Interval Agreement Weighted Average – Sensitivity to Data Set Features

Yu Zhao[1], Christian Wagner[2], Brendan Ryan[3], Direnc Pekaslan[4], and Javier Navarro[5]
[1,2,4]*Lab for Uncertainty in Data and Decision Making (LUCID), School of Computer Science*
[3]*Human Factors Research Group, Faculty of Engineering*
[1,2,3,4]University of Nottingham, United Kingdom
[5]*School of Engineering and Sciences,* Technological Institute of Higher Studies of Monterrey, Mexico
Email: [1,2,3,4]{carina.zhao, christian.wagner, brendan.ryan, direnc.pekaslan1}@nottingham.ac.uk, [5]fj.navarro.barron@tec.mx

***Abstract*—The growing use of intervals in fields like survey analysis necessitates effective aggregation methods that can summarize and represent such uncertain data representations. The Interval Agreement Approach (IAA) addresses this by aggregating interval responses into Fuzzy Sets (FSs), capturing both intra- and inter-participant agreement, while minimizing information loss. While offering a powerful modeling tool, the IAA does not natively offer a measure of central tendency, which is itself an interval of particular utility in real-world applications. In contrast, the Interval Weighted Average (IWA) has been used for directly measuring the central tendency of intervals. While straightforward and effective, it is not designed, nor able to, summarize interval data in terms of their agreement, as the IAA does. To bridge this gap, this paper introduces Interval Agreement Weighted Average (IAWA), which is specifically designed to reflect both the central tendency and agreement. This is achieved by first modeling interval agreement as FSs using the IAA, and then transforming these FSs into intervals using the IWA. We demonstrate the approach by conducting sensitivity analyses to explore the behavior of the proposed approach in detail. Our findings suggest that the IAWA is a highly effective measure of central tendency. Additionally, it also partially inherits IAA's ability to reflect the agreement of sets of intervals. We conclude by highlighting the potential and growth of the use of intervals in information elicitation, within, and beyond survey research, underpinning a new degree of understanding of both intra- and inter-source uncertainty.**

***Index Terms*—Intervals, Uncertainty, Survey, Interval Agreement Approach**

I. Introduction

Collecting comprehensive quantitative survey responses accurately and effectively is key to understanding insights from people in various fields. This enables the aggregation and statistical analysis of responses, thereby facilitating an understanding of the collective viewpoints of specific groups and subgroups within the surveyed population. Since the early twentieth century, the primary approach for capturing responses in surveys has largely remained the use of single-point scales, such as Likert, numeric, semantic differential, and visual analogue [1]–[4]. While discrete response scales have advantages such as being easy to collect and analyze, by limiting responses to a single choice along a given continuum, these scales are potentially restrictive to participants' authentic expression [5]–[8]. Specifically, these scales are limited in their capacity to capture both forms of uncertainty: the complex viewpoints or uncertainty within *individual* participants (intra-participant uncertainty), and the diverse opinions among *different* participants (inter-participant uncertainty).

For intra-participant uncertainty, this is particularly problematic due to the inherent cognitive uncertainty and imprecision in individuals' subjective judgments [9]–[12]. For example, in a railway context, responding to a question like 'What is your overall satisfaction with the train station during your journey?', participants are compelled to select a single term, despite potentially having varied and complex experiences at different stations. Requiring a single selection in such scenarios can limit the natural expression of intra-participant uncertainty, leading to information loss and distortion.

For inter-respondent uncertainty, previous research emphasises that it is vital to determine whether survey response variations accurately reflect participants' diverse opinions or if they are influenced by survey method [7]. However, the requirement for a single response can obscure or exaggerate individual differences. Consider a scenario where two individuals faced the same train delay. When asked about the delay duration, for instance, 'How many minutes of delay did you experience?' their responses may differ, as one person might be clear about the delay time, while another may not have paid close attention to it. This difference might not truly represent distinct experiences but could instead be a result of the forced precision of the survey scale. Such nuances in responses can lead to an inaccurate capture of inter-participant uncertainty, potentially distorting the authenticity of survey findings.

The interval-valued (IV) response mode was proposed as an alternative to traditional single-point response modes [3], [13]. It allows participants to provide their answers by specifying an interval on a continuous scale, for example by circling an area of the scale. Crucially, the response mode does not restrict respondents to IV responses of a pre-determined size. As a result, respondents have full freedom in their expression, affording the effective and efficient capture of both intra- and inter-participant uncertainty [3]. The usability and efficacy of the IV scale have been tested in a variety of areas; from vulnerability assessments in cybersecurity, to consumer

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/S023305/1] and by Rail Safety and Standards Board.

perceptions of products, patient understanding in healthcare, and environmental management [14]–[20].

Fuzzy Sets (FSs), known for their ability to model uncertainty in human perceptions, have been widely used in interval analysis methods, enhancing the modeling of uncertainty in IV data. These include the Interval Approach (IA) and the Interval Agreement Approach (IAA), along with extensions: the Enhanced Interval Approach (EIA) and Hao-Mendel Approach (HMA) as the extensions for IA, and Efficient IAA (eIAA) for IAA [13], [21]–[25]. IA creates normal, convex Interval type-2 (IT2) FSs to represent a set of intervals, while the IAA is designed to model agreement amongst samples and/or sources using type-1 or type-2 FSs, which can be non-normal and non-convex. Notably, IAA-based methods are designed to avoid the need for data preprocessing, model assumptions (e.g. Gaussian distribution) or the elimination of outliers [22].

While valuable, the IAA does not natively provide an IV measure of central tendency for interval data, which is, however, desirable in practice. For example, to succinctly communicate data to stakeholders. In contrast, a method like the Interval Weighted Average (IWA) [26] was designed to generate an IV central tendency by directly calculating the weighted average of the endpoints of intervals. While straightforward and effective, IWA is not designed for, nor capable of, capturing the agreement among respondents as IAA does.

To bridge this gap, in this paper, we put forward the Interval Agreement Weighted Average (IAWA) as a means to succinctly generate a measure of central tendency for a set of intervals *while* capturing agreement. While this approach will have general utility, for example in fuzzy set defuzzification–which we are exploring in a future publication–this paper will focus on introducing the method and analyzing its behavior in detail through sensitivity analysis. Discussion will be limited to type-1 IAA sets in this paper due to space limitations.

