# Usability of Concordance Indices in *FAST-GDM* Problems

Marcelo Loor[1,2], Ana Tapia-Rosero[2] and Guy De Tré[1]

[1]*Dept. of Telecommunications and Information Processing, Ghent University, Sint-Pietersnieuwstraat 41, B-9000, Ghent, Belgium*

[2]*Dept. of Electrical and Computer Engineering, ESPOL Polytechnic University, Campus Gustavo Galindo V., Km. 30.5 Via Perimetral, Guayaquil, Ecuador*

Keywords: Flexible Consensus Reaching, Group Decision-Making, Intuitionistic Fuzzy Sets, IFS Contrasting Charts.

Abstract: A *flexible attribute-set group decision-making* (FAST-GDM) problem boils down to finding the most suitable option(s) with a general agreement among the participants in a decision-making process in which each option can be described by a flexible collection of attributes. The solution to such a problem can involve a consensus reaching process (CRP) in which the participants iteratively try to reach a general agreement on the best option(s) based on the attributes that are relevant for each participant. A challenging task in a CRP is the selection of an adequate method to determine the level of concordance between the evaluations given by each participant and the collective evaluations computed for the group. To gain insights in this regard, we performed a pilot test in which a group of persons were asked to estimate the level of concordance between individual and collective evaluations obtained while other participants tried to solve a FAST-GDM problem. The perceived concordance levels were compared with several theoretical concordance indices based on similarity measures designed to compare intuitionistic fuzzy sets. This paper presents our findings on how each of the chosen theoretical concordance indices reflected the perceived concordance levels.

## 1 INTRODUCTION

Consider a situation in which a group of viniculturists are trying to reach a consensus about the best grapevine(s) for winemaking from a collection of grapevines that have been developed for the industry. In this situation, a consensus can be reached using a process where the viniculturists, under the supervision of a moderator, can iteratively reconsider their evaluations to be in agreement with the group and, thus, to decide on the best grapevine(s) for winemaking (Bouyssou et al., 2013). Assuming that the viniculturists have a similar expertise, the moderator can ask them to evaluate each grapevine (i.e., each option) using a predefined collection of attributes, which denote the features or characteristics inherent to any of the grapevines under evaluation. As such, this situation can be deemed to be an example of a *multi-attribute group decision-making* (MA-GDM) problem (Dong et al., 2016; Liu et al., 2016).

A different situation is one in which a heterogeneous group of participants have different opinions on how the attributes (or features) of the given options should be evaluated. In this case, the problem of finding a consensus with a flexible collection of attributes comes to light. As an example, one can consider another situation where three untrained viniculturists, say Alice, Bob and Chloe, are trying to reach a consensus on the best grapevine(s) for winemaking from the aforementioned collection of grapevines: while Alice considers that one of the grapevines, say '*GV1*', is almost the best for winemaking due to its reddish color, Bob considers this grapevine to be unacceptable for winemaking because it is expensive; meanwhile, Chloe considers that '*GV1*' is a good grapevine for winemaking because of its appetizing aroma, but it is not the best due to its strong flavor. Notice in this example that, since they are untrained viniculturists, Alice, Bob and Chloe evaluate '*GV1*' according to the attributes that each of them might consider to be relevant on this grapevine for deciding whether or not it is the best for winemaking. In this case, a consensus can be reached using a process where these persons iteratively reconsider their evaluations based on the attributes of the grapevines that were initially unobserved by some of them but observed by others. This last situation can be seen as an example of a *flexible attribute-set group decision-making* (FAST-GDM) problem (Loor et al., 2018).

The solution to such group decision-making pro-



Loor, M., Tapia-Rosero, A. and De Tré, G.
Usability of Concordance Indices in *FAST-GDM* Problems.
DOI: 10.5220/0006956500670078
In *Proceedings of the 10th International Joint Conference on Computational Intelligence (IJCCI 2018)*, pages 67-78
ISBN: 978-989-758-327-8

blems can involve a *consensus reaching process* (CRP) in which the participants (experts or non-experts) iteratively try to reach a collective agreement on the best option(s) (Kacprzyk and Fedrizzi, 1988; Herrera et al., 1996; Kacprzyk and Zadrożny, 2010). A challenge in this process is to find a consistently good method by which one can determine the level of concordance between the individual evaluations given by a participant and the collective evaluations computed for the group.

To gain insights about what can be accepted as a good indicator of the level of concordance in FAST-GDM problems, we conducted a pilot test in which an easy-to-reach group of persons were asked to make an estimation of the level of concordance that they perceive between the individual and the collective evaluations obtained during the iterations of a CRP in a FAST-GDM problem. To make such estimations, those persons were provided with *IFS contrasting charts*, a novelty of this work, which depict the evaluations characterized as *intuitionistic fuzzy sets* (IFSs) (Atanassov, 1986; Atanassov, 2012). Then, those estimations were compared with several theoretical concordance indices to determine the usability of each index. By means of such comparisons, we aim to determine how well those theoretical concordance indices could reflect a perceived level of concordance in a FAST-GDM problem.

It is worth mentioning that, although the CRP proposed for FAST-GDM problems is based on an augmented variant of IFSs called *augmented Atanassov intuitionistic fuzzy sets* (Loor and De Tré, 2017a), in this test we use similarity measures designed to compare traditional IFSs because the computation of concordance indices can be done by using only the membership and nonmembership levels (Loor et al., 2018). A practical motivation in this regard is to study the applicability of the tools included in the IFS framework to the solution of decision-making problems where participants having different expertise are given the freedom to perform positive or negative evaluations according to what they consider to be relevant.

To present the results of the pilot test, this paper has been structured as follows. In Section 2, we present some preliminary concepts, as well as the formulation of a FAST-GDM problem. Then, in Section 3, we describe the pilot test and introduce the novel *IFS contrasting charts*. Before concluding this paper, we present the results and our findings in Section 4 and some related work in Section 5.

# 2 PRELIMINARIES

When there are different opinions on how the attributes (or features) of a predefined collection of potential options should be evaluated, an evaluation might be accompanied by some suggestions about what have been focused on during the evaluation process. For instance, in the second introductory example Alice has considered that *'GV1'* is almost the best grapevine for winemaking due to its reddish color. To characterize this kind of evaluations, the idea of an *augmented appraisal degree* (AAD) has been introduced in (Loor and De Tré, 2017a). In the context of decision-making, such an augmented appraisal degree idea can be described as follows:

Consider a discrete collection $X = \{x_1, \cdots, x_n\}$ of potential solutions, called *options*, for a particular problem, where each $x_i \in X$ has a collection of features $\mathcal{F}_i$. Consider also that a collection of *suitable options* for this problem in $X$ is denoted by $A$, i.e., $A \subseteq X$. Finally, consider a person $P$ who has been asked to evaluate the level to which an option $x_i \in X$ satisfies the proposition "$p : x_i$ is a member of $A$." With these considerations, an augmented appraisal degree of $x_i$, say $\hat{\mu}_{A@P}(x_i)$, is a pair $\langle \mu_{A@P}(x_i), F_{\mu_A@P}(x_i) \rangle$ that denotes the level $\mu_{A@P}(x_i)$ to which $x_i$ satisfies the proposition $p$, as well as the collection of features $F_{\mu_A@P}(x_i) \in \mathcal{F}_i$ considered by $P$ while appraising the proposition $p$.

For instance, consider the collection of grapevines $X = \{$*'GV1'*, *'GV2'*, *'GV3'*$\}$. Consider also a unit interval scale where 0 and 1 represent the lowest and the highest level of satisfiability respectively. In this context, after denoting (the collection of) *the best grapevine(s) for winemaking* by the letter $A$, one can characterize Alice's evaluation by $\hat{\mu}_{A@Alice}($*'GV1'*$) = \langle 0.9, \{$*'reddish color'*$\} \rangle$.

