<sup>1</sup> or collective judgments designating only the winner/loser object, etc. For the sake of simplicity, in the case study the expected collective judgment is represented by a *complete* ranking. The analyst must be aware that the choice of input/output data for the problem has implications for both the subsequent formulation of rankings (in Sect. 3.2) and the choice of aggregation technique (in Sect. 3.3).

### 3.2 Collection of expert rankings

This stage begins with a detailed explanation of the problem to experts, who need to understand exactly which objects are to be evaluated, the attribute against which the evaluation is to be made, and how to formulate individual rankings.

<sup>1</sup> A *scoring* can be seen as the synthesis of multiple rankings through scores associated with objects, according to a conventional rule. The scoring is not necessarily a measurement of the degree of the attribute of an object (Fransceschini et al. 2022).

In order to make this formulation less laborious, especially when the number of objects being compared is large, experts can formulate ratings<sup>2</sup> of the objects, which can then be converted into a complete ranking (see example in Fig. 3).

Returning to the case study, Fig. 4 reports the resulting (complete) expert rankings, which include relationships of strict preference (“ $o_i > o_j$ ”) and indifference (“ $o_i \sim o_j$ ”) between objects. At this stage, it must be ensured that the experts’ rankings are formulated consistently with the expected type; if necessary, the formulation must be corrected/revised (see feedback loop from block 2.3 in Fig. 2).

Going into the rankings of the eight experts in the case study, it should come as no surprise that they sometimes differ from each other because they are often based on complementary perspectives. For example, experts  $e_4$  and  $e_7$  seem to express two radically different evaluations of item  $o_1$ . Probably influenced by their different backgrounds and training, these experts have developed very different perceptions of the cobot model  $o_1$ , resulting in evaluations in opposite directions.

#### 3.2.1 Concordance among expert rankings

Evaluating the *concordance*<sup>3</sup> among expert rankings is a preliminary check of the plausibility of input data, which is useful to prevent difficulties, such as excessive heterogeneity in the selection of experts, poor understanding of the problem, errors in the formulation of rankings, or other potential obstacles to achieving consensus. The scientific literature includes various statistical indicators, which can be used depending on the problem characteristics (Agresti 2010; Gibbons and Chakraborti 2010; Sato and Tan 2023). Since the present case is characterized by complete expert

<sup>2</sup> A *rating* is defined as a set of categories designed to elicit information about an attribute. In social sciences and psychology, a common example is the 5-level Likert response scale, in which an expert selects the number in the ordinal category that is believed to reflect the perceived attribute of an object.

<sup>3</sup> In statistics, the notion of *concordance* is also referred to as *degree of association* (Ross 2009).

rankings with equally-important experts, the Kendall’s  $W$  and Spearman’s  $\rho$  can be used (Franceschini et al. 2022).

$W$ , known as *coefficient of concordance*, is a *multivariate* statistic that applies at the level of expert rankings and is related to the dispersion of the ranks associated with each object (Ross 2009; Franceschini et al. 2022). This measure belongs to the range  $[0, 1]$ , where 1 indicates perfect concordance and 0 indicates independence (Legendre 2010).

Returning to the case study, each ranking can be translated into a set of ranks—that is, permutations of the integers  $\{1, 2, 3, 4, 5\}$ —which are then organized into a so-called *rank table*, i.e., a bidirectional matrix of size  $m \times n$ , with row and column labels designating experts and objects (see Fig. 4b). In the case of tied objects—i.e., pairs of objects with *indifference* relationships, e.g., “ $o_i \sim o_j$ ”—we conventionally use the *average ranks* that each set of bound objects would occupy if a preference could be expressed (Gibbons and Chakraborti 2010); for example, in a ranking where objects  $o_1$  and  $o_3$  are tied for 3rd and 4th place (e.g., see the ranking by  $e_6$  in Fig. 4a), the average rank of  $(3 + 4)/2 = 3.5$  would be assigned to both.

$W$  is defined as:

$$W = \frac{\sum_{j=1}^n (R_j - \bar{R})^2}{[m^2 \cdot n \cdot (n^2 - 1) - m \cdot \sum_{i=1}^m T_i] / 12}, \tag{1}$$

being

$n$  the number of objects (i.e., 5 here);

$m$  the number of experts (i.e., 8 here);

$R_j$  the column total related to the  $j$ -th column of the rank table;

$\bar{R} = m \cdot (n + 1)/2$  the average column total (i.e., 24 here);

$T_i = \sum_{k=1}^{g_i} (t_k^3 - t_k)$  a correction factor for ties, in which  $t_k$  is the number of tied ranks in the  $k$ -th group of tied ranks (where a group is a set of values having constant tied rank) and  $g_i$  is the number of groups of ties in the set of ranks (ranging from 1 to  $n$ ) for expert  $i$ . This correction factor ensures that, in the case of perfectly concordant rankings with ties (as all rankings coincide),  $W = 1$  (or 100%) is obtained (Gibbons and Chakraborti 2010).

With reference to the case study (cf. expert rankings and related object ranks in Fig. 4), it is obtained  $W = 22.4\%$ , denoting a relatively low level of concordance. Not surprisingly, a significance test to the *null hypothesis of independence between rankings* yields the following parameter:

$$Q = W \cdot m \cdot (n - 1) = 7.2 < \chi_{n-1,\alpha}^2 = 9.49, \tag{2}$$

$\chi_{n-1,\alpha}^2 = \chi_{4,5\%}^2$  being a chi-square ( $\chi^2$ ) variable with  $n - 1$  degrees of freedom, corresponding to a conventional significance level of  $\alpha = 5\%$ . Equation (2) indicates that the null hypothesis cannot be rejected with a confidence level of  $1 - \alpha = 95\%$  (Ross 2009; Gibbons and Chakraborti 2010).

To further investigate the reasons for this low inter-expert concordance, the *bivariate* perspective of Spearman’s correlation coefficient<sup>4</sup> related to each possible pair of rankings ( $\rho$ ) can be considered. Table 1 contains the  $\rho$  coefficients between all the possible  $\binom{m}{2} = \frac{m \cdot (m-1)}{2} = 28$  pairs of expert rankings under consideration (Ross 2009).