The paper is structured as follows: Section II describes IV data and its features, provides background on interval methods, including IA, IAA, α-cut, and IWA. Section III describes the details of IAWA. Section IV demonstrates the sensitivity analysis of IWA and IAWA in respect to different features of interval sets. Section V presents the summary, conclusions, and directions for future research.

II. Background

In this section, we briefly discuss IV data sets and their features. Later, we review commonly used methods for analyzing interval data, including the IA, IAA, α-cut and IWA.

A. Interval-Valued Data Sets and their Features

A closed crisp interval $\bar{a}$, defined by its endpoints—the lower bound a^- and the upper bound a^+, where $a^- \leq a^+$—can be expressed as $\bar{a} = [a^-, a^+]$. Its width $a^w = |a^+ - a^-|$ represents the interval's range. Alternatively, $\bar{a}$ can be described by its center $a^c = \frac{a^+ + a^-}{2}$ and radius $a^r = \frac{a^w}{2}$, offering a comprehensive characterization [27].

Besides a crisp interval, there is also an uncertain interval, where each endpoint is itself a crisp interval [13]. However, collecting such complex data often poses real-world administrative challenges. Hence, this paper focuses on crisp intervals.

Despite the inherent complexity of IV data sets, they can be categorized and evaluated using a collection of fundamental features: the mean range, standard deviation of individual interval ranges, and the standard deviation of interval centers, indicating positional diversity. Ferson et al. [28] defined IV data sets with narrow ranges as 'skinny' and those with wider ranges as 'puffy'. These data sets vary in disjointedness and overlap, from completely separate to partially or fully overlapping intervals [18].

B. The Interval Approach

IA is a primary method for modeling IV data as FSs [22]. It collects interval endpoint data from subjects to represent linguistic uncertainty. The process involves two stages: data pre-processing to eliminate outliers and unsuitable data, followed by using the refined data to build IT2 FS models. The EIA [24] extends IA by refining data pre-processing for better control of interval widths, while the HMA [25] improves the process of constructing a FS by identifying the most common overlap of intervals and building a FS around it. However, IA, EIA, and HMA all involve data pre-processing, which may lead to significant data loss. Moreover, these methods presume unimodality in data, potentially resulting in critical information loss in certain real-world scenarios [22].

C. The Interval Agreement Approach

IAA is another widely used method for modeling IV data as FSs [13], aiming to encompass the full range of data without removing any intervals. Unlike IA and its extensions, IAA neither pre-processes data nor assumes unimodality [13]. It follows the least commitment principle to minimize information loss [29]. IAA captures both inter- and intra-participant uncertainty, representing these into various FS types (Type-1, IT2, and GT2 FSs). IAA's process involves up to two steps: firstly, creating a Type-1 FS (or IT2 for uncertain interval responses) to model either inter- or intra-participant uncertainties, and optionally, forming a GT2 FS based on step 1 to model both uncertainties.

To create a Type-1 FS X using IAA, for a given set of intervals $\bar{X} = \{\bar{x}_1, \ldots, \bar{x}_n\}$, where $\bar{x_n}$ represents an interval with left and right endpoints x_n^- and x_n^+, its membership function (denoted as $\mu_X(x')$) is described as follows [13]:

$$\mu_X(x') = \frac{(\sum_{i=1}^{n} \mu_{\bar{x}_i}(x'))}{n}$$
$$\mu_{\bar{x}_i}(x') = \begin{cases} 1 & \text{if } x_i^- \leq x' \leq x_i^+ \\ 0 & \text{else} \end{cases} \tag{1}$$

A type-1 FS created by IAA, shown in Fig.1's upper section, aggregates three IV responses from the lower section, with higher agreement in overlapping areas.

There is also eIAA [23], an extension of IAA that simplifies IAA's membership function for enhanced efficiency. The decision between IAA and eIAA should be based on the

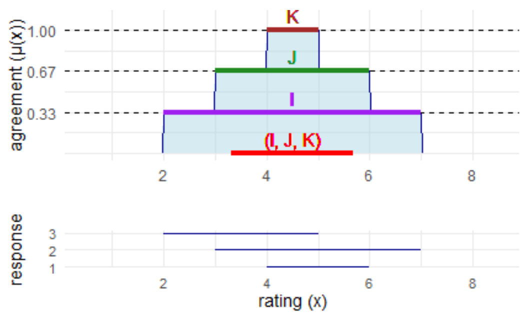


Fig. 1. Illustrative example of how the IAA (outline of the shaded area) captures the agreement between three intervals ([4,6], [3,7], and [2,5]), and how it is transformed into an interval by using α-cuts and IWA.

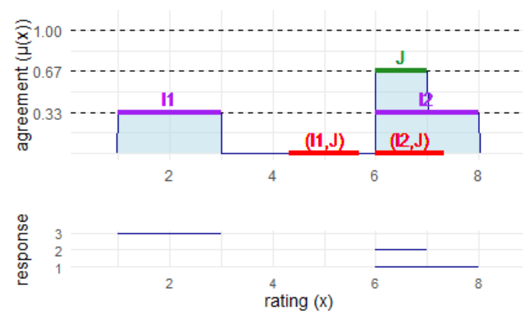


Fig. 2. Illustrative example of how the IAA (outline of the shaded area) captures the agreement between three intervals ([6,8], [6,7], and [1,3]), and how it is transformed into a discontinuous interval by using α-cuts and IWA.

application's priority—whether to retain all information or to optimize computation [22]. Given the scope of this paper, our focus will be on Type-1 FSs generated by IAA. This approach is selected for its ability to offer a direct and interpretable representation of interval data, which is particularly beneficial for real-world applications.

D. Alpha-cut

The α-cut decomposes a FS into a crisp set, represented by a single interval in convex FSs or multiple intervals for non-convex FSs, depending on the specified α level. This method is widely used to address the complexity and computational challenges in FSs [26], [30], [31]. The α-cut of a convex FS A at a level α is a crisp set (represented by a single interval) containing all the elements of the domain whose membership values in A are greater than or equal to α. In mathematical terms, the α-cut, A_α, is defined as

$$A_\alpha = \{x \in X \mid \mu_A(x) \geq \alpha\} \tag{2}$$

where $\mu_A(x)$ is the membership function of the FS A and X is the domain. The α-cuts at level α on the FS A provides a closed interval, expressed as $\overline{A_\alpha} = [A_\alpha^-, A_\alpha^+]$. Through decomposition with α-cuts at various levels, distinct intervals representing the FS are created. As illustrated in Fig.1, the FS from IAA decomposes into a set of intervals, I = $\overline{A_{0.33}}$, J = $\overline{A_{0.67}}$, and K = $\overline{A_1}$, using α-cuts with agreement levels of 0.33, 0.67, and 1.