In the second introductory example, Chloe has considered that *'GV1'* is a good grapevine for winemaking because of its appetizing aroma, but it is not the best due to its strong flavor. Notice is this case that one can provide an evaluation denoting not only *how acceptable* but also *how unacceptable* an option could be. To characterize this kind of evaluations, the inclusion of AADs into the definition of an IFS has been proposed in (Loor and De Tré, 2017a). Such augmented version of an IFS, called *augmented (Atanassov) intuitionistic fuzzy set* (AAIFS), can be described as follows:

Keep $X$, $A$, $p$ and $P$ as given above. Recall that $x_i \in X$ has a collection of features $\mathcal{F}_i$ and assume that $\mathcal{F} = \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_n$. Assume also $\mathcal{I} = [0,1]$. Let $\hat{\mu}_{A@P}(x_i) = \langle \mu_{A@P}(x_i), F_{\mu_A@P}(x_i) \rangle$ and $\hat{\nu}_{A@P}(x_i) = \langle \nu_{A@P}(x_i), F_{\nu_A@P}(x_i) \rangle$ in $\langle \mathcal{I}, \mathcal{F} \rangle$ be two AADs denoting the evaluations given by $P$ on *how acceptable* and *how unacceptable* $x_i$ is for fulfilling the proposition $p$ respectively. In this context, an AAIFS is a collection $\hat{A}_{@P}$ that describes the correspondence between each $x_i \in X$ and both $\hat{\mu}_{A@P}(x_i)$ and $\hat{\nu}_{A@P}(x_i)$ through the expression

$$\hat{A}_{@P} = \{\langle x_i, \hat{\mu}_{A@P}(x_i), \hat{\nu}_{A@P}(x_i)\rangle \mid (x_i \in X) \wedge (0 \le \mu_{A@P}(x_i) + \nu_{A@P}(x_i) \le 1)\}. \quad (1)$$

Notice in this expression that the *consistency condition*, i.e., $0 \le \mu_{A@P}(x_i) + \nu_{A@P}(x_i) \le 1$, has been inherited from the original definition of an IFS (Atanassov, 1986; Atanassov, 2012). Hence, an AAIFS can be used for the characterization of evaluations like the given by Chloe even if the evaluations are marked by hesitation. Because of this and given that an AAIFS can be used in situations when no constraint on the attributes of the collection of potential options has been established, the AAIFS concept has been used in the formulation of a FAST-GDM problem as follows (Loor et al., 2018):

Consider a discrete collection $X = \{x_1, \cdots, x_n\}$ of *potential options* for a particular problem, and consider that $A \subseteq X$ represents a collection of *suitable options* for this problem. Consider also a collection $E = \{E_1, \cdots, E_m\}$ that represent a group of participants, experts or non-experts, who were asked to evaluate to which level each option in $X$ is member of $A$. Let

$$\hat{A}_{@E_j} = \left\{\langle x_i, \hat{\mu}_{A@E_j}(x_i), \hat{\nu}_{A@E_j}(x_i)\rangle \mid (x_i \in X) \wedge \left(0 \le \mu_{A@E_j}(x_i) + \nu_{A@E_j}(x_i) \le 1\right)\right\} \quad (2)$$

be an AAIFS characterizing the *individual* evaluations given by $E_j \in E$. Let

$$\hat{A} = \{\langle x_i, \hat{\mu}_A(x_i), \hat{\nu}_A(x_i)\rangle \mid (x_i \in X) \wedge (0 \le \mu_A(x_i) + \nu_A(x_i) \le 1)\} \quad (3)$$

be an AAIFS characterizing the *collective* evaluations computed for the group of participants $E$. Finally, let $\mathrm{cix}(\hat{A}_{@E_j}, \hat{A})$ be a function, called *concordance index*, that computes the level of concordance between $\hat{A}_{@E_j}$ and $\hat{A}$, where a higher value denotes a higher concordance between $\hat{A}_{@E_j}$ and $\hat{A}$. In this context, a FAST-GDM problem boils down to finding the most suitable option(s) in such a way that the average of all the *concordance indices*, i.e., $\frac{1}{m}\sum_{E_j \in E} \mathrm{cix}(\hat{A}_{@E_j}, \hat{A})$, is maximized.

As can be noticed, the computation of a concordance index is a very influential step in a FAST-GDM problem. For this reason, the selection of an adequate method for its computation is deemed to be an important and challenging task.

An option to perform such computation is through a function, say $S$, that computes the similarity between $\hat{A}_{@E_j}$ and $\hat{A}$, i.e., a concordance index can be computed by

$$\mathrm{cix}(\hat{A}_{@E_j}, \hat{A}) = S(\hat{A}_{@E_j}, \hat{A}). \quad (4)$$

As will be shown in the next section, in this work we use similarity measures that have been designed to compare traditional IFSs.

# 3 PILOT STUDY

As was mentioned in Section 1, the aim of this paper is to study how well a theoretical concordance index could reflect a perceived level of concordance between individual and collective evaluations in FAST-GDM problems. Hence in this section, we describe a procedure to get evaluations from people with different opinions on how the attributes of the given options should be evaluated. Then, we explain how this procedure has been used to get evaluations from a heterogeneous group of participants who tried to reach consensus about the best smooth dip(s) to pair with banana chips. After that, we describe how the individual and the collective evaluations given by the aforementioned group of participants have been compared by theoretical concordance indices and perceived levels of concordance.

## 3.1 Getting Evaluations

To get evaluations from persons who express not only the level to which but also the reasons why a potential option is suitable (or unsuitable) for a problem, we use a form like the one shown in Figure 1. Notice that, through this kind of form, a person can indicate how suitable and how unsuitable an option could be along with (some of) the reasons that justify his/her appraisal.

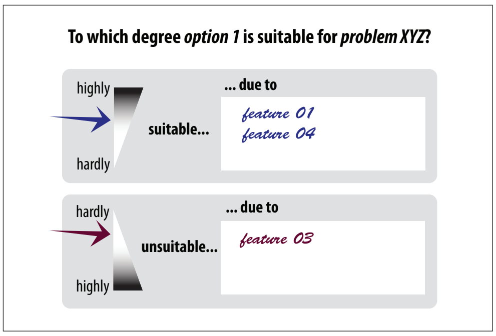



Figure 1: Form for evaluating an option.

The evaluations of the potential options filled out by a person, say $P$, in such a form can easily be characterized as an AAIFS by means of the following procedure:

Consider a discrete collection $X = \{x_1, \cdots, x_n\}$ of *potential options* for the problem under study, and consider that $A \subseteq X$ represents a collection of *suitable options* for this problem. Then, consider that the levels of suitability and unsuitability filled out by $P$ for any $x_i \in X$ can be linked to two values in a unit interval scale, say $\check{\mu}_{A@P}(x_i)$ and $\check{\nu}_{A@P}(x_i)$ respectively, where 1 denotes the highest value and 0 the lowest. Finally, consider that the features filled out for any $x_i \in X$ can be included in two collections, say $F_{\mu_{A@P}}(x_i)$ and $F_{\nu_{A@P}}(x_i)$. With these considerations, compute $\eta = \max(1, (\check{\mu}_{A@P}(x_i) + \check{\nu}_{A@P}(x_i))), \forall x_i \in X$. After that, obtain an AAIFS element $\langle x_i, \hat{\mu}_{A@P}(x_i), \hat{\nu}_{A@P}(x_i)\rangle$ for each option $x_i \in X$ such that $\hat{\mu}_{A@P}(x_i) = \langle \check{\mu}_{A@P}(x_i)/\eta, F_{\mu_{A@P}}(x_i)\rangle$ and $\hat{\nu}_{A@P}(x_i) = \langle \check{\nu}_{A@P}(x_i)/\eta, F_{\nu_{A@P}}(x_i)\rangle$.

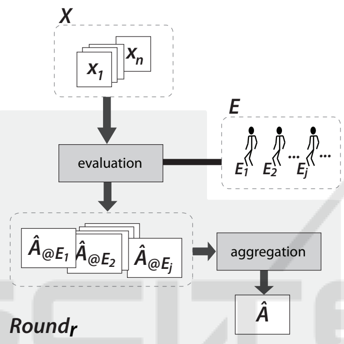


Figure 2: Getting evaluations.