Rather pronounced negative correlations (i.e.,  $\rho \leq -0.4$ ) between certain pairs of expert rankings stand out. Curiously, they often involve the ranking by  $e_5$ , denoting a sort of “countertrend” with respect to the other rankings. Upon brief investigation of the reasons for this counter-trend, it turns out that  $e_5$  misunderstood the ranking construction, formulating it in the sense of reverse preference; therefore, the correct ranking should be “ $o_3 \succ (o_1 \sim o_2 \sim o_5) \succ o_4$ ” instead of “ $o_4 \succ (o_1 \sim o_2 \sim o_5) \succ o_3$ ” (see feedback loop from block 2.7 in Fig. 2). After this correction, the new value of  $W$  is almost twice as high as the initial one (i.e.,  $W = 40.2\%$  versus 22.4%) and the significance test in Eq. (2) results into  $Q = 12.9 \geq \chi_{n-1,\alpha}^2 = 9.49$ , which leads to rejecting the null hypothesis and considering the new level of concordance as statistically significant. Simultaneously, the relatively large negative  $\rho$  values for  $e_5$  are “reabsorbed” (see Table 2, containing the new  $\rho$  values).

As exemplified, the concordance analysis can be useful in pointing out possible anomalies and “pitfalls” in the formulation of expert rankings (Franceschini et al. 2022).

### 3.3 Collective judgment and validation

At this point, it is needed to solve the ranking-aggregation problem by utilizing an appropriate aggregation technique and, subsequently, verifying the plausibility of the resulting output.

#### 3.3.1 Ranking aggregation

This is the heart of the ranking-aggregation problem and implies some knowledge of the state-of-art aggregation techniques. Far from this ambition, Table 3 simply recalls some possible aspects to be taken into account while selecting the aggregation technique (Franceschini et al. 2022).

For an overview of the aggregation techniques, we refer the reader to relevant surveys and extensive reviews (Figueira et al. 2005; Reich 2010; Herrera-Viedma et al. 2014). For example, Table 4—adapted from (Franceschini et al. 2022)—classifies nine different aggregation techniques according to the aspects listed in Table 3. It can be noted that

<sup>4</sup>  $\rho$  belongs to  $[-1, 1]$ ; a perfect  $\rho$  correlation of  $+1$  or  $-1$  occurs when the first set of rank values is a perfect monotone (respectively increasing or decreasing) function of the second one (Ross 2009).

**Table 1** Spearman’s  $\rho$  correlation table for the expert rankings in Fig. 4a

Ranking	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$
$e_1$	1							
$e_2$	0.287	1						
$e_3$	0.649	-0.177	1					
$e_4$	0.667	0.783	0.474	1				
$e_5$	<b>-0.918</b>	0.000	<b>-0.707</b>	<b>-0.447</b>	1			
$e_6$	-0.026	0.574	<b>-0.649</b>	0.103	0.344	1		
$e_7$	0.462	-0.112	-0.158	-0.200	<b>-0.447</b>	0.410	1	
$e_8$	0.865	0.412	0.250	0.527	<b>-0.825</b>	0.162	0.632	1

The most pronounced negative correlations (i.e., those with  $\rho \leq -0.4$ ) are bolded.

**Table 2** Spearman’s  $\rho$  correlation table for the expert rankings in Fig. 4, after the correction of the ranking by  $e_5$  (i.e., “ $o_3 > (o_1 \sim o_2 \sim o_5) > o_4$ ” instead of “ $o_4 > (o_1 \sim o_2 \sim o_5) > o_3$ ”)

Ranking	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$
$e_1$	1							
$e_2$	0.287	1						
$e_3$	0.649	-0.177	1					
$e_4$	0.667	0.783	0.474	1				
$e_5$	0.918	0.000	0.707	0.447	1			
$e_6$	-0.026	0.574	-0.649	0.103	-0.344	1		
$e_7$	0.462	-0.112	-0.158	-0.200	0.447	0.410	1	
$e_8$	0.865	0.412	0.250	0.527	0.825	0.162	0.632	1

The fact that most of the correlation coefficients are positive and the negative ones are relatively few and with relatively low values confirms the rather acceptable overall level of concordance.

**Table 3** Aspects to consider when selecting the aggregation technique, with reference to a specific ranking-aggregation problem (Franceschini et al. 2022)

(a) Input-data characteristics	(b) Aggregation mechanism	(c) Output-data characteristics
<ul style="list-style-type: none"> <li>• Problem size:                             <ul style="list-style-type: none"> <li>- Number of objects (<math>n</math>);</li> <li>- Number of expert rankings (<math>m</math>)</li> </ul> </li> <li>• Expert rankings:                             <ul style="list-style-type: none"> <li>- Complete;</li> <li>- Incomplete</li> </ul> </li> <li>• Expert hierarchy:                             <ul style="list-style-type: none"> <li>- Equally-important experts;</li> <li>- Importance weights;</li> <li>- Importance ranking</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Rule-based</li> <li>• Optimization-based;</li> <li>• Distribution-based</li> </ul>	<ul style="list-style-type: none"> <li>• Designation of a unique winner/loser;</li> <li>• Complete/incomplete ranking;</li> <li>• Classification in categories;</li> <li>• Collective scoring of objects;</li> <li>• Collective scaling of objects</li> <li>• ...</li> </ul>

some techniques are suited to situations with few objects/experts, while others—which can be defined as more “parsimonious” (Kabirifar et al. 2023; Corrente et al. 2024)—are also suitable for situations with a relatively large number of objects/experts. The summary in Table 4 is evidently partial and not intended to be comprehensive. In future research, we aim to provide a more comprehensive overview in this regard. Here we just point out that (i) aggregation techniques are all inherently imperfect (Arrow 2012), (ii) their success depends not only on their efficacy, accuracy, and scientific rigour but also on their simplicity of use (Oukil 2019; Sarwar et al. 2021), and (iii) in general it would be good to avoid “falling in love” with one technique and—when

possible—use multiple techniques simultaneously (cf. concept of *wisdom of crowds*) (Franceschini et al. 2022).