When FS A is non-convex, applying α-cuts at certain levels can result in multiple intervals. For instance, Fig. 2 shows that an α-cut at 0.33 results in two intervals, $I1$ and $I2$. These intervals should be conceptualized as parts of a single discontinuous interval, processed sequentially and combined by conjunction (union) [31], [32]. This ensures the original FS's features at the α-cuts level are accurately represented.

As α-cut simplifies transforming fuzzy uncertainty into interval uncertainty, it has been used to aggregate multiple FSs into one by aggregating resulting intervals at various α levels [31], [33]–[39], as well as deconstructing and reconstructing FSs [40], [41]. α-cut is also widely used for fuzzy number ranking, comparison, and defuzzification — i.e., transforming a FS to a single crisp value, which can aid decision-making. However, while useful in producing crisp values from FSs, defuzzification can lose critical uncertainty information. While effort has been made for defuzzifying fuzzy numbers into intervals to maintain some aspects of uncertainty [42], [43], there has been very little attention given to the IV defuzzification for different types of general FSs [44](i.e., FSs not bound by the specific requirements of fuzzy numbers such as normality, convexity, and unimodality [33]).

E. Interval Weighted Average

The concept of averaging, a key measure of central tendency, is foundational in data analysis. When uncertainties in data points or weights (or both) can be better represented using IV data, the IWA offers an alternative to IV central tendency. IWA can be calculated by taking the weighted average of the lower and upper bounds separately [26].

For intervals $\bar{X} = \{[x_1^-, x_1^+], \ldots, [x_n^-, x_n^+]\}$ and associated crisp weights $W = \{w_1, w_2, \ldots, w_n\}$, the IWA is given by:

$$IWA = \bar{y} = [y^-, y^+] = \left[\frac{\sum_{i=1}^n w_i x_i^-}{\sum_{i=1}^n w_i}, \frac{\sum_{i=1}^n w_i x_i^+}{\sum_{i=1}^n w_i}\right] \tag{3}$$

When all the weights are crisp and equal, i.e., $w_1 = w_2 = \cdots = w_n$, the IWA simplifies to what we refer to as the 'Interval Average', denoted as IWA_e (for 'equal' weights) in this paper, to avoid any confusion with the Interval Approach (IA). In this case, since each weight w_i is the same, the formula for IWA_e becomes:

$$IWA = \bar{y} = [y^-, y^+] = \left[\frac{\sum_{i=1}^n x_i^-}{n}, \frac{\sum_{i=1}^n x_i^+}{n}\right] \tag{4}$$

Here, the IWA_e is essentially the arithmetic mean of the interval endpoints, reflecting an equal contribution from each interval in the set X. IWA can also compute cases where the weights are intervals [26]. However, as we only use crisp values as weights in this paper, this is beyond the scope.

III. The Interval Agreement Weighted Average

This section introduces IAWA as an extension of IAA. We will begin by outlining our objectives, followed by a detailed description of our proposed method, illustrated with synthetic examples. Lastly, we will explore potential extensions of IAWA to accommodate IT2 and GT2 FSs in future research.

A. Context

A set of IV data from a single participant across multiple surveys over time can show where the participant agrees with their own previously surveyed responses (intra-participant uncertainty). In contrast, IV data from multiple participants in a single survey can reveal where they align with the opinions of the rest of the group (inter-participant uncertainty). The established IAA method [13], utilizing type-1 FSs, effectively captures both uncertainties. The objective of the proposed method is to extend IAA by introducing a step that transforms a FS into an interval (IV defuzzification). This step enables IAWA to provide a measure of the central tendency for IV datasets and concurrently reflects agreement among intervals.

B. Method

For a given IV data set from a survey or repeated surveys, IAWA firstly models intra- or inter-participant uncertainty as a type-1 FS A using IAA [13]. Under IAA, the maximum possible number of distinct agreement levels n in A are decided by the number of surveys for intra-participant uncertainty or participants for inter-participant uncertainty. Secondly, this FS A is transformed into its representative interval through α-cuts and IWA. This involves the following sub-steps:

1) **A Priori α Levels**: Define the a priori α levels as $\{\alpha_1, \ldots, \alpha_n\}$, where $0 < \alpha_n \leq 1$ and n, the number of agreement levels, is predetermined. These levels, equally spaced between 0 and 1, are represented as $\{1/n, \ldots, 1\}$. For instance, with three participants, the α levels would be $1/3$, $2/3$, and 1, ensuring all opinions are considered.
2) **α-cuts**: Decompose the Fuzzy Set A into intervals $\bar{A} = \{\overline{A_{\alpha_1}}, \ldots, \overline{A_{\alpha_n}}\}$ using α-cuts. These intervals are defined as $\{[A^-_{\alpha_1}, A^+_{\alpha_1}], \ldots, [A^-_{\alpha_n}, A^+_{\alpha_n}]\}$.
3) **IWA**: Compute the weighted average of intervals $\bar{A}$ using IWA [26], as per Equation (3). Replace the interval endpoints x_i^- and x_i^+ in Equation (3) with $A^-_{\alpha_i}$ and $A^+_{\alpha_i}$ from $\bar{A}$. The weights w_i are substituted by their respective α values, α_i. As such, the proposed IAWA can be expressed as:

$$\begin{aligned} IAWA &= \bar{y} = [y^-, y^+] \\ &= \left[\frac{\sum_{i=1}^n \alpha_i A^-_{\alpha_i}}{\sum_{i=1}^n \alpha_i}, \frac{\sum_{i=1}^n \alpha_i A^+_{\alpha_i}}{\sum_{i=1}^n \alpha_i}\right] \end{aligned} \tag{5}$$

Furthermore, for a non-convex Fuzzy Set A, applying α-cuts as in step 2 can result in discontinuous intervals. Each continuous interval within it is processed sequentially in subsequent steps and rejoined via union. This method is intended to capture the non-convexity or discontinuity of the original data in the resulting interval. The total number of combinations, denoted as N, is determined by multiplying the number of continuous intervals at each level. For example, if there were three α levels of 2, 3, and 1 intervals respectively, N would be calculated as $2 \times 3 \times 1 = 6$ total combinations. The final results are the union of all these combinations:

$$IAWA = [y_1^-, y_1^+] \cup [y_2^-, y_2^+] \cup \ldots \cup [y_N^-, y_N^+] \tag{6}$$

C. Synthetic Examples

In this section, we simulate two use cases where a participant provides opinions on a topic in three repeated surveys. As there's only one participant, we use IAWA to model and reflect intra-participant agreement uncertainty over time. The first use case results in a convex type-1 FS, and the second results in a non-convex type-1 FS, both modeled using IAA. Given the three repeated measures, we choose three levels for α-cuts: 1/3 = 0.33, 2/3 = 0.67, and 3/3 = 1.