We followed the above procedure to obtain the evaluations from 11 persons, 6 women and 5 men, who tried to reach consensus about the best smooth dip(s), among 3 potential dips, to pair with banana chips. This CRP was modeled using the notation introduced in Section 2 as follows: the collection of potential dips was denoted by $X = \{x_1, x_2, x_3\}$; the collection of the best smooth dip to pair with banana chips was denoted by $A$; the group of participants was denoted by $E = \{E_1, \cdots, E_{11}\}$; the individual evaluations given any participant $E_j \in E$ were represented by an AAIFS $\hat{A}_{@E_j}$; and the collective evaluations computed for the group were represented by an AAIFS $\hat{A}$ (see Figure 2). This means that 12 AAIFSs (1 AAIFS representing the collective evaluations and 11 AAIFSs representing the individual evaluations) were obtained in each of the 2 rounds completed by these 11 persons. In the next part, we show how these evaluations have been compared to each other.

## 3.2 Contrasting Evaluations

To compare a collection of individual evaluations, say $\hat{A}_{@E_j}$, with the collection of collective evaluations, say $\hat{A}$, obtained during each round of the CRP, we use theoretical concordance indices and perceived level of concordance as explained below.

### 3.2.1 Theoretical Concordance Indices

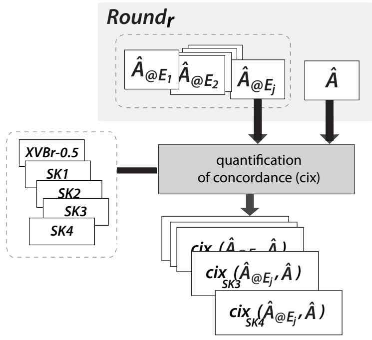


Figure 3: Quantification of theoretical concordance indices.

The concordance index between each $\hat{A}_{@E_j}$ and $\hat{A}$ was computed by means of (4), where $S$ was chosen among five (configurations of) similarity measures designed to compare traditional IFSs, namely *XVBr-0.5* (Loor and De Tré, 2017b), *SK1*, *SK2*, *SK3* and *SK4* (Szmidt and Kacprzyk, 2004) (see Figure 3). A *flat operator*, $\dashv\!|\cdot|\!\vdash$, which turns an AAIFS into an IFS by excluding the collections of features recorded in each AAIFS element, was used for converting $\hat{A}_{@E_j}$ and $\hat{A}$ into two IFSs, say $J$ and $A$ respectively, that are used as input for any of the chosen similarity measures. Regarding *XVBr-0.5*, it refers to a configuration of

$$S^{\alpha}_{@A}(J,A) = \Delta_{@A} \cdot S^{\alpha}(J,A), \qquad (5)$$

in which $\alpha$ has been set to 0.5. In this equation, $\Delta_{@A} \in [0,1]$ is a factor that was computed through the method *spotRatios* proposed in (Loor and De Tré, 2017b) and $S^{\alpha}(J,A)$ is given by

$$S^{\alpha}(J,A) = 1 - \frac{1}{n}\sum_{i=1}^{n}\Big|(\mu_A(x_i) - \mu_J(x_i)) + \alpha(h_A(x_i) - h_J(x_i))\Big|. \qquad (6)$$

With respect to *SK1*, *SK2*, *SK3* and *SK4*, they refer to following similarity measures:

$$S_{SK1}(J,A) = 1 - f(l(J,A), l(J,A^c)), \qquad (7)$$

$$S_{SK2}(J,A) = \frac{1 - f(l(J,A), l(J,A^c))}{1 + f(l(J,A), l(J,A^c))}, \qquad (8)$$

$$S_{SK3}(J,A) = \frac{(1 - f(l(J,A), l(J,A^c)))^2}{(1 + f(l(J,A), l(J,A^c)))^2} \qquad (9)$$

and

$$S_{SK4}(J,A) = \frac{e^{-f(l(J,A), l(J,A^c))} - e^{-1}}{1 - e^{-1}}. \qquad (10)$$

Herein, $A^c$ is the complement of $A$, i.e.,

$$A^c = \{\langle x_i, \nu_A(x_i), \mu_A(x_i)\rangle | (x_i \in X) \wedge (0 \le \mu_A(x_i) + \nu_A(x_i) \le 1)\}, \quad (11)$$

$l(J,A)$ has been set as the Hamming distance between $J$ and $A$ (Szmidt and Kacprzyk, 2004), i.e.,

$$l(J,A) = \frac{1}{2n}\sum_{i=1}^{n}\left(|\mu_A(x_i) - \mu_J(x_i)| + |\nu_A(x_i) - \nu_J(x_i)| + |h_A(x_i) - h_J(x_i)|\right), \quad (12)$$

and

$$f(l(J,A), l(J,A^c)) = \frac{l(J,A)}{l(J,A) + l(J,A^c)}. \quad (13)$$

The interest reader is referred to (Loor and De Tré, 2017c) for an open-source implementation of the above-mentioned similarity measures.

To obtain the perceived levels of concordance between each $\hat{A}_{@E_j}$ and $\hat{A}$, the AAIFSs were graphically represented by means of *IFS contrasting charts* as explained in the next part.

### 3.2.2 Perceived Levels of Concordance

Aiming to facilitate the interpretation of the evaluations given by a person or computed for a group during a CRP, we propose a novel visual representation of an IFS, which is called *IFS contrasting chart* or IFSCC for short.

The idea behind an IFSCC can be described through the following analogy. Consider that the evaluation of an option, say $x_i$, is a buoy floating on the surface of the sea. The air column inside this buoy corresponds to the level to which $x_i$ is a suitable option, i.e., the air column corresponds to $\mu_A(x_i)$. Likewise, the ballast in the buoy corresponds to the level to which $x_i$ is an unsuitable option, i.e., the ballast corresponds to $\nu_A(x_i)$. This means that the *buoyancy* of the buoy, say $\rho_A(x_i)$, results from the difference between $\mu_A(x_i)$ and $\nu_A(x_i)$, i.e., the buoyancy corresponds to $\rho_A(x_i) = \mu_A(x_i) - \nu_A(x_i)$. While a positive value of $\rho_A(x_i)$ suggests that $x_i$ is a *suitable option*, a negative value suggests that $x_i$ is an *unsuitable option*. The height of buoy is limited to the unit interval $[0,1]$ because of the *consistency condition*, i.e., $0 \le \mu_A(x_i) + \nu_A(x_i) \le 1$, which is expressed in the definition of an IFS. Figure 4 illustrates the evaluations of two options, $x_1$ and $x_2$, using this analogy.

We can use the above analogy to represent the appraisal levels in an AAIFS, say $\hat{A}_{@E_{11}}$, which characterizes the evaluations of the collection of potential dips performed by person $E_{11}$ during the first round. This representation is shown in Figure 5. Notice in

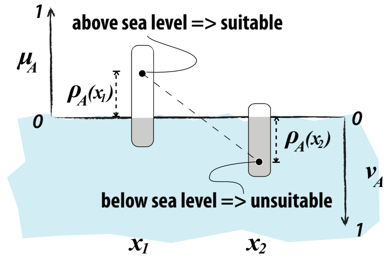



Figure 4: Idea behind an IFS contrasting chart.

this figure that the collection of potential dips is denoted by $X = \{x_1, x_2, x_3\}$. Notice also that the 'buoy' related to dip $x_1$ has equal parts of 'air' and 'ballasts', i.e., $\mu_A(x_1) = \nu_A(x_1)$, and, thus, the buoyancy of $x_1$ is 0. This suggests that $x_1$ has some 'positive' features but also some 'negative' features that make $E_{11}$ to think that $x_1$ neither satisfies nor dissatisfies the membership in the collection of the best smooth dips to pair with banana chips. In a similar way, notice that the 'buoy' related to dip $x_2$ has less 'air' than 'ballasts' and, thus, $x_2$ has a negative buoyancy ($\rho_A(x_2) = -0.48$). This indicates that $x_2$ might not be included by $E_{11}$ in the collection of the best smooth dips. Finally, notice that the 'buoy' related to dip $x_3$ has more 'air' than 'ballasts' and, thus, $x_3$ has a positive buoyancy ($\rho_A(x_3) = 0.39$). This suggests that $E_{11}$ might include $x_3$ in the collection of the best smooth dips.

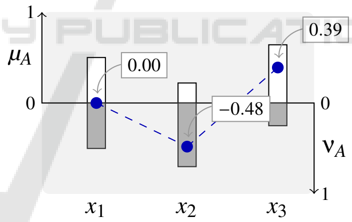



Figure 5: Evaluations of the dips $x_1$, $x_2$ and $x_3$ given by $E_{11}$.