In line with this consideration, two relatively simple aggregation techniques are applied for the problem of interest:

- Borda count (BC). For each expert ranking, the first object accumulates one point, the second two points, and so on (Borda 1781; Saari 2011). In case of ties, the *average ranks* described in Sect. 3.2 can be used. The collective score (BC) of one object can be calculated by cumulating the scores related to each ranking; in this sense, the BC method implements the concept of

**Table 4** Synthetic comparison among nine aggregation techniques illustrated in (Franceschini et al. 2022), according to the aspects in Table 3. The first and fourth techniques will be used for the case study

Aggregation technique	(a) Input-data characteristics			(b) Aggreg. mechan	(c) Output-data characteristics
	Problem size ( $n$ and $m$ )	Expert rankings	Expert hierarchy		
1. Best of the best (BoB)	$n$ and $m$ large as desired	Complete	Equally-important experts	Rule-based	Collective scoring of objects
2. Best two (BTW)					
3. Instant-Runoff Voting (IRV)					
4. Borda Count (BC)	$n$ and $m$ large as desired	Complete	Equally-important experts	Rule-based	Collective scoring of objects
5. ELECTRE-II	$n$ and $m$ of the order of a few tens at most	Complete	Importance weights of experts	Rule-based	Collective (complete or incomplete) ranking of objects
6. Yager’s algorithm (YA)	$m$ large as desired; $n$ of the order of a few tens at most	Complete	Rank-ordered experts	Rule-based	Collective complete ranking of objects
7. Enhanced Yager’s algorithm (EYA)	$m$ large as desired; $n$ of the order of a few tens at most	Incomplete	Rank-ordered experts	Rule-based	Collective complete ranking of objects
8. Thurstone’s law of comparative judgment (LCJ)	$m$ large as desired; $n$ of the order of a few tens at most	Complete	Equally-important experts	Distribution-based	Collective <i>interval</i> scaling of objects
9. $ZM_{II}$	$m$ large as desired; $n$ of the order of a few tens at most	Incomplete including "anchor" objects	Equally-important experts	Distribution-based	Collective <i>ratio</i> scaling of objects

“average rank position”. BC is used in various contexts, such as engineering design, the “RoboCup” robot soccer competition, the “Eurovision” song contest, etc. (Dym et al. 2002; Franceschini et al. 2022).

- Best of the best (BoB). For each expert ranking, the most preferred object obtains one point. In case of a tie between leading objects, the point is *fractionalized*, dividing it by the number of objects themselves (e.g., ½ if there are 2 objects, 1/3 if there are 3, and so on). In some contexts, the BoB method is also referred to as “Plurality Voting” or “First Past the Post” (Blais 2008).

Figure 5a, b respectively show the results of the application of the BC and BoB techniques to the expert rankings (after the correction of the ranking by  $e_5$ , cf. Section 2.2). These two aggregation techniques—which are simple and well suited to complete rankings by equally-important experts—here result in two similar collective rankings (see the bottom of Fig. 5). Both techniques lead to the same “trio” of most suitable cobot models:  $o_3$  (*Universal Robots UR10E*) followed by  $o_5$  (*Kinova Link 6*) and again  $o_1$  (*Techman Robot TM5-700*).

### 3.3.2 Consistency analysis

Every aggregation technique surely provides a result; but how does one know whether it is plausible? Certainly, the

rationale of the aggregation technique represents a conceptual guarantee that it is capable of producing reasonable results. However, the aggregation technique that most consistently reflects expert rankings cannot be assessed *ex ante*, but only *ex post* and on a case-by-case basis (Chiclana 2002; Arrow 2012; McComb et al. 2017).

Studies have focused on the concept of consistency of the collective judgment with respect to input data, defined as “*the ability of a collective judgment to reflect the rankings of experts, while taking the importance hierarchy into account*” (Franceschini et al. 2022). Among the available tools to assess the degree of consistency of the solution to a certain ranking-aggregation problem,  $p$ -indicators are very versatile, as they can be adapted to a variety of contexts, such as those in which expert rankings are (i) not necessarily complete, (ii) equally important, or (iii) characterized by an importance hierarchy (Franceschini et al. 2022). In general,  $p$ -indicators can be divided into two families:

- $p_j$  indicators of *local consistency*, which are based on the comparison of each  $j$ -th expert’s ranking with the collective judgement.

A preliminary operation for determining  $p_j$  is constructing “*a paired-comparison table*” in which each ranking (i.e., those from experts and that one deduced from the collective judgment) is transformed into sets of paired-comparison

**Fig. 5** **a** Expert rankings, **b** scoring/ranking resulting from the application of the *Borda Count* (BC), and **c** scoring/ranking resulting from the application of the *Best of the Best* (BoB)

Experts	(a) Rankings	(b) BC scores					(c) BoB scores				
		<i>o</i> <sub>1</sub>	<i>o</i> <sub>2</sub>	<i>o</i> <sub>3</sub>	<i>o</i> <sub>4</sub>	<i>o</i> <sub>5</sub>	<i>o</i> <sub>1</sub>	<i>o</i> <sub>2</sub>	<i>o</i> <sub>3</sub>	<i>o</i> <sub>4</sub>	<i>o</i> <sub>5</sub>
<i>e</i> <sub>1</sub>	<i>o</i> <sub>3</sub> > ( <i>o</i> <sub>1</sub> ~ <i>o</i> <sub>5</sub> ) > <i>o</i> <sub>2</sub> > <i>o</i> <sub>4</sub>	2.5	4.0	1.0	5.0	2.5	-	-	1	-	-
<i>e</i> <sub>2</sub>	<i>o</i> <sub>5</sub> > <i>o</i> <sub>1</sub> > ( <i>o</i> <sub>2</sub> ~ <i>o</i> <sub>3</sub> ~ <i>o</i> <sub>4</sub> )	2.0	4.0	4.0	4.0	1.0	-	-	-	-	1
<i>e</i> <sub>3</sub>	( <i>o</i> <sub>1</sub> ~ <i>o</i> <sub>3</sub> ) > <i>o</i> <sub>2</sub> > ( <i>o</i> <sub>4</sub> ~ <i>o</i> <sub>5</sub> )	1.5	3.0	1.5	4.5	4.5	0.5	-	0.5	-	-
<i>e</i> <sub>4</sub>	<i>o</i> <sub>1</sub> > <i>o</i> <sub>5</sub> > <i>o</i> <sub>3</sub> > <i>o</i> <sub>2</sub> > <i>o</i> <sub>4</sub>	1.0	4.0	3.0	5.0	2.0	1	-	-	-	-
<i>e</i> <sub>5</sub>	<i>o</i> <sub>3</sub> > ( <i>o</i> <sub>1</sub> ~ <i>o</i> <sub>2</sub> ~ <i>o</i> <sub>5</sub> ) > <i>o</i> <sub>4</sub>	3.0	3.0	1.0	5.0	3.0	-	-	1	-	-
<i>e</i> <sub>6</sub>	<i>o</i> <sub>5</sub> > <i>o</i> <sub>4</sub> > ( <i>o</i> <sub>1</sub> ~ <i>o</i> <sub>3</sub> ) > <i>o</i> <sub>2</sub>	3.5	5.0	3.5	2.0	1.0	-	-	-	-	1
<i>e</i> <sub>7</sub>	<i>o</i> <sub>3</sub> > <i>o</i> <sub>5</sub> > <i>o</i> <sub>4</sub> > <i>o</i> <sub>2</sub> > <i>o</i> <sub>1</sub>	5.0	4.0	1.0	3.0	2.0	-	-	1	-	-
<i>e</i> <sub>8</sub>	( <i>o</i> <sub>3</sub> ~ <i>o</i> <sub>5</sub> ) > ( <i>o</i> <sub>1</sub> ~ <i>o</i> <sub>2</sub> ) > <i>o</i> <sub>4</sub>	3.5	3.5	1.5	5.0	1.5	-	-	0.5	-	0.5
Collective scoring:		22.0	30.5	16.5	33.5	17.5	1.5	0	4.0	0	2.0
Collective ranking:		<i>o</i> <sub>3</sub> > <i>o</i> <sub>5</sub> > <i>o</i> <sub>1</sub> > <i>o</i> <sub>2</sub> > <i>o</i> <sub>4</sub>					<i>o</i> <sub>3</sub> > <i>o</i> <sub>5</sub> > <i>o</i> <sub>1</sub> > ( <i>o</i> <sub>2</sub> ~ <i>o</i> <sub>4</sub> )				