1) convex: The response intervals and IAA for case 1 are depicted in Fig.1. Applying α-cuts at the selected α levels decomposes the FS into three intervals: $I = \overline{A_{0.33}} = [2, 7]$, $J = \overline{A_{0.67}} = [3, 6]$, and $K = \overline{A_1} = [4, 5]$. In Fig.1, IAWA is denoted as (I, J, K) and is calculated as follows:

$$\begin{aligned} IAWA &= [y^-, y^+] \\ &= \frac{0.33}{0.33 + 0.67 + 1} \times [2, 7] \\ &+ \frac{0.67}{0.33 + 0.67 + 1} \times [3, 6] \\ &+ \frac{1}{0.33 + 0.67 + 1} \times [4, 5] \\ &= \left[\frac{0.33}{2} \times 2 + \frac{0.67}{2} \times 3 + \frac{1}{2} \times 4, \right. \\ &\quad \left. \frac{0.33}{2} \times 7 + \frac{0.67}{2} \times 6 + \frac{1}{2} \times 5\right] \\ &= [3.33, 5.67] \end{aligned}$$

2) non-convex: The response intervals and IAA for case 2 are depicted in Fig.2. Applying α-cuts at the selected α levels decomposes the FS into three intervals. At the 0.33 agreement level, due to non-convexity, there are $I1 = \overline{A_{0.33-I}} = [1, 3]$ and $I2 = \overline{A_{0.33-II}} = [6, 8]$; at the 0.67 level, $J = \overline{A_{0.67}} = [6, 7]$; and none for the 1 level. In Fig.2, IAWA is represented as $(I1, J) \cup (I2, J)$ and is calculated as follows:

$$\begin{aligned} IAWA &= [y_1^-, y_1^+] \cup [y_2^-, y_2^+] \\ &= \left(\frac{0.33}{0.33 + 0.67} \times [1, 3] + \frac{0.67}{0.33 + 0.67} \times [6, 7]\right) \cup \\ &\quad \left(\frac{0.33}{0.33 + 0.67} \times [6, 8] + \frac{0.67}{0.33 + 0.67} \times [6, 7]\right) \\ &= \left[\frac{0.33}{1} \times 1 + \frac{0.67}{1} \times 6, \frac{0.33}{1} \times 3 + \frac{0.67}{1} \times 7\right] \cup \\ &\quad \left[\frac{0.33}{1} \times 6 + \frac{0.67}{1} \times 6, \frac{0.33}{1} \times 8 + \frac{0.67}{1} \times 7\right] \\ &= [4.35, 5.68] \cup [6.00, 7.33] \end{aligned}$$

D. Potential Extension

It is noteworthy that IAA is capable of constructing IT2 and GT2 FSs from IV data responses. These FSs offer a more nuanced representation of participant uncertainties, but their application is frequently constrained by the perceived complexity of computation. The use of α-cuts to decompose FSs into intervals presents a solution by simplifying the computational process from handling FSs to computing intervals. Given that the proposed IAWA method specializes in interval analysis, extending IAWA to accommodate these FSs shows promise. For instance, within IAA, a GT2 FS, which captures both inter- and intra-participant uncertainties, is generated by combining multiple Type-1 FSs derived from various IV data sets. By applying IAWA to these Type-1 FSs, we can effectively transform multiple IV datasets into a single dataset. This dataset can then be processed by IAA, or alternatively by IAWA again, to reflect the combined uncertainties.

Due to the limited scope of this paper, we will cover the details of these extensions along with the IV defuzzification method for different types of FSs — not just those built using IAA — in a future publication.

IV. Demonstration and Analysis

This section presents a sensitivity analysis to evaluate and compare IWA and IAWA. Given the direct influence of IAA on IAWA, it is also visualized and discussed within this analysis. The primary comparison focuses on IWA and IAWA, as both methods offer interval-based measures for assessing the central tendency of IV sets. Following the sensitivity analysis, the cross-set results from IWA and IAWA are visualized and compared to highlight how IV central tendency enables comparisons across various IV datasets. For further exploration or application of these methods, the source code and synthetic data used in this analysis are available on GitHub: https://github.com/Carina-YuZhao/IAWA.git.

The sensitivity analysis follows three key features of IV data sets:

1) Increase in the mean range (size) of intervals (μ_{X^w});
2) Gradual increase in the standard deviation of ranges (σ_{X^w});
3) Gradual dispersion in the positioning of intervals, denoted by an increased standard deviation in the centers of intervals (σ_{X^c}).

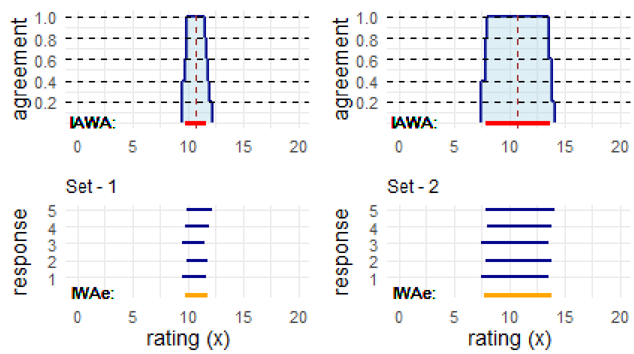
For this study, ten synthetic interval data sets, adapted from [27], were utilized to simulate a scenario in which five participants provide their opinions on a specific topic across ten different surveys over time. Each survey was modeled using IAWA and IWA with equal weight, denoted as IWA_e, to ensure all participants' opinions are treated equally. These data sets, within the range of [0,20], share the same mean of center (μ_{X^c}) for their intervals. They are categorized into two main types: 'skinny' and 'puffy', based on their mean range of intervals (feature 1). Set 1 and Set 2 serve as the foundational 'skinny' and 'puffy' data sets, respectively. For the 'skinny' type, Sets 1, 3, and 4 were used to analyze the second feature, while Sets 1, 7, and 8 were employed for the third feature. Conversely, for the 'puffy' type, Sets 2, 5, and 6 were utilized for the second feature analysis, and Sets 2, 9, and 10 for the third feature. This structure facilitates a comprehensive comparison and analysis across different data sets and features. All data sets and aggregated results are visualized.