It is worth mentioning that, although the *hesitation margin*, i.e., $h_A(x_i) = 1 - (\mu_A(x_i) + \nu_A(x_i))$ (Atanassov, 1986; Atanassov, 2012), is not explicitly depicted in an IFSCC, it can be inferred. For instance, Figure 6 depicts the evaluations given by $E_{10}$. Notice that $\mu_A(x_1)$ and $\nu_A(x_1)$ are equal to 0. This suggests that the evaluation of dip $x_1$ has been missed or that $E_{10}$ did not try dip $x_1$ during the first round. Hence, considering $x_1$ as a member (or not) of the collection of the best dips is marked by a high hesitation in this case.

Notice in Figure 5 and Figure 6 that an IFSCC shows a holistic view of the evaluations given by a participant in a CRP. Thus, to obtain the perceived level of concordance between the individual and collective evaluations in the CRP about the best smooth

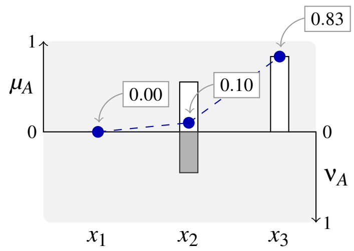



Figure 6: Evaluations of the dips $x_1$, $x_2$ and $x_3$ given by $E_{10}$.

dips, we built the IFSCCs of the 24 AAIFSs (22 representing individual evaluations and 2 representing collective evaluations) obtained after the 2 rounds of this CRP.

We made use of those IFSCCs, which are depicted in Figures 12 and 13 (see Appendix), to make 22 pairs of IFSCCs in such a way that the individual evaluations given by each participant could be compared with the computed collective evaluations in each round – i.e., 11 pairs of IFSCCs corresponding to pairs of AAIFSs $\langle \hat{A}_{@E_j}, \hat{A} \rangle$, $j = 1, \cdots, 11$, were built for each round.

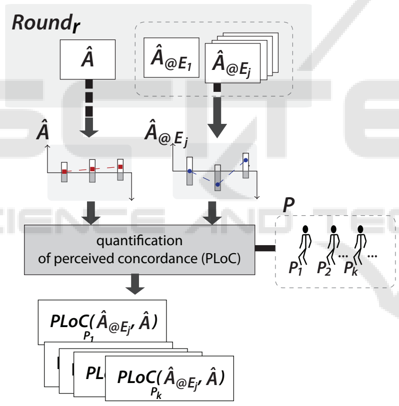



Figure 7: Quantification of perceived levels of concordance.

Using those 22 pairs of IFSCCs, we asked a group of 13 persons having managerial roles to quantify the level of concordance that each of them perceives between each pair of IFSCCs (see Figure 7). To indicate so, the use of a unit interval scale where 1 and 0 represent the highest and the lowest level of concordance respectively was recommended.

### 3.2.3 Theoretical vs. Perceived Concordance

To compare the perceived levels of concordance given by the group of 13 people with the theoretical concordance indices computed by the similarity measures mentioned in Section 3.1, we followed two approaches: a *macro* approach in which the average of the perceived levels of concordance is contrasted with each concordance index; and a *micro* approach in which each perceived level of concordance is contrasted with each concordance index. These comparisons are as follows. Consider that $P = \{P_1, \cdots, P_{13}\}$ represents the group of the 13 persons. Consider also that $\mathrm{PLoC}_{P_k}(\hat{A}_{@E_j}, \hat{A})$ represents the level of concordance between $\hat{A}_{@E_j}$ and $\hat{A}$ perceived by any $P_k \in P$. With these considerations, in the macro comparison we first compute the average of the perceived levels for each pair $\langle \hat{A}_{@E_j}, \hat{A} \rangle$ by means of

$$\overline{\mathrm{PLoC}}(\hat{A}_{@E_j}, \hat{A} \rangle) = \frac{1}{|P|} \sum_{P_k \in P} (\mathrm{PLoC}_{P_k}(\hat{A}_{@E_j}, \hat{A})), \quad (14)$$

where $|P|$ denotes the number of persons having managerial roles. Then, we compute the absolute error between the average of the perceived level of concordance and a given concordance index, say cix, for each pair $\langle \hat{A}_{@E_j}, \hat{A} \rangle$ by means of

$$\Delta_{\mathrm{cix}}(\hat{A}_{@E_j}, \hat{A}) = |\overline{\mathrm{PLoC}}(\hat{A}_{@E_j}, \hat{A}) - \mathrm{cix}(\hat{A}_{@E_j}, \hat{A})|, \quad (15)$$

where $|\cdot|$ denotes the absolute value. After that, we compute a *macro mean absolute error* by means of

$$\overline{\Delta}_{\mathrm{cix}} = \frac{1}{|E|} \sum_{E_j \in E} \Delta_{\mathrm{cix}}(\hat{A}_{@E_j}, \hat{A}), \quad (16)$$

where $|E|$ represents the number of participants in $E$. In this case, a value of $\overline{\Delta}_{\mathrm{cix}}$ close to 0 means that cix reflects well the perceived level of concordance.

Regarding the micro comparison, we compute for a given person in $P$, say $P_k$, the absolute error between the perceived level of concordance and a given concordance index, say cix, for each pair $\langle \hat{A}_{@E_j}, \hat{A} \rangle$ by means of

$$\delta_{\mathrm{cix},P_k}(\hat{A}_{@E_j}, \hat{A}) = |\mathrm{PLoC}_{P_k}(\hat{A}_{@E_j}, \hat{A} \rangle) - \mathrm{cix}(\hat{A}_{@E_j}, \hat{A})|. \quad (17)$$

Then, we aggregate all these absolute errors through a *micro mean absolute error* computed by

$$\overline{\delta}_{\mathrm{cix}} = \frac{1}{|E| \times |P|} \sum_{E_j \in E, P_k \in P} \delta_{\mathrm{cix},P_k}(\hat{A}_{@E_j}, \hat{A}). \quad (18)$$

A value of $\overline{\delta}_{\mathrm{cix}}$ close to 0, as analogous to $\overline{\Delta}_{\mathrm{cix}}$, means that cix reflects well the perceived level of concordance.

The results obtained after performing the aforementioned comparisons are shown in the next section.

## 4 RESULTS AND DISCUSSION

In this section, we present the results and our findings on how well each of the chosen theoretical concordance indices reflected the perceived levels of concordance.

Table 1: Theoretical Concordance Indices and Average of Perceived Levels of Concordance (Round 1).

| Pair | *XVBr-0.5* | *SK1* | *SK2* | *SK3* | *SK4* | $\overline{\text{PLoC}}$ |
|---|---|---|---|---|---|---|
| $\langle\hat{A}_{@E_1},\hat{A}\rangle$ | 0.75 | 0.58 | 0.41 | 0.17 | 0.46 | 0.54 |
| $\langle\hat{A}_{@E_2},\hat{A}\rangle$ | 0.19 | 0.48 | 0.32 | 0.10 | 0.36 | 0.20 |
| $\langle\hat{A}_{@E_3},\hat{A}\rangle$ | 0.59 | 0.41 | 0.26 | 0.07 | 0.29 | 0.50 |
| $\langle\hat{A}_{@E_4},\hat{A}\rangle$ | 0.00 | 0.42 | 0.27 | 0.07 | 0.30 | 0.29 |
| $\langle\hat{A}_{@E_5},\hat{A}\rangle$ | 0.76 | 0.59 | 0.42 | 0.18 | 0.47 | 0.53 |
| $\langle\hat{A}_{@E_6},\hat{A}\rangle$ | 0.45 | 0.55 | 0.38 | 0.14 | 0.43 | 0.62 |
| $\langle\hat{A}_{@E_7},\hat{A}\rangle$ | 0.88 | 0.55 | 0.38 | 0.14 | 0.42 | 0.59 |
| $\langle\hat{A}_{@E_8},\hat{A}\rangle$ | 0.50 | 0.53 | 0.36 | 0.13 | 0.41 | 0.32 |
| $\langle\hat{A}_{@E_9},\hat{A}\rangle$ | 0.25 | 0.47 | 0.31 | 0.10 | 0.35 | 0.35 |
| $\langle\hat{A}_{@E_{10}},\hat{A}\rangle$ | 0.74 | 0.54 | 0.37 | 0.14 | 0.42 | 0.63 |
| $\langle\hat{A}_{@E_{11}},\hat{A}\rangle$ | 0.43 | 0.52 | 0.35 | 0.12 | 0.40 | 0.60 |

Table 2: Theoretical Concordance Indices and Average of Perceived Levels of Concordance (Round 2).