**Table 5** Scoring system used in the construction of the “consistency table”

Case	Conv. score
1. <i>Full consistency</i> , i.e., identical relationship of strict preference (“>”) or indifference (“~”)	1
2. <i>Weak consistency</i> , i.e., consistency with respect to a weak preference relationship only (“> or ~” and “< or ~”, i.e., strict preference or indifference); e.g., when comparing the relationship <i>o</i> <sub>1</sub> > <i>o</i> <sub>2</sub> with <i>o</i> <sub>1</sub> ~ <i>o</i> <sub>2</sub>	0.5
3. <i>Inconsistency</i> (with respect to both strict and weak preference relationships); e.g., when comparing the relationship <i>o</i> <sub>1</sub> > <i>o</i> <sub>2</sub> with <i>o</i> <sub>2</sub> > <i>o</i> <sub>1</sub>	0

relationships (see symbols “>” and “~” in Tables 6a, 7a. Next, a “consistency table”—which turns the paired-comparison relationships of each expert into scores, according to the scoring system in Table 5—is constructed; the conventional assignment of 0.5 points in the case of *weak consistency* is justified by the fact that this is the intermediate case between that of *full consistency* (with score 1) and that of *inconsistency* (with score 0) (Franceschini et al. 2022). The consistency table also reports the sum of the scores (*x<sub>j</sub>*) obtained by each *j*-th expert ranking. Tables 6b and 7b exemplify two consistency tables related to the case study of interest, for both the aggregation techniques (BC and BoB respectively). Tables 6c and 7c show that both techniques result in collective rankings that are generally consistent with the single expert rankings. The least consistent expert rankings (i.e., those with lower *p<sub>j</sub>* values) appear to be those formulated by *e<sub>4</sub>* and *e<sub>6</sub>*, although the distinction is small.

Next, for each *j*-th expert, the portion of “consistent” paired-comparisons can be calculated as:

$$p_j = \frac{x_j}{\binom{n}{2}} = \frac{x_j}{10}, \tag{3}$$

being

*x<sub>j</sub>* the total score related to the *j*-th expert;  
 $\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}$  the overall number of paired comparisons (i.e., 10 here).

- *p*, i.e., indicator of *global consistency*. In the case of equally-important experts, the *p<sub>j</sub>* values are aggregated through the arithmetic average (Franceschini et al. 2022):

$$p = \frac{1}{m} \cdot \sum_{j=1}^m p_j, \quad p \in [0, 1]. \tag{4}$$

In this particular case, the two aggregation techniques result in two relatively close *p*-values: i.e., 75.0% for BC and 73.8% for BoB (see Tables 6c and 7c). This confirms that both techniques yield collective rankings that are relatively consistent with the input data (and *vice versa*), with a slight predominance of BC over BoB. In the case of non-equally-important experts and/or incomplete expert rankings, the formulation of *p*-indicators is more complex (Franceschini et al. 2022).

Besides the *p*-indicators, another tool for assessing consistency is  $W_k^{(m+1)}$ , i.e., an indicator inspired by Kendall’s *W* (cf. Equation 1), which is nothing more than *W* itself applied to (*m* + 1) rankings consisting of: (i) the *m*-expert rankings, and (ii) the collective ranking obtained after the application of a given aggregation model (*k*) to the previous expert rankings. Consistency between collective ranking and expert rankings is assessed in *relative* terms, by comparing  $W_k^{(m+1)}$  with the traditional *W*.  $W_k^{(m+1)} \geq W$  denotes consistency (or *positive* consistency) between the collective ranking and the *m*-rankings, while  $W_k^{(m+1)} < W$  denotes inconsistency (or *negative* consistency) (Franceschini and Maisano 2021). The latter situation can occur when a collective ranking is

**Table 6** **a** Paired-comparison table, **b** consistency table, and **c** *p*-indicators related to the BC technique (which resulted into the collective ranking:  $o_3 > o_5 > o_1 > o_2 > o_4$ , cf. Fig. 5b)

(a) Paired-comparison table

Paired comparison		Relationship								Collective judgment
		$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$	
1	$o_1, o_2$	>	>	>	>	~	>	<	~	>
2	$o_1, o_3$	<	>	~	>	<	~	<	<	<
3	$o_1, o_4$	>	>	>	>	>	<	<	>	>
4	$o_1, o_5$	~	<	>	>	~	<	<	<	<
5	$o_2, o_3$	<	~	<	<	<	<	<	<	<
6	$o_2, o_4$	>	~	>	>	>	<	<	>	>
7	$o_2, o_5$	<	<	>	<	~	<	<	<	<
8	$o_3, o_4$	>	~	>	>	>	<	>	>	>
9	$o_3, o_5$	>	<	>	<	>	<	>	~	>
10	$o_4, o_5$	<	<	~	<	<	<	<	<	<