In each figure, the lower section displays the intervals for the respective data set and the result from IWA_e. The upper section depicts the result from IAWA and the FS modeled by IAA, with a vertical dotted line indicating its centroid (i.e., position), dotted horizontal lines across the upper section mark the chosen agreement levels for the α-cuts, set from 0.2 to 1.0 in increments of 0.2 as there are 5 participants.

In this section, the term 'position,' refers to the center of the interval for IWA_e and IAWA, and centroid for IAA.

A. Increase in the mean range (size) of intervals

For this analysis, we utilize Set-1 ('skinny') and Set-2 ('puffy'), as illustrated in Fig.3. In both sets, intervals are overlapping, with identical central positions and the same standard deviations within their ranges, to control for their potential impact. The primary distinction between them is in their mean range of intervals; notably, Set-2 ('puffy') has a larger mean range than Set-1 ('skinny'). For all three methods, the positions of the results remain largely the same, whereas the range of the results from all methods widens as the width of the original data set increases from 'skinny' to 'puffy'. Furthermore, the shape of the IAA remains consistent, displaying a convex, nearly uniform distribution for both sets.



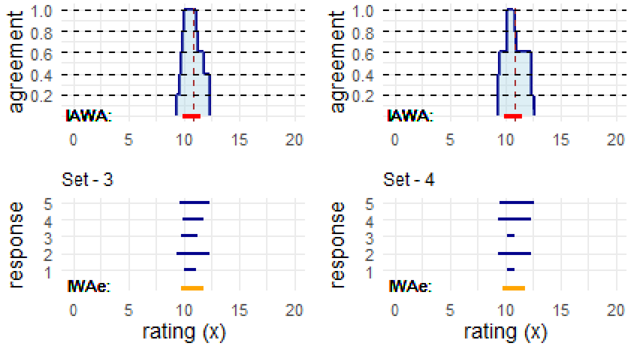
Set 1	Set 2
$\mu_{X^c} = 10.75$ $\mu_{X^w} = 2.06$	$\mu_{X^c} = 10.75$ $\mu_{X^w} = 6$
$\sigma_{X^c} = 0.22$ $\sigma_{X^w} = 0.16$	$\sigma_{X^c} = 0.22$ $\sigma_{X^w} = 0.16$
IAWA: [9.8, 11.67]	IAWA: [7.84, 13.65]
IWA_e : [9.72, 11.78]	IWA_e : [7.75, 13.75]

Fig. 3. Visualization of 'skinny' and 'puffy' IV data sets, illustrating the impact of different mean range of intervals on aggregation result.

B. Gradual increase in standard deviation of interval ranges

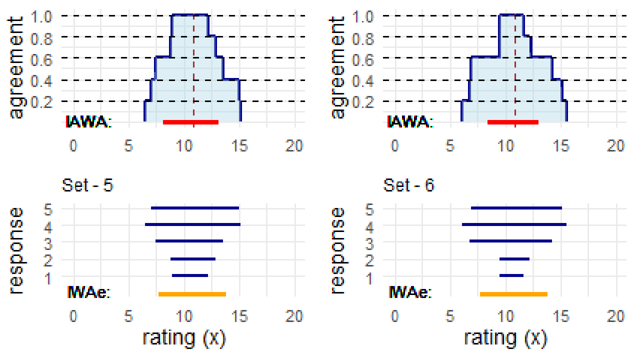
In Fig. 4, we explore two additional 'skinny' synthetic cases, Set-3 and Set-4. These were created by gradually increasing the standard deviation of the ranges from the baseline established by Set-1, as shown in Fig. 3. The aim here is to analyze how an increase in standard deviation influences the aggregation methods when the mean range of intervals and the centers of the intervals remain unchanged.

Likewise, in Fig. 5 for the 'puffy' synthetic scenarios, we assess two further sets, Set-5 and Set-6. These sets are derived from Set-2, also initially presented in Fig. 3, by similarly increasing the standard deviation of the ranges. This allows us to evaluate the impact of greater standard deviation within 'puffy' sets on our aggregation methods.



Set 3	Set 4
$\mu_{X^c} = 10.75$ $\mu_{X^w} = 2.06$	$\mu_{X^c} = 10.75$ $\mu_{X^w} = 2.06$
$\sigma_{X^c} = 0.22$ $\sigma_{X^w} = 0.79$	$\sigma_{X^c} = 0.22$ $\sigma_{X^w} = 1.31$
IAWA: [9.83, 11.54]	IAWA: [9.9, 11.45]
IWA_e : [9.72, 11.78]	IWA_e : [9.72, 11.78]

Fig. 4. Visualization of ‘skinny’ IV data sets, illustrating the impact of an increase in standard deviation of interval ranges on the aggregation result.



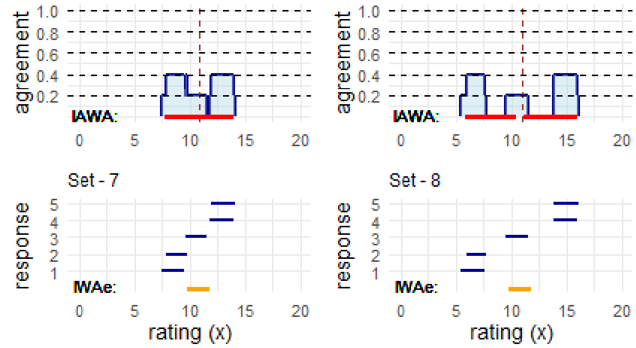
Set 5	Set 6
$\mu_{X^c} = 10.75$ $\mu_{X^w} = 6$	$\mu_{X^c} = 10.75$ $\mu_{X^w} = 6$
$\sigma_{X^c} = 0.22$ $\sigma_{X^w} = 2.38$	$\sigma_{X^c} = 0.22$ $\sigma_{X^w} = 3.28$
IAWA: [8.19, 13.2]	IAWA: [8.37, 13.04]
IWA_e : [7.75, 13.75]	IWA_e : [7.75, 13.75]

Fig. 5. Visualization of ‘puffy’ IV data sets, illustrating the impact of an increase in standard deviation of interval ranges on the aggregation result.