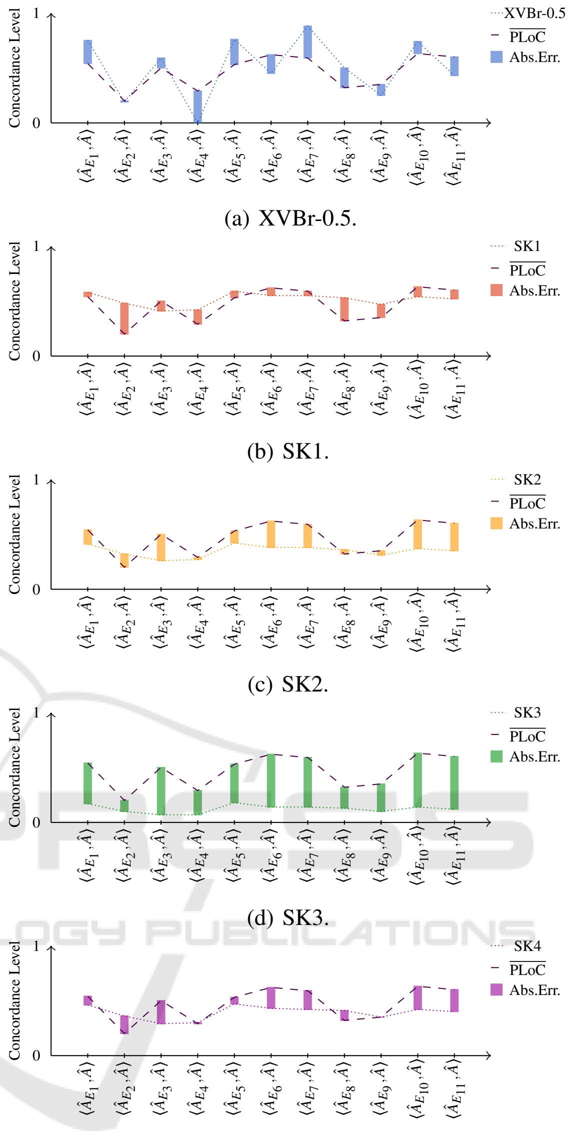
| Pair | *XVBr-0.5* | *SK1* | *SK2* | *SK3* | *SK4* | $\overline{\text{PLoC}}$ |
|---|---|---|---|---|---|---|
| $\langle\hat{A}_{@E_1},\hat{A}\rangle$ | 0.43 | 0.49 | 0.33 | 0.11 | 0.37 | 0.43 |
| $\langle\hat{A}_{@E_2},\hat{A}\rangle$ | 0.40 | 0.52 | 0.35 | 0.13 | 0.40 | 0.18 |
| $\langle\hat{A}_{@E_3},\hat{A}\rangle$ | 0.00 | 0.34 | 0.21 | 0.04 | 0.24 | 0.15 |
| $\langle\hat{A}_{@E_4},\hat{A}\rangle$ | 0.20 | 0.50 | 0.34 | 0.11 | 0.38 | 0.14 |
| $\langle\hat{A}_{@E_5},\hat{A}\rangle$ | 0.68 | 0.61 | 0.43 | 0.19 | 0.48 | 0.43 |
| $\langle\hat{A}_{@E_6},\hat{A}\rangle$ | 0.55 | 0.65 | 0.48 | 0.23 | 0.53 | 0.68 |
| $\langle\hat{A}_{@E_7},\hat{A}\rangle$ | 0.54 | 0.66 | 0.49 | 0.24 | 0.54 | 0.76 |
| $\langle\hat{A}_{@E_8},\hat{A}\rangle$ | 0.53 | 0.45 | 0.29 | 0.08 | 0.33 | 0.36 |
| $\langle\hat{A}_{@E_9},\hat{A}\rangle$ | 0.50 | 0.61 | 0.44 | 0.19 | 0.49 | 0.46 |
| $\langle\hat{A}_{@E_{10}},\hat{A}\rangle$ | 0.53 | 0.49 | 0.33 | 0.11 | 0.37 | 0.52 |
| $\langle\hat{A}_{@E_{11}},\hat{A}\rangle$ | 0.56 | 0.65 | 0.48 | 0.23 | 0.53 | 0.50 |

## 4.1 Results

The computed theoretical concordance indices, identified by *XVBr-0.5*, *SK1*, *SK2*, *SK3* and *SK4*, along with the macro average of the perceived levels of concordance, which is denoted by $\overline{\text{PLoC}}$ and computed by (14), are shown in Table 1 and Table 2: while the data in the former table correspond to the 11 pairs of IFSCCs obtained during the first round (see Figure 12 in Appendix), the data in the latter table correspond to the 11 pairs of IFSCCs obtained during the second round (see Figure 13 in Appendix).

We make use of the data in the aforementioned tables as inputs of (15) to compute $\Delta_{\text{cix}}(\hat{A}_{@E_j},\hat{A})$ for each of the theoretical concordance indices. The results are depicted as bars in Figure 8 (Round 1) and Figure 9 (Round 2). For instance, in Figure 8(a) the lowest and the highest absolute errors between *XVBr-0.5* and $\overline{\text{PLoC}}$ are related to the pairs $\langle\hat{A}_{E_2},\hat{A}\rangle$ and $\langle\hat{A}_{E_4},\hat{A}\rangle$, which result from $\Delta_{\text{XVBr-0.5}}(\hat{A}_{@E_2},\hat{A}) = |0.19-0.20| = 0.01$ and $\Delta_{\text{XVBr-0.5}}(\hat{A}_{@E_4},\hat{A}) = |0-0.29| = 0.29$ respectively.

For the sake of a better macro comparison, a frequency distribution of the absolute errors corresponding to the two rounds are depicted in Figure 10. As an example of the frequency distribution, in Figure 10(b) it is shown that, while 12 out of 22 (i.e., 54.54%) of the computed absolute errors between SK1 and $\overline{\text{PLoC}}$ are placed in the interval $[0,0.1]$, 6 out of 22 (i.e., 27.27%) are located in the interval $(0.1,0.2]$. The values computed by (16) are also shown in Figure 10. For instance, the macro mean absolute error between SK1 and $\overline{\text{PLoC}}$, i.e., $\overline{\Delta}_{\text{SK1}} = 0.13$, is indicated in the center of Figure 10(b).




(a) XVBr-0.5.

(b) SK1.

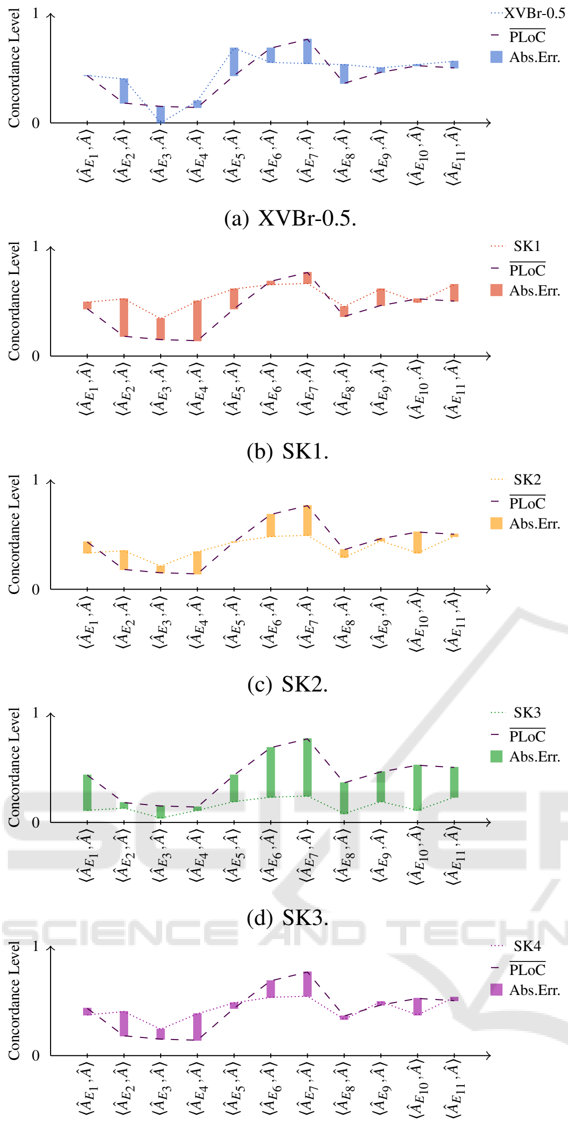
(c) SK2.