(b) Consistency table

Paired comparison		Scores							
		$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$
1	$o_1, o_2$	1	1	1	1	0.5	1	0	0.5
2	$o_1, o_3$	1	0	0.5	0	1	0.5	1	1
3	$o_1, o_4$	1	1	1	1	1	0	0	1
4	$o_1, o_5$	0.5	1	0	0	0.5	1	1	1
5	$o_2, o_3$	1	0.5	1	1	1	1	1	1
6	$o_2, o_4$	1	0.5	1	1	1	0	0	1
7	$o_2, o_5$	1	1	0	1	0.5	1	1	1
8	$o_3, o_4$	1	0.5	1	1	1	0	1	1
9	$o_3, o_5$	1	0	1	0	1	0	1	0.5
10	$o_4, o_5$	1	1	0.5	1	1	1	1	1
	$x_j$	9.5	6.5	7	7	8.5	5.5	7	9

(c) *p*-indicators

$p_{>j}$	95%	65%	70%	70%	85%	55%	70%	90%	$p =$ 75.0%
----------	-----	-----	-----	-----	-----	-----	-----	-----	----------------

somehow conflicting with the *m*-rankings. To make the consistency assessment easier, another synthetic indicator can be used:

$$b_k^{(m)} = \frac{W_k^{(m+1)}}{W^{(m)}}, \quad b_k^{(m)} \in ]0, +\infty]. \tag{5}$$

For a specific set of *m* rankings,  $b_k^{(m)} \geq 1$  indicates that the aggregation model (*k*) provides a somehow consistent collective ranking (positive consistency), while  $b_k^{(m)} < 1$  indicates that it provides a somehow inconsistent collective ranking (negative consistency). Table 8 exemplifies the calculation of indicators  $W_k^{(m+1)}$  and  $b_k^{(m)}$  for the case study, considering the BC and BoB aggregation techniques respectively.

Positive consistency is observed for both techniques, with a slight predominance of BC over BoB (e.g., consider the  $b_k^{(m)}$  value of 1.13 for BC versus 1.12 for BoB), confirming the result obtained through *p*-indicators.

### 3.3.3 Robustness of the solution

The formulation of rankings is often affected by inherent variability, which can "propagate" onto the variability of the output (Saltelli et al. 2006). Only a very few aggregation techniques associate the resulting collective judgment with a corresponding estimate of variability (Franceschini and Maisano 2020). In general, it may be useful to perform a *sensitivity analysis* to assess the robustness of the

**Table 7** **a** Paired-comparison table, **b** consistency table, and **c** *p*-indicators related to the BoB technique (which resulted into the collective ranking:  $o_3 > o_5 > o_1 > (o_2 \sim o_4)$ , cf. Fig. 5c)

(a) Paired-comparison table

Paired comparison		Relationship								Collective judgment
		$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$	
1	$o_1, o_2$	>	>	>	>	~	>	<	~	>
2	$o_1, o_3$	<	>	~	>	<	~	<	<	<
3	$o_1, o_4$	>	>	>	>	>	<	<	>	>
4	$o_1, o_5$	~	<	>	>	~	<	<	<	<
5	$o_2, o_3$	<	~	<	<	<	<	<	<	<
6	$o_2, o_4$	>	~	>	>	>	<	<	>	~
7	$o_2, o_5$	<	<	>	<	~	<	<	<	<
8	$o_3, o_4$	>	~	>	>	>	<	>	>	>
9	$o_3, o_5$	>	<	>	<	>	<	>	~	>
10	$o_4, o_5$	<	<	~	<	<	<	<	<	<

(b) Consistency table

Paired comparison		Scores							
		$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$
1	$o_1, o_2$	1	1	1	1	0.5	1	0	0.5
2	$o_1, o_3$	1	0	0.5	0	1	0.5	1	1
3	$o_1, o_4$	1	1	1	1	1	0	0	1
4	$o_1, o_5$	0.5	1	0	0	0.5	1	1	1
5	$o_2, o_3$	1	0.5	1	1	1	1	1	1
6	$o_2, o_4$	0.5	1	0.5	0.5	0.5	0.5	0.5	0.5
7	$o_2, o_5$	1	1	0	1	0.5	1	1	1
8	$o_3, o_4$	1	0.5	1	1	1	0	1	1
9	$o_3, o_5$	1	0	1	0	1	0	1	0.5
10	$o_4, o_5$	1	1	0.5	1	1	1	1	1
	$x_j$	9	7	6.5	6.5	8	6	7.5	8.5

(c) *p*-indicators

$p_j$	90%	70%	65%	65%	80%	60%	75%	85%	$p = 73.8\%$
-------	-----	-----	-----	-----	-----	-----	-----	-----	--------------

solution against small variations in the input data (Saltelli et al. 2006). An example of sensitivity analysis follows.

Table 9 contains three sets of expert rankings: (i) the initial one (cf. Fig. 5a) and (ii, iii) two additional ones, obtained by applying small distortions to the initial one. These distortions can be achieved automatically in multiple ways. In the present case, a procedure described in the following four steps was adopted.

1. Each expert ranking is translated into a scoring corresponding to the *average ranks* of individual objects. For example, the ranking by  $e_5$ , i.e.,  $o_3 > (o_1 \sim o_5) > o_2 > o_4$ , is translated into the scores ( $s$ )  $o_1 = 2.5, o_2 = 4, o_3 = 1, o_4 = 5, o_5 = 2.5$  (cf. Fig. 5b).

2. Next, the score ( $s$ ) of each object is distorted by adding to it an error ( $\epsilon$ ) given by a zero-mean random variable, *uniformly* distributed within the interval  $[-1, +1]$ , i.e.,  $\epsilon \sim U(-1, +1)$ . Translating this into a formula:

$$s' = s + \epsilon, \tag{6}$$

being  $s'$  the resulting distorted score. The above interval of variability seems in line with the idea of small (positive or negative) variations of input rankings.