The results suggest that for both ‘skinny’ and ‘puffy’ sets, an increase in the standard deviation of ranges does not influence the IWA_e outcomes, with no change at all in both position and range. For IAA, there is no substantial change in its position with an increase in the standard deviation of the ranges. However, the shape shifts noticeably; it becomes sharper and more triangular as the standard deviation increases, leading to a smaller range at higher agreement levels and a wider range at lower levels. IAWA also shows no substantial change in its position, while its range gradually decreases when the IAA shifts to a more triangular form, with a pointier peak.

C. Gradual dispersion of interval centers

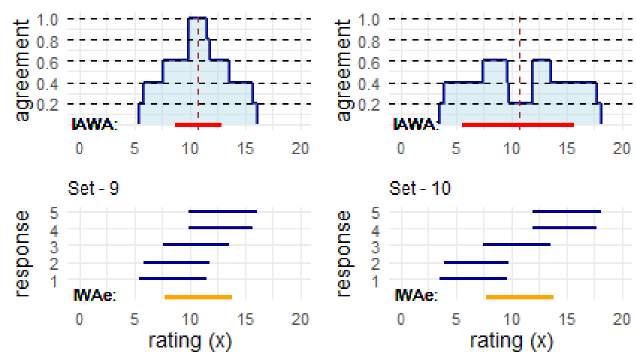
In Fig. 6, we further introduce two new synthetic cases, Set-7 and Set-8. These sets build upon the original ‘skinny’ set, Set-1 (introduced in Fig. 3). Set-7 is designed to be slightly scattered, while Set-8 is more extensively scattered. To create these sets, we adjust the centers of the intervals, effectively increasing their spread (or standard deviation) while keeping their range and the standard deviation of their range constant. This approach allows us to explore the impact of the relative positioning of intervals (whether scattered or densely placed) on our aggregation methods.



Set 7	Set 8
$\mu_{X^c} = 10.75$ $\mu_{X^w} = 2.06$	$\mu_{X^c} = 10.75$ $\mu_{X^w} = 2.06$
$\sigma_{X^c} = 2.13$ $\sigma_{X^w} = 0.16$	$\sigma_{X^c} = 4.12$ $\sigma_{X^w} = 0.16$
IAWA: [7.77, 14]	IAWA: $[5.79, 10.44] \cup [11.09, 15.99]$
IWA_e : [9.72, 11.78]	IWA_e : [9.72, 11.78]

Fig. 6. Visualization of ‘skinny’ IV data sets, illustrating the impact of dispersion of interval centers on the aggregation result.

Similarly, for the ‘puffy’ set analysis, we developed Set-9 and Set-10, as shown in Fig. 7, as counterparts to the ‘puffy’ Set-2. To ensure comparability between ‘skinny’ and ‘puffy’ sets, Sets-7 and Set-9 have identical interval centers, as do Sets-8 and Set-10. This setup enables us to examine whether the range of intervals affects the aggregation methods when the relative positioning of intervals becomes more scattered.



Set 9	Set 10
$\mu_{X^c} = 10.75$ $\mu_{X^w} = 6$	$\mu_{X^c} = 10.75$ $\mu_{X^w} = 6$
$\sigma_{X^c} = 2.13$ $\sigma_{X^w} = 0.16$	$\sigma_{X^c} = 4.12$ $\sigma_{X^w} = 0.16$
IAWA: [8.62, 12.88]	IAWA: [5.63, 15.68]
IWA_e : [7.75, 13.75]	IWA_e : [7.75, 13.75]

Fig. 7. Visualization of ‘puffy’ IV data sets, illustrating the impact of dispersion of interval centers on the aggregation result.

The results suggest that for both ‘skinny’ and ‘puffy’ sets, an increase in the standard deviation of center positions does not influence the IWA_e outcomes, with no change at all in both position and range. For IAA, with both ‘skinny’ and ‘puffy’ sets, there is no substantial change in position when the standard deviation of the center increases. However, for the ‘skinny’ sets, the shape changes significantly; the FS becomes non-convex and discontinuous, developing more peaks with lower heights (agreement level) as the sets become more scattered. For the 'puffy' sets, the initial change mirrors that in Fig.5, transitioning from a more rectangular to a triangular shape. As the data set becomes more scattered, the FS develops into a non-convex, continuous form, with multiple peaks exhibiting lower heights (agreement level). For IAWA, with both 'skinny' and 'puffy' sets, there is no substantial change in its position. In the 'skinny' set, IAWA's range increases significantly when multiple peaks appear in IAA. Additionally, IAWA produces a disjoint result in response to further multiple peaks with lower levels of agreement. For the 'puffy' set, its range initially decreases as the IAA shape shifts from rectangular to triangular. Subsequently, as multiple peaks emerge in IAA, the range of IAWA increases significantly.

D. Cross-sets comparison and discussion

A visualization of the results from both methods across ten sets is displayed in Fig.8. This section will discuss how IWA_e reflects changing features and represents central tendencies in these IV sets, then use this as a basis to compare how IAWA's results differ.

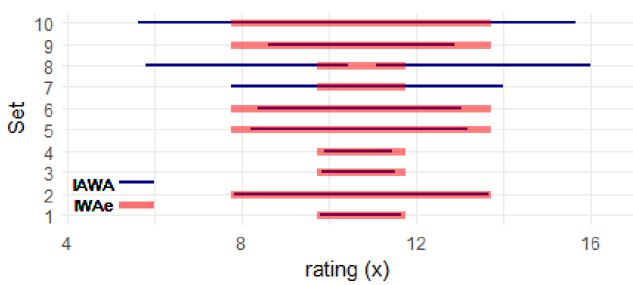


Fig. 8. IWA_e (red) and IAWA (blue) results for all ten synthetic data sets.

For IWA_e, Set 1, 3, 4, 7, and 8 have identical small ranges, from all five 'skinny' sets, while Set 2, 5, 6, 9, and 10 have the same, wider range, from all five 'puffy' sets. The trend is simple and straightforward, suggesting that the only feature of the IV dataset influencing the result of IWAe is the mean range (size) of intervals. This is as expected, since IWAe is essentially the equivalent of the mean for IV data.