(d) SK3.

(e) SK4.

Figure 8: Absolute Errors (Round 1).

Regarding the comparisons using a micro approach, a frequency distribution of the results computed by (17) and (18) are depicted in Figure 11. In Figure 11(c), e.g., it is shown that, while 68 out of 286 (i.e., 23.77%) of the computed absolute errors between SK1 and PLoC are located in the interval $[0,0.1]$, 74 out of 286 (i.e., 25.87%) are placed in the




(a) XVBr-0.5.

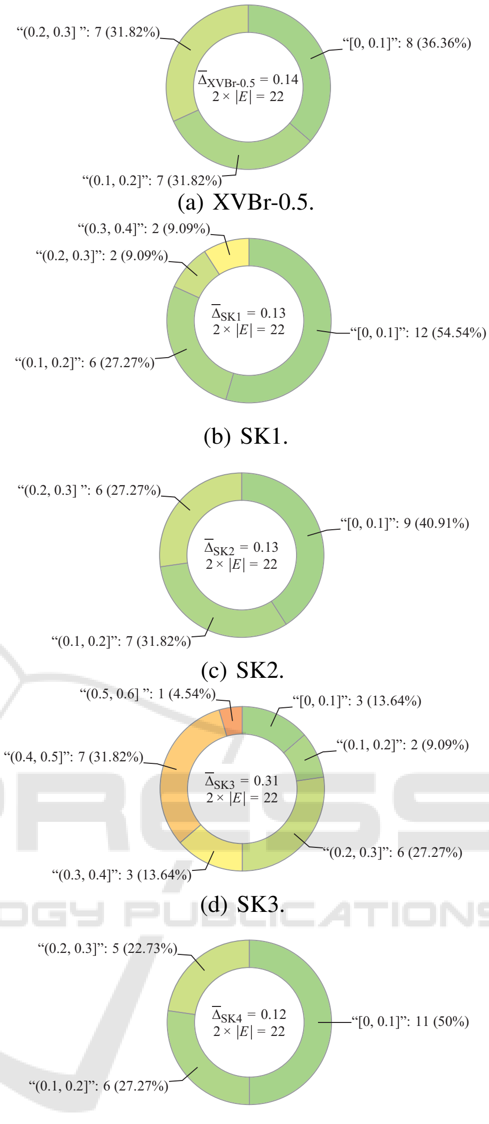
(b) SK1.

(c) SK2.

(d) SK3.

(e) SK4.

Figure 9: Absolute Errors (Round 2).




(a) XVBr-0.5.

(b) SK1.

(c) SK2.

(d) SK3.

(e) SK4.

Figure 10: Distribution of Abs. Errors (Macro comparisons).

Table 3: Macro and micro mean absolute errors.

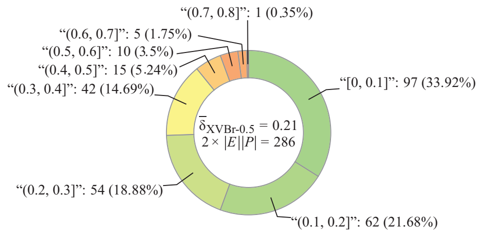
| | *XVBr-0.5* | *SK1* | *SK2* | *SK3* | *SK4* |
|---|---|---|---|---|---|
| $\overline{\Delta}_{\mathrm{cix}}$ | 0.14 | 0.13 | 0.13 | 0.31 | **0.12** |
| $\overline{\delta}_{\mathrm{cix}}$ | **0.21** | **0.21** | 0.22 | 0.33 | **0.21** |

interval $(0.1, 0.2]$. In this case, the micro mean absolute error between SK2 and PLoC, i.e., $\overline{\delta}_{\mathrm{SK2}} = 0.22$, is indicated in the center of Figure 11(c). The results are summarized in Table 3.

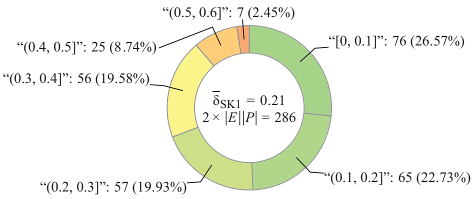
## 4.2 Discussion

The results listed in Table 3 suggest that the concordance indices (based on the similarity measures) *XVBr-0.5*, *SK1*, *SK2* and *SK4* reflect to an acceptable extent the perceived level of concordance between the individual and the collective evaluations.

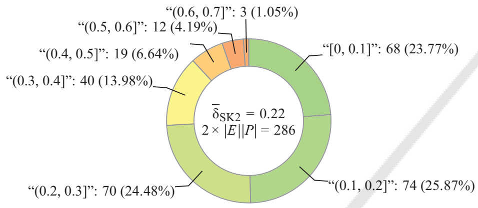
As far as one can see, *SK4* slightly outperforms *XVBr-0.5*, *SK1* and *SK2* according to both the macro and the micro mean absolute errors. Notice that, while the expression $(\overline{\Delta}_{\mathrm{SK4}} = 0.12) < (\overline{\Delta}_{\mathrm{SK1}} = 0.13) \leq (\overline{\Delta}_{\mathrm{SK2}} = 0.13) < (\overline{\Delta}_{\mathrm{XVBr\text{-}0.5}} = 0.14)$ holds for the macro mean absolute errors, the expression $(\overline{\delta}_{\mathrm{SK4}} = 0.21) \leq (\overline{\delta}_{\mathrm{SK1}} = 0.21) \leq (\overline{\delta}_{\mathrm{XVBr\text{-}0.5}} = 0.21) < (\overline{\delta}_{\mathrm{SK2}} = 0.22)$ holds for the micro mean absolute errors. However, such a slightly advantage of *SK4* can disappear if someone focuses only on the frequency of the abso-



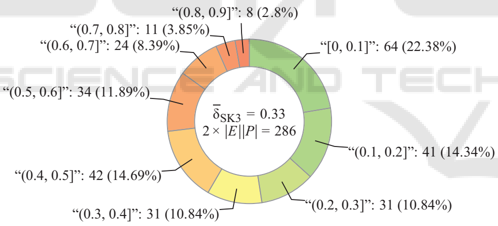

(a) XVBr-0.5.



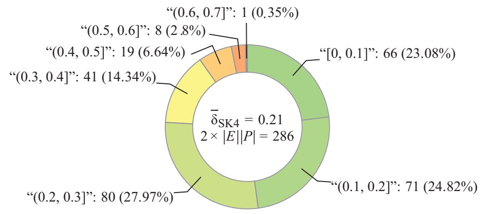

(b) SK1.




(c) SK2.




(d) SK3.




(e) SK4.

Figure 11: Distribution of Abs. Errors (Micro Comparisons).

lute errors between the concordance indices and the average of perceived levels of concordance located in the interval $[0,0.1]$ (i.e., $\Delta_{\text{cix}} \in [0,0.1]$) – notice in Figure 10 that the frequency of $\Delta_{\text{SK1}} \in [0,0.1]$, i.e., 12, is slightly greater than the frequency of $\Delta_{\text{SK4}} \in [0,0.1]$, i.e., 11. That advantage can also disappear when someone only takes into account the frequency of the absolute errors between the concordance indices and the perceived levels of concordance located in the interval $[0,0.1]$ (i.e., $\delta_{\text{cix}} \in [0,0.1]$) – notice in Figure 11 that the frequency of $\delta_{\text{XVBr-0.5}} \in [0,0.1]$, i.e., 97, is greater than the frequency of $\delta_{\text{SK4}} \in [0,0.1]$, i.e., 66. Hence, we can say that, although *SK4* has a slightly advantage, in this pilot study the concordance indices based on the similarity measures *XVBr-0.5*, *SK1*, *SK2* and *SK4* are comparatively alike when reflecting the perceived levels of concordance between the individual and the collective evaluations. A practical implication of these results is that these concordance indices can be accepted as good indicators of the level of concordance in FAST-GDM problems.