3. Next, the score ( $s'$ ) of each object is rounded to the nearest integer, resulting in the new score:

$$s'' = \text{round}(s'), \tag{7}$$

**Table 8**  $W$ ,  $W_k^{(m+1)}$ , and  $b_k^m$  indicators for the collective rankings resulting from the application of BC and BoB aggregation techniques to the problem of interest

Aggr. technique	Coll. judgment (scoring)	Collective ranking	$W$	$W_k^{(m+1)}$	$b_k^m$
BC	$o_1 = 22.0, o_2 = 30.5, o_3 = 16.5, o_4 = 33.5, o_5 = 17.5$	$o_3 > o_5 > o_1 > o_2 > o_4$	40.2%	45.4%	<b>1.13</b>
BoB	$o_1 = 1.5, o_2 = 0, o_3 = 4, o_4 = 0, o_5 = 2$	$o_3 > o_5 > o_1 > (o_2 \sim o_4)$	40.2%	45.2%	<b>1.12</b>

**Table 9** Set of rankings used for sensitivity analysis

Experts	(i) Initial set of rankings	(ii) 1st additional set	(iii) 2nd additional set
$e_1$	$o_3 > (o_1 \sim o_5) > o_2 > o_4$	$(o_3 \sim o_1) > o_5 > (o_2 \sim o_4)$	$o_5 > o_3 > o_1 > o_4 > o_2$
$e_2$	$o_5 > o_1 > (o_2 \sim o_3 \sim o_4)$	$o_1 > o_5 > (o_2 \sim o_3) > o_4$	$(o_5 \sim o_1 \sim o_2) > o_4 > o_3$
$e_3$	$(o_1 \sim o_3) > o_2 > (o_4 \sim o_5)$	$(o_1 \sim o_3) > o_5 > (o_2 \sim o_4)$	$(o_1 \sim o_3 \sim o_2) > o_4 > o_5$
$e_4$	$o_1 > o_5 > o_3 > o_2 > o_4$	$o_1 > (o_5 \sim o_3) > o_4 > o_2$	$o_5 > o_1 > (o_3 \sim o_4 \sim o_2)$
$e_5$	$o_3 > (o_1 \sim o_2 \sim o_5) > o_4$	$o_3 > (o_1 \sim o_2) > o_5 > o_4$	$o_3 > o_2 > o_1 > (o_5 \sim o_4)$
$e_6$	$o_5 > o_4 > (o_1 \sim o_3) > o_2$	$o_5 > o_4 > (o_1 \sim o_2) > o_3$	$o_4 > o_5 > (o_1 \sim o_3 \sim o_2)$
$e_7$	$o_3 > o_5 > o_4 > o_2 > o_1$	$o_5 > o_3 > o_2 > o_4 > o_1$	$o_3 > o_4 > o_5 > (o_2 \sim o_1)$
$e_8$	$(o_3 \sim o_5) > (o_1 \sim o_2) > o_4$	$(o_3 \sim o_5) > o_1 > o_2 > o_4$	$o_1 > (o_3 \sim o_5) > (o_2 \sim o_4)$

where  $round(\cdot)$  is an operator that rounds a certain score to the nearest integer.

For example, applying the distortion in Eq. 6 to the scores ( $s$ ) at step one, we get the scoring ( $s'$ ):  $o_1 = 2.1, o_2 = 4.6, o_3 = 1.6, o_4 = 4.8, o_5 = 2.6$ ; then, applying the rounding in Eq. 7, we get the new scoring ( $s''$ ):  $o_1 = 2, o_2 = 5, o_3 = 2, o_4 = 5, o_5 = 3$ .

4. Subsequently, the set of  $s''$  scores are translated into an “additional” ranking, with relationships of strict preference (“>”) and indifference (“~”), similarly to the transformation from rating to ranking in Fig. 3. Returning to the above example, the  $s''$  scores at the following step are transformed into the (additional) ranking:  $(o_3 \sim o_1) > o_5 > (o_2 \sim o_4)$ . The procedure was extended to all initial rankings and repeated twice, resulting in the two additional sets of rankings in Table 9(ii), (iii).

For each set (initial and additional), the collective scoring/ranking was determined by applying the BC and BoB aggregation techniques (see results in Table 10). Next, the average dispersion in the rank position of individual objects can be used as a proxy for the robustness of the resulting collective rankings (see Table 11). In this specific case—BC provides somewhat more robust results than BoB (i.e., lower mean standard deviation of 0.44 against 0.60). However, both solutions appear relatively robust (i.e., mean standard deviation lower than 1), therefore no revision of the aggregation techniques seems necessary (cf., feedback loop from block 3.6 of Fig. 2).

### 4 Conclusion

This paper focused on the ranking-aggregation problem, highlighting its significance due to the variety of potential applications in the field of manufacturing. By adopting a pragmatic approach based on a case study, the paper has elucidated a sequential and iterative operational methodology to address the problem at various levels:

- Checking the plausibility of expert rankings in terms of *concordance*, through *multivariate* and *bivariate* statistical measures;
- Guiding the aggregation-technique selection, depending on the desired types of input and output data;
- Evaluating the *consistency* and *robustness* of the resulting collective judgment.

The case study has demonstrated that approaching the problem systematically necessitates multiple iterations and corrections at the aforementioned levels. Notably, the application of the aggregation technique is just one component of the proposed methodology, with various verifications and corrections required prior to the aggregation phase.

This study not only enhances the understanding of the complexity of the ranking-aggregation problem but also provides practical tools to tackle it in a structured and efficient manner. The outcomes hold value for both scientists and practitioners in the manufacturing domain who encounter decision-making challenges related to ranking aggregation. It is worth mentioning these actors may not have extensive expertise to deal with the