In comparison, for IAWA, the results for Sets 1 - 6, and Set 9 are very close to, and slightly narrower than, those of IWA_e. Considering that the IAA FSs generated from these sets are all normal, convex, and unimodal, which indicates a high level of agreement among participants, this suggests that when IAWA produces a range similar to or slightly narrower than IWA_e, it reflects a high level of participant agreement. While seemingly trivial, the slightly narrower range of IAWA compared to IWA_e suggests a more specific or certain agreement among participants, as represented by a narrower peak in the IAA.

Furthermore, for Sets 7, 8, and 10, the results of IAWA are noticeably wider than those of IWA_e. The IAA FSs generated from these sets are all abnormal, nonconvex, and multimodal, indicating a lower level of agreement among participants (i.e., no single value has been unanimously selected or satisfies everyone). In particular, the result of IAWA for Set 8 is discontinuous. This is by design, as this discontinuity arises from the α-cuts of non-convex FS. Such discontinuity can therefore reflect the non-convex nature or the diverse agreements among participants as represented in Set 8 (i.e., participants are divided into three different agreements, represented by three peaks). It is worth noting that the discontinuity in IAWA results must come from the non-convexity of the IAA FS, but not the other way around. This is because the multiple intervals generated by a non-convex FS can result in a continuous interval through union, as demonstrated in Set 10.

Therefore, while IWA_e provides a straightforward measure akin to an average of the crisp data, reflecting the central tendencies with simplicity, IAWA emerges as a more intricate method, capable of further indicating details regarding participant agreement within IV set. In addition, for both methods, by compiling results from all ten sets, we can observe the changes in responses of all participants as a whole over repeated measures. Since this is effectively an new IV set, it can be further analysed using IWA, IAA, or IAWA as required.

V. Summary, Conclusion and Future Work

This paper introduced an extension of the IAA method called IAWA, which reflects the IV central tendency and agreement of IV data set. We conducted a sensitivity analysis evaluating IAWA and compared it to IWA. Our focus was on how these methods respond to various features of IV data sets. Specifically, we explored the effects of (1) wider mean range (i.e., from 'skinny' to 'puffy' data set), (2) higher standard deviation of ranges, and (3) higher standard deviation of centers on the aggregated results.

Our findings demonstrate distinct impacts of IV data set features on these two methods. IWA (with equal weight) directly mirrors the mean position and range of the interval set, unaffected by the standard deviation of ranges or centers. It effectively summarizes intra-participant uncertainty, i.e., reflecting 'skinny' or 'puffy' data sets and the uncertainty or flexibility within individual responses. IWA's utility is evident in real-world scenarios, where it aids in comparing intra-participant uncertainty across diverse data sets. Conversely, IAWA, while lacking the detailed distribution found in IAA, largely retains IAA's capability to represent either intra- or inter- participant agreement among responses. It is particularly sensitive to the width and height (i.e., the agreement level) of peaks in IAA. This sensitivity means that focused agreements lead to shorter ranges in IAWA, while dispersed agreements result in wider or even discontinuous ranges.

In conclusion, as an extension of IAA, IAWA provides a valuable measure of IV central tendency to the methods available for analyzing, comparing, and summarising IV data. These methods are particularly relevant for capturing information from real-world, which tends to be more complex than the scenarios presented in our examples. The next step in our research involves applying these methods to the analysis of real-world data, with a specific focus on enhancing service quality measurement. Given that IV data is specifically designed for capturing uncertainty and imprecise human subjective judgment, we believe it is well-suited to address the inherent challenges in measuring the intangible and heterogeneous nature of services. More broadly, considering IAA can also model intervals in IT2 and GT2 FSs, we expect to explore the proposed approach in a more general defuzzification context while also conducting broader experimental evaluation in future work.

References

[1] R. Likert, “A technique for the measurement of attitudes. ,” *Archives of psychology*, 1932.

[2] C. B. Aitken, “Measurement of feelings using visual analogue scales,” *Proceedings of the royal society of medicine*, vol. 62(10), 1969.

[3] Z. Ellerby, C. Wagner, and S. B. Broomell, “Capturing richer information: On establishing the validity of an interval-valued survey response mode,” *Behavior Research Methods*, 2021.

[4] N. Seth, S. G. Deshmukh, and P. Vrat, “Service quality models: A review,” *International Journal of Quality and Reliability Management*, vol. 22, pp. 913–949, 2005.

[5] Z. S. Chen, X. L. Liu, K. S. Chin, W. Pedrycz, K. L. Tsui, and M. J. Skibniewski, “Online-review analysis based large-scale group decision-making for determining passenger demands and evaluating passenger satisfaction: Case study of high-speed rail system in china,” *Information Fusion*, vol. 69, pp. 22–39, 5 2021.

[6] B. Hesketh, R. Pryor, M. Gleltzmat, and T. Hesket, “Practical applications and psychometric evaluation of a computerized fuzzy graphic rating scale,” *Advances in Psychology*, 1988.

[7] T. Hesketh, R. Pryor, and B. Hesketh, “An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences,” *International Journal of Man-Machine Studies*, vol. 29, pp. 21–35, 1988.

[8] M. A. Lubiano, M. Montenegro, S. Pérez-Fernández, and M. Ángeles Gil, “Analyzing the influence of the rating scale for items in a questionnaire on cronbach coefficient alpha,” *Information and Data Sciences: A Tribute to Leandro Pardo*, 2023.

[9] S.-H. Tsaur, T.-Y. Chang, and C.-H. Yen, “The evaluation of airline service quality by fuzzy mcdm,” *Tourism Management*, vol. 23, pp. 107–115, 2002.

[10] Y. C. Hu and P. C. Liao, “Finding critical criteria of evaluating electronic service quality of internet banking using fuzzy multiple-criteria decision making,” *Applied Soft Computing Journal*, vol. 11, pp. 3764–3770, 6 2011.

[11] A. Awasthi, S. S. Chauhan, H. Omrani, and A. Panahi, “A hybrid approach based on servqual and fuzzy topsis for evaluating transportation service quality,” *Computers and Industrial Engineering*, vol. 61, pp. 637–646, 10 2011.

[12] A. Mardani, A. Jusoh, E. K. Zavadskas, Z. Khalifah, and K. M. Nor, “Application of multiple-criteria decision-making techniques and approaches to evaluating of service quality: a systematic review of the literature,” *Journal of Business Economics and Management*, vol. 16, pp. 1034–1068, 9 2015.