Regarding the concordance index based on the similarity measure *SK3*, the computed macro and micro mean absolute errors suggest that this index might not reflect well the perceived level of concordance. Notice in Figure 10(d) and Figure 11(d) that roughly 50% of the computed absolute errors are located in the interval $(0.30,1]$.

It is worth mentioning that, although in the first round the individual evaluations characterized by $\hat{A}_{@E_4}$ (see Figure 12(e)) and the collective evaluations characterized by $\hat{A}$ (see Figure 12(a)) can lead to potential opposite decisions about the options $x_1$, $x_2$ and $x_3$, the average of the perceived level of concordance related to the pair $\langle \hat{A}_{@E_4}, \hat{A} \rangle$, i.e., $\overline{\text{PLoC}}(\hat{A}_{@E_4}, \hat{A}) = 0.29$, is appreciably greater than the expected theoretical value for this case, i.e., 0 (Loor and De Tré, 2017b). A potential explanation for this result might be that a more clear indication of the potential decision that can be taken after studying the evaluations represented in an IFSCC is needed. In this regard, conducting a follow-up study in which the perceived levels of concordance would be obtained by comparing IFSCCs that have been augmented with additional information (e.g., an explicit ranking of the options, or a linguistic summary about a potential decision (Kacprzyk and Zadrożny, 2005)) is suggested.

Another suggested study concerns the use of scales of measurement formed from linguistic labels such as *'highly concordant'* or *'hardly concordant'* to map or express the results of the concordance indices, as well as the perceived levels of concordance (Herrera et al., 1996; Herrera et al., 1997). Since the macro averages computed for the concordance indices *XVBr-0.5*, *SK1*, *SK2* and *SK4* are roughly 0.12, we foresee that a scale consisting of 5 linguistic labels, for instance *'hardly concordant'*, *'not specially concordant'*, *'slightly concordant'*, *'fairly concordant'* and *'highly concordant'*, can further improve the level to which those concordance indices reflect the perceived levels of concordance.

## 5 RELATED WORK

Methods that obtain the level of concordance (or agreement) between the evaluations given by two individuals by means of a similarity (or distance) measure defined in the IFS framework can be found in the literature. For instance, in (Szmidt and Kacprzyk, 2002; Szmidt and Kacprzyk, 2003; Szmidt and Kacprzyk, 2004) Szmidt and Kacprzyk proposed the use of either similarity or distance measures between two IFSs to compute the level of agreement between two participants whose evaluations have been characterized as IFSs. However, to the best of our knowledge, no empirical study oriented to determine how well the results computed by such similarity measures reflect the perceived levels of concordance in a consensus reaching process has been published so far.

Regarding the visual representations of IFSs, several geometrical interpretations of IFSs are available in the literature. The standard interpretation, in which the membership and the nonmembership components, i.e., $\mu_A(x)$ and $\nu_A(x)$, are depicted in a common region (or band) of height 1, is the most accepted visual representation of IFSs according to Atanassov (Atanassov, 2012). A variant of the standard interpretation is one in which is depicted $(1-\nu_A(x))$ instead of $\nu_A(x)$. This variant leads to a representation where each element in an IFS is depicted a unit segment. Another representation, called *IFS-interpretational triangle*, is based on a right triangle having two sides of length 1: one for $\mu_A(x)$ and the other for $\nu_A(x)$ (Atanassov, 1986; Atanassov, 2012). While an advantage of the IFS-interpretational triangle is that an operation over an IFS element can, in general, be easily visualized, a disadvantage of this representation is that a holistic view of all the IFS elements can be unclear. As a variant of the IFS-interpretational triangle, the idea of a *'unit cube'* to additionally represent the hesitation margin has been introduced in (Szmidt and Kacprzyk, 2004). Using a different approach, the representation of an IFS by means of radar charts has been proposed in (Atanassova, 2010). Since radar charts are typically used in a business-oriented environment, an IFS radar chart can be considered as a business-oriented representation of an IFS. However, the representation of the two components in the same circle (or band) can be confusing when the representation of the buoyancy of the elements of an IFS is needed. Because of this, we propose a novel business-oriented representation of an IFS by means of an IFSCC. As could be noticed throughout this paper, two bands, one for $\mu_A(x)$ and the other for $\nu_A(x)$, are used within an IFSCC to depict in a holistic way the buoyancy of each element in an IFS.

## 6 CONCLUSIONS

In this paper, we have described a pilot study in which several theoretical concordance indices based on similarity measures designed to compare *intuitionistic fuzzy sets* (IFSs) have been tested to determine their usability in *flexible attribute-set group decision-making* (FAST-GDM) problems.

During the study, the evaluations obtained from a group of participants who tried to find a solution for a FAST-GDM problem were characterized as *augmented Atanassov intuitionistic fuzzy sets* (AAIFSs). Then, each of those AAIFSs was graphically represented by means of a novel business-oriented representation called *IFS contrasting chart* (IFSCC). After that, a group of persons having managerial roles were asked to estimate the level of concordance between the individual and the collective evaluations depicted respectively in two IFSCCs. The perceived levels of concordance given by this group were then compared to the values computed by the theoretical concordance indices.

The results of this pilot study suggest that, among the five theoretical concordance indices chosen for the test, four reflected to an acceptable extent the perceived level of concordance between the individual and the collective evaluations. However, a more extended study should corroborate the usability of these concordance indices in practical situations involving FAST-GDM problems.

It was found a case in which the perceived level of concordance between the individual and collective evaluations is not the lowest even though such evaluations lead to complete opposite decisions. A possible explanation for this case might be that a more clear indication of the potential decisions is needed in the two IFSCCs representing such evaluations. Hence, a follow-up study in which additional information about a potential decision is incorporated into an IFSCC is suggested. Another suggested study concerns the usability of scales of measurement formed from linguistic labels to quantify both theoretical and perceived concordance levels.

## REFERENCES

Atanassov, K. T. (1986). Intuitionistic fuzzy sets. *Fuzzy sets and Systems*, 20(1):87–96.

Atanassov, K. T. (2012). *On Intuitionistic Fuzzy Sets Theory*, volume 283 of *Studies in Fuzziness and Soft Computing*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Atanassova, V. (2010). Representation of fuzzy and intuitionistic fuzzy data by radar charts. *Notes on Intuitionistic Fuzzy Sets*, 16(1):21–26.

Bouyssou, D., Dubois, D., Prade, H., and Pirlot, M. (2013). *Decision Making Process: Concepts and Methods*. John Wiley & Sons.

Dong, Y., Xiao, J., Zhang, H., and Wang, T. (2016). Managing consensus and weights in iterative multiple-attribute group decision making. *Applied Soft Computing*, 48:80 – 90.

Herrera, F., Herrera-Viedma, E., and Verdegay, J. L. (1996). A model of consensus in group decision making under linguistic assessments. *Fuzzy Sets and Systems*, 78(1):73 – 87.

Herrera, F., Herrera-Viedma, E., and Verdegay, J. L. (1997). A rational consensus model in group decision making using linguistic assessments. *Fuzzy Sets and Systems*, 88(1):31 – 49.

Kacprzyk, J. and Fedrizzi, M. (1988). A 'soft' measure of consensus in the setting of partial (fuzzy) preferences. *European Journal of Operational Research*, 34(3):316 – 325.

Kacprzyk, J. and Zadrożny, S. (2005). Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools. *Information Sciences*, 173(4):281 – 304. Dealing with Uncertainty in Data Mining and Information Extraction.

Kacprzyk, J. and Zadrożny, S. (2010). Soft computing and web intelligence for supporting consensus reaching. *Soft Computing*, 14(8):833–846.

Liu, Y., Fan, Z.-P., and Zhang, X. (2016). A method for large group decision-making based on evaluation information provided by participators from multiple groups. *Information Fusion*, 29:132–141.