**Table 10** Rank tables and collective scorings/rankings resulting from sensitivity analysis

		(a) Rank table								(b) Collective scoring (and rank)			
(i) Initial set of rankings		$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$	BC		BoB	
	$o_1$	2.5	2	1.5	1	3	3.5	5	3.5	22.0	(3.0)	1.5	(3)
	$o_2$	4	4	3	4	3	5	4	3.5	30.5	(4.0)	0.0	(4.5)
	$o_3$	1	4	1.5	3	1	3.5	1	1.5	16.5	(1.0)	4.0	(1)
	$o_4$	5	4	4.5	5	5	2	3	5	33.5	(5.0)	0.0	(4.5)
	$o_5$	2.5	1	4.5	2	3	1	2	1.5	17.5	(2.0)	2.5	(2)
Collective ranking:										$o_3 > o_5 > o_1 > o_2 > o_4$		$o_3 > o_5 > o_1 > (o_2 \sim o_4)$	
		(a) Rank table								(b) Collective scoring (and rank)			
(ii) 1st additional set		$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$	BC		BoB	
	$o_1$	1.5	1	1.5	1	2.5	3.5	5	3	19.0	(3.0)	4.0	(1)
	$o_2$	4.5	3.5	4.5	5	2.5	3.5	3	4	30.5	(4.0)	0.0	(4.5)
	$o_3$	1.5	3.5	1.5	2.5	1	5	2	1.5	18.5	(2.0)	2.5	(2.5)
	$o_4$	4.5	5	4.5	4	5	2	4	5	34.0	(5.0)	0.0	(4.5)
	$o_5$	3	2	3	2.5	4	1	1	1.5	18.0	(1.0)	2.5	(2.5)
Collective ranking:										$o_5 > o_3 > o_1 > o_2 > o_4$		$o_1 > (o_3 \sim o_5) > (o_2 \sim o_4)$	
		(a) Rank table								(b) Collective scoring (and rank)			
(iii) 2nd additional set		$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$	BC		BoB	
	$o_1$	3	2	2	2	3	4	4.5	1	21.5	(2.5)	1.67	(3)
	$o_2$	5	2	2	4	2	4	4.5	4.5	28.0	(4.5)	0.67	(5)
	$o_3$	2	5	2	4	1	4	1	2.5	21.5	(2.5)	2.33	(2)
	$o_4$	4	4	4	4	4.5	1	2	4.5	28.0	(4.5)	1.00	(4)
	$o_5$	1	2	5	1	4.5	2	3	2.5	21.0	(1.0)	2.33	(1)
Collective ranking:										$o_5 > (o_1 \sim o_3) > (o_2 \sim o_4)$		$(o_3 \sim o_5) > o_1 > o_4 > o_2$	

**Table 11** Results of sensitivity analysis, in terms of mean standard deviation of the objects' rank positions

		(i) Initial set of rankings	(ii) 1st additional set	(iii) 2nd additional set	St. dev
BC	$o_1$	3.0	3.0	2.5	0.29
	$o_2$	4.0	4.0	4.5	0.29
	$o_3$	1.0	2.0	2.5	0.76
	$o_4$	5.0	5.0	4.5	0.29
	$o_5$	2.0	1.0	1.0	0.58
<b>Mean st. dev</b>					<b>0.44</b>
BoB	$o_1$	3.0	1.0	3.0	1.16
	$o_2$	4.5	4.5	5.0	0.29
	$o_3$	1.0	2.5	1.5	0.76
	$o_4$	4.5	4.5	4.0	0.29
	$o_5$	2.0	2.5	1.5	0.50
<b>Mean st. dev</b>					<b>0.60</b>

Bold indicates objects rank positions

ranking-aggregating problem comprehensively; thus, the proposed procedure helps to fill this gap.

The proposed methodology can be considered *modular* in that it is able to combine several practical tools

interchangeably; however, the discussion provided in this paper, to avoid excessive length, was limited to exemplifying a few specific tool (e.g.,  $\rho$  and  $W$  as indicators of concordance of experts' rankings, and  $p$ -indicators as measures of consistency between input and output data). Furthermore, the authors acknowledge that the choice of the aggregation technique remains perhaps the most delicate aspect, which was only marginally addressed in this paper. Future research plans include establishing an extensive taxonomy of aggregation techniques and analytical tools to facilitate their selection for specific problems. It is envisaged to create a step-by-step procedure that, based on the problem's characteristics specified by the user, will guide the selection of appropriate aggregation techniques tailored to the specific case.