[13] C. Wagner, S. Miller, J. M. Garibaldi, D. T. Anderson, and T. C. Havens, “From interval-valued data to general type-2 fuzzy sets,” *IEEE Transactions on Fuzzy Systems*, vol. 23, pp. 248–269, 4 2015.

[14] Z. Ellerby, J. Mcculloch, J. Young, and C. Wagner, “Decsys-discrete and ellipse-based response capture system.” IEEE, 2019, pp. 1–6.

[15] Z. Ellerby and C. Wagner, “Do people prefer to give interval-valued or point estimates and why?” 2021, pp. 1–6.

[16] Z. Ellerby, O. Miles, J. McCulloch, and C. Wagner, “Insights from interval-valued ratings of consumer products-a decsys appraisal,” 2020.

[17] J. Navarro, C. Wagner, U. Aickelin, L. Green, and R. Ashford, “Exploring differences in interpretation of words essential in medical expert-patient communication,” 7 2016.

[18] S. Kabir, C. Wagner, and Z. Ellerby, “Towards handling uncertainty-at-source in ai – a review and next steps for interval regression,” *IEEE Transactions on Artificial Intelligence*, 4 2023.

[19] K. J. Wallace, C. Wagner, and M. J. Smith, “Eliciting human values for conservation planning and decisions: A global issue,” *Journal of Environmental Management*, vol. 170, pp. 160–168, 4 2016.

[20] K. J. Wallace, C. Wagner, D. J. Pannell, M. K. Kim, and A. A. Rogers, “Tackling communication and analytical problems in environmental planning: Expert assessment of key definitions and their relationships,” *Journal of Environmental Management*, vol. 317, 9 2022.

[21] F. Liu and J. M. Mendel, “Encoding words into interval type-2 fuzzy sets using an interval approach,” *IEEE Transactions on Fuzzy Systems*, vol. 16, pp. 1503–1521, 2008.

[22] J. McCulloch, Z. Ellerby, and C. Wagner, “On comparing and selecting approaches to model interval-valued data as fuzzy sets.” IEEE, 2019.

[23] T. C. Havens, C. Wagner, and D. T. Anderson, “Efficient modeling and representation of agreement in interval-valued data,” *Institute of Electrical and Electronics Engineers*, 2017.

[24] D. Wu, J. M. Mendel, and S. Coupland, “Enhanced interval approach for encoding words into interval type-2 fuzzy sets and its convergence analysis,” *IEEE Transactions on Fuzzy Systems*, vol. 20, pp. 499–513, 2012.

[25] M. Hao and J. M. Mendel, “Encoding words into normal interval type-2 fuzzy sets: Hm approach,” *IEEE Transactions on Fuzzy Systems*, vol. 24, pp. 865–879, 8 2016.

[26] F. Liu and J. M. Mendel, “Aggregation using the fuzzy weighted average as computed by the karnik-mendel algorithms,” *IEEE Transactions on Fuzzy Systems*, vol. 16, pp. 1–12, 2 2008.

[27] S. Kabir and C. Wagner, “Interval-valued regression - sensitivity to data set features.” Institute of Electrical and Electronics Engineers Inc., 7 2021.

[28] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, and L. Ginzburg, “Experimental uncertainty estimation and statistics for data having interval uncertainty,” 2007.

[29] X. Hu and P. Mordohai, “Least commitment, viewpoint-based, multi-view stereo,” *Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, 2012.

[30] D. Dubois and J. Fortin, “A generalized vertex method for computing with fuzzy intervals hêî ene fargier,” 2004.

[31] W. M. Dong and F. S. Wong, “Fuzzy weighted averages and implementation of the extension principle,” *Fuzzy Sets and Systems*, vol. 21, pp. 183–199, 1987.

[32] C. Wagner, D. T. Anderson, and T. C. Havens, “Generalization of the fuzzy integral for discontinuous interval-and non-convex interval fuzzy set-valued inputs,” 2013.

[33] A. Pourabdollah, J. M. Mendel, and R. I. John, “Alpha-cut representation used for defuzzification in rule-based systems,” *Fuzzy Sets and Systems*, vol. 399, pp. 110–132, 11 2020.

[34] P. Zhou, J. Luo, F. Cheng, S. Yüksel, and H. Dinçer, “Analysis of risk priorities for renewable energy investment projects using a hybrid it2 hesitant fuzzy decision-making approach with alpha cuts,” *Energy*, vol. 224, 6 2021.

[35] D. Gadjiev, A. Rustanov, and I. Kochetkov, “The advanced defuzzification methods of the convex a - cut fuzzy sets,” vol. 376. EDP Sciences, 3 2023.

[36] S. Mabuchi, “An approach to the comparison of fuzzy subsets with an a-cut dependent index,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 18, 1988.

[37] C.-B. Chen and C. M. Klein, “A simple approach to ranking a group of aggregated fuzzy utilities,” *IEEE transactions on systems, man, and cybernetics, Part B (Cybernetics)*, vol. 27, 1997.

[38] Y. Y. Yang, X. W. Liu, and F. Liu, “Trapezoidal interval type-2 fuzzy topsis using alpha-cuts,” *International Journal of Fuzzy Systems*, vol. 22, pp. 293–309, 2 2020.

[39] S. M. Zhou, F. Chiclana, R. I. John, and J. M. Garibaldi, “Alpha-level aggregation: A practical approach to type-1 owa operation for aggregating uncertain information with applications to breast cancer treatments,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, pp. 1455–1468, 2011.

[40] D. Pekaslan and C. Wagner, “Alpha-cut based compositional representation of fuzzy sets and exploration of associated fuzzy set regression.” Institute of Electrical and Electronics Engineers Inc., 2022.

[41] L. I. Perlovsky, R. Deming, and R. Ilin, *Emotional cognitive neural algorithms with engineering applications : dynamic logic : from vague to crisp*. Springer, 2011.

[42] O. M. Poleshuk and E. G. Komarov, “New defuzzification method based on weighted intervals,” 2008.

[43] S. Heilpern, “The expected value of a fuzzy number,” *Fuzzy Sets and Systems*, vol. 47, pp. 81–86, 1992.

[44] L. Anzilli, G. Facchinetti, and G. Mastroleo, “A parametric approach to evaluate fuzzy quantities,” *Fuzzy Sets and Systems*, vol. 250, pp. 110–133, 9 2014.