Loor, M. and De Tré, G. (2017a). On the Need for Augmented Appraisal Degrees to Handle Experience-Based Evaluations. *Applied Soft Computing*, 54C:284 – 295.

Loor, M. and De Tré, G. (2017b). In a Quest for Suitable Similarity Measures to Compare Experience-Based Evaluations. In Merelo, J. J., Rosa, A., Cadenas, J. M., Correia, A. D., Madani, K., Ruano, A., and Filipe, J., editors, *Computational Intelligence: International Joint Conference, IJCCI 2015 Lisbon, Portugal, November 12-14, 2015, Revised Selected Papers*, volume 669 of *Studies in Computational Intelligence*, pages 291–314. Springer International Publishing.

Loor, M. and De Tré, G. (2017c). An open-source software package to assess similarity measures that compare intuitionistic fuzzy sets. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6.

Loor, M., Tapia-Rosero, A., and De Tré, G. (2018). Refocusing attention on unobserved attributes to reach consensus in decision making problems involving a heterogeneous group of experts. In Kacprzyk, J., Szmidt, E., Zadrożny, S., Atanassov, K. T., and Krawczak, M., editors, *Advances in Fuzzy Logic and Technology 2017*, pages 405–416, Springer International Publishing.

Szmidt, E. and Kacprzyk, J. (2002). Evaluation of agreement in a group of experts via distances between intuitionistic fuzzy preferences. In *Proceedings First International IEEE Symposium Intelligent Systems*, volume 1, pages 166–170 vol.1.

Szmidt, E. and Kacprzyk, J. (2003). A consensus reaching process under intuitionistic fuzzy preference relations. *International Journal of Intelligent Systems*, 18(7):837–852.

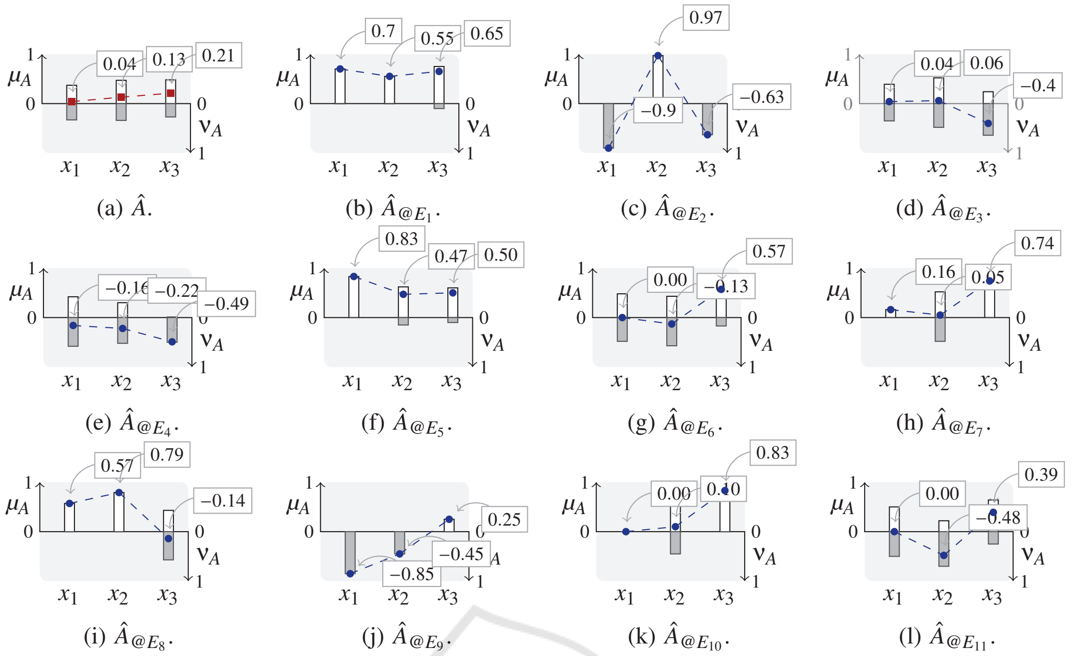
Szmidt, E. and Kacprzyk, J. (2004). A concept of similarity for intuitionistic fuzzy sets and its use in group decision making. In *IEEE International Conference on Fuzzy Systems*, pages 1129–1134.

## APPENDIX

### Evaluations about the 'best smooth dip'

This appendix presents the evaluations given by 11 persons who tried to reach a consensus about the best smooth dip(s), among 3 potential dips, to pair with banana chips. These evaluations have been graphically represented by means of IFS contrasting charts (IFSCCs) (see Section 3.2).

Figure 12 shows the IFSCCs corresponding to the evaluations given or computed during the first round of the consensus reaching process: while Figure 12(a) represents the collective evaluations computed for the group, Figures 12(b)-12(l) represent the individual evaluations given by these 11 persons respectively. In a similar way, Figure 13 shows the IFSCCs corresponding to the evaluations given or computed during the second round.

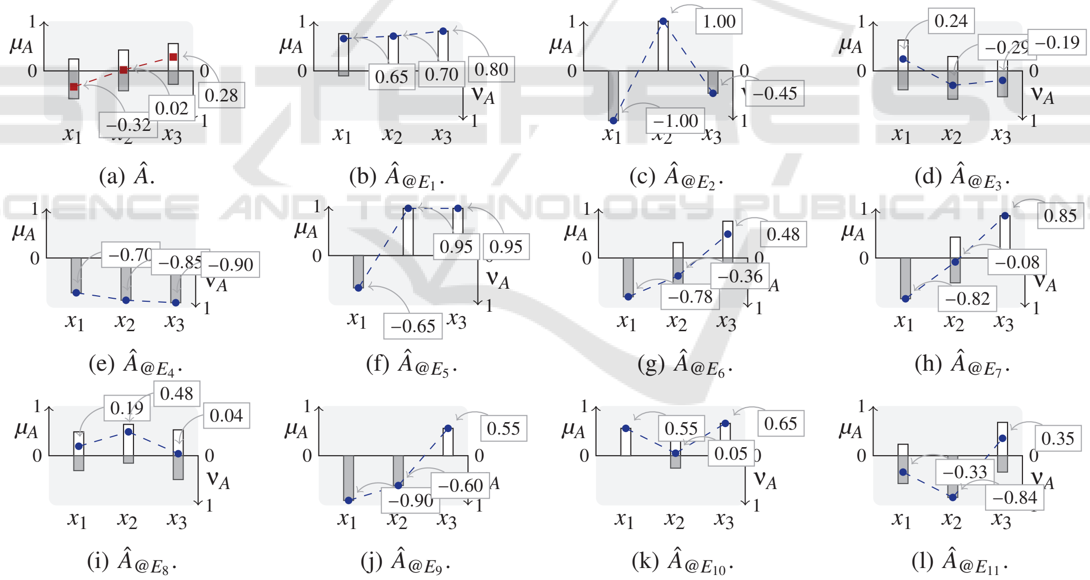



(a) $\hat{A}$. (b) $\hat{A}_{@E_1}$. (c) $\hat{A}_{@E_2}$. (d) $\hat{A}_{@E_3}$.

(e) $\hat{A}_{@E_4}$. (f) $\hat{A}_{@E_5}$. (g) $\hat{A}_{@E_6}$. (h) $\hat{A}_{@E_7}$.

(i) $\hat{A}_{@E_8}$. (j) $\hat{A}_{@E_9}$. (k) $\hat{A}_{@E_{10}}$. (l) $\hat{A}_{@E_{11}}$.

Figure 12: Evaluations obtained during Round 1.




(a) $\hat{A}$. (b) $\hat{A}_{@E_1}$. (c) $\hat{A}_{@E_2}$. (d) $\hat{A}_{@E_3}$.

(e) $\hat{A}_{@E_4}$. (f) $\hat{A}_{@E_5}$. (g) $\hat{A}_{@E_6}$. (h) $\hat{A}_{@E_7}$.

(i) $\hat{A}_{@E_8}$. (j) $\hat{A}_{@E_9}$. (k) $\hat{A}_{@E_{10}}$. (l) $\hat{A}_{@E_{11}}$.

Figure 13: Evaluations obtained during Round 2.