**Author contributions** The authors have provided an equal contribution to the drafting of the paper.

**Funding** Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.

**Availability of data and materials** Not applicable.

## Declarations

**Conflict of interest** The authors do not have conflict of interest.

**Ethical approval** The authors respect the Ethical Guidelines of the Journal.

**Consent to participate** Not applicable.

**Consent to publish** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agresti A (2010) Analysis of ordinal categorical data, 2nd edn. John Wiley & Sons, New York
- Arrow KJ (2012) Social choice and individual values, 3rd edn. Yale University Press, New Haven
- Bana e Costa CA (ed) (2012) Readings in multiple criteria decision aid. Springer Science & Business Media, Berlin Heidelberg
- Belton V, Stewart T (2002) Multiple criteria decision analysis: an integrated approach. Springer Science & Business Media, New York
- Blais A (ed) (2008) To keep or to change first past the post? the politics of electoral reform. Oxford University Press, Oxford
- Borda JC (1781) Mémoire sur les élections au scrutin comptes rendus de l'académie des sciences. Translated by Alfred de Grazia as mathematical derivation of an election system. *Isis* 44:42–51. <https://doi.org/10.1086/348187>
- Chatterjee P, Chakraborty S (2014) Flexible manufacturing system selection using preference ranking methods: a comparative study. *Int J Ind Eng Comput* 5(2):315–338. <https://doi.org/10.5267/IJIEC.2013.10.002>
- Chen S, Liu J, Wang H, Augusto JC (2012) Ordering based decision making—a survey. *Inf Fusion* 14(4):521–531. <https://doi.org/10.1016/j.inffus.2012.10.005>
- Chiclana F, Herrera F, Herrera-Viedma E (2002) A note on the internal consistency of various preference representations. *Fuzzy Sets Syst* 131(1):75–78. [https://doi.org/10.1016/S0165-0114\(01\)00256-1](https://doi.org/10.1016/S0165-0114(01)00256-1)
- Chuu SJ (2009) Group decision-making model using fuzzy multiple attributes analysis for the evaluation of advanced manufacturing technology. *Fuzzy Sets Syst* 160(5):586–602. <https://doi.org/10.1016/j.fss.2008.07.015>
- Corrente S, Greco S, Rezaei J (2024) Better decisions with less cognitive load: the parsimonious BWM. *Omega* 126:103075
- Dym CL, Wood WH, Scott MJ (2002) Rank ordering engineering designs: pairwise comparison charts and Borda counts. *Res Eng Design* 13:236–242. <https://doi.org/10.1007/s00163-002-0019-8>
- El Zaatari S, Marei M, Li W, Usman Z (2019) Cobot programming for collaborative industrial tasks: an overview. *Robot Auton Syst* 116:162–180. <https://doi.org/10.1016/j.robot.2019.03.003>
- Figueira J, Greco S, Ehrgott M (2005) Multiple criteria decision analysis: state of the art surveys. Springer, New York
- Franceschini F, Maisano D (2018a) A new proposal to improve the customer competitive benchmarking in QFD. *Qual Eng* 30(4):730–761. <https://doi.org/10.1080/08982112.2018.1437178>
- Franceschini F, Maisano D (2018b) Classification of objects into quality categories in the presence of hierarchical decision-making agents. *Accred Qual Assur* 23(1):5–17. <https://doi.org/10.1007/s00769-017-1291-7>
- Franceschini F, Maisano D (2019) Design decisions: concordance of designers and effects of the Arrow's theorem on the collective preference ranking. *Res Eng Design* 30(3):425–434. <https://doi.org/10.1007/S00163-019-00313-9>
- Franceschini F, Maisano D (2020) Aggregation of incomplete preference rankings: robustness analysis of the ZM II-technique. *J Multi-Criteria Decis Anal* 27(5–6):337–356. <https://doi.org/10.1002/mcda.1721>
- Franceschini F, Maisano D (2021) Aggregating multiple ordinal rankings in engineering design: the best model according to the Kendall's coefficient of concordance. *Res Eng Design* 32(1):91–103. <https://doi.org/10.1007/s00163-020-00348-3>
- Franceschini F, Maisano D, Mastrogiacomo L (2022) Rankings and decisions in engineering: conceptual and practical insights. International series in operations research & management science series, vol 319. Springer International Publishing, Cham
- Friedman M (1940) A comparison of alternative tests of significance for the problem of m rankings. *Ann Math Stat* 11(1):86–92. <https://doi.org/10.1214/AOMS/1177731944>
- Geramian A, Abraham A, Ahmadi NM (2019) Fuzzy logic-based FMEA robust design: a quantitative approach for robustness against groupthink in group/team decision-making. *Int J Prod Res* 57(5):1331–1344. <https://doi.org/10.1080/00207543.2018.1471236>
- Gervasi R, Mastrogiacomo L, Maisano DA, Antonelli D, Franceschini F (2022) A structured methodology to support human–robot collaboration configuration choice. *Prod Eng Res Devel* 16:435–451. <https://doi.org/10.1007/s11740-021-01088-6>
- Giachetti RE (1998) A decision support system for material and manufacturing process selection. *J Intell Manuf* 9(3):265–276. <https://doi.org/10.1023/A:1008866732609>
- Gibbons JD, Chakraborti S (2010) Nonparametric statistical inference, 5th edn. CRC Press, Boca Raton. <https://doi.org/10.1201/9781439896129>
- Hakimi-Asl A, Amalnick MS, Hakimi-Asl M (2018) Proposing a graph ranking method for manufacturing system selection in high-tech industries. *Neural Comput Appl* 29:133–142. <https://doi.org/10.1007/s00521-016-2420-7>
- Herrera-Viedma E, Cabrerizo FJ, Kacprzyk J, Pedrycz W (2014) A review of soft consensus models in a fuzzy environment. *Inf Fusion* 17:4–13
- Kabirifar K, Ashour M, Yazdani M, Mahdiyar A, Malekjafarian M (2023) Cybernetic-parsimonious MCDM modeling with application to the adoption of circular economy in waste management. *Appl Soft Comput* 139:110186
- Kendall MG (1962) Ranks and measures. *Biometrika* 49(1/2):133–137
- Köksalan M, Wallenius J, Zionts S (2013) An early history of multiple criteria decision making. *J Multi-Criteria Decis Anal* 20(1–2):87–94
- Lagerspetz E (2016) Social choice and democratic values. Springer, Heidelberg
- Legendre P (2010) Coefficient of concordance. *Encycl Res Design* 1:164–169
- Leo Kumar SP (2019) Knowledge-based expert system in manufacturing planning: state-of-the-art review. *Int J Prod Res* 57(15–16):4766–4790. <https://doi.org/10.1080/00207543.2018.1424372>

- McComb C, Goucher-Lambert K, Cagan J (2017) Impossible by design? Fairness, strategy and Arrow's impossibility theorem. *Des Sci* 3:1–26
- Nahm YE, Ishikawa H, Inoue M (2013) New rating methods to prioritize customer requirements in QFD with incomplete customer preferences. *Int J Adv Manuf Technol* 65(9–12):1587–1604
- Nestic S, Lampón JF, Aleksic A, Cabanelas P, Tadic D (2019) Ranking manufacturing processes from the quality management perspective in the automotive industry. *Expert Syst* 36(6):e12451
- Oukil A (2019) Embedding OWA under preference ranking for DEA cross-efficiency aggregation: issues and procedures. *Int J Intell Syst* 34(5):947–965. <https://doi.org/10.1002/int.22082>
- Qin Y, Qi Q, Scott PJ, Jiang X (2020) An additive manufacturing process selection approach based on fuzzy Archimedean weighted power Bonferroni aggregation operators. *Robot Comput-Integr Manuf* 64:101926
- Reich Y (2010) My method is better! *Res Eng Design* 21(3):137–142
- Ross SM (2009) Introduction to probability and statistics for engineers and scientists. Academic Press, New York
- Saari DG (2011) Decision and elections. Cambridge University Press, Cambridge
- Saltelli A, Ratto M, Tarantola S, Campolongo F (2006) Sensitivity analysis practices: strategies for model-based inference. *Reliab Eng Syst Saf* 91(10–11):1109–1125
- Sarwar M, Akram M, Liu P (2021) An integrated rough ELECTRE II approach for risk evaluation and effects analysis in automatic manufacturing process. *Artif Intell Rev* 54(6):4449–4481. <https://doi.org/10.1007/s10462-021-10003-5>
- Sato Y, Tan KH (2023) Inconsistency indices in pairwise comparisons: an improvement of the consistency index. *Ann Oper Res* 326(2):809–830
- Spohn W (2009) A survey of ranking theory. In: Huber F, Schmidt-Petri C (eds) Degrees of belief. Springer, Dordrecht, pp 185–228
- Yu Q, Hou F (2016) An approach for green supplier selection in the automobile manufacturing industry. *Kybernetes* 45(4):571–588
- Zeleny M (ed) (1976) Multiple criteria decision making Kyoto 1975 (Vo. 123). Lecture notes in economics and mathematical systems. Springer Science & Business Media, Heidelberg